

Biopolymer Gene Discovery and Characterization using Metagenomic Libraries

Colin Walter Ohlhoff

Thesis presented in partial fulfillment of the
requirements for the degree of

Master of Science

in

Plant Biotechnology

in the

Faculty of Science

Stellenbosch University

Dr JR Lloyd

Dr R Bauer

Prof JM Kossmann

December 2008

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 owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 19 November 2008

Abstract

Traditional methods used for the discovery of novel genes have previously relied upon the ability to culture the relevant microbes and then demonstrate the activity of a specific enzyme. Although these methods have proved successful in the past, they severely limit our access to the genomes of organisms which are not able to be cultured under laboratory conditions. It was therefore the aim of this project to use metagenomic strategies for the identification of novel polymer-producing genes with the prospect of commercial exploitation.

In this study, soil-derived metagenomic libraries were functionally screened for potential β -glucan producing clones using aniline blue staining. Positive reacting clones were selected and sequenced. Initial sequencing revealed a gene with high homology to previously described glucan synthases, the products of these genes all having significant industrial value. The clone was transformed into a suitable bacterial host, cultured and allowed to produce the polymer of interest. The polysaccharide was purified and subjected to various chemical analyses so as to confirm its monosaccharide composition. Data suggests that this polymer is composed mainly of glucose units and that it may be secreted out of the cell. Purification of the active enzyme was attempted using classical protein purification methods with faint activity being detected using Native polyacrylamide gel electrophoresis (PAGE). Further attempts to demonstrate activity were made through the construction of a GST (glutathione S-transferase) tagged fusion protein.

The second part of this study focuses on the construction and screening of a metagenomic DNA library from whey, a by-product of the cheese manufacturing process. It was envisaged that this could provide a resource for the identification of high value polymers when lactose is provided as a sole carbon source. The library was screened for function using Congo Red for the detection of extra-cellular polysaccharides.

Opsomming

In die verlede het tradisionele metodes vir die ontdekking van nuwe gene vertrou op die vermoë om die relevante mikrobies te kultiveer en dan die aktiwiteit van 'n spesifieke ensiem te demonstreer. Alhoewel die metodes in die verlede suksesvol was, beperk dit ons toegang tot die genoom van organismes wat nie onder laboratorium kondisies gekultiveer kan word nie. Die doel van die projek was dus om metagenomiese strategieë te gebruik om nuwe polimeer-vervaardigende gene te identifiseer met die doel vir kommersiële benutting.

In hierdie studie is metagenomiese biblioteke afkomstig van grond d.m.v. anilien blou kleuring geïsoleer vir potensieel β -glukaan produserende klone. Positiewe reagerende klone is geselekteer en die basispaaropeenvolging bepaal. Tydens die oorspronklike DNS basispaaropeenvolgingsbepaling is 'n geen met hoë homologie aan vorige gekarakteriseerde glukaan sintases gevind. Die produk van hierdie gene het almal belangrike industriële waarde. Die kloon is in 'n geskikte bakteriële gasheer getransformeer, gekultiveer en toegelaat om die polimeer van belang te vervaardig. Die polisakkaried is gesuiwer en aan verskeie chemiese toetse onderwerp om die monosakkariede samestelling te bevestig. Die data dui daarop dat die polimeer hoofsaaklik uit glukose eenhede bestaan en dat dit moontlik deur die sel uitgeskei word. Daar is gepoog om die aktiewe ensiem d.m.v. klassieke proteïen suiweringsmetodes te isoleer terwyl lae aktiwiteit op nie-denaturerende PAGE waargeneem kon word. Verdere pogings om aktiwiteit te demonstreer is d.m.v. die konstruksie van 'n GST-gemerkte fusie proteïen uitgevoer.

Die tweede deel van hierdie studie fokus op die konstruksie en assessering van 'n metagenomiese DNS biblioteek afkomstig van dikmelkwater, 'n nuwe produk van die kaas vervaardigingsproses. Die visie is dat dit 'n bron kan verskaf vir die identifisering van hoë waarde polimere wanneer laktose as die enigste koolstof bron verskaf word. Die biblioteek is geskandeer vir funksie deur Congo Rooi te gebruik vir die deteksie van ekstrasellulêre polisakkariede.

Acknowledgements

I am particularly grateful for the advice and guidance of my supervisors, Dr James Lloyd and Dr Rolene Bauer, without which this study would not have been possible.

Much gratitude is bestowed to Prof Jens Kossmann for providing me with the opportunity to study at the Institute for Plant Biotechnology as well as the valuable support he has provided during my study.

Many thanks are also due to Dr Jan Bekker, Stanton Hector and Fletcher Hiten for their advice and technical assistance throughout this study.

To all staff and students at the IPB, thank you for all your contributions in making it such a great place to work. The friendships made over the past few years are greatly valued.

Friends and family for support and love.

To my parents, for always setting an immaculate example and providing years of endless encouragement. This work is dedicated to you.

Table of Contents

Declaration	2
Abstract	3
Opsomming	4
Acknowledgements	5
Table of Contents	6
List of Abbreviations	10
List of Figures	13
List of Tables	14

Chapter 1

Introduction	15
---------------------------	----

Chapter 2

Review of Literature

2.1. Introduction.....	18
2.2. Metagenomics.....	18
2.3. Construction of metagenomic libraries.....	19
2.4. Industrial biocatalysts from the metagenome.....	20
2.5. Analysis of metagenomic libraries.....	23
2.5.1. Function-driven analysis.....	23
2.5.2. Sequence-driven analysis.....	25
2.6. Biopolymers.....	28
2.7. β -glucans.....	29

Chapter 3

Identification and Analysis of a Putative β -(1-4)-Glucan Synthase Isolated from a Metagenomic Library

3.1. Introduction.....	32
3.2. Materials and Methods.....	33
3.2.1. Organisms.....	33
3.2.1.1. <i>Escherichia coli</i> strains.....	33
3.2.2. Chemicals and kits.....	33
3.2.3. Plasmids.....	33
3.2.4. Recombinant DNA techniques.....	34
3.2.5. Transformation of DH5 α <i>E.coli</i>	34
3.2.6. Culture conditions.....	34
3.2.6.1. Solid media.....	34
3.2.6.2. Liquid media.....	34
3.2.6.3. Substrates.....	35
3.2.7. Staining for β -(1-4)-glucan polymers.....	35
3.2.8. Determination of cloned insert sizes.....	35
3.2.9. Polymer extraction from plate cultures.....	36
3.2.10. Polymer extraction from liquid cultures.....	36
3.2.11. Hydrolyzation of polymers.....	36
3.2.12. Sample derivitization for GC-MS.....	37
3.2.13. GC-MS monosaccharide analysis.....	37
3.2.14. Size exclusion chromatography.....	37
3.2.15. Phenol-H ₂ SO ₄ assay.....	38
3.2.16. DNS assay.....	38
3.2.17. Protein extraction.....	38
3.2.18. Protein quantification.....	39
3.2.19. Lysis of bacteria.....	39
3.2.20. Electrophoresis.....	39
3.2.20.1. SDS PAGE.....	39
3.2.20.2. Native PAGE.....	40

3.2.20.3. Staining of PAGE gels.....	40
3.2.21. PCR.....	40
3.2.21.1. Colony PCR.....	40
3.2.22. Sequencing of DNA.....	41
3.2.23. Protein purification.....	41
3.2.23.1. Crude protein extraction.....	41
3.2.23.2. Protein fusion construct using pGEX.....	42
3.2.23.3. Ammonium sulphate protein precipitation.....	42
3.2.23.4. Desalting of precipitated proteins.....	42
3.2.23.5. Schiff staining.....	43
3.2.23.6. Thrombin digestion of GST fusion protein.....	43
3.2.24. Determination of glucan synthase enzyme activity.....	43
3.3. Results and Discussion.....	44
3.3.1. Library screening.....	44
3.3.2. Hydropathy plot.....	46
3.3.3. Polymer investigation.....	49
3.3.3.1. Monosaccharide analysis by GC-MS.....	49
3.3.3.2. Size exclusion chromatography.....	50
3.3.3.3. Estimation of the number of glucose units.....	51
3.3.4. Activity measurements in crude extracts.....	52
3.3.5. Expression of a recombinant protein.....	56
3.3.6. Protein purification by affinity chromatography.....	57
3.3.7. Cleavage of the GST tag.....	58
3.3.8. Assay for activity determination of fusion protein.....	60
3.4. Conclusion.....	62

Chapter 4

Construction of a Metagenomic Library from Whey and Screening for Commercially Exploitable Polymers

4.1. Introduction.....	64
4.2. Materials and Methods.....	66

4.2.1. Sample collection.....	66
4.2.2. Organisms.....	66
4.2.2.1. <i>Escherichia coli</i> strains.....	66
4.2.3. Chemicals and kits.....	67
4.2.4. Transformation of competent <i>E.coli</i>	67
4.2.5. Genomic DNA extraction from whey.....	67
4.2.6. Partial digestion of genomic DNA.....	68
4.2.7. Construction of the genomic library.....	68
4.2.8. Library amplification.....	68
4.2.9. Mass excision.....	69
4.2.10. Library screening methods.....	69
4.3. Results and Discussion.....	70
4.3.1. Extraction of genomic DNA.....	70
4.3.2. Library construction and characterization.....	71
4.3.3. Functional screening of the library.....	72
4.3.4. Screening whey for exopolysaccharide producing bacteria...	75
4.4. Conclusion.....	76

Chapter 5

General Discussion and Conclusion.....	77
---	-----------

Chapter 6

Literature Cited.....	80
------------------------------	-----------

List of Abbreviations

°C	Degrees centigrade
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
dH ₂ O	Deionized water
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
EDTA	Ethylenediamine tetraacetic acid
EtOH	Ethanol
<i>g</i>	Gravitational force
GC/MS	Gas chromatography/Mass spectrometry
Glc	Glucose
GlcN	Glucosamine
GST	Glutathione-S-transferase
IPTG	Isopropyl- β -D-thiogalactopyranoside
kb	Kilo bases
kDa	Kilo daltons
LB	Luria broth

LBA	Luria broth with Ampicillin
L	Litre
μg	Microgram
μL	Microliter
mL	Milliliter
mm	Millimeter
mM	Millimolar
M	Molar
min	Minute
MSTFA	N-Methyl-N-(trimethylsilyl) trifluoroacetamide
<i>m/z</i>	Mass per charge
N	Normal
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre of Biotechnological Information
nm	Nanometer
NMR	Nuclear magnetic resonance
OD ₆₀₀	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis

PCR	Polymerase chain reaction
PEP	Phosphoenol pyruvate
pfu	Plaque forming unit
PMSF	Phenylmethylsulfonyl fluoride
rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium dodecyl sulfate
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
U	Enzyme units
UDP	Uridine diphosphate
V	Volts
vol	Volume
v/v	Volume per volume
w/v	Weight per volume
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

List of Figures

Figure 1.	Schematic diagram demonstrating the metagenomic approach to obtain novel biocatalysts.	27
Figure 2.	Structural formula for cellulose: a β -1,4-glucan polymer chain.	30
Figure 3.	Aniline blue functional screening of transformants.	45
Figure 4.	Kyte-Doolittle hydrophathy plot representing the peptide sequence obtained for the glucan synthase clone.	47
Figure 5.	Chromatogram of the hydrolyzed polymer sample.	49
Figure 6.	Phenol-sulphuric acid assay for detection of carbohydrate sugars in polymer extracts.	50
Figure 7.	DNS assay used to determine the amount of free glucose units.	52
Figure 8.	SDS-PAGE gel containing protein extract from the empty vector negative control and the glucan synthase clone.	53
Figure 9.	Native PAGE gel containing crude protein extract from the glucan synthase clone.	55
Figure 10.	The pGEX 4T-1::Glucan synthase fusion construct.	56
Figure 11.	Colony PCR to confirm the presence of the glucan synthase gene in the pGEX 4T-1 vector and purification of the GST fusion protein using glutathione agarose.	58
Figure 12.	Gradient SDS-PAGE gels of the GST-tagged fusion protein.	59
Figure 13.	Schematic diagram showing the formation of free NAD and L-lactate through a pyruvate kinase/lactate dehydrogenase coupled assay.	61
Figure 14.	Activity determination of the purified fusion protein following enzymatic cleavage of the GST tag.	62
Figure 15.	Purified genomic DNA extracted from the whey sample.	70
Figure 16.	<i>Sau3A1</i> partial enzymatic digestion of genomic DNA.	71

Figure 17.	Schematic diagram indicating the monosaccharides, D-galactose and D-glucose produced following the hydrolysis of lactose.	73
Figure 18.	Transformed clones following activity staining with Congo Red.	74

List of Tables

Table 1.	Examples of metagenomic discoveries based on the functional screening of libraries.	22
Table 2.	Oligonucleotides designed for amplification of the glucan synthase gene.	41
Table 3.	Nucleotide and predicted amino acid sequence obtained for the glucan synthase clone.	45
Table 4.	BLAST hits showing significant alignment with the glucan synthase clone sequence.	46

Chapter 1

Introduction

With the vast economic prospects that biotechnology offers modern society, various industries have different motivations to probe the enormous resource which uncultivated microbial diversity presents. Venter *et al.* (2004) reported that more than one million novel open reading frames, many of which encode putative enzymes, were identified in a single study which sampled marine prokaryotic plankton from the Sargasso Sea, thereby revealing an almost inexhaustible genetic resource for biomolecules with possible industrial value. Artificial or “man-made” products have been placed under scrutiny and are often treated with much skepticism by our ever health-conscious population. Customer satisfaction, along with an increasing demand, has therefore placed pressure on industry to find some alternative, natural means of formulating its products. These include vitamins, antibiotics and, in particular, novel biocatalysts for use in the production of flavors, agrochemicals, pharmaceuticals and high-value fine chemicals.

Metagenomics is a new, exciting field of research in which methods have been developed to exploit the genomes of previously “unculturable” microbes by unveiling functional genes with enormous biotechnological potential. The diverse array of metagenome-encoded enzymes was demonstrated in a study by Miller (2000). Total DNA was extracted from an alkaline desert sample, fragmented and cloned into a suitable expression vector. Whilst screening for lipase and esterase activity, 120 novel enzymes were discovered. These could further be classed into 21 protein families. This is not surprising given that a single organic soil sample was found to contain collective genomes which were the equivalent of 6000 to 10 000 *E.coli* genomes in size (Torsvik *et al.* 1998). This emphasizes the extent to which classical microbiological techniques are limited as far as the exploration of microbial populations is concerned, mostly because the vast majority of bacterial cells are considered to be uncultivable (Amann *et al.* 1995).

'Industrial' or *White Biotechnology* is a term which was coined by the European Association for Bioindustries (EuropaBio). Based on a case study report, it incorporates all industrially harnessed bio-based processes that are not covered by the *Red Biotechnology* (medical) or *Green Biotechnology* (plant) labels (Schepens *et al.* 2003). White biotechnology is the application of biotechnology for the processing and production of materials, chemicals and energy. Its roots can be traced back in ancient human history and its products are becoming more evident in everyday life. Medicines, vitamins, bioplastics, biofuels to bakery and dairy products, even enzymes in detergents all fall under this label.

Market analysts have indicated that white biotechnology has the potential to affect industrial production processes on a global scale. The main long-term goals include replacing fossil fuels with renewable resources, bioprocessing instead of conventional processes as well as creating new high-value bioproducts such as nutraceuticals, performance chemicals and bioactives (Lorenz & Eck. 2005). It has been predicted by the McKinsey consultancy that by 2010, between 10% and 20% of all chemicals sold could be produced through biotechnology (this amounts to approximately \$160 billion) and that about 60% of all fine chemicals (medium-volume products used as intermediates in the production of pharmaceuticals, flavors, fragrances etc.) could potentially be produced by biotechnology (Schepens *et al.* 2003).

Scientists are becoming increasingly aware of the endless possibilities which genetic access to our microbial diversity could provide. To convert metagenomic technologies into commercial success is of utmost importance. Metagenomics, coupled with *in vitro* evolution and high-throughput screening, presents industry with the opportunity to bring biomolecules into industrial application (Lorenz & Eck. 2005).

The aim of this study was two-fold. Firstly, to investigate a polymer synthesizing clone isolated from a soil-derived metagenomic library and secondly, to construct a genomic library from a selected environmental sample and screen for high-value biopolymer production. Chapter 2 provides an overview of the metagenomic process and its potential as a means of novel gene discovery. Screening methods and industrial prospects are also covered in this chapter. Chapter 3 focuses on a putative β -(1-4)-glucan synthase gene which was isolated from a soil-derived metagenomic library by functional screening. Attempts to purify and characterize the polymer, as well as the protein responsible for its synthesis, are discussed. Finally, a genomic library was constructed from whey extract and was functionally screened for clones able to produce galactan polymers when supplemented with the appropriate substrates (Chapter 4).

Chapter 2

Literature Review

2.1. Introduction

The dawn of the 21st century brought new challenges to the field of modern biotechnology, particularly with an ever increasing demand for novel biocatalysts. This sparked the development of various innovative technologies and methods to sustain the needs of this exciting industry. Although it was applied with limited success during the early to mid 1990's, metagenomics is now rapidly advancing the discovery of novel genes able to produce novel enzymes, antibiotics and biopolymers.

This overview discusses the use of metagenomic techniques as a direct means of accessing the genetic diversity of an environmental sample. The construction of metagenomic libraries as well as associated screening methods are highlighted. In addition, biopolymers and more specifically, β -glucan polymers, will be described.

2.2. Metagenomics

The word 'genomics' was originally used to describe a specific scientific discipline in genetics which incorporates the mapping, sequencing and analysis of genomes, where a genome refers to the complete set of genes and chromosomes in an organism (Xu, 2006). The term has since become more widely used both by the scientific community as well as the general public. In recent times, the use of genomics has expanded to such an extent that it has also be used for the functional analysis of entire genomes. These functional analytical aspects have now diversified to include whole genome RNA transcripts (transcriptomics), proteins (proteomics) and metabolites (metabolomics) (Xu, 2006). One newer field within genomics is 'metagenomics' which describes the study of collective genomes within an environmental community.

