

Characterisation of the Glycoside Hydrolase Domain of a Novel Bi-functional Metagenomic Clone for use in the Biofuel Production Industry

Craig C. Swanepoel



Thesis presented in partial fulfillment of the requirements for the degree of Master of Science in Plant Biotechnology in the Faculty of Natural Sciences at Stellenbosch University.

Supervisor: Dr. Shaun Peters
Co-supervisors: Dr. Bianke Loedolff, Prof. Jens Kossmann

December 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2017

Abstract

The current practice of 1st generation biofuel production is marred by several hurdles, namely concerns over food security and the moral dilemma created by using edible feedstocks as a biofuel production source. Second generation biofuel production methods stand to take the forefront and address the world's need for a renewable liquid fuel source without affecting food security and adding value to the abundant lignocellulosic biomass available worldwide. However, in order to achieve this, 2nd generation biofuel production methods need to become more efficient at liberating fermentable glucose from lignocellulosic biomass. Metagenomic sampling and novel enzyme discovery are the most promising sources of finding new, more efficient enzymes from unculturable microorganisms that can degrade lignocellulosic biomass into fermentable glucose more ably than current enzymes used in industry. Industrial enzymes within the biofuels scope are required to be thermotolerant, pH tolerant, resistant to product inhibition and resistant to denaturation by industrial solvents. A novel, bi-functional, β -glucosidase and β -galactosidase enzyme, termed Clone 3L, was identified via a metagenomic sampling approach. This study outlines the characterization of clone 3L, including the kinetics of the enzyme and an assessment of the suitability of the enzyme within an industrial, 2nd generation biofuel production pipeline. Clone 3L was found to be a promising candidate for use in 2nd generation biofuel production schemes, with the enzyme exhibiting high activity and affinity for its cellobiose substrate and a high degree of tolerance to various solvents, glucose inhibition and a high activity across a wide pH range.

Acknowledgements

Firstly I would like to thank Prof. Jens Kossmann for the opportunity to complete both my honours and my masters at his institute. It has been a truly memorable three years, and the skills that I have learned at the IPB will serve me well in the years to come.

Next, I would like to thank everyone in the IPB family, staff & students, for the laughs, advice, assorted shoulders to cry on, and support.

I would like to thank my family for their unwavering support throughout my postgraduate career, their love and care made the whole ordeal much more bearable.

To my partner in science and in crime, I wouldn't have made it without you.

Finally, and most importantly, I would like to extend my most sincere gratitude to Dr. Bianke Loedolff & Dr. Shaun Peters. For unparalleled support, knowledge, advice, and understanding. I owe it all to both of you. Thank you from the bottom of my heart.

Table of Contents

Abstract	3
Acknowledgements	4
List of Abbreviations	6
Introduction	7
1.1 General Introduction	7
1.2 The 4 Main Sectors of the Biofuel Industry.....	7
1.3 The Current Scope of the South African Biofuels Industry	10
1.4 The Quest to Create a Feasible 2 nd Generation Biofuel Platform	12
Objectives of this Study	14
2. Materials & Methods	16
Bacterial Strains and Vectors used within this Study:	16
2.1 Metagenomic Library Creation and Identification of Clone 3L.....	17
2.2 Bioinformatic Analyses of the Sequence of Clone 3L and β -glucosidase Activity Screens.....	17
2.3 Heterologous Expression of Clone 3L in the <i>E. coli</i> (DH5 α Δ LacZ) Mutant Background	17
2.4 Determining Protein Concentration.....	18
2.5 Enzyme Activity Determination.....	18
2.5.1 Determining the Optimal pH for Enzyme Activity.....	18
2.5.2 Determining the Optimal Temperature for Enzyme Activity.....	18
2.5.3 Determining the Enzyme Kinetics.....	18
2.5.4 Examination of Enzymatic Co-Factors	19
2.5.5 Determining the Glucose Tolerance of the Enzyme	19
2.5.6 Determining Enzyme Stability in Solvents/Surfactants	19
2.5.7 Determining the Bond Specificity of the Enzyme.....	19
2.6 Liberated Glucose Quantification	19
3. Results	21
3.1 Clone 3L Consistently Displays both β -galactosidase and β -glucosidase Activities on Plate-Based Functional Screens.....	21
3.2 <i>In Silico</i> Analyses of Clone 3L, Initially Identified from Functional Screens of Metagenomic Clones as Containing a Putative β -galactosidase Open Reading Frame.....	24
3.3 Enzymatic Characterization of the β -glucosidase Activity of Clone 3L.....	24
3.4 Evaluation of the Industrial Resilience of 3L.....	27
4. Discussion	31
Conclusion & Future Work	38
5. References	39
7. Appendix	43
7.1 Vector Map of the pRSETa::3L Construct	43

List of Abbreviations

β -gal: β -galactosidase

β -gluc: β -glucosidase

BSA: Bovine serum albumin

$^{\circ}\text{C}$: Degrees Celsius

CAZy database: Carbohydrate active enzymes database

ddH₂O: Distilled deionized water

DTT: Dithiothreitol

E. coli: *Escherichia coli*

EDTA: Ethylenediaminetetraacetic acid

g: Gram

Gal: Galactose

GH: Glycosyl Hydrolase

Glc: Glucose

HEPES: 4-(2-hydroxymethyl)-1-piperazineethanesulfonic acid

IPTG: Isopropyl β -D-1-thiogalactopyranoside

Kb: Kilobase

K_m : Michaelis Menten constant

LacZ: *Escherichia coli* β -galactosidase

LB: Luria broth

M: Molar

MES: 2-(N-morpholino)-ethanesulfonic acid

ml: Millilitre

mM: Millimolar

Mol: Molecules

RPM: Revolutions per minute

SDM: Site-directed mutagenesis

SDS: Sodium dodecyl sulfate

μg : Microgram

μl : Microliter

v/v: Volume/Volume

V_{max} : Maximum rate

w/v: Weight/Volume

X-gal: 5-bromo-4-chloro-3-indoyl- β -galactopyranoside

Introduction

1.1 General Introduction

The modern world has an extreme dependency on fossil fuels for a multitude of applications, from industry to automobiles (Lü *et al.*, 2011; Aro, 2015; Alaswad *et al.*, 2015). The earth hosts a finite supply of fossil fuels, thus, the key to a sustainable future lies in balancing fuel production with consumption. Currently, the human population exceeds 7 billion and is increasing rapidly; with this in mind it is clear that an extreme deficit in production and consumption of fossil fuels is imminent (Lü *et al.*, 2011). Various industries and the automotive realm have longstanding infrastructures that are intrinsically linked to fossil fuels, making a paradigm shift to an entirely new source of energy unfeasible and costly to implement. Humankind is dependent on fossil fuels for survival, but our global supplies are being exhausted rapidly, thus sustainable alternatives need to be created and implemented in such a manner so as to integrate into existing infrastructure with biofuels being the most promising source of alternative, sustainable energy (Aro, 2015)

Current trends in fossil fuel consumption indicate that the planet's reserves could be depleted by 2050. Our long-standing dependence on fossil fuels as a primary energy source has resulted in a depletion of a large portion of our fossil reserves in under a century. It is paramount that alternative fuel/energy sources be developed into the future. However, alternative fuel sources need to perform similarly to petroleum or diesel and must also be renewable in order to achieve a sustainable means of production to meet the ever-increasing global demand for fossil fuels (Sindhu *et al.*, 2016). One of the proposed solutions to the pending energy crisis is the development and use of biofuels.

Biofuels is a collective term used to refer to bioethanol, biodiesel and biogas. These are combustible fuel sources that are created sustainably by utilizing plant biomass as the input for production. The core mechanism in biofuel production hinges on the liberation of fermentable glucose from plant biomass. During subsequent fermentative processes, this glucose is converted into bioethanol, biodiesel and methane (Amnuaycheewa *et al.*, 2016). Since the year 2000, biofuels production has steadily increased across the globe, with forerunners such as Brazil and the USA setting the precedent for the production of sustainable fuel sources. As of 2015, global bioethanol production alone totaled 115 billion liters (Singh and Srivastava, 2016).

1.2 The 4 Main Sectors of the Biofuel Industry

Currently, there are 4 main biofuel production sectors globally:

- 1st generation biofuels – the production of biofuels utilizing glucose-rich plant storage organs (e.g. The production of bioethanol from the fermentation of starch from maize, sorghum, etc.) (Maity *et al.*, 2014),

- 2nd generation biofuels – the production of biofuels utilizing lignocellulosic biomass (e.g. the saccharification and fermentation of sugarcane bagasse into bioethanol (Sims *et al.*, 2010),
- 3rd generation biofuels – the production of biofuels performed by algae (mostly biogas) by means of anaerobic digestion known as anoxygenic photosynthesis (Alaswad *et al.*, 2015),
- 4th generation biofuels – the production of biofuels utilizing genetically modified algal material that has been engineered to produce biogas, biodiesel, or bioethanol. (Lü *et al.*, 2011).

Each biofuel production strategy has its own unique set of caveats and advantages. The 1st generation biofuels are currently the most commonly produced and involve the harvesting of edible, glucose-rich plant material from crops, and fermenting these to produce biofuels. This approach yields a large concentration of biofuels from the source material, with a lower input and research cost when compared to other biofuel production methods. The practice is, however, rife with controversy as certain groups continually lobby against the use of edible plant material to produce fuel (Salim *et al.*, 2015). This brings the issue of food security into the realm of biofuel production and has given rise to the dilemma that has been termed the “food vs. fuel” argument (Mohr and Raman, 2013). It stands to reason that various groups and governing bodies would be reluctant to support 1st generation biofuel initiatives in order to mitigate the pressure that increased food production would result in, both for agricultural sectors and nations experiencing severe food insecurity (Maity *et al.*, 2014). Furthermore, despite the high glucose yield from edible plant material, it is believed that certain characteristics of these plant materials lead to suboptimal biofuel production. Recent studies have reported that the high phytate content in edible plant matter results in lower yields of bioethanol (Chen *et al.*, 2015), further emphasizing the need for a global shift away from 1st gen biofuel production methods.

The 2nd generation biofuel production strategies are considered the most viable from a sustainability standpoint. The crux of a 2nd generation biofuel production platform is to utilize what would be normally considered as “plant waste” to generate biofuels. The most abundant natural biopolymer is cellulose (Tanimura *et al.*, 2016), and lignocellulosic biomass remains the most abundant but underutilized natural resource on a global scale (Sindhu *et al.*, 2016). Fermentable sugars can be liberated from the cellulose and hemicellulose regions within the plant cell wall matrix (Fig. 1, Chapter 1), and these regions account for 60% of the composition of lignocellulosic biomass, with the remaining 15-25% consisting of lignin (Ali *et al.*, 2016). Within the sugarcane, rice, and maize production industries alone, vast quantities of sugarcane bagasse, corn stover, and rice straw are left behind after farmers have recovered the sugar, maize, and rice respectively. This leftover plant material is considered useless and is usually burnt or left to degrade *in situ* in order to replenish nutrients in farmland soils or combusted as an energy source (Singh and Srivastava, 2016).

2nd generation biofuels remain an unattractive means of biofuel production due to the recalcitrant nature of the structural materials that make up the plant cell wall. The plant cell wall consists of 3 main layers,

cellulose, hemicellulose, and lignin. These layers are tightly interwoven to form the plant cell wall, with the cellulose and hemicellulose layers being the most attractive components of the plant cell wall for biofuel applications. More specifically, it is the crystalline cellulose structures in the plant cell wall, which consist of long chains of β -1,4 linked glucose monomers that form the region of interest in terms of yielding high quantities of fermentable glucose.

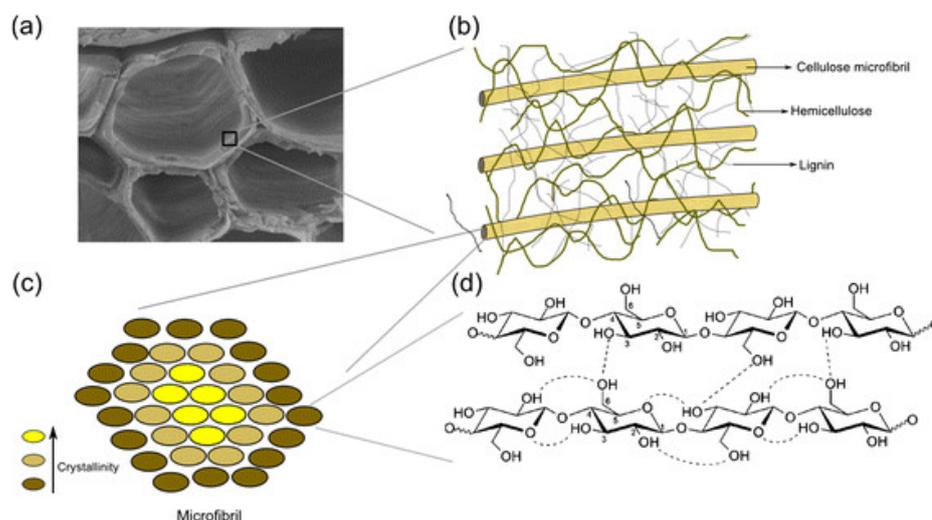


Figure 1: The structure of the plant cell wall (Tsuchida, 2017). a) Scanning electron microscope imaging of the xylem. b) The spatial arrangement of the 3 most prevalent plant cell wall constituents; cellulose, hemicellulose, and lignin. c) Colour-coded graphic depicting the increase of crystalline cellulose from the outer rings to the inner core of the microfibril structure. d) The β 1,4 – linked glucose monomers that make up the crystalline cellulose structure.

These glucose monomers are an ideal source of glucose for fermentation. However, plants have adapted their cell walls over millions of years to be as recalcitrant as possible to provide stability, structure, and to resist herbivory (Ali *et al.*, 2016). The tightly packed layers and hydrophobic regions within the plant cell wall are resistant to enzymatic degradation, meaning that the initial saccharification of lignocellulosic biomass must be conducted by physically breaking apart the material by means of wet/dry milling, degrading it by means of potent solvents and ionic liquids, or heating the material by means of steam explosion or microwave treatments in order to break apart the tightly packed layers within the cell wall matrix to release the glucose (Shirkavand *et al.*, 2016). These pretreatment steps are essential to the saccharification of lignocellulosic biomass as the subsequent enzymatic hydrolysis can yield up to 80% less fermentable sugars if not pretreated (Sindhu *et al.*, 2016). These factors translate into 2nd generation biofuel schemes being viewed as economically unfavourable in comparison to their 1st generation counterparts, despite the abundance of plant material that is virtually cost-free and readily available. Indeed, one third of the costs incurred by 2nd gen biofuel schemes are attributed to the pretreatment of the lignocellulosic biomass (Ali *et al.*, 2016).

