

**ISOLATION AND CHARACTERISATION OF NOVEL ENZYMES PRODUCING  
FRUCTO- AND GALACTO-OLIGOSACCHARIDES THROUGH A METAGENOMIC  
APPROACH**

**Marthinus Luther Smith**



UNIVERSITEIT  
iYUNIVESITHI  
STELLENBOSCH  
UNIVERSITY

100  
1918 · 2018

Dissertation presented for the degree of Doctor of Science in the Faculty of Sciences at  
Stellenbosch University

**Supervisor: Professor Jens Kossmann**

**Co-supervisor: Dr James R Lloyd**

March 2018

## DECLARATION

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

Date: March 2018

---

**Marthinus Luther Smith**

Copyright © 2018 Stellenbosch University  
All rights reserved

Copyright © 2017 Stellenbosch University

All rights reserved

## ABSTRACT

Prebiotic oligosaccharides and polymers are a form of dietary carbohydrate that are selectively fermented by the microbiota of the lower gastro intestinal tract and confer a health benefit to the host other than that of nutrition. Galacto-oligosaccharides (GOS) are a prime example of a dietary prebiotic and closely imitate human milk oligosaccharides. They consist of galactose moieties, often linked to a terminal glucose. GOS promote the growth of beneficial bacteria that aid in the prevention and treatment of gastro intestinal dysbiosis which has been linked to various diseases such as auto-immune disorders, certain cancers and metabolic disorders. They are often synthesised from lactose by the glycoside hydrolase enzyme,  $\beta$ -galactosidase. A cheap substrate for the production of GOS is whey permeate, an industrial waste product. Another prebiotic dietary carbohydrate is Fructo-oligosaccharides (FOS), which consist of fructose moieties often linked to a terminal glucose, and much like GOS is associated with the numerous health benefits of prebiotics. FOS is synthesised from the substrate sucrose by fructosyl-transferases. There is a great demand for new enzymatic biocatalysts with improved capabilities of producing GOS and FOS. In this study the construction and screening of numerous metagenomic libraries for prebiotic oligosaccharide and polymer synthesising enzymes are presented. The metagenomic DNA was sampled from locations rich in the substrate used for the production of the before mentioned prebiotics. The isolation, amplification and *in silico* analysis of three novel genes encoding proteins belonging to the Glycoside Hydrolase families 1, 2, and 68 are presented here. Their respective polypeptides were heterologously expressed in *E. coli*, purified and subsequently characterised biochemically *in vitro* through various means, and so too their respective products. Glycoside hydrolases often have two innate but distinct activities namely hydrolysis and trans-glycosylation. All three the glycoside hydrolases described here, MS- $\beta$ Gluc1, HW- $\beta$ Gal1 and SAS-Ls, had high trans-glycosylation activity apart from their hydrolytic activity. They are all excellent candidates for the industrial production of prebiotics, whether it be GOS (MS- $\beta$ Gluc1 and HW- $\beta$ Gal1) or FOS (SAS-Ls).

## OPSOMMING

Prebiotiese oligo-sakkariede en polimere is 'n tipe voedings koolhidraat wat selektief gefermenteer word deur die mikro-biota van die laer gastro-intestinale sisteem, en lewer 'n positiewe gesondheids eienskap bydrae aan die gasheer anders as die oog op liggende voedings waarde. Galakto-oligosakkariede (GOS) is 'n uitstekende voorbeeld van prebiotiese oligo-sakkariede en is 'n nouliks nabootsing van menslike melk oligosakkariede. Hulle bestaan uit opeenvolgende galaktose eenhede wat baie keer aan 'n terminale glukose molekule gekoppel is. GOS bevorder die groei van voordelige bakterieë wat help met die voorkoming en behandeling van gastro-intestinale disbiose wat verwant hou met verskeie siektetoestande soos auto-immuniteit wanorde, sekere kankersoorte asook metaboliese wanorde. GOS word vervaardig vanaf melk laktose deur die glikosied hidrolase ensiem  $\beta$ -galaktosidase. 'n Goedkoop substraat vir die produksie van GOS bestaan in die vorm van industriële wei invulling. Nog 'n goeie voorbeeld van 'n prebiotiese voedings koolhidraat is Frukto-oligosakkariede (FOS), wat bestaan uit opeenvolgende fruktose eenhede wat dikwels gekoppel is aan 'n terminale glukose molekule. Net soos GOS word FOS geassosieer met die verskeie gesondheids voordele van prebiotiese aanvulling. FOS word gesintetiseer vanaf die substraat sukrose deur ensieme genaamd fruktosiel-transferases. Daar is 'n groot aanvraag vir nuwe bio-katalise met verbeterde eienskappe vir die produksie van GOS en FOS. In hierdie studie lê ons ten toon die konstruksie en toetsing van verskeie meta-genoom biblioteke waarvan die meta-genoom DNA verkry is van gebiede wat ryk is in die substrate vir die produksie van prebiotiese oligo-sakkariede en polimere, naamlik laktose en sukrose. Vêrder beskryf ons ook die isolasie, amplifikasie en *in silico* analise van drie eensoortige glikosied hidrolase familie 1, 2 en 68 proteïene, onderskeidelik. Die drie proteïene genaamd HW- $\beta$ gal1, MS- $\beta$ gluc1 en SAS-Ls is heteroloog uitgedruk in *E. coli*, gesuiwer en biochemies gekarakteriseer tesame met hul onderskeie substraat produkte. Glikosied-hidrolases het dikwels twee ingebore onafhanklike en onderskeibare aktiwiteite genaamd hidrolise en trans-glikolisering. Al drie die glikosied-hidrolases wat hier beskryf word, het besondere hoë trans-glikolisering aktiwiteite hetsy vir GOS (MS- $\beta$ Gluc1 and HW- $\beta$ Gal1) of vir FOS (SAS-Ls), wat hul uitstekende kandidate maak vir industriële implementering en toepassing.

## ACKNOWLEDGEMENTS

To all the staff and students and most notably my family that stood by and supported me through this study.

**TABLE OF CONTENTS**

<b>DECLARATION</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>iii</b>
<b>OPSOMMING</b> .....	<b>iv</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>v</b>
<b>TABLE OF CONTENTS</b> .....	<b>vi</b>
<b>LIST OF FIGURES</b> .....	<b>xi</b>
<b>LIST OF TABLES</b> .....	<b>xvi</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>xvii</b>
<b>CHAPTER 1: General Introduction</b> .....	<b>1</b>
1.1 Background.....	1
1.2 Brief review of objectives and methodological approach.....	3
1.3 Dissertation layout .....	3
<b>CHAPTER 2: Literature Review</b> .....	<b>5</b>
2.1 The human microbiota and microbiome.....	5
2.2 The core Microbiota / Microbiome .....	6
2.3 Contribution of the core towards health .....	7
2.3.1 Metabolic Functions .....	8
2.3.2 Mucosal epithelium effect and defence barrier.....	9
2.3.3 Host immune modulation .....	9
2.4 Dysbiosis .....	10
2.4.1 Effect of microbial metabolites on health.....	10
2.4.2 Dysbiosis and GI tract disorders .....	11
2.4.3 Dysbiosis and systemic diseases.....	12
2.4.4 The gut-brain-axis: Dysbiosis and central nervous system disorders .....	13
2.5 Manipulating the gut microbiota to maintain health and treat disease.....	14
2.5.1 Probiotics.....	14

2.5.2 Microbial faecal transplants .....	15
2.5.3 Prebiotics .....	16
2.5.4 Synbiotics.....	16
2.6 Introductory review of Prebiotic synthesis by microbial enzymes and their classification .....	17
2.6.1 Structural characteristics and catalytic mechanisms of microbial Glycoside Hydrolases .....	21
2.6.2 Fructose, Fructans and Fructo-oligosaccharides (FOS), sources and synthesis .....	21
2.6.3 Galacto-oligosaccharides (GOS) as prebiotic, sources and synthesis. ....	27
<b>CHAPTER 3: Metagenomic library construction, screening and preliminary analysis. ....</b>	<b>32</b>
3.1 Introduction.....	32
3.2 Materials and Methods .....	34
3.2.1 Sample collection and DNA isolation .....	34
3.2.2 Metagenomic library construction and in vivo mass excision .....	36
3.2.3 $\beta$ -galactosidase mutant construction.....	40
3.2.4 Library screening and media used.....	42
3.2.5 Sequencing and bioinformatic analysis .....	45
3.2.6 Cloning, expression and protein analysis .....	45
3.2.7 Nitrophenyl based biochemical characterisation of positive clones.....	46
3.2.8 Product characterisation by spectrophotometric analysis of free reducing sugars .....	47
3.2.9 Product characterisation by thin layer chromatography .....	48
3.3 Results and discussion.....	48
3.3.1 Library construction and screening in the mutant.....	48
3.3.2 Sequencing and bioinformatic analysis.....	53
3.3.3 Protein analysis and biochemical characterisation .....	55
3.3.4 Product analysis.....	61
3.4 Conclusion .....	63
<b>CHAPTER 4: The biochemical characterisation of two novel glycoside hydrolases, a <math>\beta</math>-galactosidase and <math>\beta</math>-glucosidase, belonging to the Glycoside Hydrolase families 1 &amp; 2 respectively, both with high trans-galactosylation activity. ....</b>	<b>64</b>
4.1 Abstract .....	64

4.2 Introduction.....	65
4.3 Materials and methods .....	66
4.3.1 Strains, plasmids and materials.....	66
4.3.2 Library construction and screening, clone selection, sequencing and in silico analysis .....	67
4.3.3 Cloning, expression and recombinant protein purification .....	67
4.3.4 Nitrophenyl based assays and enzyme linked assays for enzymatic characterisation. ....	68
4.3.5 Biochemical characterization with the substrate lactose, through enzyme linked assay.....	69
4.3.6 Enzymatic synthesis of GOS.....	70
4.3.7 Carbohydrate analysis .....	70
4.4 Results and discussion.....	72
4.4.1 Library construction, screening, isolation and preliminary characterisation.....	72
4.4.2 In silico analysis of HW- $\beta$ gal1 and MS- $\beta$ gluc1 .....	72
4.4.3 Heterologous expression and purification of the HW- $\beta$ gal1 and MS- $\beta$ gluc1 proteins .....	74
4.4.4 Substrate specificity of HW- $\beta$ gal1 and MS- $\beta$ gluc1 .....	75
4.4.5 The effects of temperature and pH on HW- $\beta$ gal1 and MS- $\beta$ gluc1 activity .....	76
4.4.6 The effects of co-factors on HW- $\beta$ gal1 and MS- $\beta$ gluc1 activity .....	77
4.4.7 Reaction kinetics for HW- $\beta$ gal1 and MS- $\beta$ gluc1.....	78
4.4.8 Product characterisation .....	80
4.5 Conclusion .....	82
<b>CHAPTER 5: Production of a <math>\beta</math>-2,6 extracellular fructan by a novel metagenome derived levansucrase</b> .....	<b>83</b>
5.1 Abstract .....	83
5.2 Introduction.....	83
5.3 Materials and methods .....	85
5.3.1 Strains, plasmids and materials.....	85
5.3.2 Library construction and screening, clone selection, sequencing and in silico analysis .....	85
5.3.3 Recombinant protein production and purification .....	85
5.3.4 Biochemical characterisation of SAS-Ls.....	88
5.3.5 Enzymatic synthesis of levan and carbohydrate analysis.....	89

5.4 Results and discussion .....	90
5.4.1 Library construction, screening and isolation of SAS-Ls.....	90
5.4.2 In silico analysis of SAS-Ls.....	90
5.4.3 Heterologous expression and purification of SAS-Ls.....	92
5.4.4 The effects of temperature and pH on SAS-Ls activity.....	93
5.4.5 The effects of co-factors on SAS-Ls activity.....	94
5.4.6 Reaction kinetics of SAS-Ls .....	95
5.4.7 NMR analysis and product characterisation .....	96
5.5 Conclusion .....	98
<b>CHAPTER 6: General discussion and future prospects .....</b>	<b>99</b>
6.1 The metagenomic approach.....	99
6.1.1 Origins and context of metagenomic analysis for novel product discovery .....	99
6.1.2 Environmental sampling and diversity assessment .....	100
6.1.3 Metagenomic DNA isolation and preparation .....	102
6.1.4 The construction of the metagenomic libraries.....	103
6.1.5 Approaches toward metagenomic bioprospecting.....	104
6.2 Enzymatic production of prebiotic galacto-and fructo-oligosaccharides .....	106
6.2.1 The bifidogenic effect and prebiotic relevance of enzymatically produced oligosaccharides .....	106
6.2.2 Composition of the GOS synthesised by $\beta$ -galactosidases and $\beta$ -glucosidases .....	109
6.2.3 Hetero-oligosaccharides synthesised by glycoside hydrolases.....	110
6.3 Summary and concluding remarks.....	112
<b>APPENDIX .....</b>	<b>115</b>
A1    pBK-CMV vector map .....	115
A2    pRSET A, B and C vector map. ....	116
A3    GB-BGal1 sequence .....	117
A4    HW-Bgal1 sequence .....	118
A5    HW-Bgal2 sequence .....	119
A6    HW-Bgal3 sequence .....	120

A7	HW-Bgal4 sequence .....	121
A8	HW-Bgal5 sequence .....	122
A9	HW-Bgluc1 sequence.....	123
A10	HW-Bgluc2 sequence.....	124
A11	MS-Bgal1 sequence .....	125
A12	MS-Bgal2 sequence .....	126
A13	MS-Bgal3 sequence .....	127
A14	MS-Gluc1 sequence.....	128
A15	MS-Gluc2 sequence.....	129
A16	SAS-Bgal1 sequence .....	130
A17	SAS-BGal2 sequence.....	131
A18	SAS-BGal3 sequence.....	132
A19	SAS-LS sequence.....	133
<b>BIBLIOGRAPHY .....</b>		<b>134</b>

## LIST OF FIGURES

- Figure 2-1 The respective covers of Nature (Vol 464 Number 7285), Scientific American (Vol 306 Issue 6), Science (Vol 336 Issue 6086) and The Economist (Aug 18<sup>th</sup> 2012), indicating the relevance of the human microbiota and microbiome. .... 6
- Figure 2-2 The core human microbiome (red) is the set of genes present in the majority of humans. The variable human microbiome (blue) is the set of genes present in smaller subset of humans. This variation could result from a combination of factors such as host genotype, host physiological status, host pathobiology (disease status), host lifestyle, host environment and the presence of transient populations of microorganisms that cannot persistently colonize a habitat. .... 7
- Figure 2-3 The comparative and integrated hepatic metabolism of glucose and fructose: ATP (adenosine triphosphate); ADP (adenosine diphosphate); DHA (dihydroxyacetone); DHAP (dihydroxyacetone phosphate); fruc-1-P (fructose-1-phosphate); fruc-6-p (fructose-6-phospahte); fruc-1,6-bisP (fructose-1,6-bisphosphate); G-6-P (glucose-6-phospahte); Glc-6-Pase (Glucose-6-phosphatase); FDPase (Fructose-1,6-diphosphatase); GAP (D-glyceraldehyde 3-phosphate); PFK-1 (phosphofructokinase); Pi (phosphate) ..... 22
- Figure 2-4 Fructan biosynthesis of the five sub-types of fructans in plants: Glc (glucose); Fruc (fructose); 1-SST (sucrose: sucrose 1-fructosyl transferase); 1-FFT (fructan: fructan 1-fructosyl transferase); 6G-FFT (fructan: fructan 6G-fructosyl transferase); 6-SST (sucrose: sucrose 6-fructosyl transferase); 6-FFT (sucrose: fructan 6-fructosyl transferase). Note that bacterial and yeast levansucrase activity (red box and arrows) is seen as a combination of 6-SST/6-FFT activity when compared to that of plants. .... 24
- Figure 2-5  $\beta$ -galactosidase activity mechanism: EnZ (enzyme); EnZ-Lactose (enzyme lactose complex); EnZ-Gal+ (enzyme and galactose covalently bonded); EnZ-Gal-O-Sug (enzyme and galactose as well as acceptor molecule covalently bonded) Sug-OH (sugar acceptor molecule presenting –OH group). .... 28
- Figure 3-1 Visual representation of two of the five library sampling sites i.e. the MS library (A1 and A2) and the Gansbay intertidal kelp degradation library (B). Several sub-samples were collected (red arrows) from each sampling site, and were eventually amalgamated into one, from which DNA were extracted. This was done so as to achieve broader genetic representation. The pH was measured for each individual sub-sample but none differed markedly from each other within each sampling zone. .... 50

Figure 3-2 The partial digestion and column fractionation of the soil extracted metagenomic DNA that was used for the construction of the LacZ library. In Fig A lanes (1-5) is representative of the digestion of 5µg of DNA with successive 2x dilutions of 0.2 units Sau3AI per µl in a 15µl reaction volume, in 1x reaction buffer for 30min. An enzyme concentration between lanes 1 and 2 were chosen for up-scaled digestion. In Fig B lanes (1-22) is representative of the Sepharose cl-2B column fractionation of the up-scaled digestion. Lanes 6 and 7 were pooled and used for library construction. .... 51

Figure 3-3 The various putative genes for the β-galactosidase screen, cloned into the pRSET\_A vector system and grown on MDA-lac plates containing X-gal. The two rows are the same clones photographed on different backgrounds so as to accentuate the visual effect. (-K) and (+K) is representative of the Δβ-gal mutant negative and positive controls, containing the empty pRSET\_A vector and pRSET with the complete E. coli β-galactosidase coding sequence, respectively. (1) HW-βgal1 (2) HW-βgal2 (3) HW-βgal3 (4) HW-βgal4 (5) MS-βgal3 (6) SASRI-βgal1 (7) SASRI-βgal2 (8) SASRI-βgal3 (9) GB-βgal1 (10) HW-βgal5 (11) HW-βgluc1 (12) HW-βgluc2 (13) MS-βgal2 (14) MS-βGluc1 (15) MS-βGluc2. Take note that MS-βgal2 was accidentally omitted, hence only 15 and not 16 putative genes are represented (refer to table 3-3)..... 55

Figure 3-4 Levansucrase gene isolated from the SASRI metagenomic library clearly showing a visible phenotype on a matt of plated out E. coli transformants. .... 55

Figure 3-5 The various β-galactosidases and β-glucosidases that were identified through screening of metagenomic libraries on minimal media containing lactose and X-gal, all cloned into pRSET and expressed in the Δβ-gal mutant. Proteins are grouped according to family, namely (A) GH2 (B) GH42 (C) GH1 respectively. The individual clones are (-K) empty control (+K) positive control E. coli β-gal (A1) HW-βgal1 (A2) HW-βgal2 (A3) HW-βgal3 (A4) SASRI-βgal1 (A5) GB-βgal1 (6) MS-βgal1 (B1) HW-βgal4 (B2) MS-βgal3 (B3) SASRI-βgal2 (C1) HW-βGluc1 (C2) MS-βGluc1 (C3) MS-βGluc2. For the predicted molecular weights of the individual proteins refer to table 3-3 on page 54. ... 56

Figure 3-6 The substrate specificity (A), pH optimum (B) and optimum temperature (C) of the various putative genes (2-13) with E.coli β-galactosidase (1) as a positive control. The Δβ-gal mutant with empty pRSET vector used as negative control but no significant activity could be detected (results not shown). The spectrophotometric hydrolysis was measured at 405 nm and free nitrophenol deduced from a standard linear curve that was established with 4-nitrophenol, ranging between 0.0025 µM and 0.04 µM. All activity was expressed as a percentage, relative to the highest amount, which was taken to be 100% Temperature and pH was characterised with either 2-β-D-gal-NP or 4-β-D-gal-NP depending on which showed the highest activity for the respective constructs. The samples assayed were (2) HW-βgal1; (3) HW-βgal2; (4) HW-βgal3; (5) SASRI-βgal1; (6) GB-βgal1;

(7) MS- $\beta$ gal1; (8) HW- $\beta$ gal4; (9) MS- $\beta$ gal3; (10) SASRI- $\beta$ gal2; (11) HW- $\beta$ Gluc1; (12) MS- $\beta$ Gluc1; (13) MS- $\beta$ Gluc2..... 60

Figure 3-7 The spectrophotometric measurement of free glucose, fructose and galactose is coupled to the reduction of NAD<sup>+</sup> to NADH which is measurable at 340nm wavelength. For this to be observed glucose first needs to be phosphorylated and then reduced to 6-phosphogluconate through the actions of Hexokinase/Glucose-6-phosphate Dehydrogenase (HK-G6DH). For the measurement of fructose, it first needs to be phosphorylated to fructose-6-phosphate and then isomerised to glucose-6-phosphate through the combined action of Hexokinase/Phosphoglucose-Isomerase (PGI), which in turn can then be reduced by HK-G6DH. For the measurement of galactose it needs to be reduced to galactono-1,4-lactone through the actions of Galactose Dehydrogenase (GDH).  
..... 62

Figure 3-8 The product formation of the respective enzymes visualised by means of thin layer chromatography. 0.5 $\mu$ l of an 10x dilution of a 30% lactose(w/v) solution incubated overnight with 1 U.mL<sup>-1</sup> at the optimum temperature and pH for each of the respective enzymes. The individual lanes are (Glc) glucose (gal) galactose (lac) lactose (GGL) glucose/galactose/lactose (VGOS) Vivinal GOS commercial preparation of GOS (-K) empty control (+K) positive control E. coli  $\beta$ -gal (A1) HW- $\beta$ gal1 (A2) HW- $\beta$ gal2 (A3) HW- $\beta$ gal3 (A4) SASRI- $\beta$ gal1 (A5) GB- $\beta$ gal1 (6) MS- $\beta$ gal2 (B1) HW- $\beta$ gal4 (B2) MS- $\beta$ gal4 (B3) SASRI- $\beta$ gal2 (C1) HW- $\beta$ Gluc1 (C2) MS- $\beta$ Gluc1 (C3) MS- $\beta$ Gluc2. The bottom row below the table is the ratio of free glucose to galactose as determined spectrophotometrically by NAD<sup>+</sup> linked enzyme assay..... 62

Figure 4-1 SDS-PAGE protein profiles of fractions collected during the HIS-tag IMAC purification of the HW- $\beta$ gal1 (A) and MS- $\beta$ gluc1 (B) respective recombinant proteins. Lane (M) is protein MW ladder, lane (1) is  $\Delta\beta$ -gal mutant negative control carrying the empty pRSET A expression vector with no insert, lane (2) is the overnight expression by auto induction of the respective expression vector clones and lane (3) is the purified proteins. .... 74

Figure 4-2 Substrate specificity for the purified HW- $\beta$ gal1 and MS- $\beta$ gluc1 expressed relative (%) to the highest activity which was taken to be 100. No activity was detected for ONP- $\alpha$ -D-galactopyranoside, ONP- $\beta$ -D-maltoside, ONP- $\beta$ -D-xylopyranoside, ONP- $\beta$ -D-gluconide, ONP- $\alpha$ -D-glucopyranoside, ONP- $\beta$ -D-glucopyranoside, ONP- $\beta$ -L-arabinopyranoside, ONP- $\alpha$ -L-rhamnopyranoside, ONP- $\beta$ -D-cellobioside for either HW- $\beta$ gal1 and MS- $\beta$ gluc1. .... 76

Figure 4-3 The effect of temperature and pH on both the hydrolytic and trans-galactosylation reaction of HW- $\beta$ gal1 (A) and MS- $\beta$ gluc1 (B). The effect is expressed as a percentage relative to that of the

maximal activity. Generally the trans-galactosylation reaction occurs at a slightly higher temperature and pH, than that of the maximal hydrolytic activity..... 77

Figure 4-4. The effects of the various co-factors on the hydrolytic reaction of HW- $\beta$ gal1 (A) and MS- $\beta$ gluc1 (B). The influence of activity is expressed as a percentage relative to the activity in ddH<sub>2</sub>O with no added metals or salts, which was taken as a 100%. The influence of the co-factors on the trans-galactosylation reaction is statistically the same as that of the hydrolytic reaction (data not shown)..... 78

Figure 4-5. The Michaelis Menten kinetic representation of the hydrolytic and trans-galactosylation reactions of HW- $\beta$ gal1 and MS- $\beta$ gluc1 in  $\mu\text{mole}\cdot\text{mg}^{-1}\cdot\text{min}^{-1}$  against the substrate concentration of lactose (M) ..... 80

Figure 4-6 Thin layer chromatographic analysis of the overnight incubation of purified HW- $\beta$ gal1 with a 30% lactose solution in 20 mM Tris-HCl pH 8.3 at 37°C. lane (1) is glucose, (2) is galactose, (3) is lactose, (4) is all three standards combined, and lane (5) is the overnight incubation product with GOS visible..... 80

Figure 4-7 The relative % of GOS populations according to degrees of polymerisation as determined by the quantitative analysis of the overnight product by LC-MS. The majority of the GOS were tri-saccharides (60%) for both HW- $\beta$ gal1 and MS- $\beta$ gluc1, whilst approximately 30% and 20% were disaccharides respectively for HW- $\beta$ gal1 and MS- $\beta$ gluc1. Fewer than 10% were higher DP GOS for HW- $\beta$ gal1 and approximately 20% for MS- $\beta$ gluc1..... 81

Figure 4-8 The HILIC LC-MS/MS Chromatogram of the final overnight equilibration reaction of the purified protein with lactose. Here the MS- $\beta$ gluc1 reaction is shown as example. The statistical representation for both reactions is shown in Fig. 4-7. (A) is representative of the total chromatogram of the overnight incubation whilst (B) is the mass extracted chromatogram for the di-saccharides and (C) is mass extracted chromatograms for the tri-saccharides. .... 82

Figure 5-1 Phylogenetic tree of microbial levansucrases. Analysis was performed by the neighbour-joining method using the Jukes-Cantor protein distance measure. The number at the nodes indicates confidence bootstrap percentages of 500 repeats. Sequences are represented by the names of the organisms they derive from except for SAS-Ls. All sequences were obtained from the UniprotKB/Swiss-Prot. database and were validated. The accession numbers are as follow *Bacillus subtilis* (P05655), *Gluconacetobacter diazotrophicus* (Q43998), *Pseudomonas syringae* (Q88BN6), *Rahnella aquatilis* (O54435), *Leuconostoc mesenteroides* (Q5IS34), *Lactobacillus reuteri* (Q8GGV4), *Lactobacillus sanfranciscensis* (Q70XJ9), *Zymomonas mobilis* (Q60114)..... 92

- Figure 5-2 SDS-PAGE protein profiles of fractions collected during the HIS-tag IMAC purification of the SAS-Ls protein. Lane (M) is protein MW ladder, lane (1) is BI21 pLysS with the empty pRSET\_A vector, lane (2) is the overnight expression by auto induction of the Hiss-SAS-Ls in BI21-pLysS, and lane (3) is the purified proteins..... 92
- Figure 5-3 The effect of temperature and pH on both the hydrolytic and trans-fructosylation reaction of SAS-Ls. The effect is expressed as a percentage relative to that of the maximal activity which was taken to be 100. .... 94
- Figure 5-4 The effects of the various co-factors on the hydrolytic and trans-fructosylation reaction of SAS-Ls. The influence of activity is expressed as a percentage relative to the activity in ddH<sub>2</sub>O with no added metals or salts, which was taken as a 100%. The influence of the co-factors on the trans-fructosylation reaction is statistically the same as that of the hydrolytic reaction (data not shown). ..... 95
- Figure 5-5 The Michaelis Menten kinetic representation of the hydrolytic and trans-fructosylation reactions of SAS-Ls in  $\mu\text{mole.mg}^{-1}.\text{min}^{-1}$  against the substrate concentration of sucrose (mM), together with the Lineweaver-Burk double reciprocal plot. .... 96
- Figure 5-6 The <sup>13</sup>C and <sup>1</sup>H spectra of the levan produced by SAS-Ls compared to that of the commercially available *Zymomonas mobilis* levan. As can be seen both spectra are nearly identical for both <sup>13</sup>C and <sup>1</sup>H..... 97
- Figure 6-1 The five primary steps involved in the general metagenomic approach is presented in the grey central boxes. The blue peripheral call-out boxes represent key issues and decisions relevant to each of the central steps. Figure adapted from Kowalchuk et al. (2007) [374]. ..... 100

**LIST OF TABLES**

Table 2-1 Microbial glycoside hydrolase enzymes most frequently used for the synthesis of food bioactive oligosaccharides (adapted from Diez-Municio et al 2014 [101] .....	20
Table 2-2 Bacterial sources of glycoside hydrolases used in the production of GOS (adapted from Torres et al.,2010 [117]) .....	31
Table 3-1 Primers used for the LacZ lambda red recombination. The underlined blue text is the annealing sites for the amplification of the chloramphenicol resistance gene on the pKD3 donor plasmid, whilst the preceding sequence is the $\beta$ -galactosidase recombineering recognition sequence. The lacZ-frw and lacZ-rev are internal $\beta$ -galactosidase internal annealing sites to pcr test for the presence of $\beta$ -galactosidase.....	42
Table 3-2 Lists the libraries constructed and carbon source screened on, average insert size determined as per restriction digestion a well as the number of putative genes discovered through functional screening and subsequent sequencing.....	52
Table 3-3 The 17 putative genes that were identified and amplified from four metagenomic libraries. The table shows which glycosyl hydrolase family each putative gene product belongs to, based on in silico analysis as well as the the nucleotide length of the gene and the predicted molecular weight of the protein. It also shows the closest related sequence based on BlastP analysis. ....	54
Table 4-1 Summary of some of the more well known microbial $\beta$ -galactosidases and their kinetic parameters, compared to that of HW- $\beta$ gal1. ....	79
Table 4-2 Summary of some of the GOS forming enzymatic sources, reaction conditions, initial lactose concentration and GOS formed expressed as a percentage (%) of the initial lactose substrate concentration.....	81
Table 5-1 Conserved motifs among bacterial fructosyl transferases. Table adapted from Meng and Fütter [328] .....	91
Table 5-2 Summary of some of the more well known microbial levansucrases and their kinetic parameters, compared to that of SAS-Ls. ....	96
Table 5-3 Chemical $^{13}\text{NMR}$ shifts of microbial levans from various sources compared to the levan of SAS-Ls .....	97

## LIST OF ABBREVIATIONS

$\Delta\beta$ -Gal	beta-galactosidase null mutant
1-FFT	fructan: fructan 1-fructosyl transferase
1-SST	sucrose: sucrose 1-fructosyl transferase
6-FFT	sucrose: fructan 6-fructosyl transferase
6G-FFT	fructan: fructan 6G-fructosyl transferase
6-SST	sucrose: sucrose 6-fructosyl transferase
AD	<i>Anno Domini</i>
ADP	adenosine diphosphate
AMP	adenosine monophosphate
APCs	antigen presenting cells
Approx.	approximately
ATP	adenosine triphosphate
BDMs	carbohydrate binding domains
bp	base pairs
CAZy	Carbohydrate Active Enzyme database
CD	Crohn's disease
CNS	central nervous system
CTAB	N,N,N,N-cetyltrimethylammoniumbromide
DHA	dihydroxyacetone
DHAP	dihydroxyacetone phosphate
DHAP	dihydroxyacetone phosphate
DMF	dimethyl formamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DP	degree of polymerisation
EcN	<i>Escherichia coli</i> Nissle 1917
EDTA	ethylenediaminetetraacetic acid
ENS	enteric nervous system
FDA	Food and Drug Administration
FDPase	fructose-1,6-bisphosphatase
FMT	faecal microbial transplant
FOS	fructo-oligosaccharides
Fruc	fructose
Fruc-1,6-bisP	fructose-1,6-bisphosphate
Fruc-1-P	fructose-1-phosphate
Fruc-6-P	fructose-6-phosphate
Gal	galactose
GalA	galacturonic acid
GAL-DH/M	galactose dehydrogenase/mutatarose
GAP	D-glyceraldehyde 3-phosphate
GF	germ free
GHCl	guanidinium hydrochloride
GHs	glycoside hydrolases
GI	gastro intestinal tract
Glc	glucose
Glc-6-P	glucose-6-phosphate
Glc-6-Pase	glucose-6-phosphatase
GOS	galacto-oligosaccharides
GRAS	generally regarded as safe

GTs	glycosidases and/or glycosyltransferases
HFCS	high-fructose corn syrup
HK/G6PDH	hexokinase/ glucose-6-phosphate-dehydrogenase
HMO	human milk oligosaccharides
HMP	Human Microbiome Project
IBD	inflammatory bowel diseases
IgA	Immunoglobulin A
IMAC	immobilised metal affinity chromatography
IND	Investigational New Drug
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
IUBMB	International Union of Biochemistry and Molecular Biology
$K_m$	Michaelis Menten constant
KOAc	potassium acetate
LB	lysogeny broth
Malt	maltose
MetaHIT	Metagenomics of the Human Intestinal Tract
MOI	multiplicity of infection
MWCO	molecular weight cut off
NaCl	sodium chloride
NAFLD	non-alcoholic fatty liver disease
NCBI	National Centre for Biotechnology Information
NKCs	natural killer cells
ORF	open reading frames
PCR	polymerase chain reaction
PDI	polydispersity index
PEP	phosphoenolpyruvate
PFK-1	phosphofructokinase
PFK-1	phosphofructokinase
PGI	phosphoglucose isomerase
Pi	phosphate
PVPP	polyvinylpyrrolidone
RE	restriction enzyme
Rha	rhamnose
ROS	reactive oxygen species
RT	room temperature (approx. 23°C)
SCFAs	short chain fatty acids
SDS	sodium dodecyl sulphate
ssDNA	single strand deoxyribonucleic acid
STE	sodium tris EDTA buffer
TE	tris(hydroxymethyl)aminomethane ethylenediaminetetraacetic acid
TMAO	trimethylamine N-oxide
Tris	tris(hydroxymethyl)aminomethane
u	units
UC	ulcerative colitis
USA	United States of America
WHO	World Health Organisation
$xg$	centrifugal force
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside
XOS	xylo-oligosaccharides
Xyl	xylose
$\beta$ -Gal	beta-galactosidase
$\beta$ -Glc	beta-glucosidase

## CHAPTER 1: General Introduction

### 1.1 Background

The human microbiota refers to the collection of micro-organisms that reside either on, or in the human body and is composed of bacteria, fungi, archaea and their interacting viral particles. It has become increasingly evident that both the composition of the microbiota inhabiting the human gastrointestinal tract (GI), and its interaction with its surrounding environment, plays a pivotal role in human health [1]. Perturbations of the healthy biota due to various factors including, but not limited to, host genetics, antibiotics, diet and infection, can result in reduced microbial diversity and/or the loss of beneficial microbes as well as pathobiont expansion. This is referred to as dysbiosis [2].

Over the last decade numerous scientific studies have linked dysbiosis of the gut microbiota to the pathogenesis of various intestinal and extra-intestinal diseases and disorders. These include inflammatory bowel diseases (IBD) like Crohn's disease, ulcerative colitis, eczema, allergies, colorectal cancer, chronic fatigue syndrome, obesity and a range of autoimmune disorders including celiac disease, multiple sclerosis and rheumatoid arthritis [3]. Even cardiovascular health and autistic spectrum disorders have been linked with dysbiosis [4, 5], as have disorders of cognitive state like depression, anxiety and Alzheimer's disease [6].

The intestinal microbiota varies between healthy and diseased individuals and in excess of 25 diseases and syndromes have been scientifically associated with an altered intestinal microbiome [7]. The major anaerobic bacterial genera associated with the GI tract are *Bacteroides*, *Eubacteria*, *Fusobacterium*, *Bifidobacterium*, *Peptostreptococci*, *Clostridium*, *Lactobacillus* and *Streptococcus*, with a higher representation of *Bifidobacterium* and *Lactobacillus* species being associated with a healthy and beneficial GI microbiota [8]. By targeting the intestinal microbiota novel avenues for the prevention and/or treatment of the above mentioned diseases could potentially become available. There are two broad means by which the composition of the intestinal microbiota can be altered, namely probiotics and prebiotics.

Probiotics are defined as "live microorganisms that when administered in adequate amounts confer a health benefit to the host" [9]. With probiotics, contrary to popular belief, the effects are mostly transient as the administered microbes only rarely establish self-propagating cultures if a favourable environment is also created. None the less, the advantages and health benefits of probiotics are numerous and have been extensively reviewed [10-15].

An alternative and/or additive approach to microflora management *via* diet is by means of prebiotics. These are defined as “selectively fermented ingredients that are non-digestible by the host, that result in specific changes in the composition and/or activity of the gastrointestinal microbiota, thus conferring benefits upon host health” [16]. Any dietary material that enters the GI tract that is non-digestible by the host is a potential prebiotic and this includes complex carbohydrate polymers and oligosaccharides. The difference between these are that oligosaccharides are defined as carbohydrates having a degree of polymerisation (DP) of between three and nine saccharide units, while polymers have a DP of ten or more. However, current prebiotic research mainly focuses on non-digestible oligosaccharides as these have been shown to have a fermentative bias towards *Lactobacillus* and *Bifidobacterium*, the two main genera associated with a healthy microbiota [11, 14, 17].

The health benefits associated with prebiotics are numerous and wide ranging. They include anti-inflammatory, anti-carcinogenic and immunomodulatory effects, as well as the lowering of cardiovascular risk factors and positive cognitive and mental health modulation [17-19]. They have also been associated with enhanced bio-availability and mineral uptake and the promotion of satiety and weight loss and the prevention of obesity [17, 20].

Fructo-oligosaccharides (FOS), galacto-oligosaccharides (GOS), lactulose, malto-oligosaccharides and inulin together with its hydrolytic derivatives, are the major prebiotics. Such prebiotic oligosaccharides can be obtained by three different means namely (i) extraction from plants, (ii) microbial/enzymatic synthesis, and (iii) enzymatic hydrolysis of polymers [21]. Plant sources of prebiotics include onions, chicory, garlic, bananas, asparagus, artichokes, leeks and tomatoes. FOS and GOS are currently the leading prebiotics available. GOS is synthesised from the simple sugar lactose by means of a  $\beta$ -galactosidase that also possess trans-galactosylation activity, whilst FOS is synthesised from sucrose by fructofuranosidase/invertase that also possess trans-fructosylation activity [22, 23]. The respective enzymes are utilized in either whole cell, free or immobilized formats for the production of prebiotic oligosaccharides [24].

Functional foods are defined as foods that, by virtue of the presence of physiologically active components, provide a health benefit beyond basic nutrition [25]. They were previously based primarily on the fortification of foods with vitamins and minerals, however more recently they have shifted towards the addition of probiotics and prebiotics due to their synergistic effect on human health. Conventional prebiotics have Generally Regarded As Safe (GRAS) status and novel ones need only be generally assessed for risk only if they differ substantially [26]. The global prebiotic market is developing in a multi-billion dollar industry, driven largely by infant formula due to prebiotic oligosaccharides becoming one of its major components in an attempt to imitate human milk oligosaccharides (HMO) [27]. The search for novel prebiotics and the enzymes that make them, greatly add to this field.

## 1.2 Brief review of objectives and methodological approach

This study set out to discover and characterise enzymes that produce novel polymers and/or oligosaccharides comprised mainly of either galactose or fructose. The enzymes that facilitate this belong to the glycoside hydrolyse type family proteins and possess both hydrolytic as well as trans-glycosidic activity. They act on simple sugars like lactose and sucrose by transferring a glycosyl moiety to an acceptable donor after the initial hydrolysis.

The search for these enzymes were mediated through the construction and use of metagenomic libraries from niche environments that have an abundance of either galactose or fructose present. In the case of galactose the first environment was one where milk and whey are disposed and which consequently contains ample amounts of the galactose containing disaccharide lactose. Secondly, a site of kelp degradation was sampled which contains structural and functional macro algae polysaccharides consisting primarily of galactose. With regards to galactose containing saccharides and polymers, the screening of these metagenomic libraries were facilitated by the use of a novel mutagenised *Escherichia coli* strain that has no endogenous  $\beta$ -galactosidase protein. Metagenomic libraries were screened using the  $\Delta\beta$ -Gal mutant on minimal media plates, containing lactose as the sole carbon source. Only cells with metagenomic inserts that contain open reading frames (ORF) encoding proteins capable of hydrolysing the  $\beta$ -1,4 glycosidic linkage between the glucose and galactose moieties, should be able to grow.

For fructose containing environments a metagenomic library constructed from process runoff at a sugar refinery was screened for novel enzymes producing fructose containing polymers and/or oligosaccharides. Screening for these enzymes were done solely based on the visual phenotype of a laboratory *E. coli* strain grown on sucrose.

## 1.3 Dissertation layout

**CHAPTER ONE** is a brief introduction which aims to provide an overview of the relevant research field as well as the aims and objectives of this research project. It furthermore provides a chapter outlay of this dissertation.

**CHAPTER TWO** is a literature review that aims to give a comprehensive exploration of the relevant literature with regards to prebiotic oligosaccharides and polymers, their properties and uses, as well as the enzymes that synthesise them.

**CHAPTER THREE** is a results chapter that describes the initial library construction, screening and clone selection, as well as the preliminary analysis of the positive clones for their potential oligosaccharide/polymer production capabilities. Of the sixteen positive clones and their resultant translational protein products and concomitant catalytic products, that were rudimentary characterised in this chapter, three were selected for further detailed characterisation. The next two chapters will therefore focus on the three before mentioned clones and be presented for purposes of publication.

**CHAPTER FOUR:** is titled 'The biochemical characterisation of two novel glycoside hydrolases, a  $\beta$ -galactosidase and  $\beta$ -glucosidase, belonging to the Glycoside Hydrolase families 1 & 2 respectively, both with high trans-galactosylation activity.' This chapter focuses on two of the novel enzymes discovered through the metagenomic library approach described in chapter three, and entails a more in depth characterisation of the purified enzymes and their products. They both have high trans-galactosylation activity when incubated with the substrate lactose, and produce GOS. Both enzymes have potential industrial relevance for the production of prebiotic oligosaccharides. Of the two, the  $\beta$ -glucosidase is most promising and the article rendered will possibly submitted to the journal 'Applied Microbiology and Biotechnology' [ISSN: 01075-7598 (print); 1432-0614 (web)] for publication.

**CHAPTER FIVE:** is titled 'The characterisation of a novel levansucrase with high trans-fructosylation activity.' This chapter focussed on the only clone isolated from the metagenomic libraries that exhibits significant trans-fructosylation activity. In this chapter the levansucrase enzyme is further purified and characterised in vitro, and so too its product. The data is significant resultant article will again be submitted to 'Applied Microbiology and Biotechnology' [ISSN: 01075-7598 (print); 1432-0614 (web)], or alternatively to the journal 'Bioscience, Biotechnology, and Biochemistry' [ISSN: 0916-8451 (print); 1347-6947 (web)] for publication.

**CHAPTER SIX** is the final and concluding chapter. This chapter provides a critical discussion of the research conducted and results achieved. It provides an in depth review of the metagenomic approach employed together with its advantages and challenges, compared to other viable approaches. It places the research in the broader scientific context of what's already 'out there' and discusses future research and potential alternative application.

**BIBLIOGRAPHY** at the end of this dissertation the bibliography is presented for all the literature cited. The bibliographic style is based on the 'BMC Biochemistry' journal, published by Biomed Central. For author guidelines refer to <http://www.biomedcentral.com>

## CHAPTER 2: Literature Review

### 2.1 The human microbiota and microbiome

The microbiome comprises all the genetic material within a specific microbiota and can also be referred to as the metagenome of the microbiota. Technically speaking, the human microbiome is a reference to all the genes of the microorganisms and viruses that reside either on or in our bodies, however it is colloquially a reference to the metagenome of the microbiota, and its associated viral particles, of the lower GI tract. The distal GI tract is the most densely populated natural occurring microbial community known to man.

Only since the advent of affordable next generation sequencing platforms, has the scientific community become fully aware that the commensal microorganisms residing in our bodies are intractably linked to our well-being and physiological functioning. This topic has spilled over into the mainstream media in seeing that it is relevant and current (see Fig 2-1). Recent revelations of the role of the microbiome in our health have begun to shake the foundations of medicine and nutrition. Some would argue that a paradigm shift is in occurrence. Many scientist now view humans as collective supra-organisms and that our biology is determined by both the genes of our genome as well as those of the microbiome meaning that, from this perspective, human health is a form of ecology [28]. A compelling argument can be made for this as the human body contains tenfold ( $10^{14}$ ) more microbial cells than human cells (approx.  $30 \times 10^{12}$ ) and the microbiome exceeds the number of human genes by at least two orders of magnitude [1, 29]. The importance of the GI microbiota to our health is further supported by the presence of human milk oligosaccharides (HMO) in human milk. HMO, which are short chain oligosaccharides of varying lengths and composition, can make up to 20% of the total nutritive value of human milk [30]. They are however non-digestible by human infants and serve as prebiotics to mostly bifidogenic bacteria. Therefore, in the earliest and most vulnerable stage of infant life, one fifth of the total nutritive component of human milk is dedicated towards the growth and maintenance of the human microbiota [31].

The research interest in the human microbiome and its effects on humans have led to the establishment of international consortiums dedicated towards its investigation. Large projects such as the *Human Microbiome Project* (HMP) and *Metagenomics of the Human Intestinal Tract* (MetaHIT) were specifically established to investigate the microbiota and its microbiome as well as its influence on human health [32-34].



Figure 2-1 The respective covers of Nature (Vol 464 Number 7285), Scientific American (Vol 306 Issue 6), Science (Vol 336 Issue 6086) and The Economist (Aug 18<sup>th</sup> 2012), indicating the relevance of the human microbiota and microbiome.

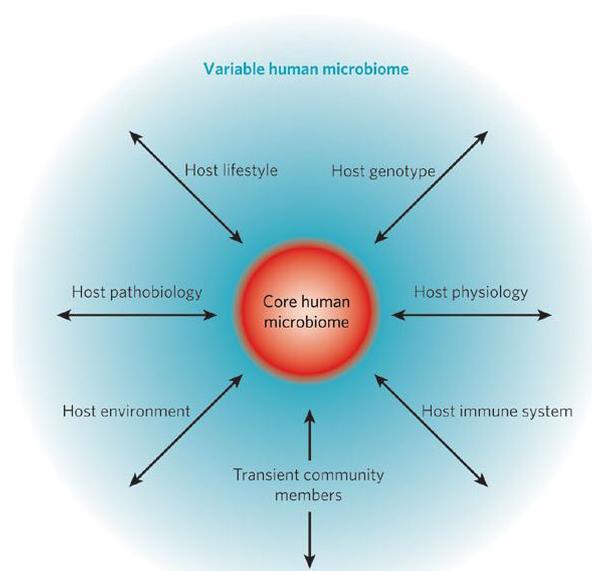
## 2.2 The core Microbiota / Microbiome

The complex communities of microbiota that inhabit the lower GI tract have been demonstrated to have a direct impact on human health, which can be either beneficial or adverse. The bacterial biomass of the entire GI tract ranges from  $10^2$  to  $10^{11}$  cells/ml [35]. Recently, studies *via* culture independent sequencing techniques have revealed that, while the exact GI microbiota composition is highly individual specific, fewer than ten bacterial phyla dominate the make-up of the GI tract in humans [32, 36]. Of these, Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria and Verrucomicrobia, constitutes the majority of the microbiota by far, with the first two being most prominent constituting almost 90% of the total microbial gene catalogue [32, 37].

The human intestinal core microbiota is defined as the number and the identity of bacteria that are shared among different individuals [37]. Schippa *et al.* (2014) hypothesised that deviations from, and altered abundance in the core microbiota is associated with the development of complex diseases like atherosclerosis, type 2 diabetes, obesity, and colorectal cancer, yet defining the exact makeup of the core has proven to be difficult [38].

On a species level Peer Bork *et al.* (2011) hypothesised the existence of three cluster types (enterotypes) based on comparative metagenome sequence analysis of different individuals, from different countries on different continents [39]. They found that enterotypes are not dictated by age, gender, body weight, or national divisions, but there are indications that long-term diet has a large influence. Type 1 is characterized by high levels of the *Bacteroides* species (diet high in animal protein and saturated fat), type 2 has few *Bacteroides* but *Prevotella* species are common (diet high in carbohydrates), and type 3 has high levels of *Ruminococcus* species (diet rich in alcohol and polyunsaturated fats) [40]. However, the year after the publication of that article new research brought into question the clear-cut boundaries of the 'enterotypes hypothesis' [41].

Perhaps, if a functional core cannot be adequately defined by the presence of certain species, it would be better served on a genetic level. Qin *et al.* (2010) as part of the Human Metagenome Project, catalogued a common ‘functional core’ of genes, conserved in each subject and among different bacterial species [32]. This reflects the genes that relate to the survival of the gut microbiota and also those major microbial functions that relate to human health [32]. For example, some of the functions included in the ‘core of genes’ are linked to the degradation of complex polysaccharides, synthesis of short-chain fatty acids (SCFAs), amino acids and vitamins. The core microbiome is present in all humans, but a smaller subset of genes known as the ‘variable microbiome’ is only associated with certain subsets of the human population. The variable microbiome depends mostly on host-specific factors, such as genotype, physiological status, host pathologies, lifestyle, diet, and also environmental factors [32, 42]. Together, the core and variable components of the human microbiome influence a multitude of different aspects of human health, all of which is of major scientific interest and at the forefront of current research (see Fig 2-2).



**Figure 2-2** The core human microbiome (red) is the set of genes present in the majority of humans. The variable human microbiome (blue) is the set of genes present in smaller subset of humans. This variation could result from a combination of factors such as host genotype, host physiological status, host pathobiology (disease status), host lifestyle, host environment and the presence of transient populations of microorganisms that cannot persistently colonize a habitat.

### 2.3 Contribution of the core towards health

The gut microbiome contributes towards human health through various mechanisms that can be broadly categorised under (i) metabolic, (ii) mucosal epithelium defence barrier and, (iii) immune modulation. In the following sections each of the sub-headings will be discussed briefly.

### 2.3.1 Metabolic Functions

The gut microbiota can be seen as a 'metabolic organ' that interacts with the human host and impacts on a myriad of essential functions and helps maintain human health [43]. Humans lack the necessary enzymes to degrade the bulk of dietary fibres, which pass through the upper digestive tract to the lower one where these complex carbohydrates are broken down and fermented to mainly SCFAs and gasses (carbon dioxide and methane) [44]. SCFAs are saturated aliphatic organic acids that consist of one to six carbons of which acetate (C2), propionate (C3) and butyrate (C4) are the most abundant ( $\geq 95\%$ ) in the gut [45]. The ratio of acetate: propionate: butyrate is normally 3:1:1, however this may vary depending on the initial dietary carbon source of fermentation, the dominant microbiota as well as host specific factors [44]. To the microbial community SCFAs are a necessary waste product required to balance redox equivalent production in the anaerobic environment of the gut, to humans however they are much more [45]. Approximately 95% of SCFAs are absorbed by the human body and only 5% is secreted in the faeces [45]. SCFAs are absorbed in the colon and have been shown to play a role in regulation of the immune system, colonic gene expression, gut motility, anti-carcinogenesis, obesity, insulin resistance and metabolic regulation [11, 43].

On the most rudimentary level SCFAs serve as an energy source to the human host. In diets that are high in legumes, fruits and vegetables the dietary fibre content may reach approximately 60 grams per day. Fermentation of this fibre will yield roughly 400–600 mmol SCFAs per day, which accounts for about 10% of the adult human caloric requirements [45]. Acetate, butyrate and propionate are absorbed through the intestinal lumen. Colonocytes are epithelial cells that form the primary barrier between the lower GI lumen and human body and it has been estimated that between 60-70% of the total energy requirements of colonocytes are derived from SCFA oxidation, specifically butyrate [45]. After acetate is absorbed it can be not only metabolised in the liver for energy production, but also serves as substrate for the synthesis of cholesterol, long-chain fatty acids, glutamine and glutamate. Propionate acts as a precursor for gluconeogenesis in the liver [46].

SCFAs also play a role in regulating the balance between fatty acid synthesis, fatty acid oxidation, and lipolysis. Fatty acid oxidation is activated by SCFAs, while its synthesis as well as lipolysis are inhibited by them, resulting in a reduction of free fatty acid plasma concentrations and, ultimately a decrease in body weight [45]. Apart from fatty acid metabolism, SCFAs also lead to a decrease of plasma glucose levels *via* multiple mechanisms [47]. Furthermore, SCFAs have also been shown to reduce plasma concentrations of cholesterol in rodents and humans [48, 49].

Other than its effects on primary metabolism, the GI microbiota produce vitamins, synthesize amino acids, influence ion absorption and transform potentially carcinogenic compounds to inactive forms [28].

### 2.3.2 Mucosal epithelium effect and defence barrier

Enterocytes are the epithelium barrier cells of the small intestine and also act as nutritive adsorptive cells. Together with the colonocytes of the colon they make up the barrier of the GI tract. The GI microbiota contribute towards maintenance of intestinal epithelial barrier integrity not only through SCFAs but also through direct signalling, active maintenance of cell-to-cell junctions, promotion of epithelial repair, as well as playing a role in the regulation of enterocytes and colonocytes turnover [50]. A healthy microbiota furthermore creates a “colonisation resistance” which can be described as the ability of the microbiota to prevent pathogenic colonisation by competing for attachment sites and nutrients and through the production and secretion of antimicrobials [51].

### 2.3.3 Host immune modulation

The microbiota plays a critical role in the establishment, adaptation, maintenance and proper function of the human immune system through a multitude of complex and dynamic host/microbiota mechanisms (for review see [52]). It is this unique set of immune-regulatory mechanisms that prevent the unnecessary activation of the immune system against non-injurious antigens, including the commensal microbiota themselves. Pivotal to this immune homeostasis is the interface and interplay between epithelial/mucosal barrier and microbiota [3]. As mentioned previously, the microbiota contribute towards the innate immune system directly through modulating this epithelial mucosal barrier. However, intestinal antigen presenting cells (APCs), macrophages, neutrophils and natural killer cells (NKCs) are all modulated by the intestinal microbiota [11]. Autoimmune and inflammatory diseases are associated with dysregulated immune responses and are often linked to disruption of this barrier [53]. Keeping a delicate balance in the immune system by eliminating invading pathogens, while still maintaining self-tolerance to avoid autoimmunity, is critical for health. The GI microbiota also influence the adaptive immune system in that they play an important role in the development of CD4<sup>+</sup> T cells, both within and outside the intestine [53]. Some of the best evidence of the pivotal role the microbiota plays in the immune system comes from studies utilising germ-free (GF) and gnotobiotic mice. GF animals are sterile with regards to intestinal microbiota whilst gnotobiotic ones have been seeded with

a singular bacterial species allowing investigation of species related traits [53]. It has also been suggested that the microbiota plays a pivotal role in the production of antibodies in the intestinal mucosal membranes. It is estimated that 0.8g of Immunoglobulin A (IgA), which are antibodies associated with mucosal membranes, are secreted per day per meter intestine by Peyer's patch immunoglobulin secreting plasma cells [54]. Significant reductions in IgA and plasma cells were observed in GF-mice, which was restored once the gut was recolonized with microbiota demonstrating that the gut microbiota is the driving force for mucosal IgA production and is therefore pivotal to the intestinal immune response [55].

## 2.4 Dysbiosis

There is mounting evidence for the association between various intestinal and extra-intestinal disorders, and the dysbiosis of the gut microbiota. Intestinal disorders are usually associated with over-reactive inflammatory responses characteristic of diseases like irritable bowel syndrome (IBS), coeliac disease and Crohn's disease. Extra-intestinal disorders include allergy, asthma, obesity, metabolic syndrome and cardiovascular disease. Even disorders of the cognitive state have been associated with the conversion of certain dietary compounds (by a subset of microbiota) to metabolites that are absorbed and which then affect the phenotype of the host and influence the risk of disease [56].

As mentioned above, even though the association between dysbiosis and health have been proven, it is none the less difficult to establish precise relations with the occurrence and relative abundance of specific microbial species. In future though, specific changes in compositional or functional diversity may serve as biomarkers for health or specific diseases. Below is a brief discussion of dysbiosis in disease under the headings of (i) Effect of microbial metabolites on health, (ii) Dysbiosis and GI tract disorders, (iii) Dysbiosis and systemic diseases, and (iv) Dysbiosis and central nervous system disorders.