The use of traditional cultivation techniques for screening novel biocatalysts from isolated microorganisms presents numerous limitations in exploration of the vast genetic diversity of environmental microorganisms, largely because more than 99% of microbes present in environments cannot be cultured (Schloss and Handelsman, 2003). In fact, it is safe to say that most of the species in most environments have never been described, and this situation will not change unless new culture technologies come to the fore. Amann *et al* (1995) reported that 0.001-0.1% of the microorganisms in seawater, 0.25% in fresh water, 0.25% in sediments and 0.3% of soil microorganisms are currently cultivatable. A method to access the genomes of uncultured microorganisms involves the direct screening of novel biocatalysts from a metagenomic library. This involves the isolation of genomic DNA from communities of microbes isolated directly from ecosystems and then ligating them into appropriate vectors producing large insert libraries (Daniel, 2004). Metagenomics follows two lines of activity. The first uncovers novel enzymes and molecules for pharmaceutical and biotechnological applications while the second aims to generate new knowledge on the microbial ecology of the relevant niches (Streit *et al.* 2004).

2.3. Construction of metagenomic libraries

A metagenomic library should theoretically contain clones representing the entire genetic complement of a single habitat; however this is largely dependent on the efficiency of the DNA extraction and the cloning techniques. Contamination of purified DNA with polyphenolic compounds is a major difficulty associated with the metagenome approach, as these compounds are often co-purified with the DNA and are difficult to remove (Tsai & Olson, 1992). Following DNA isolation and purification, DNA libraries are prepared by using suitable cloning vectors and host strains. The classical approach involves the construction of small insert libraries (typically less than 10 kb) in a standard sequencing vector and using *Escherichia coli* as a host strain (Henne *et al.* 1999). Sometimes however, these small insert libraries prevent the detection of large gene clusters or operons. This limitation can be circumvented through the use of large insert libraries,

such as cosmid DNA libraries with average insert sizes ranging from 25-35 kb (Entcheva *et al.* 2001) or bacterial artificial chromosome (BAC) libraries with insert sizes up to about 200 kb (Rondon *et al.* 2000). Presently, *E. coli* is still the preferred host for the cloning and expression of metagenome-derived genes, although recently other hosts such as *Streptomyces lividans* have been used for the identification of genes involved in the synthesis of novel antibiotics (Courtois *et al.* 2003).

The information contained within a metagenomic library can be applied to determine both community diversity and activity, the presence of specific microorganisms or biosynthetic pathways as well as identification of individual genes (Steele and Streit, 2005). Subsequently, many environments have been the focus of metagenomics, including soil, the oral cavity, faeces, aquatic habitats, as well as the hospital metagenome, a term intended to cover the genetic potential of organisms in hospitals that contribute to various public health concerns such as antibiotic resistance and nosocomial infections (Coque *et al.* 2002).

2.4. Industrial biocatalysts from the metagenome

Metagenomics can further be defined as a culture-independent approach which can be applied to uncover novel biocatalysts for both pharmaceutical and biotechnological applications. Despite there being numerous limitations in screening metagenomic libraries, such as the functional expression of foreign genes in a heterologous, screening host (Lee *et al.* 2004), much success has been achieved. The metagenomic-based strategy has subsequently led to the identification and isolation of various novel biocatalysts including lipases, esterases, proteases, nitrilases and amylases (Riesenfeld *et al.* 2004).

Lipases and esterases form part of the most widely used group of biocatalysts in organic chemistry as they remain active in organic solvents, do not require cofactors and readily display chemo-, regio- and stereo-selectivities. It is due to these chemical properties, that numerous studies have been aimed at isolating this class of hydrolase (Jaeger & Reetz, 1998). Similarly, metagenomic screens have targeted polysaccharide-modifying enzymes

as they are invaluable in the food industry (Voget *et al.* 2003). Glucoamylases, debranching enzymes and α -1-4-amylases for example, are used to convert readily available raw materials such as starch to corn syrup. Oxidoreductases are yet another example of useful biocatalysts. Their high enantioselectivity makes them ideal for the synthesis of carbonyl compounds, hydroxyl acids, amino acids and chiral alcohols (Davis & Boyer, 2001). The application of the metagenomic approach builds on recent advances in microbial genomics and in the polymerase chain reaction (PCR) amplification and cloning of genes that share sequence similarity directly from environmental samples (Pace *et al.* 1985).

In comparison to PCR amplification, which requires prior knowledge about the sequence of the gene to design primers for amplification, direct isolation and cloning of DNA can theoretically allow genes of any sequence or function to be accessed. Direct cloning of genomic DNA also provides the opportunity to capture operons or genes encoding pathways that direct the synthesis of more complex molecules. In future, application of metagenomic analysis can be implemented to reconstruct the genomes of uncultured organisms through identifying overlapping fragments in metagenomic libraries and ultimately re-assembling each chromosome (Schloss and Handelsman, 2003). The concept of cloning DNA directly from an environment was initially suggested by Pace *et al.* (1985) and first implemented by Schmidt *et al.* (2004), who constructed a λ phage library from a seawater sample and screened it for 16S rRNA genes.

Metagenomic analyses of soil was more difficult to develop than with water due to the technical difficulty of isolating DNA from the complex matrix of soil, containing numerous compounds which bind DNA or have a role in inhibiting the enzymatic reactions required for cloning. In recent years however, significant progress has been made, resulting in the production of libraries that have dramatically improved understanding the functions in the soil community (Rondon *et al.* 1996).

If we consider the high diversity of prokaryotic life in soil environments, soil metagenomic libraries would offer one of the best sources when searching for a wide

range of biocatalysts. These libraries can then be searched using direct sequencing of clones and comparison of sequences with the databases, or by functional analysis, where the library is screened for a specific activity (Streit *et al.* 2004). Initially, early screening campaigns were focused on the cloning of genes encoding phylogenetically conservative molecular traits, such as small subunit rRNA (Schmidt *et al.* 1991) or heat shock proteins (Yap *et al.* 1996). Later studies were however able to recognize novel microbial diversity, thereby demonstrating the value of metagenome cloning for retrieving novel enzymes.

Table 1. Examples of metagenomic discoveries based on functional screening of libraries (adapted from Riesenfeld *et al.* 2004).

Environment	Number of clones	Insert size (kb)	Activity of Interest	Reference
Soil	n.s.	Cosmid	Fatty acid enol esters	Brady <i>et al.</i> 2002
Soil	700 000	Cosmid	Antimicrobials	Brady & Clardy. 2000
Marine	825 000	Plasmid	Chitinases	Cottrell <i>et al.</i> 1999
Faeces and soil	4 x 6000 - 35 000	30 – 40 kb	Biotin biosynthesis	Entcheva <i>et al.</i> 2001
Soil	3 x ~300 000	5 - 8 kb	Lipases	Henne <i>et al.</i> 2000
Soil and river sediment	4 x 100 000	3 - 6 kb	Alcohol oxidoreductase	Knietsch <i>et al.</i> 2003
Human mouth	450	Plasmid	Antibiotic resistance	Diaz-Torres <i>et al.</i> 2003
Soil	n.s.	30 kb	Novel biocatalysts	Voget <i>et al.</i> 2003
Geothermal sediment	37 000	5 kb	Pigments	Wilkinson <i>et al.</i> 2002

Once new genes have been cloned and screened for activity, there is another hurdle which faces the manufacturer: expression of a sufficiently pure protein in sufficient quantity and at a reasonable cost. Obviously, cheap and efficient enzyme production in high-performance expression systems involving bacilli or filamentous fungi is vital for the process to be considered a success, particularly when the enzyme functions as part of the final product as is the case with detergents (Langer *et al.* 2006).

In a recent estimate, Roberts (2004) proposed that in 10% of processes, biocatalysis could provide a superior synthetic solution over classical chemistry. It has even been stated that the availability of an appropriate biocatalyst is now regarded as being the limiting factor for any biotransformation process (Schmid *et al.* 2001). From the above mentioned examples, it can be deduced that there is ample demand for novel enzymes, biocatalysts and biopolymers, and metagenomics is currently thought to be one of the most appropriate technologies to provide the necessary candidate molecules (Schloss and Handelsman, 2003).

2.5. Analysis of metagenomic libraries

2.5.1 Function-driven analysis

There are two approaches which allow for the extraction of biological information from metagenomic libraries. These have been classified as function-driven and sequence-driven analysis respectively. Function-driven analysis relies on the expression of a desired trait which in turn, allows for the identification of DNA sequences coding for active proteins. Further biochemical and sequence analysis can then be employed to characterize these sequences. This technique has proved successful in the identification of numerous clones which have potential industrial applications (Schloss & Handelsman, 2003). These include novel as well as previously described antibiotics (Brady *et al.* 2001), enzymes such as chitanases (Cottrell *et al.* 1999) as well as membrane proteins (Majernik *et al.* 2001). Flanking DNA of clones can be sequenced, in so doing unveiling a gene or a group of genes which could be used to derive the phylogenetic affiliation of the organism from which the DNA was isolated. Often, as in the case of a gene encoding 16S rRNA, RecA or DNA polymerase, a highly conserved gene is not present. In such cases, phylogenetic inferences can be made by sequence alignment of gene clusters with genes in the databases (Schloss & Handelsman, 2003).

Irrespective of the success rate of this approach, it has some limitations. The most notable of these being a suitable screening system, for example complementation of a mutant as well as the presence of all the genes required for the function. Furthermore, due to the very low frequency of active clones, a high-throughput assay for the function of interest is required which can be applied to large numbers of bacterial colonies. Presently, improved heterologous gene expression systems are being developed by utilizing shuttle vectors that facilitate screening of the metagenomic DNA in diverse host species. It is envisaged that this modification could expand the range of gene expression quite significantly (Schloss & Handelsman, 2003).

High throughput screening (HTS) technology, using sophisticated picking and pipetting robotics, is often used to perform functional searches. This method however has been mostly inefficient as its application has been linked to the low detection frequencies observed in functional screens. This is highlighted by the need to analyze several thousand clones, only to detect about ten with activity (Henne *et al.* 1999). The apparent inefficiency of finding functionally active proteins encoded by metagenomic DNAs can be attributed to several reasons: lack of efficient transcription of metagenome-derived genes; poor translation often followed by poor secretion of the heterologous protein; incorrect folding of the protein due to a lack of essential chaperones; a lack of cofactor synthesis or insertion into the recombinant metagenomic protein; as well as different codon usage of the expression host strain (Streit *et al.* 2004).

Application of metagenomics enables the detection and characterization of a vast range of biocatalysts. This process can often be extremely time-consuming and as mentioned previously, it usually requires the screening of thousands of clones before even a small number of positives clones can be detected. The development of DNA micro-arrays now offers a more rapid mean of screening large numbers of clones (Sebat *et al.* 2003). By using this approach, clones generated from non-cultivable microorganisms can be identified, thereby narrowing the range of clones to be sequenced and analyzed further.

There are currently several types of micro-arrays which have been developed and evaluated specifically for bacterial detection and microbial community analysis. Examples include phylogenetic oligonucleotide arrays that contain signature sequences from rRNA of specific groups of organisms, functional gene arrays that contain conserved domains of genes involved in specific metabolic pathways and community genome arrays that contain specific gene sequences from known cultured microbial species (Zhou, 2003). Another innovative approach to screening, involves substrate induced gene expression by screening for catabolic genes. Here, metagenomic libraries are generated using an operon-trap *gfp*-expression vector where the cloning site divides the *lac* promoter and the *gfp* structural gene. The library is then grown in liquid media supplemented with the substrate of interest and fluorescence-activated cell sorting is used to find the GFP-expressing clones containing the genes of interest (Uchiyama *et al.* 2005).

2.5.2. Sequence-driven analysis

Sequence analysis of metagenomic libraries is dependent on the use of conserved DNA sequences which are used to design hybridization probes (PCR primers) to screen for clones containing sequences of interest. Random sequencing of metagenomic clones have also led to significant discoveries, albeit this means of identification is more cumbersome than that described previously. Once a gene of interest has been identified, phylogenetic anchors (eg. 16S rRNA) can be sought in the flanking DNA which could then provide a link between the phylogeny and the function of the gene. Stein *et al* (1996) showed that sequence analysis guided by the identification of phylogenetic markers, produced the first genomic sequence linked to a 16S rRNA gene of an uncultured archaeon. One of the most notable findings to have emerged since the inception of metagenomics resulted from the sequencing of a clone which was isolated from seawater. The clone was initially identified because it carried a 16S rRNA gene, however further sequence analysis by Beja *et al.* (2000) unveiled a gene with very high similarity to bacteriorhodopsin genes. This result revealed the first evidence that rhodopsins are not limited to the Archea, as was previously believed.

It has been stated in a paper by Schloss & Handelsman (2003), that the sequences of most genes of practical importance are too divergent, thereby making the identification of new homologues by PCR or hybridization nearly impossible. There are however, a few classes of genes which contain sufficiently conserved regions to allow for their identification by sequence rather than functional analysis. Two such examples are the genes encoding polyketide synthases (PKSs) and peptide synthetases, both contributors to the synthesis of complex antibiotics. The PKSs have repeating domains containing divergent regions that produce the variation in chemical structures of the products. These regions are flanked by highly conserved regions, thus allowing for the design of probes to screen for PKSs genes from metagenomic clones (Courtois *et al.* 2003).

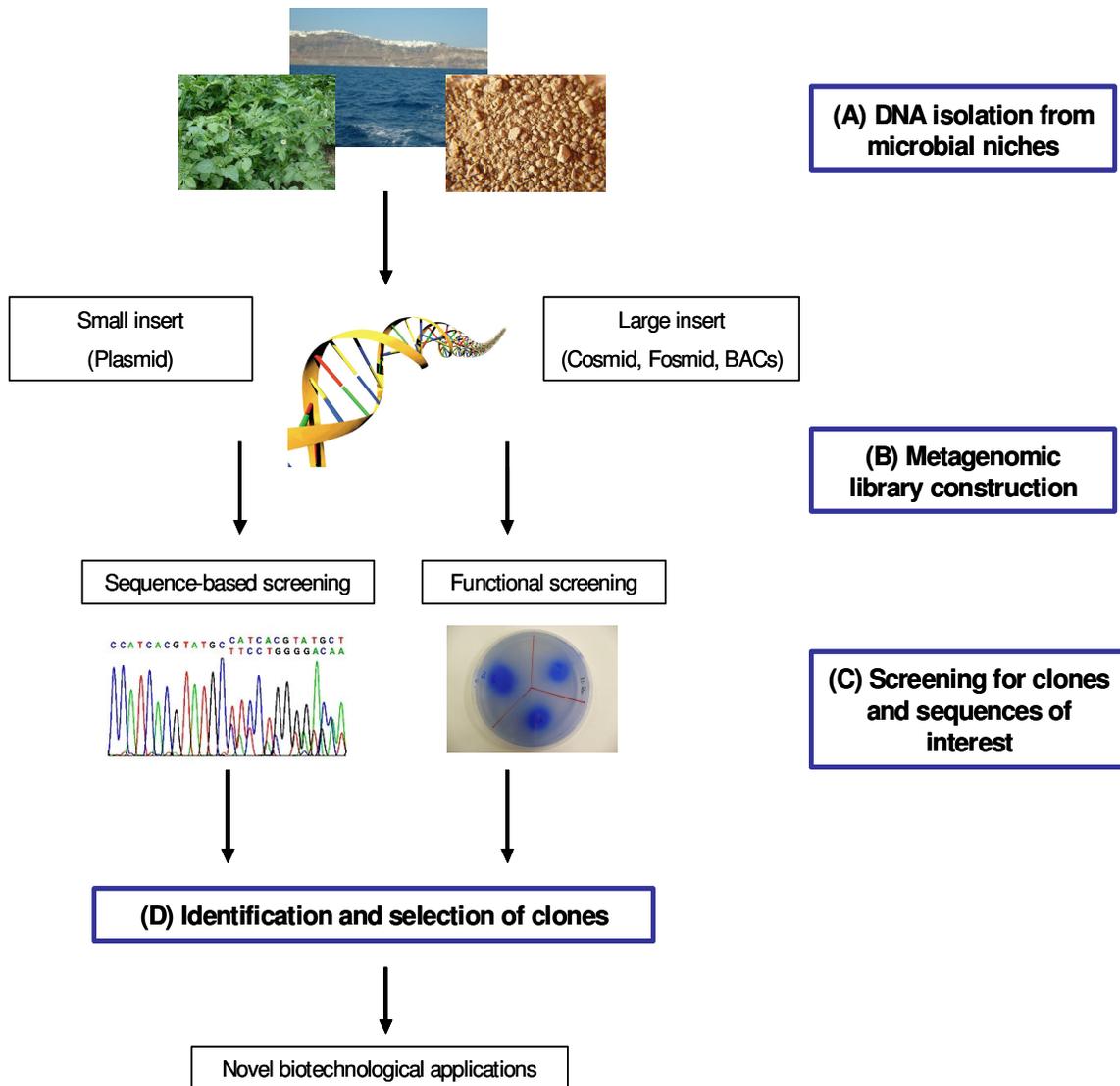


Figure 1. A schematic diagram demonstrating the metagenomic approach to obtain novel biocatalysts. This involves four major steps: (A) isolation of DNA from an environmental sample; (B) the construction of genomic DNA libraries; (B) screening for clones of interest by functional or sequence-based methods and; (D) selection of the desired clones and DNA sequences (adapted from Streit and Schmitz, 2004).

2.6. Biopolymers

Polymers have a pivotal role in natural environments as well as in modern industrial economies. Naturally occurring polymers, such as nucleic acids and proteins, encode essential biological information, while other polymers such as the polysaccharides, act as a source of energy to drive cellular activity and provide structural integrity to living systems. Advancement in the respective fields of chemistry and materials science, have resulted in the production of numerous novel synthetic polymers throughout the past century. Examples such as nylon, polyethylene and polyurethane have made a profound impact on our modern age and are evident in nearly all areas of society. The commercial prospects of biologically derived polymers are continuously increasing, particularly due to the technical advancement in the field of genetic engineering. The application of recombinant DNA techniques provide scientists with the ability to gain control over the purity and specific properties of polymers (U.S. Congress, Office of Technology Assessment, Sept 1993).

Polymers are usually identified structurally as large or complex molecules consisting of individual building blocks linked together to form long chains. Monomers, are simple molecules which can be chemically bound together to form polymers. A homo-polymer is typically composed of only one type of monomer, whereas co- or hetero- polymers are formed when two or more different monomers are linked together. Polymerization is the process by which the monomers are assembled into polymers, and this can occur either chemically or biologically. The diverse application of enzymes is becoming ever popular for use in polymer synthesis (Kobayashi *et al.* 1995) due to their rapid catalytic rates and substrate specificities (Whitesides *et al.* 1985). This has become evident in the preparation of oligosaccharides, where enzymes are being used to catalyze the glycosylation reaction (Kren and Thiem, 1997). Once formed, a polymer can be distinguished by the chemical properties of its monomeric units, the bonds which link these units together, and the size or molecular weight of the polymer. These parameters collectively contribute to the physical properties of the polymer product (U.S. Congress, Office of Technology Assessment).