The 3rd generation of biofuels is a very exciting prospect and is being researched and implemented globally. The premise for 3rd generation biofuels is based upon the cultivation of algae for use in biofuel production, algae are cultivated in vast offshore algal farms that are set up along coastlines (Puspawati

et al., 2015). Algae are a rich source of fermentable glucose, and are very readily degraded into bioethanol with little energy input, since macroalgal cell walls contain a far lower lignin content than higher plants (Alaswad *et al.*, 2015), whereas microalgae are more commonly used in biogas and biodiesel production as they are capable of both a high biomass and high lipid accumulation. These lipids are used to create biodiesel by means of transesterification (Chen *et al.*, 2015). The lower degree of lignification within macroalgal biomasses result in a less dense cell wall, thus enzymatic degradation can take place more readily without the need for physical or chemical disruption of the cell wall, resulting in the algae exhibiting a lower degree of recalcitrance to saccharification without investing as much energy into partially breaking down the cell wall when compared to higher plants (Takagi *et al.*, 2017). Typically, algal biomasses photosynthesise more efficiently than terrestrial plants (3 – 8% versus 0.5%) and thus grow faster and produce a larger biomass over time in comparison to higher or terrestrial plants (Chen *et al.*, 2015), (Alaswad *et al.*, 2015). However, the establishment of these algal farms is a costly endeavor, with both macroalgae (seaweeds) and microalgae (cyanobacteria) requiring very specific water temperatures and pHs, resulting in high cost and maintenance when it comes to cultivating algae. Although this can be done along coastal algae farms or in ponds and lakes in closed systems, both are susceptible to contamination from competing aquatic micro-and macroorganisms (Alaswad *et al.*, 2015), resulting in suboptimal conditions for algal cultivation due to the inherent nature of multiple organisms competing for food sources within a closed environment.

4th generation biofuels are linked to 3rd generation biofuels in that both are involved in “algae to biofuels” (Lü *et al.*, 2011) biofuel production. This involves the metabolic engineering of algae in order to utilize oxygenic photosynthesis to synthesise biofuels. “Algae” is often used as a blanket term for both prokaryotic cyanobacteria and eukaryotic microalgae; both are utilized in 3rd and 4th generation biofuel production methods. In 4th generation approaches, cyanobacteria and microalgae are genetically engineered to release high quantities of H₂ (biogas), have gene cassettes from *S. cerevisiae* transformed into them to create ethanologenic algae for bioethanol production, and have their lipid synthesis pathways altered in order to accumulate lipids for biodiesel production.

1.3 The Current Scope of the South African Biofuels Industry

The current state of the South African biofuels industry is mostly centered on 1st generation biofuel production methods. However, this has created a negative public opinion surrounding biofuel production due to the state of the South African agricultural sector. The South African agricultural sector is currently experiencing declining yields caused by political factors and shifts in prevailing weather patterns and rainfall, which are the main sources of the current national food deficit (The South African Department of Science and Technology (DST), 2017). It is clear that a nation that is struggling to feed its people cannot afford to divert resources into the production of biofuels, in fact, it is estimated that over 14 million households within this nation are vulnerable to the threat of food insecurity (Adeyemo and Wise, 2010).

The South African Department for Science and Technology (DST) has outlined these concerns within the “2017 Bioenergy Atlas”.

The Bioenergy Atlas encompasses all of the potential sources of sustainable energy within South Africa, including the possible barriers to bioenergy (particularly biofuels) initiatives, such as food insecurity. Additional concerns raised within the Atlas surrounding the growth of the South African biofuels industry centers around protecting and maintaining the valuable ecosystems and high degree of biodiversity that our country plays host to. The Atlas outlines several ecosystems across South Africa that are currently threatened by invasive alien plants (IAPs) and increased monoculture agricultural practices. Despite these concerns, the government has recently introduced new legislation that calls for a mandatory blending of bioethanol and biodiesel into the national liquid fuel supply based upon a white paper that was proposed for the year 2013. The paper calls for the mandatory blending of a minimum 2% bioethanol and biodiesel into the national liquid fuel supply (Blanchard *et al.*, 2011). More than ever before, it is crucial for us to discover novel ways to make 2nd generation biofuels a more feasible alternative to 1st gen biofuels in order to capitalize on the newly implemented liquid fuel regulatory guidelines and reduce our country’s overall fossil fuel consumption and dependency.

The Bioenergy Atlas potentially outlines the use of various crops such as, sugarcane, and sweet sorghum for use in 1st generation biofuel production platforms. The implementation of 1st generation platforms would usually have the potential to create detrimental knock-on effects to various sectors of the South African agricultural sector and economy. The Atlas highlights the need to shift the biofuel production industry in South Africa in such a manner so as to mitigate any potential sources of food insecurity. Whereas maize was once considered a prime candidate for 1st generation biofuels in South Africa, the production of biofuels using maize would not only have increased food prices, import rates, and food availability within the country, but also limited feedstock availability for livestock, placing our meat and wool production industries under pressure (Cloete and Idsardi, 2012). The newly implemented strategies call for the use of sugars extracted from sweet sorghum and sugarcane, crops that are currently produced abundantly enough within South Africa, for use in 1st generation biofuel initiatives. This synergises with the call for more 2nd generation biofuel approaches to utilise the bagasse created after harvesting both sweet sorghum and sugarcane. Currently, sugarcane bagasse is mostly utilised in “low-efficiency energy generation” with only 33% being utilised efficiently, thus, within the South African context, there exists an abundance of lignocellulosic biomass that can be used for bioethanol production (The South African Department of Science and Technology 2017). This forms part of new endeavour that has been termed the “valorization” of lignocellulosic material, assigning value to something that was previously worthless.

1.4 The Quest to Create a Feasible 2nd Generation Biofuel Platform

Second generation biofuels have long been considered as an economically costly method of biofuel production, due to the high-energy inputs required (pretreatments) compared to the relatively low recovery of bioethanol (Zhang *et al.*, 2016). This is due to the recalcitrance of the plant cell wall, since the matrix of lignin, hemicellulose, and cellulose is extremely tightly packed and hydrophobic (Sindhu, *et al.*, 2016), creating a structural and spatial barrier to enzymatic action (to release glucose). Thus energy and cost intensive measures are usually employed to pretreat the lignocellulosic biomass prior to enzymatic degradation and subsequent fermentation to yield bioethanol (Xing *et al.*, 2012). These methods are generally steam pretreatment, mechanical compression and break-up of the material, submersing the material in ionic liquids and organic solvents to degrade parts of the matrix, or a combination of all the abovementioned (Amnuaycheewa *et al.*, 2016).

This high labour and energy input is what makes 2nd generation biofuels economically unfavourable, however, there are ways to overcome this barrier by adding specific enzyme cocktails or genetically engineering “super” yeast strains that code for a selection of highly active and targeted enzymes (Wu *et al.*, 2017). These targeted enzymatic methods form part of what is known as consolidated bioprocessing (CBP), wherein lignocellulolytic enzymes and glycosyl hydrolases are combined in a one-step reaction that yields fermentable sugars for conversion into bioethanol by means of anaerobically respiring yeast cells. These are generally specialized strains of *S. cerevisiae* that have been bioengineered and transformed with plasmids coding for highly stable and highly active lignocellulytic enzymes (Wang *et al.*, 2016). CBP relies on the action of three key enzymes in order to liberate fermentable glucose monomers from lignocellulosic biomass (Zhang *et al.*, 2016).

Firstly, long chain, amorphous crystalline cellulose chains are broken into shorter chain cellooligosaccharides by an endoglucanase (EC 3.2.1.4). Endoglucanases are unique in the sense that they are able to bind to these amorphous, long-chain crystalline cellulose structures and perform successful cleavage of the crystalline structures. Secondly, cellobiose moieties are liberated from these shorter chain cellooligosaccharides by an exoglucanase (EC 3.2.1.91) that binds to the reducing ends of the chain. Finally, a β -glucosidase (EC 3.2.1.21) enzyme then cleaves these cellobiose moieties into two glucose monomers liberating them for fermentation into bioethanol.

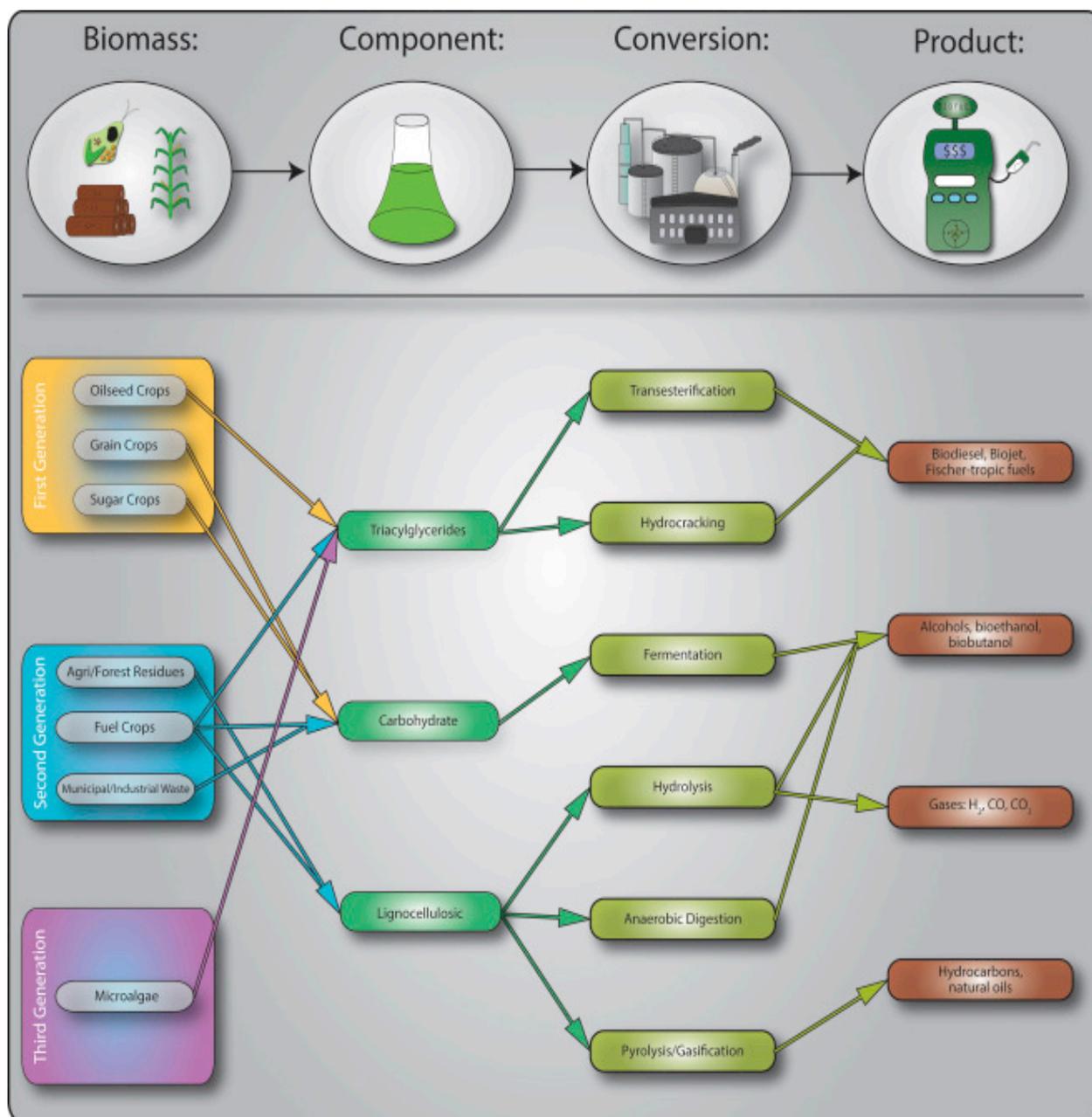


Figure 2: Basic schematic workflow of the current biofuel production methods (Brief Overview of Biofuel Production Workflows, 2017). The figure depicts (i) the 3 major generations of biofuel production, (ii) the source material used in each generation, (iii) the desired biomass for biofuel production, (iv) the main methods of converting specific biomasses into biofuels, (v) and the final biofuel product yielded.

Within the realm of 2nd generation biofuels, there exist various methods of pretreatment as previously mentioned in order to degrade the lignin content within the cell wall matrix to maximize the saccharification of the cell wall. There is therefore a need to identify cellulolytic and cellobiohydrolytic enzymes that are robust enough to withstand industrial processes and conditions whilst still maintaining a high specific activity for its preferred substrate. These enzymes would be required to remain stable under high temperatures, varying pH ranges, high concentrations of solvents, surfactants, and detergents and high concentrations of ethanol and glucose (Xing *et al.*, 2012).

Identifying novel enzymes can be approached in multiple ways and one of the most effective methods is to sample the metagenome of a specific site, chosen for its conditions, in order to isolate and identify previously undiscovered enzymes from micro-organisms that cannot be cultured using standard laboratory practices (Zhang *et al.*, 2016). Metagenomic sampling is an invaluable tool within the realm of novel enzyme discovery, enabling one to potentially discover a vast amount of novel activities from a single sample. The creation of one metagenomic library holds the potential to yield a multiplicity of potentially applicable enzymes. This practice has been termed bioprospecting and has quickly become the most utilized method of novel enzyme discovery, mostly due to the high-throughput next generation sequencing technologies that are now available. These highly advanced sequencing technologies facilitate the novel enzyme discovery platform by providing the tools to perform *de novo* assemblies of large genetic data sets and rapidly identify open reading frames within assembled data sets (Mirete *et al.*, 2016).

Objectives of this Study

Clone 3L is a metagenomic clone that was recovered from a library created utilizing samples that were obtained from dairy run-off. Initially, the library was created as a means to identify clones that are able to produce biopolymers utilizing various carbon sources as substrates. Clone 3L was shown to be able to produce biopolymers, however, following domain analysis it was also discovered that 3L was a metagenomic clone that possessed 2 discrete active domains. Activity assays using chromogenic substrates elucidated that clone 3L shows activity upon two substrates, cellobiose and lactose.