### 2.4.1 Effect of microbial metabolites on health

The complimentary effect of SCFAs to host health was discussed above, however metabolic fermentation products of other microbiota have been associated with the development of diseased states in humans. In recent years, the gut-liver axis and the impact of intestinal microbiota on liver function has come to the forefront. The liver receives approximately 70% of its blood supply from the intestinal portal vein and is, therefore, the organ that is exposed first to the absorbed metabolites and

antigens produced by the gut microbiota. Non-alcoholic fatty liver disease (NAFLD) is a prevalent and severe disease that can lead to cirrhosis, liver carcinoma and ultimately death. It has a prevalence of 16-30% among the general populace and between 50-90% in obese individuals [57]. Several mechanisms relate NAFLD, and its concomitant deposition of hepatic fat, to dysbiosis and the microbiota. These include the dysregulation of gut permeability, bile acid metabolism, immune balance and low-level inflammation. In addition, increases in endogenous ethanol production by bacteria, most notably through the metabolism of choline are also thought to play a role [58]. Mice fed a high-protein/high-fat diet had a higher rate of metabolising choline to methylamines. Due to this conversion by the microbiota, the bioavailability of choline is reduced, resulting in the inability to synthesise phosphatidylcholine giving rise to the accumulation of triglycerides in the liver. This mimics choline-deficient diets which have been consistently associated with hepatic steatosis. The bacterial methylamines are absorbed and transported to the liver where they are oxidised to trimethylamine N-oxide (TMAO). TMAO is a pro-atherogenic compound and used as an indicator for cardiovascular disease [59].

#### 2.4.2 Dysbiosis and GI tract disorders

Crohn's disease (CD) and ulcerative colitis (UC) are the most prevalent forms of inflammatory bowel disease (IBD), characterised by recurrent or chronic inflammation of the mucosal intestinal barrier. IBD is a class of autoimmune disease, in which the body's immune system attacks elements of the digestive tract giving rise to the inflammatory response. CD is generally associated with the upper GI tract whilst UC is associated more with the colon, but sometimes a definitive diagnosis cannot be made due to idiosyncrasies in the presentation. The exact causation of either disease is unknown but there is increasing evidence that intestinal microbial dysbiosis has a role in the development of IBD [60]. Both diseases are characterised by an overall decrease of between 30-50% in microbiota diversity most specifically *Firmicutes*, and an increase in *Bacteroidetes* and other facultative anaerobes such as *Enterobacteriaceae* [61]. People with IBS are at higher risk for developing colorectal cancer and can also present with extra-intestinal manifestations such as liver problems, arthritis, skin manifestations and eye problems [56].

### 2.4.3 Dysbiosis and systemic diseases

Worldwide obesity is on the rise and has reached epidemic levels [62]. According to the World Health Organisation (WHO) globally in 2014, 39% people older than 18 years were overweight and 19% of the total global population were obese. Obesity manifests its comorbidities in metabolic syndrome which includes increased risk for coronary heart disease, diabetes mellitus type 2, high blood pressure, high blood cholesterol, insensitivity to insulin, high triglyceride levels, and various cancers [63]. Obesity is most commonly caused by a combination of excessive food energy intake embodied in the 'western diet', sedentary lifestyle, lack of physical activity and, to a lesser extent, genetic susceptibility [63, 64].

Gut microbiota have been linked to obesity [65] based on two observations. Firstly, the ratio of the two dominant bacterial genera, namely *Bacteroides* and *Firmicutes*, are altered in obese people (fewer *Bacteroides* than *Firmicutes*) compared to lean people (more *Bacteroides* than *Firmicutes*) [65]. Secondly, the observation that obesity phenotype could be triggered by transplanting microbiota from obese animals to germ free mice [66, 67]. The intestinal microbiome affects glucose and lipid metabolism as well as satiety. Germ-free and conventionally colonised mice differ in that conventional mice have higher serum glucose levels, higher triacylglycerol levels in the liver and, also, higher levels of the hormones leptin and insulin which are positively correlated with body weight and with adipose mass. These data support the idea that the microbiota could potentially contribute to weight gain [64].

Two mechanisms have been hypothesised for the link between weight-gain and the GI-microbiota. Firstly, fermentation of indigestible carbohydrates by intestinal microbiota leads to an increased absorption of SCFAs followed by increased hepatic lipogenesis. Secondly, high-fat diets trigger an increased transfer of bacterial lipopolysaccharides (LPS) from the intestinal lumen to the blood causing metabolic endotoxaemia and low-grade inflammation [64] which, in turn, contributes to adipose deposition. The role of SCFAs on metabolism and plasma glucose levels were discussed above. Theoretically SCFA adsorption could contribute to obesity development; however epidemiological studies indicate that a diet rich in fibre correlates with a lower incidence of obesity and symptoms of the metabolic syndrome [17, 68]. This apparent paradox is further supported through studies linking dietary fibre to increased satiety, reduced weight gain, increased metabolic rate, increased insulin sensitivity, decreased serum cholesterol levels, overall beneficial impact on health and the presentation of fewer markers for metabolic syndrome and its associated comorbidities (for review see [17]).

In healthy individuals the intake and expenditure of metabolic energy is balanced, but there is evidence that the microbiota can influence this [64]. Current scientific evidence indicates that high-fat diets contribute towards the development of a microbiome that promotes increased harvest of energy over its expenditure from the diet, thereby promoting obesity [64]. The SCFAs produced by fermentation can

either provide additional energy to the host and can be channelled towards lipogenesis and gluconeogenesis, or, can lead to increased satiety and insulin sensitivity as well as reduced overall energy intake and increased expenditure. Which of these effects predominates, and how they are affected by the type of dietary fibre, is under current investigation.

#### 2.4.4 The gut-brain-axis: Dysbiosis and central nervous system disorders

The enteric nervous system (ENS) consists of approximately 500 million neurons and is embedded in the lining of the GI-tract, forming a mesh-like system of neurons. It governs the function of the gastrointestinal system and is often referred to as the second brain. The link between the CNS and ENS is referred to as the 'gut-brain-axis'. There is bidirectional communication between the CNS and ENS, thereby linking the emotional and cognitive centres of the brain with the GI tract. The mechanisms underlying gut-brain-axis communication involve neural pathways as well as immune and endocrine mechanisms. Recent research describes the importance of gut microbiota in influencing these interactions and intestinal dysbiosis has been associated with diseases that impact on the 'gut-brain' axis and thereby effecting on the central nervous system [69].

Perhaps some of the best evidence of the impact of microbiota on behaviour comes from animal studies. The occurrence of locusts congregating in vast swarms that can devastate crops is dependent on pheromones. Dillon *et al.* (2000) identified that locust gut microbiota were pivotal in the production of these swarming pheromones [70]. Germ free locusts did not produce the pheromones but once the bacterium *Pantoea agglomerans* was introduced into the gut, pheromone production was restored [70]. Even in humans the gut microbiota can act as a source of various biologically active molecules normally associated with neurotransmission and GABA, serotonin, melatonin, histamine, and acetylcholine have all been shown to be produced by bacteria [71]. In a study by Sharon *et al.* (2004) two isolated populations of *Drosophila* fruit flies were reared on two different feeding mediums, one containing molasses and the other starch. When the isolated populations were mixed, molasses fed flies showed a preference to mate with other molasses fed flies, while the starch fed flies showed a preference for other starch fed flies. The differential mating preference occurred after only one generation of rearing on the separate growth media and could be maintained for at least thirty seven generations. Antibiotic treatment abolished the diet-induced mating preference, but it was re-established when the flies were seeded with the original gut microbiota. These data suggest that the fly microbiota was responsible for the preferential diet-induced mating phenomenon [72]. Through the study of germ free and gnotobiotic mice several studies have shown that microbiota influences stress reactivity and anxiety-like behaviour [73]. Furthermore, germ-free mice demonstrated decreased memory function as well as increased

serotonin turnover [74, 75]. In humans there is also evidence of microbiota influencing anxiety and depressive-like behaviours. Research suggests early developmental association with beneficial microbiota is crucial later on in life to reduce the risk of depression and anxiety related disorders [76, 77]. So too has dysbiosis in autistic patients shown to present specific microbiota alterations in accordance to the severity of the disease [78, 79]. The topic is, however, very contentious and research is ongoing.

## 2.5 Manipulating the gut microbiota to maintain health and treat disease

As has been mentioned, the intestinal microbiota composition varies between healthy and diseased individuals. Generally, high microbial diversity is thought to be associated with healthy gut microbiota, while loss of diversity seems to correlate with disease [7]. Although the relationship between the alterations in the gut microbiota and disease is not always clear-cut, targeting the intestinal microbiota might offer new possibilities for preventative health and treatment.

There are three possible mechanisms by which the microbiota of the gut can be modulated, namely the administration of probiotics, prebiotics and lastly microbial transplants. Dietary prebiotics will be discussed in depth in the next section whereas probiotics and microbial transplant will be discussed briefly as sub-headings in this section.

### 2.5.1 Probiotics

Since ancient times lactic acid bacteria has been utilized to preserve milk and improve its digestibility. An early Persian translation of the Bible credits the prophet Abrahams' longevity to the daily consumption of sour milk. Legend also has it that the French king Francois 1<sup>st</sup> cured his chronic diarrhoea with the use of a Turkish 'yogurt'. Even Hippocrates, the father of modern medicine famously stated that 'all disease begin in the gut' [80]. During the First World War Alfred Nissle isolated a non-pathogenic strain of *Escherichia coli* from a battlefield soldier, who was one of a few that did not fall ill during a severe outbreak of Shigellosis. This strain *E. coli* Nissle 1917 (EcN) has since been used to treat a variety of gastrointestinal diseases including salmonellosis and shigellosis, thereby heralding the modern era of probiotic treatment [81]. A multitude of publications have shown that specific probiotic strains (or combinations of them) can be utilized to treat different diseases. For example *Oxalobacter formigenes* is the key bacterium responsible for the degradation of oxalate in humans, the accumulation of which

is the main cause of kidney stone formation [82, 83]. The bacterium is extremely sensitive to antibiotics and individuals who have undergone antibiotic treatment in their lifetime often contain reduced level of *O. formigenes*. Probiotic treatment with *O. formigenes* have been shown to temporarily reduce serum levels of oxalate [84]. Currently the dominant species of probiotics used for human and animal livestock are from the genera *Lactobacillus*, *Streptococcus* and *Bifidobacterium* [85]. By transiently colonizing the GI tract, probiotics serve to correct dysbiosis that contributes to disease. The effects that these bacteria exert can either be direct *via* the probiotic bacterium itself, or indirect *via* its interaction with the commensal microbiota. These include the secretion of antimicrobial peptides and bacteriocins, modulation of lumen pH, influencing quorum sensing in other bacteria which alters their virulence and pathogenesis, enhancing the production of mucins from gut epithelial cells, exerting an anti-inflammatory effect in the gut and promoting the secretion of IgA which binds pathogens [7, 18, 81]. It should be noted however that, in all likelihood, the probiotics exert their effect in the small intestine and not the lower GI tract of the large intestine. In the small intestine there is a lower concentration of microbiota, for the relative small number of dietary acquired probiotika to exert an effect during their transit. In the large intestine the probiotics will be vastly outnumbered by the resident microbiota and, even though they can still have an effect, its efficacy will be greatly diminished.

### 2.5.2 Microbial faecal transplants

A faecal microbiota transplant (FMT) is the process of transplantation of faecal bacteria from a healthy individual into a recipient so as to restore a healthy microbiota [86]. Evidence for the use of faecal bacteriotherapy dates back to the 4<sup>th</sup> century AD when it was reported that nomadic Bedouin used camel dung for the treatment of diarrhoea [7]. More recently FMT came to prominence for the treatment of recurrent *Clostridium difficile* infections [87]. Today, *C. difficile* is a debilitating disease and the leading cause of antibiotic associated diarrhoea and pseudomembranous colitis. Antibiotic treatments fail in up to 30% of cases and it can be a life threatening illness. Mortality rates are as high as 24% in cases involving critically ill patients, especially the elderly, and it has reached epidemic levels in North America and Europe [87]. So effective is FMT for the treatment of resistant *C. difficile* infection that the original double blind placebo controlled study was halted, for ethical reasons, so that the placebo group could also receive the treatment. In the firsts trial 85% were effectively cured and of the remaining 15% a further 90% were cured with a second treatment [88]. Due to the above mentioned success FMT is being investigated for the treatment of other dysbiosis associated diseases most notably inflammatory bowel diseases as well as metabolic and cardiovascular diseases [86]. Furthermore, the federal Food and Drug Administration (FDA) of the United States of America (USA) now recognizes FMT

as an 'Investigational New Drug' (IND) and have published proposed guidelines for its use and called for comments (<http://www.regulations.gov>) [89]

### 2.5.3 Prebiotics

The current definition of a prebiotic states that, a prebiotic is a selectively fermented ingredient that results in specific changes in the composition and/or activity of the GI microbiota, thereby conferring benefits upon host health [16]. Contrary to probiotics which aim to introduce new bacteria, prebiotics stimulate the growth and activity of particular bacteria that are already present in the GI tract and, for this reason, is considered to be one of the more effective modulators of the GI microbiota [90]. The main genera that are stimulated by prebiotics are *Bifidobacterium* and *Lactobacillus*, and the effect that prebiotics exert is often referred to as the bifidogenic effect. Inulin type polysaccharides, fructo-oligosaccharides (FOS), galacto-oligosaccharides (GOS) and lactulose are the dominant prebiotics that have garnered the most scientific attention thus far [17, 91]. Other molecules that have recently been classified as prebiotics are lactitol, mannitol, maltodextrin, raffinose, lactulose, sorbitol, isomalto-oligosaccharides, mannan-oligosaccharides, xylo-oligosaccharides (XOS) and wheat bran derived arabinoxyloligosaccharides, all of which is currently under investigation as putative and promising prebiotics [90]. Different prebiotics can selectively stimulate the range of beneficial bacteria to varying degrees. FOS were less selective than GOS as a fermentative growth substrate, while even fewer bacteria were able to use starch and long-chain inulin for fermentation [92]. It has also emerged through the metagenomic studies of the human microbiome that there is a higher representation of genes encoding GOS degrading enzymes compared with those encoding FOS degrading polypeptides [93]. The positive effect promoted by prebiotics mostly relates to SCFAs (as discussed above) and include, immunomodulation, anti-inflammation, contribution towards cardiovascular health, enhancement of the bioavailability and uptake of minerals and vitamins, promotion of satiety and weight loss, prevention of obesity and prevention of certain cancers [94, 95].

### 2.5.4 Synbiotics

Synbiotics is the combination of prebiotics and probiotics, that then work in a synergistic way, so as to beneficially effect the host by improving the survival of the live microbial probiotic dietary supplement, in the GI tract. This is achieved by the prebiotics selectively stimulating the growth and/or metabolism of one or more of the health promoting probiotic bacteria [96, 97]. Synbiotics were especially developed

to overcome the before mentioned (section 2.5.1) limitations of prebiotic survivability through the upper GI tract as several factors like pH, H<sub>2</sub>O<sub>2</sub>, organic acids, oxygen and moisture stress has been shown to effect the viability of probiotic supplementation [98]. The health benefits claimed by synbiotic treatment includes, (i) increased levels of *Lactobacillus*, *Bifidobacterium* species and an overall balanced gut microbiota (ii) improvement of liver function in cirrhotic patients, (iii) improvement of immunomodulating function, (iv) and prevention of bacterial translocation and reduced incidences of nosocomial infections in surgical patients [99].

## 2.6 Introductory review of Prebiotic synthesis by microbial enzymes and their classification

As mentioned in section 2.5.3, prebiotics enhance the growth of beneficial bacteria in the GI system, thereby establishing all the beneficial properties listed above. The specific prebiotic properties of oligosaccharides and polymers are determined by their monosaccharide composition, glycosidic linkages and degree of polymerisation [100]. Due to the scientifically backed health benefits already reviewed, there is a high demand for functional food prebiotics leading to a constant drive to develop more efficient and affordable methods of production for these types of compounds. There is also a drive to identify novel prebiotics with unique properties [101]. The two substrate sources of disaccharide carbohydrate often envisaged to be used in the production of more complex carbohydrate compounds are whey permeate and cane sugar [102]. Whey permeate is a by-product from cheese manufacturing and is high in lactose. Annually, millions of tons of cheese are manufactured leading to whey water often becoming an industrial pollutant if not adequately disposed of. Therefore, its use makes not only economic sense but is also environmentally beneficial. So too sugar, which is isolated predominantly from sugarcane throughout the tropics and sub-tropics, is a readily available and cheap source. It is therefore not coincidental that the use of the above mentioned substrates for the production of more complex derivatives of lactose and sucrose is currently the focus of research and is an on-going challenge for industry.

The production of prebiotics by means of enzymatic synthesis has great potential as they generally exhibit high yields, substrate specificity as well as regio-and stereo-specificity [21, 103]. Such enzymes can be obtained from a variety of sources including microorganism, plant and animal sources. In general however, microbial enzymes are preferred for industrial application for various reasons including (i) ease of microbial handling and extraction, (ii) higher microbial enzymatic yields, (iii) high microbial enzymatic product synthesis, (iv) ease of use of microbial genes in recombinant technology relative to those of higher complex eukaryotic genes, (v) economic feasibility in terms of fermentation biology, (vi) versatility with regards to acceptor substrate, (vii) better enzymatic stability and finally (viii), regularity

of supply in that it's not dependent on seasonal availability of extraction source [24, 101, 104, 105]. Consequently many bacterial, yeast and fungal enzymes have been utilized in the production of dietary oligosaccharides and polymers and the dominant ones are summarised in table 2-1.

With regards to the classification of enzymes two systems predominate, namely the *International Union of Biochemistry and Molecular Biology* (IUBMB) enzyme nomenclature, and the *Carbohydrate active enzymes database* (CAZy). Commonly and traditionally though, enzymes involved with the synthesis of bioactive polymers and oligosaccharides are colloquially referred to as glycoside hydrolases (GHs) and/or glycosyltransferases (GTs) with the two terms often interchanged depending on which nomenclature system are being employed. These enzymes are capable of transferring a glycosyl moiety from one compound (donor) to another compound (acceptor) thereby elongating the oligosaccharide/polymer [106].

With the IUBMB system it is the reaction, rather than the enzyme itself, that are classified and this is done primarily on the basis of their substrate specificity which is expressed as the Enzyme Commission (EC) number of the enzyme (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>). In accordance to this system GHs/GTs (enzymes which hydrolyse compounds containing *O*- and *S*- glycosidic bonds) are given the serial code 3.2.1.X where X is representative of the substrate specificity but can also occasionally be indicative of either the enzymatic molecular mechanism or the glycosidic linkage type [107]. These GHs/GTs can further be subdivided into groups according to the glycosyl moiety that they actually transfer i.e. hexosyltransferases (EC. 2.4.1), pentosyltransferases (EC 2.4.2) and those transferring other glycosyl moieties (EC 2.4.99) [108]. There are however several limitations to the IUBMB system of enzyme classification, most notably that the numbering system does not allow for enzymes which act on more than one substrate. It also does not relay any inherent structural information or give insight to the actual mechanistic action of the enzyme [109]. The GHs/GTs were then further classified into three mechanistic groups, based on the characteristics of the donor substrate. Firstly (i) there are the Leloir-type glycosyltransferases which require activated sugar nucleotides to synthesise elongation e.g. monosaccharide di-phosphonucleotides. Secondly (ii), there are the non-Leloir type glycosyltransferases which utilise substrates like sugar- phosphates, sugar-pyrophosphates as well as the respective polyprenol phosphates and pyrophosphates. Finally (iii) are the transglycosidases which employ non-activated disaccharide sugars like lactose and sucrose and consequently utilise the energy released from the initial hydrolysis to catalyse the transfer of the sugar moiety to the donor [110].

Contrary to the IUBMB, the CAZy system of nomenclature and classification exploit amino acid sequence similarities rather than the reaction characteristics in order to group enzymes in families. Henrissat *et al.* (1991, 1997) hypothesised that in principle sequence and structure are inherently related and, therefore, structural and mechanistic properties of enzymes can be inferred by means of this sequence

based system [111, 112]. Henrissat's initial system of classification evolved into the CAZy database (<http://www.cazy.org>) which describes the different families with structurally related catalytic and functional domains (Carbohydrate Binding Domains, BDMs) to which the various enzymes involved in the degradation, modification and/or creation of glycosidic bonds belong [113]. Currently CAZy describes over 300 protein families which can be further subdivided into classes of carbohydrate active enzymes namely (i) *Glycoside Hydrolases* (GHs, hydrolyses and/or rearrangement of glycosidic bonds), (ii) *Glycosyltransferases* (GTs, formation of glycosidic bonds), (iii) *Polysaccharide Lyases* (non-hydrolytic cleavage of glycosidic bonds), (iv) *Carbohydrate esterases* (hydrolysis of carbohydrate esters), and (v) *Auxiliary Activities* (redox enzymes that work in conjunction with CAZymes).

It is noteworthy that in accordance with CAZy, but not with the IUBMB nomenclature system, a clear distinction is being made between glycoside hydrolases and glycosyltransferases. The former refers to the original transglycosidases (utilising disaccharides as substrates) with the latter being a combination of the Leloir (nucleotide sugars) and non-Leloir (sugar derivatives) transferases as described in the three-pronged mechanistic and substrate donor based IUBMB denoted system [109]. Even though bi-enzymatic systems utilising glycosyltransferases have been reported, it is glycoside hydrolases which are most often employed in the production of prebiotic oligosaccharides and polymers as they do not require rare activated sugar nucleotides or monosaccharide derivatives for synthesis [114].

Table 2-1 Microbial glycoside hydrolase enzymes most frequently used for the synthesis of food bioactive oligosaccharides (adapted from Diez-Municio et al 2014 [101])

	Structure / linkage <sup>a</sup>	Enzymes	EC number <sup>b</sup>	Enzyme source	GH family <sup>b</sup>	Substrate	References
Fructo-oligosaccharides	(Fruc) <i>n</i> -Glc [(β-2,1),(β-2,6)]	β-Fructofuranosidase Inulosucrase Levansucrase	EC 3.2.1.26 EC 2.4.1.9 EC 2.4.1.10	Fungi ( <i>Aspergillus niger</i> ; <i>A japonicus</i> ; <i>A.oryzae</i> ; <i>Aureobasidium pullulans</i> ; <i>Penicillium citrinum</i> ) Bacteria ( <i>B. macerans</i> ; <i>Z. mobilis</i> ; <i>Lactobacillus reuteri</i> ; <i>Arthrobacter</i> sp.)	32, 68, 100	Sucrose (transfructosylation)	[115, 116]
Galacto-oligosaccharides	(Gal) <i>n</i> -Glc or (Gal) <i>n</i> -Gal [(β-1,3),(β-1,4),(β-1,6)]	β-Galactosidase	EC 3.2.1.23	Fungi ( <i>A.oryzae</i> ; <i>A. niger</i> ; <i>A. aculeatus</i> ) Bacteria ( <i>Bacillus</i> sp.; <i>Streptococcus thermophilus</i> ; <i>L. acidophilus</i> ; <i>L. reuteri</i> ; <i>Bifidobacterium</i> sp.) Yeast ( <i>Kluyveromyces lactis</i> ; <i>K. marxianis</i> ; <i>Saccharomyces fragilis</i> ; <i>Cryptococcus laurentii</i> )	1,2,3,35, 42, 50	lactose (transgalactosylation)	[24, 117, 118]
Galacto-oligosaccharides	(Gal) <i>n</i> -Fruc or (Gal) <i>n</i> -Gal or (Gal) <i>n</i> -Fruc-Gal [(β-1,1),(β-1,4),(β-1,6)]	β-Galactosidase	EC 3.2.1.23	Fungi ( <i>A. oryzae</i> ; <i>A. aculeatus</i> ) Yeast ( <i>K.lactis</i> )	1,2,3,35, 42, 50	lactulose (transgalactosylation)	[119, 120]
Lactosucrose derived from lactulose	(β-Gal-1,4)-(α-Glc-1,2)-β-Fruc	β-Fructofuranosidase Levansucrase β-Galactosidase	EC 3.2.1.26 EC 2.4.1.10 EC 3.2.1.23	Bacteria ( <i>Arthrobacter</i> sp. <i>K-1</i> ; <i>Z. mobilis</i> ; <i>Bacillus subtilis</i> ; <i>B. natto</i> ; <i>B. circulans</i> )	32, 68, 100, or 1, 2, 3, 35, 42, 50	transfructosylation of lactose or transgalactosylation of sucrose	[121-125]
Lactulosucrose	(β-Gal-1,4)- (β-Fruc-2,1)-α-Glc	Dextransucrase	EC 2.4.1.5	Bacteria ( <i>Leuconostoc mesenteroides</i> )	70	lactulose and sucrose transglucosylation	[126]
2-α-glucosyl-lactose	(β-Gal-1,4)-(α-Glc-2,1)-α-Glc	Dextransucrase	EC 2.4.1.5	Bacteria ( <i>Leuconostoc mesenteroides</i> )	70	lactulose and sucrose transglucosylation	[127]
Isomalto-oligosaccharides	(Glc) <i>n</i> -[α-1,6]	A) α-Amylase with Pullulanase B) α-Glucosidase		Fungi ( <i>Aspergillus</i> sp.; <i>Aureobasidium pullulans</i> ) Bacteria ( <i>Bacillus subtilis</i> ; <i>B. licheniformis</i> ; <i>B. stearothermophilus</i> )	A) 13, 14, 57, 119 ( B) 4, 13, 31, 63, 97, 122	Starch [hydrolysis (A) and transglucosylation (B)]	[128, 129]
Gluco-oligosaccharides	(Glc) <i>n</i> -[(α-1,6),(α-1,3), (α-1,4), (α-1,6)]	Dextransucrase	EC 2.4.1.5	Bacteria ( <i>L. mesenteroides</i> ; <i>L. citreum</i> )	70	Maltose and sucrose (transglucosylation)	[130-132]
Gentio-oligosaccharides	(Glc) <i>n</i> -[β-1,6]	β-Glucosidase or Gluco-endo-1,6-β-Glucosidase	EC 3.2.1.21 EC 3.2.1.75	Fungi ( <i>Penicillium multicolor</i> ; <i>A. oryzae</i> )	1, 3, 5, 9, 30, 116 or 5, 30	Pustulan (hydrolysis) or Gentiobiose (transglucosylation)	[133, 134]
Pectic-oligosaccharides	(GalA) <i>n</i> -[α(1,4)] (GalA-Rha) <i>n</i> -[α(1,2); α(1,4)] *GalA can be partially esterified and Rha ramified	Polygalacturonase Rhamnogalactronan galacturonohydrolase	EC 3.2.1.15 EC3.2.1.173	Fungi ( <i>Fusarium moniliforme</i> ; <i>Aspergillus pulverulentus</i> ; <i>A. aculeatus</i> ; <i>Kluyveromyces fragilis</i> ) Bacteria ( <i>B. licheniformis</i> )	28	Pectin (hydrolysis)	[135, 136]
Xylo-oligosaccharides	(Xyl) <i>n</i> [β(1,4)]	Endo-1,4-β-xylanase	EC 3.2.1.8	Fungi ( <i>Trichoderma</i> sp.; <i>A. oryzae</i> )	5, 8, 9, 10, 11, 12, 16, 26, 30, 43, 44, 51	Xylan (hydrolysis)	[128]
Maltosyl-fructosides	(Fruc) <i>n</i> -Malt [β(2,1), β(2,6)]	β-Xylosidase Levansucrase Inulosucrase	EC 3.2.1.37 EC 2.4.1.10 EC 2.4.1.9	Bacteria ( <i>B. subtilis</i> ; <i>Lactobacillus gasseri</i> )	32, 68	Maltose and sucrose (transfructosylation)	[137, 138]

a. Glc, glucose; Fruc, fructose; Gal, galactose; GalA galacturonic acid; Rha rhamnose; Xyl, xylose; Malt, maltose

b. According to the CAZY database

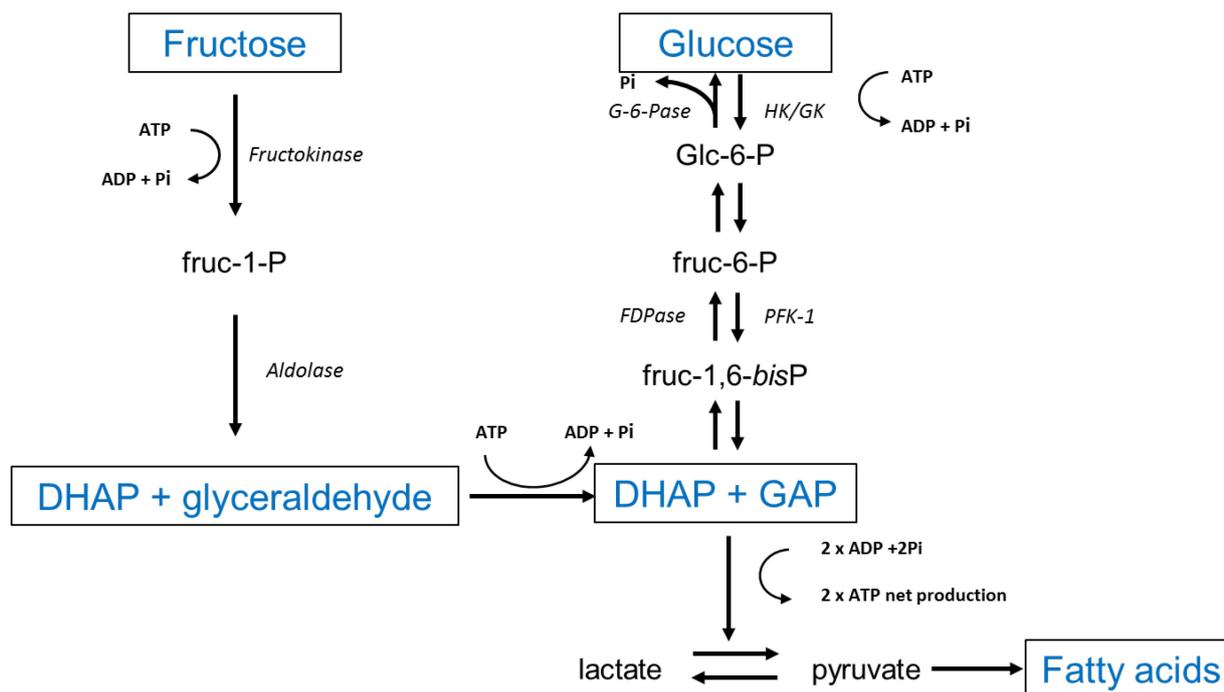
### 2.6.1 Structural characteristics and catalytic mechanisms of microbial Glycoside Hydrolases

The mode of action and substrate specificity of GHs are determined by their tertiary, rather than their secondary structure [112]. In this regard three dimensional structural studies by means of crystallography are crucial in identifying the relevant amino acid residues involved in catalysis and, therefore by default the mechanistic action of any enzyme. Furthermore, site directed mutagenesis studies have established that a few crucial amino acids in the active site of GHs are crucial with regards to establishing linkage specificity [139]. More detailed structural and catalytic mechanisms will be discussed in the following section under either fructose or galactose containing oligosaccharides and polymers.

### 2.6.2 Fructose, Fructans and Fructo-oligosaccharides (FOS), sources and synthesis

Fructose ( $C_6H_{12}O_6$ ) is a hexose monosaccharide and a ketose type reducing sugar. In comparison with fructans, fructose either in its free form or as part of sucrose can be metabolised by humans. Before the advent of the modern diet the only sources of fructose were from fruits and honey, both of which contain relatively low levels. The western diet however is high in refined sugars, fructose sweeteners and high-fructose corn syrup (HFCS) [140]. A high fructose diet is associated with insulin resistance and many other features of the metabolic syndrome including obesity, inflammation and increased risk of type-II diabetes and cardiovascular disease [141]. Once fructose and glucose are absorbed, they are translocated to the liver where the hepatic enzyme system metabolises fructose, converting it into glyceraldehyde and dihydroxyacetone phosphate (DHAP). This is the point where fructose and glucose metabolism converge (Fig.3) and the most fundamental difference between these is that glucose metabolism does not rely on the activity of phosphofructokinase (PFK-1)(Fig.3). PFK-1 converts fructose 6-phosphate to fructose 1,6-bisphosphate and is a pivotal regulatory step in the glycolytic flow being governed by allosteric control. The PFK-1 enzyme is regulated by the concentration of the metabolites AMP, ADP and fructose 2,6-bisphosphate which act as activators and ATP, phosphoenolpyruvate (PEP) and citrate that inhibit it. Furthermore in contrast to glucose metabolism, insulin is unable to influence the hepatic metabolism of fructose and its subsequent conversion to triose phosphates. Therefore, fructose enters the glycolytic pathway unabated and generates excess energy caused by the low  $K_m$  of the enzyme fructokinase and the lack of negative feedback from ATP or citrate. The end result is that the triose phosphates produced from fructose can be converted into pyruvate and oxidized to  $CO_2$  and

H<sub>2</sub>O in the tricarboxylic acid cycle, or can be converted to fatty acids before entering de novo lipogenesis (Fig 3).



**Figure 2-3** The comparative and integrated hepatic metabolism of glucose and fructose: ATP (adenosine triphosphate); ADP (adenosine diphosphate); DHA (dihydroxyacetone); DHAP (dihydroxyacetone phosphate); fruc-1-P (fructose-1-phosphate); fruc-6-p (fructose-6-phosphate); fruc-1,6-bisP (fructose-1,6-bisphosphate); G-6-P (glucose-6-phosphate); Glc-6-Pase (Glucose-6-phosphatase); FDPase (Fructose-1,6-diphosphatase); GAP (D-glyceraldehyde 3-phosphate); PFK-1 (phosphofruktokinase); Pi (phosphate)

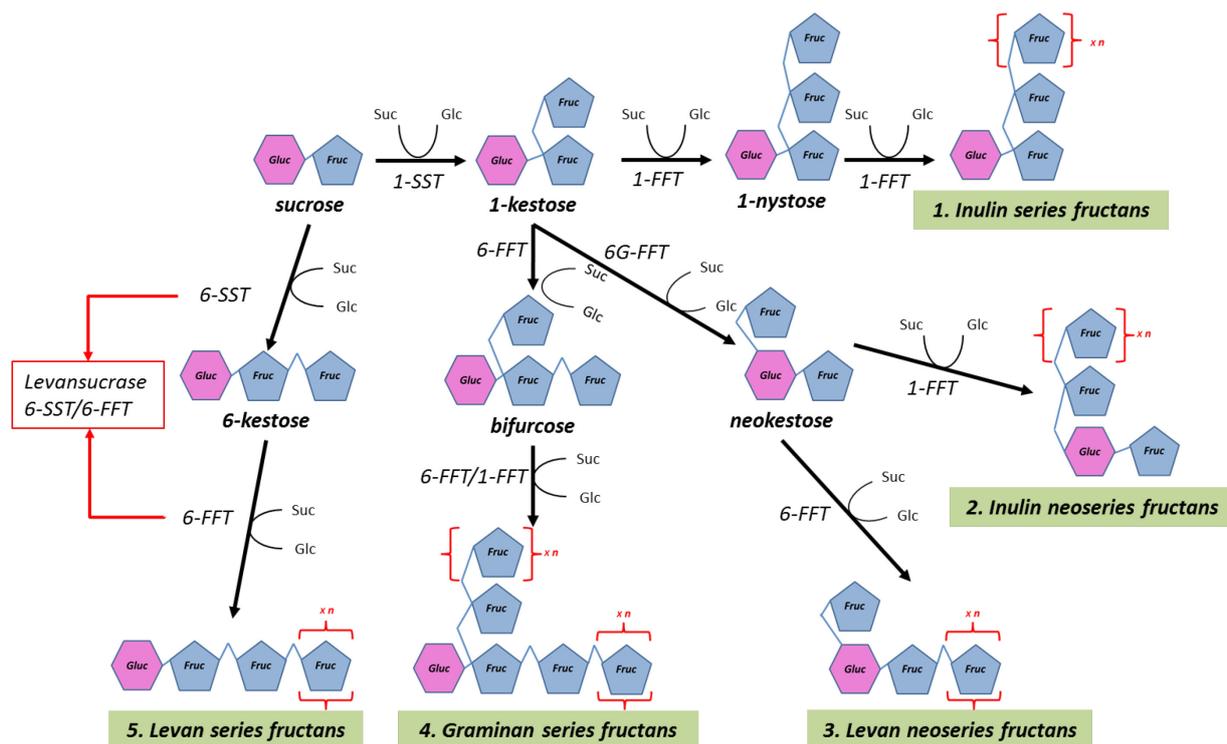
It is necessary that a clear distinction be made between dietary ‘free fructose’ (HFCS, sucrose and fruit sugars) together with its negative health associations described above, and that of fructans, as the two are often confused by the general public. Fructans are regarded as functional foods and have been shown to act as prebiotics and are therefore, not metabolised by the human gastric system, but instead fermented by the GI microbiota. By definition fructans are homopolymeric oligo- and polysaccharides comprised of fructose. They have varying chain lengths and are generally synthesised using a starter sucrose unit which then takes the place of the reducing terminus. The structure of fructans is unusual in the sense that the sugar rings are to the exterior of the polymeric chain of carbon and oxygen atoms [142]. The five membered furanose rings of the chain are linked by -O-CH<sub>2</sub> groups, rendering the structure more flexible than that of the six membered pyranose rings linked by an solitary oxygen, as is found in starch [143]. Apart from the enhanced flexibility induced by the ring structure, it also confers a left-hand twist on the tertiary helix structure [144]. Fructans can also differ from each other based on their structural isomerism linkage types [91]. Together the plethora of different structural characteristics of fructans gives rise to their unique biological characteristics. Broadly speaking a wide range of health benefits have been ascribed to them, including inducing satiety sensing and

counteracting lipogenesis in the liver together with all the other prebiotic benefits described above, as well as anti-viral and antioncogenic traits [91]. Recently fructans have also been recognised as having antioxidative properties and may be useful in disease prevention by actively reducing concentrations of reactive oxygen species (ROS) [145].

With regards to structure there are five main fructan subgroups which differ on the basis of the positional placement of the glucose moiety and on the linkage type between their fructosyl residues. The five sub-classes that can be discerned are, inulin (linear), levan (linear), graminan (branched), neo-inulin (branched) and neo-levan type (branched) fructans. The tri-saccharides 1-kestose, 6-kestose, neokestose as well as the tetra-saccharide bifurcose, are building blocks from which higher degree of polymerization (DP) fructans are synthesized. The five main fructan subgroups and the enzymes that synthesise them are as follows and are illustrated in Figure 2-4.

- I. The best known and well characterised fructan is inulin, which consists of a terminal glucose and a series of  $\beta$ -2,1 linked fructosyl units (Fig 2-4). 1-Kestose is the primary building block of these type of fructans and is formed when a fructosyl moiety is transferred to an acceptor sucrose substrate from a suitable donor by the enzyme *sucrose: sucrose 1-fructosyl transferase* (1-SST). *Fructan: fructan 1-fructosyl transferase* (1-FFT) then polymerizes 1-kestotriose into higher DP inulin-type fructans by continuously adding fructosyl residues in a  $\beta$ -2,1 linkage (Fig 2-4). The best dietary food sources of inulin are garlic, leeks and onions which can contain between 3-18% of their fresh weight as inulin.
- II. The enzyme *fructan: fructan 6G-fructosyl transferase* (6G-FFT) synthesizes 6G-kestotriose (neokestose) from 1-kestotriose as donor substrate and sucrose as acceptor substrate, thereby forming an internal glucose molecule flanked by two fructosyl moieties. Elongation of neokestose by 1-FFT leads to the formation of inulin neo-series type fructan which is abundant in asparagus.
- III. Neokestose is formed as above, but when it is elongated by 6-SFT it leads to the formation of levan neo-series type fructans.
- IV. *Fructan: fructan 6-fructosyl transferase* (6-FFT) transfers a fructosyl moiety from sucrose to the first fructosyl monomeric unit of 1-kestotriose, thereby producing 1,6-kestotetraose (bifurcose). Bifurcose is the smallest graminan-type fructans of mixed linkages and is most commonly found in cereals. Bifurcose can subsequently be further extended by 6-FFT and 1-FFT leading to higher degree of polymerization (DP) graminans.

- V. Linear levan-type fructans are found in animal feed grasses. In plants levan type fructans are not nearly as abundant as the inulin type [142]. Much like inulin, levan in plants is synthesised by two distinct enzymes *sucrose: sucrose 6-fructosyl transferase* (6-SST) which forms the initial 6-kestose upon which 1-FFT elongate the fructan chain in a  $\beta$ -2,6 manner. Levan is however also produced by certain bacteria and yeast by the enzyme levansucrase. Levansucrase can be seen as having both 6-SST and 1-FFT innate activity in that it initially transfers a fructosyl moiety from a sucrose donor to another sucrose molecule thereby forming 6-kestotriose, after which it elongates the linear levan with 6-FFT activity in a  $\beta$ -2,6 manner.



**Figure 2-4** Fructan biosynthesis of the five sub-types of fructans in plants: Glc (glucose); Fruc (fructose); 1-SST (*sucrose: sucrose 1-fructosyl transferase*); 1-FFT (*fructan: fructan 1-fructosyl transferase*); 6G-FFT (*fructan: fructan 6G-fructosyl transferase*); 6-SST (*sucrose: sucrose 6-fructosyl transferase*); 6-FFT (*sucrose: fructan 6-fructosyl transferase*). Note that bacterial and yeast levansucrase activity (red box and arrows) is seen as a combination of 6-SST/6-FFT activity when compared to that of plants.

Fructans and FOS in human diets comes mainly from plant sources most notably wheat (*Triticum aestivum*), rye (*Secale cereale*), barley (*Hordeum vulgare*), oat (*Avena sativa*), leek (*Allium ampeloprasum*), lettuce (*Lactuca sativa*), Belgian endives (*Cichorium intybus*), salsify (*Scorzonera hispanica*), onion (*Allium cepa*), garlic (*Allium sativum*), globe artichoke (*Cynara scolymus*) and asparagus (*Asparagus officinalis*) [58]. The latter four species contain high fructan concentrations and have been extensively researched and utilized in medicine with reported immunomodulatory and antiviral properties [91].

Fructans are recognized as one of the principle storage forms of energy in approximately 15% of higher plants [146]. In dicots, inulin accumulates as a long term reserve carbohydrate in underground roots and tubers whereas in grasses, graminan, levan, and neokestose-derived fructans function predominantly as short-term storage compounds [142]. Of all the fructans, by far the most researched is that of the plant derived inulin, especially with regards to its prebiotic and other health associated benefits [147]. Inulin is commercially manufactured from chicory root, where it comprises roughly one-fifth of the fresh weight of which approximately 55% has a DP range of 2-19, 28% has a DP of 20-40 and 17% has a DP higher than 40, giving it a good polydispersity index (PDI) [115]. This is a measure of distribution of molecular mass within a given polymer sample and is calculated by the weight average molecular weight divided by the number average molecular weight. Different chain lengths of inulin are fermented at different rates according to their DP [148]. Inulin type oligosaccharides and inulin with a lower DP are fermented rapidly in the proximal part of the colon where they exert the characteristic bifidogenic effect. Higher DP inulins are fermented more slowly and are able to reach the distal parts of the colon where readily fermentable carbohydrates are in short supply [149]. This creates a low carbohydrate environment where the microbiota experiences a metabolic shift towards predominantly proteolytic fermentation due to protein being the only remaining nutrient supply, resulting in the production of toxic putrefactive metabolites. High DP fructans are able to reduce the proteolytic activity in favour of a beneficial saccharolytic activity in these areas [150, 151].

An alternate method of obtaining both FOS and high molecular weight fructans is through microbial production of levan with the enzyme levansucrase (EC 2.4.1.10). Whereas in plants the functional role of fructans relates to bridging the temporal gap between resource availability and energy demands, in microorganisms their evolutionary role relates more to exopolysaccharide production for microbial biofilm formation in response to environmental and competitive stressors [152]. Plants require at least two distinct and separate enzymes namely *sucrose: sucrose 6-fructosyl transferase* (6-SST) and *Sucrose: fructan 6-fructosyl transferase* (6-SFT) to catalyse the initial priming and subsequent chain elongation in the synthesis of levan type fructans (Fig 2-4). The low availability of these in plants compared with inulin makes plants an unfeasible source for commercial extraction. Microbial levan biosynthesis, however, is catalysed by the single enzyme, levansucrase, which simultaneously both primes and elongates using sucrose predominately as a substrate. This, therefore, makes it a great deal more attractive for industrial application [153, 154]. Numerous levansucrases that produces either levan or levan type FOS have been identified from a wide range of microorganisms [155]. They can be utilized either in a crude cell extract, partially or completely purified, produced in a recombinant system, or immobilized to produce levan on an reactor based scale [156]. The levan yield as well as the DP and polydispersity index vary greatly depending on the source of levansucrase, and the specific reaction conditions such as temperature, pH, initial substrate concentration and ionic strength [154, 157, 158].

As mentioned above, in comparison with  $\beta$ -2,1 linkages of inulin obtained from plant sources, microbial levan consist primarily of  $\beta$ -2,6 linked fructosyl units, leading it to have additional and different biological and commercial characteristics. Together with its popular properties, such as renewability, flexibility, bio-degradability and industrial application as a natural adhesive and surfactant, levan also has numerous biomedical properties. These include acting as anti-inflammatory, anti-viral, anti-carcinogenic, fibrinolytic, hypolipidemic, immunostimulatory and anti-oxidant agents as well as being a hyperglycaemic inhibitor [159-164]. Furthermore levan has found applications in the food and cosmetic industry as emulsifiers and thickening agents and, in addition, levan type FOS are being utilized as a health beneficial artificial sweeteners [165]. Recently more novel applications for levan has been proposed in terms of nanotechnology based drug delivery systems and wound dressings [155]. By utilising a combination of synthetic biology and microbial fermentation techniques levan can also be effectively converted to biofuel [166]. One of the most useful characteristics of levansucrase is that it has been shown to hold a broad range of acceptor specificities to which it can transfer the fructosyl moiety after the initial sucrose hydrolysis. These include numerous other monosaccharides, disaccharides and oligosaccharides, thereby creating scope for the synthesis of valuable, novel or rare fructosylated saccharides e.g. lactosucrose (lactose-Fruc) and sucrose analogues (D-Gal-Fruc, D-Man-Fruc, D-Xal-Fruc and D-Fruc-Fruc) [155]. Other fructosylated compounds can also be formed when the fructosyl moiety is transferred to donors other than saccharides and the synthesis of alkyl-fructosides (methyl-fructoside) and the fructosylation of aromatic alcohols like hydroquinone have been reported [123, 167, 168].

With regards to levan as a prebiotic substrate surprisingly little research has been conducted compared to that examining inulin, with only six recent papers addressing the subject [144, 169-173]. The prebiotic potential of any complex carbohydrate is dependent on numerous factors including, but not limited to, its degree of polymerization, monomeric composition, linkage type, crystallinity, solubility and its relationship with other substrates [174]. From a microbial perspective the most rudimentary requirement is whether or not the resident microbiota are in possession of the necessary metabolic machinery to degrade the complex carbohydrates. Among GI microbiota, the Bacteroidetes, Firmicutes and Actinobacteria phyla are the most prominent in the degradation of plant fibre polysaccharides. Indeed, inulin administration has been shown to increase *Bifidobacterium* populations both *in vivo* and *in vitro*, but this would be expected as *Bifidobacterium* are known to possess  $\beta$ -fructofuranosidase activity that specifically cleaves the  $\beta$ -2,1 linkage type of inulin [175, 176]. Whether the same metabolic machinery is available for the degradation of microbial derived  $\beta$ -2,6 linked levan is still being investigated, although compelling evidence has been presented. As mentioned above microbial levan can either be in the form of FOS or larger polymers than can have a DP of several thousand [177]. Several studies have concluded that FOS with  $\beta$ -2,6 linkage type selectively enhance *Bifidobacteria* and

*Lactobacilli* numbers presumably by providing a good growth substrate that consequently lead to the production of lactic and acetic acid (For review see [144]). In addition, certain *Bacteroides* species has been shown to possess extra-cellular endo-levanase activity capable of degrading the longer chain levan into oligosaccharides which are more readily fermented [178]. Visnapuu *et al.* (2015) hypothesised that even if some probiotic microbiota cannot hydrolyse the longer chain  $\beta$ -2,6 fructans, they are still stimulated by it due to the ability of *Bacteroides* species and/or other bacteria with extracellular levan-degrading capabilities to break it down [144]. To further substantiate levan's proposed application as a prebiotic, Adamberg *et al.* (2015) showed most recently that high molecular weight levan enhances the associated growth of levan-degrading (e.g. *Bacteroides*) and butyric acid-producing (e.g. *Faecalibacterium*) taxa in levan supplemented media and that the main products of levan metabolism were acetic, lactic, butyric, propionic and succinic acids as well as carbon dioxide [173].

Therefore levan has a broad array of applicability including that of a prebiotic fibre. The search for novel enzymatic catalysts with enhanced activities and specificities that synthesise them has great potential to improve their functionalities.

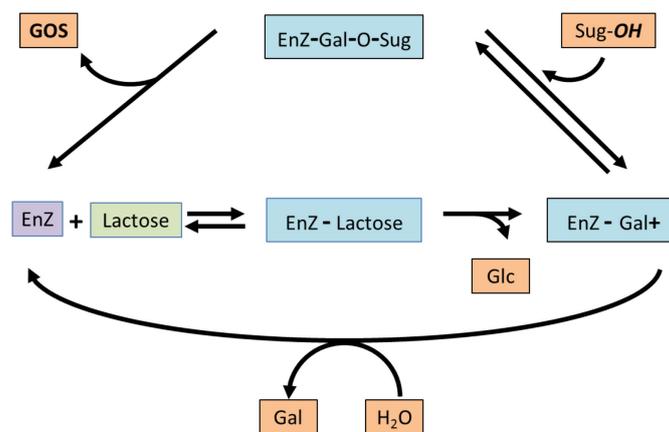
### 2.6.3 Galacto-oligosaccharides (GOS) as prebiotic, sources and synthesis.

The hydrolysis of lactose in milk is one of the most important technological procedures in the food processes industry. The reasons for this are several fold namely (i) to diminish problems of lactose intolerance that are present in more than half the global adult population [179, 180], (ii) formation of galacto-oligosaccharides so as to enhance the growth of beneficial intestinal microbiota [181], (iii) enhancement of the sensorial characteristics of hydrolysed milk products (milk with high lactose content crystallise in ice-cream and the hydrolysed glucose sweetens the milk) [182, 183], and (iv) improvement of the degradability of hydrolysed waste whey water [184]. Lactose can be hydrolysed either by means of acid hydrolysis or enzymatically using  $\beta$ -galactosidase, however it is only by means of the latter that GOS can be synthesised [182]. Additionally the enzymatic process doesn't add any by-products or alter the nutritional content or colour of milk, making it the preferred method. Several advances have been made in the field and lactose can be hydrolysed/transformed either with the addition of whole cell extracts, or by immobilised  $\beta$ -galactosidase on membranes by means of cross-linking, thereby making the enzymes reusable [183, 185, 186].

The enzyme  $\beta$ -galactosidase (EC 3.2.1.23) is of special historical significance in molecular biology. It played a pivotal role in Jacob and Monod's development of the operon model for the regulation of gene expression and today the *lac* operon is synonymous with transcriptional regulation [187]. Furthermore,

its ability to phenotypically manifest its presence by producing an easily recognizable blue/indigo product through its activity on X-gal has made it a workhorse in molecular cloning *via* blue/white colour selection. The basis of screening on X-gal is  $\alpha$ -complementation in the presence of IPTG. In early studies of  $\beta$ -galactosidase, it was observed that deletion of certain amino acid residues near the amino-terminus caused the tetrameric enzyme to dissociate into inactive dimers thereby abolishing activity. By co-expressing some or all of the missing peptide residues it was possible to reconstitute the tetrameric form of the enzyme and thereby its inherent catalytic activity [188].

There are two intrinsic competitive activities inherent in  $\beta$ -galactosidase namely those of hydrolysis and trans-galactosylation.



**Figure 2-5**  $\beta$ -galactosidase activity mechanism: EnZ (enzyme); EnZ-Lactose (enzyme lactose complex); EnZ-Gal+ (enzyme and galactose covalently bonded); EnZ-Gal-O-Sug (enzyme and galactose as well as acceptor molecule covalently bonded) Sug-OH (sugar acceptor molecule presenting –OH group).

The trans-galactosylation reaction usually takes place in the presence of a high lactose substrate concentration [189]. Three distinct steps can be discerned for the general reaction mechanism of  $\beta$ -galactosidase, the last of which allows for either hydrolysis or trans-galactosylation (Fig 2-5):

1. enzyme + lactose  $\rightarrow$  enzyme -- lactose (complex)
2. enzyme -- lactose (complex)  $\rightarrow$  galactosyl -- enzyme + glucose
3. galactosyl -- enzyme + acceptor  $\rightarrow$  galactosyl – acceptor + enzyme

In the first step the lactose substrate is docked in the active site of the tetrameric enzyme and an enzyme substrate complex is formed. In the second step, an amino group in the active site acts as a general acid and donates a proton to the glycosidic oxygen. Within the active site there is also another negatively charged group that then stabilizes the positively charged carbonium galactosyl intermediate by forming a temporary covalent bond [181]. The third and final step of the reaction mechanism involves the galactosyl transfer to a nucleophilic acceptor molecule. All reducing sugars can act as acceptors resulting

in a final mixture of di-, tri-, tetra-, and even higher saccharides. However, acceptor molecules act in a competitive manner and, if H<sub>2</sub>O is the acceptor molecule, hydrolysis occurs.

As described above, GOS are galactose containing oligosaccharides that usually take the form of  $\alpha\text{-Glc-1,4}[\beta\text{-Gal-1,3/4/6}]_n$ , where  $n = 2\text{-}9$ . The initial acceptor molecule is primarily lactose, to which consecutive galactosyl moieties are added through the trans-galactosidase activity of the enzyme  $\beta$ -galactosidase [181]. However,  $\text{Gal}_n\text{-Gal}$  structures have also been shown to be formed when a galactosyl moiety is transferred to free galactose and branched structures of GOS have also been identified. Di-, tri-, tetra-, and penta-saccharides are the main end products with the disaccharide allolactose ( $\beta\text{Gal-1,6-Glc}$ ) and galactobiose ( $\beta\text{Gal-1,6-Gal}$ ) making up the majority of the trans-galactosylation product [190].

The concentration and yield of GOS obtained, as well as the different structural linkage types formed during their enzymatic synthesis, depend on various factors, most notably the reaction conditions and the specific enzyme used. The overall efficiency of the trans-galactosylation reaction is kinetically controlled and depends on the specific biochemical attributes of the enzyme. The GOS themselves are potential substrates and can be hydrolysed by the enzyme. The enzymatic reaction eventually hydrolyses all lactose and reaches an equilibrium with the resultant mixture of monosaccharides, disaccharides and oligosaccharides dependent on the conditions and enzyme used [191]. To produce GOS of a higher DP the enzyme would need to have a higher propensity to catalyse the trans-galactosylation reaction relative to that of the hydrolysis reaction and, and/or have a lower affinity for the hydrolysis of the already formed GOS relative to that of lactose [181].

As already noted GOS can differ in their saccharide composition, type of regio-chemical glycoside linkages, as well as the degree of polymerization [189]. These differences impart physicochemical and biological attributes that affect their prebiotic capabilities and are also relevant to other food applications. *In vitro* evidence suggests that the prebiotic microbiota differentially ferment oligosaccharides with different structures [192]. All prebiotics enhance the growth and/or activity of the beneficial microbiota but GOS has, furthermore, been shown to be a potent receptor decoy to prevent adherence of pathogenic bacteria [193]. Most recent research has emphasised the profound effect GOS have on inflammation, hypercholesterolemia as well as its ant-carcinogenic effects [194-198]. Special interest is also given to GOS as they closely resemble human milk oligosaccharides (HMO) and all evidence indicate that they have a large bifidogenic effect if added to infant formula. The GI microbiota of infants receiving GOS supplemented formula, more closely resemble that of breastfed infants [27, 199]

As mentioned the preceding sections there are two possible enzymatic systems for the production of GOS from lactose, namely glycosyltransferases (EC 2.4) and glycoside hydrolases (EC 3.2.1).