2.7. β -glucans

In industry, β -glucans are recognized as high-value polymeric compounds particularly in texturizing as fat substitutes. They have also been identified as having an important positive health impact, largely due to their benefits in coronary heart disease, cholesterol lowering and reduction of the glycemic response. These health benefits can be attributed to its high viscosity although it may be that some of these effects are due to appetite suppression also (Burkus and Temelli, 2005).

High molecular weight β -glucans are viscous due to labile cooperative associations whereas lower molecular weight β -glucans can form soft gels as the chains are easier to rearrange to maximize linkages. β -Glucans form 'worm'-like cylindrical molecules containing up to about 250,000 glucose residues that may produce cross-links between regular areas containing consecutive cellotriose units (Roubroeks *et al.* 2001).

Cellulose has been identified as a crystalline β -(1-4)-glucan, and is formed by the repeated connection of these D-glucose building blocks. It is the world's most abundant biopolymer, making its biomass a global carbon sink and renewable energy source, as well as its crystallinity providing mechanical properties central to plant morphogenesis and the fiber industries (Arioli *et al.* 1998). Cellulose has been used for approximately 150 years as a chemical raw material. The formation of cellulose nitrate by reaction with nitric acid and the corresponding technical synthesis of the first thermoplastic polymer material called celluloid by the Hyatt Manufacturing Company in 1870, demonstrated that novel materials could be produced on an industrial scale by the chemical modification of cellulose (Balsler *et al.* 1986). This pioneering work led to further interest, which resulted in increased use of synthetic fibers based on wood cellulose, instead of native cellulose fibers for textiles and technical products.

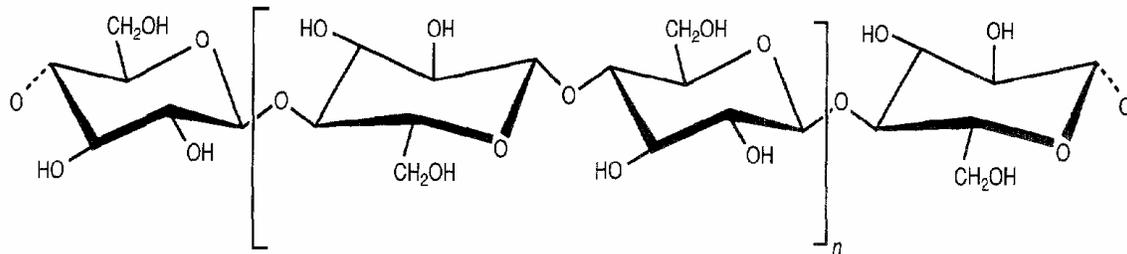


Figure 2. Structural formula for cellulose: a β -(1-4)-glucan polymer chain (Brown *et al.* 1996).

Wood pulp is at present, the most important raw material source for the processing of cellulose, most of which is used for the production of paper and cardboard. Approximately 2% (3.2 million tons in 2003) were used for the production of cellulose regenerate fibers and films, as well as for the synthesis of a large number of cellulose esters and ethers. These cellulose derivatives are produced on an industrial scale and are used for coatings, laminates, optical films and sorption media, as well as for property-determining additives in building materials, pharmaceuticals, foodstuffs and cosmetics (Klemm and Fink *et al.* 2005).

β -Glucan synthases are membrane-bound enzymes involved in cell wall morphogenesis. Much effort has been aimed towards characterization of properties and the function of these enzymes in bacteria, fungi, and green plants over the past few years (Fevre *et al.* 1988), however due to their transmembrane location, these enzymes are notoriously difficult to purify. This observation, accompanied by the fact that the enzyme is often not a single polypeptide but rather occurs as part of a larger complex protein composed of several peptides, greatly complicates analysis of enzyme activity (Selitrennikoff, 1995). Cellulose and hemicellulose are cell wall polysaccharides in green plants. It has been shown that activation of latent enzymes of the plasma membrane or conversion of 1,4- β -glucan synthase by moderate proteolysis could lead to the deposition of 1,3- β -glucans (Delmer, 1987). In the fungus *Saprolegnia*, 1,3- β -glucans and cellulose are integral parts of the cell wall. Using this organism, Fevre and Rougier (1981) demonstrated that isolated membrane fractions exhibit *in vitro* glycosyl transferase activities producing 1,3- β -glucan or 1,4- β -glucan synthesis according to the assay conditions.

Various techniques including gradient density centrifugation, column chromatography and electrophoresis have been applied to show that glucan synthases are large protein complexes (>450 kDa) with the possibility that numerous protein subunits ranging from 18 to 83 kDa may be involved in glucan synthesis (Eiberger and Wasserman, 1987). Bulone *et al.* (1990), proceeded to partially purify 1,3- β -glucan and 1,4- β -glucan synthases from the fungus *Saprolegnia*. They achieved this by using an entrapment procedure, where enzymes were pelleted and solubilized with the reaction product, following purification by density gradient centrifugation.

Chapter 3

Identification and Analysis of a Putative β -(1-4)-Glucan Synthase Isolated from a Metagenomic Library

3.1. Introduction

The functional screening of metagenomic libraries is an effective tool for exploiting the biocatalytic potential of microorganisms present within environmental samples, and has subsequently resulted in the discovery of numerous genes with industrial applications. Although this method is considerably less laborious than cultivation-based techniques, it relies on the heterologous expression of a foreign gene within a suitable host organism and therefore often requires the screening of thousands of clones before the function of interest is identified.

The use of soil for metagenomic screening has been particularly successful in the discovery of new antibiotics and has shown that they are far more prevalent in the uncultivated population (Gillespie *et al.* 2002). Furthermore, the significance of metagenomics for biotechnology is emphasized when one considers that up to 10 000 different microbial species can be present in a single gram of soil with that same gram housing approximately 6.1×10^7 genes (Streit *et al.* 2004).

For the purpose of this study, soil-derived metagenomic libraries were functionally screened for β -(1-3) and β -(1-4) glucan activity using aniline blue, which stains specifically for polymers comprised of these bonds. These polymers could be applied to various industries, particularly the medical sector as they have been shown to possess numerous medicinal advantages. The libraries were received from Bayer BioSciences and were constructed using genomic DNA extracted from various permafrost soil samples collected in Kamtchatka, Russia.

3.2. Methods and Materials

3.2.1. Organisms

3.2.1.1. *Escherichia coli* strains

DH5 α strain: F⁻*endA1 hsdR17*(r_k⁻m_k⁺) *supE44 thi-1 recA1 gyrA* (Nal^r) *relA1* Δ (*lacZYA-argF*) *deoR* (Φ 80*dlac* Δ (*lacZ*)*M15*) (Promega).

BL21 strain: F⁻ *ompT gal dcm lon hsdS_B*(r_B⁻ m_B⁻) λ (DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*]) (Promega).

3.2.2. Chemicals and kits

Enzymes, chemicals and kits were purchased from Sigma (St. Louis, Missouri, USA), Roche Diagnostics (Mannheim, Germany), Invitrogen (Carlsbad, California, USA), Promega (Madison, Wisconsin, USA), Stratagene (La Jolla, California, USA), CalBiochem (Merck Biosciences, Darmstadt, Germany), or QIAGEN (Hilden, Germany).

3.2.3. Plasmids

pCR 2.1-TOPO: contains ampicillin and kanamycin resistance markers, *lacZ* reporter gene, T7 promoter and *f1* origin of replication. This allows for the efficient selection of bacterial colonies that take up the vector plasmid during transformation (Invitrogen).

pGEX 4T-1: contains a Thrombin cleavage site and is designed for prokaryotic expression of proteins as fusion products with Glutathion-S-transferase (GST) (Amersham Biosciences).

3.2.4. Recombinant DNA techniques

Standard procedures in molecular biology were used for preparation of plasmid DNA, restriction enzyme digestion, DNA agarose gel electrophoresis, DNA ligation, and the transformation of bacteria according to Sambrook *et al.* (1989).

3.2.5. Transformation of DH5 α *E. coli*

Competent DH5 α *E. coli* cells were transformed by the standard heat-shock method (Sambrook *et al* 1995), with slight modifications. Fifty μ L of cells were combined with 1 μ L plasmid DNA in a microcentrifuge tube. The mixture was kept on ice for 20 min followed by incubation at 37°C for 90 s. The mixture was again placed on ice for a further 2 min after which 200 μ L Luria broth (LB) medium was added. Following incubation at 37°C for 30 min, 125 μ L of the transformation mixture was plated out.

3.2.6. Culture conditions

3.2.6.1. Solid Media

Top-agar plating was used to allow optimal production and visualization of polymers. This requires a bottom layer of LB supplemented with 50 μ g/ml ampicillin (LBA), 1.2% (w/v) agar with 1% (w/v) glucose as substrate. Transformants were added to 5 mL LBA containing 0.5% (w/v) agar and poured on top of the underlying layer. Plates were then incubated overnight at 37°C, followed by a period of 4 days at room temperature before being stained.

3.2.6.2. Liquid Media

Cultures grown in liquid media were used for the production and extraction of polymers. Conical flasks containing 1L LBA with 1% (w/v) glucose were inoculated using 1 mL overnight cultures. The flasks were incubated for 7 days at RT with agitation under aerobic conditions.

3.2.6.3. Substrates

Initially, various sugar substrates including D-glucose, D-glucosamine, D-galactose, D-fructose, mannitol, sorbitol, sucrose, myo-inositol, galacturonic acid as well as glucaronic acid were independently added to LB at a concentration of 1% (w/v). Once clones had been plated onto the respective substrate-containing media, it was possible to detect the substrate of preference by monitoring the ability of the mutant strains to manufacture polymers.

3.2.7. Staining for β -(1-4)-glucan polymers

Aniline Blue stains specifically for β -(1-4)-glucans and hence was preferred for the purpose of this study. Plates were stained by flooding with 0.1% Aniline Blue solution for 10 min, followed by destaining using 1 M NaCl for 30 min. The destain step was repeated to remove any background which hindered the visualization of polymers. Colonies surrounded by a blue halo were selected for further analysis, and the plasmids contained in them were isolated.

3.2.8. Determination of cloned insert sizes

Plasmid DNA was extracted from *E. coli* clones using an alkaline lysis method described by Sambrook (1989). Plasmids were digested with *Bam*HI (Roche), and insert sizes determined by separating the insert from the plasmid using agarose gel electrophoresis and comparing the inserts with a molecular DNA marker.

3.2.9. Polymer extraction from plate cultures

Colonies which stained positive for the production of β -(1-4)-glucan polymers were selected and cultured overnight in 2 mL LBA at 37°C. Two-hundred and fifty μ L was then plated onto LBA containing 1% glucose. Plates were incubated for 4 days at 28°C to allow for the production of polymers. Following the incubation period, cultures were scraped from the plates using surgical blades and resuspended in 25 mL cell disruption buffer (20 mM Tris-HCl pH 7.5, 20 mM NaCl, 0.1 mM EDTA, 15 mM β -Mercaptoethanol, 100 mM PMSF). The suspension was sonicated for 30 s intervals on ice. Cell debris was collected by centrifugation at 7000 g for 20 min. The supernatant was removed and precipitated in 80% EtOH for 2 hrs at 4°C. Precipitate was again collected by centrifugation and washed extensively with 80% ethanol after which it was dried under vacuum.

3.2.10. Polymer extraction from liquid cultures

Colonies which stained positive for the production of β -(1-4)-glucan polymers were selected and cultured overnight in 2 mL LBA at 37°C. Conical flasks containing 1L LBA with 1% Glc (w/v) were inoculated with overnight culture and allowed to grow at RT for 6-7 days. Cells were pelleted from the culture, after which the culture supernatants were precipitated in 80% EtOH overnight at -20°C. Precipitate was transferred to 50 mL Corning tubes and polymer was collected by centrifugation at 8000 g for 10 min. The polymer was washed using 80% EtOH and 100% methanol respectively and dried under vacuum.

3.2.11. Hydrolysis of polymers

The dried polymer sample (0.001 mg) was added to 500 μ L 4M trifluoroacetic acid. These were mixed thoroughly by vortexing and then incubated at 120°C for 1hr. Following heating, samples were dried under vacuum overnight. The pellet was washed

by resuspension in 100% methanol and again dried under vacuum. The wash step was repeated to provide a pure product.

3.2.12. Sample derivitization for GC-MS analysis

The polymer sample (0.001 g) was derivitized using 0.008 g methoxyamine hydrochloride (MeOx) and 400 μL pyridine. The mixture was incubated for 90 min at 30°C with shaking. Following incubation, 140 μL N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was added and the mixture was placed at 37°C for 30 min with intermittent vortexing. The sample was incubated for 120 min at RT prior to injection.

3.2.13. GC-MS monosaccharide analysis

The analytical system used for monosaccharide analysis consisted of an AS 2000 autosampler, a trace GC and a quadropole trace MS (ThermoFinnigan). Gas chromatography was conducted on a 30 m Rtx®-5Sil MS column (RESTEK) with Integra Guard (inner diameter of 0.25 mm and 0.25 mm film thickness). Samples were injected with a splitless injection in 1 μL volumes. A flow rate of 1 mL min^{-1} was used with the injection temperature set at 230°C and ion source temperature of 200°C. The following temperature program was applied: 5 min at 70°C, then 1°C min^{-1} oven ramp to 76°C and a second ramp of 6°C min^{-1} up to 300°. Before injection of the next sample, the system was temperature equilibrated at 70°C. Mass spectra were captured at two scans per sec with a scanning range of 25 550 m/z . Xcalibur software version 1.2 (Finnigan Corporation 1998-2000) was used for evaluation of mass spectra and chromatograms.

3.2.14. Size exclusion chromatography

Ten mL sepharose CL-6B column (Sigma) was washed with one volume dH_2O followed by five volumes of running buffer (50 mM Sodium Acetate, 500 mM Sodium Chloride, pH 7) with a flow rate of 0.5 mL/sec . The sample was weighed (0.01 g) and dissolved in

1 mL running buffer followed by addition to the column. The column separates molecules with molecular weights ranging between 1×10^4 and 1×10^6 Da. Fractions passing through the column were collected and stored for further analysis.

3.2.15. Phenol-H₂SO₄ assay

This method was used to obtain visual confirmation (by means of color change) of carbohydrate presence in the polymer samples. 25 μ L sample was mixed with 25 μ L 5% phenol and 125 μ L H₂SO₄. Reactions were performed in microtiter wells, with a positive test indicated by the sample changing to a dark yellow color.

3.2.16. DNS assay

The presence of free glucose in hydrolyzed polymer samples was quantified using 3,5-dinitro salicylic acid (DNS). One mg of the polymer extract was digested with 5 M TFA for 3 hours at 100°C to ensure complete hydrolysis. Reactions were performed in 96-well microtiter plates by adding 90 μ l sample with 10 μ l of 5 mM fructose and 100 μ l DNSA reagent (1% w/v 3,5-dinitro salicylic acid, 0.5 M KOH, 1 M K/Na-tartrate). The reaction mixture was incubated in a hybridization oven for exactly 10 min at 83°C. The presence of a reducing sugar was indicated by a shift in absorbance at 560 nm and this was detected using a Powerwave X Microplate Spectrophotometer (Bio-Tek Instruments, Winooski, Vermont, USA). Glucose concentration was determined using a standard curve between 10 and 60 mM glucose.

3.2.17. Protein extraction

Cultures were inoculated in 20 mL LBA and grown overnight at 37°C with shaking. IPTG was added to a final concentration of 1 mM and tubes were incubated at 28°C. Cultures were then harvested 2 hrs and 4 hrs after the addition of IPTG respectively. Cells were collected by centrifugation at 6000 g for 5 min. The supernatant was discarded and the pellet resuspended in 10 mL protein extraction buffer (50 mM Tris-HCl pH 7.5,

100 mM NaCl, 1 mM β -mercaptoethanol, 5% (w/v) sucrose, 1 mM EDTA, 1 mM PMSF). The mixture was sonicated for 30 s on ice repeated 5 times. Unbroken cells and cell debris were removed by centrifugation at 6000 g for 10 min. The supernatant was divided into 200 μ L aliquots and stored at -80°C.

3.2.18. Protein quantification

The protein content of samples was determined by the method of Bradford (1976) using BioRad protein assay reagent. Bovine Serum Albumin (BSA) was used as the standard and absorbance was measured at 595 nm with a Powerwave X Microplate scanning spectrophotometer (Bio-Tek Instruments, Vermont, USA).

3.2.19. Lysis of bacteria for SDS-PAGE

One mL of an overnight culture was harvested by centrifugation at 12000 g for 30 s. The pellet was resuspended in 500 μ L 50 mM Tris (pH 7) and again collected by centrifugation. The supernatant was discarded and 25 μ L dH₂O and 25 μ L 2x SDS gel-loading buffer (100 mM Tris pH 6.8, 200 mM DTT, 4% (w/v) SDS, 0.2% (w/v) Bromophenol blue, 20% (v/v) glycerol) was added. The mixture was then incubated in a boiling water bath for 5 min. Chromosomal DNA was sheared by sonication at 30 s intervals for 2 min on ice. Debris was pelleted by centrifugation at 10 000 g for 10 min at RT. The lysate was loaded onto SDS-PAGE gel.

3.2.20. Electrophoresis

3.2.20.1. SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Protein separation was performed by the method of Laemmli (1970). A BioRad Protean minigel apparatus (BioRad Laboratories GmbH, Munich, Germany) was used with gels of 11 x 7 cm in size and 0.75 mm in thickness. Samples were mixed with 0.4 volumes of 3x Laemmli loading dye (2.4 mL 1 M Tris pH 6.8, 3 mL 20% (w/v) SDS, 3 mL 100%

glycerol, 1.6 mL β -mercaptoethanol, 0.006 g Bromophenol blue) and boiled for 3 min before being loaded onto the gel. Gels were run at 200 V (constant voltage) at RT in 1x SDS-PAGE running buffer (10x buffer: 30.3 g/L Tris base, 144 g/L glycine, 10 g/L SDS). Protein marker (SDS 7B2, Sigma) was loaded onto gels as a standard.

3.2.20.2. Non-denaturing (Native) PAGE

Crude protein extracts from cultures were loaded directly onto PAGE gels and run at 4°C under non-denaturing conditions. In-gel protein activity was then tested by incubation in substrate containing solutions. Substrates were typically dissolved in sodium acetate buffer (pH 6.5) to a final concentration of 0.1% m/v and incubation periods varied from 6 – 18 hours.