Clone 3L proved to be intriguing due to this dual activity, particularly its activity upon cellobiose, outlining 3L as a β -glucosidase. With new legislation being implemented in South Africa regarding biofuels and cutting-edge biofuel technology being researched at both the Microbiology and Process Engineering faculties of Stellenbosch University, we sought to:

1. Comprehensively illustrate the bi-functional nature of clone 3L, utilizing plate-based screening assays.
2. Biochemically characterize the β -glucosidase domain of clone 3L utilizing its natural substrate.
3. Assess the ability of 3L to tolerate unfavourable conditions and thus gauge its suitability within a 2nd gen bioethanol production pipeline.

In order to achieve these objectives;

- Clone 3L was screened *in vivo* within a bacterial strain that carries no native β -glucosidase or β -galactosidase activity, Δ LacZ strain of *E. coli*.

- The maximal activity of 3L, affinity for cellobiose, temperature, and pH optimum was determined via biochemical assays.
- Clone 3L was utilized within incubations containing various added elements that were used to gauge its ability to retain affinity for, and activity on its substrate under industrial conditions.

2. Materials & Methods

Bacterial Strains and Vectors used within this Study:

DH5α (Invitrogen)

F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* ϕ 80d*lacZ* Δ M15 Δ (*lacZYA-argF*)U169, *hsdR17*($r_K^- m_K^+$), λ^-

DH5α Δ LacZ (Kyle Willard, IPB)

F' *proA+B+ lacI^q Δ LacZ fhuA2 Δ (lac-proAB) glnV gal R(zgb-210::Tn10)Tet^S endA1 thi-1 Δ (hsdS-mcrB)*

A previous member of the IPB, Kyle Willard, created the Δ LacZ line by inserting a chloramphenicol resistance cassette into the endogenous *E. coli LacZ* gene. The insertion of the resistance cassette was achieved by means of phage transduction, interrupting the *LacZ* gene and abolishing native β -galactosidase activity. Screening both the Δ LacZ mutant and the wild type DH5 α on LB agar media supplemented with IPTG & X-Gal results in a blue hue forming within the wild type colony, whereas no colour-shift is observed within the Δ LacZ colony. The wild type DH5 α strain is already LacZ deficient, however, the creation of the Δ LacZ mutant was deemed necessary. This was to prevent possible alpha-complementation that may lead to false positives when screening metagenomic libraries for β -galactosidase activity. Furthermore, the insertion of the chloramphenicol resistance cassette allows for a more stringent antibiotic control when screening metagenomic libraries or constructs with different plasmids. This allows the Δ LacZ strain to be useful when utilising different plasmids that carry different antibiotic resistance markers, and allows the user to perform antibiotic counter-selection as a means of insert verification.

The wild type exhibits this blue colouration due to the intact *LacZ* gene, enabling the bacteria to code for a functional β -galactosidase enzyme that is able to cleave the X-Gal substrate and release the 5-bromo-4-chloro-3-indolyl moiety that results in the blue staining of the colony. The Δ LacZ mutant lacks a functional *LacZ* gene due to the insertion of the chloramphenicol resistance cassette, preventing the formation of an active β -galactosidase enzyme within the Δ LacZ strain. Lacking the β -galactosidase enzyme, the Δ LacZ mutant is inactive upon the X-Gal substrate. For downstream protein expression applications, the Clone 3L coding sequence was cloned into the pRSETa protein expression vector (Thermo Fisher Scientific, South Africa). The pRSETa vector was chosen for many reasons; its availability within the Institute for Plant Biotechnology, its low-levels of basal (leaky) expression, His-tagging for protein purification, and its ampicillin resistance marker to counter-select against the chloramphenicol resistance of the Δ LacZ strain (see appendix 7.1).

2.1 Metagenomic Library Creation and Identification of Clone 3L

The metagenomic library, from which Clone 3L was identified, was derived from sampling the dairy slurry run-off at the Stellenbosch University experimental farm at Welgevallen (33.9427° S, 18.8664° E). The library was constructed by means of phage transduction, using the Agilent Gigapack III Gold kit (Agilent Technologies, USA). The resultant clones were initially screened for β -galactosidase activity. Plate-based screening for β -galactosidase activity was conducted by utilizing Luria-Bertani agar plates (1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1.5% (w/v) bacterial agar, 40 μ g/mL X-gal, 0.1 mM IPTG) and the appropriate antibiotics. 5 mL cultures of clone 3L in the Δ LacZ strain, the Δ LacZ complemented with a functional *LacZ* gene construct, the empty Δ LacZ strain, and wild type DH5 α , were grown to an OD₆₀₀ of 0.4 (37°C, at 200 RPM agitation) and 3 μ L of each culture was spotted onto the plates in triplicate. The plates were left to incubate overnight at 37°C. Positive activity resulted in a blue colouration of the bacterial colony. Clones that yielded positive activity in the *E. coli* (DH5 α Δ LacZ) mutant background were then sequenced and putative open reading frames (ORFs) identified using the CLC Genomics Workbench software package (v.10, Qiagen Bioinformatics, inqaba biotec™, South Africa). Putative ORFs were then sub-cloned into the pRSETa protein expression vector (Thermo Fisher Scientific, South Africa) for downstream experiments. These subcloned ORFs were retransformed into *E. coli* (DH5 α Δ LacZ) and retested for β -galactosidase activity in plate based functional screening assays.

2.2 Bioinformatic Analyses of the Sequence of Clone 3L and β -glucosidase Activity Screens

ExPASy domain analysis of Clone 3L displayed two discrete glycosyl hydrolase 1 (β -glucosidase) family domains, thus 3L was also screened for β -glucosidase activity using the esculin – ferric citrate plate-based assay. 5 mL cultures of; clone 3L in the Δ LacZ strain, the Δ LacZ complemented with a functional *LacZ* gene construct, the empty Δ LacZ strain, and wild type DH5 α , were grown to an OD₆₀₀ of 0.4 (37°C at 200 RPM agitation) and 3 μ L of each culture was spotted in triplicate on LB Agar plates supplemented with esculin and ferric citrate (1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1.5% (w/v) bacterial agar, 0.1% (w/v) esculin, 0.05% (w/v) ferric citrate) and the appropriate antibiotics. The plates were left to incubate overnight at 37°C. Clones that tested positive for β -glucosidase activity were identified by a dark brown colouration of the bacterial colony and a dark brown halo formation around the colony.

2.3 Heterologous Expression of Clone 3L in the *E. coli* (DH5 α Δ LacZ) Mutant Background

Heterologous protein expression was conducted within the Δ LacZ strain. 5 mL overnight starter cultures were inoculated into 500 mL flasks containing 200 mL LB supplied with the appropriate antibiotics. Flasks were incubated at 37°C, with agitation, until an OD₆₀₀ of 0.4 was attained. At an OD₆₀₀ of 0.4, protein induction was initiated by the addition of IPTG at a final concentration of 0.2 mM. Induction was

allowed to proceed for 4 hours, at which point the cultures were divided and transferred to 50 mL conical tubes and pelleted by centrifugation for 10 minutes at 4°C and 5 000 *g*.

The bacterial pellets were then resuspended in protein extraction buffer (50 mM HEPES-KOH, 5 mM MgCl₂, 1 mM EDTA, 20 mM DTT, 0.1% (v/v) Triton-X 100, 1 mM benzamidine, 1 mM PMSF, 50 mM sodium ascorbate, 2% (w/v) PVP), at which point lysozyme (0.1% (w/v)) was added to the resuspension mix and incubated on ice for 30 minutes, with gentle agitation. Following lysozyme treatment, the resuspended pellets were subjected to 3 rounds of sonication consisting of a 10 second pulse and a 15 second rest per round. Finally, the bacterial lysate was pelleted by means of centrifugation for 15 minutes at 4°C and 10 000 *g* in order to separate the bacterial debris from the protein suspended in the supernatant.

2.4 Determining Protein Concentration

A Bradford assay was used to determine protein concentration spectrophotometrically using the BioRad Bradford Protein Assay (Bio-Rad, USA). 5 μ L of crude protein extract was added to 200 μ L of Bradford reagent in one well of a 96-well microplate. The mixture was incubated for 5 minutes and the absorbance measured at 595 nm using a VersaMax ELISA microplate reader and the accompanying SoftMax Pro software.

2.5 Enzyme Activity Determination

The biochemical incubations were conducted within this study were executed by means of 100 μ L reactions consisting of 10 μ L cellobiose (50 mM final concentration), 40 μ L of crude enzyme extract (\approx 35 μ g of crude total protein and \approx 4.6 μ g crude 3L protein per reaction), and 50 μ L of buffer unless stated otherwise. The incubations were allowed to proceed for 60 minutes, following which the reactions were terminated by heat inactivation at 95°C for 10 minutes.

2.5.1 Determining the Optimal pH for Enzyme Activity

The pH optimum of 3L was determined by measuring the concentration of glucose liberated from cellobiose across a range of values from pH 3.0 – 10.0 in intervals of 1.0. The following buffers at 100 mM concentrations were utilized in order to establish the pH ranges; Mcllvaine buffer (pH 3.0 – 4.0), MES –KOH (pH 5.0 – 6.0), HEPES-KOH (pH 7.0 – 8.0), and sodium carbonate (pH 9.0 – 10.0).

2.5.2 Determining the Optimal Temperature for Enzyme Activity

The temperature optimum of the enzyme was determined by measuring the concentration of glucose liberated from cellobiose at a temperature range of 20-70°C, in 10°C increments.

2.5.3 Determining the Enzyme Kinetics

The V_{max} and K_M of 3L was determined by measuring the concentration of glucose liberated from cellobiose using a set of incubations with various cellobiose concentrations, from 0-10 mM, in 1 mM increments. The V_{max} and K_M were determined using an allosteric sigmoidal modeling software package within GraphPad Prism 5.

2.5.4 Examination of Enzymatic Co-Factors

The activity of the enzyme with various ions and additives was determined by measuring the concentration of glucose liberated from cellobiose. Additional incubations that were supplemented with various ions (EDTA, $CoCl_2$, KCl, NaCl, SDS, and $CaCl_2$) were set up at two concentration points, 5 mM and 10 mM.

2.5.5 Determining the Glucose Tolerance of the Enzyme

In order to ascertain the glucose tolerance of 3L, the concentration of glucose liberated from cellobiose was determined by set of incubations supplemented with a glucose concentration spanning 0 – 0.8 M. The glucose concentration gradient was established by adding various volumes of a 3 M glucose stock to each of the incubations.

2.5.6 Determining Enzyme Stability in Solvents/Surfactants

Further incubations used to determine the stability of 3L when exposed to ethanol, methanol, Triton-X 100, Tween 20, Glycerol, and DMSO, were executed by measuring the concentration of glucose liberated from cellobiose when supplementing the incubations with a % (v/v) gradient of each chemical ranging from 0%-40% (v/v).

2.5.7 Determining the Bond Specificity of the Enzyme

The bond specificity analysis of 3L was conducted via a set of incubations with β -D-glucopyranoside, α -D-glucopyranoside, β -D-galactopyranoside, and α -D-galactopyranoside, with 10 μ L of 2 mg/mL stock added to the incubation in place of the 10 μ L cellobiose substrate. The bond specificity was determined by measuring the absorbance of each incubation at 405 nm using the VersaMax ELISA microplate reader and the SoftPro software suite.

2.6 Liberated Glucose Quantification

The glucose liberated within the incubations was quantified utilizing the Boehringer Mannheim D-Glucose assay kit (Roche, Switzerland) via a modification of the manufacturer's protocol. The manufacturer's protocol was scaled down by a factor of 5 in order to adapt the protocol from a cuvette-based assay into a 96-well microplate format. The assay was conducted by adding 20 μ L of incubation sample to a mixture of 200 μ L Solution 1 (triethanolamine buffer pH 7.6 16% (w/v), NADP (2.4 μ g/mL), ATP (5.8 μ g/mL), and magnesium sulphate) and 380 μ L of ddH₂O. Initial absorbance of the mixture was measured at 340 nm using a VersaMax ELISA microplate reader using the SoftMax Pro software suite.

The reaction was initialized with the addition of 4 μL of Solution 2 (Hexokinase (0.3 U/ μL) and glucose-6-phosphate dehydrogenase (0.16 U/ μL). The reaction was incubated for 15 minutes at 25°C, at which point, 200 μL of each reaction was pipetted into 3 separate microplate wells. The absorbance of each reaction was measured in triplicate at a wavelength of 340 nm using the VersaMax ELISA microplate reader and the SoftMax Pro software suite. Each biochemical assay was conducted independently with triplicate technical repeats ($n = 3$).

All graphs within the results section are plotted with standard error and error bars, excluding the graphs generated using relative activity (%). The relative activity data was calculated from previous graphs that contained 3 technical repeats across two independent experiments. If error bars are not visible on certain data points, this is due to the standard error value being too low for the error bar to extend above or below the data point.

3. Results

3.1 Clone 3L Consistently Displays both β -galactosidase and β -glucosidase Activities on Plate-Based Functional Screens

Following the identification of two putative open reading frames in clone 3L, the ORF was subcloned into the pRSETa expression vector. In a series of subsequent experiments, both the β -galactosidase and β -glucosidase activities were tested for in plate based screening assays. For the β -galactosidase functional screens the Δ LacZ mutant background transformed with clone 3L tested weakly positive for β -galactosidase activity when compared to the Δ LacZ mutant background transformed with a known β -galactosidase gene. Importantly neither the DH5 α background, nor the Δ LacZ mutant background (created in the DH5 α background) showed any β -galactosidase activity (Fig. 4, Chapter 3). We confirmed the respective backgrounds by introducing the relevant antibiotics into subsequent experiments (Fig. 4, Chapter 3), demonstrating that neither the DH5 α background, nor the Δ LacZ were able to grow on ampicillin. However since the pRSET vectors imparts this resistance both the positive control and clone 3L (in Δ LacZ) grew and displayed β -galactosidase activity (Fig. 4, Chapter 3)

For the β -glucosidase functional screens only the Δ LacZ mutant background transformed with clone 3L showed strong β -glucosidase activity when screen on plates containing esculin and ferric citrate. Importantly, the Δ LacZ mutant background transformed with a known β -galactosidase gene showed no β -glucosidase activity confirming the presence of a functional domain in clone 3L (Fig. 6, Chapter 3). We similarly confirmed the respective genetic backgrounds by introducing the respective antibiotics and confirmed that activity was due to the presence of clone 3L in the respective Δ LacZ mutant background (Fig. 7, Chapter 3).