Galactosyltransferases catalyse the transfer of sugar moieties from an activated sugar nucleotide to an acceptor molecule leading to an increasing DP [114]. Galactosyltransferases are highly selective with regards to acceptor molecules and bond formation and can produce high yields of GOS and longer chain polysaccharides [200]. Even though novel bi-enzymatic systems have been developed with regards to incorporating the nucleotide sugar production systems, as well as the galactosyl-transferase system, their combined metabolic burden makes the requirement too cumbersome for it to be feasible on an industrial level [114].

Many organisms exhibit endogenous  $\beta$ -galactosidase activity and it can be extracted from a variety of sources including microorganisms, plant and animal cells [118]. The  $\beta$ -galactosidase utilized by industry originates from various microbial sources. Torres *et al.* (2010) compiled a comprehensive list of both the fungal and bacterial sources of enzymes as well as their optimum reaction conditions and yield [117]. Traditionally the main sources are the yeast *Kluyveromyces lactis* and the fungus *Aspergillus oryzae.*, due to their easy cultivation and GRAS status [118]. However, increasingly bacterial sources are being investigated especially in regards to recombinant DNA technology. *Bifidobacterium*, *Lactobacillus* and *Bacillus* species are some of the bacteria utilized thus far (Table 2-2). There remains, however, a need to find or develop better production mechanisms for prebiotic GOS as well as FOS and longer chain polysaccharides. One way of accomplishing this is through the construction and screening of metagenomic libraries for enzymes with improved capabilities.

As mentioned already in section 1.2. of the general introduction the aim of this study was for the discovery and subsequent objective characterisation of enzymes that produce novel polymers and/or oligosaccharides comprised of either galactose or fructose. The enzymes that facilitate this belong to the glycoside hydrolyse type family proteins and possess both hydrolytic as well as trans-glycosidic activity as already reviewed. The search for these enzymes was mediated through the construction and use of metagenomic libraries from niche environments that have an abundance of either galactose or fructose present.

**Table 2-2** Bacterial sources of glycoside hydrolases used in the production of GOS (adapted from Torres *et al.*,2010 [117])

Bacteria	Enzyme & preparation	Synth.	GH	T/°C	pH	L <sub>0</sub>	Y <sub>max</sub>	Ref
<b>→ Actinobacteria</b>								
<i>Saccharopolyspora rectivirgula</i>	β-Galactosidase, purified		GH2	70	7.0	60	44	[201]
<i>Bifidobacterium adoloescentis</i>	crude enzyme fraction			55	7.0	30	43	[202]
<i>Bifidobacterium angulatum</i>	crude enzyme fraction			55	7.5	5-30	44	[202]
<i>Bifidobacterium bifidum</i>	β-Galactosidase (BIF3), mutated		GH2	37	6.0	10-40	44	[203]
<i>Bifidobacterium bifidum</i>	crude enzyme fraction			55	7.5	30	38	[202]
<i>Bifidobacterium bifidum</i>	Cells, resting cells			39	6.8	50	20	[204]
<i>Bifidobacterium bifidum</i>	Cells, toluene treated, resting cells			40	6.2-6.8	45-50	44	[205]
<i>Bifidobacterium bifidum</i>	Cells, toluene treated, resting cells			40	5.0-5.5	Whey, 47	38	[206]
<i>Bifidobacterium longum s. infantis</i>	β-Galactosidase(Bgal I), cloned		GH2	30	6.5	36	49	[207]
<i>Bifidobacterium longum s. infantis</i>	β-Galactosidase(Bgal II), cloned		GH2	30-60	7.5	20-30	68	[208]
<i>Bifidobacterium longum s. infantis</i>	β-Galactosidase(Bgal III), cloned		GH42	37	7.5	20	10	[209]
<i>Bifidobacterium longum s. infantis</i>	crude enzyme fraction			55	7.5	30	48	[202]
<i>Bifidobacterium longum s. infantis</i>	Cells, in culture			60	7.5	Milk, 35	43	[210]
<i>Bifidobacterium longum</i>	crude enzyme fraction			45	6.8	5-50	35	[211]
<i>Bifidobacterium pseudolongum</i>	crude enzyme fraction			55	7.5	30	27	[202]
<b>→ Deinococcus-Thermus</b>								
<i>Thermus sp.</i>	β-Glucosidase, cloned		GH1	70	7.0	7.5-30	40	[212]
<i>Thermus caldophilus</i>	β-Glucosidase (BglA), cloned		GH1	70-80	6.0	30-50	77	[213]
<b>→ Thermotoga</b>								
<i>Thermotoga maritima</i>	β-Galactosidase (LacZ), cloned		GH2	80	6.0	20-50	19	[214]
<b>→ Firmicutes</b>								
<i>Bacillus circulans</i>	β-Galactosidase (I), purified			40	6.0	4.6	6	[215]
<i>Bacillus circulans</i>	immobilized enzyme extract	SG	GH2	40	6.0	4.3	40	[216]
<i>Bacillus circulans</i>	β-Galactosidase (I), purified		GH3	40	6.0	4.6	41	[215]
<i>Bacillus circulans</i>	immobilized enzyme extract	SG, PBR	GH4	40	6.0	4.6-20	48	[217]
<i>Bacillus circulans</i>	immobilized enzyme extract	PR, PBR	GH5	40	6.0	4.6	35	[216]
<i>Bacillus circulans</i>	crude enzyme fraction			40-55	6.0	6.5-36	43	[218]
<i>Bacillus sp.</i>	crude enzyme fraction			50-55	5.0	33-36	43	[219]
<i>Bacillus sp.</i>	immobilized enzyme extract	Chitosan		55	5.5	36	41	[220]
<i>Geobacillus stearothermophilus</i>	β-Galactosidase (BgaB), cloned		GH42	37	6.5	17	2.4	[221]
<i>Geobacillus stearothermophilus</i>	β-Galactosidase (BgaB), mutated		GH42	37	6.5	17	30	[221]
<i>Caldicellulosiruptor saccharolyticus</i>	β-Glucosidase (BglA), cloned		GH1	65-80	6.3	10-72	63	[222]
<i>Lactobacillus acidophilus</i>	β-Galactosidase (LacL+LacM), cloned		GH2	30	65.0	21	39	[223]
<i>Lactobacillus reuteri</i>	β-Galactosidase (LacL+LacM), cloned		GH2	23	6.0	21	26	[224]
<i>Lactobacillus reuteri</i>	β-Galactosidase (LacL+LacM), purified		GH2	30-37	6.0-6.5	4.6-21	38	[225]
<i>Lactobacillus reuteri</i>	immobilized enzyme extract	CMAR	GH3	37	6.0	21	29	[226]
<i>Streptococcus thermophilus</i>	β-Galactosidase (LacZ), purified		GH2	37		Milk, 5.3	25	[227]
<b>→ Proteobacteria</b>								
<i>Enterobacter agglomerans</i>	β-Galactosidase (LacZ), cloned		GH2	50	7.5	5 to 13	38	[228]
<i>Enterobacter cloacae</i>	β-Galactosidase (Bga/LacZ), cloned		GH2	50	6.5	28	49	[228]
<i>Enterobacter cloacae</i>	Cells, resting cells			50	6.5	28	55	[229]
<i>Escherichia coli</i>	β-Galactosidase (LacZ), purified		GH2	30 to 37	6.5 to 7.2	2.2 to 24	56	[230]
<i>Escherichia coli</i>	immobilized enzyme extract	RM		37	7.5	2.2	44	[230]

GH, glycoside hydrolase family. Synth., synthesis method. L<sub>0</sub>, initial lactose concentration (%w/v). Y<sub>max</sub>, maximum GOS yield reported (%) = GOS concentration/L<sub>0</sub> x 100. SG, immobilized support gel. PR, phenol-formaldehyde resin. PBR, packed bed reactor. CMAR, membrane assisted reactor. RM, reverse micelles.

## CHAPTER 3: Metagenomic library construction, screening and preliminary analysis.

### 3.1 Introduction

Despite their promising performance in the laboratory, the application of enzymatic biocatalysts on industrial or semi-industrial level often ends in failure [231]. This limitation is primarily due to the lack of availability of microbial enzymes that can perform the desired chemical reactions under industrial conditions. Even if a catalytic specific enzyme is available, the enzymatic fitness itself may be inadequate and the industrial processes redesigned so as to fit the characteristics of that enzyme [232]. Therefore, the quest for suitable, robust and need specific enzymatic biocatalysts is constant.

Over the last decade, the ever advancing next generation sequencing platforms together with their mass data acquisition and concurrent bioinformatic sequence analysis has taken centre stage, resulting in over 190 billion nucleotides of sequence data being available in the public domain by 2010 [233]. Currently there are over a trillion nucleotides on GenBank and tenfold more from whole genome sequencing projects (Genbank statistics; <https://www.ncbi.nlm.nih.gov/genbank/statistics/>).

However, several limitations do exist with regards to the search for functional biocatalysts *via* next generation sequencing. Accurate functional assignment of genes is a major challenge with mass bioinformatic analysis. Direct comparison of functional and metagenomic analyses have laid bare these limitations in that functional misannotation or over-representation often results in true positives being missed. Although environmental metagenome sequencing is being conducted by means of next generation platforms, the majority of next generation sequencing taking place is on the single organism level. Most industrial enzymes currently being utilized are of microbial origin, however fewer than 1% of bacteria present in most environments can be cultured under standard laboratory conditions, resulting in many potential targets being missed. Taking all of the above into consideration, it is not surprising that the functional screening of metagenomic libraries still plays a prominent, and often central, role in enzyme discovery [234].

Metagenomic library construction and screening is however not without its own set of limitations. The probability of identifying a specific gene depends on multiple factors including, source metagenomic DNA, host system, vector system, target gene size, its relative abundance in the source metagenome, efficiency of heterologous gene expression in host and, most importantly, the functional screening method itself [234].

The most critical consideration for the construction of a successful and functional metagenomic library is the sampling site, in order to maximise the likelihood of the presence of the desired genetic elements.

For the purposes of this study the main interest lies in glycoside hydrolases that demonstrate both hydrolytic and trans-glycosidic activities, thus bypassing the need for complex bi-enzymatic systems utilising glycosyltransferases and activated sugar nucleotide substrates. Therefore, in the pursuit of microbial glycosidases that produce either a galactose or fructose containing oligosaccharide/polymer, logic dictates that environments that are abundant in disaccharidic substrates containing the desired monosaccharides moieties i.e. lactose and sucrose should be targeted. In addition to targeting substrate rich environments, environments where desirable enzymatic characteristics (like temperature stability and a broad pH range) could develop would serve as the main parameter for environmental sampling selection.

For the construction of metagenomic libraries most researchers utilise *E. coli* as a host system. Almost all laboratory strains of this bacterium lack the homologous recombination genes *recA* and *recBC*, as well as the functional restriction enzyme genes *mcrA* and *mcrBC*, making it useful for cloning modified foreign DNA. Its main drawback though, is that even though *E. coli* is efficient for the expression of a great quantity of microbial genes, the functional expression of eukaryotic genes is often low or absent due to introns. Therefore, functional screens in *E. coli* using metagenomic libraries constructed from environmental samples will mainly identify prokaryotic genes.

The choice of a vector depends largely on the size of the inserts and whether single genes or whole operons are targeted. Plasmids are suitable for cloning genetic fragments up to 12 kb, while cosmids (20-35 kb), fosmids (20-40 kb) and BACs (50-200 kb) are used to clone larger fragments. The targets for novel enzyme discovery are usually single genes thus making plasmid based bacteriophage library systems the method of choice for most gene discovery. Plasmid vector systems can also have the advantage of having a high copy number and strong promoters [235, 236].

The methodology used to functionally screen metagenomic libraries can also be critical in terms of identifying industrially useful enzymes. Enzymatic activities are usually assayed on agar plates supplemented with specific substrates. By plating a metagenomic library on the before mentioned plates, positive clones can be identified through visual screening for the presentation of a phenotype or colour reaction.

For the purposes of this study a dual approach was taken towards the screening of the metagenomic libraries based on the dual nature of the required glycosidases, i.e. it's hydrolytic and trans-glycosylation activities. Firstly, depending on the source of the library, it will be screened on minimal media plates containing either lactose or sucrose as the sole carbon source. Cells should only grow if they contain an inserted sequence that encodes a functional enzyme capable of hydrolysing the disaccharides, thereby liberating the glucose hexose moiety that can enter glycolysis. The simultaneous presence of an exopolysaccharide, caused by polymer formation using the residual monosaccharide obtained from the

disaccharide after hydrolysis, will further serve to identify a positive phenotype. Additionally, with regards to screening for glycoside hydrolases that utilise lactose as a substrate, the well-known molecular technique of blue/white screening on X-gal will be exploited. In normal molecular techniques blue/white selection is utilized to establish whether a vector containing the coding sequence for the lacZ-alpha monomer is uninterrupted, or if a sequence of DNA has been inserted into it. If it is uninterrupted then a functional peptide is expressed and the LacZ-alpha protein forms a hetero-enzyme complex with LacZ-omega, consequently restoring activity and giving the cell the ability to hydrolyse X-gal resulting in the phenotypic indigo blue colour [188]. In this study, an *E. coli* mutant was constructed lacking both the LacZ-alpha and LacZ-omega subunits, thereby abolishing any and all  $\beta$ -galactosidase activity even if a vector derived  $\alpha$ -subunit were to be present. Colour formation on X-gal will only occur, therefore, if an exogenous coding sequence is present in the library derived vector, capable of its hydrolysis.

Once positive clones have been identified and confirmed, they can be sequenced through conventional methods, after which they can be cloned and partially characterised so as to establish potential suitability for industrial applications.

### 3.2 Materials and Methods

All chemicals were of molecular biology grade. Restriction endonucleases, DNA ligase, and RNase were purchased from Thermo Scientific (Waltham, Massachusetts, United States). GoTaq was purchased from Promega (Madison, Wisconsin, United States). Galactose dehydrogenase/mutatarose, hexokinase/glucose-6P-dehydrogenase were obtained from Megazyme (Bray, Ireland). All other chemicals were purchased from Sigma (St. Louis, Missouri, United States).

#### 3.2.1 Sample collection and DNA isolation

Several libraries were constructed from metagenomic DNA extracted from environmental samples. For galactose containing polymers/oligosaccharides soil was collected at the Faircape dairy farm located in the Western Cape province of South Africa during early autumn on the 25<sup>th</sup> April 2014. Sampling was done at a place where milk runoff occurs from a storage container into the soil (GPS coordinates: 33°45'01.2"S 18°36'44.9"E). The ambient temperature was approximately 20°C . Several 'interesting' sub-sites were sampled in and around the general site, and collected and sealed in plastic containers

(Fig. 3-1). Samples were taken to the lab where their pH's were measured individually at room temperature, before being amalgamated into one general sample for each respective library. For the second library, sampling was done at an intertidal zone where galactose containing coastal kelp degradation takes place at Gansbaai on the Western Cape coast, South Africa (GPS coordinates 34°34'26.9"S 19°20'45.2"E). Samples were collected during autumn on May the 15<sup>th</sup> 2014, early afternoon. The ambient temperature during time of sampling was approximately 19°C. Much like the sampling for the previous library, several samples around the general sampling site were collected and taken to the laboratory for further analysis and extraction (Fig 33-1). For the library constructed for the purposes of screening for fructose containing polymers/oligosaccharides samples were collected from onsite runoff occurring at the Illovo sugar refinery Durban, KwaZulu-Natal, South Africa (GPS coordinates: 30°05'28.5"S 30°49'30.8"E) during September 2013. It should be noted that the author himself didn't conduct the sampling of the latter site, and no other physical parameters are known other than the fact that the refinery conditions are generally hot and humid with temperatures usually in excess of 30°C.

A modified protocol was utilized for the extraction of all genomic DNA based on the protocol of Verma *et al.* (2011) and Dos Reis Falcão *et al.* (2008) [237, 238]. A total of 100 g of soil/sludge was extracted for each sample. In brief; Soil (1 g) was suspended with activated charcoal (0.4 g) and 20 µl of proteinase K (10 mg.ml<sup>-1</sup>) was added together with 2 ml of modified extraction buffer [(1%; w/v) N,N,N,N-cetyltrimethylammoniumbromide (CTAB), (2%; w/v) polyvinyl-poly pyrrolidone (PVPP), 1.5 M NaCl, 100 mM EDTA, 0.1 M TE (Tris-HCL EDTA) buffer (pH 8.0), 0.1 M sodium phosphate buffer (pH 8.0), and 100 µl RNase A (10 mg.ml<sup>-1</sup>). The sample was incubated at 37°C with orbital shaking at 200 rpm for a total of 30 min. Subsequently, 200 µl of SDS (10%; w/v) was added to the homogenate and further incubated at 60°C for 2 hrs with intermittent shaking. DNA was precipitated by adding 1 ml PEG 8000 (30%; w/v in 1.6 M NaCl) and left at room temperature for an hour. The precipitated DNA was collected by centrifugation at 12,000 xg at 4°C. The supernatant was discarded, and the pellet resuspended in 1 mL TE buffer (pH 8.0) after which 100 µl of 5 M potassium acetate was added and incubated at 4°C for 30 min. The supernatant was collected after centrifugation at 12,000 xg and twice extracted with an equal volume of a phenol-chloroform : isoamyl alcohol (25:24:1; v/v). This was followed by another extraction with an equal volume of a chloroform/isoamyl alcohol (24:1; v/v) only, before being centrifuged at 10,000 xg for 15 min. The aqueous layer was transferred to another tube and treated with a 0.7 volume of isopropanol for 1 hour at room temperature. The DNA was pelleted by centrifugation at 12,000 xg for 20 min at 4°C after which the supernatant was decanted and the remaining pellet carefully washed with 1 ml of 70% (v/v) ethanol and then briefly dried at room temperature. The dried pellet was resuspended in 100 µl of sterile Milli Q water. To further purify and remove polysaccharide contamination an equal volume of GHCl extraction buffer [6.5 M guanidinium

hydrochloride, 100 mM Tris-HCl (pH 8.0), 0.1 M sodium acetate (pH 5.5), 0.1 M  $\beta$ -mercaptoethanol, 0.2 M KOAc] was added and gently mixed. The sample was then centrifuged at 12,000  $g$  for 15 min to selectively precipitate the polysaccharides, and the supernatant transferred to a new tube. The DNA was again precipitated with two volumes of ethanol, pelleted by centrifugation at 12,000  $g$  for 15 min and suspended in 100  $\mu$ l ddH<sub>2</sub>O. After DNA extraction the quality and quantity of the DNA was assessed by agarose gel electrophoresis on a 1% gel. The purity was assessed spectrophotometrically and nuclease activity on the extracted DNA was qualified visually by restriction enzyme (RE) digestion of 3  $\mu$ g of extracted DNA with 3 units XhoI in the supplied 1X buffer in a final volume of 50  $\mu$ l.

### 3.2.2 Metagenomic library construction and *in vivo* mass excision

Libraries were constructed with the Lambda ZAP express pre-digested vector kit according to the manufacturer's recommended protocol (Agilent Technologies). In brief: 5  $\mu$ g genomic DNA was digested with *Bsp*143I RE ranging between 0.2  $u \cdot \mu$ l<sup>-1</sup> and 0.0125  $u \cdot \mu$ l<sup>-1</sup> in a final 15  $\mu$ l reaction volume for 30 min after which it was heat inactivated at 70 °C for 10 min and then chilled on ice. One unit of enzyme activity is defined as the amount of enzyme needed to digest 1  $\mu$ g of DNA in a 50  $\mu$ l volume in one hour. The digestion products were separated by agarose gel electrophoresis (1%; w/v) in order to identify the correct enzyme concentration for optimal partial digestion, where the majority of the digested fragments were larger than 2 kb in size. After establishing the desired RE to DNA ratio in the correct volume, the reaction was up-scaled so as to digest 100  $\mu$ g of total DNA. The digested DNA was then fractionated on a 20 cm long 1 ml Pasteur pipette, packed with Sepharose CL-2B, by the gravity fed method. The column was pre-washed with 10 ml of STE buffer (100 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl pH 8.0). The sample was loaded together with 1  $\mu$ l of saturated bromophenol blue solution to serve as a tracking dye. Three drop fractions were collected in microfuge tubes up until the dye ran through the column. Five microliter of each fraction was analysed by agarose gel electrophoresis (1%; w/v). Fractions containing DNA of the appropriate size (2 kb and higher) were pooled, precipitated with 2 volumes ethanol, washed with ethanol (70%; v/v) and reconstituted in 20  $\mu$ l Tris buffer (5 mM; pH 7.5). A total of 4  $\mu$ g of size fractured DNA (avg. size ~ 3 kb) was ligated with a roughly estimated equimolar amount of lambda ZAP vector in an overnight reaction, so as to attain the ligated concatemeric DNA required for optimal packaging, with the aid of T4 DNA ligase. The ligation reaction was then heat inactivated at 60°C for 10 min and immediately cooled down on ice. Commercial packaging extracts were used to encapsulate recombinant lambda phage DNA with high efficiency within their proteinaceous viral capsules. The extract itself is an extremely temperature sensitive viral protein preparation that is free of any foreign DNA. It is restriction minus thus allowing methylated DNA

to be packaged (HsdR<sup>-</sup> McrA<sup>-</sup> McrBC<sup>-</sup> McrF<sup>-</sup> Mrr<sup>-</sup>). The ligated DNA was packaged with Gigapack III packaging extract according to the manufacturer's recommended protocol (Agilent technologies). In brief: the packaging extracts were removed from the -80°C freezer and briefly thawed between the finger tips before being placed on ice. The ligated DNA preparation was carefully added to the packaging extract and gently stirred with a pipette tip whilst taking care not to introduce any bubbles. The tube was briefly spun down and incubated at 22°C for two hours. The reaction was stopped by adding 500 µL of SM buffer [25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 8 mM MgSO<sub>4</sub>, gelatin (0.01%; w/v)], and 20 µL of chloroform. The tube was briefly spun down to sediment debris. The primary library was now ready for titering and stored at 4°C until needed.

The primary library was titered and amplified using the XL1 blue MRF' *E. coli* strain according to the manufacturer's recommended protocol (Agilent technologies). In brief: lysogeny broth (LB) supplemented with 10 mM MgSO<sub>4</sub> and maltose (0.02%; w/v) was inoculated with a single colony from a fresh overnight streak of XL1 blue MRF' cells that was grown on LB agar tetracycline (30 µg.mL<sup>-1</sup>) plates. The cells were then grown with rigorous shaking at 37°C until an OD<sub>600</sub> of 1.0 was reached, and spun down at 500 *xg* for 5 min. The supernatant was decanted and the cells gently resuspended in 10 mM MgSO<sub>4</sub> and diluted down to a final OD<sub>600</sub> of 0.5 (approx. half the original volume). Five sequential 2x serial dilutions of the primary library was prepared, of which 1 µL of each were incubated with 200 µL of the diluted XL1 Blue MRF' cells in a 37°C water bath for 15 min, so as to allow the phage to attach to the cells. Three millilitres of LB top agar (0.7%; w/v) at approx. 48°C were added to the cell preparation, briefly vortexed, and spread out on pre-warmed LB agar plates. The plates were incubated upside-down overnight at 37°C following which the plaques were counted to determine the titer in plaque forming units per millilitre (pfu/ml).

After the primary library was titered it was amplified so as to attain a stable high-titer stock of the library. In brief: a fresh culture of XL1 Blue MRF' cells in MgSO<sub>4</sub> (OD<sub>600</sub> = 0.5) was prepared as previously described. Aliquots of the primary library, each containing approx. 5x10<sup>4</sup> pfu of bacteriophage, were incubated with 600 µL of XL1 Blue MRF' cells at 37°C in a water bath for 15 min following which 6.6 mL of NZY top agar [0.5% NaCl (w/v); 0.2% MgSO<sub>4</sub> (w/v); 0.5% yeast extract (w/v); 1% NZ amine (w/v) 0.7% agar (w/v) at approx. 48°C] was added, briefly vortexed, and then spread out on 150 mm diameter freshly poured NZY agar plates. To amplify 1x10<sup>6</sup> plaques, a total of 20 aliquots (each containing 5x10<sup>4</sup> plaques/150 mm plate) needs to be plated. The inverted plates were incubated at 37°C for 8 hrs. The plaques should not be larger than 1-2 mm and should just be touching. Each plate was overlaid with 10 ml of SM buffer and gently rocked at 4°C on an orbital inclinor-shaker overnight, allowing the phage to diffuse into the SM buffer. The phage containing buffer was decanted and incubated with chloroform (5%; v/v) for 15 min at RT. Cellular debris were removed by centrifugation at 5,000 *xg* for 5 min. The

supernatant was recovered and chloroform added (0.3%; v/v) and stored at 4°C. For long term storage aliquots were stored with DMSO (7%; v/v) at -80°C.

The ZAP express vector is designed to allow simple efficient *in vivo* excision and recircularization, of the cloned DNA insert, into the pBK-CMV phagemid vector thereby effectively converting lambda phage DNA into a utilizable plasmid based vector [236, 239]. The protocol to achieve this is detailed in the following paragraph, whilst this paragraph aims to describe the process itself. This complex and ingenious operation of *in vivo* excision is dependent on the placement of the cloned DNA within the lambda phage, and the presence of a variety of proteins, including certain filamentous (M13) derived bacteriophage proteins. These M13 proteins recognizes two important regions of the 'origin of replication' on the lambda DNA, namely the (i) site of initiation of, and (ii) termination of DNA synthesis. The target lambda DNA (Lambda ZAP) is made accessible to the before mentioned M13 proteins by simultaneously co-infecting an *E. coli* strain with both the Lambda ZAP phage vector, and the M13 helper phage. Within the *E. coli* the M13 derived 'helper' proteins recognize the initiation site and then nicks one of the DNA strands and initiates rolling replication downstream of the 3' nicking site. DNA synthesis continues through the cloned DNA region, up until the the termination sequence downstream of the 3' nicking site is reached. The ssDNA is then circularized by the gene II product of the M13 helper phage, thus forming the complete single stranded pBK-CMV phagemid vector that contains the cloned DNA as well as all the necessary phagemid genes, including the f1 origin of replication. Signals for the packaging of the newly created phagemid is linked to the f1 origin sequence, resulting in the circularized ssDNA being packaged into phagemid particles and secreted by the *E. coli*. Following secretion, the *E. coli* used for *in vivo* excision can be killed by heat treatment at 70°C. The encapsulated phagemids aren't affected by the heat and can be utilized for downstream processes including single strand DNA rescue. The phagemids can also be used to infect *E. coli*, upon which it will be converted to its double stranded plasmid form, and can then be plated on selective media to form colonies. The plasmid DNA from the colonies can be extracted and used for analysis of the insert DNA including DNA sequencing, sub-cloning and restriction mapping.

After the primary library was amplified it was again titered as described previously and *in vivo* mass excision conducted according to the manufacturer's recommended protocol (Agilent technologies). The ExAssist helper phage in combination with the *E. coli* XL0LR strain were specifically designed to efficiently excise the pBK-CMV phagemid vector from the ZAP Express vector, while eliminating problems associated with helper phage co-infection. The ExAssist helper phage contains an amber mutation that prevents replication of the helper phage genome in a non-suppressing *E. coli* strain such as XL0LR. This allows only the excised phagemid to replicate in the host, removing the possibility of co-infection from the ExAssist helper phage. The protocol in brief are as follow: both XL1-Blue MRF' and XL0LR cells were grown till log phase, pelleted through centrifugation at 5,000 xg, and resuspended in

10 mM MgSO<sub>4</sub> at a final OD<sub>600</sub> of 1.0 (approx. 8x10<sup>8</sup> cells.mL<sup>-1</sup>). In a 50 mL conical tube bacteriophage from the amplified library was combined with XL1-Blue MRF' cells at a multiplicity of infection (MOI) of 1:10 lambda ZAP phage-to-cell ratio. A hundred fold more lambda ZAP phage was excised, than was the primary size of the library, so as to ensure proper statistical representation of the excised clones. To the same conical tube the ExAssist helper phage was added to a MOI of 10:1 helper phage-to-cell ratio, thus ensuring that every cell is co-infected with both lambda ZAP phage and helper phage. The tube was incubated at 37°C in a water bath for 15 min following which 20 mL of LB broth with supplements [10 mM MgSO<sub>4</sub>, 0.2% maltose (w/v)] was added, and then subsequently shaken at 37°C in an incubator for 2.5 hrs. The tube was heated to 70°C for 20 min to lyse the cells and then briefly cooled on ice. The cellular debris was pelleted by centrifugation at 10,000 *xg* for 10 min and the supernatant decanted to a sterile tube. To titer the phagemid, 1 µL of three sequential 5x serial dilutions of the phagemid preparation were combined with 200 µL of the previously prepared XL0LR cells in a microcentrifuge tube, and incubated at 37°C in a heating block for 15 min. Forty microliters of a 5x NZY broth were then added (final concentration 1x) and shaken at 37°C for 1 hrs to allow for the kanamycin resistance gene to be expressed before 100 µL was plated out on selective media (50 µg.mL<sup>-1</sup> kanamycin).

After the titer was accurately obtained and noted, a large scale transduction of the phagemid into the XL0LR cells was performed. Cells were plated out on multiple large LB agar square plates (28x28 cm) and care was taken to achieve at least 10x higher clonal representation as the original library size. The colony density was such that the colonies just touched, but weren't overgrown in a complete matt, allowing for approximately a hundred thousand colonies per large plate. After the overnight incubation at 37°C the plates were cooled at 4°C for several hours and the cells then scraped of with a scalpel blade and pooled, taking care to avoid excess agar.

The phagemid plasmids were then extracted from the cell mass by means of large scale alkaline lysis as described by Sambrook *et al.* (2001) to yield a final plasmid library that can be used later with the aid of conventional transformation methods, e.g. electroporation and chemical, to screen in either the constructed Δβ-gal mutant or *E.coli* strains [240]. The protocol in brief: After the cells were scraped, they were first washed by suspending them in 200 mL 0.1x TE (1 mM Tris-HCl, 0.1 mM EDTA; pH 8.0). The cell suspension was equally divided between 50 mL conical tubes to facilitate ease of use. The cells were pelleted at 10,000 *xg* and the supernatant decanted. The cell pellets were suspended in 5 mL ice-cold solution I [50 mM glucose; 25 mM Tris-HCl (pH 8.0), 10 mM EDTA] and vigorously vortexed. To this, 10 mL of freshly prepared solution II (0.2 N NaOH; 1% SDS) were added and gently inverted several times making sure all the surface area got covered. The tubes were stored on ice for 15 min. Seven and a half millilitres of solution III (3M KOAc, pH 6.0) were then added and the tubes gently inverted several times, and again stored on ice for 10 min. The tubes were centrifuged at 12,000 *xg* for 15 min and the supernatant carefully transferred to a new tube. The solution was the twice extracted with an equal

volume of phenol-chloroform : isoamyl alcohol (25:24:1), and once only with an equal volume of chloroform. The DNA was precipitated with 2 volumes of ethanol, briefly air dried and suspended in 200  $\mu$ L of TE (pH 8.0). The quality and quantity of the isolated DNA was ascertained by gel electrophoresis (1%; w/v) and spectrophotometric analysis.

### 3.2.3 $\beta$ -galactosidase mutant construction

The *E. coli* DH5 $\alpha$  strain employed for the screening of the metagenomic libraries seeking galactose containing oligosaccharides/polymers was mutated so as to contain no endogenous  $\beta$ -galactosidase sequence at all, through lambda red recombineering according to the methodology of Datsenko and Wanner (2000) [241]. To modify the target DNA, one needs to electroporate a linear donor dsDNA into *E. coli* expressing the  $\lambda$ -red enzymes. These  $\lambda$ -red enzymes then catalyse the homologous recombination of the donor DNA with that of the target DNA, based on their shared flanking homologous sequences, that was incorporated into the donor DNA through the polymerase chain reaction (PCR). Expressing the lambda red genes from a plasmid allows for a mobile recombineering system, but tight regulation of expression is required for a successful experiment, otherwise non specific recombination can occur. The pKD46-RecA (Nature Technology) plasmid utilized in this procedure carry the  $\lambda$ -red genes behind the *araBAD* promoter, which allows for rigid expression control. Expression of the  $\lambda$ -red genes is sufficiently induced by adding L-arabinose (0.1%; w/v) to the culture media. The plasmid itself is temperature-sensitive and can be cured from the strain by growing it at elevated temperatures above 30°C. By transforming any *E. coli* K-12 strain with pKD46-RecA it will be turned into a highly efficient recombineering strain when grown in the presence of arabinose.

The recombineering plasmid pKD46-RecA was transformed into chemical competent CaCl<sub>2</sub> DH5 $\alpha$  cells that was prepared according to the methods of Tu *et al.* (2005) [242], taking care though never to elevate the the culture temperature above 30°C, once the cells were transformed with the plasmid. For the chemical competent cell preparation 'Chemical Competent' (CC) solution (10 mM Pipes, 55 mM MnCl<sub>2</sub>, 15 mM CaCl<sub>2</sub>, 250 mM KCl) was prepared by mixing all components except for MnCl<sub>2</sub>, and adjusting the pH to 6.7 with KOH. Only then was the MnCl<sub>2</sub> added and the CC solution filter-sterilized through a pre-rinsed 0.45  $\mu$ m filter unit, and stored at 4°C until later use. For the preparation of the cells a fresh DH5 $\alpha$  colony from an overnight streak was inoculated into SOC broth [bacto-tryptone (2%; w/v), yeast extract (0.5%; w/v), 10 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 20 mM glucose], and incubated at 37°C with vigorous shaking till an OD<sub>600</sub> of 0.4 was reached. The cells were pelleted through centrifugation at 5,000  $xg$  for 5 min at 4°C, and resuspended in one-half volume of sterile ice-cold CC solution, and incubated on ice for 30 min. The cells were again pelleted through centrifugation at

5,000  $xg$  for 5 min at 4°C and resuspended in one-tenth the volume CC buffer. Cells can be stored at 4°C for up to 3 days, or sterile glycerol added (20%; w/v) and then aliquoted and stored at -80°C for later use. For the transformation with pKD46-RecA, 1  $\mu$ L of plasmid was added to 100  $\mu$ L competent cells together with 1  $\mu$ L DMSO, and incubated on ice for 30 min. The cells were heat shocked at 42°C for 90 s in a heating-block and placed back on ice for 2 min, after which 400  $\mu$ L of SOC medium was added and shaken in an incubator at 30°C for 1 hrs. The transformation mixture was spread on prewarmed SOC agar plates containing 100  $\mu$ g.mL<sup>-1</sup> ampicillin and incubated overnight at 30°C.

For the actual recombineering event electrocompetent cells need to be prepared from the previously generated DH5 $\alpha$  cells containing the pKD46-RecA plasmid. This was done according to the methodology of Sambrook *et al.* (2001) [240]. The protocol in brief: an overnight colony from a freshly streaked DH5 $\alpha$  (pKD46-RecA) grown on LB agar (100  $\mu$ g.mL<sup>-1</sup> ampicillin) at 30°C, was inoculated into selective SOC broth with arabinose [100  $\mu$ g.mL<sup>-1</sup> ampicillin; 0.1% arabinose (w/v)] into a baffled flask, and grown with vigorous shaking at 30°C till an OD<sub>600</sub> of 0.6 was reached. The flask was immediately submerged into an ice-water slurry and kept there for 10 min with occasional swirling. The cells were pelleted through centrifugation at 5,000  $xg$  for 5 min at 4°C, and then resuspended in one-seventh the original volume ice-cold sterile ddH<sub>2</sub>O. The cells were washed twice more with ice-cold sterile ddH<sub>2</sub>O and then reconstituted in one-twentieth the original volume of ice-cold ddH<sub>2</sub>O containing 10% glycerol. The cells were again pelleted as described above and resuspended in one-hundredth the original volume 10% glycerol. The cells were either immediately used for electroporation or stored at -80°C for later use.

Primers were designed using the CLC-genomics workbench software suite (Qiagen, Hilden, Germany) and synthesised by Inqaba Biotech (Pretoria, South Africa). Both the forward and reverse primer contained two important sequential sequence segments namely (i), on the 5' end of each primer a 70bp sequence homologous to the region adjacent the *E. coli*  $\beta$ -galactosidase sequence to be removed. And (ii), on the 3' of each primer a 23 bp sequence homologous to the flanking sequence of the chloramphenicol resistance gene on the pKD3 donor plasmid, that allows for amplification by PCR of the chloramphenicol resistance gene that will replace the  $\beta$ -galactosidase sequence on the *E. coli* genome in the actual recombination event (Table 3-1).

For the recombineering event, the linear dsDNA PCR product was cleaned using the PCR purification kit (Qiagen, Hilden, Germany) and electroporated (pulse controller = 200 $\Omega$ ; capacitor = 25  $\mu$ F; voltage = 2.5 kV) into the DH5 $\alpha$ (pKD46-RecA) cells with a Bio-Rad Gene Pulser Xcell electroporator (Bio-Rad Industries, USA), using disposable 1 mm cuvettes. Warm 30°C SOC broth with arabinose (0.1%; w/v) were immediately added to the cuvette, before being transferred to a conical tube and incubated in an incubator-shaker at 30°C for 1 hour with vigorous shaking before being plated-out on selective LB agar plates (10  $\mu$ g.ml<sup>-1</sup> chloramphenicol and 50  $\mu$ g.mL<sup>-1</sup> ampicillin) and grown for two days

at 30°C. Positive colonies were tested by PCR for the presence of  $\beta$ -galactosidase utilizing primers internal to the sequence that was homogenously removed. The temperature sensitive pKD46 plasmid was cured by inoculating a single colony into liquid culture and incubating overnight at 40°C. Cells were streaked out on selective plates (10  $\mu\text{g}\cdot\text{ml}^{-1}$  chloramphenicol) and subsequently negatively tested by streaking out individual colonies on plates containing 50  $\mu\text{g}\cdot\text{mL}^{-1}$  ampicillin plates. Cells that did not grow were taken to have dropped the plasmid. The  $\Delta\beta$ -gal mutant was further assayed for the complete abolition of any endogenous  $\beta$ -galactosidase activity, even through  $\alpha$ -complementation, by transforming the pBluescript (sk-) plasmid into the newly generated mutant and testing on X-Gal. Once it was confirmed that  $\Delta\beta$ -gal was a complete null-mutant for any endogenous  $\beta$ -galactosidase activity, and it did not contain the pKD46-RecA recombineering plasmid anymore, both chemical competent and electrocompetent cells were prepared as described previously and stored at -80°C for later use.

**Table 3-1** Primers used for the LacZ lambda red recombination. The underlined blue text is the annealing sites for the amplification of the chloramphenicol resistance gene on the pKD3 donor plasmid, whilst the preceding sequence is the  $\beta$ -galactosidase recombineering recognition sequence. The lacZ-frw and lacZ-rev are internal  $\beta$ -galactosidase internal annealing sites to pcr test for the presence of  $\beta$ -galactosidase.

Primer name	Sequence	Tm (°C)	Binding site	NCBI ACNO
LzF Recomb	ATCCCATACGGTCAATCCGCCGTTTGTCCACGGAGAATCCGACGGGT <u>gtgtagcctggagctgcttc</u>	58.1	317-366	NC_000913.3
LzR Recomb	CTTCCAGCGTTCGACCCAGGCGTTAGGGTCAATGCGGGTCGCTTCACTTA <u>atgggaattagccatggctc</u>	54.8	2387-2436	NC_000913.3
lacZ-frw	TCGATGAGCGTGGTGGTT	56.8	839-856	NC_000913.3
lacZ-rev	GCTGCTGGTGTGTTTGCTT	54.3	1859-1876	NC_000913.3

### 3.2.4 Library screening and media used

For the screening of the metagenomic libraries, the *in vivo* mass excised pBK-CMV plasmid libraries were electroporated into either the  $\Delta\beta$ -gal DH5 $\alpha$  mutant, or normal DH5 $\alpha$  electrocompetent cells. The before mentioned choice was dependant on whether screening for  $\beta$ -galactosidases or fructosyltransferases. Technically speaking all of the screening could have been conducted in the  $\Delta\beta$ -gal mutant as the removal of the endogenous  $\beta$ -galactosidase activity should have no bearing on screening for fructosyltransferase activity, but so as to avert any imperceivable consequence this protocol was adhered to, even though the libraries were cross-screened for the different activities.

The pBK-CMV plasmid preparation was desalted prior to electroporation with a Amicon Ultra 0.5 mL centrifugal filter according to the manufacturer's recommendations (Merck-Millipore, Massachusetts, USA). After electroporation (pulse controller = 200  $\Omega$ ; capacitor = 25  $\mu\text{F}$ ; voltage = 2.5 kV) 1 mL

prewarmed SOC broth was immediately added to the cuvette before being transferred to a conical tube and incubated in an incubator-shaker for 40 min at 37°C with vigorous shaking. Cells were immediately placed on ice for several minutes to halt growth before being pelleted at 5,000  $xg$  for 5 min at 4°C, and then resuspended in an equivolume minimal MDA-5052 media (see paragraph below). The cells were titered on normal LB agar selective plates (50  $\mu g \cdot mL^{-1}$  kanamycin) so as to establish a representative matt plate-out titre where individual colonies can barely be distinguished from the lawn of cells. The cells suspension can be kept at 4°C for 10 days without the loss of cell viability or the titre being affected.

All minimal and auto-induction media used for library screening and heterologous protein expression, were prepared according to methodology of Studier *et al.* (2014) [243]. Several stock solutions were made from which all final media preparations were composed. For convenience all of the preparation of the stock solutions employed in the Studier methodology are described here. All stock solutions were autoclaved and stored at RT unless specified otherwise. In brief:

- 50 x M: 80 ml water, 17.75 g  $Na_2HPO_4$ , 17 g  $KH_2PO_4$ , 13.4 g  $NH_4Cl$ , 3.55 g  $Na_2SO_4$ .  
1x concentration: 25 mM  $Na_2HPO_4$ , 25 mM  $KH_2PO_4$ , 50 mM  $NH_4Cl$ , 5 mM  $Na_2SO_4$ , (pH ~6.7).
- 50 x 5052: 25 g glycerol (weigh in beaker), 73 ml water, 2.5 g glucose, 10 g  $\alpha$ -lactose monohydrate. 1x concentration: 0.5% glycerol, 0.05% glucose, 0.2%  $\alpha$ -lactose.
- 25% aspartate: 84 ml water, 25 g aspartic acid, 8 g NaOH (pH ~7).
- 50% lactose: 500 g lactose, warm ddH<sub>2</sub>O to 1 L. Stirred until dissolved with occasional heating. Either filter sterilised or autoclave. Can be reheated if crystallisation occurs, but not once added to final media composition.
- 30% sucrose: 300 g sucrose, warm ddH<sub>2</sub>O to 1 L. Stirred until dissolved with occasional heating. Either filter sterilised or autoclave. Can be reheated if crystallisation occurs, but not once added to final media composition.
- 40% glucose: 40 g glucose, ddH<sub>2</sub>O to 100 mL.
- 1M MgSO<sub>4</sub>: 24.65 g  $MgSO_4 \cdot 7H_2O$  ddH<sub>2</sub>O to 100 mL.
- 1000x metals: 50 mM  $FeCl_3$ , 20 mM  $CaCl_2$ , 10 mM  $MnCl_2$ , 10 mM  $ZnSO_4$ , 2 mM  $CoCl_2$ , 2 mM  $CuCl_2$ , 2 mM  $NiCl_2$ , 2 mM  $Na_2MoO_4$ , 2 mM  $Na_2SeO_3$ , 2 mM  $H_3BO_3$ . The trace metal mix was assembled from autoclaved stock solutions of the individual components except for  $FeCl_3$ , which was added from the 0.1 M acid solution (see immediately below), do not autoclave once added together.
- 0.1 M  $FeCl_3$ : in ~0.12 M HCl: 99 ml water, 1 ml concentrated HCl (~12 M), 2.7 g  $FeCl_3 \cdot 6 H_2O$ , do not autoclave.
- ZY: 1 litre ddH<sub>2</sub>O, 10 g N-Z amine (casein hydrolysate), 5 g yeast extract.
- X-gal: 200 mg in 10 mL DMSO, aliquot and store at -20°C.

- Kanamycin: 1 g kanamycin sulphate in 10 mL ddH<sub>2</sub>O, aliquot and store at -20°C.

For the final preparation of the minimal media agar plates either containing 30% lactose (MDA-lac; w/v) or 20% sucrose (MDA-suc; w/v), 17 g agar was autoclaved in approx. 250 mL ddH<sub>2</sub>O and allowed to cool to ~60°C, following which, components of the sterile pre-prepared stock solutions were then added. Below are the final compositions of the solid and liquid media used in subsequent experiments. A more in-depth review of the auto-induction methodology will be conveyed in chapter five.

- MDA-lac defined minimal selective medium: 2 mL 1 M MgSO<sub>4</sub>, 200 µl 1000x metals, 4 mL 25% aspartate, 20 mL 50xM, 600 mL 50% lactose, 2 mL X-gal, 1 mL kanamycin, autoclaved ddH<sub>2</sub>O to 1 L. Final composition: 25 mM Na<sub>2</sub>HPO<sub>4</sub>, 25 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NH<sub>4</sub>Cl, 5 mM Na<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.2x metals, 0.1% aspartate, 30% lactose (w/v), 40 µg.mL<sup>-1</sup> X-gal, 100 µg.mL<sup>-1</sup> kanamycin.
- MDA-suc defined minimal selective medium: 2 mL 1 M MgSO<sub>4</sub>, 200 µl 1000x metals, 4 mL 25% aspartate, 20 mL 50xM, 667 mL 30% sucrose, 1 mL kanamycin, autoclaved ddH<sub>2</sub>O to 1 L. Final composition: 25 mM Na<sub>2</sub>HPO<sub>4</sub>, 25 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NH<sub>4</sub>Cl, 5 mM Na<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.2x metals, 0.1% aspartate, 20% sucrose (w/v), 40 µg.mL<sup>-1</sup> X-gal, 100 µg.mL<sup>-1</sup> kanamycin.
- MDA-5052 defined minimal media 2 mL 1 M MgSO<sub>4</sub>, 200 µl 1000x metals, 4 mL 25% aspartate, 20 mL 50xM, 20 mL 50x 5052, 2 mL X-gal, 1 mL kanamycin, autoclaved ddH<sub>2</sub>O to 1 L. Final composition: 25 mM Na<sub>2</sub>HPO<sub>4</sub>, 25 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NH<sub>4</sub>Cl, 5 mM Na<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.2x metals, 0.1% aspartate, 0.5% glycerol, 0.05% glucose.
- ZYM-5052 complex auto-inducing medium: 957 mL ZY, 2 mL 1 M MgSO<sub>4</sub>, 200 µl 1000x metals, 20 mL 50 x 5052, 20 mL 50 x M. Final composition: 1% N-Z-amine, 0.5% yeast extract, 25 mM Na<sub>2</sub>HPO<sub>4</sub>, 25 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NH<sub>4</sub>Cl, 5 mM Na<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.2x metals, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose.

The minimal selective agar media with their respective carbon sources were poured in large square plates (28 x28 cm) and allowed to cool. The previously prepared titered cells were plated out at the desired titre and the plates were incubated at 37°C for several days. The experiment was repeated several times and incubated at varying temperatures (4°C, 10°C and 20°C) once the colonies were formed, to screen for differential activity. All positive colonies that either presented a blue colour for X-gal hydrolysis or presented a visible phenotype in terms of exopolysaccharide production, were isolated and the plasmid extracted by alkaline lysis as described in section 3.2.2, and re-transformed into either the Δβ-gal mutant or DH5α so as to confirm the phenotype.

### 3.2.5 Sequencing and bioinformatic analysis

In order to de-replicate the multiple positive clones that were found for each library screen, a restriction mapping approach was firstly employed, after which a sequencing based approach was used. Several restriction enzymes that were mutually compatible in their activity buffers were employed, in various combinations. The restriction analysis were done with the restriction map and profile of the pBK-CMV vector taken into consideration (Appendix A1 on page 115). Based on the restriction profiles, the clones were grouped together in restriction classes. A single plasmid from each class was selected for standard Sanger capillary sequencing, using a commercial service (Macrogen, Amsterdam, Netherlands). All sequencing reactions were repeated at least three times, so as to reduce the impact of sequencing misreads and avoid assembly conflicts. For continued internal sequencing, primers were designed for primer walking using the CLC genomic workbench software package (Qiagen, Hilden, Germany). Continuous mass assembly analysis after each new round of sequencing revealed whether the different clones, from the different restriction classes, belonged in fact to the same contig assembly after which and if so, results were pooled and sequencing continued until complete end-to-end sequencing of the combined contig was achieved. Following assembly, the multiple ORFs in each assembled contig were identified with the CLC genomics workbench software and the deduced nucleotide and translated protein sequences were analysed *in silico* with the BLASTX and BLASTP algorithms respectively, on the NCBI webpage [244].

### 3.2.6 Cloning, expression and protein analysis

After *in silico* identification of putative ORFs and bioinformatic analysis, primers were designed to amplify the relevant genes, so as to enable cloning into protein expression vectors. Primers included RE sites compatible with that of the MCS of the pRSET-A vector (ThermoFisher Scientific) (Appendix A2 on page 116), so as to facilitate in frame and directional cloning with the 6x histidine leader sequence. PCR products were restriction digested with selected RE's and cloned into dephosphorylated pRSET-A that was also cut with the same REs. In-frame cloning and PCR specificity was confirmed through Sanger sequencing. In the case of putative  $\beta$ -galactosidases, the respective protein expression vectors were transformed back into the  $\Delta\beta$ -gal mutant and expressed overnight in auto-induction medium according to the methodology of Studier *et al.* (2014) [243] (see section 3.2.4). Putative fructosyl transferases were transformed into the Bl21 pLysS *E.coli* protein expression strain and expressed overnight in ZYM-5052 auto-induction medium (see section 3.2.4). dialysed overnight against 0.1x TE (1 mM Tris-HCl,

0.1 mM EDTA; pH 8.0) with snakeskin dialysis tubing (Thermo Scientific) with a 3.5K molecular weight cut off (MWCO).

The resulting crude protein extract was used for all subsequent experiments in the rudimentary characterisation of the respective catalyst and their products. Protein extract from cells transformed with empty pRSET-A and pRSET-A containing the *E. coli*  $\beta$ -galactosidase were used as negative and positive controls respectively. Proteins were quantified spectrophotometrically and electrophoretically separated by denaturing polyacrylamide gel electrophoresis using the methodology of Bradford *et al.* (1976) and Laemmli *et al.* (1970) [245, 246].

### 3.2.7 Nitrophenyl based biochemical characterisation of positive clones

A standard linear curve with 4-nitrophenol ranging between 0.0025  $\mu$ M and 0.04  $\mu$ M was established spectrophotometrically and measured at 405 nm on a VersaMax ELISA microplate reader (Molecular Devices). To test for substrate specificity 10 mM of various nitrophenyl based substrates dissolved in Z-buffer (20 mM phosphate buffer pH 7.0, 5 mM  $MgCl_2$ ) were incubated with 1  $\mu$ g total protein of the respective cell extracts, and activity was measured continuously at 405 nm for a period of 30 min at 37°C. All measurements were done in triplicate. The nitrophenyl based substrates utilized for substrate specificity were the following : ONP- $\alpha$ -D-galactopyranoside, ONP- $\beta$ -D-galactopyranoside, ONP- $\beta$ -D-maltoside, ONP- $\beta$ -D-xylopyranoside, ONP- $\beta$ -D-glucuronide, ONP- $\alpha$ -D-glucopyranoside, ONP- $\beta$ -D-glucopyranoside, ONP- $\beta$ -D-fucopyranoside, ONP- $\beta$ -L-arabinopyranoside, ONP- $\alpha$ -L-rhamnopyranoside, ONP- $\beta$ -D-cellobioside. All nitrophenyl based substrates were either obtained from Sigma-Aldrich (Missouri, USA), or Carbosynth (Berkshire, UK).

The substrate that showed the highest relative activity for each of the respective protein extracts were further utilized in subsequent experiments for the characterisation of the temperature (1°C - 55°C) and pH optima (pH 5.2 - pH 10.8). The biological buffers used for the pH characterisation were, acetic acid (pH 5.2), sodium phosphate (pH 6.4, pH 7.0, pH 7.6), Bis-Tris-propane (pH 8.2, pH 9.1) and sodium carbonate (pH 10.8). For temperature measurements where the continuous spectrophotometric assay method could not be employed, the reactions were stopped by heating at 80°C for 10 min followed by the addition of an equal volume of 1M  $Na_2CO_3$ , after which an endpoint measurement was taken at wavelength 405 nm.

### 3.2.8 Product characterisation by spectrophotometric analysis of free reducing sugars

After the establishment of substrate specificity together with that of the temperature and pH optima, the protein extracts were incubated with either sucrose or lactose as substrate (30%; w/v), under their respective optimal conditions overnight in a rotary shaker. The extracts were incubated at a concentration of  $1 \text{ u.mL}^{-1}$  in a total volume of 10 mL. One unit of activity is defined as the amount of total crude protein extract needed to liberate  $1 \text{ }\mu\text{mole}$  of free reducible glucose from the disaccharidic substrate in 1 minute. Following the overnight incubation the samples were twice extracted with an equal volume of phenol-chloroform : isoamyl alcohol (25:24:1), and once only with an equal volume of chloroform, after which the reducing sugars were assayed. The results were expressed as the ratio between the amount of the free reducible monosaccharide constituents of each substrate, i.e. glucose : galactose for the lactose substrate, and glucose : fructose for sucrose substrate. This is to give a representation of the hydrolytic reaction versus that of the trans-glycosylation reaction of galactose and fructose. Free reducing glucose, fructose and galactose were determined using an enzyme linked assay based on the reduction of  $\text{NAD}^+$  to NADH at 340nm on a VersaMax ELISA microplate reader (Molecular Devices). Hexokinase/glucose-6-phosphate dehydrogenase (HK-G6DH) was coupled to the measurement of free glucose, while fructose was measured with HK-G6DH in conjunction with phosphoglucose isomerase (PGI). Galactose-dehydrogenase/mutatarose (GAL-DH/M) coupled with HK-G6DH was used for the measurement of free galactose (see Fig 3-7 under results and discussion for a visual representation).

The combined protocol in brief: to each well of a clear-bottom UV transparent microtiter plate,  $200 \text{ }\mu\text{L}$  of either Buffer<sub>Gluc/Fruc</sub> (100 mM Tris-HCl pH 7.4; 5 mM  $\text{MgCl}_2$ ; 1 mM  $\text{NAD}^+$ ; 1 mM ATP), or Buffer<sub>Gluc/Gal</sub> (100 mM Tris-HCl pH8.5, 1 mM  $\text{NAD}^+$ ; 1 mM ATP) was added, taking care not to introduce any air bubbles. The respective samples were then added to each well in a total volume of  $50 \text{ }\mu\text{L}$ , making the final volume a combined  $250 \text{ }\mu\text{L}$ . A continuous assay was then established on the spectrophotometer so as to determine when baseline was reached, after which an endpoint background measurement was taken of which the value was designated  $G_0$ . For the measurement of glucose  $10 \text{ }\mu\text{L}$  of a 40x dilution of a HK-G6DH enzyme preparation from Megazyme was added to each well (approx. 1 U hexokinase and 0.5 U G6-DH per well). Again a continuous assay was established on the spectrophotometer and an endpoint measurement taken when baseline was reached when all the free glucose was reduced (approx. 20 min). This measurement was designated  $G_{gluc}$ . Depending on which substrate was originally assayed,  $10 \text{ }\mu\text{L}$  of a 50x dilution of either GAL-DH/M (for the measurement of galactose), or PGI (for the measurement of fructose) was added to each well (approx. 1 U PGI or GAL-DH per well). A final measurement was taken after baseline was established and the final result designated either  $G_{gal}$  or  $G_{fruc}$  (approx. 20 min). A standard curve was established for the NADH coupled reduction of free glucose on

the spectrophotometer in a range of 0.02  $\mu\text{mole}$  to 0.2  $\mu\text{mole}$  per well (10  $\mu\text{L}$  of 1.7 mM - 20 mM glucose stock solution). Because glucose, galactose and fructose all occur at equimolar ratios within the lactose and sucrose substrates, the standard curve of glucose can be extrapolated for the measurement of free galactose and fructose as well. To calculate free glucose,  $G_{gluc} - G_0$  gives the change in absorbance due to the reduction of glucose, whilst  $G_{gal} - (G_{gluc} - G_0)$  and  $G_{ fruc} - (G_{gluc} - G_0)$  gives the change in absorbance due to the reduction of free galactose and fructose respectively. The actual amounts can be deduced from the standard curve.