3.2.20.3. Staining of PAGE gels

The Colloidal Coomassie Blue Staining kit (Invitrogen) was used to stain SDS-PAGE gels according to the manufacturer's guidelines. Native PAGE gels were stained with 0.1% aniline blue for 20 min at RT. All gels were destained in distilled H₂O.

3.2.21. PCR

3.2.21.1 Colony PCR

PCR was also used to screen transformed *E. coli* colonies for the presence of the desired insert after cloning. Colonies were picked using a sterile toothpick and were incubated in 10 μ L 0.2% (v/v) Triton X-100 for 10 min before addition of PCR reagents as this assisted cell disruption during the first denaturation cycle.

The PCR conditions were set as follows: 2 min denaturation at 94°C; 30 cycles of 30 s at 94°C; 30 s at 54°C; 90 s at 72°C; final elongation at 72°C for 8 min.

Table 2. Oligonucleotides designed for amplification of the glucan synthase gene

Name	Sequence 5' - 3'
GS Fwd	CCGGATCCATGCCCGTAAAATATTTG
GS Rev	TCAGGCTGCGCAACTGTT

3.2.22. Sequencing of DNA

DNA sequencing was performed by the DNA sequencing facility (Central Analytical Facility, University of Stellenbosch, Stellenbosch, South Africa) using an Applied Biosystems ABI Prism 373 Genetic Analyzer in conjunction with an ABI BigDye™ terminator cycle sequencing ready reaction kit according to the manufacturer's guidelines (Perkin-Elmer, Boston, Massachusetts, USA). Open reading frames (ORF) were analyzed using the ORF search tool provided by the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Homology searches were carried out against the GenBank database using the BLASTX and BLASTP algorithms (<http://www.ncbi.nlm.nih.gov/BLAST>).

3.2.23. Protein purification

3.2.23.1. Crude protein extraction

Overnight starter cultures were used to inoculate 40 mL LBA. Cultures were then incubated at 37° C for 6 hours after which expression was induced by the addition of IPTG to a final concentration of 0.1 mM. Incubation at 37° C was continued for a further 3 hours. Cells were harvested by centrifugation at 6000 g for 10 min and resuspended in

4 mL 20 mM sodium acetate buffer (pH 5.0) containing Complete protease inhibitor (Roche). Cells were disrupted by sonication and the debris collected by centrifugation at 10 000 g for 10 min. The supernatant was mixed with 1/10 volume of glycerol and aliquots were stored at -20° C for further use.

3.2.23.2. Protein fusion construct using pGEX

PCR primers were designed to amplify the glucan synthase gene out of the pCR2.1 TOPO vector. The forward primer contained a BamHI restriction site immediately upstream of the methionine start codon. Following amplification, the PCR product was purified using a commercially available kit (Qiagen) and ligated into the pGEM®-T Easy vector (Promega). The gene was then excised from pGEM®-T Easy using the restriction enzymes *BamHI* and *NotI*. These enzymes were also used to digest the pGEX 4T-1 vector. The glucan synthase gene was ligated overnight into the *BamHI* and *NotI* restriction sites of the pGEX 4T-1 vector.

3.2.23.3. Ammonium sulphate precipitation of proteins

A beaker containing the protein solution was placed inside a cooling bath (containing ice slurry) on top of a magnetic stir plate. While agitating gently, ammonium sulphate was added slowly to a final concentration of 80% (w/v). The mixture was kept stirring for 4 hours until the ammonium sulphate had dissolved completely. Proteins were collected by centrifugation at 7000 g for 20 min, resuspended in dH₂O and stored at -20°C.

3.2.23.4. Desalting of ammonium sulphate precipitated proteins

Protein samples were sealed inside Membra-Cel 32 mm flat width dialysis tubing (Sigma) and incubated in 50 mM sodium acetate buffer (pH 5.5) for 2 hours at 4°C. The buffer was then replaced and the samples were dialyzed further overnight.

3.2.23.5. Schiff staining

Protein samples were applied to native gels. Following electrophoresis, gels were fixed in 40% (v/v) EtOH – 7% (v/v) acetic acid solution for 30 min. This wash was repeated 3 times followed by overnight incubation in fresh fixing solution. Polysaccharides in the gel were oxidized by immersion in a solution of 1% (w/v) periodic acid – 3% (v/v) acetic acid for 60 min. Oxidized gels were washed 10 times in dH₂O for 10 min per wash to remove traces of periodic acid. Gels were incubated in Schiff's reagent for 60 min in the dark. Following staining, background was eliminated by washing 3 times in a 0.58% (w/v) potassium metabisulfite – 3% (v/v) acetic acid solution for 30 min intervals.

3.2.23.6. Thrombin digestion of GST fusion protein

Thrombin (Sigma) was used for site-specific cleavage of the GST affinity tag from the fusion protein. Approximately 2 mg of the fusion protein was incubated with 4 µg thrombin (3 U/ µg) for 20 min at RT in thrombin cleavage buffer (50 mM Tris pH 8.0, 150 mM NaCl, 2.5 mM CaCl₂ and 0.1% (v/v) β-mercaptoethanol).

3.2.24. Determination of glucan synthase enzyme activity

Reactions were performed in 96 well microtiter plates with a total reaction volume of 210 µL per well. The reaction volume consisted of 129 µL dH₂O; 12.5 µL Tris-HCl (1 M, pH 7.0); 25 µL KCl (1 M); 37.5 µL MgCl₂·6H₂O (100 mM); 5 µL NADH (7.5 mM); 25 µL PEP (50 mM); paramylon (1 mg/mL w/v); 30 µL UDP-glucose (100 mM) and 10 µL protein extract (4.36 µg/ µL). One µL of each of the two coupling enzymes, pyruvate kinase (2 U/µL) and lactate dehydrogenase (6 U/µL), were finally added to each reaction. End-point readings were taken at 340 nm using path-length correction on a Powerwave X Microplate Spectrophotometer (Bio-Tek Instruments, Winooski, Vermont, USA). ADP (10 µL of a 25 mM stock) was used for the control reaction.

3.3. Results and Discussion

3.3.1. Library screening

Functional screening of the metagenomic libraries was performed by plating out the library in top agar. The library was transformed by electroporation into the DH5 α strain of *E. coli* and transformants were then added to 3 mL of LBA containing 0.6% (w/v) agar. This mixture was then allowed to solidify on-top of LBA plates containing a number of different carbohydrate substrates. Following incubation at RT for two to five days, the plates were stained with an aniline blue (methyl blue) solution to allow for the detection of β -1,3 or -1,4 polymers (Nakanishi *et al.* 1974). This is a particularly useful method of screening for β -glucan activity as the stain is sensitive, reliable and easy to perform. Of the numerous substrates used, only glucose and glucosamine provided positive staining colonies. Positive colonies were identified as either exhibiting a blue halo surrounding the individual colony, or by darker staining of the individual colony itself (Figure 1). To eliminate the selection of false-positives, the plasmids from all positive colonies were isolated, re-transformed and stained as before to confirm the phenotype.

Positive clones were again identified from the replica plates, subjected to plasmid DNA isolation and then sequenced (figure 3A). From the BLAST results, a clone was isolated which showed significant nucleotide homology to previously identified glucan synthase genes (table 4), providing further hope that the staining polymer was in fact a glucan. This clone was then used for further analysis during this study. When retransformed into DH5 α *E. coli*, clones expressing the gene would produce dark staining halos when treated with aniline blue as seen in figure 3B.

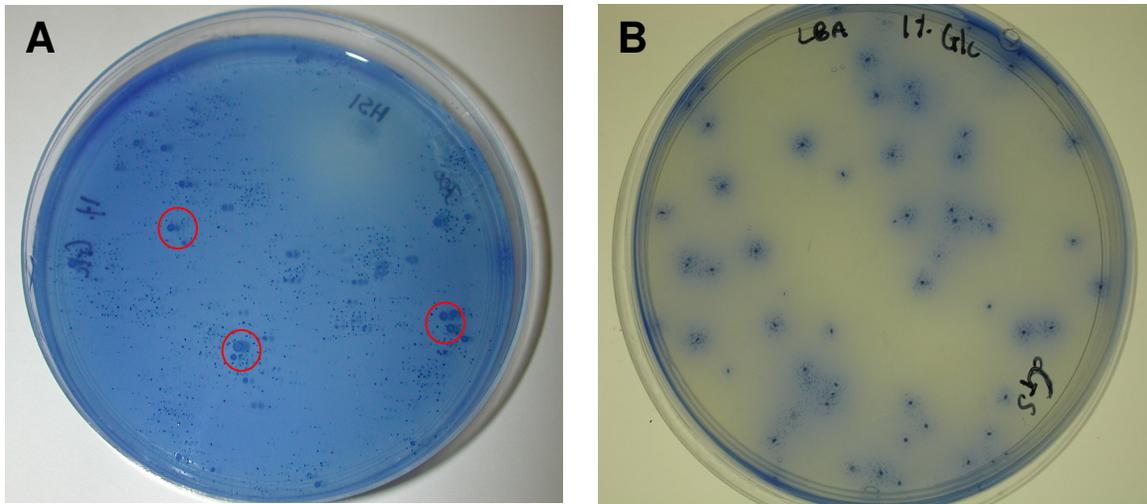


Figure 3. Following transformation of the library into DH5 α *E. coli*, colonies staining with aniline blue (as indicated in **A**) were selected and those containing plasmids were sequenced. Retransformation of the positive glucan synthase clone shows halo-staining colonies (visible in **B**) with aniline blue. This suggests the production of a β -glucan polymer.

Table 3. The 447 bp nucleotide sequence obtained for the glucan synthase clone. The predicted amino acid sequence is given below the nucleotide sequence in the standard one-letter code. The stop codon is marked by an asterisk.

1	ATG	CCC	GTA	AAA	TAT	TTG	CGG	AGA	AAC	CGC	CTG	GTC	AAA	AGG	CAG	45
1	M	P	V	K	Y	L	R	R	N	R	L	V	K	R	Q	15
46	CGC	CAG	TTT	GTC	TGG	CAC	GGT	GTT	GTC	CAT	GAA	TAT	TTG	GAG	GTC	90
16	R	Q	F	V	W	H	G	V	V	H	E	Y	L	E	V	30
91	GCG	GGC	AAG	CTT	TTC	ACA	AGA	TAT	CCG	CCA	CGC	ATC	GCA	AAG	AGA	135
31	A	G	K	L	F	T	R	Y	P	P	R	I	A	K	R	45
136	AGC	CGT	ACA	CCG	ACC	GCA	ACC	TGC	AAA	TTT	ATT	TGC	AGC	GCA	AGG	180
46	S	R	T	P	T	A	T	C	K	F	I	C	S	A	R	60
181	AGC	GGC	AAG	AGC	CGT	TTT	CGC	CGA	GAG	ATC	AAG	GGC	AAT	CGA	ATT	225
61	S	G	K	S	R	F	R	R	E	I	K	G	N	R	I	75

226	CCC	GCG	GCC	GCC	ATG	GCG	GCC	GGG	AGC	ATG	CGA	CGT	CGG	GCC	CAA	270
76	P	A	A	A	M	A	A	G	S	M	R	R	R	A	Q	90
271	TTC	GCC	CTA	TGG	TCA	GTC	GTA	TTA	CAA	TTC	ACT	GGC	CGT	CGT	TTT	315
91	F	A	L	W	S	V	V	L	Q	F	T	G	R	R	F	105
316	ACA	ACG	TCG	GGA	CTG	GGA	AAA	CCC	TGG	CGT	TAC	CCA	ACT	TCA	TCG	360
106	T	T	S	G	L	G	K	P	W	R	Y	P	T	S	S	120
361	CCT	TGC	AGC	ACA	TCC	CCC	TTT	CGC	CAG	CTG	GCG	TCA	TGG	CGA	AGA	405
121	P	C	S	T	S	P	F	R	Q	L	A	S	W	R	R	135
406	GGC	CCG	CAC	CGA	TCG	CCC	TTC	CCA	ACA	GTT	GCG	CAG	CCT	TGA		447
136	G	P	H	R	S	P	F	P	T	V	A	Q	P	*		

Table 4. BLAST hits showing significant alignment with the glucan synthase clone sequence

GenBank accession no.	Name
YP_001643660	Glycosyl transferase family 2 [Bacillus weihenstephanensis KBAB4]
ZP_02645843	Glycosyl transferase family 2 [Paenibacillus sp. JOR-2]
ZP_02601664	Glycosyltransferase, possible beta 1,4 glucosyltransferase [Bacillus cereus 03BB108]
YP_893938	Glycosyltransferase, possible beta 1,4 glucosyltransferase [Bacillus thuringiensis str. Al Hakam]
YP_173976	Beta 1,4 glucosyltransferase [Bacillus clausii KSM-K16]
ZP_02526185	Beta 1,4 glucosyltransferase [Bacillus cereus AH1134]
YP_001420378	Putative beta 1,4 glucosyltransferase [Bacillus amyloliquefaciens FZB42]
ZP_02615484	Beta 1,4 glucosyltransferase [Clostridium botulinum Bf]

3.3.2. Hydropathy plot

The determination of protein function can often be assisted by gaining insight into its structure. For example, the application of hydropathy plots predicts the location of potential transmembrane or surface regions in proteins (Kyte & Doolittle, 1982). The peptide sequence of a protein is analyzed and each amino acid is given a hydropathy

score between 4.6 (most hydrophobic) and -4.6 (most hydrophilic). These values can then be used to predict the structure of the protein in question with transmembrane regions generally peaking at values of 1.6 or greater. The nucleotide sequence obtained for the glucan synthase clone was translated into protein using ExPASy translate tool (1). The 447 bp nucleotide sequence therefore provided a 149 amino acid peptide sequence. This sequence was entered into a Kyte-Doolittle hydropathy plot (Figure 4).

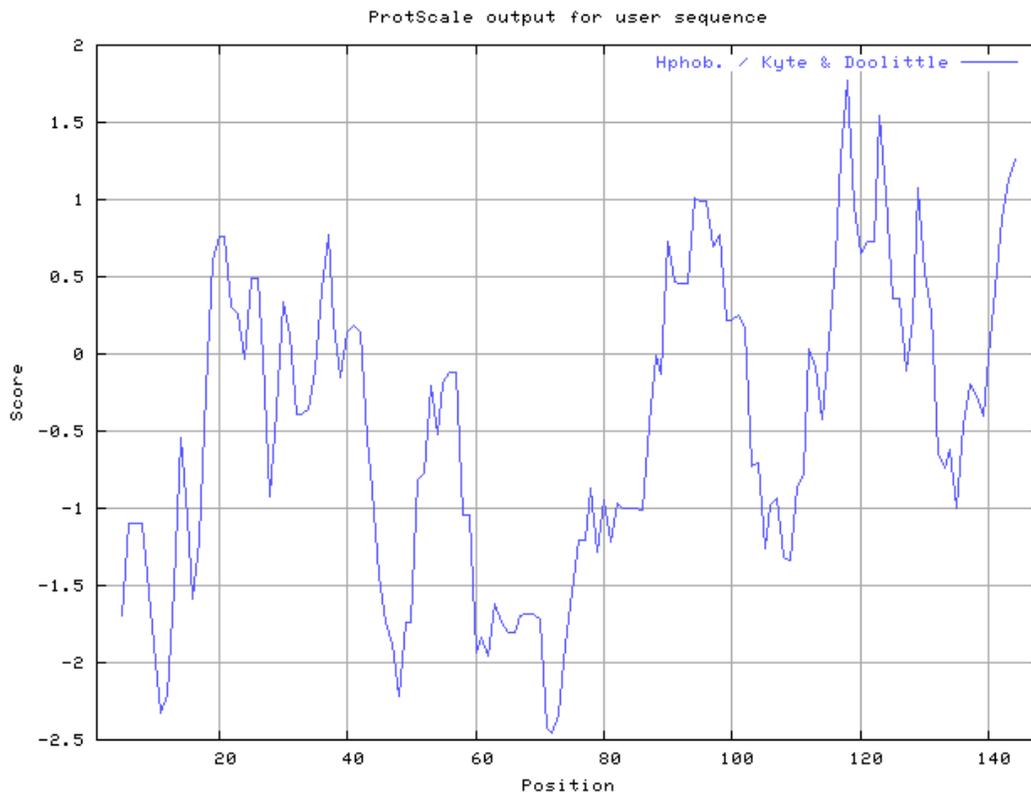


Figure 4. Kyte-Doolittle hydropathy plot representing the 149 amino acid peptide sequence obtained for the glucan synthase clone.

The plot indicates greater peaks in the hydrophilic region with numerous domains exceeding scores of -1.5. This is in contrast to the hydrophobic region which suggests merely 2 peaks above 1.5 on the scale. The hydrophobicity data for this protein therefore strongly suggests that it is not membrane bound but rather soluble. This is interesting as the glucan produced by the protein appears to accumulate outside of the cell.

It might be possible that the protein is present in the plasma membrane and manufactures the polymer there, secreting it into the media. However, given that the protein appears to be soluble, it is likely that the protein is secreted out of the cell and catalyses a polymerization reaction in the media.

The first step to gaining a better understanding of this clone would be to purify the polymer that it produces and determine its monosaccharide composition. Initial attempts at purification were performed using cultures grown on solid media containing the appropriate substrates. Following incubation, clones were scraped off the plates and polymer extracts were made. This method eventually proved unsuccessful as precipitation of the final extracts revealed insufficient quantities of polymer for GC-MS analysis. It is likely that the poor yield could be attributed to the polymer being manufactured in the surrounding medium thereby negating the scraping method employed on top of the solid media. The next step was to find an alternative method for collection of the polymer.

In an attempt to isolate larger amounts of polymer, the positive clones were inoculated in liquid culture. In addition, a control of the same *E. coli* strain containing the empty vector was also inoculated into another flask. These were again cultivated over a period of six days at room temperature under aerobic conditions, with antibiotics being added every 36 hours to prevent contamination. Cells were first removed from the culture by centrifugation and polymers were revealed by precipitation of the culture supernatant by addition of ethanol to a final concentration of 80% (v/v). Due to the fact that initial staining of these clones provided an indication that extracellular polymers were being produced, it was decided to use the culture supernatant and not the cells to perform polymer extractions. This method resulted in the clear visualization of a precipitate, which was not present in the control, and suggested that a significant quantity of polymer was being produced.