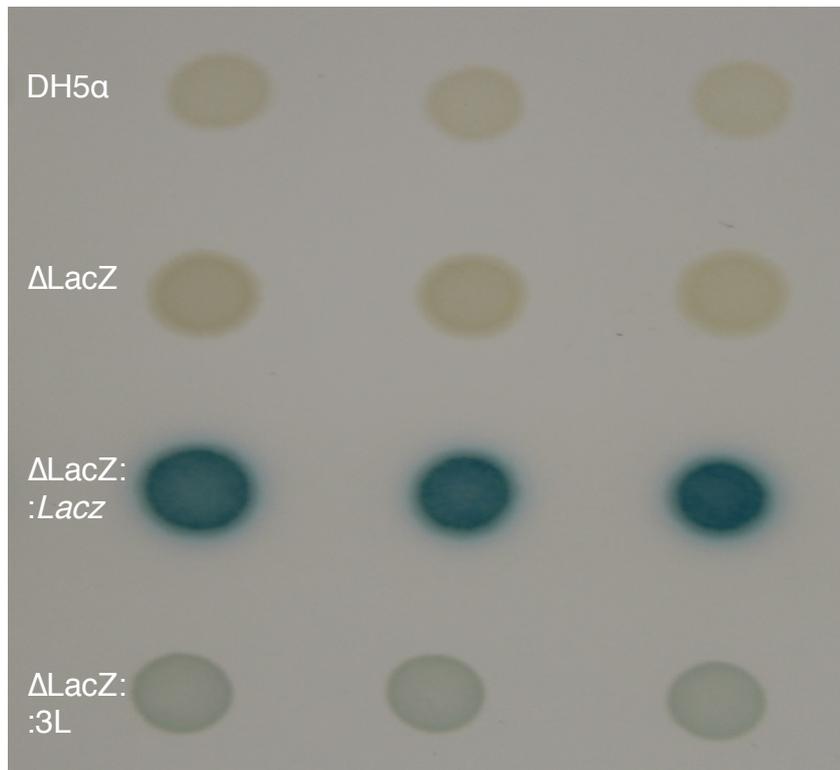


Figure 4: LB agar plates supplemented with IPTG and X-Gal were used to screen for the β -galactosidase activity once clone 3L had been subcloned into the pRSETa expression vector. Cultures (5 mL) were grown to an OD_{600} of 0.4 and 3 μ L aliquots spotted into the plates, followed by overnight incubation at 37°C. The DH5 α strain and the Δ LacZ mutant (created in the DH5 α background) were used as negative controls. For the positive control, β -galactosidase activity was recovered in the Δ LacZ mutant by cloning a known β -galactosidase into the pRSETa expression vector.

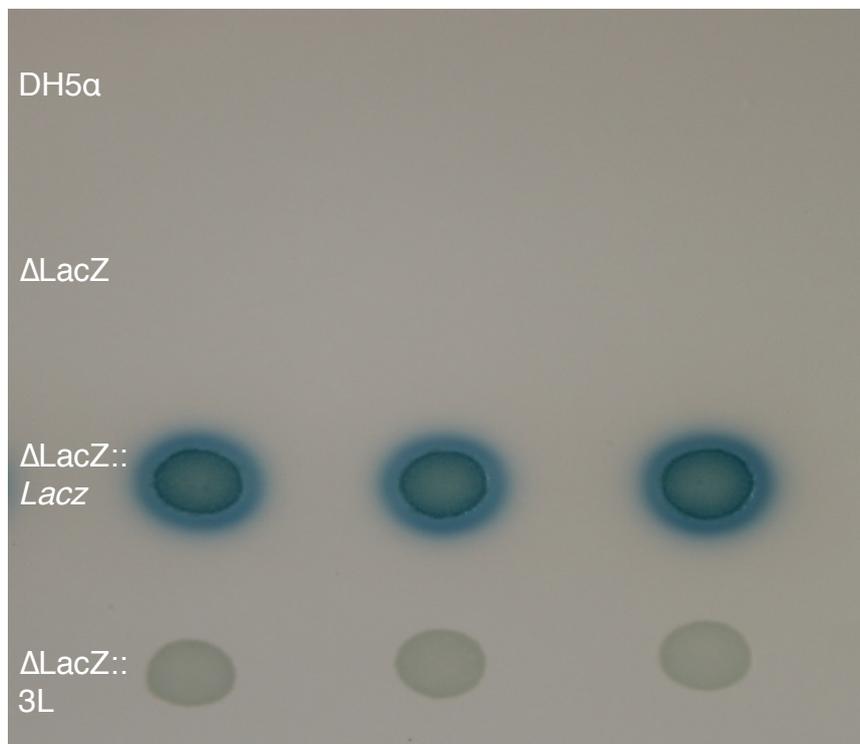


Figure 5: LB agar plates supplemented with IPTG and X-Gal in addition to Amp¹⁰⁰ and Chi³⁴ were used to screen for activity and confirm the backgrounds of the various strains. 3 μ L of each culture was spotted on the plate from 5 mL cultures at an OD_{600} of 0.4.

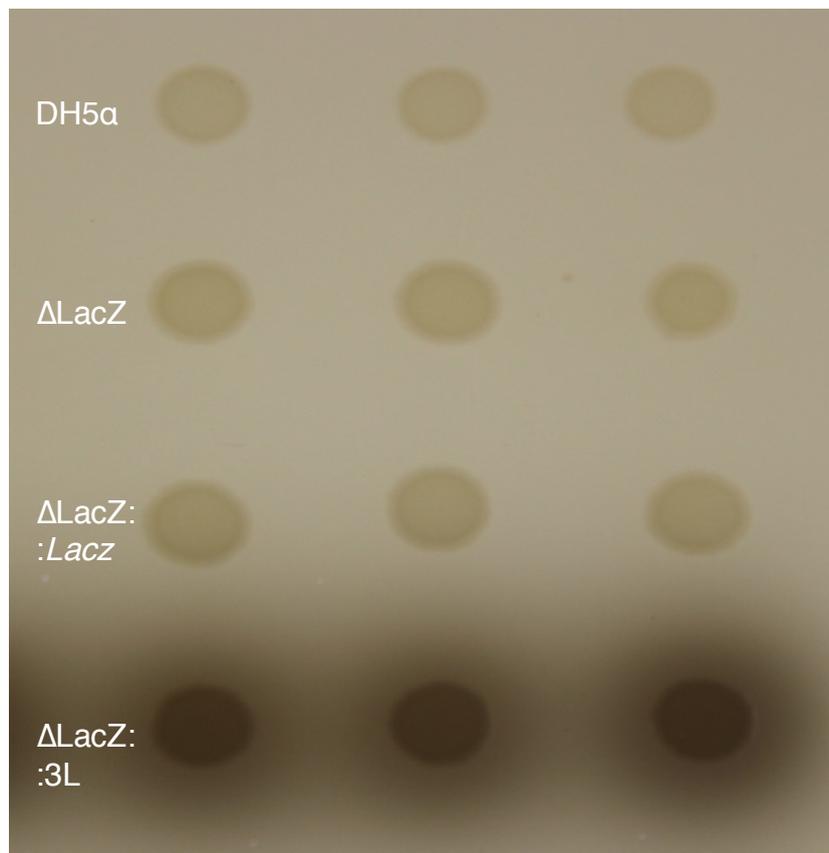


Figure 6: LB agar plates supplemented with esculin and ferric citrate with no antibiotic selection were used to screen for the β -glucosidase. Each culture was grown in 5 mL LB until an OD_{600} of 0.4. 3 μ L of each culture was spotted onto the plate and incubated at 37°C overnight.

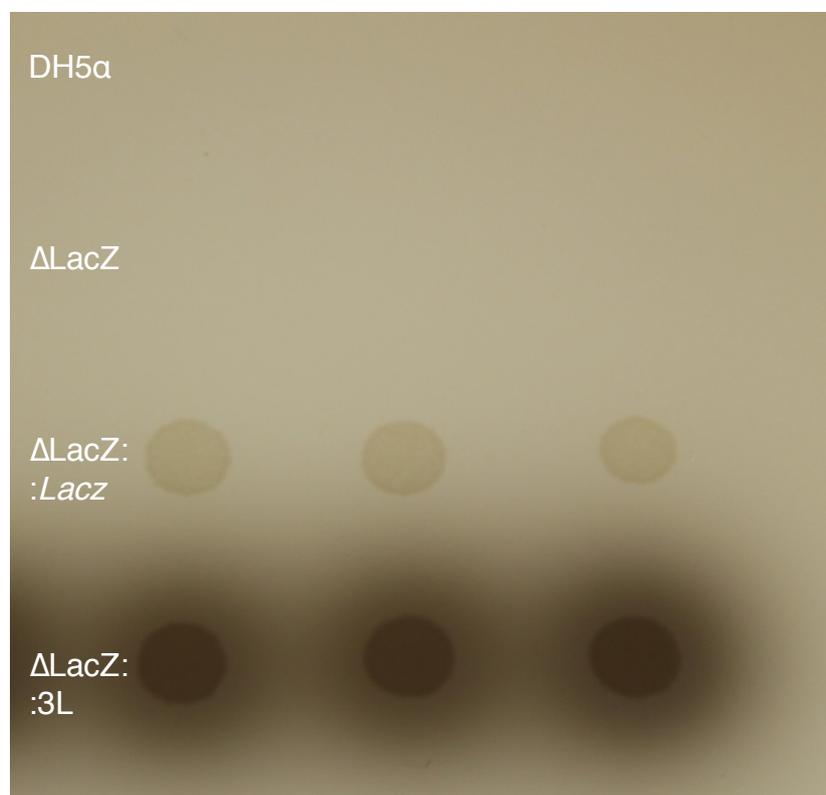


Figure 7: LB agar plates supplemented with esculin, ferric citrate, Amp¹⁰⁰ and Chl³⁴ were used to screen for β -glucosidase activity and confirm the backgrounds of the various strains. 3 μ L of each culture was spotted on the plate from 5 mL cultures at an OD_{600} of 0.4.

3.2 *In Silico* Analyses of Clone 3L, Initially Identified from Functional Screens of Metagenomic Clones as Containing a Putative β -galactosidase Open Reading Frame

Once the metagenomic clone from which 3L was sourced, was identified as containing a putative β -galactosidase on plate based screening assays (Figs. 4 – 7, Chapter 3) and the putative ORF (for 3L) identified from the clone following sequencing, the sequence was analysed using the ExPASy PROSITE domain finder (<http://prosite.expasy.org/>). The *in silico* domain analyses conducted on the amino acid sequence of clone 3L yielded two discrete glycosyl hydrolase family 1 domains situated at the amino acids positions 13 to 27 and 347 – 355 (Fig. 3, Chapter 3) of the predicted protein (total predicted size in amino acids). These domains typed as a β -glucosidase and a β -galactosidase, respectively when clone 3L was subjected to plate-based functional screening assays.

hits by patterns: [2 hits (by 2 distinct patterns) on 1 sequence]

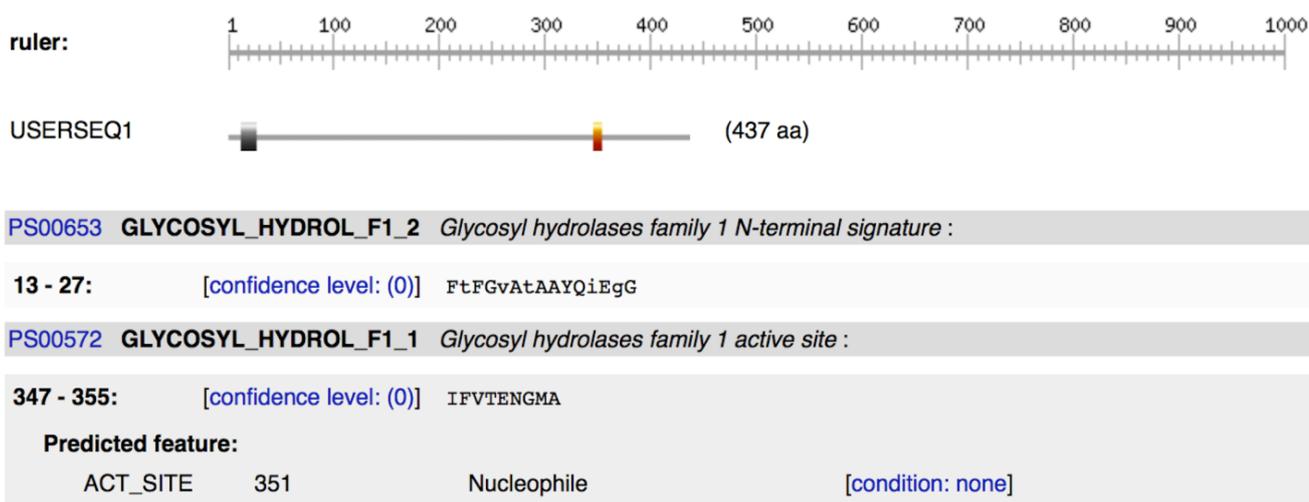


Figure 3: The ExPASy PROSITE domain finder analysis of the clone 3L open reading frame (ORF) displaying the two discrete domains found within the ORF. The analysis software classed both domains as GH1 domains, the upstream domain consisting of 14 residues and the downstream, 8 residues. The upstream domain carries an N-terminal signature, with the downstream domain classified as the active site.

3.3 Enzymatic Characterization of the β -glucosidase Activity of Clone 3L

The kinetic parameters of 3L were determined using 3 assays, temperature, pH, substrate affinity (K_m), and maximal reaction velocity (V_{max}). The temperature optimum of 3L (Fig. 8, Chapter 3) was determined across a temperature range of 20 – 70°C, in 10°C increments, with the empty Δ LacZ strain utilized as a negative control. Activity increases gradually from 20 – 40°C, with a maximal activity of 56.85 μ M glucose. m^{-1} . μ g protein $^{-1}$ observed at 40°C. Activity decreased 4-fold at a temperature of 50°C and was abolished at a temperature of 70°C. Having determined 40°C to be the optimal temperature all subsequent experiments were conducted at 40°C.

The pH optimum of 3L was ascertained across a pH range of 3.0 – 10.0 (Fig. 9, Chapter 3), with the empty Δ LacZ strain acting as a negative control. The pH optimum of 3L was at pH 8.0 with a maximal activity of $39.46 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$. Clone 3L displays a wide active pH range across pH 6.0 – 9.0, with activity being abolished below a pH of 4.0 and above a pH of 9.0. After determining the pH optimum, all subsequent experiments were conducted at a pH of 8.0.

The kinetic properties of clone 3L were determined via an assay incorporating cellobiose across a 0.0 – 10.0 mM concentration range in increments of 1.0 mM (Fig. 10, Chapter 3). The data points were modeled with an allosteric sigmoidal model, mapped to an accuracy (R^2) of 0.952. The allosteric sigmoidal modeling package incorporated into GraphPad Prism 5 was used to determine the V_{max} and K_m . Following the statistical analyses, the V_{max} was calculated at $74.98 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$ with a K_m of 1.842 mM.