### 3.2.9 Product characterisation by thin layer chromatography

For a visual characterisation of the products formed through the enzymatic reactions, when the crude extracts were incubated with substrates, thin layer chromatography was employed. The samples were prepared the same as described above in section 3.2.8, only the resulting supernatant was diluted 10 fold with ddH<sub>2</sub>O. Included in the TLC is the commercially available GOS preparation, Vivinal GOS from Friesland Campina (Amersfoort, The Netherlands). Thin layer chromatography was conducted according to the protocol of Wang *et al.* (2014) [247]. Of each sample 0.5  $\mu\text{L}$  was evenly spotted on a pencilled horizontal straight line, on aluminium backed silica gel 60 TLC plates (Merck, Germany). The plates were briefly dried in an oven, before being run in a sealed glass chamber that was equilibrated for an hour with the mobile phase, consisting of butanol/ethanol/water (5:3:2). The plates were run until the mobile phase reached  $\frac{3}{4}$  the way up. Again the plates were oven dried before being developed by spray misting a solution containing 0.5% (w/v) 3,5-dihydroxytoluene (orcinol) (Sigma-Aldrich, USA) and 20% (v/v) sulphuric acid, onto them with the aid of a gravity fed pressurised air micro-spray gun. The silica plates were clamped between two glass plates and baked in the oven for a several minutes at 100°C until product development could be visualised, after which they were taken out and photographed.

## 3.3 Results and discussion

### 3.3.1 Library construction and screening in the mutant

For the purposes of this study altogether three libraries were constructed and screened. One previously constructed library (Hotwood), which was established in our laboratory for the screening of novel  $\alpha$ -glucosidases, was also utilised in this study and cross-screened for novel  $\beta$ -galactosidases. The three

newly constructed libraries were designated 'MS' (place of milk-runoff into soil; high lactose environment), 'SASRI' (South African Sugarcane Research Institute library, from factory runoff at sugarcane refinery; high sucrose environment) and 'Gansbay' (place of intertidal coastal kelp degradation; high galactose polymer environment).

Primarily the location of the sampling sites were chosen based on the prevalence of galactose or fructose containing substrates e.g. lactose, sucrose and galactose rich polysaccharides. Because these substrates contain the monosaccharide constituents galactose or fructose, they can be hydrolysed and their moieties potentially incorporated into polymers/oligosaccharides through trans-glycosylation. The hypothesis is that the occurrence and prevalence of the before mentioned substrates in the sampling sites will create a niche environment, that would select for and enhance the probability for the incidence of the desired enzymatic biocatalyst, that synthesise galactose or fructose containing oligosaccharides/polymers.

Secondary to the substrate occurrence, other factors were taken into consideration, or at the very least noted, that could influence or have bearing on the characteristics of the isolated enzymes. These include factors like environmental pH and temperature. The Hotwood library, constructed from wood pulp process runoff at a sawmill, represents a thermophilic sampling site rich in glycosidic linkages. No other environmental data, other than the fact that the sampling site was hot and humid by nature, were noted for the Hotwood library. Figure 3-1 shows the sites that were sampled for the construction of the MS (Fig 3-1A) and Gansbay (Fig. 3-1B) libraries. Several sub-sites were sampled at each main-site, so as to increase the diversity of the sampled microbiome. Each sub-site was initially sampled individually and their pH's measured so as to make sure they don't differ significantly, before each main-site's samples were amalgamated and their collective DNA isolated. The sample for the SASRI library was obtained from a single bucket, placed under the conveyer belt of the sugarcane refinery. Both the MS and SASRI samples were slightly acidic (pH 6.4-6.8), with none of their respective sub-sites differing markedly. This is expected due to the presence of lactic acid in milk, which acts as a proton donor, and also due to the high probability of anaerobic fermentation of both sucrose and lactose, resulting in an acidic environment. The pH for the sampling site at Gansbay was near neutral. The sub-sites at the Gansbay sampling site contained sludge from stagnant pools that could potentially differ markedly in temperature during day and night. This is due to the sun heating the shallow pools to extreme temperatures during the daytime, which then cools rapidly during the night. At the time of the Gansbay sampling, the temperature was moderate but at the very least this site is potentially an extreme environment due to the occurrence of coastal storms and fluctuating temperatures. No temperature was noted for the SASRI sample at collection, but this too is representative of a hot and humid environment due to it being in a refinery at a sub-tropical location.



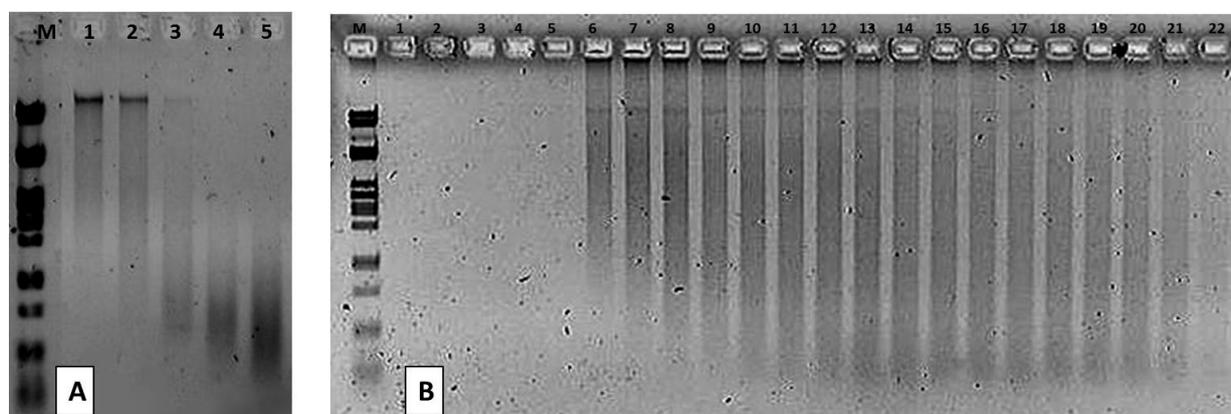
**Figure 3-1** Visual representation of two of the five library sampling sites i.e. the MS library (A1 and A2) and the Gansbay intertidal kelp degradation library (B). Several sub-samples were collected (red arrows) from each sampling site, and were eventually amalgamated into one, from which DNA were extracted. This was done so as to achieve broader genetic representation. The pH was measured for each individual sub-sample but none differed markedly from each other within each sampling zone.

DNA isolated from environmental sources for metagenomic purposes are often degraded and contaminated with polyphenolics, polysaccharides and various other contaminants, rendering them unusable for downstream application [248, 249]. For this purpose, great care was employed to design and optimise the protocol for DNA isolation so as to ensure that highly representative, intact DNA of high purity was obtained. The DNA that was isolated from the respective metagenomic sources, by the modified and combined protocols of Verma *et al.* (2011) and Dos Reis Falcão *et al.* (2008) [237, 238], were demonstrated to be of high quality and purity with little degradation and polysaccharide contamination as shown by the 260/280 spectrophotometric ratios and separation by agarose gel electrophoresis. All the DNA isolated, readily digested with restriction endonucleases and was used successfully as template in PCR (results not shown).

The protocol of Verma *et al.* (2011) [237] focuses on the removal of charged and small contaminants by incubation with PVPP and activated charcoal. This is especially true for polyphenolics and their quinone oxidation products, that intercalates with DNA, causing it to break-up with freeze-thaw cycles as well as obstructing enzymatic action [250].

The protocol of Dos Reis Falcão *et al.* (2008) [238] was originally developed for the extraction of high quality RNA from polysaccharide rich brown algae, and focuses on the selective precipitation of high molecular weight contaminants (such as polysaccharides) by potassium acetate. The sites sampled all had major polysaccharide constituents in their respective environments, whether it be from plant/algae breakdown, or visible biofilm formation. Initial attempts failed to isolate adequately pure DNA, and only once the potassium acetate selective precipitation step was included, was high purity DNA obtained. The guanidinium hydrochloride also included in the protocol, is probably the best known protein denaturant and instantly denatures any environmental nucleases present in solution, thereby shielding the nucleic acids from proteolytic degradation.

The metagenomic libraries were constructed with the Lambda ZAP express pre-digested vector kit according to the manufacturer's recommendations (Agilent Technologies). Fig 3-3A show the results of the partial digestion trial, of the MS metagenomic DNA sample that was previously isolated, with the *Bsp143I* (aka *Sau3AI*) restriction enzyme. This trial was conducted so as to establish the optimal nucleic acid and RE concentration for the desired length of incubation time. The fractions visualised in lanes 1 and 2 contained the majority of their restriction product as fragments larger than 2 kb, and subsequently an intermediate RE concentration was chosen for the large scale partial digestion. A total of 100 µg of DNA was then digested at a concentration of 0.333 mg.mL<sup>-1</sup>, with 10 U.mL<sup>-1</sup> *Bsp143I*, for 30 min. Fig 3-3B shows the CL-2B column fractionation of the partially digested DNA. In order to reduce cloning bias towards shorter fragments and interfering nucleotides, only lanes 6 and 7, which showed a marked reduction in shorter fragments, were pooled and utilized for the subsequent MS-library construction. The same partial digestion trial and column fractionation procedure was repeated for the construction of the other two libraries. The fractionated and pooled DNA was ligated into the lambda ZAP express bacteriophage vector [236, 239]. The *Bsp143I* used to partially digest the metagenomic DNA recognizes the four base pair (bp) sequence 5'...AT<sup>+</sup>CGAT...3' and generates compatible ends with the *Bam*HI restriction sites of the lambda ZAP express vector.



**Figure 3-2** The partial digestion and column fractionation of the soil extracted metagenomic DNA that was used for the construction of the LacZ library. In Fig A lanes (1-5) is representative of the digestion of 5µg of DNA with successive 2x dilutions of 0.2 units *Sau3AI* per µl in a 15µl reaction volume, in 1x reaction buffer for 30min. An enzyme concentration between lanes 1 and 2 were chosen for up-scaled digestion. In Fig B lanes (1-22) is representative of the Sepharose cl-2B column fractionation of the up-scaled digestion. Lanes 6 and 7 were pooled and used for library construction.

Good titres were obtained for all libraries constructed ranging between  $5 \times 10^6$  and  $1 \times 10^8$  pfu/µg Lambda DNA, with the exception of the Gansbay library ( $3 \times 10^4$  cfu/µg), which relatively had a lower titer than the other libraries (see Table 3-2). The reason for the lower titer could be multifaceted, ranging from less pure DNA to faulty ligation, to name but two possible explanations. The construction of

lambda phage libraries will always be a hit and miss procedure as a multitude of factors eventually influence success, and good titres are never guaranteed. After amplification of the primary libraries, the bacteriophage vectors were converted to plasmids by *in vivo* mass excision according to the manufacture's recommendations and as described previously, yielding usable plasmid libraries that can be electroporated into the desired *E. coli* strain for screening [239].

**Table 3-2** Lists the libraries constructed and carbon source screened on, average insert size determined as per restriction digestion as well as the number of putative genes discovered through functional screening and subsequent sequencing.

	<i>library name</i>	<i>metagenome source</i>	<i>carbon source</i>	<i>average insert size</i>	<i>library size</i>	<i>putative genes</i>
1	<b>MS</b>	soil milk runoff	Lactose	4.5 kb	5 x 10 <sup>6</sup>	6
2	<b>SASRI</b>	runoff at sugar refinery	lactose and sucrose	3.8 kb	1 x 10 <sup>8</sup>	6
3	<b>Gansbay</b>	coastal kelp degradation	Lactose	4.1 kb	3 x 10 <sup>4</sup>	1
4	<b>Hotwood</b>	wood mill high temperature	Lactose	3.6 kb	unknown	4

The  $\Delta\beta$ -gal mutant construction was efficacious in that all endogenous  $\beta$ -galactosidase activity was abolished. No activity was observed with  $\alpha$ -complementation when the empty pBluescript (sk-) plasmid was transformed into the  $\Delta\beta$ -gal mutant and screened on X-gal, nor could any pcr product be detected for primers complementary to the internal  $\beta$ -galactosidase sequence that was homologously removed through the  $\lambda$ -Red recombineering event. Only when the  $\Delta\beta$ -Gal mutant was transformed with the pRSET vector containing the complete coding sequence for the *E. coli*  $\beta$ -galactosidase, was the indigo blue phenotype restored when plated out on X-gal containing plates (se Fig 3-3, the K- and K+ controls).

The *in vivo* mass excised plasmid libraries were used to screen for  $\beta$ -galactosidase activity in the  $\Delta\beta$ -gal mutant on MDA-lac minimal media (30% lactose; w/v). All four libraries were screened for the presence of novel  $\beta$ -galactosidases but only the SASRI-library was screened for novel fructosyl-transferases on MDA-suc minimal media plates (30% sucrose; w/v). This approach was adopted due to the ubiquitous nature of the  $\beta$ -galactosidase enzyme. When screening for fructosyl-transferases the normal unmutated DH5 $\alpha$  strain was used.

The various libraries were screened for different time intervals and at different temperatures, resulting in the cumulative discovery of several hundred positive clones. The Positive clones were identified either through presentation of a blue colour when grown on X-Gal in the case of  $\beta$ -galactosidase screen (MDA-lac), or by the visible production of exopolysaccharide for either screen (MDA-lac + MDA-suc).

### 3.3.2 Sequencing and bioinformatic analysis

The tedious task of de-replicating the positive clones was undertaken firstly, by restriction analysis of the alkaline lysis isolated plasmid preparations, and secondly by sequencing the plasmid RE classes. RE classification reduced the initial positive clone count from several hundred to just 48. As the sequencing progressed, mass sequence analysis and assembly proved that several of the clones that at first appeared distinct through the initial restriction analysis, in fact belonged to the same continuous contigs, and as a consequence were grouped together and assembled as one from that point forward. This reduced the total number of positive contigs to 13 in total, with three originating from the SASRI library, five from the Hotwood library, four from the MS library and one from the Gansbay library. Of these 13 contigs, 12 were obtained from the screening for novel  $\beta$ -galactosidase activity and identified based on the production of a blue phenotype when screened on MDA-lac supplemented with X-gal. One contig was from the trans-fructosylation screen and was identified through the visible production of an exopolysaccharide.

The 13 contigs were sequenced end to end and analysed *in silico*. The ORFs were identified and analysed with the BLASTX and BLASTP protocols on the NCBI website which, revealed that several of the contigs contained more than one putative gene that could potentially account for the observed phenotype (Table 3-3). In all, 17 putative genes were identified, that could potentially either catalyse the production of galactose (16 in total) or fructose (1 in total) containing oligosaccharides/polymers (Table 3-3; See appendix for sequence information).

The Carbohydrate-Active Enzymes database (CAZy; <http://www.cazy.org>) provides a sequence-based family classification system, linking the sequence to the specificity of enzymes that synthesise, modify or hydrolyse oligo- and/or polysaccharides [113]. The putative genes isolated, and the protein products they encode, belong to a range of GH families.  $\beta$ -Glucosidases (EC 3.2.1.21) and  $\beta$ -galactosidases (EC 3.2.1.23) catalyse the selective cleavage of glycosidic bonds from the non-reducing ends of their substrates and some of them have been shown to have trans-glycosylation activity. Moreover some  $\beta$ -glucosidases have been shown to be promiscuous with regards to substrate specificity and not only hydrolyse lactose, but also catalyse trans-galactosylation [212]. It is therefore not surprising that a number of GH1 family members, which are predominantly  $\beta$ -glucosidases, have been identified through functional screen designed primarily to identify  $\beta$ -galactosidase activity.

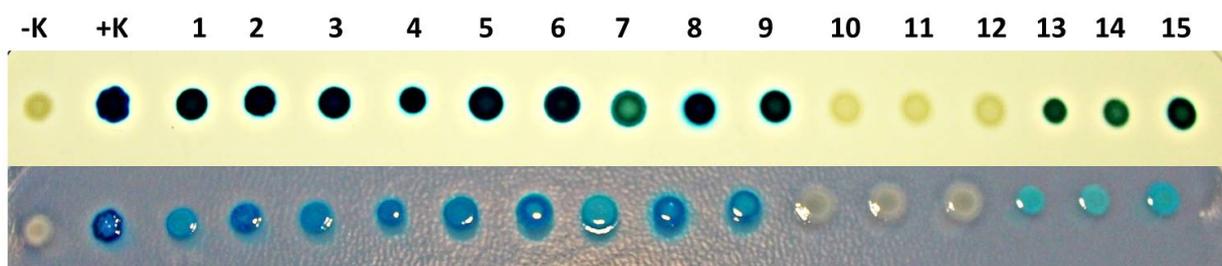
For the  $\beta$ -galactosidase screen, the principal GH families isolated belonged to families 2 and 42, families which are known to produce GOS [23]. Glycoside hydrolase family 2 comprises enzymes with several known activities i.e.  $\beta$ -galactosidase (EC 3.2.1.23);  $\beta$ -mannosidase (EC 3.2.1.25);  $\beta$ -glucuronidase (EC 3.2.1.31) [251]. These enzymes contain a conserved glutamic acid residue which has been shown in

*E. coli*  $\beta$ -galactosidase, to be the general acid/base catalyst in the active site of the enzyme [251]. GH42 are known  $\beta$ -galactosidases (EC 3.2.1.23), however, other commonly found activities are  $\alpha$ -L-arabinosidase (EC 3.2.1.55) and  $\beta$ -D-fucosidase (EC 3.2.1.38) (see Fig. 3-6; 8A-10A). Three GH53  $\beta$ -galactosidases were also isolated, all of which were in combined contigs, but only SAS- $\beta$ gal3 could be associated with the blue phenotype of x-gal hydrolysis. The only known activity of GH53 family proteins are  $\beta$ -1,4-galactanase activity and they are usual related to microbial degradation of galactans and arabinogalactans in the pectic component of plant cell walls [252].

**Table 3-3** The 17 putative genes that were identified and amplified from four metagenomic libraries. The table shows which glycosyl hydrolase family each putative gene product belongs to, based on *in silico* analysis as well as the nucleotide length of the gene and the predicted molecular weight of the protein. It also shows the closest related sequence based on BlastP analysis.

#	name	family	Xgal	Contig #	size	aa	Mw	%	similarity
1	HW- $\beta$ gal1	GH 2	yes	HW_C1/C4	3108 bp	117.2	kDa	99%	Klebsiella sp.
2	HW- $\beta$ gal2	GH 2	yes	HW_C2/C3	3075 bp	116.1	kDa	97%	Kosakonia sp.
3	HW- $\beta$ gal3	GH 2	yes	HW_C5/C6	3108 bp	118.5	kDa	72%	Mangrovibacter sp
4	HW- $\beta$ gal2	GH 42	yes	HW_X1	2058 bp	77.3	kDa	94%	Klebsiella sp.
5	HW- $\beta$ gal3	GH 53	no	HW_X1	1203 bp	44.9	kDa	95%	Klebsiella sp.
6	HW- $\beta$ gluc1	GH 1	no	HW_X2	1335 bp	41.8	kDa	92%	Bacillus
7	HW- $\beta$ gluc2	GH 16	no	HW_X2	768 bp	30.6	kDa	65%	Caldicellulosiruptor sp.
8	MS- $\beta$ gal2	GH 53	no	MS_17A	1203 bp	44.8	kDa	99%	Raoultella sp.
9	MS- $\beta$ gal3	GH 42	yes	MS_17A	2067 bp	77.6	kDa	99%	Enterobacteriaceae sp.
10	SAS- $\beta$ gal1	GH 2	yes	SAS_C1/C2/C3	3072 bp	115.8	kDa	97%	Enterobacter sp.
11	SAS- $\beta$ gal2	GH42	yes	SAS_C4/C5/C6	2108 bp	77.5	kDa	91%	Enterobacter sp.
12	SAS- $\beta$ gal3	GH 53	yes	SAS_C4/C5/C6	1188 bp	43.9	kDa	99%	Enterobacter sp.
13	GB- $\beta$ gal1	GH 2	yes	GB_NV5	3120 bp	115.8	kDa	97%	Enterobacter sp.
14	MS- $\beta$ gal1	GH 2	yes	MS_C32	3396 bp	128.6	kDa	79%	Proteiniphilum sp.
15	MS- $\beta$ gluc1	GH 1	yes	MS_3L	1314 bp	46.5	kDa	89%	Devosia sp.
16	MS- $\beta$ gluc2	GH 1	yes	MS_4L	1299 bp	49.3	kDa	51%	Candidatus sp.
17	SAS-LS	GH 68	no	SAS_LS	1290 bp	47.1	kDa	77%	Komagataeibacter sp.

After sequencing, all of the putative genes were amplified by PCR and ligated in frame to a 6x histidine tag in the the pRSET\_A vector system, and again sequenced to confirm correct frame and orientation. As stated previously, some of the assembled contigs contained more than one putative gene that could account for the observed phenotype (HW- $\beta$ gal2 & 3; HW- $\beta$ gluc1 & 2; MS- $\beta$ gal2 & 3 and SAS- $\beta$ gal2 & 3). In these cases both genes were amplified out and ligated separately into pRSET\_A and then re-transformed into the mutant to ascertain which genes is responsible for the observed phenotype, thus revealing the true positives (Fig 3-3 and Table 3-3) For the GH68 putative levansucrase identified, a clear visible exopolysaccharide was observed (Fig 3.4).



**Figure 3-3** The various putative genes for the  $\beta$ -galactosidase screen, cloned into the pRSET\_A vector system and grown on MDA-lac plates containing X-gal. The two rows are the same clones photographed on different backgrounds so as to accentuate the visual effect. (-K) and (+K) is representative of the  $\Delta\beta$ -gal mutant negative and positive controls, containing the empty pRSET\_A vector and pRSET with the complete *E. coli*  $\beta$ -galactosidase coding sequence, respectively. (1) HW- $\beta$ gal1 (2) HW- $\beta$ gal2 (3) HW- $\beta$ gal3 (4) HW- $\beta$ gal4 (5) MS- $\beta$ gal3 (6) SASRI- $\beta$ gal1 (7) SASRI- $\beta$ gal2 (8) SASRI- $\beta$ gal3 (9) GB- $\beta$ gal1 (10) HW- $\beta$ gal5 (11) HW- $\beta$ gluc1 (12) HW- $\beta$ gluc2 (13) MS- $\beta$ gal2 (14) MS- $\beta$ Gluc1 (15) MS- $\beta$ Gluc2. Take note that MS- $\beta$ gal2 was accidentally omitted, hence only 15 and not 16 putative genes are represented (refer to table 3-3)

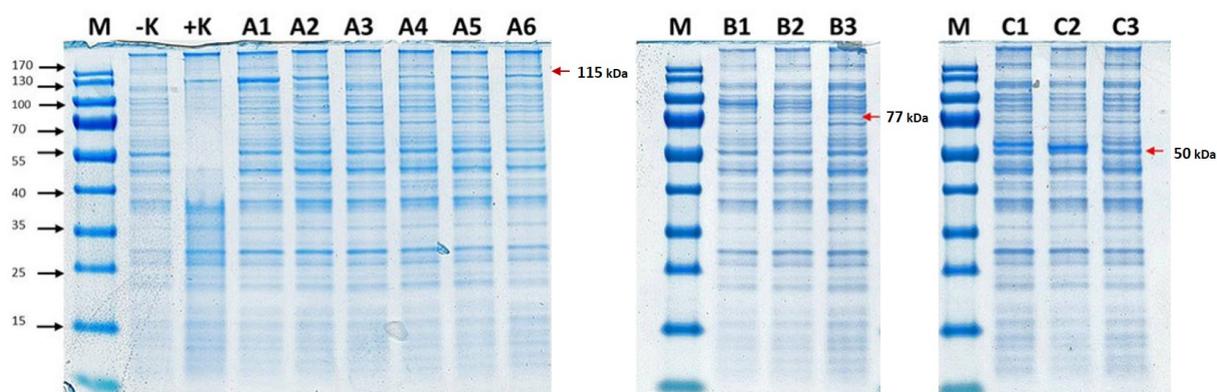


**Figure 3-4** Levansucrase gene isolated from the SASRI metagenomic library clearly showing a visible phenotype on a matt of plated out *E. coli* transformants.

### 3.3.3 Protein analysis and biochemical characterisation

To ascertain whether the respective protein catalysts have any potential for the production of oligosaccharides/polymers, their biochemical characteristics needed to be determined to identify optimal production conditions. For the purposes of the rest of this chapter, focus will be given only to those positive clones isolated from screening on lactose containing medium, the putative  $\beta$ -galactosidases and  $\beta$ -glucosidases. The putative GH68 levansucrase that produced copious amounts of exo-polymer (Fig. 3-4) and its sequence will be investigated more thoroughly in chapter 5. All of the protein constructs were expressed in the  $\Delta\beta$ -gal mutant to remove any background activity resulting from the endogenous *E. coli*  $\beta$ -galactosidase (Fig.3-4). Protein expression in the pRSET vector is driven by the T7 promoter and is therefore dependent on the host cell providing the T7 RNA polymerase, usually through the induction via the *lac* promoter [253, 254]. The DH5 $\alpha$  host cells that were utilized for the creation of the mutant do not contain the DE3 prophage insertion or any other insertions that code for the T7 RNA polymerase. Contrary to expectations, more than adequate protein production was achieved in the mutant, enabling the characterisation of the proteins. It is hypothesised that other polymerases bind to the expression vector for the expression of other vector born genes, and these somehow facilitate the expression of the T7 driven genes. It was also observed that expression of any

of the putative GH2  $\beta$ -galactosidases in several different BL21 protein expression strains [BL21(DE3), BL21(DE3) pLysS and BL21AI] led to cells that appeared to lyse and flocculate during mid-log phase, and no expression could be achieved as the plasmid was dropped mid log phase. Under normal circumstances the native GH2 *E. coli*  $\beta$ -galactosidase is a tetrameric protein and I propose that the endogenous  $\beta$ -galactosidase complexes with the newly introduced protein, resulting in disruption of cellular integrity and cell lysis. Tangible support for this hypothesis comes from an experiment where the GH2  $\beta$ -galactosidase genes isolated from the various libraries, were expressed in the un-mutated DH5 $\alpha$  strain. In all the cases, protein expression could not be achieved as was done with the mutant strain. It is hypothesised that as was the case in the BL21 strains, the vector born  $\beta$ -galactosidase complexed with the truncated form of the endogenous *E. coli*  $\beta$ -galactosidase, resulting in the plasmid ultimately being dropped due to the stress effected on the cells. This problem was not observed for the other  $\beta$ -glucosidases and  $\beta$ -galactosidases from families GH1, GH42 and GH53. The overnight incubations of the respective protein expression constructs were extracted through sonication, their protein concentrations were quantified through the Bradford method (1976), and then separated on SDS denaturing polyacrylamide gel electrophoresis [245, 246]. Only the protein constructs that produced a blue phenotype when grown on X-gal were grown overnight for further investigation. The protein extracts were grouped according to the three GH families i.e. GH2 (6 in total), GH42 (3 in total) and GH1 (3 in total) for the PAGE analysis. This was done for simplicity and visual aid as all the GH2 proteins were approx. 115 kDa in size, whilst the GH42 and GH1 proteins were approx. 77 kDa and 50 kDa in size, respectively (see Fig. 3-5 and Table 3-3). Even though it was difficult to discern the individual protein bands in some cases, activity was confirmed for all extracts relative to that of the controls with the aid of nitrophenyl based substrates (Fig. 3-6)



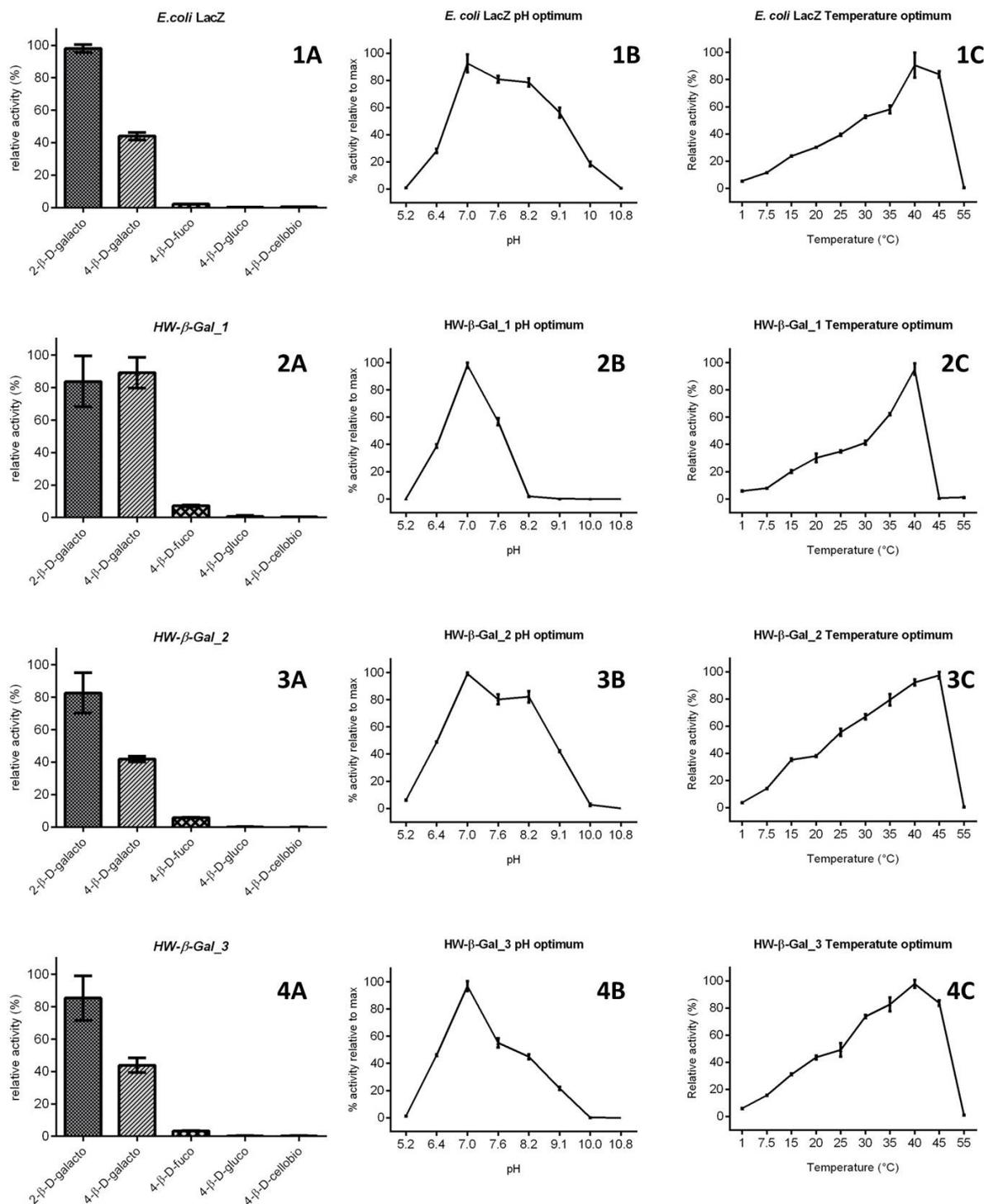
**Figure 3-5** The various  $\beta$ -galactosidases and  $\beta$ -glucosidases that were identified through screening of metagenomic libraries on minimal media containing lactose and X-gal, all cloned into pRSET and expressed in the  $\Delta\beta$ -gal mutant. Proteins are grouped according to family, namely (A) GH2 (B) GH42 (C) GH1 respectively. The individual clones are (-K) empty control (+K) positive control *E. coli*  $\beta$ -gal (A1) HW- $\beta$ gal1 (A2) HW- $\beta$ gal2 (A3) HW- $\beta$ gal3 (A4) SASRI- $\beta$ gal1 (A5) GB- $\beta$ gal1 (A6) MS- $\beta$ gal1 (B1) HW- $\beta$ gal4 (B2) MS- $\beta$ gal3 (B3) SASRI- $\beta$ gal2 (C1) HW- $\beta$ Gluc1 (C2) MS- $\beta$ Gluc1 (C3) MS- $\beta$ Gluc2. For the predicted molecular weights of the individual proteins refer to table 3-3 on page 54.

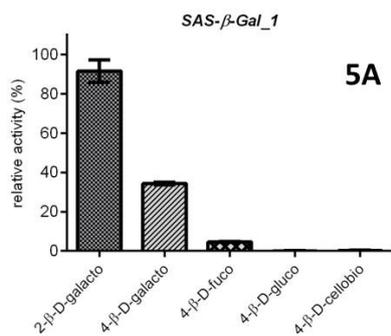
The overnight protein extracts that were visualised through SDS-PAGE, were used to characterise the pH, temperature and substrate specificities of the respective proteins (Fig. 3-6). Because crude extracts were utilised and not the purified proteins itself, specific activities of the individual proteins could not be determined. Consequently the level of activity would be directly proportional to the level of heterologous expression of the individual proteins, which is greatly variable between the different constructs. The activities, whether it be for substrate specificity, pH or temperature optima, were all expressed as a percentage relative to the maximum activity, which was taken to be 100 (Fig. 3-6). Chromogenic nitrophenyl based substrates allows for a relatively easy colorimetric assay to determine substrate specificity as well as the pH and temperature optima of the various proteins.

For an enzyme to be used in the industrial production of GOS and/or polymers it should ideally exhibit a broad pH and temperature range to accommodate any robust production needs. Generally there are two needs with regards to temperature optima, namely a high activity at low temperatures for the production of GOS in milk based products (sub 15°C), or a high level of activity at elevated temperatures for GOS production in saturated solutions of whey permeate lactose (above 35°C) [191, 255]. GOS production has been shown to have a linear relation with substrate concentration, thus the higher the lactose concentration, the higher the rate of GOS production. Lactose itself is not readily soluble at high concentrations at temperatures below 30°C, and therefore thermophilic enzymes are often preferred when it is used as a substrate [103, 256]

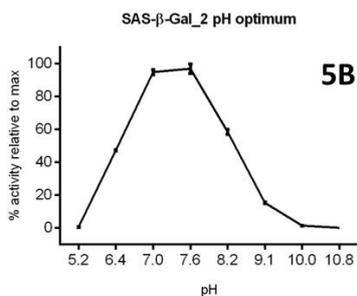
The characterisation of the crude protein extracts were again grouped together, as was done for the SDS PAGE analysis, into the GH2, GH42 and GH1 family groups (Fig. 3-6). All of the GH2 extracts had similar pH optima, with the highest activity being observed at slight acidic to near neutral conditions (pH 6.8) for all the extracts (Fig. 3-6; 2B-7B). So too were the temperature activity spectrum very much the same, with a gradual increase in activity to approx. 40°C, after which a sharp drop in activity was observed. The exception to this was GB- $\beta$ gal1 which showed a maximum activity at 25°C. Interestingly, all the GH2  $\beta$ -galactosidases showed a higher affinity towards 2- $\beta$ -D-galactosyl nitrophenyl based substrates, than for 4- $\beta$ -D-galactosyl nitrophenyl based substrates. The opposite was observed for GH42  $\beta$ -galactosidases and GH1  $\beta$ -glucosidases. GH42 enzymes have been shown to have high specificity for axial C4-OH groups [257]. In addition, all of the GH42  $\beta$ -galactosidases and GH1  $\beta$ -glucosidases demonstrated significant fucosidase activity, this has previously been shown in the literature and relates to glycan degradation of the plant cell walls, often together with that of GH53 proteins, as mentioned previously. Moreover they are often found together with cellulosomes, which further cements their role in glycan degradation [258, 259]. GH42 enzymes are often used in the production of lactose free milk [260]. As was the case for the GH2 extracts, the GH42 and GH1 pH optima was near neutral at around pH 6.8, except for MS- $\beta$ gluc2 which had a more acidic optima of pH 6.0, which then suddenly dropped with almost no activity at pH 7.0. In general the GH42 extracts had a broader pH spectrum than that of

the GH1 extracts, and was more comparable with that of the GH2 extracts. The temperature optima for the GH42 and GH1 extracts also gradually increased towards approx. 40°C after which a sudden drop was observed. The temperature and pH ranges of the isolated protein extracts correlate with two of the most prominent sources of biocatalyst for the production of GOS, namely *Kluyveromyces lactis* and *Bacillus circulans*, and depending on their product formation capabilities, any one of them could have potential industrial relevance [185, 261].

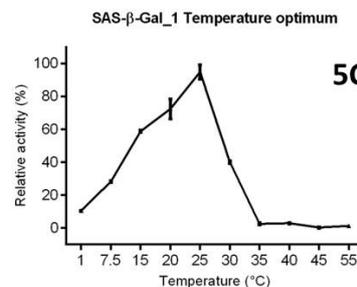




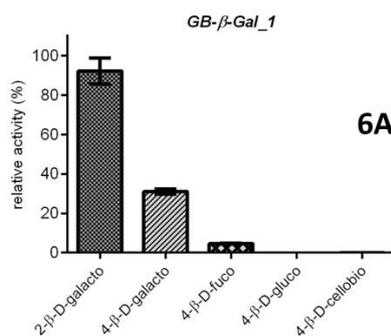
**5A**



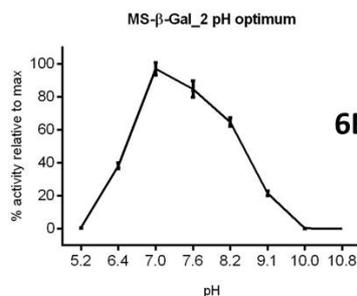
**5B**



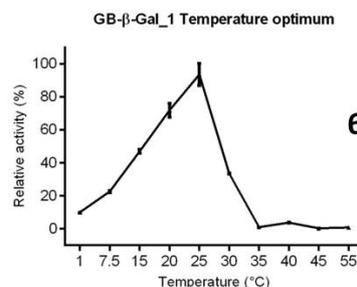
**5C**



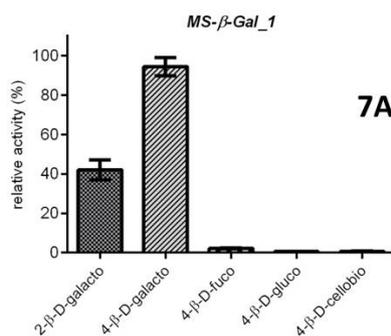
**6A**



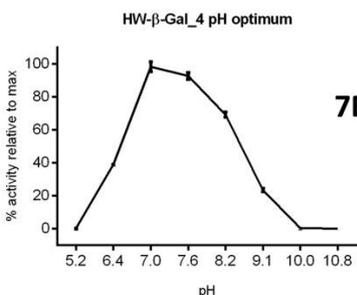
**6B**



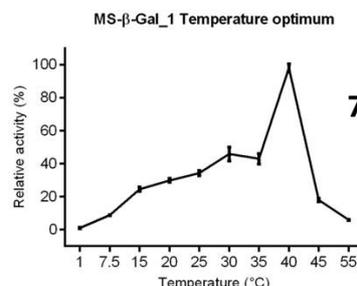
**6C**



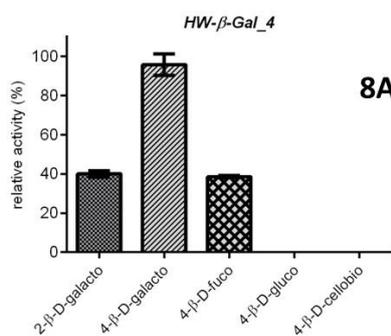
**7A**



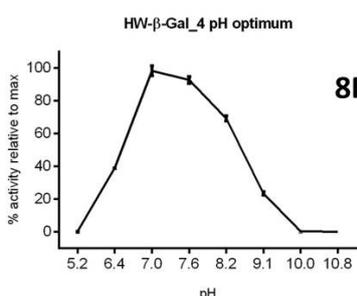
**7B**



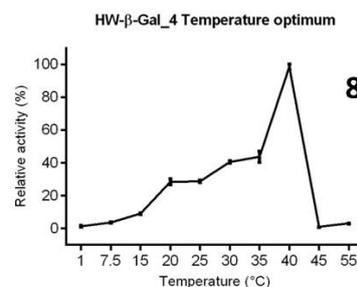
**7C**



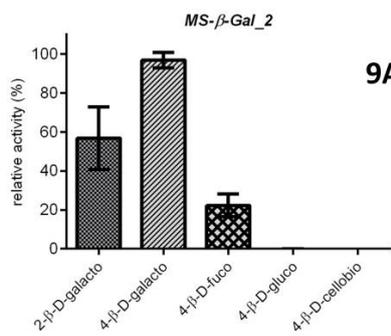
**8A**



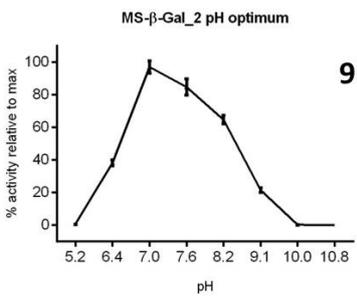
**8B**



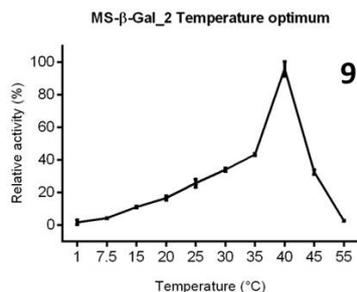
**8C**



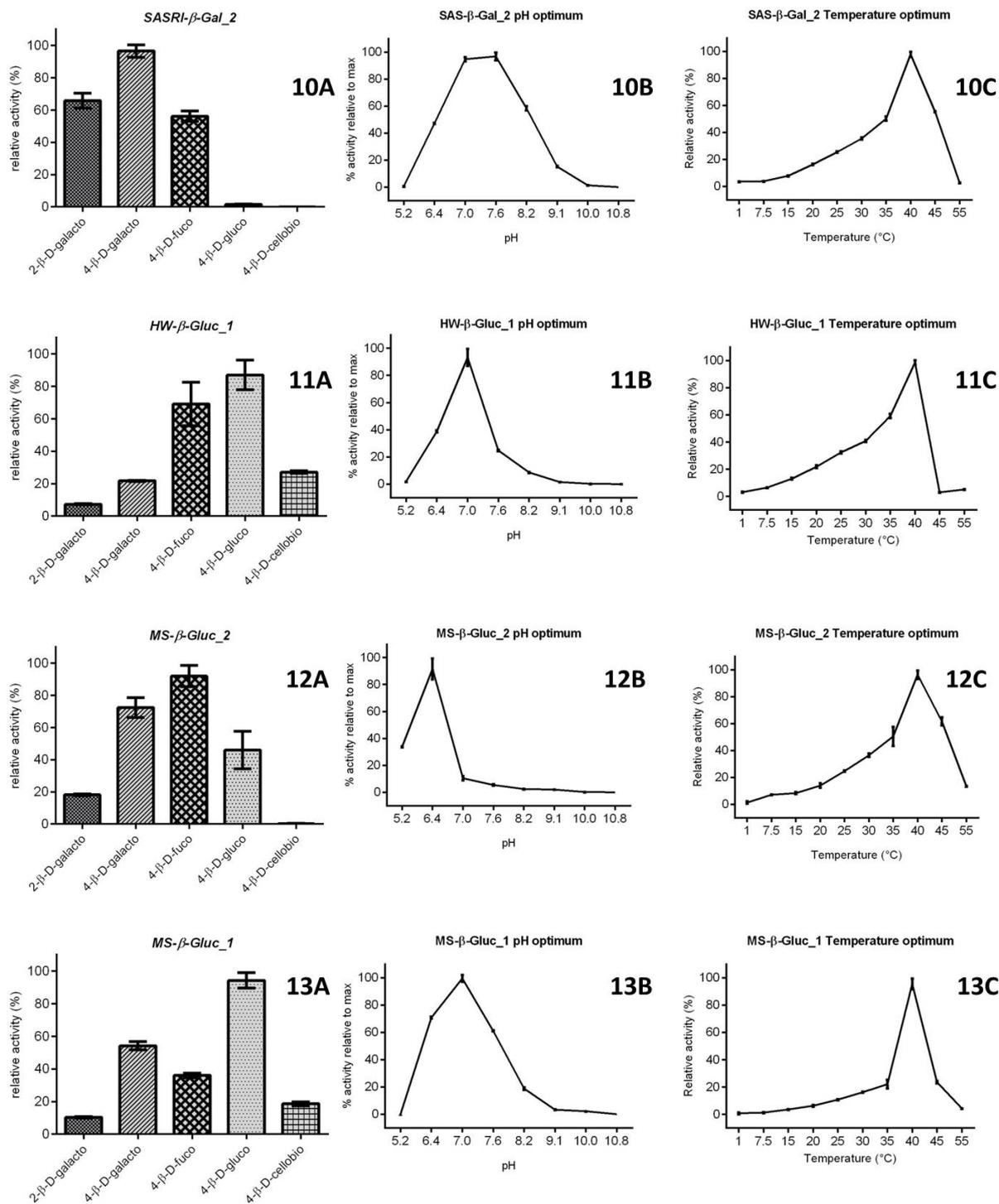
**9A**



**9B**



**9C**



**Figure 3-6** The substrate specificity (A), pH optimum (B) and optimum temperature (C) of the various putative genes (2-13) with *E. coli*  $\beta$ -galactosidase (1) as a positive control. The  $\Delta\beta$ -gal mutant with empty pRSET vector used as negative control but no significant activity could be detected (results not shown). The spectrophotometric hydrolysis was measured at 405 nm and free nitrophenol deduced from a standard linear curve that was established with 4-nitrophenol, ranging between 0.0025  $\mu$ M and 0.04  $\mu$ M. All activity was expressed as a percentage, relative to the highest amount, which was taken to be 100%. Temperature and pH was characterised with either 2- $\beta$ -D-gal-NP or 4- $\beta$ -D-gal-NP depending on which showed the highest activity for the respective constructs. The samples assayed were (2) HW- $\beta$ gal1; (3) HW- $\beta$ gal2; (4) HW- $\beta$ gal3; (5) SASRI- $\beta$ gal1; (6) GB- $\beta$ gal1; (7) MS- $\beta$ gal1; (8) HW- $\beta$ gal4; (9) MS- $\beta$ gal3; (10) SASRI- $\beta$ gal2; (11) HW- $\beta$ Gluc1; (12) MS- $\beta$ Gluc1; (13) MS- $\beta$ Gluc2.

### 3.3.4 Product analysis

The industrial relevance of any potential biocatalyst obtained from the library screens will always be underscored by the product formation capabilities and kinetic characteristics of the enzyme. The crude characterization of the respective enzymes with nitrophenyl based substrates enabled us to establish the optimal conditions for the hydrolytic reaction, but it provides no insight for the trans-glycosylation reaction or the products formed. It has always been assumed that the optimal conditions for trans-glycosylation is the same as that for the hydrolysis reaction, in seeing that the two reaction are interlinked, and all throughout the literature it is treated as such [262-264].

The product forming capabilities of each of the protein extracts were investigated by overnight incubation in a lactose substrate solution (30%; w/v) under their respective optimal conditions, with regards to pH and temperature. The extracts were incubated at a concentration of 1 u.mL<sup>-1</sup> in a total volume of 10 mL. One unit of activity is defined as the amount of total crude protein extract needed to liberate 1  $\mu$ mole of free reducible glucose from the disaccharidic lactose substrate in 1 minute. The enzyme linked NADH reduction assay for the measurement of the free reducing sugars glucose, galactose and fructose is visually explained in Fig. 3-7 and described in more detail under the material and methods section. The standardised overnight reactions will allow for all the reactions to equilibrate to endpoint, and enable the final product formations to be visualised through TLC and the ratio of free glucose and galactose to be determined.

The product formation visualised by TLC is shown in Fig. 3-8. Included in the TLC is the commercially available GOS preparation, Vivinal GOS, from Friesland Campina (Amersfoort, The Netherlands). It should be noted that this industrially produced commercial preparation is a concentrated form that was treated by various means so as to rid it of as much free monosaccharide and lactose as possible through various fermentation and purification methods. It should be used as a direct comparison of catalytic efficiency only in the most extreme case [265]. An ideal enzymes would be one that shows the highest propensity for trans-galactosylation over hydrolysis together with a high rate of substrate conversion. TLC is a simple and rudimentary chromatography method which gives no insight to linkage type. The main three focal points of the TLC in Fig. 3-8 is the final equilibrium amounts of, (i) the free glucose and galactose as visualised on top part of the TLC. (ii) The reduction in the substrate lactose in the middle of the TLC, indicative of substrate conversion. (iii) The formation of higher DP oligosaccharides towards the lower part of the TLC. To further aid in the characterisation of the trans-galactosylation reaction, the ratio of free glucose to galactose is shown at the bottom of each TLC sample (Fig 3-8) . From the figure below it can be seen that the two enzymatic preparations that appears to produce the largest

amount of GOS with the highest equilibrium endpoint towards product formation is (A1) HW-βgal1, and (C2) MS-βGluc1. Even though their ratio of free glucose: galactose isn't the highest, they appear to have the highest substrate conversion to GOS, by far. The other protein extract preparations appear to reach an endpoint equilibrium where either they continuously hydrolyse and trans-glycosylate the GOS already formed, or they produce a disaccharidic product that appear similar to the substrate lactose.

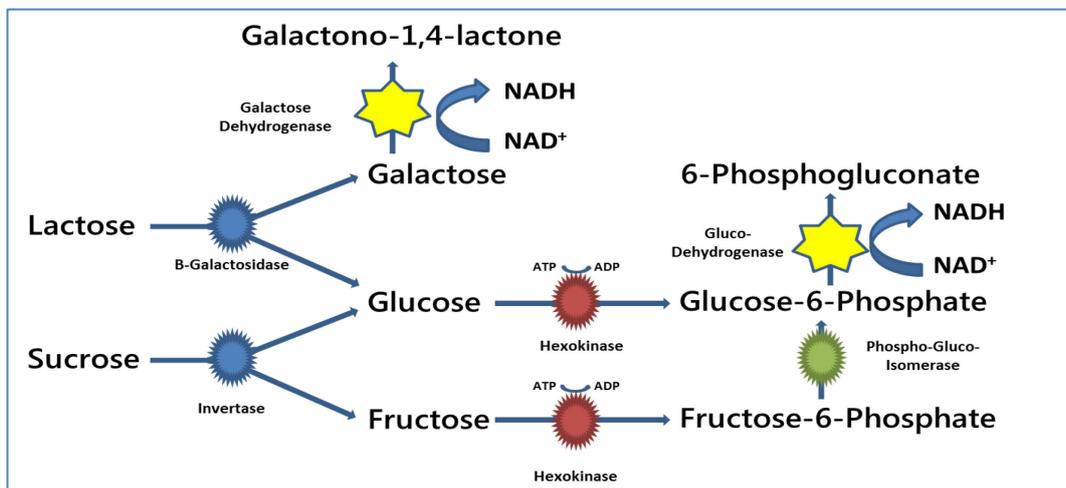


Figure 3-7 The spectrophotometric measurement of free glucose, fructose and galactose is coupled to the reduction of NAD<sup>+</sup> to NADH which is measurable at 340nm wavelength. For this to be observed glucose first needs to be phosphorylated and then reduced to 6-phosphogluconate through the actions of Hexokinase/Glucose-6-phosphate Dehydrogenase (HK-G6DH). For the measurement of fructose, it first needs to be phosphorylated to fructose-6-phosphate and then isomerised to glucose-6-phosphate through the combined action of Hexokinase/Phosphogluco-Isomerase (PGI), which in turn can then be reduced by HK-G6DH. For the measurement of galactose it needs to be reduced to galactono-1,4-lactone through the actions of Galactose Dehydrogenase (GDH).

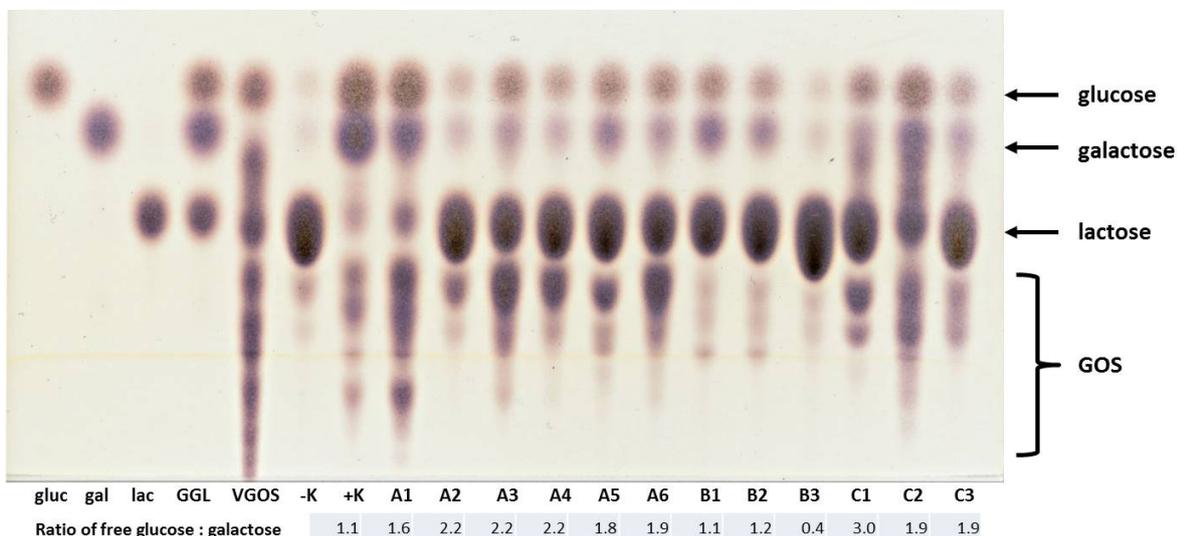


Figure 3-8 The product formation of the respective enzymes visualised by means of thin layer chromatography. 0.5μl of an 10x dilution of a 30% lactose(w/v) solution incubated overnight with 1 U.mL<sup>-1</sup> at the optimum temperature and pH for each of the respective enzymes. The individual lanes are (Glc) glucose (gal) galactose (lac) lactose (GGL) glucose/galactose/lactose (VGOS) Vivinal GOS commercial preparation of GOS (-K) empty control (+K) positive control E. coli β-gal (A1) HW-βgal1 (A2) HW-βgal2 (A3) HW-βgal3 (A4) SASRI-βgal1 (A5) GB-βgal1 (6) MS-βgal2 (B1) HW-βgal4 (B2) MS-βgal4 (B3) SASRI-βgal2 (C1) HW-βGluc1 (C2) MS-βGluc1 (C3) MS-βGluc2. The bottom row below the table is the ratio of free glucose to galactose as determined spectrophotometrically by NAD<sup>+</sup> linked enzyme assay.

### 3.4 Conclusion

The metagenomic DNA isolation was successful and three libraries with decent titers were constructed with the lambda ZAP express bacteriophage library system. A  $\beta$ -galactosidase deletion mutant was constructed, through  $\lambda$ -Red recombineering, to facilitate the screening for novel biocatalyst that hydrolyse lactose and x-gal and potentially produce GOS. Altogether five libraries were screened for novel enzymes that possesses either trans-galactosylation or trans-fructosylation activity. Through functional screening of the metagenomic libraries, seventeen putative genes were identified that could potentially produce prebiotic oligosaccharides/polymers. All were sequenced and *in silico* analysed. They were PCR amplified and cloned into the protein expression vector pRSET\_A, and heterologously expressed in *E. coli*. Of these original seventeen, thirteen were again positively associated with the phenotype, and further characterised on a rudimentary level. Their substrate specificities, pH and temperature optimums were established through the spectrophotometric analysis of their hydrolytic action on various nitrophenyl based substrates. Their product forming capabilities were investigated by overnight incubation with the substrate lactose, under optimal conditions. The equilibrium end reactions were visualised through TLC, and the free reducing sugars assayed through the spectrophotometric measurement of the enzyme linked reduction of NADH. Preliminary data of the biochemical characteristics and product formation capabilities shows promise but further and more detailed analysis is needed of the purified proteins. The putative  $\beta$ -galactosidase HW- $\beta$ gal1 isolated from the Hotwood-library, together with the putative  $\beta$ -glucosidase MS- $\beta$ Gluc1 isolated from the MS-library, and the putative levansucrase SASRI-LS2 isolated from the SASRI-library, showed the greatest promise with regard to product formation, and were selected for further analysis that will be discussed in detail in the next two chapters.