3.3.3. Polymer investigation

3.3.3.1. Monosaccharide analysis by GC-MS

The polymer was prepared for monosaccharide analysis by hydrolyzing it to its monosaccharide components using TFA, followed by several washes with 80% (v/v) methanol. The dried sample was derivatized before being run against internal standards. The chromatogram presented a very large peak at a retention time of approximately 27 min, which represents glucose. This would suggest that the polymer is composed predominantly of glucose monomers. Lesser amounts of levoglucosan (23 min), glucosamine (27.7 min) and maltose (40 min) were also present in the sample. It is possible that these sugars are breakdown products of the polymer under investigation.

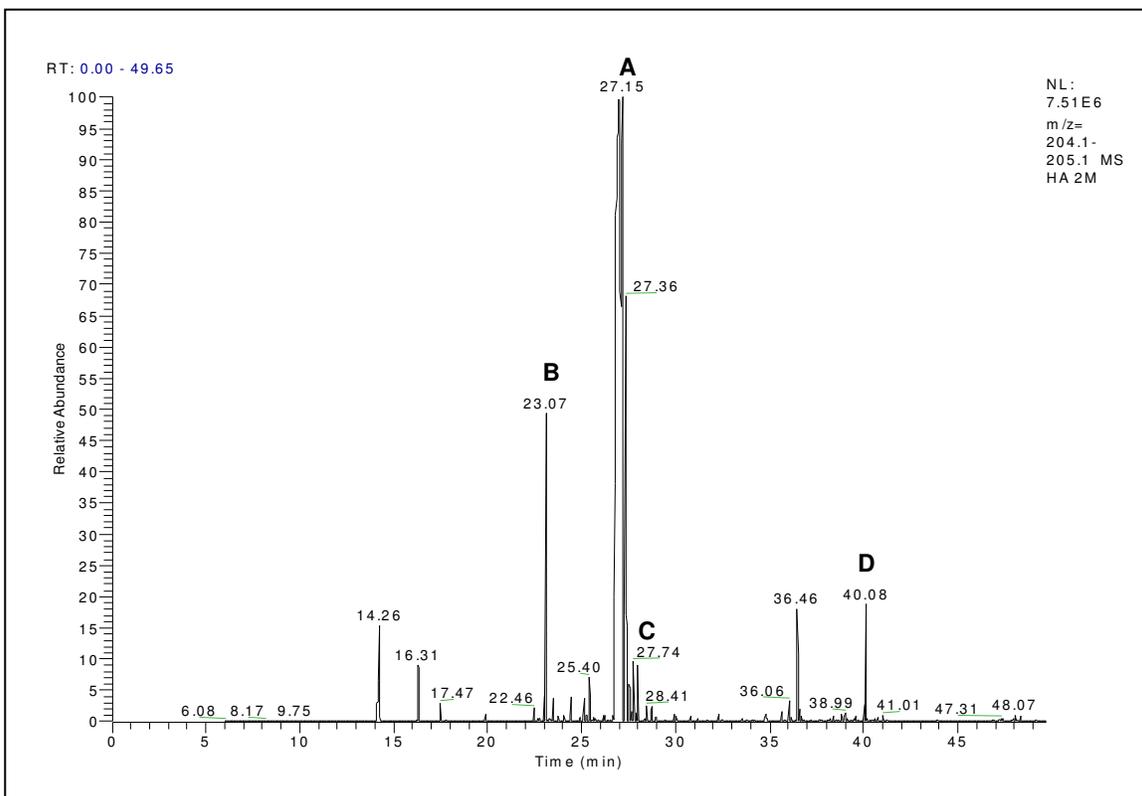


Figure 5. Chromatogram of the hydrolyzed polymer with peak A representing glucose, B - levoglucosan, C - glucosamine and D - maltose.

3.3.3.2. Separation of carbohydrates by size-exclusion chromatography

To examine the molecular size of the polymer, it was passed through a size exclusion column containing Sepharose CL-6B. Carbohydrate was detected in fractions which eluted from the column using the phenol-sulphuric acid method. The pellet extracted from a negative control culture (containing empty vector only) was similarly purified through the column. As can be seen in Figure 6, the yellow color-change, demonstrating the presence of carbohydrate, was only present in fractions from the pellet isolated from bacteria containing the putative glucan synthase gene, but not in the negative control. As the polymer didn't elute in the void volume, but rather was separated by the Sepharose column, the size of the polymer can be estimated to be between 1×10^4 and 1×10^6 Da as this is the size range that Sepharose CL-6B is known to separate.

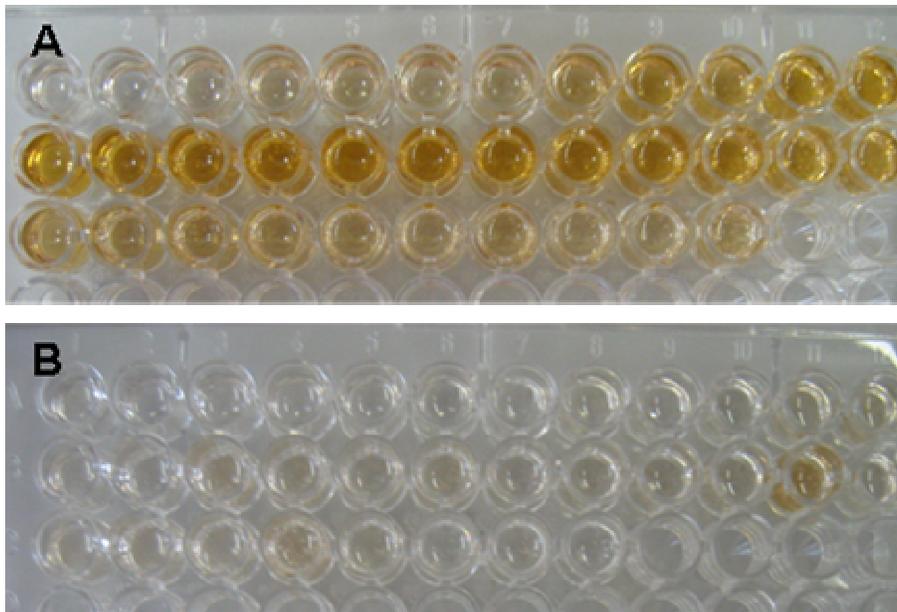


Figure 6. Phenol-sulphuric acid assay for detection of carbohydrate sugars in polymer extracts. **A** indicates the fractions collected for the glucan synthase clone compared to **B**, which are the negative control fractions. The color change from clear to yellow indicates the presence of sugars in the sample.

3.3.3.3. Estimation of the number of glucose units present in the polymer

To estimate the average number of glucose moieties in each molecule of the polymer, reducing end numbers were estimated using the 3,5-dinitro salicylic acid assay. A polymer will have one reducing end, but when it is hydrolyzed all the monomers will have a reducing end. By estimating the number of reducing ends in the polymer before and after hydrolysis the average number of monomers in the polymer can be calculated.

Results showed the ratio of free glucose to be 0.86 mg per 1.00 mg of polymer, thereby correlating with the large glucose peak observed on the GC-MS chromatogram. The number of reducing ends in the undigested polymer was approximately five fold less than in the digested polymer, indicating that the polymer is a pentasaccharide, which would release five glucose monosaccharides upon hydrolysis. If a single glucose unit has a molecular weight of 180.16 Da, this then suggests that the polymer under investigation is approximately 900 Da in size. This however does not correlate with the size estimate achieved from the CL-6B column which would exclude molecules below 10^4 Da. It is therefore likely that the polymer was not hydrolyzed completely thus resulting in the 5:1 monosaccharide ratio. These contradictory results could also indicate that the polymer is composed of more than one type of monomeric unit thereby accounting for its size of greater than 10^4 Da.

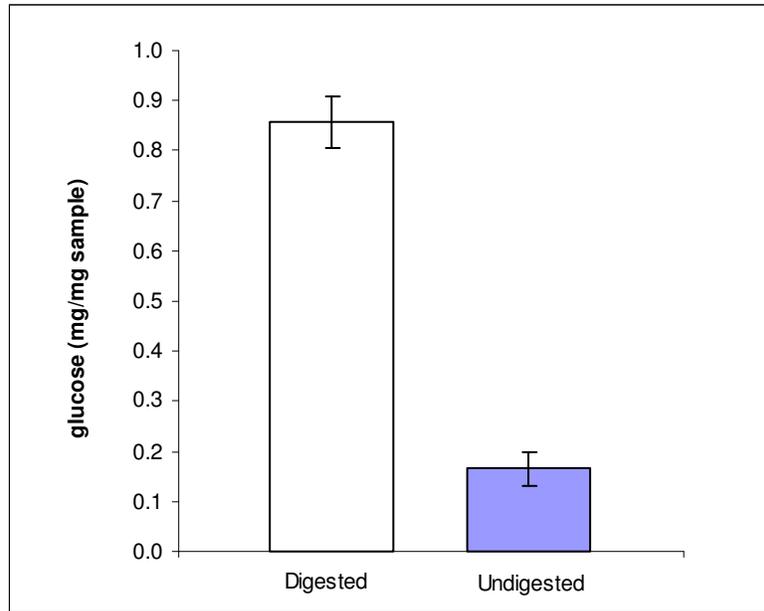


Figure 7. DNS assay was used to determine the amount of free glucose present in the polymer sample following TFA hydrolyzation. The polymer was measured to be composed of 0.86 mg glucose per 1.00 mg of sample. This was approximately five times more than the untreated sample.

3.3.4. Activity measurements in crude extracts

The gene being investigated in this study was identified from a metagenomic library and thus no information is available as to its biological origin. In order to gain a more practical insight into the function of this synthase clone, it was expressed in *E. coli* and soluble extracts were made from the bacteria.

SDS-PAGE is an effective method to separate proteins based on their molecular weight. This technique can be applied for screening and identifying the presence of a protein being expressed relative to a negative control lacking that particular activity. Although application of this technique for homogenous purification of a single protein itself would be extremely difficult due to the copious amounts of proteins present on the gel, it can still provide sufficient evidence that the protein of interest is present thereby validating

further purification methods. This technique was initially applied in attempts to identify protein activity in extracts from the glucan synthase clone. Protein extracts from the clone were run on SDS-PAGE gels along with proteins extracted from empty vector control cultures. Following Coomassie blue staining of the gels, protein bands were visible in both of the samples, however there was insufficient evidence to show a differentially expressed protein between the control and clone extracts.

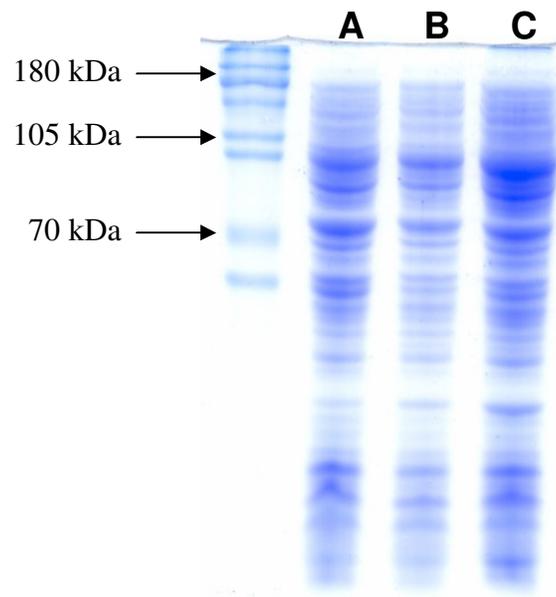


Figure 8. SDS-PAGE gel containing protein extract from empty vector negative control in lane A and the glucan synthase clone in lane B and C. The gel provided no evidence of a differentially expressed protein in the glucan synthase samples when compared to the negative control.

Native PAGE was used in an attempt to show in-gel glucan synthase activity. Crude protein extracts were prepared and separated under non-denaturing conditions. The gel was then incubated in a sodium acetate buffer (pH 6.5) containing glucose as a substrate for 12 hours to allow for β -glucan polymer production by the active enzyme. Aniline blue was used to stain the gels to detect the formation of any β -1,3 or β -1,4 compounds, however, repeated attempts failed to reveal a staining band in the gel. As this was potentially a novel enzyme, no prior knowledge of its mode of action could be applied to exploit activity.

It is possible that a primer was required by the enzyme to initiate polymer synthesis and would therefore need to be added to the gel. For this purpose paramylon, a β -1,3-glucan from *Euglena gracilis*, was considered as a possible initiating compound. The paramylon was incorporated into the PAGE gels to a final concentration of 0.1% (w/v) and the protein extracts were loaded and run as before. Sequence homology for the glucan synthase clone suggests that it may function by glycosyltransferase action, whereby it transfers sugar residues from an activated sugar donor to a growing carbohydrate group. For this purpose, UDP-glucose was selected as a likely substrate donor. This is because the glucose used for the initial functional screening is not an activated sugar and would therefore be an ineffective substrate for the biosynthesis of possible glucan polysaccharides. A buffer pH of 6.5 was initially used for the incubation however it was thought that this might influence potential enzyme activity. This was tested using a range of buffers with pH varying from 5.0 to 8.0. At buffer pH 6.0 and in the presence of UDP-glucose, a faint band was observed for the glucan synthase clone. Numerous unsuccessful attempts were made to repeat this in the hope of achieving a significant increase in intensity of the staining band.

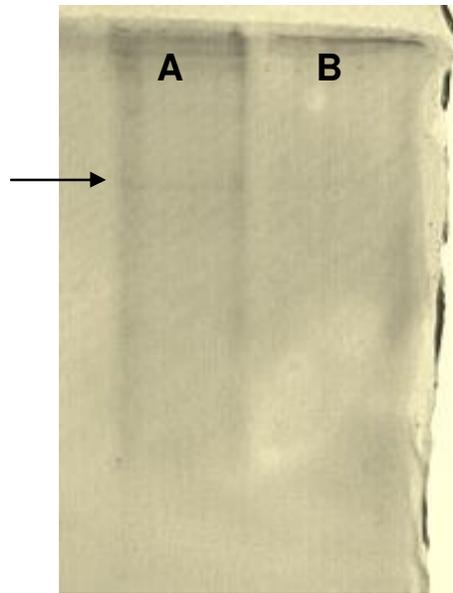


Figure 9. Native PAGE gel containing crude protein extract from the glucan synthase clone in lane A and extract from empty vector control in lane B. The arrow indicates a faint band produced upon aniline blue staining following incubation in buffer containing UDP-glucose.

Schiff staining of PAGE gels using fuchsin-sulphite reagent (Sigma) was attempted to detect the presence of glycoproteins that may have formed. This method relies on the oxidation of glycoprotein bands using a 1% (w/v) periodic acid – 3% (v/v) acetic acid solution. Protein extracts were separated under non-denaturing conditions and gels were incubated in Tris buffer (pH 6.5) overnight. UDP-Glucose was added to the buffer at a final concentration of 100 mM to serve as a substrate for the enzyme. It was expected that the higher detection sensitivity of this stain might enable the visualization of a polymer being formed within the gel. Various parameters likely to affect the enzyme activity were adjusted to reach optimal activity conditions. Substrate concentration, incubation time, incubation temperature and buffer pH were varied, yet staining for the glycoproteins still remained unsuccessful.

3.3.5. Expression of a recombinant protein

A further attempt to purify the protein by recombinant expression was performed using the pGEX vector system. For this purpose the glucan synthase clone was ligated in-frame with the glutathione transferase (GST) tag in the pGEX 4T-1 vector. This tag would result in the expression of a polypeptide fusion rather than an endogenous protein. PCR primers were designed to amplify the gene out of the pCR2.1 TOPO vector with a *Bam*HI site incorporated immediately upstream of the start codon. This would enable the gene to be ligated into the expression vector and prevent a frame-shift from taking place.

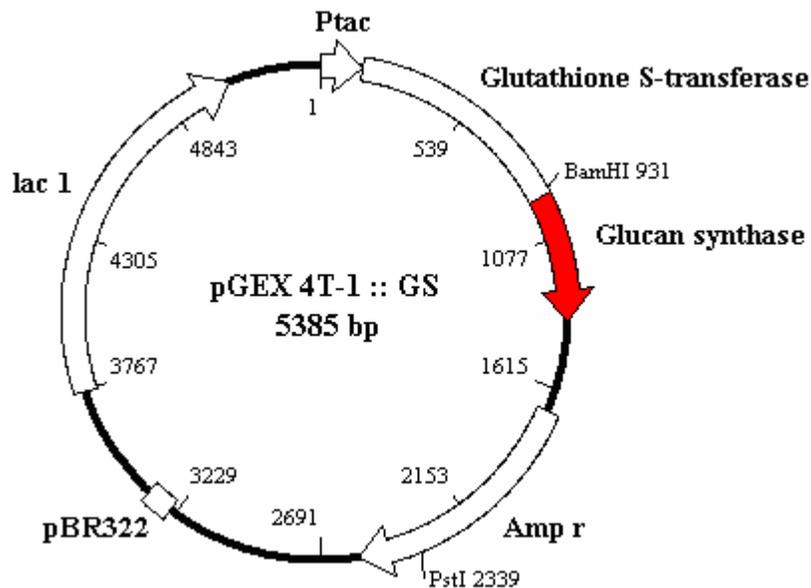


Figure 10. The pGEX 4T-1::Glucan synthase fusion construct. The 447 bp glucan synthase gene (GS) was cloned into the *Bam*HI and *Not*I restriction sites in the sense orientation as indicated on the map above.

The fusion construct was transformed into the *E. coli* BL21 strain and colony PCR was conducted on transformants to confirm the presence of the gene. A positive clone was then cultured to allow over-expression of the recombinant protein. Ammonium sulphate precipitation was used to extract proteins from the culture supernatant while sodium chloride (100mM pH 2.0) was used to wash possible membrane bound proteins from the cells. Samples were desalted by overnight dialysis in sodium acetate buffer (50mM pH 5.5) and protein concentrations determined by the method of Bradford (1976).

3.3.6. Purification of a recombinant protein by affinity chromatography

GST fusion proteins are readily purified from bacterial lysates using glutathione agarose as an affinity medium. Impurities can be washed through the column, allowing the fusion protein to be eluted using reduced glutathione under non-denaturing conditions and in so doing maintaining its function. Following elution, a concentration of 1.84 $\mu\text{g}/\mu\text{L}$ was achieved for the fusion protein and 0.97 $\mu\text{g}/\mu\text{L}$ for the GST control. Samples were then run on SDS-PAGE to confirm a size difference between the fusion protein and the GST tag. Initially, 12.5% gels were used and this did not provide sufficient separation between the two samples. This could be attributed to the small size of the protein of interest which would only be approximately 5 kDa in size. The GST tag itself measures 26 kDa, therefore we were expecting a fusion protein in the vicinity of 31 kDa.