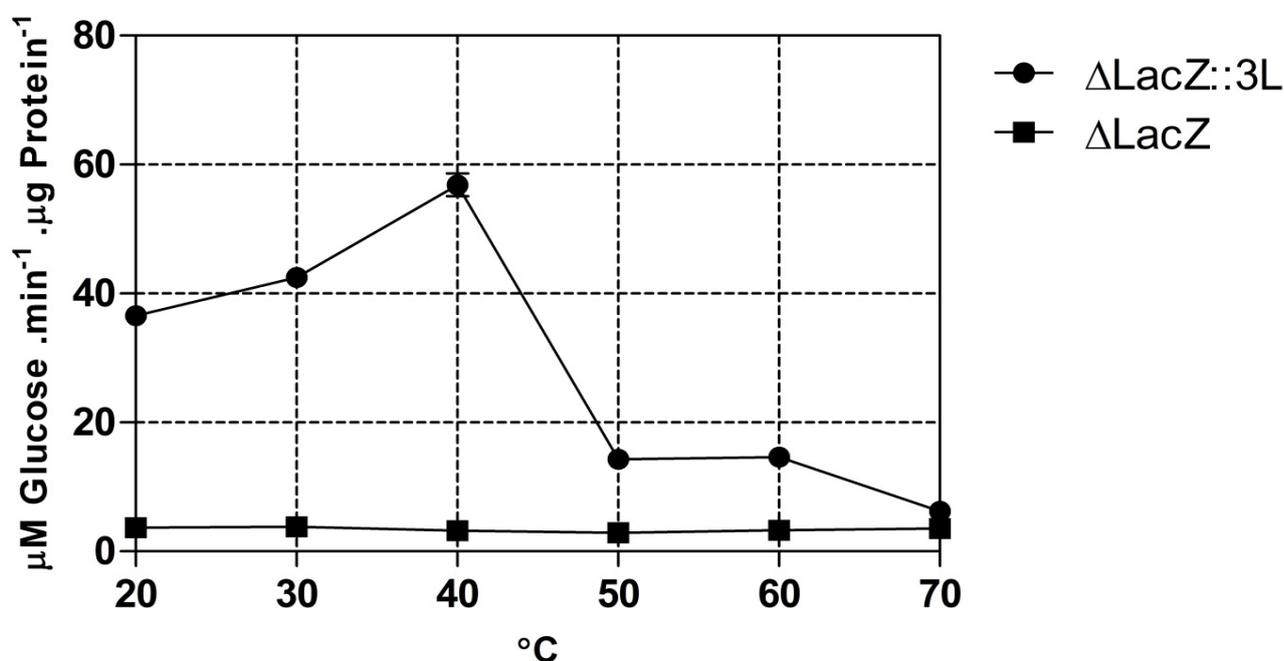


Figure 8: The temperature optimum of 3L was determined across a temperature range of 20 – 70°C, with the empty Δ LacZ strain acting as a negative control. The assay was conducted with 50 mM of cellobiose substrate (10 μL in the reaction), 40 μL of crude enzyme extract, and 50 μL of 100 mM HEPES-KOH pH 8.0 for 60 minutes.

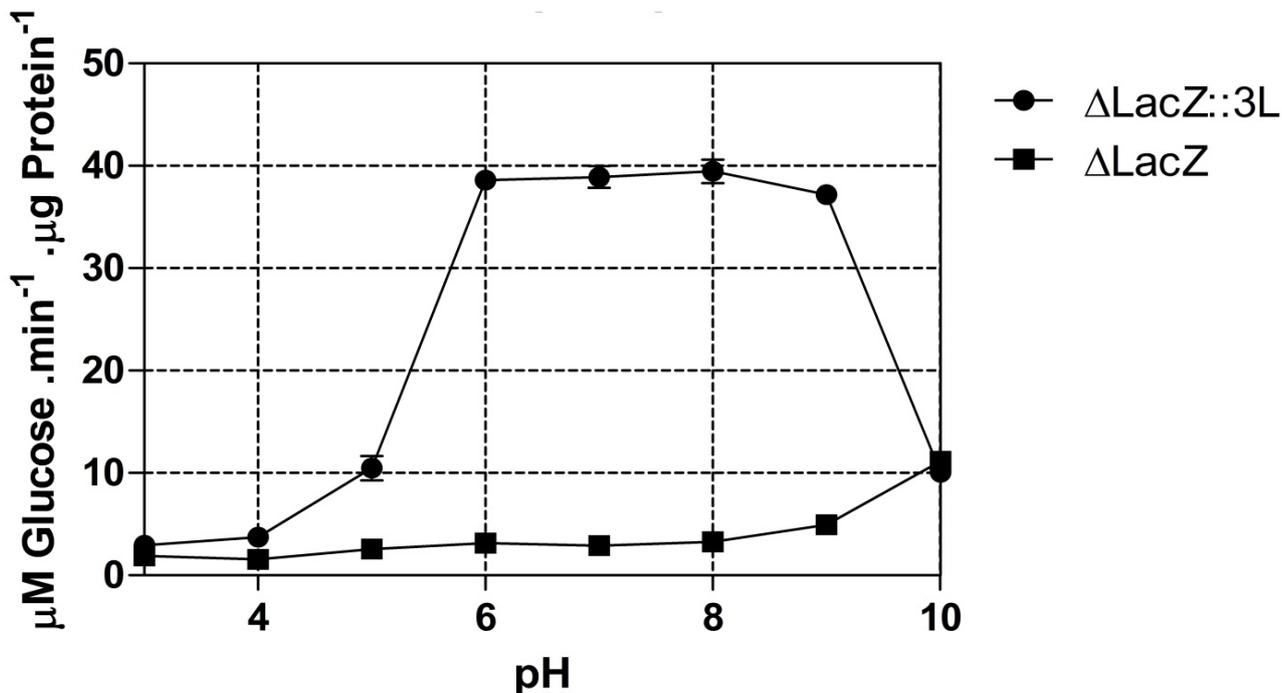


Figure 9: The pH optimum of 3L was ascertained across a pH range of 3.0 – 10.0. The empty ΔLacZ strain was utilized as a negative control. The assay was conducted with 50 mM cellobiose (10 μL in the reaction), 40 μL of crude enzyme extract, and 50 μL of buffer, at a temperature of 40°C for 60 minutes. The buffers utilized were; Mcllvaine Buffer (pH 3.0 – 4.0), MES-KOH (pH 5.0 – 6.0), HEPES-KOH (pH 7.0 – 8.0), and sodium carbonate (pH 9.0 – 10.0) all at 100 mM.

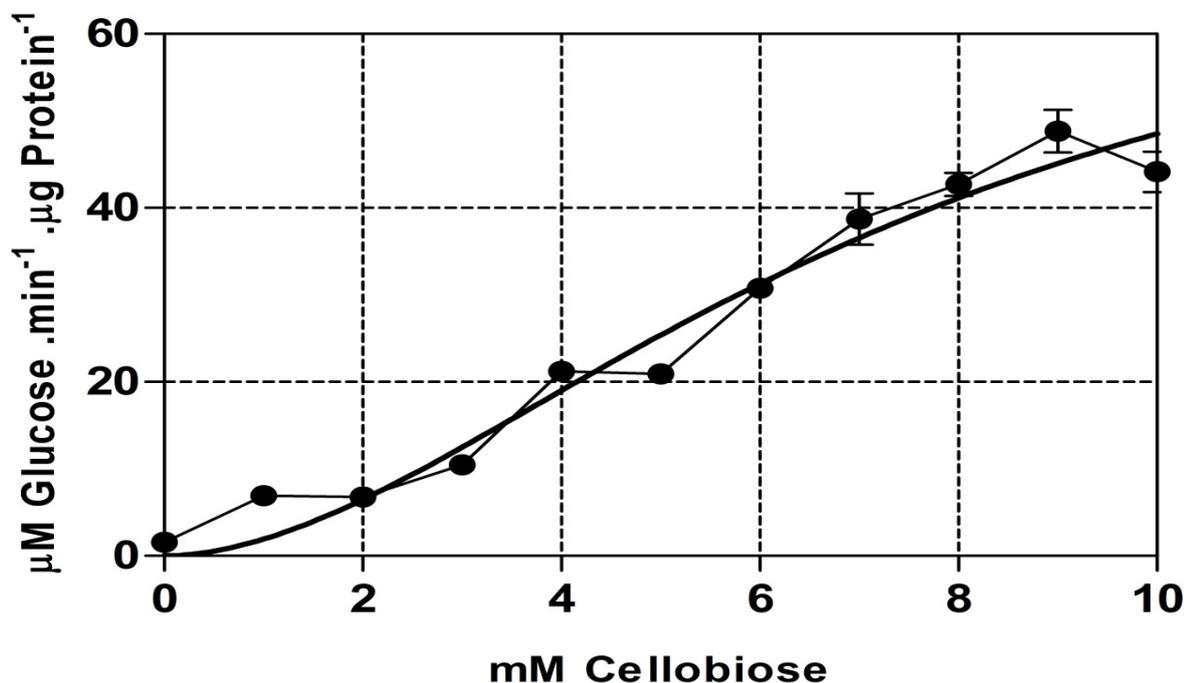


Figure 10: The kinetic properties of 3L were determined by assaying the activity across a cellobiose concentration gradient, ranging from 0.0 – 10.0 mM cellobiose. The incubations consisted of 10 μL cellobiose at various concentrations, 40 μL crude enzyme extract, 50 μL 100 mM HEPES-KOH pH 8.0, at a temperature of 40°C for 60 minutes. The kinetic properties of 3L were calculated using the allosteric sigmoidal modeling package incorporated in the GraphPad Prism 5 suite.

3.4 Evaluation of the Industrial Resilience of 3L

In order to establish the suitability of 3L for use in a biofuel production pipeline, a battery of assays were required with supplemented ions, reagents, solvents, and sugars. Assays were conducted with supplemented ions, CoCl_2 , KCl , NaCl , CaCl_2 , a chelating agent, EDTA, and the detergent, SDS. The abovementioned components were supplemented at both 5 mM and 10 mM concentrations, with the resulting activity compared to the control activity for 3L ($48.82 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$).

At a 5 mM concentration (Fig. 11, Chapter 3), clone 3L maintains 71.65% of its activity in the presence of EDTA with an activity of $34.98 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$. At the 10 mM concentration, 3L displays increased activity in the presence of EDTA, maintaining 75% of its activity at $37.09 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$. When exposed to SDS at both a 5 and 10 mM concentration, activity is reduced to 22% ($10.75 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$) and 14% ($6.83 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$), respectively. At both 5 mM and 10 mM concentrations, none of the supplemented ions provided an increase in activity, with all ions resulting in a 20% - 43% reduction in activity relative to the control activity.

The glucose tolerance of 3L was determined across a glucose concentration range of 0.0 – 0.8 M glucose supplemented to the incubations (Fig. 12, Chapter 3). Clone 3L displays a 4.8-fold reduction in activity at a 0.2 M glucose concentration, with activity abolished at the higher concentrations.

Clone 3L was assayed for methanol and ethanol tolerance across a 0 – 40% (v/v) concentration spectrum with 5% increments (Fig. 13, Chapter 3). At 5% (v/v) ethanol, 3L exhibits a 4.18-fold ($40.647 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$ to $9.716 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$) reduction in activity whilst only exhibiting a 1.08-fold ($40.647 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$ to $37.678 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$) reduction in activity at 5% (v/v) methanol. Activity is virtually abolished at 10% (v/v) ethanol, whilst a 1.64-fold ($40.647 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$ to $24.752 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$) reduction in activity is observed at 10% (v/v) methanol. Abolishment of activity in the methanol incubations is only noted at 30% (v/v) methanol. Thus, 3L exhibits a greater tolerance to methanol than to ethanol across a broad concentration range

The bond specificity of 3L was determined by using 4 chromogenic substrates, pNP β -D-glucopyranoside, pNP α -D-glucopyranoside, pNP β -D-galactopyranoside, and pNP α -D-galactopyranoside. The activity of 3L upon each chromogenic substrate was determined spectrophotometrically at 405 nm, with a higher absorbance denoting a greater concentration of pNP moieties liberated, and thus the preferred substrate and bond configuration (Fig. 14, Chapter 3). The greatest absorbance at 405 nm was observed utilizing the pNP β -D-glucopyranoside substrate, with an absorbance of 0.84. A 60% reduction in activity can be observed when pNP α -D-glucopyranoside is utilized in the incubation, with both galactopyranoside substrates displaying negligible or background

activity. This result confirms that clone 3L behaves as a β -D-1,4-cellobiohydrolytic enzyme and adheres to the characteristics of an EC 3.2.1.21 enzyme, according to the CAZy database.

In order to ascertain the resilience of clone 3L when exposed to common industrial solvents and surfactants, incubations were set up with Glycerol, DMSO, Tween 20, and Triton-X 100 across a 0 – 40% (v/v) concentration gradient in 10% increments (Fig. 15, Chapter 3). The relative activity of 3L was compared to the activity of the control 3L reaction ($40.647 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$). With regards to the surfactants, Tween 20 and Triton-X 100, clone 3L remains stable in Tween 20, with the lowest relative activity of 71% noted across the concentration gradient. Activity spikes at 20% (v/v) Tween 20 with an observed relative activity of 105%. Clone 3L maintains 98% of its activity in the presence of 10% (v/v) Triton-X 100, with activity reducing to 36% at 20% (v/v), 11% at 30% (v/v), and 22% at 40% (v/v). Overall, clone 3L retains a higher stability in the presence of Tween 20 in comparison to Triton-X 100.

At 10% (v/v) glycerol, 3L maintains 38% of its activity, this increases to 51% at 20% (v/v) with activity decreasing again to 36% at 30% (v/v) and increasing to 56% at the final 40% (v/v) concentration. Across the DMSO concentration gradient, a proportional decline in activity is observed at each concentration point, with 3L exhibiting 73% relative activity in the presence of 10% DMSO, with the relative activity decreasing by $\approx 10\%$ at each concentration level to a final relative activity of 30% at the 40% (v/v) concentration. Observing the assay results, it is clear that 3L exhibits a greater and more linear tolerance to DMSO, despite observing a higher activity in 40% (v/v) glycerol in comparison to DMSO.