## CHAPTER 4: The biochemical characterisation of two novel glycoside hydrolases, a $\beta$ -galactosidase and $\beta$ -glucosidase, belonging to the Glycoside Hydrolase families 1 & 2 respectively, both with high trans-galactosylation activity.

*This chapter is partially written in a style aimed at publication. Of the two enzymes characterised here, the  $\beta$ -glucosidase is most promising and the article extracted will possibly be submitted to the journal 'Applied Microbiology and Biotechnology' [ISSN: 01075-7598 (print); 1432-0614 (web)]. So as to avoid repetition, some of the 'Material and Methods' and 'Results' covered in chapter three will be omitted.*

### 4.1 Abstract

Genes encoding two novel glycoside hydrolases named MS- $\beta$ gluc1 and HW- $\beta$ gal1, were isolated from two separate metagenomic libraries, as described previously in chapter three. The libraries were constructed with DNA extracted from soil at either a place where milk runoff occurs at a dairy farm or from process runoff collected at a sawmill. In the previous chapter, the crude protein extracts were rudimentary biochemically characterized with nitrophenyl based substrates, and analysed for their prebiotic GOS synthesis abilities. In this chapter the proteins were purified and characterised more thoroughly with the substrate lactose, and so too their products. Both genes were heterologously expressed in *E. coli* and their respective proteins purified to apparent homogeneity by immobilized metal affinity chromatography (IMAC). HW- $\beta$ gal1, a GH2 type  $\beta$ -galactosidase had a 1035 amino acid sequence and a predicted molecular mass of 117 kDa. For the substrate lactose it had a  $K_m$  of 32 mM, a  $K_{cat}$  of 625  $s^{-1}$  and a catalytic efficiency of 19.5  $mM^{-1}.s^{-1}$ . The optimum pH and temperature for the hydrolysis reaction of HW- $\beta$ gal1 was determined to be pH 7.5 and 35°C respectively. The second enzyme, MS- $\beta$ gluc1, was determined to be a GH1 type  $\beta$ -glucosidase and had a 438 amino acid sequence and a predicted molecular mass of 46.5 kDa. For the substrate lactose it had a  $K_m$  of 41 mM, a  $K_{cat}$  of 111  $s^{-1}$  and a catalytic efficiency of 2.7  $mM^{-1}.s^{-1}$ . The optimum pH and temperature for the hydrolysis reaction of MS- $\beta$ gluc1 was determined to be pH 6.0 and 45°C respectively. Both proteins were shown to have a high trans-galactosylation activity with HW- $\beta$ gal1 yielding approximately 43% galacto-oligosaccharides (GOS), mainly consisting of tri-saccharides of the  $\beta$ -D-1,6-lactose linkage type. MS- $\beta$ gluc1 yielded approximately 40% GOS, also consisting mainly of trisaccharides of both the  $\beta$ -D-1,3-lactose and  $\beta$ -D-1,6-lactose linkage type. The results indicate that both HW- $\beta$ gal1 and MS- $\beta$ gluc1 are attractive candidates for lactose conversion and galacto-oligosaccharide production based on high activity and stability within a broad pH and temperature range.

## 4.2 Introduction

Prebiotics are non-digestible oligosaccharides and polymers that are selectively fermented by the commensal microbiota of the human gastro intestinal tract and confer benefits to the hosts, other than nutritional value [14, 266, 267]. The benefits conferred are numerous and range from improving mineral absorption, inhibiting pathogenic adhesion and prevention of colon cancer, to modulating the hosts immune system, regulating metabolism and serum lipid profiles [7, 18, 197, 268]. Prebiotic intake through diet is important in the maintenance of a healthy GI microbiota [11, 269]. Perturbations of that can be caused by various factors, such as host genetics, antibiotic treatment and infection. Reduced microbial diversity and/or the loss of beneficial microbes and pathobiont expansion is referred to as dysbiosis [2, 3, 56]. This has been linked to various diseases including metabolic syndrome, dyslipidemia, coronary heart disease, inflammatory diseases like coeliac and Crohn's disease and even diseases of the cognitive state like depression, anxiety and dementia [4, 56, 94, 270]. Prebiotics ingestion have been proven to be an effective counter for dysbiosis and its resulting effects, by promoting the growth of beneficial bacteria [90, 197, 266, 267]

Galacto-oligosaccharides (GOS) are a form of prebiotic oligosaccharide, derived from lactose and consist of galactose monomers often linked to a terminal glucose [271, 272]. GOS can vary in both chain length and linkages and normally consist of between three and nine degrees of polymerisation [189, 273]. GOS closely resemble human milk oligosaccharides, which are a large constituent of mammalian milk specifically aimed at stimulating the growth of beneficial bacteria. This stimulatory growth effect is colloquially referred to as the bifidogenic effect, even though they are not exclusively linked to Bifidobacterium growth and effect various beneficial commensal bacteria either directly or indirectly [174, 266]. For this reason GOS are often added to infant formula [31, 95, 274].

GOS can be produced from lactose by the enzyme  $\beta$ -galactosidase (EC 3.2.1.23), through its trans-galactosylation activity [23, 275, 276]. They often catalyse two competing reactions, namely hydrolysis and trans-galactosylation. After it has broken the  $\beta$ -1,4-D linkage between the glucose and galactose, an enzyme substrate complex is formed between the galactose monomer and the active site. If the galactosyl moiety is then transferred to H<sub>2</sub>O, hydrolysis occurs, but if it's transferred to another carbohydrate, trans-galactosylation results [118, 277]. Eventually a dynamic equilibrium is formed between hydrolysis and trans-galactosylation. Enzymes which catalyse a reaction that favours trans-galactosylation over hydrolysis will result in higher yields of GOS and are highly sought after by industry [128, 191]. Traditionally the main industrial sources of  $\beta$ -galactosidases utilised for the production of GOS, has been from the fungal species *Kluyveromyces* and *Aspergillus* [118, 190].

However, increasingly the identification of other prokaryotic sources of these enzymes has been sought after due to their ease of use in microbial expression system [118, 256, 278]. Some  $\beta$ -glucosidases have also been shown to facilitate trans-galactosylation when presented with lactose as a substrate and, as they are often smaller in size and more robust than the classical  $\beta$ -galactosidases, they make excellent candidates for industrial application [212, 279-281].

The combination of metagenomics and functional screening provide a powerful tool for the discovery of new catalytic enzymes that can aid in the betterment of industrial processes for the production of prebiotic GOS [233, 249]. In this chapter we build upon the work conducted in the previous chapter. The heterologous expression of two metagenomic derived genes, encoding a GH1 and GH2  $\beta$ -galactosidase and  $\beta$ -glucosidase, referred to as HW- $\beta$ gal1 and MS- $\beta$ gluc1 respectively, were successfully carried-out. Ultimately their protein products were purified to near homogeneity by means of IMAC. The purified proteins were biochemically characterised. Analysis of their kinetic properties and their enzymatic products formed, indicate that both have relatively high trans-galactosylation activity.

## 4.3 Materials and methods

### 4.3.1 Strains, plasmids and materials

All chemicals and enzymes were purchased from Sigma (St. Louis, MO, USA) unless stated otherwise. All restriction enzymes, T4 DNA ligase, DNA polymerase and corresponding buffers were purchased from Thermofisher Scientific (Massachusetts, USA). The enzymes galactose dehydrogenase/mutarase (GAL-DH/M) and hexokinase/glucose-6-phosphate-dehydrogenase (HK-G6DH) were obtained from Megazyme (Wicklow, Ireland). The lambda ZAP express vector kit and packaging extract was purchased from Agilent technologies (California, USA). The pRSET protein expression vector system was obtained from Life technologies (California, United States). The oligosaccharide standards 1,3-galactobiose ( $\alpha$ -Gal-1,3-Gal); 1,4-galactobiose ( $\beta$ -Gal-1,4-Gal); 1,6-galactobiose ( $\beta$ -Gal-1,6-Gal); galactotriose ( $\alpha$ -Gal-1,3- $\beta$ -Gal-1,4-Gal); galactotetraose ( $\alpha$ -Gal-1,3- $\beta$ -Gal-1,4- $\alpha$ -Gal-1,3-Gal); lactose ( $\beta$ -Gal-1,4-Glc); allolactose ( $\beta$ -Gal-1,6-Glc); mellibiose ( $\alpha$ -Gal-1,6-Glc); lactulose ( $\beta$ -Gal-1,6-Fru); 3' galactosyl-lactose ( $\beta$ -Gal-1,3- $\beta$ -Gal-1,4-Glc); 4' galactosyl-lactose ( $\beta$ -Gal-1,4- $\beta$ -Gal-1,4-Glc); 6' galactosyl-lactose ( $\beta$ -Gal-1,6- $\beta$ -Gal-1,4-Glc); maltobiose ( $\alpha$ -Glc-1,4-Glc); maltotriose ( $\alpha$ -Glc-(1,4)<sub>2</sub>-Glc); maltotetraose ( $\alpha$ -Glc-(1,4)<sub>3</sub>-Glc); maltopentose ( $\alpha$ -Glc-(1,4)<sub>4</sub>-Glc); maltohexaose ( $\alpha$ -Glc-(1,4)<sub>5</sub>-Glc) and maltoheptaose ( $\alpha$ -Glc-(1,4)<sub>6</sub>-Glc), were purchased from either Carbosynth (Compton, United Kingdom) or Dextra

laboratories (Reading, United Kingdom). Aluminium backed silica gel 60 thin layer chromatography plates were purchased from Merck (Darmstadt, Germany). The  $\Delta\beta$ -galactosidase DH5 $\alpha$  mutant lacking the endogenous  $\beta$ -galactosidase gene in its entirety was created previously through lambda red recombineering according to the protocols Datsenko and Wanner (2000), and as described in chapter three [241].

#### 4.3.2 Library construction and screening, clone selection, sequencing and *in silico* analysis

The methods of metagenomic DNA isolation, digestion and library construction were conveyed in great detail in sections 3.2.1 to 3.2.2 on p34 - p40. The  $\Delta\beta$ -gal mutant construction was described in section 3.2.3 on page 40 - p42, and the screening of the libraries on minimal media plates were covered in section 3.2.4 on page 42 - p45. The approach to positive clone selection, de-replication sequencing and bioinformatic analysis were covered in section 3.2.5 on page 45.

#### 4.3.3 Cloning, expression and recombinant protein purification

The complete coding sequence for both the HW- $\beta$ gal1 and the MS- $\beta$ gluc1 genes were PCR amplified and cloned directionally in-frame into the pRSET\_A (Invitrogen) protein expression vector, as described in section 3.2.6 on p45. The HW- $\beta$ gal1 and MS- $\beta$ gluc1 expression vectors were transformed into the  $\Delta\beta$ -gal mutant and their respective proteins expressed according to the protocols of Studier *et al.* (2014) [282]. In brief: a single colony from a fresh overnight streak was inoculated into 500 mL of ZYM-5052 auto induction media (1% N-Z-amine, 0.5% yeast extract, 25 mM Na<sub>2</sub>HPO<sub>4</sub>, 25 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NH<sub>4</sub>Cl, 5 mM Na<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.2x metals, 0.5% glycerol, 0.05% glucose, 0.2%  $\alpha$ -lactose) and incubated overnight at 37°C in a 2 L baffled flask with vigorous shaking at approx. ~300 rpm (see section 3.2.4 for media preparation).

The cells were spun down at 10,000  $xg$  for 10 min, washed once with 50 mL of ddH<sub>2</sub>O before being pelleted again as before, and reconstituted in 50 ml of protein extraction buffer (50 mM imidazole, 1.5M NaCl, 20 mM Tris-HCl pH 7.2). The cell suspension was sonicated five times on ice in 30 second burst with 1 minute intervals in-between, with intermittent swirling. The cell debris was pelleted by centrifugation at 12,000  $xg$  for 10 min and the supernatant transferred to a new tube. The supernatant was force-filtered through a 0.22 $\mu$ m disposable filter connected to a 50 mL syringe, and the sample kept on ice till IMAC purification.

The recombinant proteins were individually purified, with the aid of a HisTrap HP 5ml column from Life Technologies (California, United States) on an ÄKTA FPLC (GE Life Sciences) protein purification system, according to the manufacturer's recommendations. The proteins were eluted in 500  $\mu$ L fractions over an imidazole gradient (50 mM - 500 mM), and a 50 mL elution volume. The resultant fractions were tested for *o*-Nitrophenyl- $\beta$ -D-galactopyranoside activity as described below in section 4.3.4. Fractions demonstrating activity were pooled and dialysed overnight against 0.1x TE (1 mM Tris-HCL pH 7.0, 0.1 mM EDTA). Protein concentrations were determined according to the method of Bradford (1976) using bovine serum albumin as the standard, before being analysed on denaturing SDS polyacrylamide gel electrophoresis so as to assess the purity [245].

#### 4.3.4 Nitrophenyl based assays and enzyme linked assays for enzymatic characterisation.

The nitrophenol based substrate specificity assays conducted in chapter three with the crude protein extracts were repeated, but this time only with the purified protein obtained in section 4.3.3. This was done in order to ascertain if there was any aberrant effect introduced on the substrate specificity by the crude extracts, rather than with the purified proteins.

The protocol in brief: a standard linear curve with 4-nitrophenol ranging between 0.0025  $\mu$ M and 0.04  $\mu$ M was established spectrophotometrically by measuring at wavelength 405 nm on a VersaMax ELISA microplate reader (Molecular Devices). To test for substrate specificity, 10 mM solutions of various nitrophenyl based substrates dissolved in Z-buffer (10 mM phosphate buffer pH 7.0, 5 mM  $MgCl_2$ ), were incubated with 0.05  $\mu$ g of the respective purified proteins. The activity was measured continuously at 405 nm for a period of 30 min at 37°C. If the assayed activity was out of the standard range by either being too high or too low, the protein concentration was adjusted accordingly so as to allow adequate spectrophotometric analysis of the enzymatic reaction. All measurements were done in triplicate. The nitrophenyl based substrates utilized to assay the substrate specificity of HW- $\beta$ gal1 and MS- $\beta$ gluc1 were the following: ONP- $\alpha$ -D-galactopyranoside, ONP- $\beta$ -D-galactopyranoside, ONP- $\beta$ -D-maltoside, ONP- $\beta$ -D-xylopyranoside, ONP- $\beta$ -D-glucuronide, ONP- $\alpha$ -D-glucopyranoside, ONP- $\beta$ -D-glucopyranoside, ONP- $\beta$ -D-fucopyranoside, ONP- $\beta$ -L-arabinopyranoside, ONP- $\alpha$ -L-rhamnopyranoside, ONP- $\beta$ -D-cellobioside. All nitrophenyl based substrates were either obtained from Sigma-Aldrich (Missouri, USA), or Carbosynth (Berkshire, UK). For the nitrophenyl based substrates, one unit of activity is defined as the amount of purified enzyme needed to produce 1  $\mu$ mole of nitrophenol in one minute under standard assay conditions.

#### 4.3.5 Biochemical characterization with the substrate lactose, through enzyme linked assay.

In order to appropriately biochemically characterize HW- $\beta$ gal1 and MS- $\beta$ gluc1, their kinetics with regards to the substrate lactose needs to be quantified. For the enzymatic characterisation using lactose as substrate, free reducing glucose and galactose were determined using an enzyme linked assay based on the reduction of NAD<sup>+</sup> to NADH at 340nm, on a VersaMax ELISA microplate reader (Molecular Devices). Hexokinase/glucose-6-phosphate dehydrogenase (HK-G6DH) is coupled to the measurement of free reducing glucose. Galactose-dehydrogenase/mutararose (GAL-DH/M) together with HK-G6DH is coupled to the measurement of free reducing galactose. A standard curve was established for the NADH coupled reduction of glucose on the spectrophotometer in a range of 0.02  $\mu$ mole to 0.2  $\mu$ mole per well. Because glucose and galactose occur at equimolar ratios within the substrate lactose, the standard curve of glucose can be extrapolated for the measurement of free galactose also.

The protocol in brief: to each well of a clear-bottom UV transparent microtiter plate, 200  $\mu$ L of assay buffer (100 mM Tris-HCl pH8.5, 1 mM NAD<sup>+</sup>; 1 mM ATP) was added. The samples were added to the wells in a total volume of 50  $\mu$ L, making the final volume a combined 250  $\mu$ L per well. A continuous assay was established on the spectrophotometer at 340 nm until baseline was reached, and an endpoint background measurement taken. For the measurement of glucose, 10  $\mu$ L of a 40x dilution of a HK-G6DH enzyme preparation from Megazyme was added to each well (approx. 1 U hexokinase and 0.5 U G6-DH per well). Again an continuous assay was established and an endpoint measurement taken when baseline was reached (approx. 20 min). The change in OD<sub>340</sub> from the first measurement, is directly proportional to the amount of free glucose in solution, and can be deduced from the standard curve previously established. For the measurement of free galactose, 10  $\mu$ L of a 50x dilution of GAL-DH/M was added to the same well (approx. 1 U GAL-DH per well). Again a measurement was taken after baseline was reached. The difference in optical density at 340 nm between the last two endpoint measurements is directly proportional to the amount of free galactose present in solution, and can be calculated from the standard curve. For the substrate lactose, one unit of enzyme activity is defined as the amount of protein needed to produce 1  $\mu$ mole of NADH, through the coupled reduction of glucose, after it was liberated from the substrate lactose, per minute, under the standard assay condition.

Lactose was used as substrate for all enzymatic characterisation, other than the initial substrate specificity utilising the nitrophenyl based substrates. The amount of free glucose released by either HW- $\beta$ gal1 or MS- $\beta$ gluc1 after incubation with lactose was used to characterise the hydrolytic reaction, while the the amount of free galactose subtracted from the amount of free glucose was used to characterise the trans-galactosylation reaction. For the characterisation of the pH, temperature and co-factor optima with lactose as substrate, an endpoint stopped assay was employed. One unit of

enzyme activity was incubated in a 500µl lactose solution (20%; w/v) for 20 minutes, after which the samples were flash frozen in liquid N<sub>2</sub> and subsequently incubated at 90°C for 5 min. Characterisation of optimum pH was performed using the following buffers: acetic acid/sodium acetate (pH 5.2; 50mM), sodium phosphate Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> (pH6.4, pH7.0; 50mM), bis-tris propane/citric acid (pH7.6, pH8.2, pH9.1; 50mM) and sodium carbonate (pH10.3; 50mM). For the characterisation of the temperature optima of HW-βgal1 and MS-βgluc1, temperatures ranging from 10°C to 55°C were evaluated. For the co-factors the divalent cations Cu<sup>2+</sup>, Ca<sup>2+</sup>, Co<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup> as well as NaCl were tested against H<sub>2</sub>O and EDTA in a final concentration of both 1 mM and 5m M. Michaelis–Menten kinetics were assayed in lactose concentrations ranging from 8 mM to 1.1 M using the optimum pH, temperature and co-factors as determined previously.

#### 4.3.6 Enzymatic synthesis of GOS

The synthesis of GOS was carried out using lactose monohydrate as the substrate and either HW-βgal1 or MS-βgluc1 as the catalyst. The reaction was carried out overnight in 10ml final volume lactose solution (30%; w/v) containing 10 units of purified enzyme (1 U.mL<sup>-1</sup>). HW-βgal1 was assayed at pH 8.5 (20 mM Tris-HCl) and 35°C, whilst MS-βgluc1 was assayed at pH 6.0 (20 mM Tris-HCl) and 43°C, both in a thermostatic oscillator. Following the overnight incubation the samples were twice extracted with an equal volume of phenol-chloroform : isoamyl alcohol (25:24:1), and once only with an equal volume of chloroform. Each time the sample was centrifuged at 12,000 *xg* for 5 min at 4°C and the supernatant transferred to a new tube. The final transfer was diluted with ddH<sub>2</sub>O to a final concentration of 5% of the original substrate concentration and used for further analysis.

#### 4.3.7 Carbohydrate analysis

The analysis of the final equilibrium reaction of the overnight incubation with the substrate lactose by TLC was repeated as in chapter three. This time however the purified enzyme was used as catalyst, rather than the crude protein extract. This was done in order to assign the GOS synthesis solely to the respective purified proteins, and negate the potential effect to any other enzymatic catalyst in the crude extract. Thin layer chromatography was conducted according to the protocol of Wang *et al.* (2014) [247].

Of the previously obtained 6x diluted overnight sample described above in section 4.3.6, 0.2  $\mu\text{l}$  was spotted next to 0.2  $\mu\text{l}$  of the standards solutions (10% w/v of glucose, galactose and lactose). The samples and standards were spotted just above a pencilled horizontal straight line, on aluminium backed silica gel 60 TLC plates (Merck, Germany). The plates were briefly dried in an oven, before being run in a sealed glass chamber that was equilibrated prior for an hour with the mobile phase, consisting of butanol/ethanol/water (5:3:2). The plates were run until the mobile phase reached  $\frac{3}{4}$  of the way up. Again the plates were oven dried briefly before being developed by spray misting an orcinol developing solution (0.5% 3,5-dihydroxytoluene orcinol, w/v; 20% sulphuric acid, v/v), onto them with the aid of a gravity fed pressurised air micro-spray gun. The silica plates were clamped between two glass plates and baked in the oven for a several minutes at 100°C, until product development could be visualised, and then taken out and photographed.

However, to fully characterise the end-products of the dynamic equilibrium reaction, a much more detailed analysis is needed. This thorough quantitative and qualitative characterisation of the GOS products was conducted by means of LC-MS/MS. Oligosaccharide samples were first desalted utilising AG 1-X8 formate and AG 50W resin (Bio-Rad, USA) packed spin columns. Columns were packed, washed and centrifuged twice, with 2 ml ddH<sub>2</sub>O at 1,400  $xg$ . The deproteinated, purified and desalted carbohydrate samples were then separated on a Waters Acquity UPLC system (Hewlett-Packard, USA), coupled to a Waters Synapt G2 QTOF (Hewlett-Packard, USA). Samples (2  $\mu\text{l}$ ) were injected and separated with a UPLC BEH Amide column (1.7  $\mu\text{m}$  particle size; 2.1 x 150 mm; 35 °C) at a flow rate of 0.17 ml/minute. Elution was over a gradient utilising ddH<sub>2</sub>O and acetonitrile as mobile phases, both containing NH<sub>4</sub>OH (0.1%; v/v). The gradient changed from 20 : 80 (water : acetonitrile; v/v) to 50 : 50 over 22 minutes. The column was then returned to the initial running conditions for 8 minutes so to equilibrate, before the next sample was loaded and ran. Electrospray ionization was operated in negative mode, under the following MS conditions: nebulizing gas (N<sub>2</sub>) pressure 260 kPa, cone voltage of 40V. The data was analysed in MassLynx version 4.0 software package (Hewlett-Packard, USA).

For the quantitation of different DP (DP2-7) populations, the following standards were used: lactose (DP2), 3'-galactosyl-lactose (DP3), maltotetraose (DP4), maltopentose (DP5), maltohexaose (DP6) and maltoheptaose (DP7). Standard curves were established, ranging between 0.086  $\mu\text{g}$  and 1.44  $\mu\text{g}$  total carbohydrate injected. The area under the chromatograms were used to extrapolate concentration of the samples, and ultimately calculate the actual amount of the different DP populations.

For qualitative linkage and MS/MS breakdown analyses, the following standards were injected at different amounts (0.5  $\mu\text{g}$  to 2  $\mu\text{g}$ ), depending on the highly individualistic MS signal peak strength: 1,3-galactobiose ( $\alpha$ -Gal-1,3-Gal); 1,4-galactobiose ( $\beta$ -Gal-1,4-Gal); 1,6-galactobiose ( $\beta$ -Gal-1,6-Gal); galactotriose ( $\alpha$ -Gal-1,3- $\beta$ -Gal-1,4-Gal); galactotetraose ( $\alpha$ -Gal-1,3- $\beta$ -Gal-1,4- $\alpha$ -Gal-1,3-Gal); lactose

( $\beta$ -Gal-1,4-Glc); allolactose ( $\beta$ -Gal-1,6-Glc); mellibiose ( $\alpha$ -Gal-1,6-Glc); lactulose ( $\beta$ -Gal-1,6-Fru); 3'-galactosyl-lactose ( $\beta$ -Gal-1,3- $\beta$ -Gal-1,4-Glc); 4'-galactosyl-lactose ( $\beta$ -Gal-1,4- $\beta$ -Gal-1,4-Glc); 6'-galactosyl-lactose ( $\beta$ -Gal-1,6- $\beta$ -Gal-1,4-Glc); maltobiose ( $\alpha$ -Glc-1,4-Glc); maltotriose ( $\alpha$ -Glc-(1,4)<sub>2</sub>-Glc); maltotetraose ( $\alpha$ -Glc-(1,4)<sub>3</sub>-Glc); maltopentose ( $\alpha$ -Glc-(1,4)<sub>4</sub>-Glc); maltohexaose ( $\alpha$ -Glc-(1,4)<sub>5</sub>-Glc) and maltoheptaose ( $\alpha$ -Glc-(1,4)<sub>6</sub>-Glc).

## 4.4 Results and discussion

### 4.4.1 Library construction, screening, isolation and preliminary characterisation

The results for the metagenomic DNA isolation, library construction,  $\Delta\beta$ -gal mutant generation, library screening, positive clone selection, clonal de-replication and sequencing were all presented and discussed in sections 3.3.1 to 3.3.2 on p48 - p55, and will not be repeated here again, save to say that all efforts led to the identification and cloning of HW- $\beta$ gal1 and MS- $\beta$ gluc1 into their respective protein expression constructs.

### 4.4.2 *In silico* analysis of HW- $\beta$ gal1 and MS- $\beta$ gluc1

Even though a basic *in silico* analysis for all the positive clones isolated from the various libraries were presented in chapter three, a more in-depth analysis is offered here for HW- $\beta$ gal1 and MS- $\beta$ gluc1, and furthermore explained in context of the libraries from which they were isolated.

For both libraries, the numerous putative glycoside hydrolase genes isolated from the metagenomes, are indicative of their environments. For the HW-library metagenome, the presence of glycoside hydrolase activity would facilitate the breakdown and use of plant cell wall components as nutrients and contribute towards microbial fitness by providing an energy source. For the MS-library it would allow for the breakdown of an abundant energy source, i.e. lactose, conferring to it the competitive edge. Initially the HW-library was constructed for the purposes of discovering novel thermophilic glucosidases that act upon cellulose, but was screened for  $\beta$ -galactosidase activity also [22, 200]. The HW-library yielded altogether five  $\beta$ -galactosidases and only two  $\beta$ -glucosidases, whilst the MS-library yielded three  $\beta$ -galactosidases and two  $\beta$ -glucosidases. The expectation was that the HW-library will yield more  $\beta$ -glucosidase encoding clones, and the MS-library  $\beta$ -galactosidase encoding ones, however the opposite was found. Several factors could account for this. It could relate to the nature of

$\beta$ -galactosidases and  $\beta$ -glucosidases in that their biological function are overlapping with regards to the hydrolysing of glycosidic bonds [110, 283]. Further support for this comes from the fact that these enzymatic catalyst often populate the same GH families, e.g. GH1, GH2 and GH42 [283, 284]. Ultimately, the probability of recovering a certain gene depends on its abundance in the environmental DNA used for library construction. The respective metagenomes could quite possibly have vastly different representation of microbial diversity thereby accounting for observed results [285]

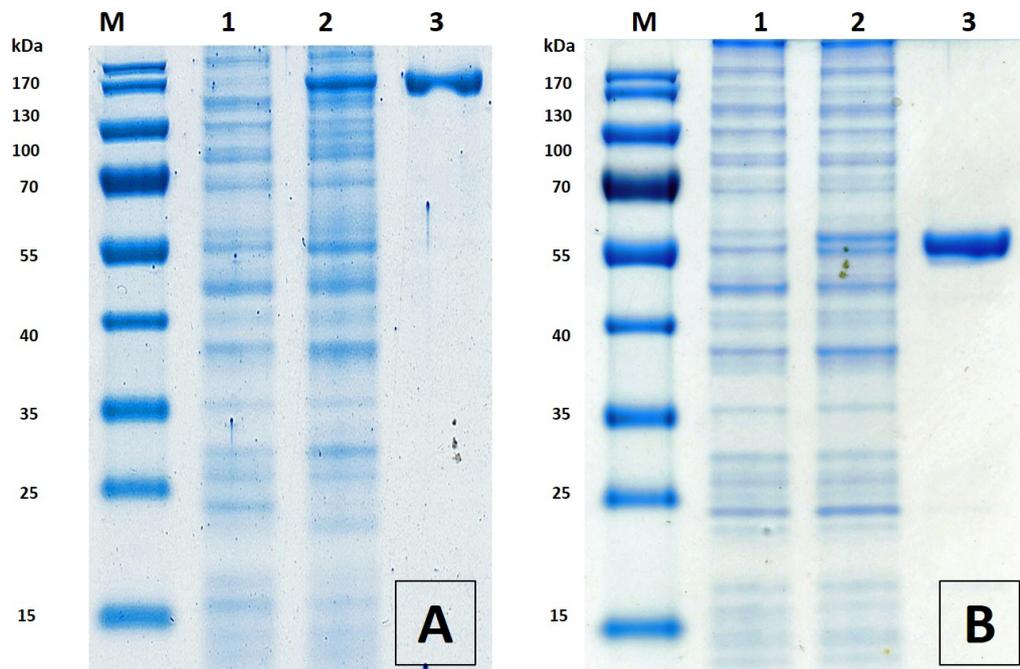
The entire HW- $\beta$ gal1 reading frame consists of 3108 bp and encodes 1035 amino acid residues with a predicted molecular mass of 117.18 kDa and a theoretical pI of 5.53. This enzyme contained 216 charged amino acid residues (20.9% by frequency) and 542 hydrophobic amino acids (52.3%). The estimated  $\alpha$ -helix and  $\beta$ -strand contents were 23% and 32%, respectively. The exponential increase in the number of available amino acid sequences of glycoside hydrolases (GHs) over the last few decades permits the classification of these enzymes based on amino acid sequence similarities. Enzymes with  $\beta$ -galactosidase activity are grouped within the GH1, GH2, GH35, and GH42 families, with the majority of the commercial enzyme preparations originating from either GH2 or GH42. The GH42  $\beta$ -galactosidases of *Lactobacilli* and *Bifidobacteria* are frequently inhibited by high glucose concentrations, and have a lower preference for trans-galactosylation compared with GH2  $\beta$ -galactosidases from the same organisms [208, 286]. Analysis of the primary protein structure of HW- $\beta$ gal1 shows that it belongs to the glycoside hydrolase family 2 and contains the three classical conserved domains that characterise this family, namely the TIM-barrel domain stretching between amino acid (aa) 343-637, the  $\beta$ -galactosidase small chain domain (aa 760-1033), and the sugar binding domain stretching over (aa 55-225) [287]. It also contains the catalytic acid/base region between aa 455-469 with the catalytic Glu469. When the amino acid sequence was blasted against the non-redundant protein sequence database with the BlastP algorithm it showed highest similarity to several *Klebsiella* species (99%); to the best of our knowledge none of these enzymes were biochemically characterised. Relative to the better characterised GH2  $\beta$ -galactosidases, it showed 59% similarity to that of *E. coli*  $\beta$ -galactosidase, and only 32% similarity to that of *Kluyveromyces lactis*, which is one of the dominant commercially utilized GH2  $\beta$ -galactosidases [288, 289]. SignalP 4.1 server analysis furthermore does not predict any signal peptide leading sequences and Prosite analysis predicts a cytosolic location. The results presented here, therefore, suggest that the HW- $\beta$ gal1 enzyme is a cytoplasmically localized enzyme of bacterial origin [290-293].

Compared to HW- $\beta$ gal1, the MS- $\beta$ gluc1 was considerably smaller with a reading frame consisting of 1314 bp translating to 437 amino acid residues and a predicted molecular mass of 49.3 kDa and a theoretical pI of 5.79. This enzyme contained 105 charged amino acid residues (24.0% by frequency) and 224 hydrophobic amino acids (51.3%) and an overall aliphatic index of 73.4%. A Pfam conserved domain analysis showed that MS- $\beta$ gluc1 belonged to the Glycoside Hydrolase family 1 and was in all

probability a  $\beta$ -D-1,4-glucosidase which would account for its high activity on the  $\beta$ -1,4 linked substrate lactose [294]. A protein NCBI blastp analysis against the uniprotkb/swiss-prot databases revealed that MS- $\beta$ gluc1 showed highest similarity (48%) to a  $\beta$ -glucosidase (BglA) belonging to *Thermotoga maritima*, a bacteria which was reviewed in an article that investigated the distant relationship between  $\beta$ -glucosidases of the GH1 family, and other families of  $\beta$ -1,4-glycoside hydrolases [295].

#### 4.4.3 Heterologous expression and purification of the HW- $\beta$ gal1 and MS- $\beta$ gluc1 proteins

The HW- $\beta$ gal1 and MS- $\beta$ gluc1 proteins were cloned in frame into the 6x histidine tagged pRSET protein expression vector and heterologously expressed in the  $\Delta\beta$ -gal DH5 $\alpha$  mutant that contains no coding sequence for the endogenous *E. coli*  $\beta$ -galactosidase polypeptide. This was done for two reasons. Firstly, so as to avoid any co-purification of the endogenous *E. coli*  $\beta$ -galactosidase, thereby skewing results. Secondly, all attempts to express HW- $\beta$ gal1 in strains conventionally used for protein expression, such as BL21 pLysS, failed. It appears that co-expressing HW- $\beta$ gal1 with the endogenous  $\beta$ -galactosidase led to cellular instability, with the strains ultimately dropping the HW- $\beta$ gal1 vector before adequate heterologous expression could be achieved.

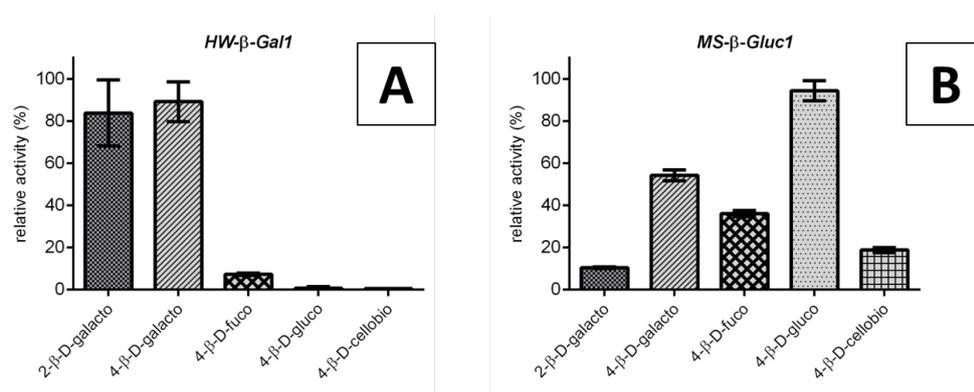


**Figure 4-1** SDS-PAGE protein profiles of fractions collected during the HIS-tag IMAC purification of the HW- $\beta$ gal1 (A) and MS- $\beta$ gluc1 (B) recombinant proteins. Lane (M) is protein MW ladder, lane (1) is  $\Delta\beta$ -gal mutant negative control carrying the empty pRSET A expression vector with no insert, lane (2) is the overnight expression by auto induction of the respective expression vector clones and lane (3) is the purified proteins.

The MS- $\beta$ gluc1 glucosidase however had no complications with protein expression in the conventional strains but was none the less expressed in the deletion mutant. Both the HW- $\beta$ gal1 and MS- $\beta$ gluc1 proteins were expressed overnight by means of catabolic suppression and auto-induction using lactose (which acts as an analogue of IPTG) according to the protocols of Studier *et al.* (2014) [243]. High levels of protein expression were achieved despite the DH5 $\alpha$  host strain used not containing the coding sequence for the T7 RNA polymerase (Fig 4-1). The proteins were purified by IMAC FPLC and eluted over an imidazole gradient. A high purity level was achieved as was visualised by SDS PAGE analysis (Fig 4-1). The collected fractions were dialysed overnight against 0.1xTE buffer (10 mM Tris-HCL pH 8.0, 1 mM EDTA) to rid the protein solution of imidazole. The samples were stored on ice until characterisation.

#### 4.4.4 Substrate specificity of HW- $\beta$ gal1 and MS- $\beta$ gluc1

GH1 and GH2 family proteins are diverse and have numerous overlapping activities including  $\beta$ -galactosidase (EC 3.2.1.23);  $\beta$ -mannosidase (EC 3.2.1.25);  $\beta$ -glucuronidase (EC 3.2.1.31) and  $\alpha$ -L-arabinofuranosidase (EC 3.2.1.55) [107]. It was decided to repeat the enzymatic specificity towards the chromogenic nitrophenyl based substrates that was first assayed with the crude protein extracts. This was done in order to eliminate the possibility of other endogenous *E. coli* proteins possibly accounting for the observed activities, as well as for the purposes of publication. Again, as was the case with the crude extracts, HW- $\beta$ gal1 showed highest activity towards *p*-nitrophenyl- $\beta$ -D-galactopyranoside and *o*-nitrophenyl- $\beta$ -D-galactopyranoside with fewer activity towards *p*-nitrophenyl- $\beta$ -D-fucosidase (Fig 4-2). MS- $\beta$ gluc1 showed highest activity towards 4- $\beta$ -D-glucopyranoside, but also demonstrated significant activity using 4- $\beta$ -D-galactopyranoside, 4- $\beta$ -D-fucopyranoside, 4- $\beta$ -D-cellobiose and 2- $\beta$ -D-galactopyranoside. Neither HW- $\beta$ gal1 and MS- $\beta$ gluc1 showed activity towards the nitrophenyl based substrates of  $\alpha$ -D-galactopyranoside,  $\beta$ -D-maltoside,  $\beta$ -D-xylopyranoside,  $\beta$ -D-glucoronide,  $\alpha$ -D-glucopyranoside,  $\beta$ -L-arabinopyranoside,  $\alpha$ -L-rhamnopyranoside. These results were the same as for the crude protein extracts as discussed in chapter three, and is further confirmation for the successful creation of the  $\Delta\beta$ -gal deletion mutant. The  $\beta$ -glucosidases are grouped in the GH1 and GH3 families, whilst  $\beta$ -galactosidases are grouped in the GH1, GH2, GH35 and GH42 families [103]. Although  $\beta$ -galactosidase activity has frequently been reported for GH1 glucosidases, no  $\beta$ -glucosidase activity has ever been reported for GH2 galactosidases, as is observed for HW- $\beta$ gal1 which showed no statistically significant activity towards either ONP-4- $\beta$ -D-glucopyranoside or ONP-4- $\beta$ -D-cellobiose (Fig 4-2) [190, 286, 287].

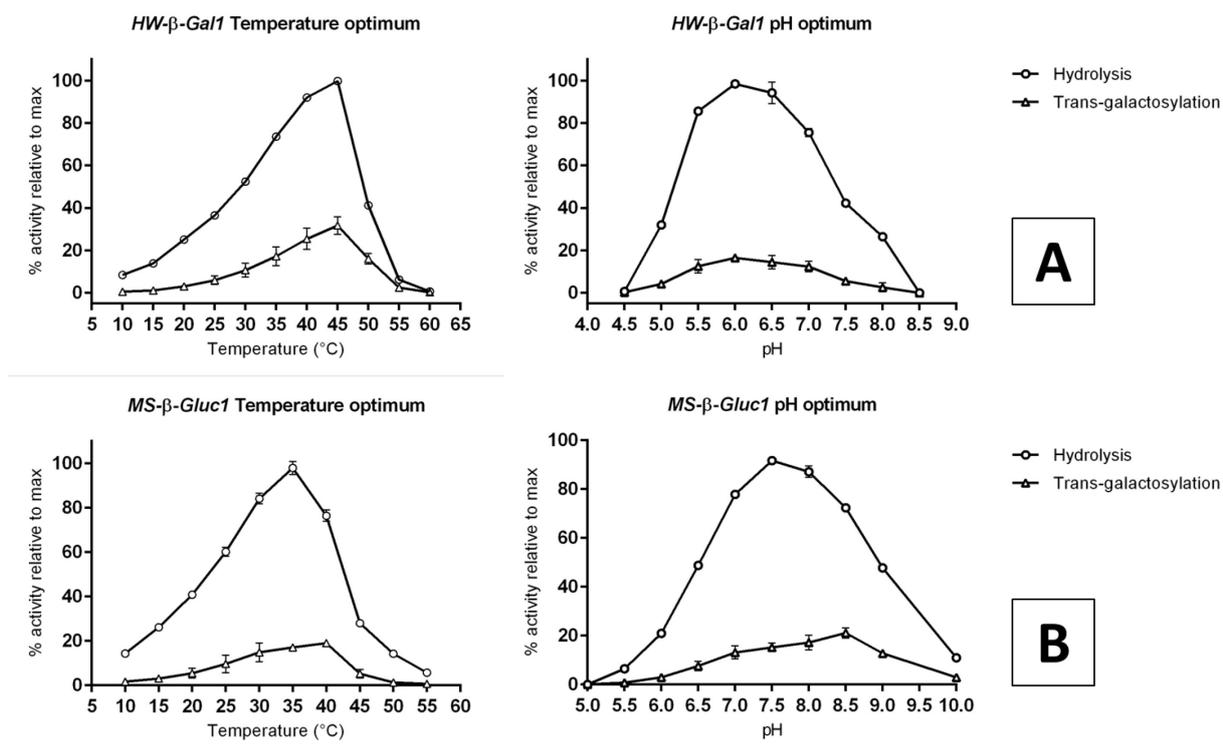


**Figure 4-2** Substrate specificity for the purified HW-βgal1 and MS-βgluc1 expressed relative (%) to the highest activity which was taken to be 100. No activity was detected for ONP-α-D-galactopyranoside, ONP-β-D-maltoside, ONP-β-D-xylopyranoside, ONP-β-D-glucuronide, ONP-α-D-glucopyranoside, ONP-β-D-glucopyranoside, ONP-β-L-arabinopyranoside, ONP-α-L-rhamnopyranoside, ONP-β-D-cellobioside for either HW-βgal1 and MS-βgluc1.

#### 4.4.5 The effects of temperature and pH on HW-βgal1 and MS-βgluc1 activity

In order to determine the optimal conditions for both the hydrolytic and trans-galactosylation reactions of HW-βgal1 and MS-βgluc1 with lactose as substrate, the release of free glucose and galactose were quantified over a broad pH and temperature range. Because there is 1:1 stoichiometric relationship between glucose and galactose within lactose, the hydrolytic reaction can be characterised by the release of free glucose, and the trans-galactosylation can be characterised by the subtracting the free galactose from that of the glucose assayed. Few reports exist that characterise both the hydrolytic and trans-galactosylation simultaneously and, to our knowledge, this is the first quantitative distinction between free glucose and galactose with regard to these two reactions. Previous studies characterising reaction parameters and kinetics were often performed with the aid of either the chromogenic substrate *o*-nitrophenyl-β-D-galactopyranoside (which serves only to characterise the hydrolytic reaction), or by determining release of reducing sugars by 3,5-dinitrosalicylic acid (DNSA), which is unable to distinguish between galactose and glucose. Other methodologies, such as HPLC are also unable to differentiate between the individual monosaccharide contributions of glucose and galactose and any quantitative analysis by means of peak integration merely group them together. It is often not clear therefore, what the relative velocities of the hydrolysis and trans-galactosylation reactions are. For both enzymes in this study the hydrolysis reaction was about 5 fold greater, at the optimal pH and temperatures, than the trans-galactosylation reaction. One mistaken assumption is that the reaction parameters for both reactions are identical. This is often thoughtlessly taken as true in the literature, with no real attempt, up to date, to distinguish between the two reaction. It is evident from the data in this study, that this is not necessarily the case. The temperature optimum for MS-βgluc1 differs between the hydrolytic (35°C) and trans-galactosylation (45°C) reactions (Fig 4-3). In addition the pH optima is

pH 7.5 for hydrolysis and 8.5 for trans-galactosylation. The higher rate of trans-galactosylation at the respective pH and temperature optima is most likely the result of three dimensional conformational changes in the active site which reduce the water potential and favour an increase in the rate of trans-galactosylation [103, 105]. The optimum temperature and pH reactions catalysed by HW- $\beta$ gal1 are however the same for both reactions, at 45°C and pH 6.0. Both HW- $\beta$ gal1 and MS- $\beta$ gluc1 can be characterised as mesophilic with a broad range of discernible activity ranging from 10°C to 55°C. MS- $\beta$ gluc1 however, demonstrates activity over a broader pH range (pH 5.5-10) than HW- $\beta$ -Gal1 (pH 5-8).

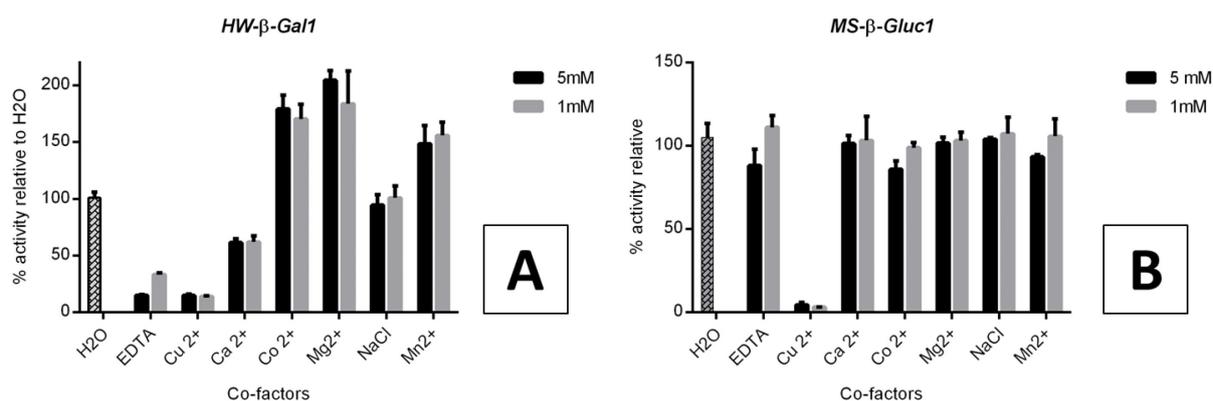


**Figure 4-3** The effect of temperature and pH on both the hydrolytic and trans-galactosylation reaction of HW- $\beta$ gal1 (A) and MS- $\beta$ gluc1 (B). The effect is expressed as a percentage relative to that of the maximal activity. Generally the trans-galactosylation reaction occurs at a slightly higher temperature and pH, than that of the maximal hydrolytic activity.

#### 4.4.6 The effects of co-factors on HW- $\beta$ gal1 and MS- $\beta$ gluc1 activity

Previous reports examining the effects of divalent cations and salts on glycoside hydrolases have indicated that they can influence enzyme activity differentially, either to the positive, negative or both depending on their actual concentration [103, 296, 297]. For this reason the co-factors were tested at final concentrations of both 1 mM and 5 mM. Several divalent cations and salts were tested for their efficacy in either promoting or decreasing the rate of lactose hydrolysis and trans-galactosylation for both the HW- $\beta$ gal1 and MS- $\beta$ gluc1 enzymes. There was however no differential effect of the co-factors

on hydrolysis and trans-galactosylation for either enzyme, and only the results for the hydrolytic reaction are henceforth shown. All co-factors were measured relative to the controls of either ddH<sub>2</sub>O with no additives, or 5 mM EDTA. The activity of lactose hydrolysis in H<sub>2</sub>O was taken as the standard against which all other reactions were expressed against. GH2 enzymes often contain several metal binding sites and between 1-10 mM of Mn<sup>2+</sup> or Mg<sup>2+</sup> are usually required for optimal activity [286, 297]. This is evident in Fig 4-4 that shows that the presence of Co<sup>2+</sup>, Mg<sup>2+</sup> or Mn<sup>2+</sup> led to a nearly twofold increase in activity. Several compounds reduced activity significantly including EDTA (-80%), Cu<sup>2+</sup> (-80%) and Ca<sup>2+</sup> (-40%), all in line with previous reports [264, 288, 298, 299]. For MS-βgluc1 no stimulatory effect was observed for any of the cations or salts indicating that no co-factor is required. Cu<sup>2+</sup> was however a strong inhibitor of the reaction. Only three previous publications have reported an inhibitory effect of Cu<sup>2+</sup> for β-glucosidase, one of which also reported trans-glycosylation activity. It should be noted though that none of these reports used lactose as a substrate [300-302]. No additive effect was observed for either the stimulatory divalent cations or the inhibitors when combined, for either enzyme (data not shown).



**Figure 4-4.** The effects of the various co-factors on the hydrolytic reaction of HW-βgal1 (A) and MS-βgluc1 (B). The influence of activity is expressed as a percentage relative to the activity in ddH<sub>2</sub>O with no added metals or salts, which was taken as a 100%. The influence of the co-factors on the trans-galactosylation reaction is statistically the same as that of the hydrolytic reaction (data not shown).

#### 4.4.7 Reaction kinetics for HW-βgal1 and MS-βgluc1

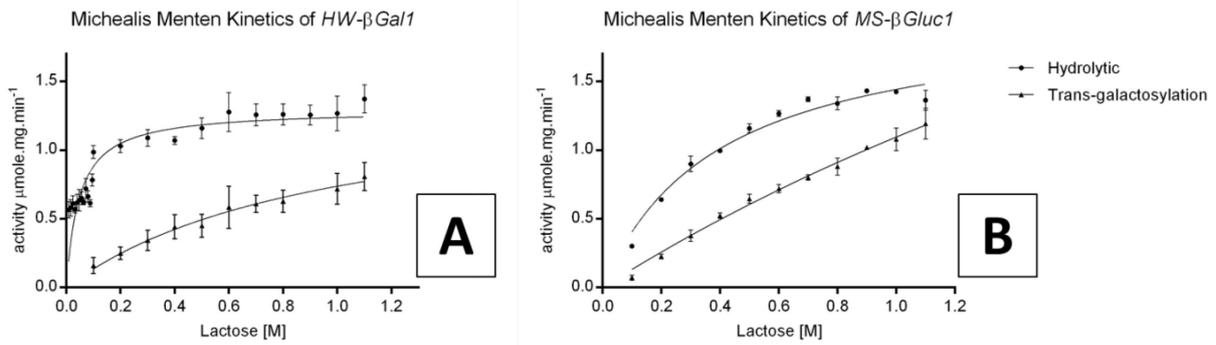
Both GH2 β-glycosidases and GH1 β-glucosidases are retaining enzymes and follow a classical Koshland double-displacement mechanism as was first evidenced in 1951 by Wallenfels, who reported the trans-galactosylation of lactose by means of an implicated glycosyl-enzyme intermediate [303]. Michaelis-Menten kinetics were determined for both the hydrolytic and trans-galactosylation reactions of HW-βgal1 and MS-βgluc1.

The trans-galactosylation reaction of both catalysts demonstrated a linear and non Michaelis-Menten correlation between substrate concentration and activity. This is indicative of the relationship between acceptor concentration, other than H<sub>2</sub>O, and trans-galactosylation. Previous studies showed an increased trans-galactosylation reaction at higher substrate concentrations due to the excess substrate reducing the water potential, thereby increasing the chance of the galactosyl moiety being transferred to another carbohydrate rather than H<sub>2</sub>O [117, 192, 278].

The reaction kinetics for the hydrolytic reaction of HW-βgal1 showed that it had a *K<sub>m</sub>* of 32 mM and a *K<sub>cat</sub>* of 625 s<sup>-1</sup> and a catalytic efficiency (*K<sub>m</sub>/K<sub>cat</sub>*) 19.5 mM<sup>-1</sup>.s<sup>-1</sup>. MS-βgluc1 had a *K<sub>m</sub>* of 41 mM and a *K<sub>cat</sub>* of 111 s<sup>-1</sup> and a catalytic efficiency (*K<sub>m</sub>/K<sub>cat</sub>*) 2.7 mM<sup>-1</sup>.s<sup>-1</sup>. The kinetic parameters of some of the other characterized microbial β-galactosidases with trans-galactosylation activities are summarized in in Table 4-1. HW-βgal1 has an especially high catalytic turnover of 625 s<sup>-1</sup> compared to the other characterised βgalactosidases (Table 4-1). Even though the enzyme's affinity for its substrate is lower than some of the other, GOS is usually industrially synthesised under high lactose concentrations so as to reduce the water potential and favour the trans-galactosylation reaction. This observation, alongside the low *K<sub>m</sub>* and the high *K<sub>cat</sub>* indicates that HW-βgal1 has a very high catalytic efficiency making it suitable for industrial application. MS-βgluc1 too is favourable for industrial application with regards to hydrolysis but ultimately it is the product synthesis capabilities of the trans-galactosylation reaction that has the most industrial relevance.

**Table 4-1** Summary of some of the more well known microbial β-galactosidases and their kinetic parameters, compared to that of HW-βgal1.

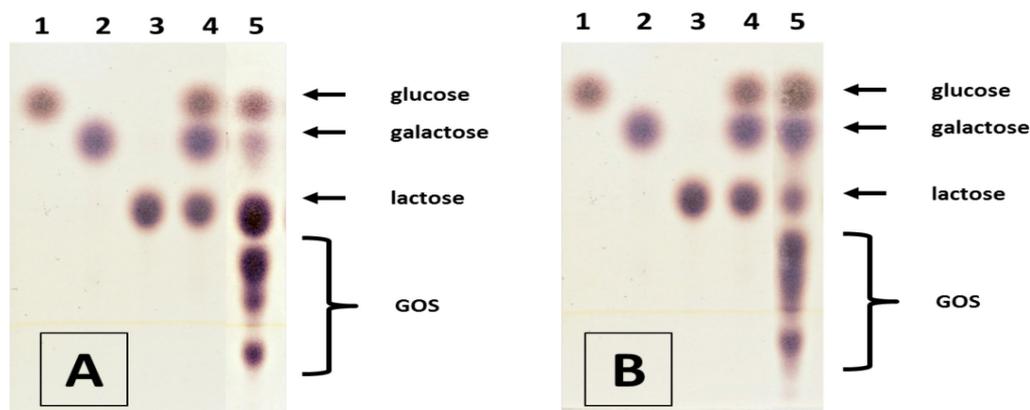
Enzyme source	substrate	<i>K<sub>m</sub></i> (mM)	<i>k<sub>cat</sub></i> (s <sup>-1</sup> )	<i>k<sub>cat</sub>/K<sub>m</sub></i> (s <sup>-1</sup> mM <sup>-1</sup> )	Reference	
<b>β-glucosidases</b>						
MS-βgluc1	lactose	41	111	2.7	This study	
<i>Halothermothrix orenii</i>	lactose	154	231	1.5	Hassan <i>et al.</i> 2014	[280]
<i>Talaromyces thermophiles</i>	lactose	18	79	4.4	Nakkharat <i>et al.</i> 2006	[281]
<b>β-galactosidases</b>						
HW-βgal1	lactose	32	625	19.5	This study	
<i>Arthrobacter psychrolactophilus</i>	lactose	42	3	0.1	Nakagawa <i>et al.</i> 2007	[281]
<i>Caldicellulosiruptor saccharolyticus</i>	lactose	30	42	1.4	Park and Oh 2009	[304]
<i>Lactobacillus acidophilus</i>	lactose	4	50	12.5	Nguyen <i>et al.</i> 2007	[223]
<i>Lactobacillus reuteri</i>	lactose	31	58	1.9	Nguyen <i>et al.</i> 2006	[305]
<i>Saccharopolyspora rectivirgula</i>	lactose	0.7	63	84.0	Nakao <i>et al.</i> 1994	[201]
<i>Sulfolobus solfataricus</i>	lactose	13	1.9	0.1	Pisani <i>et al.</i> 1990	[306]
<i>Lactobacillus delbrueckii</i>	lactose	1.05	45.7	46	Rhimi <i>et al.</i> 2009	[307]



**Figure 4-5.** The Michaelis-Menten kinetic representation of the hydrolytic and trans-galactosylation reactions of HW- $\beta$ gal1 and MS- $\beta$ gluc1 in  $\mu\text{mole.mg}^{-1}.\text{min}^{-1}$  against the substrate concentration of lactose (M)

#### 4.4.8 Product characterisation

The GOS yield, lactose conversion rate and composition of produced GOS vary depending on the enzyme used. It is difficult to analyse the lactose conversion and GOS production due to the inability to easily distinguish disaccharidic GOS from lactose in normal HPLC analysis, unless mass spectrometry is employed. Even with GOS of a higher DP the variety of linkages leads to different elution patterns off the column resulting in either co-elution or fractionated elution patterns. Furthermore, with MS analysis one has to be mindful of formate adduct formation and breakdown products if one is to thoroughly analyse GOS formation [308, 309]. The GOS produced by HW- $\beta$ gal1 and MS- $\beta$ gluc1 were characterised by hydrophobic liquid interaction (HILIC) LC-MS/MS according to the protocols of Hernandez *et al.* (2012). The quantitative standards lactose, 3'-galactosyl-lactose, maltotetraose and maltoheptahose were used to set up calibration curves for the quantitative analysis of synthesized GOS [310].