Native PAGE was also used to detect possible glucan synthase activity, however no staining resulted after treatment with aniline blue. It was possible that the presence of the much larger GST tag may have hindered the function of the protein under investigation and could account for its lack of activity.

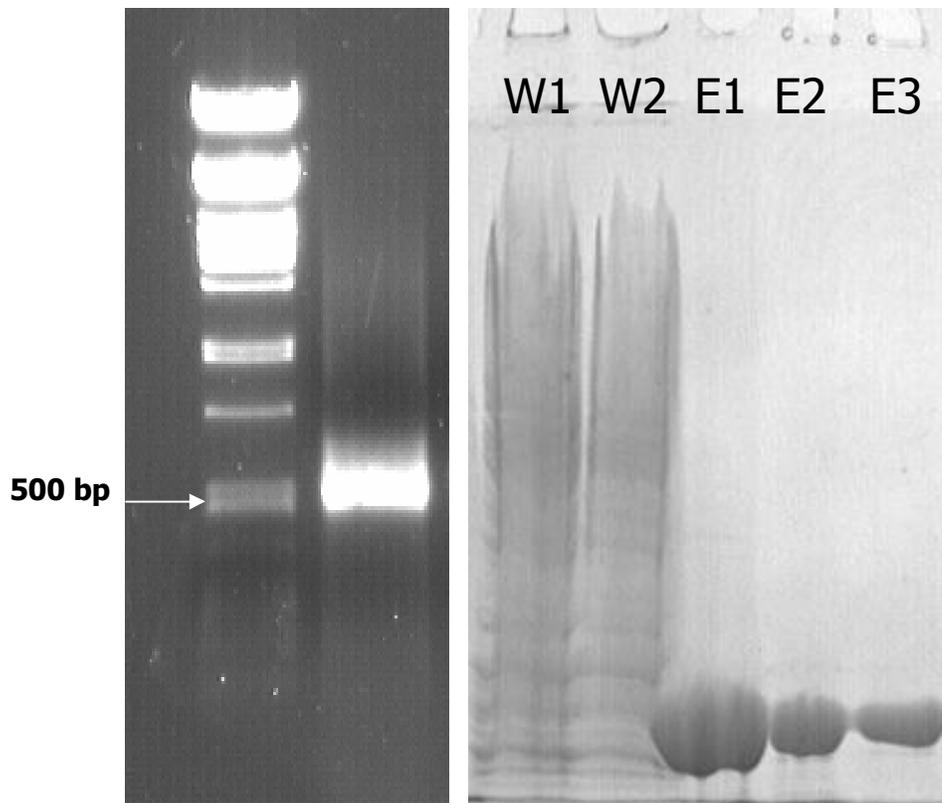


Figure 11. Colony PCR to confirm the presence of the glucan synthase gene in the pGEX 4T-1 vector. Purification of the GST fusion protein using glutathione agarose. W1 and W2 represent the column wash fractions. E1, E2 and E3 are fractions collected following addition of elution buffer. Fraction E3 contained the pure fusion protein and was used for further analysis.

3.3.7. Cleavage of the GST tag

As an attempt to determine the function of the protein it might, therefore, be necessary to remove the GST affinity tag. For this purpose thrombin, a site-specific protease which would cleave the tag from the fusion protein, was used. The thrombin-digested sample was run on SDS-PAGE alongside the fusion protein to determine whether the tag had been removed.

The resolution achieved using 12.5% gels was insufficient to accurately confirm the size of the fusion product and it was decided to run the sample on gradient PAGE gels (4-15%). These were then used to compare the samples and revealed a slight difference in size between the fusion protein and negative GST control.

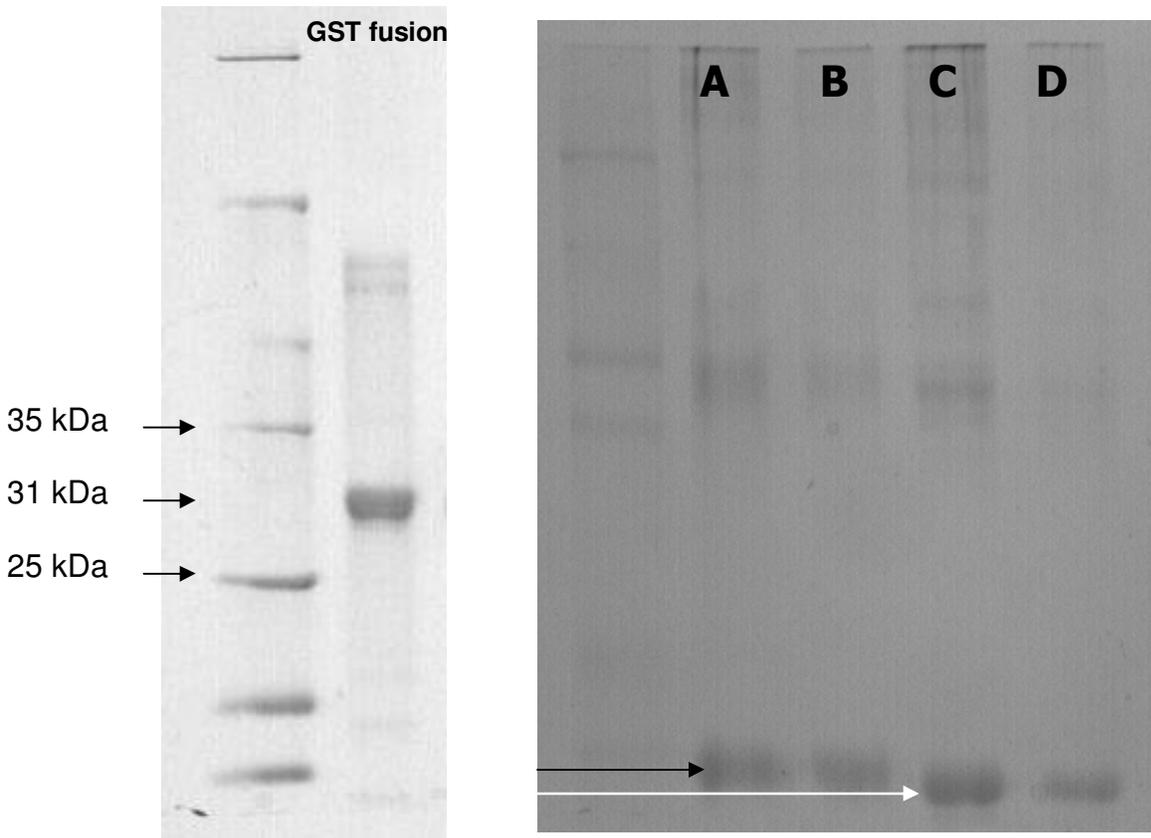


Figure 12. Gradient SDS-PAGE gel (4% - 15%) on the left suggests that the fusion protein is approximately 31 kDa in size. With the GST tag being 26 kDa, it can be concluded that the protein of interest is 5 kDa. On the right, a comparison between the GST fusion protein in lanes A and B, and the thrombin-digested GST fusion in lanes C and D. The arrows indicate a very small size difference once the protein of interest has been cleaved from the tag.

3.3.8. Assay for activity determination of the fusion protein

Following thrombin cleavage of the GST tag from the protein of interest, a coupled assay was developed to assess its activity. This was used as it should be more sensitive than techniques such as native gels. UDP-glucose was provided as a substrate for the purified protein and it was envisaged that this would result in the formation of UDP. PEP was added to the reaction to serve as a substrate for the function of the first coupling enzyme, pyruvate kinase, which would then produce pyruvate, UTP and NADH. This then followed by lactate dehydrogenase conversion of the products to L-lactate and NAD. Free NAD in the sample could then be measured on a spectrophotometer at 340 nm. Paramylon was also included in the assay buffer as a primer at a final concentration of 1 mg/mL. Activity was expressed as nmol/min/mg and the positive control reaction was performed using ADP as it can serve as a substrate for pyruvate kinase. Figure 14 indicates that very low levels of activity were observed in both the purified sample as well as the control. Results also show no significant difference in product formation when paramylon was added to the reaction as a potential precursor.

The reason as to why no activity is detectable remains unclear. The enzyme is clearly active when expressed in *E. coli* as the formation of an aniline blue-staining polymer is visible. What is unknown, however, is what the actual substrate for the enzyme is. I utilized UDP-glucose as a potential glucan donor as that is the most likely biological precursor. Other more minor activated sugars, such as GDP-glucose, were not tested however and it could be that the enzyme is specific for one of these. In addition paramylon was used as a primer, and it might be that some other compound is needed to initiate polymerization. It also remains unknown whether the enzyme requires co-factors (such as metal ions) to act as intermediate carriers of functional groups for the conversion of the substrate into product. These may have been present within the host organisms and hence allowed for product to be formed. Another influential factor is pH, which may not have been optimal under experimental conditions therefore negatively impacting on enzyme function.

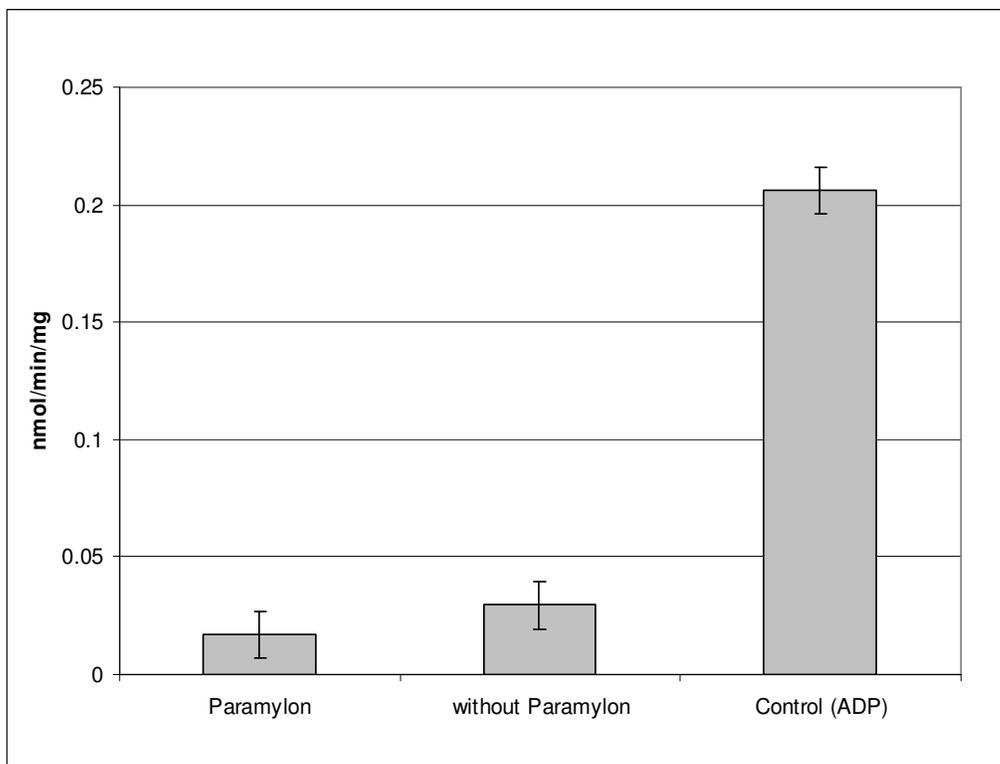


Figure 14. Activity determination of the purified fusion protein following enzymatic cleavage of the GST tag. The assay was attempted using paramylon as precursor however no increase in activity was evident. The control reaction using ADP showed a significantly higher yield thereby questioning the efficiency of the purified enzyme acting on the UDP-Glc substrate.

3.4. Conclusion

The functional screening method applied during this study proved sensitive enough to detect the exopolysaccharides being produced by the clones. Sequence homology to previously identified glucan synthases, as well as glycosyl transferase genes, suggested that the polymer was likely to be a glucan. Glucans are typically recognized as high-value polymeric compounds and, therefore, the potential exploitation of this clone could lead to the industrial production of a high value compound.

Measurement of the activity that the protein coded for by the gene has proved to be difficult. Very low levels of activity were achieved under non-denaturing conditions using PAGE, however parameter adjustments failed to provide improved expression. Results proved similar when using Schiff staining to detect the presence of glycoproteins that may have formed within the gel. This led to the application of a fusion tag system to enable purification of a homogenous protein. For this purpose, the pGEX expression vector was used as it produces a glutathione S-transferase fusion protein with the protein of interest. The glucan synthase gene was amplified out of the pCR2.1 TOPO vector using primers which were specifically designed to allow the PCR product to be ligated in-frame with the GST gene in the pGEX 4T-1 vector. Glutathione agarose purification was successfully employed to purify the GST-tagged protein and the fusion construct was measured to be approximately 31 kDa in size.

Again, demonstration of this proteins activity proved difficult. As the protein is only about 5 kDa in size, it might be that the presence of the much larger GST fused protein (26 kDa) would have interfered with the glycosylation action which the protein of interest is presumed to have. The tag was then removed by thrombin cleavage yet activity was still not detected. From this it can be deduced that either the protein is part of a larger complex and therefore does not function by itself or that it is only operational when expressed within the *E. coli* host due to the presence of co-factors, precursor molecules or pH within the organism which were not provided at the optimum levels in the assays. The determination of the activity of the enzyme remains, therefore, to be determined.

Chapter 4

Construction of a Metagenomic Library from Whey and Screening for Commercially Exploitable Polymers

4.1. Introduction

The collective genomes of all microorganisms present within a given microbial community are referred to as metagenomes. The construction and analysis of metagenomic libraries provides an invaluable tool to collect and archive environmental genetic resources (Ferrer et al. 2005). The method is naturally biased when representing the population's genomes within a given sample since dominant organisms would contribute a significantly larger proportion of the total DNA as apposed to organisms with a less frequent occurrence (Bohannan and Hughes, 2003), however it remains a rapid means of new gene discovery and its biotechnological potential is unquestionable.

Various species of lactic acid bacteria (LAB) are able to produce exopolysaccharides (EPSs), often as a means to provide protection against bacteriophage or protozoan attack as well as possible desiccation (Weiner *et al.* 1995). EPSs occur either attached to the cell surface as a capsule, or are liberated into the extracellular environment in the form of amorphous slime (Cerning, 1990). These molecules can be grouped depending on their respective sugar compositions. Homopolysaccharides are those consisting of a single type of monosaccharide, while heteropolysaccharides are composed of several types of monosaccharides (Gruter, 1992). Depending on the producing strain heteropolysaccharides may vary in chain length and degree of branching, with the structure composed of repeating units of up to eight monosaccharide residues. In some cases charged groups such as phosphate, glycerolphosphate and acetate are present (de Vuyst and Degeest, 1999).

Several species of LAB are able to produce dextran, mutan and levan using sucrose as a substrate (Sutherland, 1972). Strains belonging to the genus *Lactobacillus*, *Leuconostoc* and *Streptococcus* have been implicated in the production of dextran, a large class of extracellularly formed glucan polymers which consist primarily of α -1,6 linkages (Franz, 1986). Polymerization initially relies on the action of a glycosidase enzyme to catalyze the hydrolysis of the glycosidic bond to generate two smaller sugars. Glycosyltransferase enzymes then act by transferring the monosaccharide sugar unit to an acceptor molecule to form a polymer. These enzymes are usually dependent on the presence of metal ions such as magnesium or manganese which facilitate active-site binding. This process is demonstrated by dextransucrase, which catalyzes the synthesis of dextran outside the cell to produce D-glucose and D-fructose from sucrose. The free glucose unit is then transferred to an acceptor to form dextran. Similarly, the extracellular enzyme levansucrase, is able to produce levans by transferring D-fructose to growing fructan chains (Cerning, 1990). Heteropolysaccharides differ from homopolysaccharides as they are produced intracellularly, with nucleotide sugars acting as precursors for the synthesis of chains (Cerning, 1995). These EPSs consist largely of glucose and galactose as well as lesser amounts of fructose, mannose, rhamnose and galactosamine (van den Berg *et al.* 1995).

LAB are classified as food-grade microorganisms and have GRAS (Generally Regarded as Safe) status. These organisms are therefore particularly valuable in the food industry where EPS producers are used as viscosifying, gelling and emulsifying agents able to improve the texture and consistency of fermented foods (Cerning, 1990). Application also extends into the medical and pharmaceutical industries, as various *Lactobacillus* polysaccharides have demonstrated antitumoral (Oda *et al.* 1983) as well as cholesterol-lowering activity (Roberfroid *et al.* 1993).

The aim of this study was to construct and screen a metagenomic library from whey, a by-product of the cheese manufacturing process. As this environment hosts a vast range of microorganisms, including many LAB strains, it was envisaged that this may prove to be a viable source for the discovery of novel genes able to manufacture commercially exploitable polymers using lactose as a sole carbon source.

4.2. Materials and Methods

4.2.1. Sample collection

Whey samples (1 L) were collected from Fairview estate (Stellenbosch, Western Cape, South Africa) and Simonsberg cheese factory (Stellenbosch, Western Cape, South Africa). The samples were stored overnight at 4°C before extraction of genomic DNA.

4.2.2. Organisms

4.2.2.1. *Escherichia coli* strains

XL1-Blue MRF['] strain: $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac$ [F['] *proAB lacI^qZΔM15 Tn10 (Tet^r)*] (Stratagene).

XL0LRTM strain: $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA thi-1 recA1 gyrA96 relA1 lac$ [F['] *proAB lacI^qZΔM15 Tn10 (Tet^r)*] Su⁻ (nonsuppressing) λ^r (lambda resistant) (Stratagene).

BL21 strain: F⁻ *ompT gal dcm lon hsdS_B(r_B⁻ m_B⁻) λ(DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*])* (Promega).

4.2.3. Chemicals and kits

Enzymes, chemicals and kits were purchased from Sigma (St. Louis, Missouri, USA), Roche Diagnostics (Mannheim, Germany), Invitrogen (Carlsbad, California, USA), Promega (Madison, Wisconsin, USA), Stratagene (La Jolla, California, USA), CalBiochem (Merck Biosciences, Darmstadt, Germany), or QIAGEN (Hilden, Germany).