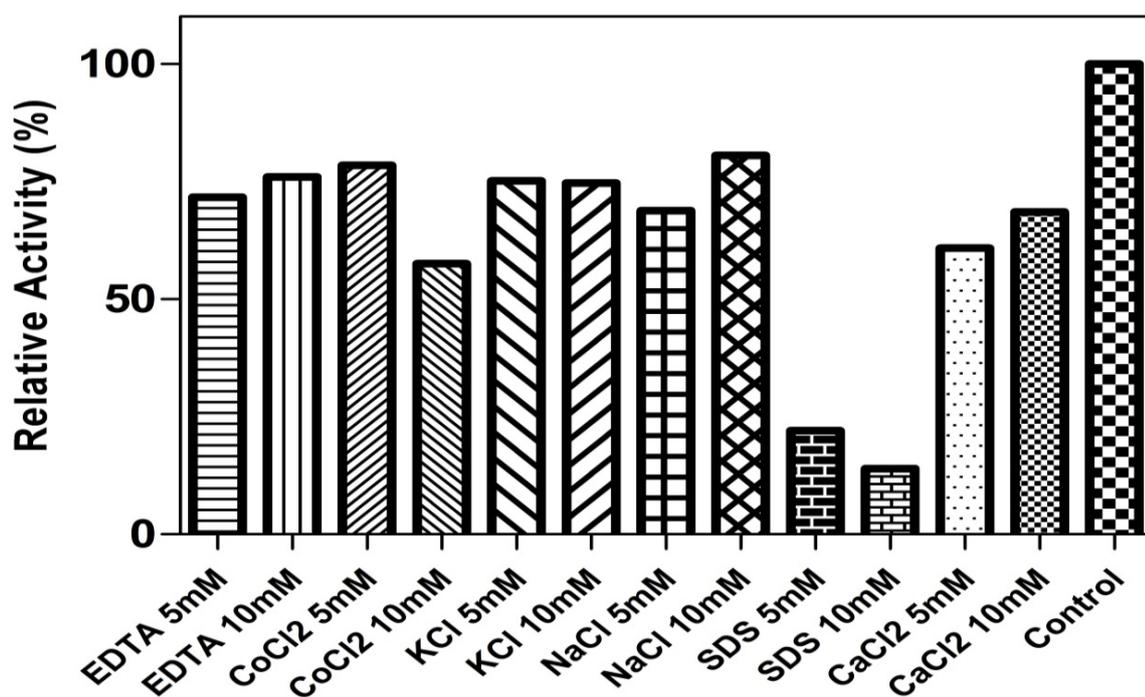


Figure 11: The relative activity of 3L was determined in the presence of supplemented ions, chelating agents, and detergents. Incubations were conducted using 50 mM cellobiose (10 μL in the reaction) 40 μL crude enzyme extract, 50 μL 100 mM HEPES-KOH pH 8.0, x grams of each reagent, at 40°C for 60 minutes.

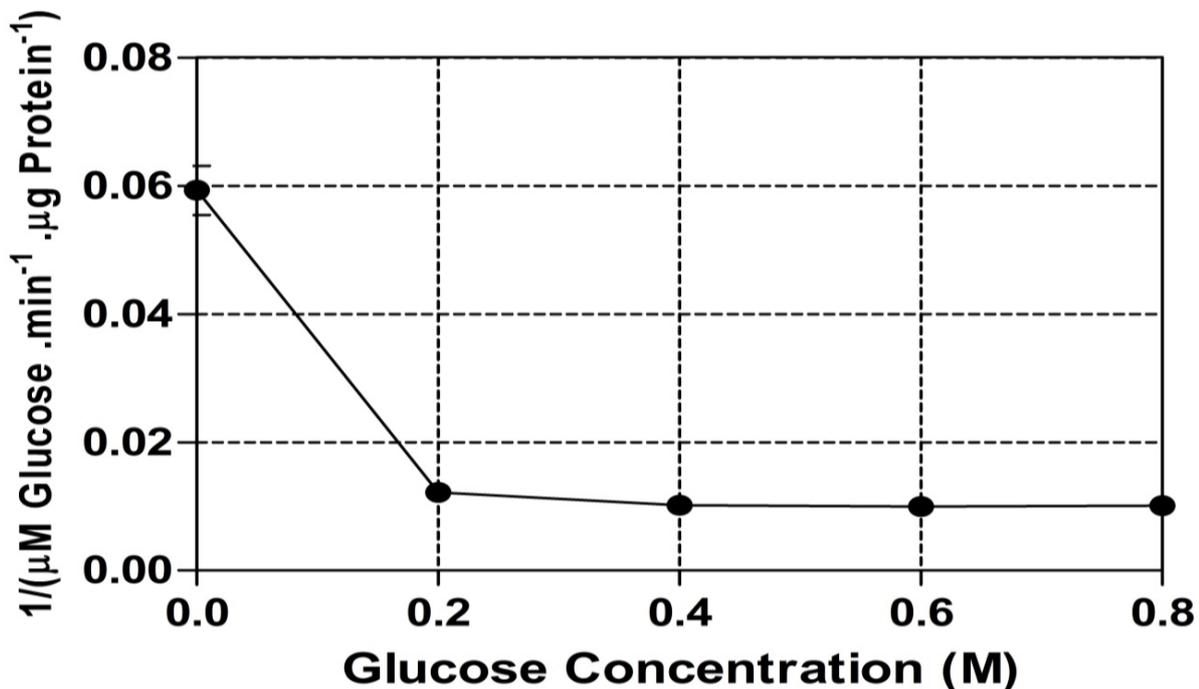


Figure 12: The glucose tolerance of clone 3L was assessed across a 0.0 – 0.8M glucose concentration gradient in 200mM increments. Incubations were conducted using 50 mM cellobiose (10 μL in reaction), 40 μL crude enzyme extract, a volume of 100 mM HEPES-KOH buffer pH 8.0 that varied to accommodate the addition of a 3 M glucose stock to adjust the final reaction concentration, at 40°C for 60 minutes.

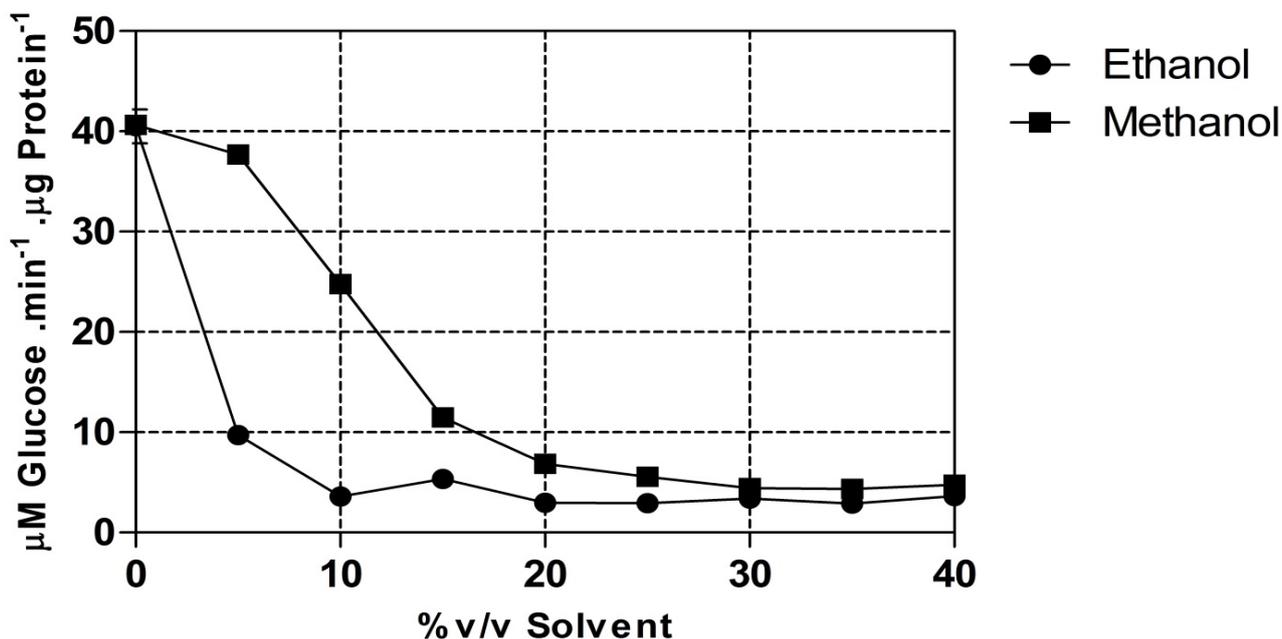


Figure 13: The tolerance of 3L to ethanol and methanol was ascertained across a 0 – 40% (v/v) concentration gradient in 5% increments. Incubations were conducted using 50 mM cellobiose (10 μL in reaction), 40 μL crude enzyme extract, a volume of 100 mM HEPES-KOH pH 8.0 that was varied with volumes of methanol and ethanol to adjust to the final reaction concentrations at 40°C for 60 minutes.

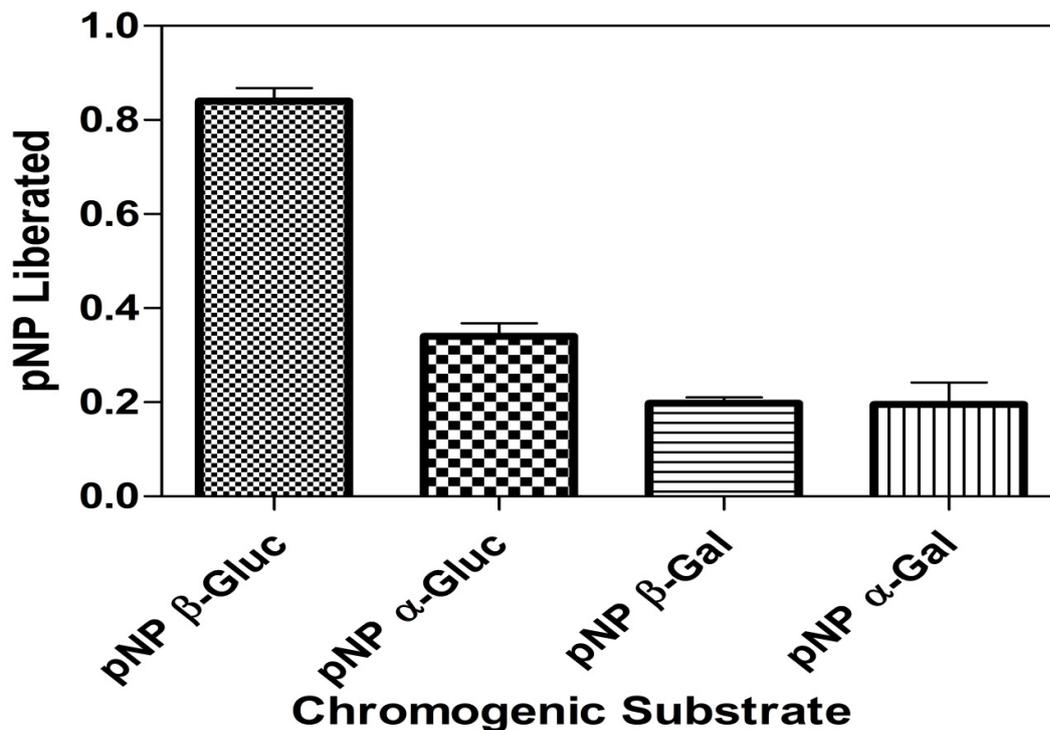


Figure 14: Chromogenic substrates were utilized to ascertain the bond preference of clone 3L, namely; pNP β-D-glucopyranoside, pNP α-D-glucopyranoside, pNP β-D-galactopyranoside, and pNP α-D-galactopyranoside. Incubations were conducted using 10 μL of 2 mg/mL chromogenic substrate, 40 μL crude enzyme extract, 50 μL 100 mM HEPES-KOH pH 8.0, at 40°C for 60 minutes.

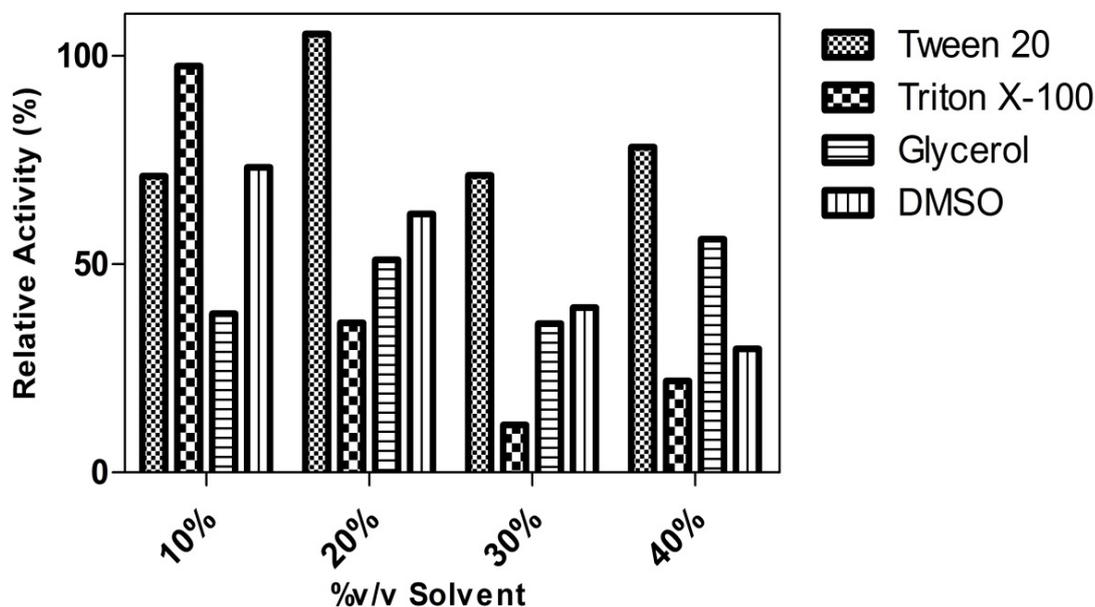


Figure 15: The resilience of clone 3L in the presence of industrial solvents and surfactants was determined across a 10 – 40% concentration range of Tween 20, Triton-X 100, Glycerol, and DMSO. Incubations were conducted using 50 mM cellobiose (10 μL in reaction), 40 μL crude enzyme extract, a volume of 100 mM HEPES-KOH pH 8.0 buffer that varied in accordance to the volume of surfactants/solvents, and at 40°C for 60 minutes.