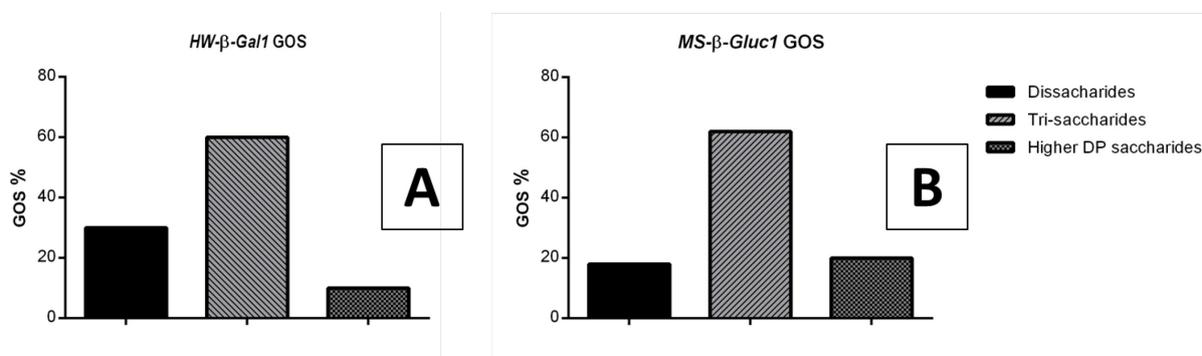


**Figure 4-6** Thin layer chromatographic analysis of the overnight incubation of purified HW- $\beta$ gal1 with a 30% lactose solution in 20 mM Tris-HCl pH 8.3 at 37°C. lane (1) is glucose, (2) is galactose, (3) is lactose, (4) is all three standards combined, and lane (5) is the overnight incubation product with GOS visible.

To date, the highest lactose conversion to GOS (63%) is achieved by a thermostable enzyme from *Bifidobacterium infantis*, whilst the highest production ( $0.318 \text{ g.L}^{-1}$ ) is achieved by *Sulfolobus solfataricus* (Table 4-2). Their thermostability allowed high solubility of lactose and, consequently, a higher reaction velocity [208]. Both HW- $\beta$ gal1 and MS- $\beta$ gluc1 and are mesophilic enzymes that operates optimally at a lower temperature, but still managed to demonstrate an approximate 43% and 45% GOS yield respectively, with an initial lactose substrate concentration of 30% (w/v) (Table 4-2). As can also be seen from thin layer chromatography of the overnight incubation conducted with the purified protein, there appears to be a high conversion of lactose with limited hydrolysis of the GOS product. (Fig 4-6).

**Table 4-2** Summary of some of the GOS forming enzymatic sources, reaction conditions, initial lactose concentration and GOS formed expressed as a percentage (%) of the initial lactose substrate concentration.

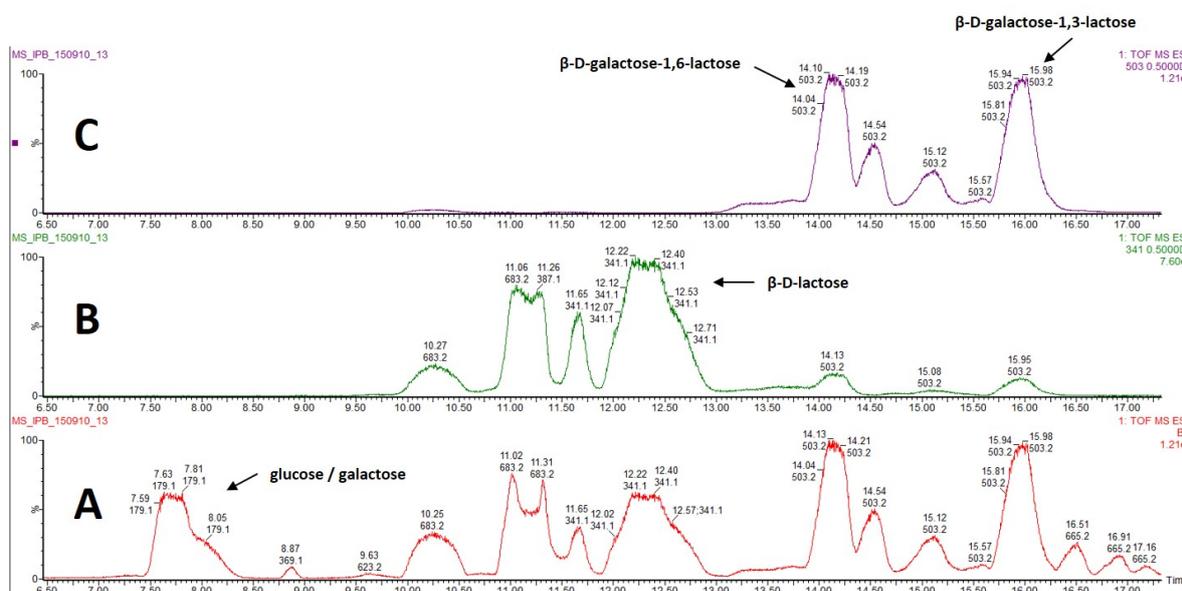
Enzyme source	Temp °C	pH	lactose $\text{g.L}^{-1}$	GOS %	GOS $\text{g.L}^{-1}$	Reference
HW- $\beta$ gal1	37	8.3	300	43	129	This study
MS- $\beta$ gluc1	40	8.5	300	45	135	This study
<i>Bifidobacterium infantis</i>	60	7.5	300	63	189	Hung <i>et al.</i> 2009 [208]
<i>Pyrococcus furiosus</i>	80	5	270	22	59.4	Bruins <i>et al.</i> 2003 [311]
<i>Sulfolobus solfataricus</i>	80	6	600	53	318	Park <i>et al.</i> 2008 [312]
<i>Geobacillus stearothermophilus</i>	73	6.5	180	23	41.4	Placier <i>et al.</i> 2009 [221]
<i>Kluyveromyces lactis</i>	40	7	400	25	100	Chockwasdee <i>et al.</i> 2004 [261]
<i>Bacillus circulans</i>	40	6	460	40	184	Bakken <i>et al.</i> 1992 [185]



**Figure 4-7** The relative % of GOS populations according to degrees of polymerisation as determined by the quantitative analysis of the overnight product by LC-MS. The majority of the GOS were tri-saccharides (60%) for both HW- $\beta$ gal1 and MS- $\beta$ gluc1, whilst approximately 30% and 20% were disaccharides respectively for HW- $\beta$ gal1 and MS- $\beta$ gluc1. Fewer than 10% were higher DP GOS for HW- $\beta$ gal1 and approximately 20% for MS- $\beta$ gluc1.

GOS structures can differ in terms of their saccharide composition, regio-chemistry of the glycosidic linkages and the degree of polymerisation. *In vitro* evidence suggests that prebiotic microbes grow differently on oligosaccharides with different structures [16]. LC-MS/MS analysis demonstrated the synthesis of 11 different di- and tri-saccharides by HW- $\beta$ gal1 and 10 by MS- $\beta$ gluc1 (Fig 4-8). Although it

is difficult to assign identities to all peaks, with regards to linkage analysis, the major populations of di and tri-saccharides could be identified. Quantitative analysis showed that the biggest population of GOS product for HW- $\beta$ gal1 were tri-saccharides making up almost 60% of the GOS whilst approximately 30% were di-saccharides and all the other higher DP population composed fewer than 10% (Fig. 4-7). MS/MS data reveal that the majority of the linkages formed were  $\beta$ -1,6 and that the major tri-saccharide present was  $\beta$ -D-galactose-1,6-lactose. MS- $\beta$ gluc1 synthesised an equal amount of disaccharides and higher DP GOS (combined, 20% of the total), and approximately 60% of the total GOS were made up of tri-saccharides of which there was an equal distribution between  $\beta$ -D-galactose-1,3-lactose and  $\beta$ -D-galactose-1,6-lactose.



**Figure 4-8** The HILIC LC-MS/MS Chromatogram of the final overnight equilibration reaction of the purified protein with lactose. Here the MS- $\beta$ gluc1 reaction is shown as example. The statistical representation for both reactions is shown in Fig. 4-7. (A) is representative of the total chromatogram of the overnight incubation whilst (B) is the mass extracted chromatogram for the di-saccharides and (C) is mass extracted chromatograms for the tri-saccharides.

#### 4.5 Conclusion

The GH2 HW- $\beta$ gal1 and GH1 MS- $\beta$ gluc1 genes isolated from metagenomic libraries were heterologously expressed and their protein products purified by IMAC. The recombinant proteins were biochemically characterised with the substrate lactose. Their kinetic properties demonstrated that the enzymes had a high catalytic efficiency. LC-MS/MS analysis of the overnight reaction demonstrated that both catalyst produced a mixed population of GOS and had a high yield. The HW- $\beta$ gal1 and MS- $\beta$ gluc1 proteins have the potential for industrial application. Contextualised relevance of the data obtained, as well as future research will be discussed in the final chapter six.

## CHAPTER 5: Production of a $\beta$ -2,6 extracellular fructan by a novel metagenome derived levansucrase

*This chapter is partially written in a style aimed at publication. The enzyme characterised here has industrial relevance and the article extracted will be submitted to the journal 'Bioscience, Biotechnology, and Biochemistry' [ISSN: 0916-8451 (print); 1347-6947 (web)]. So as to avoid repetition, some of the 'Material and Methods' and 'Results' covered in chapter three will be omitted*

### 5.1 Abstract

A novel metagenome derived levansucrase gene was cloned and heterologously expressed in *E. coli*. The resultant recombinant protein, SAS-Ls, was purified to apparent homogeneity and biochemically characterised for both its hydrolytic and trans-fructosylation reaction. The two reactions, were found to have the highest  $K_{cat}$ 's,  $7682\text{ s}^{-1}$  and  $5587\text{ s}^{-1}$  respectively, ever reported for a levansucrase. The enzyme had an  $K_m$  of 61 mM for the substrate sucrose and pH and temperature optima of pH 4.5 – 5.0 and 25°C - 35°C respectively. The fructo-oligosaccharides produced by the enzyme were identified by NMR as a levan type with  $\beta$ -2,6 linkages. These data indicate that SAS-LS could be a candidate for industrial application, most notably in the prebiotic market.

### 5.2 Introduction

Fructans are homopolymeric oligo-and polysaccharides comprised of fructose. They differ from each other, not only in chain length, but also in their structural isomerism linkage types [140]. The best characterised fructan is inulin, which consists of a terminal glucose and a series of  $\beta$ -2,1 linked fructosyl units [148]. It is a plant derived fructan, extracted mainly from chicory root, and is employed predominantly as a prebiotic fibre [91, 157]. Prebiotics are non-digestible oligosaccharides and polymers that are selectively fermented by the commensal microbiota of the human gastro intestinal (GI) tract and confer benefits to the hosts, other than nutritional value [14, 266, 267]. The benefits conferred are numerous and range from improving mineral absorption, inhibiting pathogenic adhesion and prevention of colon cancer, to modulating the hosts immune system, regulating metabolism and serum lipid profiles [7, 18, 197, 268].

Recently levan, a bacterial  $\beta$ -2,6 linked fructan, has garnered much scientific attention due to its numerous uses and prospective application. Whereas in plants the functional role of fructans relates to

bridging the temporal gap between resource availability and energy demands, in microorganisms their evolutionary role relates to exopolysaccharide production for microbial biofilm formation in response to environmental and competitive stressors [152]. It is therefore not surprising that levan has favourable rheological properties, such as being soluble in both water and oil, having a high tensile strength (990 psi) and low intrinsic viscosity ( $\eta = 0.18 \text{ dl.g}^{-1}$ ). It is also compatible with most salts, surfactants, acids and bases, has high water and chemical retention and film forming capacity (for review see [313]). These biological properties of levan lends it to be utilized as stabilising or encapsulating agents, emulsifiers, osmoregulators and cryoprotectants [314]. Apart from its industrial application levan also has numerous biomedical properties. These include acting as anti-inflammatory, anti-viral, anti-carcinogenic, fibrinolytic, hypolipidemic, immunostimulatory and anti-oxidant agents as well as being a hyperglycaemic inhibitor [159-164]. It is also being explored for application in nanotechnology based drug delivery systems and wound dressings [155]. Levan type FOS are also sold as a health beneficial artificial sweetener [165] and levan has recently been advanced as a potential prebiotic fibre, much like inulin [144, 169-173]. The prebiotic potential of any complex carbohydrate is dependent on numerous factors including, but not limited to, its degree of polymerization, monomeric composition, linkage type, crystallinity, solubility and its relationship with other substrates [174]. Several studies have concluded that FOS with  $\beta$ -2,6 linkage type selectively enhance *Bifidobacteria* and *Lactobacilli* numbers in the human GI tract, presumably by providing a good growth substrate that consequently lead to the production of lactic and acetic acid (For review see [144]). In addition, certain *Bacteroides* species has been shown to possess extra-cellular endo-levanase activity capable of degrading the longer chain levan into oligosaccharides which are more readily fermented [178]. All the above mentioned literature indicates that levan is an excellent candidate for a prebiotic fibre and research is ongoing [157].

Numerous levansucrases (EC: 2.4.1.10) that produce either levan or levan type FOS have been identified from a wide range of microorganisms including *Zymomonas*, *Streptococcus*, *Bacillus*, *Pseudomonas*, *Erwina*, *Acetobacter* and *Leuconostoc* [155, 313-316]. Despite all the previously described functional applications and potential uses of microbial levan, its actual usage remains limited due to its high production cost and low yield [160, 317, 318]. There is, therefore, a need for the discovery of a novel catalyst that will enhance their industrial production [160]. In this chapter we describe the isolation, identification, cloning and heterologous expression of a novel metagenome derived Glycoside Hydrolase Family 68 levansucrase, henceforth referred to as SAS-Ls. Analysis of its kinetic properties indicate that SAS-Ls demonstrates relatively high trans-fructosylation activity and has potential for the industrial production of microbial levan.

## 5.3 Materials and methods

### 5.3.1 Strains, plasmids and materials

All chemicals and enzymes were purchased from Sigma (St. Louis, MO, USA) unless stated otherwise. All restriction enzymes, T4 DNA ligase, DNA polymerase and corresponding buffers were purchased from Thermofisher Scientific (Massachusetts, USA). The enzymes hexokinase/glucose-6-phosphate-dehydrogenase (HK-G6DH) and phosphoglucose isomerase (PGI) were obtained from Megazyme (Wicklow, Ireland). The lambda ZAP express vector kit and packaging extract was purchased from Agilent technologies (California, USA). The pRSET protein expression vector system was obtained from Life technologies (California, United States).

### 5.3.2 Library construction and screening, clone selection, sequencing and *in silico* analysis

The methods of metagenomic DNA isolation, digestion and library construction were conveyed in great detail in sections 3.2.1 to 3.2.2 on p34 - p40. The methodologies for the screening of the library on minimal media plates were covered in section 3.2.4 on page 42 - p45. The approach to positive clone selection, dereplication, sequencing and bioinformatic analysis were described in section 3.2.5 on page 45.

### 5.3.3 Recombinant protein production and purification

The complete coding sequence for the SAS-Ls gene was PCR amplified and directionally cloned in-frame into the pRSET\_A (Invitrogen) protein expression vector, as described in section 3.2.6 on p45. The SAS-Ls expression vector was transformed into the BI21(pLysS) *E. coli* expression strain and heterologously expressed according to the protocols of Studier *et al.* (2014) [282]. His protocols are an ingenious system that utilises defined and/or complex media (see section 3.2.4 for media preparation) allowing for the maintenance of inducible cells until a metabolic switch triggers auto-induction during the late log phase, thereby producing fully induced high density cultures at saturation. Cultures can be inoculated directly from a frozen glycerol stock, rather than a 'fresh streak' for overnight expression, and routinely yields higher protein production than the tedious IPTG cell monitoring density dependent inducer method [282]. The system is briefly described below.

The inducible heterologous production of proteins by cloned genes utilising the T7 expression system in *E. coli* is a common laboratory practise. However, more often than not, premature unintended induction inadvertently give rise to cultures that produce either poor or variable results. The gene for the T7 RNA polymerase is located on the chromosome of the *E. coli* expression strains, and is under the control of the *lacUV5* promoter. Even the slightest basal induction of *lacUV5* can result in the unwarranted expression of the T7 RNA polymerase, which in turn drives the heterologous protein expression of the cloned gene on the plasmid, via the T7 promoter. This process could redirect the majority of the cellular resources towards heterologous protein production. It could place the cells under untimely stress, ultimately resulting in the plasmid being 'dropped', long before saturation is reached.

The discovery by Grossman *et al.* (1998) that growth in most complex media resulted in a high level of auto-induction as the cell culture approached saturation, defined the problem and allowed for a targeted approach [319]. Several attempts have been made to counter this 'leaky' expression. Firstly, reduction in basal expression can be achieved by introducing small amounts of lysozyme, which inhibit the transcription of the T7 RNA polymerase [320]. Secondly, transcription of the target gene can be reduced by placing a binding site for the *lac* repressor immediately after the T7 promoter (referred to as a *T7lac* promoter) [321]. And thirdly, a more recent strategy was to place the gene for the expression of the T7 polymerase under the control of the *pBAD* promoter, rather than *lacUV5*, which is induced by arabinose and is thought to have lower basal expression (BI21-AI from life Technologies). All three methods have been employed with varying success. However it was the investigations by Studier *et al.* (1991-2005) in how the composition of complex media influenced growth, cell density at saturation and heterologous protein expression that finally produced an explanation for the inadvertent induction, and laid the foundation for the development of defined non-inducing media as well as high density auto-inducing media that allows for the successful and reproducible high-level production of heterologous proteins in *E. coli* utilising the T7 expression system [243, 282, 320, 322].

Amino acids and small residual peptides provide the primary energy source in complex media such as LB and SOC broth. It is derived from the enzymatic digest of the milk protein casein (e.g. Tryptone or NZ-amine), as well as the yeast extract in the broth. Since milk is rich in lactose (an inducer of the T7 expression), variable amounts can be present in the digest depending on the specific batch. The residual amounts of lactose is however not enough to promote appreciable expression during early log-phase growth, but is more than adequate to drive expression as the cell density approaches saturation, particularly at lower rates of aeration [282]. The presence of glucose prevents such induction, but it also causes the solution to acidify when culture saturation is approached [282]. Recent work also found that small amounts of galactose present in the media, also caused unintended induction in BI21(DE3) [323]. This probably occurs because BI21 strains cannot metabolise galactose due to a lack of galacto-kinase, causing galactose (also a weak T7 inducer) to build up intracellularly to a high enough level were it

induces T7 expression [322]. However, the presence of glucose in the solution also prevents induction by galactose as was the case for lactose. This prevention of induction by both lactose and galactose, is based on the premise that as long as glucose is present it remains the preferred carbon source for metabolism. This causes only glucose transporters to be expressed and concomitantly strongly suppresses the expression of operons for the metabolism of lactose, galactose and arabinose, thereby preventing them from entering the cell and inducing T7 expression [324].

This catabolic suppression remains true for as long as there is glucose present in the solution, and as long as the medium doesn't acidify [324]. Glucose is highly effective at preventing induction and even strains that express highly toxic target proteins grow well and maintain a high fraction of inducible cells before induction begins. To grow cultures to high densities ( $OD_{600} \sim 10$ ) and cell concentrations greater than  $10^{10} \cdot \text{mL}^{-1}$ , glucose concentrations must be finely adjusted so as to prevent the pH of the culture falling below pH6. This metabolic balancing of pH requires that the culture be well aerated. A glucose concentration is also required that is just depleted and causes auto-induction during the mid to late log phase, as aeration is falling and induction of target protein is robust. A reliable carbon source in addition to amino acids is required to maintain metabolic activities when glucose becomes depleted during auto-induction. For this purpose glycerol is provided as a good carbon source that does not prevent glucose depletion during growth, glucose exclusion of inducing sugars, or the uptake of inducing sugars upon glucose depletion [243]. Non-inducing and auto-inducing media make production of proteins from cloned genes in *E. coli* reliable and convenient, and are adaptable for applications from small-scale laboratory testing to large-scale screening and protein production.

The protocol in brief: a stab from a SAS-Ls glycerol stock was inoculated into 500 mL of ZYM-5052 auto induction media (1% N-Z-amine, 0.5% yeast extract, 25 mM  $\text{Na}_2\text{HPO}_4$ , 25 mM  $\text{KH}_2\text{PO}_4$ , 50 mM  $\text{NH}_4\text{Cl}$ , 5 mM  $\text{Na}_2\text{SO}_4$ , 2 mM  $\text{MgSO}_4$ , 0.2x metals, 0.5% glycerol, 0.05% glucose, 0.2%  $\alpha$ -lactose) and incubated overnight at 37°C in a 2 L baffled flask with vigorous shaking at approx.  $\sim 300$  rpm (see section 3.2.4 for media preparation).

The cells were spun down at 10,000  $xg$  for 10 min, washed once with 50 mL of ddH<sub>2</sub>O before being pelleted again as before, and reconstituted in 50 ml of protein extraction buffer (50 mM imidazole, 1.5 M NaCl, 20 mM Tris-HCl pH 7.2). The cell suspension was sonicated five times on ice in 30 second burst with 1 minute intervals in-between, with intermittent swirling. The cell debris was pelleted by centrifugation at 12,000  $xg$  for 10 min and the supernatant transferred to a new tube. The supernatant was force-filtered through a 0.22 $\mu\text{m}$  disposable filter connected to a 50 mL syringe, and the sample kept on ice till IMAC purification. The recombinant protein was purified with the aid of a HisTrap HP 5ml column from Life Technologies (California, United States) on an ÄKTA FPLC (GE Life Sciences) protein purification system, according to the manufacturer's recommendations. The proteins were eluted in

500  $\mu$ L fractions over an imidazole gradient (50 mM - 500 mM), and a 50 mL elution volume. The resultant fractions were tested for the release of glucose when incubated with the substrate sucrose, by means of the enzyme linked reduction of NADH, as described below in section 5.3.4. Fractions demonstrating activity were pooled and dialysed overnight against 0.1x TE (1 mM Tris-HCL pH 7.0; 0.1 mM EDTA). Protein concentrations were determined according to the method of Bradford (1976) using bovine serum albumin as the standard, before being analysed on denaturing SDS polyacrylamide gel electrophoresis so as to visually assess the purity [245].

#### 5.3.4 Biochemical characterisation of SAS-Ls

Sucrose was used as substrate for all enzymatic characterisation. For the characterisation of the pH, temperature and co-factor optimums, a stopped assay was used. Zero point one units of SAS-Ls enzyme was incubated in a 500 $\mu$ L final volume of sucrose solution (10%; w/v) for 30 minutes, after which the samples were flash frozen in liquid N<sub>2</sub> before being incubated at 90°C for 5 min to dead-stop the reaction. For the substrate sucrose, one unit of enzyme activity is defined as the amount of protein needed to produce 1  $\mu$ mole of NADH, through the coupled reduction of glucose, after it was liberated from the substrate sucrose, per minute, under the standard assay condition.

Characterisation of optimum pH was performed using acetic acid/sodium acetate (pH 3.5, pH 4.0, pH 4.5, pH 5.0, 50 mM) and sodium phosphate Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> (pH 5.5, pH 6.0, pH 6.5, pH 7.0, pH7.5, 50 mM). For the characterisation of temperature optimum of SAS-Ls, temperatures ranging from 5°C to 55°C were assessed. For co-factors the divalent cations Cu<sup>2+</sup>, Ca<sup>2+</sup>, Co<sup>2+</sup>, Mg<sup>2+</sup>, Mn<sup>2+</sup> as well as NaCl were tested against H<sub>2</sub>O and EDTA in a final concentration of both 1 mM and 5 mM. Michaelis–Menten kinetics were assayed in sucrose concentrations ranging from 8 mM to 300 mM using the optimum pH, temperature and co-factors as determined above, by a continuous assay on a VersaMax ELISA microplate reader (Molecular Devices).

The amount of free glucose released by SAS-Ls after incubation with sucrose was used to characterise the hydrolytic reaction, while the the amount of free fructose subtracted from the amount of free glucose was used to characterise the trans-fructosylation reaction. The liberation of glucose and fructose from sucrose were measured spectrophotometrically in stopped assays by the enzyme linked reduction of NAD<sup>+</sup> to NADH at 340nm utilising hexokinase/glucose-6-phosphate-dehydrogenase (HK-G6DH) and phosphoglucose isomerase (PGI). A standard curve was established for the NADH linked reduction of glucose on the spectrophotometer in a range of 0.02  $\mu$ mole to 0.2  $\mu$ mole per well. Because

glucose and fructose occur at equimolar ratios within the substrate sucrose, the standard curve of glucose can be extrapolated for the measurement of free fructose also.

The protocol in brief: to each well of a clear-bottom UV transparent microtiter plate, 200  $\mu\text{L}$  of assay buffer (100 mM Tris-HCl pH 7.4; 5 mM  $\text{MgCl}_2$ ; 1 mM  $\text{NAD}^+$ ; 1 mM ATP) was added. The samples were added to the wells in a total volume of 50  $\mu\text{L}$ , making the final volume a combined 250  $\mu\text{L}$  per well. A continuous assay was established on the spectrophotometer at 340 nm until baseline was reached, and an endpoint background measurement taken. For the measurement of glucose, 10  $\mu\text{L}$  of a 40x dilution of a HK-G6DH enzyme preparation from Megazyme was added to each well (approx. 1 U hexokinase and 0.5 U G6-DH per well). Again a continuous assay was established and an endpoint measurement taken when baseline was reached (approx. 20 min). The change in  $\text{OD}_{340}$  from the first measurement, is directly proportional to the amount of free glucose in solution, and can be deduced from the standard curve previously established. For the measurement of free fructose, 10  $\mu\text{L}$  of a 50x dilution of PGI was added to the same well (approx. 1 U PGI per well). Again a measurement was taken after baseline was reached. The difference in optical density at 340 nm between the last two endpoint measurements is directly proportional to the amount of free fructose present in solution, and can be calculated from the standard curve.

### 5.3.5 Enzymatic synthesis of levan and carbohydrate analysis

Levan was produced by incubating SAS-Ls ( $0.2 \text{ U}\cdot\text{ml}^{-1}$ ) in a 10 ml sucrose solution (0.3 M sucrose, 50 mM sodium acetate buffer pH 4.6) at  $30^\circ\text{C}$  overnight in an incubator oscillator. The next day the levan was recovered by precipitation with an equal volume of ethanol before being freeze dried. Acid hydrolyses were performed on 5 mg of levan in a 500  $\mu\text{l}$  1 M trifluoroacetic acid solution (TFA) (Merck) at  $100^\circ\text{C}$  for 2 hours, after which the liquid was evaporated and washed twice with high grade methanol to remove residual TFA. The final hydrolysis product was suspended in 100  $\mu\text{l}$  ddH<sub>2</sub>O and the free glucose and fructose ratio determined as described in section 5.3.4. Levan glycosidic linkages were examined by NMR analysis. The sample was suspended in D<sub>2</sub>O ( $2 \text{ mg}\cdot\text{ml}^{-1}$ ) and the spectra were obtained on a Varian VNMRS-500 operating at 100.5 MHz for  $^{13}\text{C}$  and 499.9 MHz for  $^1\text{H}$ . The  $^{13}\text{C}$  spectrum is the accumulation of 448 transients with a  $45^\circ$  pulse width, acquisition time of 1.3 s and a recycle delay of 1 s. Lorentzian broadening of 1.0 Hz was applied before Fourier transformation. The  $^1\text{H}$  spectrum is the accumulation of four transients with a  $45^\circ$  pulse width, acquisition time of 2.0s and a recycle delay of 1s.

## 5.4 Results and discussion

### 5.4.1 Library construction, screening and isolation of SAS-Ls

The results for the metagenomic DNA isolation and SASRI library construction were all discussed in section 3.3.1 on p48 to p53. In chapter 3 however the majority of the focus was on the screening for putative  $\beta$ -galactosidases, and the analysis and screening of the SASRI library weren't discussed in great detail at all, and will consequently be conveyed here briefly.

For the SASRI library a high titre of approx.  $8 \times 10^7$  pfu/ $\mu$ g lambda DNA was obtained with an average insert size of approximately 3.8 kb. After *in vivo* mass excision the converted plasmid library was transformed into the DH5 $\alpha$  and screened on minimal MDA-suc media containing 20% sucrose (w/v) as the sole carbon source. Two hundred and fifteen positive clones, that grew vigorously and produced copious amounts of a clear, sweet smelling and sticky viscous exopolysaccharide were isolated. After plasmid isolation, preliminary restriction fingerprinting and first round sequencing was conducted so as to de-replicate the positive clones. All of the 215 positive clones proved to be related to the same library insert. A representative clone that presumably contained the full insert, was chosen and sequenced in full from both ends. The insert was analysed for the complete open reading frames (ORFs) and preliminary analysed utilising BLASTX so as to assign possible function [244]. A putative Glycoside Hydrolase 68 gene was amplified by PCR with primers containing the necessary restriction sites and directionally cloned into the pRSET\_A protein expression vector. The exopolysaccharide phenotype was reconfirmed by transforming the SAS-Ls pRSET construct back into DH5 $\alpha$  and plating out on media containing sucrose.

### 5.4.2 *In silico* analysis of SAS-Ls

In chapter 3 a rudimentary *in silico* analysis of SAS-Ls was presented whilst a more in-depth review is offered here. The entire SAS-Ls reading frame consists of 1290 bp and encoded for 429 amino acid residues with a predicted molecular mass of 47.15 kDa and a theoretical pI of 4.98. This enzyme contained 81 charged amino acid residues (18.8% by frequency), 227 hydrophobic amino acids (52.9%) and 113 (26.3%) hydrophilic amino acids. The exponential increase in the number of available amino acid sequences of glycoside hydrolases (GHs) over the last few decades permits the classification of these enzymes based on amino acid sequence similarities [113, 325]. BlastP analysis using the non-redundant Genbank peptide database showed highest similarity (77%) to a hypothetical

levansucrase from *Komagataeibacter xylinus*, and when blasted against the UniprotKB/Swiss-Prot SAS-Ls showed only a 58% similarity to that of the already characterised and well known levansucrase from *Zymomonas mobilis* [244, 326]. A Pfam domain analysis revealed that SAS-Ls belongs to the Glycoside Hydrolase Family 68 family of the hydrolase/transferase class of enzymes which contains the frucosyltransferases of levansucrase (EC 2.4.1.10), beta-fructofuranosidase (EC 3.2.1.26) and inulosucrase (EC 2.4.1.9), all of which use sucrose as donor substrate [327].

**Table 5-1** Conserved motifs among bacterial fructosyl transferases. Table adapted from Meng and Fütter [328]

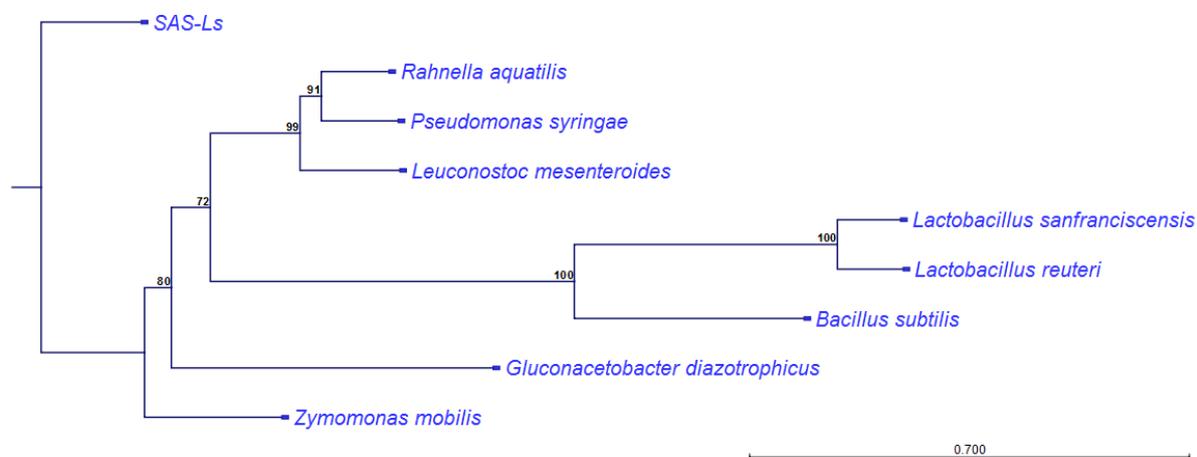
Motif number	Motif designation	Consensus sequence	Position*	Function†
I	QWDTG	(V/L)WD(T/S)(W/M)	41-45	Catalytic nucleophile, conserved in GH68 family
II	EWSGG	(E/Q)W(A/S)G(T/S)	117-121	Sucrose box I, conserved in GH68 family
III	DG	DG	165-166	Involved in acceptor recognition
IV	MFYTS	L(F/Y)T(D/C)	134-138	Sucrose box II and acceptor recognition
V	RDP	RDP	193-195	Catalytic centre, stabilises transition state
VI	YCLFE	Y(M/L)VFE	207-211	Sucrose hydrolysis and trans-fructosylation
VII	PLV	PL(V/I)	270-272	Conserved among fructosyl transferase, unknown function
VIII	DQTER	D(Q/E)(T/I)ER	279-283	Catalytic centre, acid base catalyst, conserved in GH68 family
IX	YLFT	YLF(T/S)	294-297	Conserved among fructosyl transferase, unknown function
X	YRPLN	YXP(L/M)N	327-331	Conserved among fructosyl transferase, unknown function
XI	TYS	TYS	347-349	Orientates nucleophilic residue for trans-fructosylation

\*Motif designation and amino acid positions are relative to the SAS-Ls levansucrase

† Meng and Fütterer 2003 [328], Martínez-Fleites *et al.* (2005) [329], Van Hijum *et al.* (2006) [330]

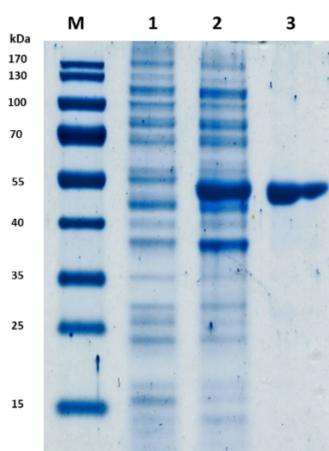
Algorithms within the SignalP 4.1 prediction program indicated the probable presence of a signal peptide on the N-terminus of SAS-Ls [293]. This prediction is supported by the observed exopolysaccharide produced by SAS-Ls in *E. coli* and the fact that many bacteria, both Gram positive and negative, secrete levansucrases [331-334]. Tajima *et al.* (2000) reported a conserved signal peptide motif in the N-terminal domain of Gram negative levansucrases, WT(R/I)ADA(L/M,) which is located between amino acids 8–14 of SAS-Ls (WTIADAL), indicating that SAS-Ls could originate from a Gram negative bacterium [335]. The catalytic domains of microbial levansucrases have 11 conserved sequence motifs that are shown in Table 5-1 [328]. Although there are some differences, most notably in the first and last amino acid of the catalytic nucleophile domain (QWDTG) and also the last amino acid of the sugar box I, SAS-Ls conforms to the spatial arrangements and motifs of the catalytic domains of the eight other best characterised and industrially relevant levansucrases (Figure 5-1) [329, 330]. A phylogenetic tree analysis was performed for these eight levansucrases together with that of SAS-Ls, by the neighbour-joining method using the Jukes-Cantor protein distance measure demonstrating that SAS-Ls

is furthest removed from Gram positive bacteria (e.g. *B. subtilis* and *G. diazotrophicus*) and its closest relative is the Gram negative bacterial levansucrase of *Z. mobilis* (Figure 5-1). Few crystallography studies has been conducted on levansucrases, therefore further studies into SAS-Ls, its catalytic nucleophile and its crystal structure could be a significant contribution to the field [329, 336].



**Figure 5-1** Phylogenetic tree of microbial levansucrases. Analysis was performed by the neighbour-joining method using the Jukes-Cantor protein distance measure. The number at the nodes indicates confidence bootstrap percentages of 500 repeats. Sequences are represented by the names of the organisms they derive from except for SAS-Ls. All sequences were obtained from the UniprotKB/Swiss-Prot. database and were validated. The accession numbers are as follow *Bacillus subtilis* (P05655), *Gluconacetobacter diazotrophicus* (Q43998), *Pseudomonas syringae* (Q88BN6), *Rahnella aquatilis* (O54435), *Leuconostoc mesenteroides* (Q51S34), *Lactobacillus reuteri* (Q8GGV4), *Lactobacillus sanfranciscensis* (Q70XJ9), *Zymomonas mobilis* (Q60114).

### 5.4.3 Heterologous expression and purification of SAS-Ls

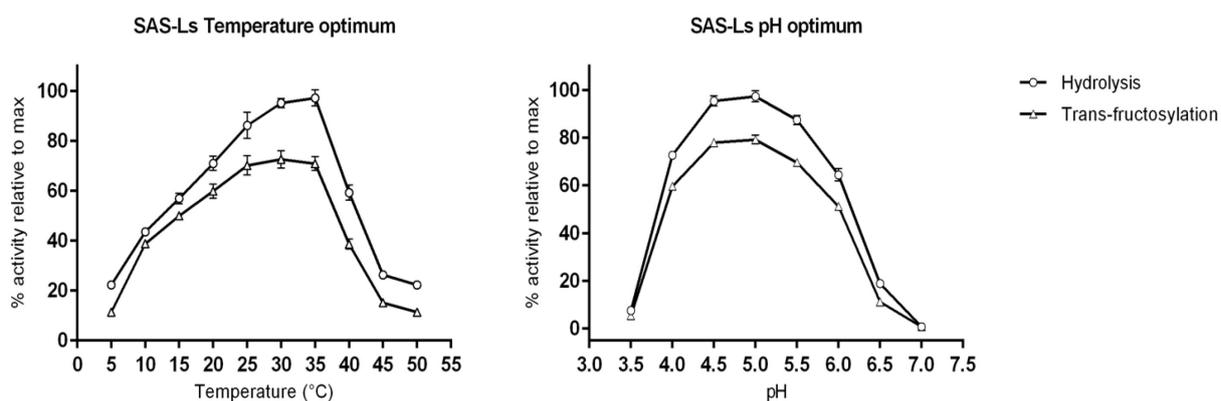


**Figure 5-2** SDS-PAGE protein profiles of fractions collected during the HIS-tag IMAC purification of the SAS-Ls protein. Lane (M) is protein MW ladder, lane (1) is BI21 pLysS with the empty pRSET\_A vector, lane (2) is the overnight expression by auto induction of the Hiss-SAS-Ls in BI21-pLysS, and lane (3) is the purified proteins.

The SAS-Ls gene was heterologously expressed in BL21(pLysS) according to the methods of Studier *et al.* (2014), as described in section 5.3.3 on p85 [243]. Copious amounts of protein was obtained, after inoculation from a frozen glycerol stock and overnight incubation, by means of catabolic suppression and auto-induction using lactose (which acts as an analogue of IPTG). The signal peptide was not removed but this showed no notable effect on cytosolic accumulation of recombinant protein as seen from the SDS-PAGE analysis (Figure 5-2). The histidine tagged polypeptide was purified by IMAC FPLC and eluted over an imidazole gradient. Fractions were tested for levansucrase activity by measuring the hydrolytic release of glucose from sucrose. A high purity level was achieved as was visualised by SDS PAGE analysis (Fig 5-2). The collected fractions were dialysed overnight in 0.1xTE buffer (10 mM Tris-HCL pH 8.0, 1 mM EDTA) to rid the protein solution of imidazole. The protein was stored on ice at 4°C for further analysis and showed no loss of activity over a 14 day period on ice in the 4°C walk in room.

#### 5.4.4 The effects of temperature and pH on SAS-Ls activity

Levansucrase can catalyse two distinct reactions depending on the fructosyl acceptor, namely (i) hydrolysis when H<sub>2</sub>O is the acceptor and (ii), trans-fructosylation/polymerisation when any other reducing molecule, including fructose acts as an acceptor [155]. Levansucrase generally shows different temperature optima for the hydrolysis and trans-fructosylation reactions. For example, the protein from *Pseudomonas syringae* pv. phasiolicola demonstrates marked optima difference of 18°C and 60°C for hydrolysis and polymerization respectively [337]. On the other hand the levansucrase for *Zymomonas mobilis*, showed maximal polymerization at 30°C whilst the highest rate of hydrolysis was at 45-50°C [338]. In general, microbial levansucrases seems to exhibit optimum levan formation activity at lower temperatures, usually below 50°C (for review see [155]). This is also true for SAS-Ls where the rates of hydrolysis and trans-fructosylation are similar at temperatures lower than 20°C, being nearly identical at 10°C (Fig 5-3). Maximum hydrolytic activity for SAS-Ls was at 35°C after which a sharp decline in activity for both the hydrolytic and trans-fructosylation reactions was observed. Overall SAS-Ls activity exhibited a broad temperature range of 12-37°C where activity was above 50% of the maximum activity for both reactions, a property that would render it suitable for industrial applications (Table 5-2) [155]. With regards to pH, SAS-Ls also shows a broad reaction viability ranging from pH 3.7-6.0 where activity is above 50% of maximal for both the hydrolytic and trans-fructosylation reaction. This is similar for previously reported levansucrase whom all tend to be more active under acidic conditions than alkaline (Table 5-2). The hydrolytic reaction was greater than the trans-fructosylation one at all temperatures and pH's.

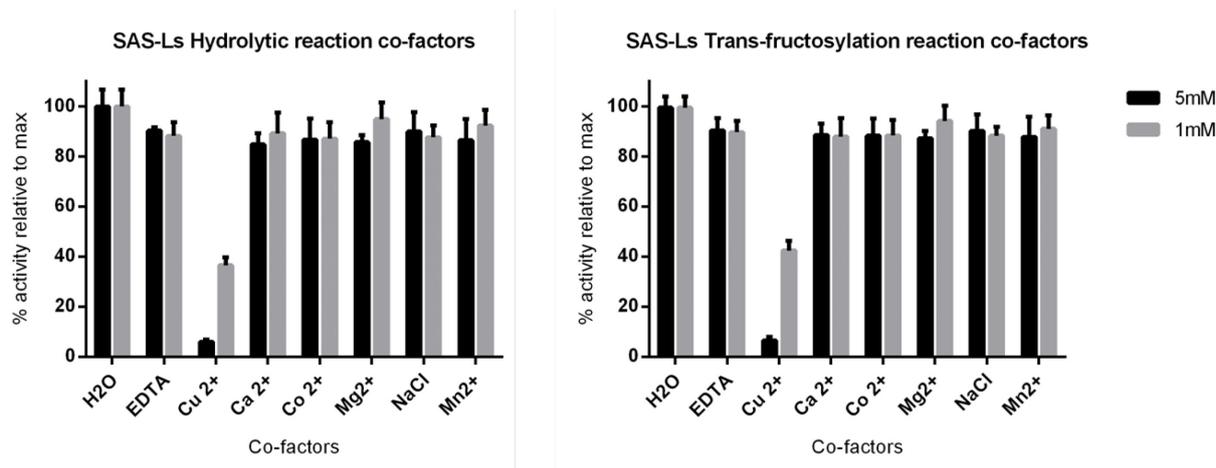


**Figure 5-3** The effect of temperature and pH on both the hydrolytic and trans-fructosylation reaction of SAS-Ls. The effect is expressed as a percentage relative to that of the maximal activity which was taken to be 100.

#### 5.4.5 The effects of co-factors on SAS-Ls activity

All co-factors were measured in conjunction with the controls of pure ddH<sub>2</sub>O or 10 mM EDTA solution. The effect of various co-factors were tested at final concentrations of 1 mM and 5 mM, for both the hydrolysis and trans-fructosylation reaction. The only metal ion to have an effect was Cu<sup>2+</sup>, which inhibited both reactions by more than 50% at a 1 mM concentration and more than 90% at a 5 mM concentration (Figure 5-4). A similar effect has also been reported for the levansucrases from *Zymomonas mobilis* and *Bacillus subtilis* [339, 340].

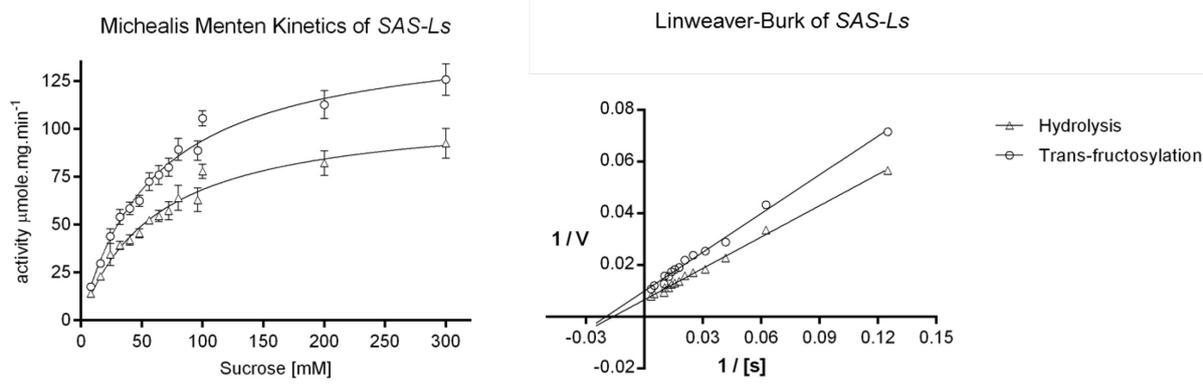
Szwengiel *et al.* (2007) reported that the presence of Mn<sup>2+</sup> increased the transferase activity of *Bacillus subtilis* levansucrase by 100% whilst lowering the hydrolytic reaction by 80% [163, 341]. Other compounds have also been shown to have a stimulatory effect on the activities of various levansucrases. Fe<sup>2+</sup> had an 4 fold increase in both reactions on the thermostable *Bacillus sp. TH4-2* [342], Co<sup>2+</sup> has been shown to have a slight enhancing effect on the levansucrase of *Gluconacetobacter diazotrophicus* for both reactions [343]. Ca<sup>2+</sup> has also been shown to play an important structural role in the levansucrases of both *Lactobacillus reuteri* and *Bacillus subtilis* [344, 345]. None however had any effect on SAS-Ls activity (Figure 5-4), which makes SAS-Ls rather unique in its co-factor independence.



**Figure 5-4** The effects of the various co-factors on the hydrolytic and trans-fructosylation reaction of SAS-Ls. The influence of activity is expressed as a percentage relative to the activity in ddH<sub>2</sub>O with no added metals or salts, which was taken as a 100%. The influence of the co-factors on the trans-fructosylation reaction is statistically the same as that of the hydrolytic reaction (data not shown).

#### 5.4.6 Reaction kinetics of SAS-Ls

The initial reaction rates of both the hydrolytic and trans-fructosylation reactions were determined at 35°C over a sucrose concentration range of 0–300 mM, where the enzyme behaviour adhered to Michaelis–Menten kinetics, as indicated by the linearity of the corresponding Lineweaver–Burk double reciprocal plots (Figure 5-5). The values of  $K_m$ ,  $K_{cat}$ , and the catalytic efficiency ( $K_{cat}/K_m$ ) were determined by linear regression of the Lineweaver-Burk plot as well as integrated non-linear regression using the GraphPad Prism software (Figure 5-5). The  $K_m$  value for hydrolysis and trans-fructosylation were calculated as  $61.07 \pm 3.59$  mM and  $61.03 \pm 4.49$  mM, the  $K_{cat}$ 's were  $7682 \pm 190.2$  s<sup>-1</sup> and  $5587 \pm 173.1$  s<sup>-1</sup>, and the catalytic efficiencies were  $7.55 \times 10^6$  M<sup>-1</sup>min<sup>-1</sup> and  $5.49 \times 10^6$  M<sup>-1</sup>min<sup>-1</sup>, respectively. Table 5-2 compares the catalytic characteristics of the other microbial levansucrases that have been characterised with that of SAS-Ls. Although SAS-Ls exhibits a relatively high  $K_m$  compared with the other characterised levansucrases, it has an extremely high catalytic turnover which at  $7682$  s<sup>-1</sup> is a 3.38 fold increase over that of the enzyme from *Bacillus megatarum*. In fact the catalytic efficiency of SAS-Ls ( $K_{cat}/K_m$ ) at  $7.55 \times 10^6$  M<sup>-1</sup>min<sup>-1</sup>, is only exceeded by that of *Bacillus megatarum* ( $5.93 \times 10^7$  M<sup>-1</sup>min<sup>-1</sup>) due to the lower  $K_m$  of the *B. megatarum* levansucrase (2.3 mM compared to 61 mM of SAS-Ls). This, however is of little consequence as industrial application of levansucrases rarely occurs at such low concentrations [346].



**Figure 5-5** The Michaelis Menten kinetic representation of the hydrolytic and trans-fructosylation reactions of SAS-Ls in  $\mu\text{mole.mg}^{-1}.\text{min}^{-1}$  against the substrate concentration of sucrose (mM), together with the Lineweaver-Burk double reciprocal plot.

**Table 5-2** Summary of some of the more well known microbial levansucrases and their kinetic parameters, compared to that of SAS-Ls.

Organism/source	Accession number	$K_m$ (mM)	$K_{cat}$ ( $s^{-1}$ )	cat efficiency ( $K_{cat}/K_m$ ) $M^{-1}.\text{min}^{-1}$	$V_{\text{max}}$	pH	Temp	Reference
SAS-Ls	NA	61.07	7682	$7.55 \times 10^6$	151.3	4.5-6	25-35	This study
<i>Bacillus amyloliquefaciens</i>	NA	47.81	NA	NA	10.8	6-6.2	25	[334, 347]
<i>Bacillus megatarum</i>	NA	2.3	2272	$5.93 \times 10^7$	NA	6-7	45	[346]
<i>Bacillus subtilis</i>	P05655	40	164.6	$2.47 \times 10^5$	203	5.2-6	30	[161]
<i>G. diazotrophicus</i>	Q43998	11.4	1	$5.26 \times 10^3$	NA	5	30-40	[348]
<i>Lactobacillus panis</i>	NA	22.5	NA	NA	NA	4-4.6	45-50	[349]
<i>Leuconostoc mesenteroides</i>	Q5IS34	26.6	NA	NA	126.6	6.2	28-30	[350]
<i>Pantoea agglomerans</i>	NA	28	NA	NA	NA	6	NA	[351]
<i>Pseudomonas savastanoi</i>	NA	160	NA	NA	298	6.5	37-60	[337]
<i>Rahnella aquatilis</i>	O54435	50	NA	NA	731	4-6	55-60	[352]
<i>Zymomonas mobilis</i>	Q60114	40	461	$6.92 \times 10^5$	NA	4-6	37-50	[353, 354]

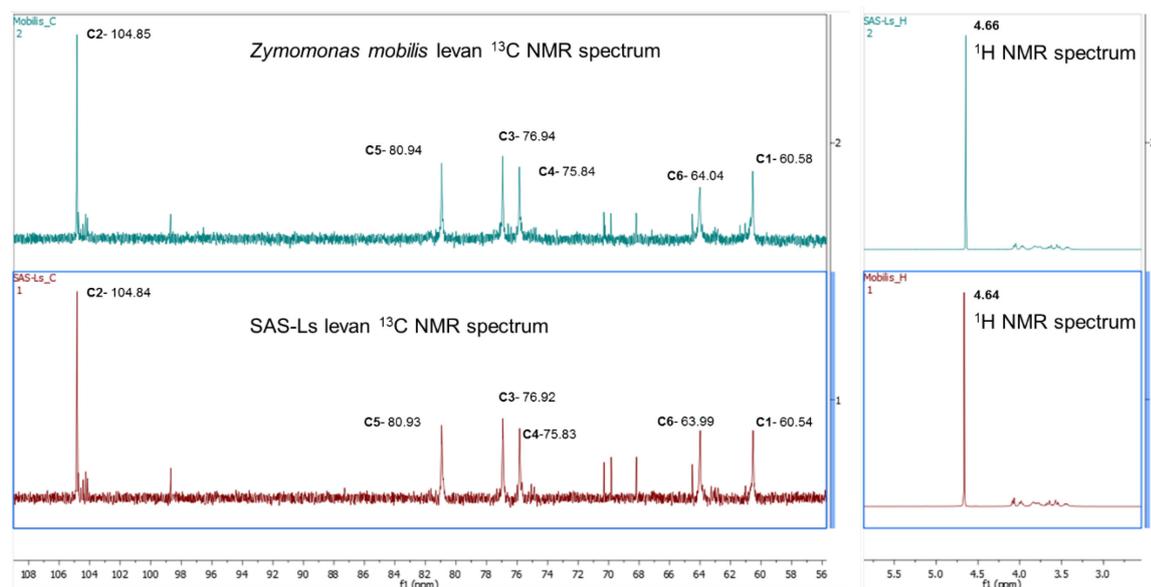
#### 5.4.7 NMR analysis and product characterisation

To characterise the structure of the polymer/oligosaccharide, the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectrums were determined for the product of the overnight incubation of SAS-Ls with sucrose. As a positive control the commercially available levan from *Z. mobilis* was used (Fig 5-6). Table 5-3 summarises the main resonances of carbon signals (in ppm) and is compared to that of other previously characterised levans [160, 355-357]. The resonances for C1 (60.5), C2 (104.8), C3 (76.9), C4 (75.8), C5 (80.9) and C6 (64.04) (Fig 5-3) aligns with that of the levan of *Z. mobilis* as originally characterised and assigned by Han and

Clarke (1990) [358]. The large C2 resonance is indicative of a  $\beta$ -fructofuranose, and overall, the data strongly indicates that the SAS-Ls product is a levan with a  $[-\rightarrow 6\text{-}\beta\text{-D-Fruc-2}\rightarrow]_n$  structure. The nearly identical peak pattern of the  $^1\text{H}$  spectra's further acts as confirmation.

**Table 5-3** Chemical  $^{13}\text{C}$ NMR shifts of microbial levans from various sources compared to the levan of SAS-Ls

Carbon atom	Chemical shifts of Levan $^{13}\text{C}$ NMR ppm					
	SAS-Ls	<i>P. Fluoresces</i> [357]	<i>Z. Mobilis</i> [357]	<i>L. Reuteri</i> [356]	<i>B. Subtilis</i> [355]	<i>B. Polymyxa</i> [355]
C-1	60.54	60.428	60.761	61.4	60.1	60.7
C-2	104.84	104.696	104.641	105.6	104.4	104.2
C-3	76.92	76.77	77.683	77.8	76.5	77
C-4	75.83	75.783	75.754	76.7	75.4	75.7
C-5	80.93	80.88	80.783	81.7	80.5	80.5
C-6	64.04	63.987	63.957	64.8	63.6	63.6



**Figure 5-6** The  $^{13}\text{C}$  and  $^1\text{H}$  spectra of the levan produced by SAS-Ls compared to that of the commercially available *Zymomonas mobilis* levan. As can be seen both spectra are nearly identical for both  $^{13}\text{C}$  and  $^1\text{H}$ .