4.2.4. Transformation of competent *E.coli* cells

A 20 μ l aliquot *E.coli* (BL21) electrocompetant cells, prepared according to the method of Ausubel *et al.* (1987), were thawed on ice and the plasmid solution added with gentle stirring. The mixture was transferred to a pre-chilled electroporation cuvette (0.1 cm width) and pulsed (25 μ F, 200 Ω , 1.8 kV) using a Gene Pulser XcellTM electroporator (Bio-Rad). The cells were immediately resuspended in 1 ml of SOC medium (2% Tryptone, 0.5% Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄ and 20 mM glucose) and incubated for one hour at 37°C. Following incubation the cells were plated onto LB medium containing the appropriate antibiotic.

4.2.5. Genomic DNA extraction from whey

As the selected samples were collected during the cheese fermentation process, a large amount of unwanted debris was present which could influence the extraction procedure. Debris was removed by filtration through a syringe containing glass wool and the filtrate was pelleted at 5000 g for 5 min to collect all microorganisms. The pellet was washed twice with 500 μ L dH₂O and resuspended in 500 μ L solution A (20% sucrose; 1 M Tris-HCl pH 8.0; 1 M EDTA; 1 M NaCl) containing 50 μ L lysozyme (50 mg/mL). After 3 hrs at 37°C, 40 μ L Proteinase K (40 mg/mL) was added and sample incubated at 37°C for 30 min. Sufficient cell lysis was achieved following the addition of 20 μ L SDS (10% w/v) and 10 μ L NaOH (10N) for 30 min at 37°C. Protein was extracted with 1 vol phenol/chloroform/isoamylalcohol (25:24:1) and centrifuged at 8000 g for 10 min. The upper phase was carefully aspirated and traces of phenol removed by the addition of 1 vol

chloroform/isoamylalcohol (24:1). After centrifugation (10 000 g for 10 min) genomic DNA was precipitated from the upper phase with 0.1 vol 7.5 M NH₄OAc and 3 vol 100% EtOH. The pellet was washed with 70% EtOH, dried and resuspended in 400 µL TE buffer (pH 8.0).

4.2.6. Partial digestion of genomic DNA

Genomic DNA was partially digested with *Sau3A1*. The reaction was performed using 60 µL gDNA (919 ng/µL) with 0.2 µL *Sau3A1* (4U/ µL) in 180 µL final reaction volume. Aliquots were removed at 30 s intervals for the first 2 min, thereafter every 1 min. Digestion was stopped by adding DNA loading buffer containing EDTA followed by a 15 min incubation step at 65°C. Samples were subjected to gel electrophoresis (0.8% agarose). Fragments in the range of 11 kb and 5 kb were excised and purified using a gel extraction kit (Qiagen).

4.2.7. Construction of the genomic library

Sau3A1 digested genomic DNA fragments were ligated using equimolar amounts of the Zap Express phage vector (Stratagene) according to the manufacturer's guidelines. Following overnight ligation at 4°C, the recombinant lamda phage was packaged using the Gigapack III Gold packaging extract (Stratagene).

4.2.8. Library amplification

The library was amplified by adding aliquots of the packaged mixture to *E.coli* XL1-Blue MRF' cells resuspended in 10 mM MgSO₄ to an OD₆₀₀ of 1.0. Phage attachment to cells was achieved by incubation of the mixture at 37°C for 15 min. Infected bacteria were plated out according to the manufacturer's guidelines and the titer of the amplified library was calculated from the number of visible plaques formed.

4.2.9. Mass excision

Overnight cultures of *E.coli* XL1-Blue MRF⁺ and XLOLR cells were grown in LB broth with supplements (10 mM MgSO₄ and 6 mM maltose) at 30°C. Cells were pelleted at 2000 g and resuspended in 10 mM MgSO₄. The OD₆₀₀ of each culture was adjusted to 1.0. Phage attachment to the cells was achieved by combining 3 µL of the amplified library with 20 µL of the XL1-Blue MRF⁺ cells (OD₆₀₀ = 1.0) and 200 µL of ExAssist helper phage. This mixture was incubated at 37°C for 15 min, added to 20 mL of LB containing supplements (10 mM MgSO₄ and 6 mM maltose) and incubated for 3 h at 37°C. Lysis of the phage particles and cells was conducted at 70°C for 20 min. Following lysis, cell debris was pelleted at 3000 g for 10 min and the supernatant, containing the excised phagemids, was stored at 4°C until further use. The mass excision was tittered by combining 3 µL of this supernatant with 200 µL XLOLR cells (OD₆₀₀ = 1.0) and incubating the mixture at 37°C for 15 min. This was then added to 40 µL 5x NZY broth (1x: 5 g NaCl, 2 g MgSO₄.7H₂O, 5 g yeast extract, 10 g NZ amine, 1L dH₂O; pH was adjusted to 7.5 with NaOH) and placed at 37°C for 45 min. Hundred µL of the cell mixture was plated onto LB plates containing kanamycin (50 µg/mL) and incubated overnight at 37°C.

4.2.10. Library screening methods

The library was initially screened by top-agar plating of transformed BL21 *E.coli* cells onto LB media with appropriate antibiotic (50 µg/ml kanamycin) and in the presence of 5% (w/v) lactose as substrate and 0.003% (w/v) galactan as precursor. Congo Red (1% w/v) was used as a stain to detect possible extracellular polymer or 'slime' production.

Colonies staining positive for EPS production were isolated and replica plated onto LB media containing kanamycin and X-gal. Blue/White selection was used to distinguish between empty vectors and those containing inserts.

4.3. Results and Discussion

4.3.1. Extraction of genomic DNA

The crude whey sample used as a source for genomic DNA was initially filtered through a syringe containing glass wool. This provided a means to remove debris and slime which was prominent in the sample. The glass wool was also effective as it allowed microorganisms to pass through whilst trapping unwanted particle matter. The glass wool was regularly replaced during filtering to avoid clogging. The filtrate, containing the microbes, was centrifuged (3000 g for 5 min) and the collected cells treated with lysozyme to rupture cellular membranes. Sufficient cell lysis was only achieved upon extended treatment (3 hrs) with lysozyme. Following phenol/chloroform extraction and ethanol precipitation, a satisfactory yield (913 ng/ μ L in a final volume of 600 μ L) of high quality gDNA was obtained (figure 15).

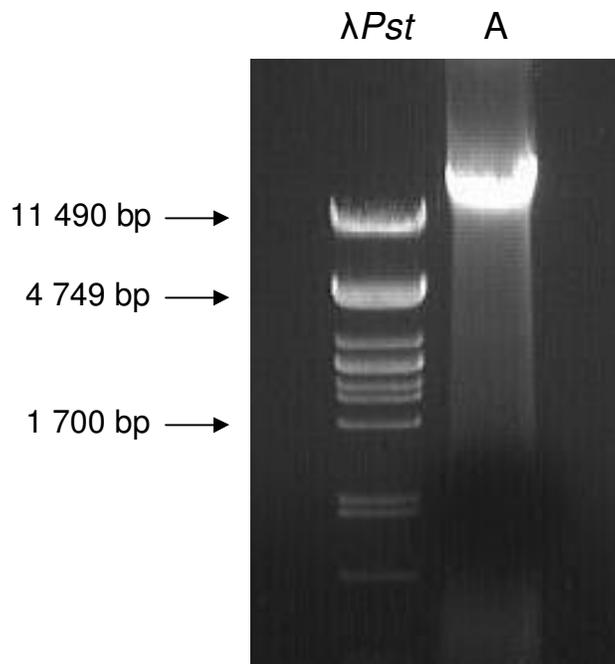


Figure 15. Lane A indicates the purified genomic DNA extracted from the whey sample.

4.3.2. Library construction and characterization

The restriction enzyme *Sau3A1* was used to fragment gDNA. Due to its 4 bp recognition site, this enzyme cuts frequently along the length of the gDNA thereby creating a large variation in fragment sizes. This in turn, allows for the construction of a library with a wide range of insert sizes. The partial digest was optimized by varying the amount of time for which the gDNA was treated with the enzyme. Intervals of 30 s were eventually sufficient for partial digestion to occur. A total of 5.48 μg of gDNA was digested with 0.75 U *Sau3A1* in a final reaction volume of 180 μL . Aliquots of 30 μL were removed every 30 s and the digestion halted. The addition of EDTA to the reaction was necessary as the heat-shock treatment alone was insufficient to prevent further enzymatic cleavage of the gDNA.

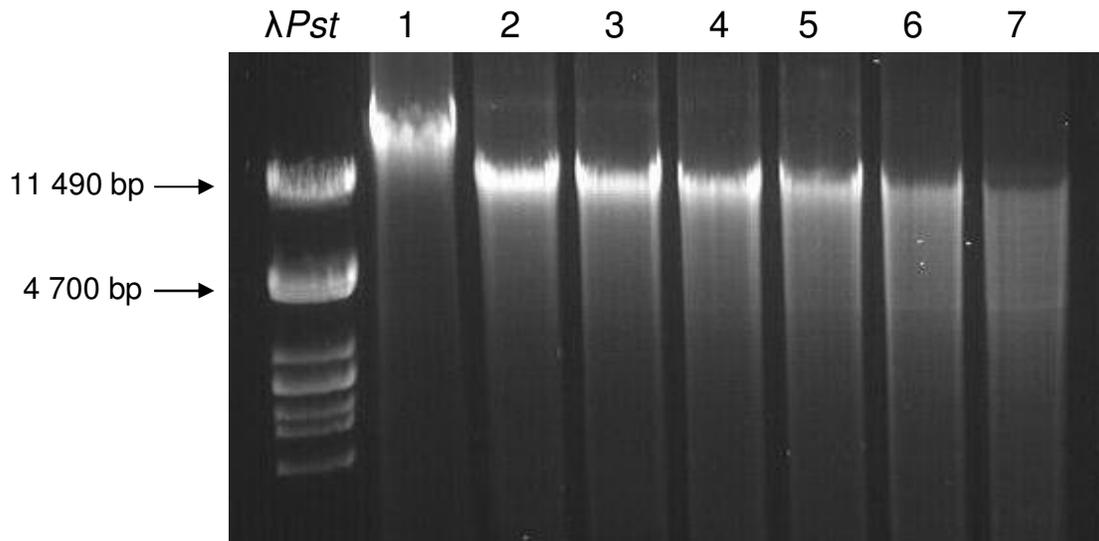


Figure 16. *Sau3A1* partial enzymatic digestion of genomic DNA. Lane 1 indicates the undigested DNA and lanes 2 – 7 show the partial digestion of the DNA over increasing 30 s time intervals.

Partial digested gDNA fragments (5 to 11 kb) purified from the agarose gel were quantified at 90 ng/ μ L using a NanoDrop® ND-1000 spectrophotometer and inserts were ligated into the pBK-CMV phagemid vector (Stratagene). Following ligation, Gigapack III Gold packaging extract (Stratagene) was used to package recombinant lambda phage. Library amplification was performed to create a large, stable quantity of high-titer stock and a final titer of 3.2×10^5 pfu/mL was obtained.

The amplified library was used to infect XL0LR *E.coli* cells for mass excision of the phagemid vector. Twenty-five colonies were selected and the plasmid DNA from each digested with the enzymes *Not1* and *Pst1*, which cut on either side of the cloned insert. The digested plasmids were run on an 0.8% agarose gel and the insert sizes measured against λ *Pst* molecular weight marker. Inserts ranged from 9.6 kb to 740 bp with the average calculated to be 2.85 kb.

4.3.3. Functional screening of the library

Whey was selected as a genetic source for this library as it is a by-product of milk fermentation in the production of cheese. It therefore houses microorganisms which thrive in a lactose-rich environment and which may polymerize the galactose moiety of lactose. Lactose is a disaccharide composed of one glucose and one galactose monosaccharide and can be hydrolyzed into these constituent sugars by the lactase enzyme, β -galactosidase.

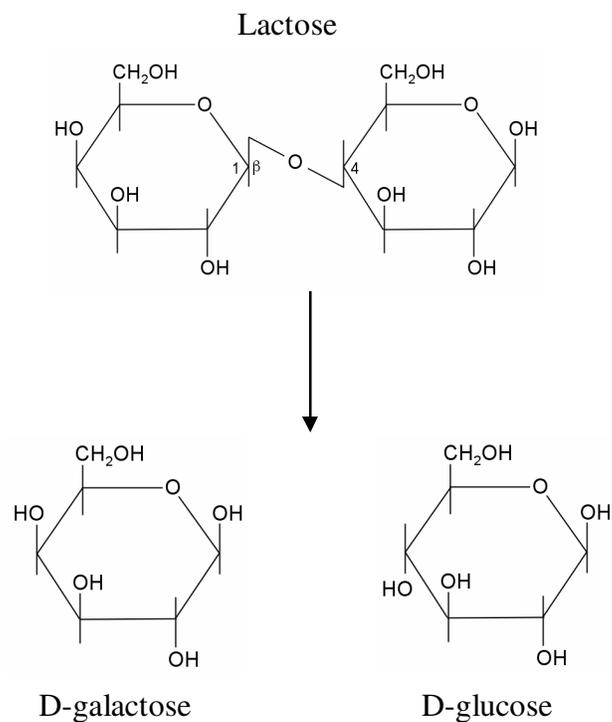


Figure 17. A schematic diagram indicating the monosaccharides, D-galactose and D-glucose produced following the hydrolysis of lactose. For the purpose of this study, the library was screened for clones exhibiting galactosyltransferase ability, capable of manufacturing a polymer composed of these galactose units.

BL21 *E.coli* cells were transformed with the library by electroporation. This method proved more efficient than heat-shock transformation and transformants had to be serially diluted before single colonies could be selected. Functional screening was applied for the detection of potential polymer-synthesizing clones using Congo Red. This stain displays a strong affinity for polysaccharides comprised of β -(1-4)-linked D-glucopyranosyl units, β -(1-3)-D-glucans as well as certain galactoglucmannans (Teather and Wood, 1982). The use of top-agar plating proved effective for the staining procedure as transformants remained embedded within the agar layer, thereby preventing spreading of colonies upon application of the stain. Colonies staining positive should be dark red compared to the much lighter orange color (Figure 18).

False positives can be eliminated by replica plating selected colonies onto LB agar containing X-gal and repeating the Congo Red stain. The presence of X-gal allows clones containing insert to remain white as apposed to empty vector containing transformants which would turn blue. Plasmid DNA could be extracted from positive clones and the inserts sequenced to aid identification.

Repeated screening attempts on 5 and 10% lactose substrates with galactan precursor provided no visible dark-red staining colonies. Following transformation, plates were incubated overnight at 37°C or until transformants appeared. These plates were stored at room temperature for a further two days before the stain was applied. Colonies were regularly inspected for noticeable morphological changes and for the production of slime. Although positive clones have yet to be identified, it is hoped that further screening of this library will produce an exopolysaccharide-producing clone on the lactose substrate. Screening for the utilization of other sugar substrates, such as sucrose, may also yield clones with the potential to manufacture novel polymers.

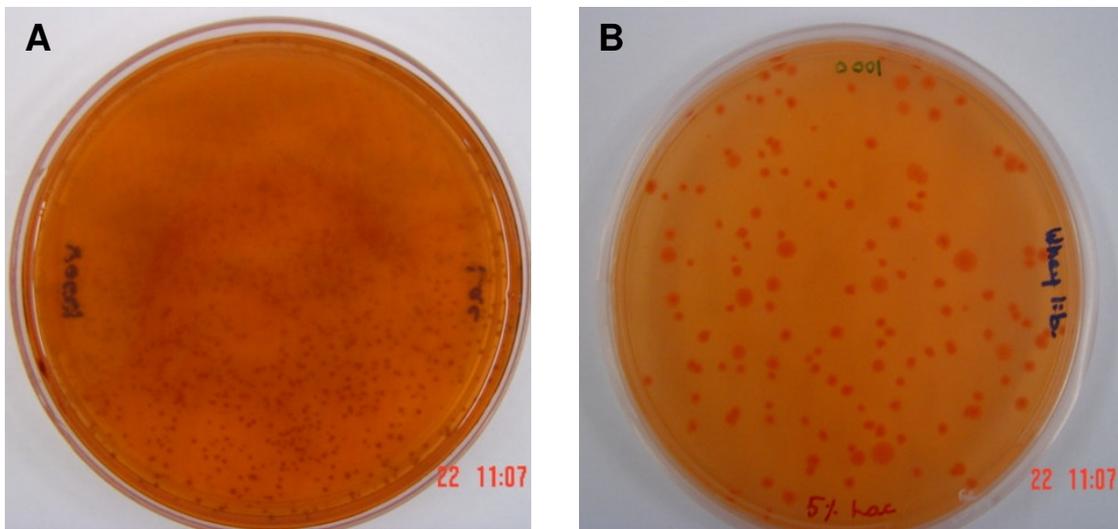


Figure 18. Transformed clones following staining with Congo Red. No visible dark-red colonies were identified. Plate B represents a 1000x dilution of transformants and provides a sufficient spread of colonies for the detection of exopolysaccharide production.

4.3.4. Screening the whey sample for exopolysaccharide producing bacteria

The method of Bauer *et al.* (2008, submitted for publication) was applied in an attempt to identify bacteria which are able to produce exopolysaccharides when provided with lactose as a sole source of carbon. Hundred mL of the whey sample was left at 25°C for 2 days before bacterial isolation. The sample was homogenized, serially diluted in sterile distilled water and plated onto MRS agar. Plates were incubated at 30°C until single colonies were visible. Five hundred single colonies were selected at random with each colony being streaked onto modified MRS agar (glucose being replaced with lactose) over an area of approximately 1cm². Isolates were screened for exopolysaccharide production after an incubation period of 3 days at 30°C.

Screening was achieved by inspecting bacterial colonies for slime (shiny colonies) and ropiness. Although the screen during this study did not result in the positive identification of EPS production on lactose, future attempts may provide success. Should a candidate be observed, I would proceed with further primary identification of the isolate. This would involve elucidation of colony morphology, pigment formation, cell morphology and arrangement, Gram-staining and catalase activity. Monosaccharide analysis of the EPS would be achieved by GC-MS.

4.4. Conclusion

The extraction of genomic DNA from the whey sample was optimized to allow for a substantial yield of quality DNA. Partial digestion of the DNA was achieved using *Sau3A1*, with fragments between 11 kb and 5 kb being used to successfully construct a metagenomic library. The library was calculated to have an average insert size of 2.85 kb. Functional screening was used to detect clones which are able to produce exopolysaccharides when provided with lactose as a carbon source. Numerous rounds of screening using Congo Red provided no evidence of polymer-producing clones, however it is hoped that future screening of this library might reveal a polymer composed of galactan moieties. The library will also be screened for other commercially exploitable polymers such as β -1,3- and β -1,4-glucans on a variety of sugar substrates including sucrose, glucose and glucosamine.