4. Discussion

Functional metagenomic strategies include the identification of “clones” on the basis of phenotype (such as enzymatic activity), heterologous complementation and detection by reporter genes (Simon & Daniel, 2009). Consequently, functional metagenomics does not rely on prior knowledge of sequence information to detect enzymes and thus holds great potential to discover novel enzymes that may be applicable to industry. A recent study reported the discovery of multiple novel β -galactosidase genes from a single functional metagenomics approach (Cheng *et al.*, 2017). Despite β -galactosidases being one of the most well characterized enzymes, that study found 19 distinct clones, which showed β -galactosidase activities and presented no obvious sequence similarity to β -galactosidases. Following extensive bioinformatics analyses and structural modeling, the study reported on the discovery of three previously unknown β -galactosidase families. Such examples reinforce the value of functional metagenomics approaches and their use in isolating novel genes that could not have been predicted from DNA sequence analysis alone.

In our study, a metagenomic library was constructed from the runoff at a dairy farm and clones were functionally screened for β -galactosidase activity in an *E. coli* Δ LacZ mutant background (null mutation leading to β -galactosidase deficiency). Following sequencing and analyses of positive clones, we identified 3L as containing a potentially unique/novel ORF. *In silico* analyses of peptide sequence using ExPASy PROSITE domain finder suggested the presence of two putative active domains (both glycosyl hydrolase family 1) encoding for both β -galactosidase and β -glucosidase (Figure 3, Chapter 3).

What can be ascertained is that the domain analysis seems to indicate that both domains are family 1 glycoside hydrolases, according to their amino acid sequences. Functionally, that is not the case, as the plate-based assays (Figures 3 – 7, Chapter 3) clearly display both β -glucosidase and β -galactosidase activity. β -glucosidases form a part of the GH1 family, however, β -galactosidases are defined as GH2 enzymes, according to their CAZy definition. This is an unprecedented result, as both domains appear to belong the GH1 family when using *in silico* analyses, but have different functional enzymatic mechanisms. Delving deeper into the relationship of the two domains will not be discussed in this study, as site-directed mutagenesis (SDM) approaches would have to be executed in order to observe the effect that abolishing activity in either domain would have on the overall behavior of clone 3L.

Whilst 3L is an interesting metagenomic clone, its duality can be explained by the nature of the site sampled for the library. 3L was derived from a library created by sampling the run-off slurry from a local dairy farm in Stellenbosch. Within the sampled microbiome, the most abundant sources of carbon for the microbial community would be lactose from the dairy itself and cellulose from the cow's regurgitated cud (Hess *et al.*, 2011). This speaks to a high degree of causality between the available carbon sources and

predominant enzyme activities identified in the library and, by extension, clone 3L. The β -glucosidase activity is most likely due to high presence of cellobiose and partially digested cellulose present in the cud and the high concentration of lactose is evidence for the β -galactosidase activity identified within the clone 3L (Qi *et al.*, 2017). There are reported instances of multifunctional enzymes that display both β -galactosidase and β -glucosidase activities being discovered within metagenomic libraries derived from bovine rumen samples (Xing *et al.*, 2012), indicating that within microbial populations, the spatial arrangement of two protein coding domains that share both synergistic and mutualistic interactions is commonplace.

The rationale behind all means of enzyme characterization within this study was to (i) comprehensively determine the optimum conditions and kinetics of the β -glucosidase domain within clone 3L, and (ii) assess the β -glucosidase activity to elucidate any advantageous characteristics with regards to 2nd generation biofuel production pipelines. The 2nd generation biofuel industry requires enzymes that possess thermotolerance, glucose tolerance, methanol and ethanotolerance, resistance to extreme pH levels (pHs below 5.0 and exceeding 8.0) and industrial solvents, while maintaining high activity and affinity for its substrate. Novel gene discovery approaches, such as metagenomic libraries can address these needs by identifying environmental niches that are conducive to the proliferation of unculturable microorganisms that produce enzymes with novel and desirable traits.

All enzymatic assays were performed using 50 mM cellobiose as a substrate, unless otherwise stated. In order to determine the temperature optimum for clone 3L, a set of incubations were conducted across a temperature gradient from 20 to 70°C. Within the realm of biofuel production, there are two desired traits pertaining to the temperature at which an enzyme is active. A high activity at a low temperature (temperatures below 30°C) is desirable in order to liberate a large concentration of product from the substrate without having to apply a large amount of energy to heat up the reaction. Conversely, enzymes that display a high activity at high temperatures (Temperatures exceeding 40°C) are also desirable for a multiplicity of reasons. A thermostable enzyme would be more resistant to the heat pretreatment of the lignocellulosic biomass, and thus can be indiscriminately added to the reaction at any given point, simplifying the overall reaction scheme (Salim *et al.*, 2015). Furthermore, these thermostable enzymes can immediately begin to liberate shorter-chain cellooligosaccharides, cellobiose, and fermentable glucose from the lignocellulosic biomass during the heat pretreatment phase, thus shortening the overall lignocellulosic biomass degradation and fermentation cycle.

The optimal temperature for clone 3L is at 40°C (Fig. 8, Chapter 3), with a peak activity of 56.85 $\mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$, with activity increasing gradually from 20 to 40°C. Activity decreased 4-fold at temperatures exceeding 40°C and was almost abolished at 70°C. This result is unsurprising given the origin of the sample used for the metagenomic library; the dairy run off mainly consists of spillover created during the bovine milking process. Thus the majority of microbial enzymes recovered in the

library would be adapted to a mammalian physiological temperature, 37°C. Metagenomically isolated β -glucosidases characterized in other studies have been found to possess similar degrees of thermostability (Zhang *et al.*, 2016). Whilst a high degree of thermostability would be preferable within current 2nd gen biofuel platforms, it has been discovered that high temperatures during the lignocellulosic pretreatment phase can form macroradicals that inhibit the successful saccharification of lignocellulosic biomass (Sindhu *et al.*, 2016). These findings show how rapidly the biofuels industry is evolving; as little as two years ago, thermostability when exposed to temperatures exceeding 80°C was still a highly sought trait for enzymes within the realm of biofuels (Salim *et al.*, 2015). Taking this into account, 3L's modest temperature optimum of 40°C may be sufficient as the biofuels industry undergoes this paradigm shift, abandoning high-temperature pretreatment schemes.

Moving on to the optimum pH for clone 3L, a pH range of 3.0 to 10.0 was selected in order to gauge the kinetics of 3L across as wide a pH range possible (Fig. 9, Chapter 3), whilst still being a realistic range when compared to the potential real-world use case scenarios. During the pretreatment phase of the lignocellulosic material, a variety of organic solvents are added to the material in order to aid the degradation of lignocellulose, hemicellulose, pectins, and lignin within the plant cell wall matrix (Amnuaycheewa *et al.*, 2016). This degradation step creates areas within the lignocellulosic material wherein endoglucanases are able to successfully bind to the crystalline cellulose structures and cleave them into shorter-chain cello-oligosaccharides. These organic solvents have the potential to cause fluctuations to the overall pH within the reaction, meaning that the enzymes utilized within the pretreatment phase must be able to withstand pHs ranging from 3.0 to 10.0.

Thus, it is clear that for any enzyme to be applicable within the scope of lignocellulosic biomass degradation, it would need to maintain its activity across a broad pH range. The pH optimum assay (Fig. 9, Chapter 3) was conducted with both Δ LacZ::3L and the untransformed control, the empty Δ LacZ strain. The untransformed control exhibits a negligible degree of activity across the pH range; this can be attributed to background absorbance from the reagents in the kit, or a very small degree of cellobiose hydrolysis during the heat-inactivation step of the assay incubation. Clone 3L displays an activity of $\approx 38.90 \mu\text{M glucose} \cdot \text{m}^{-1} \cdot \mu\text{g protein}^{-1}$ across the pH 6.0 – 9.0 range, with the peak activity at pH 8.0, and negligible activity displayed below pH 6.0 and above pH 9.0. The optimum pH of 3L is once again closely influenced by the environment the sample was isolated from, whilst a more acidic pH could have been expected due to the nature of milk being slightly acidic, subsequent breakdown of proteins within the milk run off and exposure to the outside air may have brought about a slightly more alkaline environment. What is interesting is that 3L displays peak activity across a broad pH range, which differs greatly from most enzymes that tend to have a more defined pH optimum. This makes the case for the suitability of the 3L within a lignocellulosic biomass degradation pipeline, as the enzyme would be recalcitrant to denaturation should large fluxes in pH occur during the degradation process. Interestingly, similar findings have been reported by (Xing *et al.*, 2012), wherein a β -glucosidase enzyme was

discovered to display a wide pH optimum, ranging from pH 7.0 – 9.0. This is a positive sign for the advocacy of metagenomics and functional, plate-based screening methods as an effective means to discover novel enzymes with abnormal but advantageous characteristics to the the biofuels production industry.

The kinetic parameters of 3L were ascertained using an array of incubations utilizing varying concentrations of cellobiose as substrate. (Fig. 10, Chapter 3) shows the curve obtained from the kinetic analysis, displaying enzymatic activity across a 0 – 10 mM cellobiose concentration range. The line of best fit in this case is a sigmoidal curve instead of a more traditional straight-line regression model; this is due to the inherent enzymatic mechanism of 3L. The results of the kinetic analysis of 3L indicates that 3L can be classed as an allosterically regulated enzyme, meaning that the mechanism of feedback inhibition is regulated by a ligand binding to a non-catalytic site on the protein. This binding event triggers a conformational change within the protein, altering the bond angles between amino acid residues at the catalytic site. Findings by (Yang *et al.*, 2015) corroborate the ideas behind this mechanism, as their study found that β -glucosidases that have been mutated at their entrance and middle substrate channels display a lower sensitivity to glucose inhibition. The mechanism of regulation is still unclear, it does not make sense that glucose would not be able to bind to the active site, but molecules containing glucose are able to move through the narrow entrance channel and bind with the active site. What can be concluded, however, despite this irregularity, is that GH1 enzymes display sensitivity to glucose as an allosterically-bound regulator that forms part of a negative feedback inhibition mechanism.

Allosterically regulated enzymes do not have a linear relationship between substrate and product, especially β -glucosidases such as 3L. Most naturally occurring (non-engineered) β -glucosidases (or cellobiohydrolases) are allosterically regulated by glucose, their product (Ali *et al.*, 2016). This plays an important physiological role within the organisms that play host to the enzyme, as once a sufficient concentration of glucose has been generated; the product itself is able to regulate the enzyme that produces it, giving the organism precise control of its glucose levels. Taking into account the non-linear nature of β -glucosidase enzymes, the kinetic properties of 3L were elucidated by means of a sigmoidal regression line ($R^2 = 0.9524$). 3L was discovered to possess a K_m of 1.82 μ M cellobiose and a V_{max} of 74.98 μ M glucose. $m^{-1} \cdot \mu$ g protein $^{-1}$. 3L displays a high specific activity in comparison to other β -glucosidases in literature (Yang *et al.*, 2015) and a relatively low K_m . This translates to 3L being a potentially useful enzyme for use in biofuel production schemes. The low K_m value indicates that 3L has a high-affinity for its substrate, coupled with the ability to rapidly degrade it into glucose.

Findings by (Andreini *et al.*, 2008) indicate that many enzymes of various EC classifications require metals and cofactors in order to gain maximal activity, with 39% of hydrolases requiring cofactors to aid activity. In this case the activity of clone 3L was tested with the addition of various common ions, chelating agents, and detergents added to the incubation at both 5 mM and 10 mM concentrations (Fig.

11, Chapter 3). The addition of EDTA to the incubations was used to gauge the activity of 3L in an incubation starved of divalent ions, as EDTA acts as a chelating agent, binding divalent cations (Peng *et al.*, 2017). An incubation conducted with the addition of SDS was incorporated into the graphs in order to display the effect that an added surfactant would have on the activity of 3L. More common sources of divalent cations such as ZnCl_2 , etc, were not added to the assay as the protein extraction buffer used in this study contains ZnCl_2 (See Chapter 2). The addition of various ion sources had an unfavourable effect on the activity of 3L at both concentration points when compared to the control value. The addition of EDTA as a chelating agent also resulted in a decrease in activity; however, the decrease in activity lower than anticipated. This putatively indicates to 3L being able to retain a relatively high degree of activity within the presence of a chelating agent. This is an interesting result in of itself as the pretreatment process may introduce chelating agents into the reaction mix via the solvents added during the pretreatment phase, thus 3L would be able to maintain activity and remain able to yield fermentable glucose. As can be expected, the addition of a detergent resulted a reduction in activity indicating that 3L will denature to some extent when exposed to a detergent.

Owing to the fact that 3L plays host to two discrete protein coding domains and acts as a bifunctional enzyme, it was prudent to verify the true bond specificity of 3L (Fig. 14, Chapter 3) displays the degree of pNP liberated from 4 chromogenic substrates, namely; pNP β -D-glucopyranoside, pNP α -D-glucopyranoside, pNP β -D-galactoside, and pNP α -D-galactoside. The highest activity was observed upon the pNP β -1,4-D-glucopyranoside substrate, irrefutably demonstrating that 3L can be classed as a β -1,4-cellobiohydrolytic enzyme (EC 3.2.1.21).

Glucose tolerance is a valuable trait that has been identified with regards to biofuel-specific β -glucosidases (Peng *et al.*, 2017). As previously discussed, β -glucosidases are most commonly allosterically regulated via a negative feedback loop, with glucose acting as the ligand, binding to the protein and creating a conformational shift that abolishes activity by distorting the shape of the active site (Yang *et al.*, 2015). From a bioethanol production standpoint, this is an obstacle to yielding the maximum amount of bioethanol in 2nd generation biofuel pipelines. Glucose monomers are required for the yeast fermentation process. As glucose is liberated from cellobiose by the β -glucosidase enzymes, a portion of the liberated monomers will not remain in solution, but will instead bind to the non-catalytic site on the enzyme and deactivate it. This results in not only a reduction of freely available glucose monomers for fermentation, but also an overall decrease in enzyme activity that results in a reduction of glucose monomers being liberated from cellobiose.

The high concentration of glucose within the 2nd generation biofuel production reaction means that it is important to elucidate the ability of a β -glucosidase enzyme to maintain activity in the presence of glucose in order to determine its potential feasibility within the realm of bioethanol production. (Fig. 12, Chapter 3) depicts the activity of 3L when exposed to varying concentrations of glucose, 0.0 – 0.8 M, in

0.2 M increments. A reduction in activity can be observed between 0 and 0.2 M, with the activity decreasing to a baseline level for the remaining glucose concentration levels. These results indicate that 3L has a high degree of glucose tolerance, with a concentration of 0.2 M required to bring about a near 100% reduction in activity. General assays conducted to ascertain the glucose tolerance of a β -glucosidase do not exceed a 5 mM concentration of glucose, with most β -glucosidases losing 25% of their activity in the presence of 5 mM glucose (Peng *et al.*, 2017). 3L experiences \approx 25% reduction in activity at \approx 50 mM glucose, indicating that 3L possesses at least a 10-fold higher degree of glucose tolerance when compared to other metagenomically-isolated β -glucosidase enzymes. However, the glucose tolerance of 3L is significantly lower than certain fungal-derived β -glucosidases that display glucose tolerance up 0.8 M (Yang *et al.*, 2015). 3L possesses a 4-fold reduction in glucose tolerance in the presence of 0.8 M glucose in comparison.