To estimate the average approximate chain length of the levan produced by SAS-Ls a reducing sugar assay was conducted on both hydrolysed and un-hydrolysed levan. This led to an estimation of the average chain length being  $13.2 \pm 3.2$  which indicates that SAS-Ls produces fructo-oligosaccharides rather than longer chain polymers. This is in line with previous research which found that levans produced by enzymatic systems generally consists of 10-12 residues as compared to a whole cell system where the polymer can be comprised of several million residues [160, 359]. Furthermore Gram negative bacterial levansucrases have been associated with the production of levan type fructo-oligosaccharides

as compared to the Gram positive levansucrases which are more often found to synthesise higher DP levan [360]. Although both levan polymer and levan FOS demonstrate interesting properties, levan FOS are often preferred over a levan polymer, especially in regards to prebiotic dietary supplementation [157]. This has led to the design of bi-enzymatic systems which have both a levansucrase, and an endo-levanase/inulase which generates FOS from the longer chain levan [169, 361]. A levansucrase from *Zymomonas mobilis* has been demonstrated to produce either levan polymer or FOS, depending on reaction conditions and, due to the similarity of that protein with SAS-Ls, it certainly warrants further investigation to see if the same is true for SAS-Ls [314]

## 5.5 Conclusion

In this study a novel metagenome derived levansucrase (SAS-Ls) was successfully isolated, cloned, heterologously expressed and biochemically characterised. SAS-Ls has a high and desirable catalytic turnover making it an excellent candidate for industrial application. SAS-Ls was further shown to produce levan type fructo-oligosaccharides rather than longer chain levan polymer. Because of the low extraction yield of levan and FOS from natural sources, biocatalytic approaches based on trans-fructosylation is an attractive alternative for its synthesis. The SAS-Ls levansucrase could certainly aid in this regard. Future research and alternative applications of SAS-Ls and its enzymatic products will be further discussed in the final chapter.

## CHAPTER 6: General discussion and future prospects

### 6.1 The metagenomic approach

#### 6.1.1 Origins and context of metagenomic analysis for novel product discovery

It is perhaps prudent to give a brief contextualised review on the origins and evolution of the metagenomic concept, up until its application in functional screening, so as to aid in the discussion. Metagenomics has its roots in bacterial classification and description, and to a considerable extent this still contributes to the field of study today. For nearly 300 years the study of microorganisms were based primarily on morphological features as well as selective growth [362]. Woese *et al.* (1977) was the first to propose the use of ribosomal RNA gene sequences as a universal molecular marker for life's classification [363]. In the very same year Sanger *et al.* (1977) first published his automated sequencing method. These two procedures, together with the development and controversial patenting of the PCR procedure by Kary Mullis in 1983, revolutionised the classification of microbial life in the coming years [364, 365]. All of this phylogenetic analysis were however conducted on a single species cultivation basis, making it difficult to study collective communities. In fact, based on real time microscopic cell counts in conjunction with 16s rRNA phylogenetic profiling, it was established that approx. 1% or less of all the microbial organisms can be recovered from soil samples through standard laboratory cultivation [366, 367]. Especially certain conglomerates of microbes such as those present in soil, have been shown to contain a collective gene pool of several thousand different microbial genomes per gram of soil [366, 368, 369]. This represents a treasure trove of genetic material and logic dictates that scientist would haste in their endeavour to investigate such environments and the genetic data it contained, but for the pre-metagenomic era, this large gene pool was mostly inaccessible.

In the quest to overcome this impediment, phylogenetic classification was the initial driving force, and in 1990 Giovannoni *et al.* described the first microbial community study by 16S rRNA analysis [370]. For about a decade though, albeit due to technical limitations, focus was lost on the rest of the metagenome as researchers zoomed in on phylogenetic markers and specific conserved genes of interest. Gradually, with the advances in molecular techniques especially with regards to cloning and heterologous gene expression, a window of opportunity was presented, for the attention to shift from the phylogenetic complexity to that of the functional complexity of the metagenome. By transposing these new techniques onto the microbial diversity interpretation, access was granted to this new 'uncultured world'. This resulted in the emphasis shifting from classification to that of discovery, and in 1998

Handelsman *et al.* coined the term ‘metagenomic analysis’ describing it as the study of the collection of all the genomes of a microbial community from a specific environment, with the the focus on a new frontier for natural product discovery [371]. Products such as ‘terragines’ discovered from *Streptomyces lividians* in 2000, and other broad-spectrum antibiotics were all cloned from soil-DNA libraries, heralding in this new era of biotechnology [372, 373]. Since the early 2000s functional metagenomics progressed to its prominent role, all while co-evolving with other technological advances such as the sequencing revolution. These all present their own unique set of challenges, technical limitations and inherent biases necessitating the need for a tailored metagenomic approach for each individual case. It is in this context that the the novel products discovered through a functional metagenomic approach, as presented earlier in this dissertation, will be discussed. In general however, the metagenomic approach can be broken down into several strategic stages that is best summarised in Figure 6-1, and the construction and screening of the libraries generated in this dissertation will briefly be discussed as such.

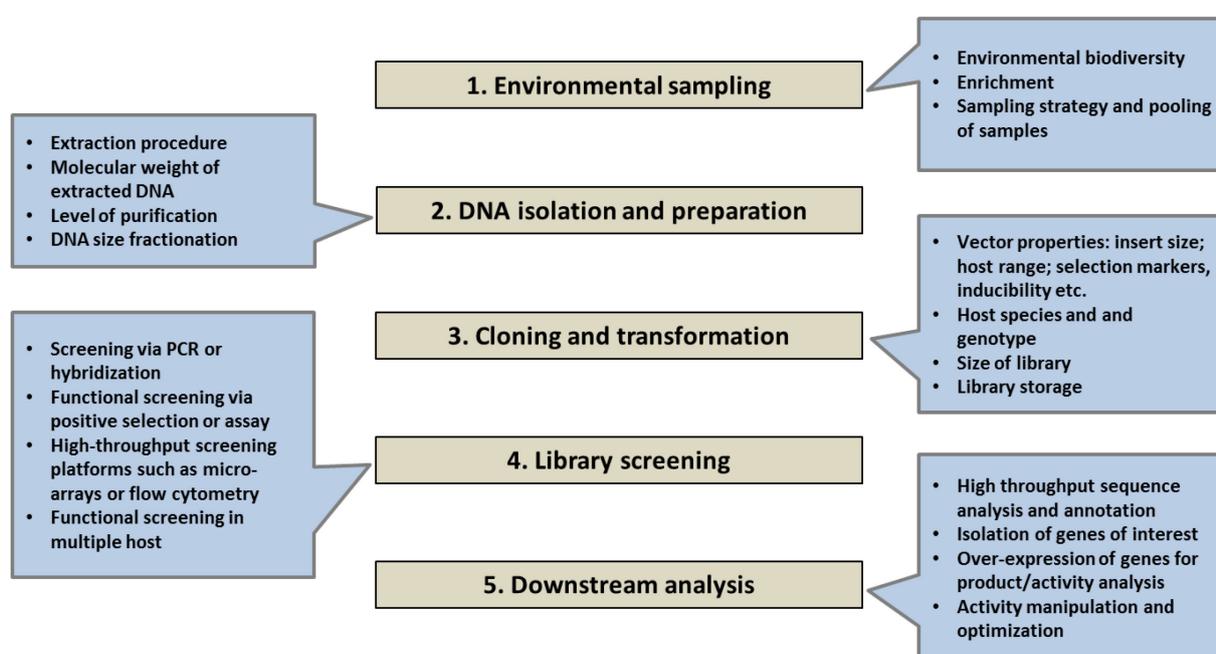


Figure 6-1 The five primary steps involved in the general metagenomic approach is presented in the grey central boxes. The blue peripheral call-out boxes represent key issues and decisions relevant to each of the central steps. Figure adapted from Kowalchuk *et al.* (2007) [374].

### 6.1.2 Environmental sampling and diversity assessment

With the environmental sampling, a multitude of factors will influence the eventual site chosen, all relating to the effect the environment exerts on the microbial community. The vast majority of our biosphere’s genetic and metabolic miscellany are contained within the microbial communities of the

world. These microbial groups colonise habitats that can be complex in nature (e.g. soils and sediments' dregs), and also extreme in both location (e.g. deep sea marine sulphur hot vents, arctic tundra, cavernous subterranean) and environmental parameters (e.g. temperature, pH, salinity, nutritional limitation, meteorological pressure). The physiological and genetic properties of the microorganisms that reside in these and other environments are habitually adapted to them. Complex and extreme environments often yield diverse and/or extreme enzymes [375, 376].

It therefore makes sense, that in consideration of the appropriate sampling site, these environmental parameters be of paramount importance to the decision making, as it can be crucial to the success when conducting the metagenomic screening for specific enzymes (Figure 6-1) [377]. To this extent the Hotwood, SASRI and Gansbay represents environments that at least temporarily are exposed to high temperatures. The interest in high temperature tolerant biocatalyst relates to the linear relationship between substrate concentration and enzymatic competence for oligosaccharide and/or polymer formation, and the fact that the substrates (i.e. lactose and sucrose) have higher saturation levels at higher temperatures. None of the enzymes isolated and characterised were however thermophilic in nature, but had high activity in the upper mesophilic range.

The foremost selection parameter in choosing the sampling sites were however that of substrate availability. Genomic diverse environments such as soil and sediment, owe their microbial diversity to the heterogenous nature of the biotic and abiotic factors. Samples for metagenomic environments are often collected from natural environments as was the case for the Gansbay library, and to a certain extent the milk-soil library (MS) as well. Often environments are pre-treated prior to sampling, but this can have a significant outcome on the metagenomic DNA isolated. Spiking with certain substrates can confer growth benefits to certain microbial species which share certain physiological traits. Care should be taken though as this can severely reduce the microbial diversity in this *quid pro quo* approach. This is also the case for where natural enrichment (high galactose content of degrading kelp) of the environmental sample might have occurred, or where anthropogenic factors (milk leakage) influenced microbial diversity as was the case for Gansbay and MS libraries respectively [378, 379]. Both the Hotwood (saw mill run-off) and SASRI libraries (sugarcane refinery run-off) represents artificial environments that were enriched to the extreme for their respective substrates.

None of the libraries were analysed for their microbial diversity, to a certain extent this was due to time and cost constraints, but also by reason of the fact that the microbial diversity is more or less expected to be restricted as a consequence of the enriched environments as discussed above. This hold true to a lesser extent for the Gansbay and MS libraries. Non the less, a reduced microbial diversity could not be interpreted as a qualitative qualifier for the constructed libraries. Microbial 16S rRNA diversity analysis would however have been insightful from a phylogenetic perspective so as to

recognize the dominant species, as this might be relevant to the search for the novel catalyst but this information can be ascertained to a limited, albeit skewed, extent from the *in silico* analysis of the isolated positive clones itself.

Instead of targeting certain specialised and enriched environments, enzyme activities may also be found in atypical environments. Against expectancy, enzymes don't necessarily share similar physiological properties with the habitat from which they originate from. One example is the interesting discovery of two novel amylases from a acid mine drainage site, which shared no homology to any of the known sequences on public databases [375]. It is for this very reason that the libraries were cross screened on the different media containing either lactose or sucrose, and indeed  $\beta$ -galactosidases were isolated from the SASRI library but this in all probability relates to the ubiquitous nature of the  $\beta$ -galactosidase gene.

### 6.1.3 Metagenomic DNA isolation and preparation

The outcome of any metagenomic library is first and foremost dependent on the quality of the nucleic acids isolated. The nucleic acids should be suitable for downstream molecular procedures and applications including enzymatic action, cloning, transformation, hybridisation and direct sequencing. Nucleic acids purified are qualified on three basic characteristics namely, (i) purity and the absence of sample site contaminants, (ii) fragment size and degree of shearing, and (iii) and concentration. Particularly DNA isolated from soil are known to be severely contaminated by humic impurities that harshly impede molecular biological applications [380].

Numerous protocols exist for the extraction of metagenomic DNA, and various commercial products are available. The nature of the sample can however differ greatly and no one solution will fit all. Water samples can be filtered through various membranes so as to concentrate the sample before nucleic acid extraction. Generally two approaches are taken with regard to solid samples namely, (i) direct lysis, or (ii) the preceding separation of matrix adherent cells. Separation of the cells from the matrix is achieved through either mechanical or chemical treatment, thereby releasing the microbial cells from the particles before being concentrated by various means. Direct lysis extract all the nucleic acid from all the biological material within the sample, whereas separation techniques only lyse cells that were separated and concentrated [381, 382]. Comparatively and generally, direct lysis yields between 10-100 times more nucleic acid, but the co-extraction of extra-cellular DNA as well as DNA from degrading plant material and such, probably contribute to this significantly. On a diversity level no real differences were observed between the two general approaches [383, 384].

The actual lysis of the cells can be accomplished through mechanical shearing, chemical disruption or enzymatic release. In this dissertation the protocol developed simultaneously took care of ridding the extracted DNA of high molecular weight polysaccharide contaminants, as well as preventing any enzymatic degradation. The protocol is based on the direct lysis method and contains the chaotropic salt guanidinium hydrochloride, as well as various other chemical agents that bind and inhibit contaminants and nucleases [237, 238]. The protocol proved to be universal for all three environmental samples from which metagenomic DNA was extracted and copious amounts of clean, intact and high quality DNA were obtained for all sites sampled.

#### 6.1.4 The construction of the metagenomic libraries

Depending on the fragmentation size of the isolated metagenomic DNA, it can be cloned into either plasmid based vectors (< 15 kb), fosmids (up to 50 kb), or bacterial artificial chromosomes (BACs; up to 200 kb). If smaller fragments are desired, various means exist to reduce them to size such as ultrasonic fragmentation, Hamilton syringe treatment, or enzymatic digestion which can yield compatible ends for cloning, as was done in this dissertation. The latter has the added advantage that it negates the potential need for blunting and adaptor ligation, thereby avoiding further induced biases through additional ligation and cloning.

The various insert-size libraries each have their own advantages and limitations. Generally though, smaller size plasmid based libraries focuses on the expression of single genes in a size range of 1 kb - 3 kb. Plasmid based libraries have the added advantage of being high-copy number and usually contain strong promoters (e.g. T7), thereby making the detection of under-active enzymes more likely. In contrast, large-inset libraries usually rely on the expression of many genes and operons that are driven by their own native promoters located on the library insert itself. Often these insert derived promoters aren't active with the cloning host's transcriptional machinery. The *F'* based *E. coli* origin of replication employed in many of these large insert vectors maintains a single copy thereby making the detection of toxic genes more achievable [385-387]. It also makes the upkeep of such a large DNA insert within the host cell more probable. Mechanisms to selectively Induce multi-copy numbers (approx. 100 copies per cell) through the employ of the *OriV* multiple origins, by TrfA (trans-replication protein) induction, are employed by several library systems [388]. This enables higher DNA recovery yields for downstream purposes, as well as enhance detection of the positive phenotype once stable growth of the host cells is achieved [377].

Several of the library systems, whether it be small or large insert based, contain *cos*-sites for the packaging of the ligated DNA into  $\lambda$ -bacteriophage heads. These phage based vectors can enter the host cell much more readily than naked DNA, particularly the larger vectors. There are several advantages to employing phage based libraries. For the smaller insert size M13 bacteriophage based libraries, storage and handling are considerably improved, as the packaged DNA can be maintained almost indefinitely in its protective capsids and the initial amplified library often yields adequate amounts of library phage making sharing amongst scientist more feasible. Moreover, most small insert size M13 based library systems like the  $\lambda$ -ZAP library system employed in this study, enables the easy conversion from the phage vector to that of a plasmid (pBK-CMV) through *in vivo* excision, thereby providing it all the benefits of plasmid based screening as discussed above. Furthermore,  $\lambda$ -based vectors enables the use of other molecular techniques, like single strand DNA recovery and hybridisation studies, more practical [236].

The main objective of this project was to discover novel enzymatic biocatalyst that could have application in the industrial production of prebiotic oligosaccharides and polymers through a metagenomic approach. It therefore makes sense that these biocatalyst should have a reduced transcriptional footprint so as to incur a minimalist burden on the host cell for its heterologous production. This can be achieved through having a higher catalytic efficiency, increased durability, being smaller in size, or a combination of all. More often than not, the more robust and stable industrial biocatalyst are also smaller in size [389]. In this regard, the smaller insert size M13 based  $\lambda$ -ZAP library construction mechanism is the perfect choice to achieve the main objective.

#### 6.1.5 Approaches toward metagenomic bioprospecting

By far the most frequent application of metagenomics is the quest for novel biocatalyst. This can be accomplished through either function-based and/or sequence-based screening techniques. Both approaches have their advantages, as well as shortcomings. In general, functional screening has the potential to identify novel enzymes based on functionality that share no homology to already known and characterised protein sequences that are available on the public databases. In contrast to the before mentioned, sequence based screening generates huge amounts of data through next generation sequencing platforms, and find potential positives based primarily on the homology to conserved sequences and domains. It therefore identifies new variants of already known enzyme families, but often fails to identify really novel gene sequences as they aren't related to known sequences, as is the case for functional screening [390, 391]. A prudent and recent example of functional vs. sequence based screening, was the discovery and biochemical characterisation of three novel  $\beta$ -galactosidases by

Cheng *et al.* (2017) from a soil metagenomic library. Although two of the sequences contained conserved domains, none showed any homology to the known glycoside hydrolase families and their ORFs and functional assignment could not be identified through standard bioinformatic analysis [392].

Never the less, sequence based screening has the potential to overcome many of the inherent biases related to library construction as well as the problems associated with heterologous gene expression, all of which are preconditions for functional screening of metagenomic libraries [233]. A sequenced based approach to metagenomic screening can also be advantageous in that the phylogenetic data will be readily available for profiling of the microbial community, should such a need exist. Furthermore the time consuming procedures of DNA manipulation, cloning, transformation and library maintenance can be circumvented. However, for the functional genes to be identified the *de novo* sequence data firsts needs to be assembled, and for huge data sets to be adequately assembled sufficient coverage needs to be achieved. To this extent the processing and meaningful interpretation of the metagenomic sequence data obtained, still remains the biggest challenge and bottleneck in a sequence based screening approach. In general two self-explanatory types of assembly methods are employed, namely reference based assembly and *de novo* assembly [393]. In seeing the high diversity often associated with microbial communities such as soil, the latter is most often employed in metagenomic analysis. After the initial assembly, downstream *in silico* analysis involves ORF prediction and functional assignment. To assign the most probable functions, expert-curated databases like SWISS-PROT, Pfam, TIGRFAMs, as well as integrated databases like the non-redundant NCBI, UniProt and InterPro are accessible [294, 394-398]. Other databases are based on the phylogenetic classification of proteins by Clusters of Orthologous Groups (COG), or by Gene Ontology (GO), or by mapping enzymatic function to biochemical pathways using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) [399-401]. Furthermore comprehensive databases with detailed information regarding the different enzymes and their respective families, like BRENDA and CAZy are also available [325, 402].

In contrast to sequence based screening, functional metagenomic screening relies on the successful heterologous expression of the gene products in the host organism, which is usually *E.coli*. Three broad differential functional screens can be distinguished namely, (i) *phenotypic detection* which is based on the actual enzyme activities that are detected by indicator agents, (ii) *modulated detection* which involves heterologous complementation that allows the metagenomic clones to grow under selective conditions, e.g. defined selective media. And lastly (iii), *induced gene expression detection* which involved substrate or product induction of reporter genes within the expression host. The functional screening approach employed in this dissertation was a combination of the first two, i.e. phenotypic and modulated detection. For the screening of  $\beta$ -galactosidase and thus potentially trans-galactosylation activity, the relative cheap and convenient indicator substrate 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (x-gal) was used. This screening method could only be employed once

all of the endogenous  $\beta$ -galactosidase background activity was removed from the screening host through recombineering. By further employing a combinational screen through an added modulated approach, i.e. screening on minimal media with distinctive carbon sources, a very selective and effective screening system was developed. The effectiveness of a functional screen lies predominately in its design and is extensively reviewed in the Taupp *et al.* article titled '*The art and design of functional metagenomic screens*' [377]. Because of the effectiveness of the screen design, and the data analysis challenges associated with a purely sequence based screening approach, it was decided to employ a functional screen for the experimental section of this dissertation. However, due to the ever advancement of data analysis platforms and methodologies, and the still falling cost of next generation sequencing, for future purposes a dual approach might very well be employed.

## 6.2 Enzymatic production of prebiotic galacto-and fructo-oligosaccharides

### 6.2.1 The bifidogenic effect and prebiotic relevance of enzymatically produced oligosaccharides

To a large extent, there is still a dominant belief that even though dietary intake may modulate certain metabolic activities associated with the intestinal microbiota, changing one's diet has little effect on the overall composition of the microbial community of the GI tract [403]. Over the last decade however compelling research has begun to dispel this inexactness to the contrary. Increasingly it is being recognised that even small changes to the diet can have a significant influence on the relative species composition and physiological traits of the GI microbiota, and thereby greatly effect the health of the human host [21]. The impact of dysbiosis of the GI tract, as well as its remediation through pre-, pro-, and synbiotics were extensively reviewed in chapter 2 section 2.4 -2.5 on pages 10 - 17.

Also in chapter 2, the prebiotic effect was defined as the particular stimulation of growth and/or activities of a selective number of microbial species of the gut microbiota that confer a health benefit to the host other than nutritional value [266]. The three key characteristics of a good prebiotic are: (i) that it is non-digestible by the human host, (ii) that it is selectively fermented by a key set of resident species of the GI microbiota, and (iii) that this consequently results in the targeted increase of the specific bacteria, which then subsequently through various means effect a health benefit on the host [21]. It is in the beforementioned context that the novel biocatalyst and their respective oligosaccharide products, described in this dissertation in chapters 4 and 5, must be interpreted.

Prebiotics are dietary derived carbohydrates and the two major carbohydrate groups that satisfy the criteria described in the preceding paragraph, are fructans and galactans e.g. FOS and GOS

(for review see chapter 2). Inulin type fructans with  $\beta$ -2,1 linkages and a terminal  $\alpha$ -linked glucose are either extracted from plants or synthesised enzymatically from sucrose, and have been shown in numerous studies to promote the prebiotic 'bifidogenic' effect (for review see Shoaib *et al.* 2016) [148]. Increasingly  $\beta$ -2,6 levan type FOS are being touted as an proper prebiotic source. This is apart from all its other biomedical, pharmaceutical and food science related applications [157]. Several studies have shown that levan and levan type FOS, selectively and more specifically stimulate bifidobacteria growth than even inulin on occasion, and thus rightfully act as a prebiotic substrate on its own [144, 157, 404].

Contrary to FOS, prebiotic GOS are galactose containing oligosaccharides with a terminal  $\beta$ -linked glucose monomer. The disaccharide lactose occurs almost exclusively in the milk of Mammals [182, 405]. Commercial lactose derivatives include GOS, lactitol, lactulose and lactosucrose [138, 297]. The GOS that are commercially derived from lactose by the enzymatic action of glycoside hydrolases, are colloquially revered to as  $\beta$ -GOS. These  $\beta$ -GOS differ their stereo- and regio-chemistry from the plant derived galacto-oligosaccharides/polymers, most commonly found in beans and pulses and which contain  $\alpha$ -linked galactose and glucose moieties and can furthermore be terminally  $\beta$ -linked to a fructose monomer (e.g. raffinose, stachyose and verbascose) [406]. Due to this diversity in linkage of the plant derived GOS, they can be fermented by any bacteria that contain the enzymatic machinery for these bonds, thus explaining the reported observation that plant GOS is less selective in its bifidobacteria growth stimulation than lactose derived  $\beta$ -GOS [407]. Prebiotic GOS research therefore increasingly focuses, much like that of FOS, on the enzymatic production rather than extraction from natural sources, due to its specificity in the promotion of bifidobacteria and other key probiotic bacterial growth [406].

The developmental origins for commercially derived  $\beta$ -GOS was modelled on human milk oligosaccharides (HMO). Breastfed infants have been shown to have an intestinal microbiota that is dominated by bifidobacteria, whereas non-breastfed infants contains less bifidobacteria and greater numbers of potentially harmful bacteria like clostridia and enterococci [408, 409]. Infants who are fed powdered milk, have also been shown to display elevated levels of detrimental constituents like ammonia, amines and phenols. [410]. Based on the analysis of human milk and its high concentration of galactose, Boehm *et al.* (2002) developed a prebiotic mixture to be added to cows milk, consisting of 10% FOS and 90% GOS, so as to simulate human milk for the use in infant formula. When this formula was fed to term and pre-term infants, it ultimately led to an increase in intestinal bifidobacteria and lactobacilli which more resembled the composition of microbiota in breastfed infants [411, 412]. HMOs contain lactose at their reducing end, much like many of the enzymatically derived  $\beta$ -GOS. Furthermore, HMOs can also be fucosylated and/or sialylated with  $\alpha$ -linkages [413]. Ultimately, when taking into consideration monomer composition, linkage types and degree of polymerisation, more than 1000 structures for HMOs exist, this however varies among individuals and is related to the maternal Lewis

blood type [413]. The quest to imitate HMO is a considerable driving force for the enzymatic production of GOS and novel oligosaccharides, creating a demand for an increased repertoire of regio and stereo specific linkages. This has led to the discovery of various catalytic sources for the production of different types of GOS, as well as different types hetero-oligosaccharides produced by glycosides hydrolases utilising lactose and other compounds as either glycosyl substrate or donor, as is discussed in sections 6.2.2 and 6.2.3 below.

In infants HMOs are specifically metabolised by *Bifidobacterium longum* ssp. *infantis*, however their benefit to health is not limited to the bifidogenic effect and include numerous other effects (for extensive review see Chapter 2), including the prevention of pathogen adhesion. So too does  $\beta$ -GOS resemble cell surface glycoconjugate receptor structures used by pathogens for adherence in GI tract [414]. The bifidogenic effect in itself is also not restricted to infants, as evidence suggest that *Bifidobacterium adolescentis* and *Bifidobacterium longum* predominate in most adults, whereas *Bifidobacterium breve* and *Bifidobacterium infantis* are the main species colonising the infant GI tract. [415]. There is also an age related observational decline in the bifidobacteria and lactobacilli content of the GI tract in the elderly [416] There are numerous feeding studies that show that bifidobacteria, as well as lactobacilli numbers can be increased by FOS and GOS supplementation [151, 417-419]. It is bifidobacteria and lactobacilli that play an important role in the eco-physiology of the GI microbiota and it is these bacteria specifically that has been associated with a healthy microbiota and the prevention of dysbiosis. Some bifidobacteria and many lactobacilli manifest strong anti-mutagenic and anti-tumour properties, and at least in animal models have prophylactic and therapeutic benefits [420]. Some strain of bifidobacteria also produce large amounts of folate, nicotinic acid, thiamine, pyridoxine and vitamin B12 [90].

But why the metabolic propensity of bifidobacteria and lactobacilli towards GOS and FOS? Due to their  $\beta$ -glycosidic bonds, GOS and FOS are neither hydrolysed nor absorbed in the acidic upper GI tract, and thus travel to the large intestine where they can be fermented [117]. Bacteria have different specificity for the prebiotics they can utilise. Gene clusters within their genomes dictate which saccharolytic machinery they possess and ultimately which prebiotic metabolites they can metabolise and to which extent. Genetic mechanisms for  $\beta$ -GOS specificity have been identified within bifidobacteria. They express higher activity for  $\beta$ -galactosidase and they often contain multiple isoforms of the enzyme, all of which explain why they show such a preference for  $\beta$ -GOS [421]. A study by Vulevic *et al.* (2009) which assayed 1% (w/v) culture additions of various prebiotic substrates, showed that enzymatically derived  $\beta$ -GOS in specific, increased the growth rates of *Bifidobacterium spp* but not other bacteria [407]. An additional possible explanation for this could be that some bifidobacteria at least, including *Bifidobacterium adolescentis*, have no transmembrane domains on some of their  $\beta$ -galactosidases and evidence suggest that they internalise oligosaccharides prior to digestion, thereby preventing the

scavenging of GOS by other bacteria. Furthermore, *in vitro* studies suggests that certain bifidobacteria preferentially utilise oligosaccharides and disaccharides before glucose in mixed cultures [422]. A randomised cross-over controlled study in 59 healthy volunteers showed that either one of two different types of enzymatically derived  $\beta$ -GOS preparations, had a significant bifidogenic effect at seven gram a day supplementation [423]. So too did a recent study by Hamdy *et al.* (2017) show that the levan produced by levansucrases from bacteria isolated from honey and bee gut, to possess a significant prebiotic effect and in specific promoted the growth of *Lactobacillus casei* and *Lactobacillus reuteri* [424]. Regardless of the exact mechanisms, it is these bacterial populations of bifidobacteria, lactobacilli and a few other species, that benefit most from GOS and FOS supplementation and ultimately account for this prebiotic aka 'bifidogenic' effect. Due to these properties of GOS and FOS as active ingredients in functional foods and as HMOS substitutes, the prebiotic market is rapidly expanding. In view of the above mentioned, in conjunction with the reasons provided in the review chapter 2, the isolation and characterisation of the MS- $\beta$ gluc1, HW- $\beta$ gal1, and SAS-Ls genes and their translational protein products as well as their respective GOS and  $\beta$ -2,6 type FOS preparations, have the potential to significantly contribute towards the field of enzymatic production of prebiotic fibres.

#### 6.2.2 Composition of the GOS synthesised by $\beta$ -galactosidases and $\beta$ -glucosidases

Commercial preparations of GOS utilises the  $\beta$ -galactosidases from *Kluyveromyces lactis*, *Bacillus circulans* (Vivinal GOS), *Bifidobacterium bifidum* (Bimuno), *Aspergillus oryzae* and *Streptococcus thermophilus* (oligomate55) [117]. These preparations differ in their substrate specificities and the range of compositional GOS they produce. In all of these commercial preparations, the residual lactose together with the GOS that range in DP2 to DP4, constitute approx. 90% of the total solids [273]. The glycosyl hydrolase family 2 type  $\beta$ -galactosidases produce mainly  $\beta$ -1,6 linked GOS with 6'-galactosyl-lactose being the major trisaccharide produced. This is exactly as was established for the major GOS constituent of the preparation by the GH2 HW- $\beta$ gal1 characterised in this dissertation, and is in line with current research for other characterised  $\beta$ -galactosidases most notably that of *Aspergillus oryzae*, *Aspergillus aculeatus*, *Kluyveromyces lactis*, *Bifidobacterium breve*, *Lactobacillus sakei*, *Lactobacillus reuteri* and *Lactobacillus plantarum* [425]. Even though these  $\beta$ -galactosidases have been shown to synthesise  $\beta$ -1,4 and  $\beta$ -1,3 linkages, this is to a much lesser degree and only *A. oryzae* really does so to any real significant extent (approx. 5% total) [190]. Originally the bifidogenic effect was attributed to mostly  $\beta$ -1,6 linkage type and more specifically the trisaccharide 6'-galactosyl-lactose [118]. However the bifidogenic effect for  $\beta$ -1,3 linkage type has also been proven since, but few true sources exist for it in view of the dominance of the  $\beta$ -1,6 and to a much lesser extent the  $\beta$ -1,4 linkage

types most commonly produced by GH2  $\beta$ -galactosidases [120, 223, 230]. It is here that the GOS synthesised by  $\beta$ -glucosidases rather than  $\beta$ -galactosidases can contribute significantly in that they appear to favour the formation of GOS with linkages other than the predominant  $\beta$ -1,6. In a study by Hassan *et al.* (2015) a thermostable  $\beta$ -glucosidase from *Halothermothrix orenii* produced a similar GOS profile in its equal distribution of 6'-galactosyl-lactose and 3'-galactosyl-lactose as did MS- $\beta$ gluc1 described in Chapter 4 [280]. In fact the recent literature shows that more and more research is being focussed on GOS production by means of  $\beta$ -glucosidases [426-428]. Several reasons for this abide but most notable is that unlike GH2  $\beta$ -galactosidases,  $\beta$ -glucosidases which are predominantly from the GH1 family are much smaller (avg. 50 kDa) in size and monomeric in nature. Furthermore, the stable ( $\alpha/\beta$ )<sub>8</sub>-barrel which accounts for the functional catalytic site of  $\beta$ -glucosidases is located to a single polypeptide chain, as apposed to the complex tetrameric active site associated with GH2  $\beta$ -galactosidases. This simplicity of  $\beta$ -glucosidases makes their protein production easier and less costly. It also lends them to mutational improvement and has provided much insight in to the catalytic mechanisms of trans-galactosylation. Recently, Yang *et al.* (2017) created seven mutant variant of the *Thermotoga naphthophila*  $\beta$ -glucosidase resulting in a 350 fold decrease in the hydrolytic reaction in the F414S variant which ultimately increased the GOS yield by 50% [428]. Also recently, another  $\beta$ -glucosidase was isolated and characterised from *T. naphthophila*. What is remarkable from this catalyst is that it exclusively synthesise a yet uncharacterised trisaccharide GOS [427]. This  $\beta$ -glucosidase has a deep catalytic pocket that prevents simultaneous access of both lactose and another trisaccharide to the catalytic site, accounting for its distinct catalytic specificity. Thus far all reported GOS preparations are mixed populations of different DPs and linkages, and this is the first time that such specificity has been shown. The production of homogenous oligosaccharide preparations have the potential to significantly reduce downstream purification cost and also provide insight into the the finer details of GOS synthesis. In all, future production of GOS might be more inclined to the exploitation of  $\beta$ -glucosidases, rather than the classical  $\beta$ -galactosidase models and further study of the MS- $\beta$ gluc1 characterised in this dissertation, including it's potential mutational analysis, could significantly contribute to this field.

### 6.2.3 Hetero-oligosaccharides synthesised by glycoside hydrolases

As stated above, the initial drive to imitate HMOs led to innovative means in the production of various hetero-oligosaccharides. Two distinct types of hetero-oligosaccharides can be produced utilising lactose and other substrates; namely the ones synthesised by  $\beta$ -Gals and  $\beta$ -Glucs with lactose as galactosyl donor, and the ones synthesised with the aid of various glycoside hydrolases where lactose act as an

acceptor. In general the specificity of  $\beta$ -galactosidases and indeed some fructosyl-transferases like levansucrase are low with respect to the acceptor molecule. Therefore, various sugars can interchangeably act as acceptor to the trans-glycosylation reaction, enabling the synthesis of unique hetero-oligosaccharides [155, 275, 297]. It is also not only carbohydrates that can act as acceptor, but other hydroxyl containing compounds like alcohols as well [429]. This opens up many possibilities for the production of various compounds and in this regards MS- $\beta$ gluc1, HW- $\beta$ gal1 and SAS-Ls should be targeted in future research for possible application. When these type of reactions occur with sucrose and lactose in the reaction medium, and depending on the glycoside hydrolase/s present, the reaction medium will contain a combination of the the hydrolytic products (i.e. glucose, galactose, fructose), GOS/FOS, as well as these other compounds of interest. Thus far, three valuable hetero-oligosaccharide compounds have been obtained by utilising this enzymatic strategy namely, lactulose, lactosucrose and lactitol.

Lactulose (Gal- $\beta$ -1,4-Fruc), can be produced either enzymatically, or through chemical isomerisation. Its main application is the treatment of hepatic encephalopathy, and also as a prebiotic source [24]. Lactulose actively reduces the ammonia levels in the blood resulting from decreased liver function by being a preferred fermentation substrate over that of proteins, resulting in less ammonia being produced by the GI microbiota which can ultimately 'fog' cognitive function [430]. In the process of lactulose synthesis the trans-galactosylation of residual glucose produced by the hydrolysis of lactose also occurs, thereby co-generating allolactose (Gal- $\beta$ -1,6-Glc) in the reaction medium [431]. The yield of the lactulose is ultimately dependant on the ratio of lactose : fructose [432]. Both lactulose and allolactose can also act as galactosyl acceptor thereby yielding fructosyl-GOS, most commonly  $\beta$ -Gal-1,6-lactulose/allolactose or  $\beta$ -Gal-1,4-lactulose/allolactose [275]. The trans-glycosylation reaction of the hetero-oligosaccharides are still related to the original enzymatic source, and in this regard the investigation into whether the  $\beta$ -glucosidase MS- $\beta$ gluc1 could potentially produce Gal- $\beta$ -1,3 linkage types for hetero-oligosaccharides as it did for its initial GOS production, could be of significance in the production of rare  $\beta$ -1,3 compounds. Currently lactulose is produced exclusively through chemical catalysis, due to the fact that its enzymatic production yields heterologous mixtures of different oligosaccharides that is difficult to purify. The enzymatic route is however a good green chemistry option that is increasingly gaining favour. Furthermore, through mutational engineering the possibility exist for creating catalyst that is specific in their substrate formation e.g. lactulose, as was described for the  $\beta$ -glucosidase from *T. naphthophila* in section 6.2.2 which would then ultimately simplify purification.

Lactosucrose ( $\beta$ -Gal-1,4-sucrose) is a very rare trisaccharide that scarcely exist in nature and is difficult to chemically synthesize. It has been shown to possess a significant bifidogenic effect and also inhibit *Clostridium difficile* growth [124]. Furthermore, it is exceptionally good at promoting mineral absorption

and the administration of lactosucrose has been shown to have the long term effect of enhancing intestinal calcium retention and absorption and reducing bone resorption in healthy young women [433]. There are reports of it having the beneficial effect of inhibiting fat accumulation and thereby preventing obesity, and in mice it has been shown to reduce serum cholesterol levels [47, 197]. In addition, lactosucrose is an outstanding excipient for spray dried powders and shows promising results in the protein stabilisation of IgG and could therefore potentially be utilised as an amorphous powder for inhalation drug delivery [434]. Lactosucrose can be produced either through the trans-fructosylation of lactose by  $\beta$ -fructofuranosidase and levansucrase, or through the trans-galactosylation of sucrose by  $\beta$ -galactosidase and  $\beta$ -glucosidase. All of MS- $\beta$ gluc1, HW- $\beta$ gal1 and SAS-Ls could potentially contribute in the synthesis of lactosucrose and further research in this regard is warranted.

Klewicki *et al.* studied the the trans-galactosylation of several polyhydroxy-alcohols most notably sorbitol, xylitol, lactitol and erythritol [435]. During these enzymatic processes GOS were co-synthesised and the production of galactosyl-sorbitol was of special interest because it is an epimer of lactitol (4-O- $\alpha$ -D-galactopyranosyl-D-sorbitol) [436]. Lactitol is an artificial synthetic sugar produced from lactose by the catalytic hydrogenation of the glucose moiety in the disaccharide [437]. Lactitol has been reported to have excellent prebiotic properties. Furthermore, it is purported to have special synbiotic efficacy by increasing the viability of several probiotic strains [438]. It is also a low calorific sweetener with use as a substitute for sucrose in the food industry [272].

Galactosidases, sialidases, glucansucrases, fructansucrases,  $\alpha$ -fucosidases and N-acetyl-glucosaminidases are all retaining glycoside hydrolases with trans-glycosylation activity. They all have the ability to transfer their corresponding sugar moieties to acceptor sugars such as lactose and sucrose, to generate a multitude of hetero-oligosaccharides, which ultimately include several fundamental structures of human milk oligosaccharides. Here special reference was given to the hetero-oligosaccharide production by  $\beta$ -glucosidases,  $\beta$ -galactosidases and levansucrases, in seeing their relevance in this dissertation. The high added value products that can potentially be obtained by the application of MS- $\beta$ gluc1, HW- $\beta$ gal1 and SAS-Ls, other than the obvious GOS and FOS production, adds even more worth to these enzymatic catalyst. It is a great incentive for continual research into these three specific enzymes.

### 6.3 Summary and concluding remarks

It is well known that the disaccharides lactose (O- $\beta$ -D-galactopyranosyl-1,4-D-glucopyranose) and sucrose (O- $\alpha$ -D-glucopyranosyl-1,2- $\beta$ -D-fructofuranoside) are two of the most common and cheapest

substrates available. Several million tons of whey derived lactose are produced annually as a by product in the cheese manufacturing industry. The disposal of industrial whey is a never-ending environmental problem with no real green solution in sight. Sucrose is an ubiquitous sweetener in the food industry and has a multi-billion dollar footprint. It is produced globally in the tropical and subtropical regions; areas which are often associated with poverty and being heavy reliant on the sugarcane industry. Increasingly though the excess consumption of sucrose, especially its high fructose corn syrup derivative, are being linked with a multitude of negative health associated risk including the current obesity epidemic and type II diabetes. This has led to a gradual global decline in the sugar industry, and alternate uses are actively being sought. To the contrary and ironically, the prebiotic oligosaccharides GOS and FOS which are enzymatically derived from lactose and sucrose, are associated with a multitude of health related benefits and are actively being promoted as functional food ingredients in an attempt to address these very concerns currently facing the global population. The utilisation of both lactose and sucrose to produce functional derivatives especially GOS and FOS, therefore presents not only an economical viable model but also addresses the environmental aspects and the quest for alternate uses of these substrates.

In CHAPTER 3: the construction and functional screening of several metagenomic libraries were presented. A collection of these libraries were screened on minimal media containing lactose in a specially constructed  $\beta$ -galactosidase deletion mutant, for the presence of lactose hydrolysing and trans-galactosylation enzymatic catalyst. A library was also constructed from runoff collected at a sugarcane refinery, and screened for the presence of trans-fructosylation enzymatic catalyst, on minimal media plates containing sucrose as the sole carbon source. In all, 17 putative glycoside hydrolase genes, from various families, were isolated, *in silico* characterised, and preliminarily analysed for their oligosaccharide/polymer producing capabilities (Table 3-3, section 3.3.2). Eventually three candidates genes were selected for further and more detailed analysis, which were then researched in the following two chapters.

In CHAPTER 4: the two genes encoding for the novel glycoside hydrolases named MS- $\beta$ gluc1 and HW- $\beta$ gal1, belonging to Glycoside Hydrolases family 1 and 2 respectively, were cloned, heterologously expressed and biochemically characterised together with their respective end products. MS- $\beta$ gluc1 had a  $K_m$  of 41 mM, a  $K_{cat}$  turnover of  $111\text{ s}^{-1}$  and a catalytic efficiency of  $2.7\text{ mM}^{-1}\cdot\text{s}^{-1}$  for the substrate lactose at its optimum pH 6 and temperature of  $45^\circ\text{C}$ . The second enzyme, HW- $\beta$ gal1 had a  $K_m$  of 32 mM, a  $K_{cat}$  turnover of  $625\text{ s}^{-1}$  and a catalytic efficiency of  $19.5\text{ mM}^{-1}\cdot\text{s}^{-1}$  for the substrate lactose at its optimal pH 7.5 and temperature of  $35^\circ\text{C}$ . Both proteins were shown to have high trans-galactosylation activity with HW- $\beta$ gal1 yielding approximately 43% galacto-oligosaccharides (GOS), mainly consisting of tri-saccharides of the  $\beta$ -1,6-galactosyl-lactose type. MS- $\beta$ gluc1 yielded approximately 40% GOS, consisting of a combination of the trisaccharides  $\beta$ -1,3-galactosyl-lactose and  $\beta$ -1,6-galactosyl-lactose.

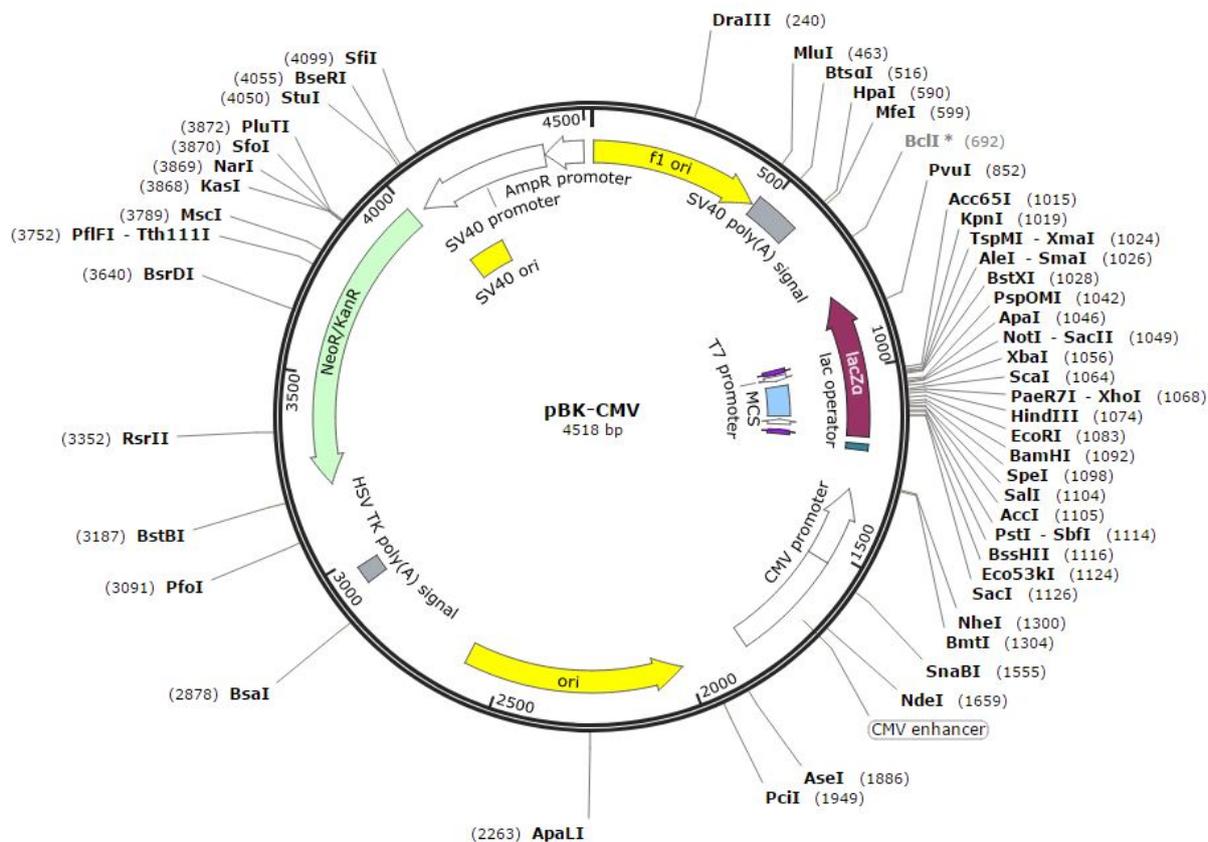
Both enzymatic catalyst have industrial relevance and especially the  $\beta$ -glucosidase has the potential for the synthesis of more novel and rare  $\beta$ -1,3 linked GOS and other hetero-oligosaccharides.

In CHAPTER 5:, a novel gene encoding for a Glycoside Hydrolase Family 68 levansucrase, *SAS-Ls*, was cloned, heterologously expressed and biochemically characterised together with its reaction end-product. *SAS-Ls* had a  $K_m$  of 61 mM for the substrate sucrose and displayed its optimum reaction conditions at pH 4.5 – 5.0 and a temperature of 35°C. Furthermore, *SAS-Ls* had an extremely high  $K_{cat}$  turnover of  $7682s^{-1}$  and  $5587s^{-1}$  for the hydrolytic and trans-fructosylation reaction respectively, which makes it one of the highest reported yet. The fructo-oligosaccharides produced by *SAS-Ls* was identified as a levan type with  $\beta$ -2,6 linkage by NMR and hydrolysis analysis.  $\beta$ -2,6 linked FOS together with GOS are proven prebiotic sources and *SAS-Ls* could significantly contribute to the production of prebiotic FOS. Levan type oligosaccharides and polymers also have a multitude of other potential applications in the food, cosmetic and medical industries, in all of which *SAS-Ls* could potentially find relevance. Like the galactosidase activity of MS- $\beta$ gluc1 and HW- $\beta$ gal1, *SAS-Ls* could furthermore possibly contribute to the production of novel and rare hetero-oligosaccharides through its trans-glycosylation activity, and this needs to be explored in future research.

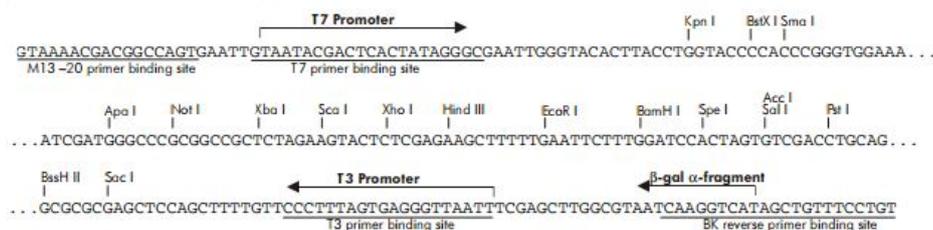
In CHAPTER 2: the various characteristics and health promoting effects of prebiotics were extensively reviewed, together with its influence on the GI microbiota and ultimately its human, host both in a diseased and healthy state. Additionally to the before mentioned, was an examination of its enzymatic production, and also a review of the unique biocatalyst that facilitate this. In CHAPTER 6: the enzymatic catalyst isolated and characterised in the three experimental chapters were thoroughly discussed, not only in the prebiotic context, but also in their potential application in the synthesis of other hetero-oligosaccharides derived from lactose and/or sucrose. All three biocatalyst have potential to be employed on an industrial level and their discovery and characterisation adds to the ever expanding field of prebiotic enzymatic synthesis. Numerous avenues exist for further research to be conducted, so as to broaden their application and relevance.

APPENDIX

A1 pBK-CMV vector map

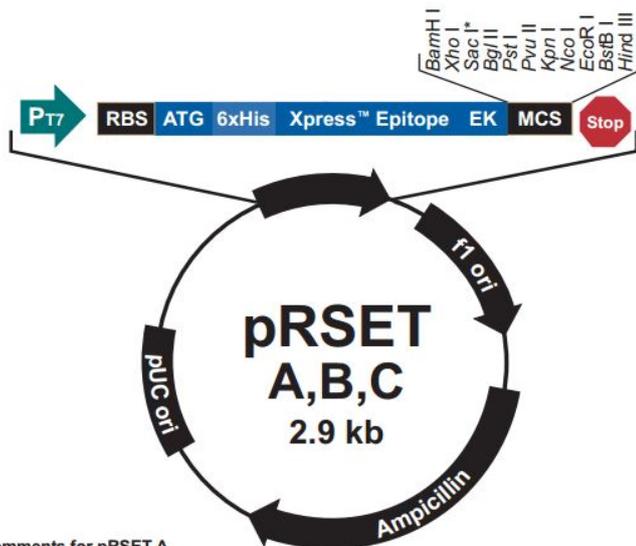


**pBK-CMV Multiple Cloning Site Region  
(sequence shown 952-1196)**



Feature	Nucleotide Position
f1 origin of ss-DNA replication	24-330
SV40 polyA signal	469-750
β-galactosidase α-fragment coding sequence (lacZ')	812-1183
multiple cloning site	1015-1122
lac promoter	1184-1305
CMV promoter	1306-1895
pUC origin of replication	1954-2621
HSV-thymidine kinase (TK) polyA signal	2760-3031
neomycin/kanamycin resistance ORF	3209-4000
SV40 promoter	4035-4373
bla promoter	4392-4518

A2 pRSET A, B and C vector map.