Chapter 5

General Discussion and Conclusion

The aim of the work presented in this thesis was to apply metagenomic strategies for the identification of novel polysaccharides with potential industrial applications. For this purpose, functional screening of metagenomic libraries was applied as it has previously been successful in the discovery of valuable biocatalysts and compounds (Lee *et al.* 2004; Brady and Clardy 2000; Voget *et al.* 2003). The major disadvantage encountered with this method of screening, is that it relies on the heterologous expression of a foreign gene within a suitable host organism. This often becomes laborious as thousands of clones need to be screened before a gene of interest is identified. Many potential open reading frames, coding for novel enzymes, would also be overlooked due to incompatible expression in the host. Screening for function, however, allows for the highly specific detection of the exact compound or enzyme of interest by using stains which bind only to the target molecule. This greatly simplifies the identification of positive clones as visual inspection is rapid and clones can readily be distinguished due to their change in color.

For the first part of this study, soil-derived metagenomic libraries were received from Bayer Biosciences as part of a collaborative project. These libraries were screened for β -1,3- and β -1,4-glucan production as these compounds have enormous biotechnological value, particularly for the medical industry. Functional screening was performed using aniline blue as this dye binds specifically to the β -1,3- and -1,4- bonds. A clone was subsequently identified which produced a positive-staining polymer when treated with aniline blue. The polymer was clearly present surrounding the clone of interest as shown by halo-staining. This therefore suggests that the compound was being excreted out of the cell wall and provided hope that the protein of interest was in fact not membrane-bound.

Isolation of the plasmid DNA and sequencing of the insert revealed high homology to previously identified β -1,4-glycosyltransferase and synthase genes, again confirming the aniline blue detection. The gene was determined to be 447 bp in size and the translated peptide sequence estimated at approximately 149 amino acids. The predicted hydrophobicity data for the membrane-spanning domains also indicated that the protein was not membrane bound.

The clone was cultivated in liquid media containing glucose substrate and allowed to manufacture the polymer. The polymer was purified, hydrolyzed and subjected to GC-MS analysis so as to determine its monosaccharide composition. This data showed that glucose was the predominant monosaccharide and the presence of carbohydrates and reducing sugars were confirmed by phenol-sulphuric acid and dinitro salicylic acid assays respectively. Further analysis of the polymer is required to elucidate its structure. This may be achieved through Fourier transform infrared spectroscopy (FTIR), nuclear magnetic resonance (NMR) and methylation analysis.

To improve our understanding of the putative glucan synthase clone, attempts were made to purify the enzyme by using classical protein purification methods. Native PAGE was applied with limited success to demonstrate in-gel protein activity, as faint detection of the active enzyme could not be reproduced. For the purpose of this study, UDP-glucose was provided as a possible substrate for the enzyme, however it could be that it required another activated sugar for polymerization. It is also possible that the enzyme was only active when expressed within the host organism due to the presence of co-factors and precursor molecules which were not provided under experimental conditions.

A further attempt to purify the protein to homogeneity was made using the pGEX expression system. The gene of interest was amplified out of the pCR2.1 TOPO library vector using sequence specific primers and cloned in-frame into the pGEX 4T-1 vector. The construct was transformed into BL21 *E.coli* to allow expression of the GST-tagged fusion protein. This homologous fusion protein was purified using glutathione agarose

and was shown to be approximately 31 kDa in size. As the GST tag is 26 kDa, the protein of interest could be estimated at 5 kDa.

Subsequent efforts to demonstrate activity of the fusion protein were unsuccessful and could be attributed to the small size of the protein of interest relative to the GST tag. Removal of the tag also had no influence on detectable activity and this could be due to conformational changes in the structure of the protein induced by the addition and/or removal of the tag.

The second part of this study focused on the construction of a metagenomic library from an environmental sample which has the potential to house novel polymer-producing microorganisms. For this purpose, whey was selected as an appropriate source as it is a by-product of the cheese manufacturing process and therefore should contain a wealth of lactic acid bacteria as well as other microbes. A method was adapted to successfully extract a large amount of quality DNA from the sample and this was used in conjunction with the ZAP Express® Library kit (Stratgene) to produce a high-titer library stock. The library was calculated to have an average insert size of 2.85 kb and this was suitable for expression screening. Initial functional screens were performed by plating clones onto a lactose-rich substrate in attempts to obtain polymers composed predominantly of galactose moieties. Congo Red was used as a stain and although no positive clones have been detected, it is hoped that future screening of this library may uncover commercially exploitable polysaccharides.

In conclusion, a putative glucan synthase gene was isolated from a soil-derived metagenomic library and a library was constructed using whey as a genomic resource. Future studies could be directed at further assessment of the biotechnological potential of the polymer as well as possible expression of the gene within a plant system thereby using the host as a bioreactor for the production of the compound.

Chapter 6

Literature Cited

1. Internet: <http://expasy.org/tools> (January 2008)
2. Internet: <http://blast.ncbi.nlm.nih.gov/Blast.cgi> (March 2007)

Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiology Reviews* 59: 143-169

Arioli, T., Peng, L., Betznet, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Hofte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J. and Williamson, R.E. (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science* 279: 717-720

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds.) (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.

Bauer, R., Bekker, J.P., van Wyk, N., du Toit, C., Dicks, L.M.T. and Kossmann, J. (2008) Exopolysaccharide production by lactose-hydrolyzing bacteria isolated from traditionally fermented milk. Submitted for publication.

Balser, K., Hoppe, L., Eichler, T., Wendel, M. and Astheimer, A-J. (1986) Ullmann's *Encyclopedia of Industrial Chemistry*, Vol. A5 (Eds.: W. Gerhartz, Y. S. Yamamoto, F. T. Campbell, R. Pfefferkorn, J. F. Rounsaville), VCH, Weinheim, pp. 419 – 459

Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M. Feldman, R.A. and Spudich, J.L. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289: 1902-1906

van den Berg, D.J.C., Robijn, G.W., Janssen, A.C., Giuseppin, M.L.F., Vreeker, R., Kamerling, J.P., Vliegthart, J.F.G., Ledebor, A.M. and Verrips, T. (1995) Production of a novel extracellular polysaccharide by *Lactobacillus sake* 0-1 and characterization of the polysaccharide. *Applied and Environmental Microbiology* 61: 2840-2844

Bergmeyer, H.U. and Bernt, E. (1974) Sucrose. In: *Methods of enzymatic analysis*, 2nd edition 1176-1179

Bohannan, B.J.M. and Hughes, J. (2003) New approaches to analyzing microbial biodiversity data. *Current Opinion in Microbiology* 6: 282-287

Bradford, M.M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248-254

Brady, S.F., Chao, C.J. and Clardy, J. (2002) New natural product families from an environmental DNA (eDNA) gene cluster. *Journal of the American Chemical Society* 124: 9968-9969

Brady, S.F., Chao, C.J., Handelsman, J. and Clardy, J. (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Organic Letters* 3: 1981-1984

Brady, S.F. and Clardy, J. (2000) Long-chain N-acetyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *Journal of the American Chemical Society* 122: 12903-12904

Brown, R.M., Saxena, I.M. and Kudlicka, K. (1996) Cellulose biosynthesis in higher plants. *Trends in Plant Science* 1(5): 149-156

Bulone, V., Girard, V. and Fevre, M. (1990) Separation and Partial Purification of 1,3- β -Glucan and 1,4- β -Glucan Synthases from *Saprolegnia*. *Plant Physiology* 94: 1748-1755

Burkus, Z. and Temelli, F. (2005) Rheological properties of barley β -glucan. *Carbohydrate Polymers* 59: 459-465

Cerning, J. (1990) Exocellular polysaccharides produced by lactic acid bacteria. *FEMS Microbiology Reviews* 87: 113-130

Cerning, J. (1995) Production of exopolysaccharides by lactic acid bacteria and dairy propionibacteria. *Lait* 75: 463-472

Coque, T.M., Oliver, A., Perez-Diaz, J.C., Baquero, F. and Canton, R. (2002) Genes encoding TEM-4, SHV-2, and CTX-M-10 extended-spectrum beta-lactamases are carried by multiple *Klebsiella pneumoniae* clones in a single hospital (Madrid, 1989 to 2000). *Antimicrobial Agents and Chemotherapy* 46:500–510

Cottrell, M.T., Moore, J.A. and Kirchman, D.L. (1999) Chitinases from uncultured marine microorganisms. *Applied and Environmental Microbiology* 65: 2553-2557

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J. and Osbourne, M.S. (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology* 69: 49-55

Daniel, R. (2004) The soil metagenome - a rich resource for the discovery of novel natural products. *Current Opinion Biotechnology* 15: 199–204

Davis, B.G. and Boyer, V. (2001) Biocatalysis and enzymes in organic synthesis. *Natural Product Reports* 18: 618-640

De Vuyst, L. and Degeest, B. (1999) Heteropolysaccharides from lactic acid bacteria. *FEMS Microbiology Reviews* 23: 153-177

Diaz-Torres, M.L., McNab, R., Spratt, D.A., Villedieu, A. and Hunt, N. (2003) Novel tetracycline resistance determinant from the oral metagenome. *Antimicrobial Agents and Chemotherapy* 47: 1430-1432

Eiberger, L.L. and Wasserman, P.B. (1987) Partial purification of digitonin-solubilized β glucan synthase from red beet root. *Plant Physiology* 83: 982-987

Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R. (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Applied and Environmental Microbiology* 65: 89-99

Ferrer, M., Martinez-Abarca, F. and Golyshin, P. N. (2005) Mining genomes and 'metagenomes' for novel catalysts. *Current Opinion in Biotechnology* 16: 588-593

Fevre, M. and Rougier, M. (1981) 1,3-, β and 1,4- β Glucan synthesis by membrane fractions from the fungus *Saprolegnia*. *Planta* 157: 232-241

Franz, G. (1986) Polysaccharides in pharmacy. *Advances in Polymer Science* 76: 1-30

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M. and Handelsman, J. (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology* 68: 4301-4306

Gruter, M. (1992) Structural characterization of polysaccharides from lactic acid bacteria. Ph.D Thesis, Utrecht University, The Netherlands.

Henne, A., Daniel, R., Schmitz, R.A. and Gottschalk, G. (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Applied and Environmental Microbiology* 65: 3901-3907

Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G. and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Applied and Environmental Microbiology* 66: 3113-3116

Jaeger, K.E. and Reetz, M.T. (1998) Microbial lipases form versatile tools for biotechnology. *Trends in Biotechnology* 16: 396-403

Kim, Y-J., Choi, G-S., Kim, S-B., Yoon, G-S., Kim, Y-S. and Ryu, Y-W. (2005) Screening of a novel esterase from a metagenomic library. In Press.

Khalameyzer, V., Fischer, I., Bornscheuer, U.T. and Altenbuchner, J. (1999) Screening, nucleotide sequence, and biochemical characterization of an esterase from *Pseudomonas fluorescens* with high activity towards lactones. *Applied and Environmental Microbiology* 65 (2): 477-482

Klemm, D., Heublein, B, Fink, H-P. and Bohn, A. (2005) Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *Angewandte Chemie* 44: 3358-3393

Knietsch, A., Waschowitz, T., Bowien, S., Henne, A. and Daniel, R. (2003) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Applied and Environmental Microbiology* 69: 1408-1416

Kobayashi, S., Shoda, S. and Uyama, H. (1995) Enzymatic polymerization and oligomerization. *Advances in Polymer Science* 121: 1-30

Kren, V. and Thiem, J. (1997) Glycosylation employing bio-systems: from enzymes to whole cells. *Chemical Society Reviews* 26: 215-238

Kyte, J. and Doolittle, R. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105-132

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685

Langer, M., Esther, M., Liebeton, K., Meurer, G., Niehaus, F., Schulze, R., Eck, J. and Lorenz, P. (2006) Metagenomics: An inexhaustible access to nature's diversity. *Biotechnology Journal* 1: 815-821

Lee, S-W., Won, K., Lim, H.K., Kim, J-C., Choi, G.J. and Cho, K.Y. (2004) Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Applied Microbiology and Biotechnology* 65: 720-726

Li, Y., Wexler, M., Richardson, D.J., Bond, P.L. and Johnston, A.W.B. (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes. *Environmental Microbiology* 7(12): 1927-1936

Lorenz, P. and Eck, J. (2005) Metagenomics and industrial applications. *Nature Reviews Microbiology* 3: 510-515

Lorenz, P., Liebeton, K., Niehaus, F. and Eck, J. (2002) Screening for novel biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Current Opinion in Biotechnology* 13: 572-577

Majernik, A., Gottschalk, G. and Daniel, R. (2001) Screening environmental DNA libraries for the presence of genes conferring $\text{Na}^+(\text{Li}^+)/\text{H}^+$ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *Journal of Bacteriology* 183: 6645-6653

Miller, C.A. (2000) Advances in enzyme discovery technology: capturing diversity. *Biotechnology Reports* 11: 489-495

Nakanishi, I., Kimura, K., Kusui, S. and Yamazaki, E. (1974) Complex formation of gel-forming bacterial (1-3)- β -D-glucans (curdlan type polysaccharides) with dyes in aqueous solution. *Carbohydrate Research* 32: 47-52

Oda, M., Hasegara, H., Komatsu, S., Kambe, M. and Tsuchiya, F. (1983) Antitumoral polysaccharide from *Lactobacillus* ssp. *Agricultural and Biological Chemistry* 47: 1623-1625

Pace, N.R., Stahl, D.A., Lane, D.J. and Olsen, G.J. (1985) Analyzing natural microbial populations by rRNA sequences. *ASM News* 51:4-12

Riesenfeld, C.S., Schloss, P.D. and Handelsman, J. (2004) Metagenomics: Genomic analysis of microbial communities. *Annual Review of Genetics* 38: 525-52

Roberfroid, M.B., Gibson, G.R. and Delzenne, N. (1993) Biochemistry of oligofructose, a non-digestible fructooligosaccharide: an approach to estimate its caloric value. *Nutrition Reviews* 51: 137-146

Roberts, S.M. (2004) Biocatalysts in synthetic organic chemistry. *Tetrahedron* 60: 499-500

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F. and Grossman, T.H. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* 66: 2541–47

Roubroeks, J.P., Andersson, R., Mastromauro, D.I., Christensen, B.E. and Åman, P. (2001) Molecular weight, structure and shape of oat (1-3),(1-4)- β -D-glucan fractions obtained by enzymatic degradation with (1-4)- β -D-glucan 4-glucanohydrolase from *Trichoderma reesei*. *Carbohydrate Polymers* 46: 275-285

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Schepens, H. *et al.* (2003) *White Biotechnology: gateway to a more sustainable future* [online] www.mckinsey.com/client-service/chemicals/pdf/BioVision_Booklet_final.pdf (May 2006)

Schloss, P.D. and Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology* 14: 303-310

Schmid, A., Dordick, J.S., Hauer, B. and Kiener, A. (2001) Industrial biocatalysis today and tomorrow. *Nature* 409: 258-268

Schmidt, T., DeLong, E. and Pace, N. (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology* 173: 4371-4378

Sebat, J.L., Colwell, F.S. and Crawford, R.L. (2003) Metagenomic profiling: microarray analysis of an environmental genomic library. *Applied and Environmental Microbiology* 69: 4927-4934

Selitrechnikoff, C.P. (1995) Antifungal drugs: (1,3)- β -glucan synthase inhibitors. R.G. Landes Company. Austin, USA.

Steele, H.L. and Streit, W.R. (2005) Metagenomics: Advances in ecology and biotechnology. *FEMS Microbiology Letters* 247: 105-111

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. and Delong, E.F. (1996) Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* 178(3): 591-599

Stokes, H.W., Holmes, A.J., Nield, B.S., Holley, M.P., Nevalainen, K.M.H., Mabbutt, B.C. and Gillings, M.R. (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Applied Environmental Microbiology* 67: 5240-5246

Streit, W.R., Daniel, R. and Jaeger, K.-E. (2004) Prospecting for biocatalysts and drugs in the genomes of non-cultured microorganisms. *Current Opinion in Biotechnology* 15: 285–290

Streit, W.R. and Schmitz, R.A. (2004) Metagenomics – the key to uncultured microbes. *Current Opinion in Microbiology* 7: 492-498

Sutherland, I.W. (1972) Bacterial exopolysaccharides. *Advances in Microbial Physiology* 8: 143-212

Teather, R.M. and Wood, P.J. (1982) Use of Congo Red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and Environmental Microbiology* 43(4): 777-780

Torsvik, V., Daae, F.L., Sandaa, R-A. and Ovreas, L. (1998) Novel techniques for analyzing microbial diversity in natural and perturbed environments. *Journal of Biotechnology* 64: 53-62

Tsai, Y.L. and Olson, B.H. (1992) Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology* 58: 2292-2295

Uchiyama, T., Abe, T., Ikemura, T. and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenomic libraries for isolation of catabolic genes. *Nature Biotechnology* 23: 88-93

U.S. Congress, Office of Technology Assessment, *Biopolymers: Making Materials Nature's Way – Background Paper, OTA-BP-E-102* (Washington, DC: U.S. Government Printing Office, September 1993)

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E. and Nelson, W. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66-74

Voget, S., Leggewie, C., Uesbeck, A., Raasch, C. and Jaeger, K.E. (2003) Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology* 69: 6235-6242

Weiner, R., Langille, S. and Quintero, E. (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *Journal of Industrial Microbiology* 15: 339-346

Whitesides, G.M. and Wong, C-H. (1985) Enzymes as catalysts in synthetic organic chemistry. *Angewandte Chemie International Edition* 24: 617-718

Wilkinson, D.E., Jeanicke, T. and Cowan, D.A. (2002) Efficient molecular cloning of environmental DNA from geothermal sediments. *Biotechnology Letters* 24: 155-161

Xu, J. (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular Ecology* 15: 1713-1731

Yap, W., Soong, T. and Davies, J. (1996) Genetic diversity of soil microorganisms assessed by analysis of hsp70 (dnaK) sequences. *Journal of Industrial Microbiology* 17: 179-184

Yun, J., Kang, S., Park, S., Yoon, H., Kom, M-J., Heu, S. and Ryu, S. (2004) Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and Environmental Microbiology* 70(12): 7229-7235

Zhou, J. (2003) Microarrays for bacterial detection and microbial community analysis. *Current Opinions in Microbiology* 6: 288-294