The next characteristic of 3L that must be evaluated is the overall stability of the enzyme in varying concentrations of ethanol and methanol. Ethanol is produced in high concentrations in consolidated bioprocessing (CPB) and batch fermentation processes, high yielding fermentation reactions can yield a 9.5% (w/v) ethanol yield after a 60 hour fermentation run (Wang *et al.*, 2016) and methanol is regularly added to the pretreatment process of lignocellulosic biomass in order to aid the saccharification of the plant material (Shirkavand *et al.*, 2016). Ethanol can also potentially act as inhibitor for most cellulolytic activities within the fermentation process. Concentrations for both ranged from 0 – 40% (v/v) in 5% increments. Referring to (Fig. 13, Chapter 3), it can be seen that 3L displays an increased tolerance to methanol, retaining over 50% of its activity even at a 10% (v/v) methanol concentration. Conversely, 3L loses \approx 75% of its activity in the presence of 5% (v/v) ethanol. These observations outline that 3L possesses characteristics that result in the enzyme being well suited to a bioethanol production platform, due to the ubiquitous use of methanol during the pretreatment of the lignocellulosic biomass. 3L experiences an extreme reduction in activity in the presence of only 5% (v/v) ethanol, thus one would assume that 3L does not possess an adequate enough degree of ethanol tolerance, taking into account that the maximal ethanol production rates within medium sized batch fermentations can reach 9.5% (w/v) ethanol after 60 hours. It must be taken into account, however, that despite the total ethanol yield exceeding 3L's ethanol tolerance, the accumulation of ethanol per hour is only 0.32% (w/v) (Wang *et al.*, 2016). Thus, 3L would remain active over a period of \approx 15.5 hours before losing activity and needing to be replenished.

Following on from the previous set of incubations the activity of 3L was ascertained when exposed to varying concentrations of industrial solvents and surfactants (Fig. 15, Chapter 3). Glycerol and DMSO are universally utilized as solvents for various reagents; glycerol in particular is widely used in the biofuels industry in a process known as glycerol thermal processing (GTP). GTP is involved in the lignocellulosic biomass pretreatment phase, wherein glycerol is added to the heated biomass in order to act as an extruding reagent, disrupting the structures within the lignocellulosic biomass in order to aid

saccharification and secure a higher yield of fermentable monomers (Zhang *et al.*, 2015). Surfactants such as Tween 20 and Triton X-100 are commonly used in the lignocellulosic biomass pretreatment process in order to disrupt the plant cell wall and aid in the saccharification of the cellulose and hemicellulose layers (Sindhu *et al.*, 2015). As shown in (Fig 15, Chapter 3), 3L is able to maintain between 70 and 100% of its activity across a 10 – 40% (v/v) Tween 20 gradient. 3L maintains 97% of its activity in the presence of 10% (v/v) Triton X-100, however, activity decreases rapidly to below 50% at higher concentrations.

3L maintains between 38 and 55% of its activity in the presence of glycerol, however it is more resistant to the presence of DMSO, maintaining 73% of its activity when exposed to 10% (v/v) DMSO. 3L only exhibits a loss of activity greater than 50 % in the presence of 30% (v/v) DMSO. Whilst 3L is adversely affected by some of the main reagents utilized in a bioethanol production platform, there are certain reagents that 3L displays a large degree of recalcitrance towards.

Conclusion & Future Work

Taking into account the results of this study, 3L shows promise within the biofuel production realm. 3L displays a good degree of activity on its natural substrate, coupled with a high affinity for its substrate. The *in vitro* results generated by the incubations display that 3L is a very promising β -glucosidase that could, putatively, form part of a bioethanol production platform. 3L possesses average thermostability, high glucose tolerance for a metagenomically-isolated enzyme, and varying degrees of resistance to industrial solvents, surfactants, and ethanol. Additionally, 3L shows above average methanol tolerance and a wide active pH range, in its current state it is very feasible within the realm of bioethanol production.

Future experimentation with regards to 3L will revolve around site-directed mutagenesis approaches to engineer the protein. The β -gal active site has been found within the downstream domain within the ORF and will be targeted and mutated to abolish β -gal activity. This will allow us to identify which domain is responsible for which enzymatic activity. The 1st domain within the ORF is suspected to be a cellulose-binding domain (CBD) due to the result from the PROSITE analysis that the upstream domain carries an N-terminal signature, which is usually associated with binding domains, If this is true it would mean that the downstream domain is responsible for both β -glucosidase and β -galactosidase activity, which is uncommon and could increase the industrial value of 3L. Furthermore, if the upstream domain proves to be a CBD with high affinity for cellulose, the CBD can be amplified and engineered upstream of a highly active cellulase that may lack an active CBD. This would be of great value to the 2nd generation biofuel industry, as one barrier to the effective degradation of the lignocellulosic biomass is the inability for cellulases to bind efficiently to the long-chain crystalline cellulose structures within the cell wall matrix.

5. References

- Adeyemo, O. and Wise, R. (2010). Bio-fuels and Food Security: A Case Study of South Africa. *Africa Insight*, 39(3).
- Alaswad, A., Dassisti, M., Prescott, T. and Olabi, A. (2015). Technologies and developments of third generation biofuel production. *Renewable and Sustainable Energy Reviews*, 51, pp.1446-1460.
- Ali, S., Nugent, B., Mullins, E. and Doohan, F. (2016). Fungal-mediated consolidated bioprocessing: the potential of *Fusarium oxysporum* for the lignocellulosic ethanol industry. *AMB Express*, 6(1).pp?
- Amnuaycheewa, P., Hengaroonprasan, R., Rattanaporn, K., Kirdponpattara, S., Cheenkachorn, K. and Sriariyanun, M. (2016). Enhancing enzymatic hydrolysis and biogas production from rice straw by pretreatment with organic acids. *Industrial Crops and Products*, 87, pp.247-254.
- Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. and Thornton, J. (2008). Metal ions in biological catalysis: from enzyme databases to general principles. *JBIC Journal of Biological Inorganic Chemistry*, 13(8), pp.1205-1218.
- Aro, E. (2015). From first generation biofuels to advanced solar biofuels. *Ambio*, 45(S1), pp.24-31.
- Blanchard, R., Richardson, D., O'Farrell, P. and Von Maltitz, G. (2011). Biofuels and biodiversity in South Africa. *South African Journal of Science*, 107(5/6).pp?
- Brief Overview of Biofuel Production Workflows. (2017). [image] Available at: <http://ars.els-cdn.com/content/image/1-s2.0-S1364032114010764-gr1.jpg> [Accessed 27 Sep. 2017].
- Chen, G., Zhao, L. and Qi, Y. (2015). Enhancing the productivity of microalgae cultivated in wastewater toward biofuel production: A critical review. *Applied Energy*, 137, pp.282-291.
- Chen, X., Xiao, Y., Shen, W., Govender, A., Zhang, L., Fan, Y. and Wang, Z. (2015). Display of phytase on the cell surface of *Saccharomyces cerevisiae* to degrade phytate phosphorus and improve bioethanol production. *Applied Microbiology and Biotechnology*, 100(5), pp.2449-2458.
- Cloete, P. and Idsardi, E. (2012). Bio-fuels and Food Security in South Africa: The Role of Indigenous and Traditional Food Crops. In: *International Association of Agricultural Economists (IAAE) Triennial Conference*. Potchefstroom, South Africa: Unit for Environmental Sciences and Management, North-West University, pp.1-30.

Hess, M., Sczyrba, A., Egan, R., Kim, T., Chokhawala, H., Schroth, G., Luo, S., Clark, D., Chen, F., Zhang, T., Mackie, R., Pennacchio, L., Tringe, S., Visel, A., Woyke, T., Wang, Z. and Rubin, E. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science*, 331(6016), pp.463-467.

Lü, J., Sheahan, C. and Fu, P. (2011). Metabolic engineering of algae for fourth generation biofuels production. *Energy & Environmental Science*, 4(7), p.2451.

Maity, J., Bundschuh, J., Chen, C. and Bhattacharya, P. (2014). Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives – A mini review. *Energy*, 78, pp.104-113.

Mirete, S., Morgante, V. and González-Pastor, J. (2016). Functional metagenomics of extreme environments. *Current Opinion in Biotechnology*, 38, pp.143-149.

Mohr, A. and Raman, S. (2013). Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. *Energy Policy*, 63, pp.114-122.

Peng, S., Cao, Q., Qin, Y., Li, X., Liu, G. and Qu, Y. (2017). An aldono-lactonase AltA from *Penicillium oxalicum* mitigates the inhibition of β -glucosidase during lignocellulose biodegradation. *Applied Microbiology and Biotechnology*, 101(9), pp.3627-3636.

Puspawati, S., Wagiman, Ainuri, M., Nugraha, D. and Haslianti (2015). The Production of Bioethanol Fermentation Substrate from *Eucheuma cottonii* Seaweed through Hydrolysis by Cellulose Enzyme. *Agriculture and Agricultural Science Procedia*, 3, pp.200-205.

Qi, T., Gu, G., Xu, L., Xiao, M. and Lu, L. (2017). Efficient synthesis of tyrosol galactosides by the β -galactosidase from *Enterobacter cloacae* B5. *Applied Microbiology and Biotechnology*, 101(12), pp.4995-5003.

Salim, T., Ratnawati, L., Agustina, W. and Sriharti (2015). Bioethanol Production from Glucose by Thermophilic Microbes from Ciater Hot Springs. *Procedia Chemistry*, (16), pp.503-510.

Shirkavand, E., Baroutian, S., Gapes, D. and Young, B. (2016). Combination of fungal and physicochemical processes for lignocellulosic biomass pretreatment – A review. *Renewable and Sustainable Energy Reviews*, 54, pp.217-234.

- Simon, C. and Daniel, R. (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology and Biotechnology*, 85(2), pp.265-276.
- Sims, R., Mabee, W., Saddler, J. and Taylor, M. (2010). An overview of second generation biofuel technologies. *Bioresource Technology*, 101(6), pp.1570-1580.
- Sindhu, R., Binod, P. and Pandey, A. (2016). Biological pretreatment of lignocellulosic biomass – An overview. *Bioresource Technology*, 199, pp.76-82.
- Sindhu, R., Gnansounou, E., Pandey, A. and Binod, P. (2015). A novel crude glycerol assisted surfactant pretreatment strategy of chili post-harvest residue for bioethanol production. *Biofuels*, 6(5-6), pp.383-390.
- Sindhu, R., Kuttiraja, M., Prabisha, T., Binod, P., Sukumaran, R. and Pandey, A. (2016). Development of a combined pretreatment and hydrolysis strategy of rice straw for the production of bioethanol and biopolymer. *Bioresource Technology*, 215, pp.110-116.
- Singh, R. and Srivastava, M. (2016). Regional Renewable Energy India: Bioethanol From Rice Straw. *Current Sustainable/Renewable Energy Reports*, 3(3-4), pp.53-57.
- Takagi, T., Sasaki, Y., Motone, K., Shibata, T., Tanaka, R., Miyake, H., Mori, T., Kuroda, K. and Ueda, M. (2017). Construction of bioengineered yeast platform for direct bioethanol production from alginate and mannitol. *Applied Microbiology and Biotechnology*, 101(17), pp.6627-6636.
- Tanimura, K., Takashima, S., Matsumoto, T., Tanaka, T. and Kondo, A. (2016). 2,3-Butanediol production from cellobiose using exogenous beta-glucosidase-expressing *Bacillus subtilis*. *Applied Microbiology and Biotechnology*, 100(13), pp.5781-5789.
- The South African Department of Science and Technology (DST) (2017). *The BioEnergy Atlas for South Africa*. South African Environmental Observation Network (SAEON), pp.1-122.
- Tsuchida, J. (2017). *Structure of the plant cell wall - A Cross section of the bagasse cell wall*. [image] Available at: https://www.researchgate.net/profile/Jefferson_Tsuchida/publication/265994893/figure/fig12/AS:271767541383198@1441805839864/Figure-12-Structure-of-the-plant-cell-wall-a-Cross-section-of-the-bagasse-cell-wall.png [Accessed 27 Sep. 2017].

- Wang, D., Li, F. and Wang, S. (2016). *Engineering a natural Saccharomyces cerevisiae strain for ethanol production from inulin by consolidated bioprocessing*. 9th ed. Qindao: BioMed Central Ltd, p.96.
- Wu, W., Davis, R., Tran-Gyamfi, M., Kuo, A., LaButti, K., Mihaltcheva, S., Hundley, H., Chovatia, M., Lindquist, E., Barry, K., Grigoriev, I., Henrissat, B. and Gladden, J. (2017). Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. *Applied Microbiology and Biotechnology*, 101(6), pp.2603-2618.
- Xing, M., Zhang, X. and Huang, H. (2012). Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnology Advances*, 30(4), pp.920-929.
- Yang, F., Yang, X., Li, Z., Du, C., Wang, J. and Li, S. (2015). Overexpression and characterization of a glucose-tolerant β -glucosidase from *T. aotearoense* with high specific activity for cellobiose. *Applied Microbiology and Biotechnology*, 99(21), pp.8903-8915.
- Yang, Y., Zhang, X., Yin, Q., Fang, W., Fang, Z., Wang, X., Zhang, X. and Xiao, Y. (2015). A mechanism of glucose tolerance and stimulation of GH1 β -glucosidases. *Scientific Reports*, 5(1).
- Zhang, G., Liu, P., Zhang, L., Wei, W., Wang, X., Wei, D. and Wang, W. (2016). Bioprospecting metagenomics of a microbial community on cotton degradation: Mining for new glycoside hydrolases. *Journal of Biotechnology*, 234, pp.35-42.
- Zhang, W., Sathitsuksanoh, N., Barone, J. and Renneckar, S. (2015). Enhanced enzymatic saccharification of pretreated biomass using glycerol thermal processing (GTP). *Bioresource Technology*, (199), pp.148-154.

7. Appendix

7.1 Vector Map of the pRSETa::3L Construct

Created with SnapGene®