Comments for pRSET A  
2897 nucleotides

\*Version C does not contain Sac I

- T7 promoter: bases 20-39
- 6xHis tag: bases 112-129
- T7 gene 10 leader: bases 133-162
- Xpress™ epitope: bases 169-192
- Multiple cloning site: bases 202-248
- T7 reverse priming site: bases 295-314
- T7 transcription terminator: bases 256-385
- f1 origin: bases 456-911
- bla* promoter: bases 943-1047
- Ampicillin (*bla*) resistance gene (ORF): bases 1042-1902
- pUC origin: bases 2047-2720 (C)

### pRSET-A MCS

T7 promoter RBS

21 AATACGACTC ACTATAGGGA GACCACAACG GTTCCCTCT AGAATAAATT TTGTTTAACT TTAAGAAGGA

Polyhistidine (6xHis) region

91 GATATACAT **ATG** CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT  
**Met** Arg Gly Ser His His His His His His Gly Met Ala Ser Met Thr

T7 gene 10 leader Xpress™ Epitope BamH I

148 GGT GGA CAG CAA ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT C**GA TGG GGA**  
 Gly Gly Gln Gln Met Gly Arg Asp Leu Tyr **Asp Asp Asp Asp Lys Asp** Arg Trp Gly  
EK recognition site EK cleavage site

Xho I Sac I Bgl II Pst I Pvu II Kpn I Nco I EcoR I BstB I Hind III

205 **TCC GAG** CTC GAG ATC TGC AGC TGG TAC CAT GGA ATT CGA AGC TTG ATC CGG CTG CTA  
 Ser Glu Leu Glu Ile Cys Ser Trp Tyr His Gly ile Arg Ser Leu Ile Arg Leu Leu

T7 reverse priming site

262 ACA AAG CCC GAA AGG AAG CTG AGT TGG CTG CTG CCA CCG CTG AGC AAT AAC TAG CAT  
 Thr Lys Pro Glu Arg Lys Leu Ser Trp Leu Leu Pro Pro Leu Ser Asn Asn \*\*\* His

## A3 GB-BGal1 sequence

LOCUS GB\_Bgal1\_GH2 1024 aa linear UNA

FEATURES Location/Qualifiers

Region 1..1023

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 47..216

/Domain bit score=174.9

/Domain bias=0.0

/Accession=PF02837.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_N

Region 219..330

/Domain bit score=56.6

/Domain bias=0.0

/Accession=PF00703.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2

Region 333..626

/Domain bit score=394.1

/Domain bias=0.0

/Accession=PF02836.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_C

Region 749..1021

/Domain bit score=277.4

/Domain bias=0.0

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

ORIGIN

```

1  MPSTLASLLS RRDWENPVVT QWHRLAAHAP MRSWRDETA RDEADSAHH SLNGIWQFSF
61  FASPEAVPEQ WLEQDCADAV AMPVPSNWQM QGFDTPIYTN VTYPVNPFP FVPQQNPTGC
121 YSLTFNVDAE AIAEQTRIV FDGVNSAFHL WCNGQWVGYS QDSRLPAEFD LSAVLRAGEN
181 RLAVMVLRW C DGSYLEDQDM WRMSGIFRDV SLQHKPALHI ADYHYTTALN AEFTRASVQV
241 TVELAGEFAG SRVTAVLWRN GEKIASGEQT PGSVVDERG NWAERLSLTL PVESPLLWSA
301 ETPNLYRLTL TLYDAQGSCV EAEACDVGFR HVEIHQGLLK LNGQPLLIRG VNRHEHHPEH
361 GQAVDEATMW RDIELMKQHN FNAVRCSHYP NHPLWYRLCD RYGLYVVDEA NIETHGMVPM
421 SRLADDPRLW PAMSERVTRM VQRDRNHPSI IIVSLGNESG HGANQDALYR WLKSTDPTRP
481 VQYEGGGANT AATDIICPMY SRVDQDQFPF AVPKWSIKKW IGMPNETRPL ILCEYAHAMG
541 NSFGGFAKYW QAFRAAPRLQ GGFVWDWVDQ ALTKTGADGE LFWAYGGDFG DTPNDRQFCM
601 NGLVFPDRTP HPALFEAQR AQQFFQFSLLS ASPLTIEVTS EYLFRTSDNE VLRWRVERDG
661 EVLAQGETAL VIAPQGVQI ALDLPALAA PGEVWLNAEV YQREATAWSA AGHRCARDQW
721 RLPAPLWVAV PDKQGEQPTL EVNDETYTVK QGNQRWAFCR QRGNLVQWWR DGQETLLTPV
781 TDCFTRAPLD NDIGVSEVTR IDPNAWMERW KAAGMYDLHA ELLSFEVQES AQEVTLRVH
841 RWIGAGKSAF ISEKTWRIDR TGGLQADIEV VVANDIPPPA RIGLVCQLAE RPAEVSWLGL
901 GPHENYPDRK LAACQGRWTQ PLAALHTPYI FPGENLRCD TRLLQCGAHQ LEGRFHFSLG
961 CYSDTQLRET THHHLLREEA GCWLHLDAFH MGVGGDDSW SPSVSPEFILQ QERVRYRLRW
1021 QQA*
//

```

## A4 HW-Bgal1 sequence

LOCUS HW\_Bgal1\_GH2 1036 aa linear UNA

FEATURES Location/Qualifiers

Region 1..1035

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 56..225

/Domain bit score=160.7

/Domain bias=0.0

/Accession=PF02837.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_N

Region 228..341

/Domain bit score=37.5

/Domain bias=0.0

/Accession=PF00703.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2

Region 344..637

/Domain bit score=401.4

/Domain bias=0.0

/Accession=PF02836.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_C

Region 761..1033

/Domain bit score=295.3

/Domain bias=0.0

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

ORIGIN

1 MPISDTSRRH APDFHAVLAR EDWQNTTITH LNRLPAHPAF ASWRDELAAR DNRPSRRRQ  
61 LDGEWQFAYA RSPFAVDAQW LTQDLPGSRG TPVPSNWQME GYDAPIYTNV RYPIDTIPPR  
121 VPEDNPTGCV SLHVAIDDAW HTDGTQTIIF DGVNSAFHLW CNGAWVGYVSDSRLPAAFDL  
181 SPFLRPGDNR LCVMMRWWSA GSWLEDQDMW RMSGIFRSVW LLNKPQRLC DVQLTPTLDA  
241 LCRDGLQVQ ATVEATEAAL AGLCVGSLW RGEQPVAHR QPLGSPAVDE RGHYAERVDF  
301 SLAVAAPAHW SAETPNCYRA VVTLWRGDEL LEAEAWDIGF RRIEADGLL RLNGKPLLIR  
361 GVNRHEHHHL RGQVSEADM VQDILLMKQN NFNAVRCSHY PNAPRWYELC NRYGLYVDE  
421 ANIETHGMVP MNRLSDDPAW LPAFSARVTR MVQSNRNHPC IIIWSLGNES GGGGNHEALY  
481 HWLKRNDPSR PVQYEGGAD TTATDIICPM YARVERDQPI PAVPKWGICK WISLPGEQRP  
541 LILCEYAHAM GNSLGNFADY WQAFREYPRL QGGFIWDWAD QAICKTFDDG SVGWAYGGDF  
601 GDKPNDRQFC MNGLVFPDRT PHPSLIEAKH AQQYFQFALL STSPLRVRIA SEYLFQSDN  
661 EALRWVQVAA GETLYHGNTL LALPPEGSDE ITLLDDLILP PGARAVWLTL EVVQPRATDW  
721 SPADHRVAWQ QFPLPAPLAL PAPTVPAGAP DLVSEEAQW IRAGTQCWTV DRRTGLLSGW  
781 SLAGQEQLL PLQDQFIRAP LDNDIGVSEV ERIDPNAWVE RWKSAGLFDL EACCVQCDQ  
841 RLANETLVDS RWHYLRGDEV VIVSHWRMRF TADGTVRLTV DGERAETLPP LPRVGLHFQV  
901 AEQQAPVSWL GLGPHENYPD RRSSACFARW EQPLAAMSTP YIFPTENGLR CDTQALDWGR  
961 WHVSGHFHFS VQPWSTRQLM ETDHWHKIQA KDGWITLDG LHMVGVGDDSDS WTPSVLPQWL  
1021 LTQTRWQYEV SLRCL\*

## A5 HW-Bgal2 sequence

LOCUS HW\_Bgal2\_GH2 1025 aa linear UNA

FEATURES Location/Qualifiers

Region 1..1024

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 48..217

/Domain bit score=167.1

/Domain bias=0.0

/Accession=PF02837.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_N

Region 220..331

/Domain bit score=54.7

/Domain bias=0.0

/Accession=PF00703.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2

Region 334..627

/Domain bit score=396.9

/Domain bias=0.0

/Accession=PF02836.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_C

Region 750..1022

/Domain bit score=255.9

/Domain bias=0.0

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

ORIGIN

```

1 MSATALSSLV SRRDWNVPVI THWHRLPAHA PLRSWRDENA ARDDAASPAR RLLNGEWRFS
61 LFGSPEAVPE RWITEDCADA VAMPVPSNWQ MQGFDTPIYT NVTYPIPVTP PFVPPQNPTG
121 CYSLTFLMEE EALAQQQTRI VFDGVNAAFY LWCNGQWIGY SQDSRLPAEF DLSQALHPGE
181 NRLAVMVLRW CDGSYLEDQD MWRMSGIFRD VSLQHKPQTH IADFHVYSEL NAELTHAQLQ
241 VNVQLAGAF AECRAVALWH DGKNIAAAQ SPGSVVDER GAWAERLCVT LPVAAPVLWS
301 AETPNLYRLT LTLDDAAGNV LEAEACDVGK RKVEICNGLL LLNGKPLLIR GVNREHHPE
361 NGQAIDEATM RRDIELMKQH NFNAVRCSHY PNHPLWYRLC DRYGLYVVDE ANIETHGMVP
421 MSRLADDPGW LPAMSERVTR MVQRDRNHPS IIIWSLGNES GHGANHDALY RWLKTDPTR
481 PVQYEGGGAN TAATDIVCPM YARVDQDQPF PAVPKWSIKK WIGLPGENRP LILCEYAHAM
541 GNSFGGFAKY WEAFFRAFPRL QGGFVWDWVD QALTKTGDDG QPFWAYGGDF GDTPNDRQFC
601 MNGLVFPDRT PHPALYEAQR AQQFFQFSL STTPLVLEVE SEYLFRAFDN EYLWRSVARD
661 GDVLAQGEIT LDIAPQGGQR IELNIPALVA APGEVWLNVD VFQRAATRWS AADHRCARDQ
721 WRLPAPLYIA PRVVQNSRPT LQASEQEFVI THQSQRWHFC RRSGLNQQWW RDEQPTLLAP
781 LSDCFSRAPL DNDIGISEVT RIDPNAWVER WKAAGMYDLS AELLYCDVEE RSTGIVVNTG
841 QRWLGAGKTA FLSHKCWRID GDGALHGDVT VQVARDIPPP ARVGLVCQLA ERHPQVSWLG
901 LGPHENYPDR QLAARQGRWT QPLSALHTPY IFPGENLRC NTRAVWYGAH QWQGFHFSL
961 GCYSKQLRE TTHHLLREE AGCWLHLDAF HMGVGGDDSW SPSVSPEFIL QDETVRYAFC
1021 WWQN*
//

```

## A6 HW-Bgal3 sequence

LOCUS HW\_Bgal3\_GH2 1036 aa linear UNA

FEATURES Location/Qualifiers

Region 1..1035

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 55..224

/Domain bit score=173.6

/Domain bias=0.0

/Accession=PF02837.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_N

Region 227..339

/Domain bit score=39.6

/Domain bias=0.0

/Accession=PF00703.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2

Region 342..635

/Domain bit score=391.6

/Domain bias=0.0

/Accession=PF02836.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_C

Region 759..1031

/Domain bit score=270.1

/Domain bias=0.0

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

ORIGIN

```

1 MSKPISSASL TGLATLLARR DWENPVVTHW HRLACHAPLA SWRTEQEARQ GAPSAQRYSL
61 NGQWAFCLYP SPESVPESWL QADEPQSVAT PVPGNWQMAG FDTVPVYTNVN YPIVATPPCV
121 PVENPTGCYS RRFVPEPDWL MSGQTRIID GVNAAFYLWC NGHWWGYSQD SRLPAEFDLT
181 PYLTAGENRL AVMVLRWCDG TWLEDQDMWR MSGIFRDVTL LHKPACRIAD YHHQVCFNSD
241 YSRASLTLTL ETEGEQPQAC QAEVSLWRDG QCIVRQTKPL GSQVIDERGN YPERVTLTLD
301 IENPLLWSAE TPHYRLVMV LQDKQGQCLD AEACWTGLRE IVIQNGLLKL NGKPLLIRGV
361 NRHEHHPEHG QVMDEATMRQ DILLMKQHNF NAVRCSHYPN HPLWYQLCDE YGLYVVDEAN
421 IETHGMQPMs RLADDPWFVG AMSERVTRMI QRDRNHACII IWSLGNESGH GANHDALWRW
481 VKTTDPSRPV QYEGGGANTA ATDIVCPMYA RVDEDQPFEQ VPKWSIKKWV GMPDEHRPLI
541 LCEYAHAMGN SLGGFYRYWQ AFRQYPRLQG GFVWDWVDQA LTRHTEQGED WWAYGGDFGD
601 KPNDRQFCLN GLVFPDRTPH PALFEAQHAQ QFFRFDLVDA HPLTVRITSE YLFRDTDNER
661 LCWAVMQDGE PVLEGHRVLN LAPEAEVTIV LGDVPVQTRP GQRWLTWVVE YAADTLWARA
721 GQACAWGEWQ LPARLFLSPE AEPMGDLPQL TTTETEYVVC HGNKQWVFD R AQGTLAQRV
781 DGAAQLYSPV VDQFVRAPLD NDIGTSEAR IDPLAWVERW KAAGMYQLTP QVVLCEAGTV
841 FGDVVIRTRH AWYAAQQCCF ISEKQWRIDR LGILHIDVDV HIAGDIPPPA RIGLCCQLAI
901 VAPRVEWLGR GPHENYPDRK LSALQGRWRL PLEEMHTPYI FPEENGLRCD TRKLCYGHHQ
961 WQGEFHFLG RYSLQQLMDV SHRHLITEEP GTWLNLD AFH MGIGGDDSW S PSVHRENILT
1021 KTHVHYQLRW SLDAM*
//

```

## A7 HW-Bgal4 sequence

LOCUS HW\_Bgal4\_GH42 686 aa linear UNA

FEATURES Location/Qualifiers

Region 1..685

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 20..393

/Domain bit score=514.6

/Domain bias=3.2

/Accession=PF02449.10

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42

Region 404..616

/Domain bit score=242.3

/Domain bias=0.0

/Accession=PF08532.5

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42M

Region 627..683

/Domain bit score=42.1

/Domain bias=0.0

/Accession=PF08533.5

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42C

ORIGIN

1 MNKFAPLHPK VNTLLHGADY NPEQWENDPD IIDKDIAMMQ QAKCNVMSVG IFSWAKLEPR  
 61 EGVDFFAWLD TILDKLYAAG IHVFLATPSG ARPAWMSQRY PQVLRVGRDR VPALHGGRHN  
 121 HCMSSPIYRE KTLKINSLLA ERYAAHPAVL GWHISNEYGG ECHCDLCQAR FRGWLKARYQ  
 181 TLENLNQAWW STFWSHTYSD WSQIESPAPQ GETSIHGLNL DWHRFNTAQV TDFCRHEIAP  
 241 LKAANAALPV TTNFMEYFYD YDYWQLAEVL DFISWDSYPM WHRDKDETAL ACYTAMYHDM  
 301 MRSLKGGQPF VLMESTPGAT NWQPTSKLKK PGMHILSSLQ AVAHGADSVQ YFQWRKSRGS  
 361 VEKFGHAVVD HVGHIDTRIG REVSKLGEIL SKLPEVRGCR TEAKVAIIFD QQNRWALDDA  
 421 QGPRNLGMEY EKTVNEHYRP FWEQGIADV IDADADLTPY RLVIAPMLYM VRDGFAARAE  
 481 AFVASGGHLV TTYWTGIVNE SDLCYLGGF GPLRNLLGIW AEEIDCLNDG EFNLVQGLAG  
 541 NQCGLQGPYQ VRHLCELIHT ESAQTLATYR DDFYAGRPV TVNGFGKGA WHVASRNDLA  
 601 FQRDFFAALS KELALPRAIA ADLPPGVVAT ARTDGESTFV FLQNYSAQSH TLSLPSGYRD  
 661 CLTDAAVSDP LTLSAWDCRI LRRHA\*

//

## A8 HW-Bgal5 sequence

LOCUS HW\_Bgal5\_GH53 401 aa linear UNA

FEATURES Location/Qualifiers

sig\_peptide 1..23

/note="Cleavage site: 23^24: SLA-AD"

/note="D-score: 0.696"

/note="D-cutoff: 0.45"

/note="Max C-score: 0.448"

/note="Max S-score: 0.922"

/note="Max Y-score: 0.603"

/note="S-mean: 0.842"

/note="Networks: SignalP-TM"

sig\_peptide 1..23

/note="Cleavage site: 23^24: SLA-AD"

/note="D-score: 0.927"

/note="D-cutoff: 0.57"

/note="Max C-score: 0.853"

/note="Max S-score: 0.975"

/note="Max Y-score: 0.905"

/note="S-mean: 0.953"

/note="Networks: SignalP-noTM"

Region 1..400

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 41..399

/Domain bit score=403.2

/Domain bias=0.8

/Accession=PF07745.8

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_53

ORIGIN

```

1 MKRFTPAWLA VCLACSFSLP SLAADALET R AFQGM PADFI KGADISTLLD AEKHGAIFYD
61 QNNQRKDPIA ILKENG VNYV RLRLWVD PQS ASGEGYGGGN NDLATTLALA KRAKAQGMKL
121 LLDHFHYSDFW TDPGKQFKPK AWEKMDYPQL KTAIH DYTRD TIARFKQAGV LPDMVQIGNE
181 INGGMLWPEG KSWGQGGGEF DRLAGLLNAA IDGLKANLRN GEQVKIMLHL AEGTKNDTFR
241 WWFDEIEKRH VPYDVIGLSM YTYWNGPISA LKANMDDISK RYNKDVIVVE AAYAYTLANC
301 DNAENSFQAK EEKDGYPAT VQGQYRYIHD LMQAVIDVPD QRGKGFYWE PTWIAVPGNT
361 WATPAGMKYI HDEWKEGNAR ENQALFDCQG KVLPSVKVFN *

```

//

## A9 HW-Bgluc1 sequence

LOCUS HW\_Gluc1\_GH1 445 aa linear UNA

FEATURES Location/Qualifiers

Region 1..444

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 2..444

/Domain bit score=542.5

/Domain bias=6.9

/Accession=PF00232.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_1

ORIGIN

```
1 MNKFSKDFIF GTATSSYQIE GAYQEDGRTP SIWDTFSRTP GKVFNDMDNGD IACDHYHLYE
61 KDIEILKTLG VDSYRFSIAW PRIFPKQGGY NEAGMDFYKR LITRLIENGI KPAVTLYHWD
121 LPMWAHEKGG WTNRESVNWF LEYAEKCFEE LDEHVEMWST HNEPWCAGFL GYHQGVHAPG
181 HTNMEEAVKA VHHMLLSHGE AVSLLKGKVV SETPIGITLN LSPMYPASNS ANDQLAANNA
241 DGYTNRWFLD PVLKGSYPAD MMNLFKYVH SFDFIQEGDL EKISVECDFF GINYNNRSLV
301 EFNSASDFLF KSAYSYPKS GMGWDISPAE FKELIHRLRK EYTNLPIYIT ENGSAFDDHV
361 SEDHRVHDSR RQDYVEKHIT AVAELNDEGM NVAGYLLWSL LDNFEWAFGY DKRFGITYVD
421 FETQERILKD SGYRYAEIIR NRSI*
//
```

## A10 HW-Bgluc2 sequence

LOCUS HW\_Gluc\_GH16 256 aa linear UNA

FEATURES Location/Qualifiers

Region 1..255

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 41..249

/Domain bit score=136.2

/Domain bias=0.8

/Accession=PF00722.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_16

ORIGIN

```
1 MMTLSKTEAE WSLVWEENFD LPDIDESKWN FVEAGTGFGN EESQFYTRRK ENARIENGML
61 VLEARNEEYN GMDYTSAKLT TRGKAAWTYG RFSIRAKLPE GQGIWPAIWM MPEDMELYTG
121 WPACGEIDIM ELIGHQPGTV YGTLHYGMPH TYTGENTYLP DGKKFSEDFH VFTLDWKPGE
181 FRWYVDDVPY ARQTEWFSQS PESAEKQAGF APFDRDFYLQ LNLAVGGKWP GYPDEKTQFP
241 QQMTVDYIKV YKKEK*
```

//

## A11 MS-Bgal1 sequence

LOCUS MS\_Bgal1\_GH2\_(32) 1132 aa linear UNA

FEATURES Location/Qualifiers

sig\_peptide 1..22

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 72..252

/Domain bit score=166.5

/Domain bias=0.4

/Accession=PF02837.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_N

Region 255..365

/Domain bit score=62.0

/Domain bias=0.0

/Accession=PF00703.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2

Region 377..661

/Domain bit score=279.1

/Domain bias=2.9

/Accession=PF02836.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_C

Region 787..905

/Domain bit score=55.2

/Domain bias=0.1

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

Region 1057..1115

/Domain bit score=26.4

/Domain bias=0.0

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

ORIGIN

```

1 MKFKFLSLVL TVGLSTIVTL HAQESRMISS QRKTPYWQDV NNVQVNKEYP RTQFMFTFDNK
61 PEAMNSRFEE SKYYISLNGT WKFYFVEGYK QLPENVTDSV VSLSGWKEIK VPGNWELQGF
121 GTPIYVNHYP EFVERDPKTR FPKFAPPYLP EKNPVGYYRR EIDIPQDWKD REIFLSIDGA
181 KSGVYVYING KEVGYSEDSK TSAEFRISKY VKPGKNSLVL KIFRWSTGSY LECQDFWRIS
241 GIERDVLWFS QPKTSLRDFR VKSTLDDSYQ NGIFELETTV SNYSPGVSYA EVFYELLDAA
301 GKTAASGSQA VSVQGGQENT VKFEAQLPNV ATWTSEHPNL YKLLISVRKE GEKSGEVVPPY
361 SVGFRRFEIK AVKTGERIDR LFLVNGQPIK LKGVNIHETN PKTGHYVPEE LMRKDFELMK
421 QNNINTVRLS HYPQARRFYE LCSEFGLYVY DEANIESHGM YYGQESLAKH PEWEQAHRDR
481 TVNMFERNKN HPSVAIWSLG NEAGNGINFF HTYKYLKDQE RNFMNRPVNY ERGLWEYNTD
541 MYVPQYPSAA WLEEVGKKGS DRPVVPEYS HAMGNSSGNL DLQWQAIYKY PNLQGAYIWD
601 WVDQGMIAVD ENGRVYTYG GDYGTDMPSD GNFLCNGIVN PDRTPHPAMA EVKYTHQNFA
661 VEAVDLTKGI FNIINRQYFS NTDNYTFKYN ITENKQISE GVLVPTLAPQ QAAHATVPVG
721 QIKARPGMEY FLNFEVVQKT ATQLIPANHI VAVEQFKLPI TLPKQAFDQK SQKPELNITS
781 SGKSIVVKSP TVNFVFDKKS GTVTSYKVTG QEYFDKGFQI QPNFWRAPND NDFGNGNPHR
841 LQVWKQSSRH FNVDVKGYA QGNNAVVETT YLLAAGNLYT IKYTVHPSGV VKVDVEFHST
901 DMQAAQLEAS EATLMATASP EATAARKASS ELVVPRIGVR FRLPASMNAV EYFGRGPGEN
961 YIDRASGSKV GLYKTTADEM YFPYVRPQEN GHRTDTRWVA LKGAGNGLMV VADETIGFNS
1021 LKNSVEDFDS EEATRRPYQW NNFSSIEIAG RDDADAKNKR PRHTHINDIT PRNFVEVCVD
1081 MKQQGVAGYN SWGARPLPEY SIPANQNYKW GFTLVPFGNA GDIQIKSVLK Y*
//

```

## A12 MS-Bgal2 sequence

LOCUS MS\_Bgal2\_GH42\_(17A) 689 aa linear UNA

FEATURES Location/Qualifiers

Region 1..688

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 23..396

/Domain bit score=515.3

/Domain bias=4.7

/Accession=PF02449.10

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42

Region 407..619

/Domain bit score=239.1

/Domain bias=0.0

/Accession=PF08532.5

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42M

Region 630..686

/Domain bit score=36.3

/Domain bias=0.0

/Accession=PF08533.5

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42C

ORIGIN

```

1 MFIMNKFAPL SPKVNALLHG ADYNPEQWEN YPGIIDQDIA MMQQAKCNVM SVGIFSWAKL
61 EPQEGVFEFG WLDSILDKLY AAGIHVFLAT PSGARPAWMS QAYPQVLRVG RDRVPALHGG
121 RHNHCMTSPV YREKIFKINS LLAERYAQHP AVLGWHISNE YGGDCHCERC QARFRDWLKA
181 RYQTLDNLNH AWWSTFWSHY YSDWSQIESP APQGEVSIHG LNLDWRRFNT AQVDFCRHE
241 IAPLKAANAD LPVTTNFMEY FYDYDYWQLA QALDFISWDS YPMWHRDKDE TTLACYTAMY
301 HDMMRSLKGG KPFVLMESTP SATNWQPTSK LKKPGMHLS SMQAVAHGAD SVQYFQWRKS
361 RGSVEKFHGA VIDHVGHLDT RVGREVSRLG DMLSRLPGVV GCRTDAKVAI IFDQQNRWAL
421 DDAQGPRNLG MEYENTVNEH YRPFWEQGIA VDIIDADGDL SAYQLVIAPM LYMVRDGFSA
481 RAEAFVADGG HLVTTYWTGI VNESDLCHLG GFPGLRNLL GIWAEIDCL NDGERNLVQG
541 LAGNEGGLQG PYQVRHLCEL IHAESARPLA TYRDDFYAGR PAVTVNHFGK GKAWHVASRN
601 DLAFQRDFFA VISRELALPR AIESELPPGV VATARTDGET TYVFLQNYSA QQHSVSLPQG
661 YQDSLGAAI SAPLTLTAWD CRILSRKA*

```

//

## A13 MS-Bgal3 sequence

LOCUS MS\_Bgal3\_GH53 401 aa linear UNA

FEATURES Location/Qualifiers

sig\_peptide 1..23

/note="Cleavage site: 23^24: ALA-AE"

/note="D-score: 0.813"

/note="D-cutoff: 0.45"

/note="Max C-score: 0.604"

/note="Max S-score: 0.967"

/note="Max Y-score: 0.745"

/note="S-mean: 0.92"

/note="Networks: SignalP-TM"

Region 1..400

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 41..399

/Domain bit score=403.5

/Domain bias=2.0

/Accession=PF07745.8

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_53

ORIGIN

1 MKRLTPALLA VCLACSFSPA ALAAEALQTR AFRALPADFI KGADISTLLD AEKHGAKFYN

61 HNNQQQDPIA ILKADGVNYV RLRLWVDPKD AQQQGYGGGD NDLAATLALA KRAKAQGMKL

121 LLDFHYSDFW TDPGKQFKPK AWEKMDYPQL KTTIH DYTRD TIARFKQEGV LPDMVQIGNE

181 INGGMLWPEG KSWGQGGGEF DRLAGLLNAA IDGLKENLQN GEQVKIMLHL AEGTKNDTFR

241 WWFDEISKRN VPYDIIGLSM YTYWNGPISA LKANMDDISR RYNKDVIVVE AAYAYTLENC

301 DNAENSFQAK EEKEGGYPAT VQGQYNYIHD LMQAVADVPD QRGKGIFYWE PTWIAVPGNT

361 WATPAGMKYI HDEWKQGNAR ENQALFDCQG KVLPSAKVFN \*

//

## A14 MS-Gluc1 sequence

LOCUS MS\_Gluc1\_GH1\_(3L) 438 aa linear UNA

FEATURES Location/Qualifiers

Region 1..437

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 6..437

/Domain bit score=498.7

/Domain bias=0.1

/Accession=PF00232.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_1

ORIGIN

```
1 MFSFDRRDFG SNFTFGVATA AYQIEGGQGD GRGQSIWDTF SATPGNVHNG DTGRDACNHY
61 ELWAQDLDLI RDGGFDAYRF SFAWPRLIPE GTGAINQAGV DFYDRLIDGM LERGIKPYAT
121 LYHWDLPSAL QDRGGWMNRD IANWFADYAS LIAEKYGDRL AATATINEPW CVAFLSHFLG
181 VHAPGYRDLR AAARAMHHVL YAHGTAIDAL RAGGAKNLGI VLNLEKSEPA TESDEDKAAC
241 NFGDALFNRV YLGGVFKGQY PKELTEWLAP YLPANYQADM DVVSRPLDWL GINYTRSLY
301 KASQMPGRPV DQVKGPLEKT DIGWEIYPKG LSDLLVRVSN EYTKLPIFVT ENGMAEVEGD
361 NDPRRVKYYE DHLKAVLAAK KDGADVRGYF AWSLLDNYEW AEGYNKRFGV VHVYETQKR
421 TPKASFRSFQ GLLHNTR*
//
```

## A15 MS-Gluc2 sequence

LOCUS MS\_Gluc2\_GH1\_(4L) 433 aa linear UNA

FEATURES Location/Qualifiers

sig\_peptide 1..23

/note="Cleavage site: 23^24: AAA-AT"

/note="D-score: 0.738"

/note="D-cutoff: 0.45"

/note="Max C-score: 0.647"

/note="Max S-score: 0.886"

/note="Max Y-score: 0.705"

/note="S-mean: 0.79"

/note="Networks: SignalP-TM"

Region 1..432

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 28..289

/Domain bit score=134.6

/Domain bias=0.0

/Accession=PF00232.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_1

Region 269..431

/Domain bit score=89.3

/Domain bias=0.0

/Accession=PF00232.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_1

ORIGIN

1 MNRRTVIGMI GASAAASAFP AAAATGKPRR TPAGFLWGTA GAAYQVEGGN VASDLWVVEQ

61 LNTPLFAEPS GDACDVYHRY EDDLELVARL GFNCHRLGIE WSRIEPRGQ ISEALAYR

121 RVLQACVRNG LKPVITFSHF TVPRWVAASG GFKDPANIEA FAAHCARLTR TMGDLIHLAA

181 TFNEPNLSTV VRWTGLGDKI RPLIESVQRA AGASQNAPKW SSPMLGGETQ FEGIAAHTR

241 AIDAIQAGG RFPiGLTLAL PADTAAGGDD GPLKRKAEM MDRWIAAPGD FIGVQTYTST

301 PVGPDGDLPP APGTELTQMG YPFTPSAVEG AVRMAAARTS KPIYITENG V ATEDDSRRIA

361 YIDGAIAGVQ RLLADGIDLR GYIHWSLLDN WEWMHGYKPK FGLVAVDRKT FKRTPKPSAS

421 HLGAIARRGG LA\*

//

## A16 SAS-Bgal1 sequence

LOCUS SAS\_Bgal1\_GH2 1024 aa linear UNA

FEATURES Location/Qualifiers

Region 1..1023

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 47..216

/Domain bit score=174.9

/Domain bias=0.0

/Accession=PF02837.13

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_N

Region 219..330

/Domain bit score=56.6

/Domain bias=0.0

/Accession=PF00703.16

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2

Region 333..626

/Domain bit score=394.1

/Domain bias=0.0

/Accession=PF02836.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_2\_C

Region 749..1021

/Domain bit score=277.4

/Domain bias=0.0

/Accession=PF02929.12

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Bgal\_small\_N

ORIGIN

```

1  MPSTLASLLS RRDWENPVVT QWHRLAAHAP MRSWRDETA RDEADSAHH SLNGIWQFSF
61  FASPEAVPEQ WLEQDCADAV AMPVPSNWQM QGFDTPIYTN VTYPVPVNP FVPQQNPTGC
121 YSLTFNVDAE AIAEQTRIV FDGVNSAFHL WCNGQWVGYS QDSRLPAEFD LSAVLRAGEN
181 RLAVMVLRW C DGSYLEDQDM WRMSGIFRDV SLQHKPALHI ADYHYTTALN AEFTRASVQV
241 TVELAGEFAG SRVTAVLWRN GEKIASGEQT PGSVVDERG NWAERLSLTL PVESPLLWSA
301 ETPNLYRLTL TLYDAQGSCV EAEACDVGFR HVEIHQGLLK LNGQPLLIRG VNRHEHHPEH
361 GQAVDEATMW RDIELMKQHN FNAVRCSHYP NHPLWYRLCD RYGLYVVDEA NIETHGMVPM
421 SRLADDPRLW PAMSERVTRM VQRDRNHPSI IIVSLGNESG HGANQDALYR WLKSTDPTRP
481 VQYEGGGANT AATDIICPMY SRVDQDQFPF AVPKWSIKKW IGMPNETRPL ILCEYAHAMG
541 NSFGGFAKYW QAFRAAPRLQ GGFVWDWVDQ ALTKTGADGE LFWAYGGDFG DTPNDRQFCM
601 NGLVFPDRTP HPALFEAQR AQQFFQFSLLS ASPLTIEVTS EYLFRTSDNE VLRWRVERDG
661 EVLAQGETAL VIAPQGVQVI ALDLPALAAA PGEVWLNAAEV YQREATAWSA AGHRCARDQW
721 RLPAPLWVAV PDKQGEQPTL EVNDETYTVK QGNQRWAFCR QRGNLVQWWR DGQETLLTPV
781 TDCFTRAPLD NDIGVSEVTR IDPNAWMERW KAAGMYDLHA ELLSFEVQES AQEVTLRVH
841 RWIGAGKSAF ISEKTWRIDR TGGLQADIEV VVANDIPPPA RIGLVCQLAE RPAEVSWLGL
901 GPHENYPDRK LAACQGRWTQ PLAALHTPYI FPGENGLRCD TRLLQCGAHQ LEGRFHFLSLG
961 CYSDTQLRET THHLLREEA GCWLHLDAFH MGVGGDDSW SPSVSPEFILQ QERVRYRLRW
1021 QQA*
//

```

## A17 SAS-BGal2 sequence

LOCUS SAS\_Bgal2\_\_GH42 686 aa linear UNA

FEATURES Location/Qualifiers

Region 1..685

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 20..393

/Domain bit score=516.5

/Domain bias=5.4

/Accession=PF02449.10

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42

Region 404..616

/Domain bit score=240.9

/Domain bias=0.0

/Accession=PF08532.5

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42M

Region 627..683

/Domain bit score=40.2

/Domain bias=0.0

/Accession=PF08533.5

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_42C

ORIGIN

```

1 MNKFAPLSPK VVSLHGDY NPEQWENYPG IIDKDIAMMQ QAKCNVMSVG IFSWSKLEPQ
61 EGVFNFAWLD EVIEKLYAAG IHIFLATPSG ARPAWMSQKY PQVLRVGRDR VPALHGGRHN
121 HCMTSPVYRE KTLKINTLLA ERYGQHPAVL GWHISNEYGG ECHCDLCQQK FRDWLKARYQ
181 TLEALNHAWW SDFWSHTYSY WSQIESPAPQ GEVSIHGLNL DWRRFNTAQV TDFCRHEVAP
241 LKAANAALPV TTNFMEYFYD YDYWQLAEAI DFISWDSYPM WHRDKDETQL ACYTAMYHDM
301 MRTLKGGKPF VLMESTPSVT NWQPTSKLKK PGMHILSSLQ AVAHGADSVQ YFQWRKSRGS
361 VEKFGAVID HVGHLDRVGV REVSALGEML SKLTPVMGCR TEARVAIIFD QQNRWALDDA
421 QGPRNKGMEY EKTVNEHYRP FWEKGVAVDV INADADLGRY HLVIAPMLYM VRDGFAERAE
481 SFVASGGHLV TSYWSGVVNE SDLCHLGGFP GPLRNLLGIW AEEIDCLDDH ERNLVQGLAG
541 NEAGLQGPYQ VRHLCELIQL EGAQPLATYR DDFYAGRPV TVNAFGKGA WHVASRNDLA
601 QRDFFANLI DTLSPRALN VELSPGVVAT AREDEQAQFV FIQNFTAQQQ SITLPPGYHD
661 CLSEAAVSGA LALKPWDCRV VRRDA*

```

//

## A18 SAS-BGal3 sequence

LOCUS SAS\_Bgal3\_\_GH53 396 aa linear UNA

FEATURES Location/Qualifiers

sig\_peptide 1..21

/note="Cleavage site: 21^22: ASA-VT"

/note="D-score: 0.639"

/note="D-cutoff: 0.45"

/note="Max C-score: 0.223"

/note="Max S-score: 0.973"

/note="Max Y-score: 0.454"

/note="S-mean: 0.928"

/note="Networks: SignalP-TM"

sig\_peptide 1..21

/note="Cleavage site: 21^22: ASA-VT"

/note="D-score: 0.892"

/note="D-cutoff: 0.57"

/note="Max C-score: 0.718"

/note="Max S-score: 0.978"

/note="Max Y-score: 0.832"

/note="S-mean: 0.959"

/note="Networks: SignalP-noTM"

Region 1..395

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 36..394

/Domain bit score=398.2

/Domain bias=1.9

/Accession=PF07745.8

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_53

ORIGIN

```

1 MKRLTPALLA VCLAASFAS AVTTRPFTQM PADFIKGADI STLLDAEKHG ATFYNDKNQP
61 QDALAILKAN GVNIVRLRLW VDPKDAQGNA YGGGDNDLAT TLALAKRAKA QGMKLLLDLH
121 YSDFWTDPGK QFKPKAWKEL DYPQLKSRVH DYTRDTIAQF KQAGVLPDMV QIGNEINGGM
181 LWPEGKSWGQ GGGEFDRLAG LLTAGIDGVK ENLNNGEQVK IMLHLAEGTK NDTFRWWFDE
241 ITKRNVPPFDV IGLSMYTYWN GPISALQANM DDISKRYNKD VIVVEAAYAY TLANCDNAEN
301 SFQAKEEKAG GYPATVQGQY DYVHDLMSV INVQGQRGKG IFYWEPTWIA VPGNGWATPA
361 GMKYINDHWK EGNARENQAL FDCQGVLPV VKVFN*

```

//

## A19 SAS-LS sequence

LOCUS SAS\_LS\_GH68\_(Ls2) 430 aa linear UNA

FEATURES Location/Qualifiers

Region 1..429

/region\_type=Extracellular

/note="TMHMM2.0 OUTSIDE"

/label=outside

Region 2..410

/Domain bit score=407.2

/Domain bias=0.1

/Accession=PF02435.11

/Predicted by="HMMER 3.1b1 (May 2013)"

/label=Glyco\_hydro\_68

ORIGIN

```
1 MTYETSAWTI ADALKVRADD PTTMPVIAQ NFPVIDEALW QWDTGALRTI RGATVTFKGW
61 YVMWALVANK ADTGATVEGW HNRNAFAYIG YYYSRDGLDW TFGGRLDKS ADLRPDEWGS
121 GLVMREGTEN VVDMFYTSVN TDTNQSVPVS SSGRILADAN GVWFDGFTST TEMFAADGVH
181 YANADEDQYF DFRDPHPFLN PADGKIYCLF EGNVAGIRGQ FVISDRERGP TPPAYDVDAG
241 AQYGAAAIGI ARLDGNYSKG EFDKWTLPP LVTALGVNDQ TERPHVVFQD GRTYLFTISH
301 HSTYTGNLGS PDGVYGFVSD KGIFGPYRPL NGSGLVLGNP SVAPYETYSH FVDPQGYVQS
361 FIDTLPAPDT VDPQNPVTYR IGGTLAPTVR ILLDGERTFL TEVHGYGQIF VQGAWPVRNT
421 PDVRPVATS*
//
```

## BIBLIOGRAPHY

1. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI: **Host-bacterial mutualism in the human intestine.** *Science* 2005, **307**(5717):1915-1920.
2. Hawrelak JA, Myers SP: **The causes of intestinal dysbiosis: A review.** *Alternative Medicine Review* 2004, **9**(2):180-197.
3. Petersen C, Round JL: **Defining dysbiosis and its influence on host immunity and disease.** *Cell Microbiol* 2014, **16**(7):1024-1033.
4. Serino M, Blasco-Baque V, Nicolas S, Burcelin R: **Far from the Eyes, Close to the Heart: Dysbiosis of Gut Microbiota and Cardiovascular Consequences.** *Curr Cardiol Rep* 2014, **16**(11):1-7.
5. Parracho HMRT, Bingham MO, Gibson GR, McCartney AL: **Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children.** *Journal of Medical Microbiology* 2005, **54**(10):987-991.
6. Cryan JF, Dinan TG: **Mind-altering microorganisms: The impact of the gut microbiota on brain and behaviour.** *Nat Rev Neurosci* 2012, **13**(10):701-712.
7. Scott KP, Antoine J-M, Midtvedt T, Hemert Sv: **Manipulating the gut microbiota to maintain health and treat disease.** *Microbial Ecology in Health & Disease* 2015, **26**.
8. Salminen S, Bouley C, Boutron-Ruault MC, Cummings JH, Franck A, Gibson GR, Isolauri E, Moreau MC, Roberfroid M, Rowland I: **Functional food science and gastrointestinal physiology and function.** *Br J Nutr* 1998, **80**(SUPPL. 1):S147-S171.
9. Nations WHOFAOotU: **Health and Nutritional Properties of Probiotics in Food including Powder Milk with Live Lactic Acid Bacteria.** In. American Córdoba Park Hotel, Córdoba, Argentina: FAO/WHO Expert Consultation; 2001.
10. Frei R, Akdis M, O'Mahony L: **Prebiotics, probiotics, synbiotics, and the immune system: Experimental data and clinical evidence.** *Curr Opin Gastroenterol* 2015, **31**(2):153-158.
11. Lin CS, Chang CJ, Lu CC, Martel J, Ojcius DM, Ko YF, Young JD, Lai HC: **Impact of the gut microbiota, prebiotics, and probiotics on human health and disease.** *Biomed J* 2014, **37**(5):259-268.
12. Hajela N, Ramakrishna BS, Nair GB, Abraham P, Gopalan S, Ganguly NK: **Gut microbiome, gut function, and probiotics: Implications for health.** *Indian Journal of Gastroenterology* 2015, **34**(2):93-107.
13. Vandenplas Y, Huys G, Daube G: **Probiotics: An update.** *J Pediatr* 2015, **91**(1):6-21.
14. Patel R, Dupont HL: **New approaches for bacteriotherapy: Prebiotics, new-generation probiotics, and synbiotics.** *Clin Infect Dis* 2015, **60**:S108-S121.
15. Pandey KR, Naik SR, Vakil BV: **Probiotics, prebiotics and synbiotics- a review.** *J Food Sci Technol* 2015, **52**(12):7577-7587.
16. Gibson GR, Scott KP, Rastall RA, Tuohy KM, Hotchkiss A, Dubert-Ferrandon A, Gareau M, Murphy EF, Saulnier D, Loh G *et al*: **Dietary prebiotics: current status and new definition.** *Food Science and Technology Bulletin: Functional Foods* 2011, **7**(1):1-19.
17. Slavin J: **Fiber and prebiotics: Mechanisms and health benefits.** *Nutrients* 2013, **5**(4):1417-1435.
18. Vitetta L, Briskey D, Alford H, Hall S, Coulson S: **Probiotics, prebiotics and the gastrointestinal tract in health and disease.** *Inflammopharmacology* 2014, **22**(3):135-154.
19. Saulnier DM, Ringel Y, Heyman MB, Foster JA, Bercik P, Shulman RJ, Versalovic J, Verdu EF, Dinan TG, Hecht G *et al*: **The intestinal microbiome, probiotics and prebiotics in neurogastroenterology.** *Gut Microbes* 2013, **4**(1):17-27.
20. Molinaro F, Paschetta E, Cassader M, Gambino R, Musso G: **Probiotics, Prebiotics, Energy Balance, and Obesity. Mechanistic Insights and Therapeutic Implications.** *Gastroenterol Clin North Am* 2012, **41**(4):843-854.

21. Al-Sheraji SH, Ismail A, Manap MY, Mustafa S, Yusof RM, Hassan FA: **Prebiotics as functional foods: A review.** *J Funct Foods* 2013, **5**(4):1542-1553.
22. Desmet T, Soetaert W: **Enzymatic glycosyl transfer: Mechanisms and applications.** *Biocatalysis and Biotransformation* 2011, **29**(1):1-18.
23. Fleuri LF, Okino-Delgado CH, Novelli PK, Lima GPP, Pedrosa VA, Francisco VB, Andriolli G, Almendra ACR: **Enzymatic production of functional oligosaccharides - A review.** *Int Sugar J* 2014, **116**(1387):496-503.
24. Panesar PS, Kumari S, Panesar R: **Biotechnological approaches for the production of prebiotics and their potential applications.** *Critical Reviews in Biotechnology* 2013, **33**(4):345-364.
25. Hasler CM: **Functional foods: Benefits, concerns and challenges - A position paper from the American Council on Science and Health.** *Journal of Nutrition* 2002, **132**(12):3772-3781.
26. Kumar H, Salminen S, Verhagen H, Rowland I, Heimbach J, Bañares S, Young T, Nomoto K, Lalonde M: **Novel probiotics and prebiotics: Road to the market.** *Current Opinion in Biotechnology* 2015, **32**:99-103.
27. Vandenplas Y, De Greef E, Veereman G: **Prebiotics in infant formula.** *Gut Microbes* 2015, **5**(6):681-687.
28. Blaser M, Bork P, Fraser C, Knight R, Wang J: **The microbiome explored: Recent insights and future challenges.** *Nature Reviews Microbiology* 2013, **11**(3):213-217.
29. Savage DC: **Microbial ecology of the gastrointestinal tract.** *Annual Review of Microbiology* 1977, **31**:107-133.
30. Oliveira DL, Wilbey RA, Grandison AS, Roseiro LB: **Milk oligosaccharides: A review.** *Int J Dairy Technol* 2015, **68**(3):305-321.
31. Walker WA, Iyengar RS: **Breast milk, microbiota, and intestinal immune homeostasis.** *Pediatric Research* 2015, **77**:220-228.
32. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T *et al*: **A human gut microbial gene catalogue established by metagenomic sequencing.** *Nature* 2010, **464**(7285):59-65.
33. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C *et al*: **The NIH Human Microbiome Project.** *Genome Research* 2009, **19**(12):2317-2323.
34. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI: **The Human Microbiome Project.** *Nature* 2007, **449**(7164):804-810.
35. Walter J, Ley R: **The human gut microbiome: Ecology and recent evolutionary changes.** In: *Annual Review of Microbiology*. vol. 65; 2011: 411-429.
36. Zoetendal EG, Rajilic-Stojanovic M, de Vos WM: **High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota.** *Gut* 2008, **57**(11):1605-1615.
37. Sankar SA, Lagier JC, Pontarotti P, Raoult D, Fournier PE: **The human gut microbiome, a taxonomic conundrum.** *Systematic and Applied Microbiology* 2015, **38**(4):276-286.
38. Schippa S, Conte MP: **Dysbiotic events in gut microbiota: Impact on human health.** *Nutrients* 2014, **6**(12):5786-5805.
39. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM *et al*: **Enterotypes of the human gut microbiome.** *Nature* 2011.
40. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R *et al*: **Linking long-term dietary patterns with gut microbial enterotypes.** *Science* 2011, **334**(6052):105-108.
41. Yong E: **Gut microbial 'enterotypes' become less clear-cut. Communities of gut bacteria may form a spectrum rather than falling into distinct groups.** In: *Nature*. 2012.
42. Turnbaugh PJ, Gordon JI: **The core gut microbiome, energy balance and obesity.** *J Physiol* 2009, **587**(17):4153-4158.
43. Tremaroli V, Bäckhed F: **Functional interactions between the gut microbiota and host metabolism.** *Nature* 2012, **489**(7415):242-249.

44. Puertollano E, Kolida S, Yaqoob P: **Biological significance of short-chain fatty acid metabolism by the intestinal microbiome.** *Current Opinion in Clinical Nutrition and Metabolic Care* 2014, **17**(2):139-144.
45. Den Besten G, Van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM: **The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism.** *Journal of Lipid Research* 2013, **54**(9):2325-2340.
46. Bergman EN: **Energy contributions of volatile fatty acids from the gastrointestinal tract in various species.** *Physiological Reviews* 1990, **70**(2):567-590.
47. Krishnan S, Alden N, Lee K: **Pathways and functions of gut microbiota metabolism impacting host physiology.** *Current Opinion in Biotechnology* 2015, **36**:137-145.
48. Kondo T, Kishi M, Fushimi T, Ugajin S, Kaga T: **Vinegar intake reduces body weight, body fat mass, and serum triglyceride levels in obese Japanese subjects.** *Bioscience, Biotechnology and Biochemistry* 2009, **73**(8):1837-1843.
49. Fushimi T, Suruga K, Oshima Y, Fukiharuru M, Tsukamoto Y, Goda T: **Dietary acetic acid reduces serum cholesterol and triacylglycerols in rats fed a cholesterol-rich diet.** *Br J Nutr* 2006, **95**(5):916-924.
50. Kobozev I, Reinoso Webb C, Furr KL, Grisham MB: **Role of the enteric microbiota in intestinal homeostasis and inflammation.** *Free Radic Biol Med* 2014, **68**:122-133.
51. Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, Reading NC, Villablanca EJ, Wang S, Mora JR *et al*: **Gut immune maturation depends on colonization with a host-specific microbiota.** *Cell* 2012, **149**(7):1578-1593.
52. Belkaid Y, Hand TW: **Role of the microbiota in immunity and inflammation.** *Cell* 2014, **157**(1):121-141.
53. Wu HJ, Wu E: **The role of gut microbiota in immune homeostasis and autoimmunity.** *Gut Microbes* 2012, **3**(1).
54. Brandtzaeg P, Halstensen TS, Kett K, Krajci P, Kvale D, Rognum TO, Scott H, Sollid LM: **Immunobiology and immunopathology of human gut mucosa: Humoral immunity and intraepithelial lymphocytes.** *Gastroenterology* 1989, **97**(6):1562-1584.
55. Hapfelmeier S, Lawson MAE, Slack E, Kirundi JK, Stoel M, Heikenwalder M, Cahenzli J, Velykoredko Y, Balmer ML, Endt K *et al*: **Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses.** *Science* 2010, **328**(5986):1705-1709.
56. Carding S, Verbeke K, Vipond DT, Corfe BM: **Dysbiosis of the gut microbiota in disease.** *Microbial Ecology in Health & Disease* 2015, **26**.
57. Vernon G, Baranova A, Younossi ZM: **Systematic review: The epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults.** *Alimentary Pharmacology and Therapeutics* 2011, **34**(3):274-285.
58. Mullin GE: **Intestinal dysbiosis: A possible mechanism of alcohol-induced endotoxemia and alcoholic steatohepatitis in rats.** *Nutr Clin Prac* 2010, **25**(3):312-313.
59. Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M *et al*: **Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome.** *Nature* 2013, **499**(7456):97-101.
60. Baumgart DC, Carding SR: **Inflammatory bowel disease: cause and immunobiology.** *Lancet* 2007, **369**(9573):1627-1640.
61. Hansen J, Gulati A, Sartor RB: **The role of mucosal immunity and host genetics in defining intestinal commensal bacteria.** *Curr Opin Gastroenterol* 2010, **26**(6):564-571.
62. Zimmet P, Magliano D, Matsuzawa Y, Alberti G, Shaw J: **The metabolic syndrome: a global public health problem and a new definition.** *Journal of atherosclerosis and thrombosis* 2005, **12**(6):295-300.
63. Haslam DW, James WPT: **Obesity.** *Lancet* 2005, **366**(9492):1197-1209.
64. Blaut M: **Gut microbiota and energy balance: Role in obesity.** *Proceedings of the Nutrition Society* 2015, **74**(3):227-234.
65. Ley RE, Turnbaugh PJ, Klein S, Gordon JI: **Microbial ecology: Human gut microbes associated with obesity.** *Nature* 2006, **444**(7122):1022-1023.

66. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR *et al*: **Gut microbiota from twins discordant for obesity modulate metabolism in mice.** *Science* 2013, **341**(6150).
67. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest.** *Nature* 2006, **444**(7122):1027-1031.
68. Slavin JL: **Dietary fiber and body weight.** *Nutrition* 2005, **21**(3):411-418.
69. Forsythe P, Kunze WA: **Voices from within: Gut microbes and the CNS.** *Cellular and Molecular Life Sciences* 2013, **70**(1):55-69.
70. Dillon RJ, Vennard CT, Charnley AK: **Pheromones: Exploitation of gut bacteria in the locust.** *Nature* 2000, **403**(6772):851.
71. Iyer LM, Aravind L, Coon SL, Klein DC, Koonin EV: **Evolution of cell-cell signaling in animals: Did late horizontal gene transfer from bacteria have a role?** *Trends in Genetics* 2004, **20**(7):292-299.
72. Sharon G, Segal D, Zilber-Rosenberg I, Rosenberg E: **Symbiotic bacteria are responsible for diet-induced mating preference in *Drosophila melanogaster*, providing support for the hologenome concept of evolution.** *Gut Microbes* 2011, **2**(3):190-192.
73. Carabotti M, Scirocco A, Maselli MA, Severi C: **The gut-brain axis: Interactions between enteric microbiota, central and enteric nervous systems.** *Annals of Gastroenterology* 2015, **28**(2):203-209.
74. Gareau MG, Wine E, Rodrigues DM, Cho JH, Whary MT, Philpott DJ, MacQueen G, Sherman PM: **Bacterial infection causes stress-induced memory dysfunction in mice.** *Gut* 2011, **60**(3):307-317.
75. Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S: **Normal gut microbiota modulates brain development and behavior.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(7):3047-3052.
76. Foster JA, McVey Neufeld KA: **Gut-brain axis: How the microbiome influences anxiety and depression.** *Trends Neurosci* 2013, **36**(5):305-312.
77. Naseribafrouei A, Hestad K, Avershina E, Sekelja M, Linløkken A, Wilson R, Rudi K: **Correlation between the human fecal microbiota and depression.** *Neurogastroenterology and Motility* 2014, **26**(8):1155-1162.
78. Mayer EA, Padua D, Tillisch K: **Altered brain-gut axis in autism: Comorbidity or causative mechanisms?** *BioEssays* 2014, **36**(10):933-939.
79. Song Y, Liu C, Finegold SM: **Real-time PCR quantitation of clostridia in feces of autistic children.** *Applied and Environmental Microbiology* 2004, **70**(11):6459-6465.
80. Savel RH, Munro CL: **From asclepius to hippocrates: The art and science of healing.** *American Journal of Critical Care* 2014, **23**(6):437-439.
81. Mizock BA: **Probiotics.** *Dis Mon* 2015, **61**(7):259-290.
82. Hoppe B: **An update on primary hyperoxaluria.** *Nature Reviews Nephrology* 2012, **8**(8):467-475.
83. Stewart CS, Duncan SH, Cave DR: **Oxalobacter formigenes and its role in oxalate metabolism in the human gut.** *FEMS Microbiology Letters* 2004, **230**(1):1-7.
84. Siva S, Barrack ER, Reddy GPV, Thamilselvan V, Thamilselvan S, Menon M, Bhandari M: **A critical analysis of the role of gut *Oxalobacter formigenes* in oxalate stone disease.** *BJU International* 2009, **103**(1):18-21.
85. Sanders ME: **Probiotics: Considerations for human health.** *Nutrition Reviews* 2003, **61**(3):91-99.
86. Merenstein D, El-Nachef N, Lynch SV: **Fecal microbial therapy: Promises and pitfalls.** *Journal of Pediatric Gastroenterology and Nutrition* 2014, **59**(2):157-161.
87. Konturek PC, Haziri D, Brzozowski T, Hess T, Heyman S, Kwiecien S, Konturek SJ, Koziel J: **Emerging role of fecal microbiota therapy in the treatment of gastrointestinal and extra-gastrointestinal diseases.** *Journal of Physiology and Pharmacology* 2015, **66**(4):483-491.
88. Eiseman B, Silen W, Bascom GS, Kauvar AJ: **Fecal enema as an adjunct in the treatment of pseudomembranous.** *Surgery* 1958, **44**(5):854-859.

89. Paramsothy S, Borody TJ, Lin E, Finlayson S, Walsh AJ, Samuel D, Van Den Bogaerde J, Leong RWL, Connor S, Ng W *et al*: **Donor recruitment for fecal microbiota transplantation.** *Inflammatory Bowel Dis* 2015, **21**(7):1600-1606.
90. Di Bartolomeo F, Startek JB, Van Den Ende W: **Prebiotics to fight diseases: Reality or fiction?** *Phytotherapy Research* 2013, **27**(10):1457-1473.
91. Peshev D, Van den Ende W: **Fructans: Prebiotics and immunomodulators.** *J Funct Foods* 2014, **8**(1):348-357.
92. Scott KP, Martin JC, Duncan SH, Flint HJ: **Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro.** *FEMS Microbiology Ecology* 2014, **87**(1):30-40.
93. Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Doré J, Henrissat B, Remaud-Siméon M, Monsan P, Potocki-Véronèse G: **Functional Metagenomics Reveals Novel Pathways of Prebiotic Breakdown by Human Gut Bacteria.** *PLoS ONE* 2013, **8**(9).
94. Cénit MC, Matzaraki V, Tigchelaar EF, Zhernakova A: **Rapidly expanding knowledge on the role of the gut microbiome in health and disease.** *Biochim Biophys Acta Mol Basis Dis* 2014, **1842**(10):1981-1992.
95. Roberfroid M: **Prebiotics: The concept revisited.** *Journal of Nutrition* 2007, **137**(3):830S-837S.
96. De Vrese M, Schrezenmeir J: **Probiotics, prebiotics, and synbiotics.** In: Edited by Stahl U, Donalies UEB, Nevoigt E, vol. 111; 2008: 1-66.
97. Cencic A, Chingwaru W: **The role of functional foods, nutraceuticals, and food supplements in intestinal health.** *Nutrients* 2010, **2**(6):611-625.
98. Romeo J, Nova E, Wärnberg J, Gómez-Martínez S, Ligia LED, Marcos A: **Immunomodulatory effect of fibres, probiotics and synbiotics in different life-stages.** *Nutr Hosp* 2010, **25**(3):341-349.
99. Zhang MM, Cheng JQ, Lu YR, Yi ZH, Yang P, Wu XT: **Use of pre-, pro-and synbiotics in patients with acute pancreatitis: A meta-analysis.** *World J Gastroenterol* 2010, **16**(31):3970-3978.
100. Sanz ML, Gibson GR, Rastall RA: **Influence of disaccharide structure on prebiotic selectivity in vitro.** *Journal of Agricultural and Food Chemistry* 2005, **53**(13):5192-5199.
101. Díez-Municio M, Herrero M, Olano A, Moreno FJ: **Synthesis of novel bioactive lactose-derived oligosaccharides by microbial glycoside hydrolases.** *Microb Biotechnol* 2014, **7**(4):315-331.
102. Yadav SA, Agte VV, Nilegaonkar SS: **Enrichment of prebiotics in foods using green chemistry approach.** *Curr Org Chem* 2014, **18**(23):2961-2971.
103. Bhatia Y, Mishra S, Bisaria VS: **Microbial  $\beta$ -glucosidases: Cloning, properties, and applications.** *Critical Reviews in Biotechnology* 2002, **22**(4):375-407.
104. Gurung N, Ray S, Bose S, Rai V: **A broader view: Microbial enzymes and their relevance in industries, medicine, and beyond.** *BioMed Research International* 2013, **2013**.
105. Filice M, Marciello M: **Enzymatic synthesis of oligosaccharides: A powerful tool for a sweet challenge.** *Curr Org Chem* 2013, **17**(7):701-718.
106. Monsan P, Paul F: **Enzymatic synthesis of oligosaccharides.** *FEMS Microbiology Reviews* 1995, **16**(2-3):187-192.
107. Henrissat B, Davies GJ: **Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics.** *Plant Physiology* 2000, **124**(4):1515-1519.
108. Plou FJ, Martín MT, Gomez De Segura A, Alcalde M, Ballesteros A: **Glucosyltransferases acting on starch or sucrose for the synthesis of oligosaccharides.** *Can J Chem* 2002, **80**(6):743-752.
109. Coutinho PM, Deleury E, Davies GJ, Henrissat B: **An evolving hierarchical family classification for glycosyltransferases.** *Journal of Molecular Biology* 2003, **328**(2):307-317.
110. Plou FJ, Segura AGD, Ballesteros A: **Application of glycosidases and transglycosidases in the synthesis of oligosaccharides.** In: *Industrial Enzymes: Structure, Function and Applications.* Springer Netherlands; 2007: 141-157.
111. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities.** *Biochemical Journal* 1991, **280**(2):309-316.
112. Henrissat B, Davies G: **Structural and sequence-based classification of glycoside hydrolases.** *Current Opinion in Structural Biology* 1997, **7**(5):637-644.

113. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013**. *Nucleic Acids Research* 2014, **42**(D1):D490-D495.
114. Palcic MM: **Glycosyltransferases as biocatalysts**. *Current Opinion in Chemical Biology* 2011, **15**(2):226-233.
115. Singh RS, Singh RP: **Production of fructooligosaccharides from inulin by endoinulinases and their prebiotic potential**. *Food Technology and Biotechnology* 2010, **48**(4):435-450.
116. Sangeetha PT, Ramesh MN, Prapulla SG: **Recent trends in the microbial production, analysis and application of Fructooligosaccharides**. *Trends in Food Science and Technology* 2005, **16**(10):442-457.
117. Torres DP, Gonçalves M, Teixeira JA, Rodrigues LR: **Galacto-Oligosaccharides: Production, properties, applications, and significance as prebiotics**. *Comprehensive Reviews in Food Science and Food Safety* 2010, **9**(5):438-454.
118. Panesar PS, Panesar R, Singh RS, Kennedy JF, Kumar H: **Microbial production, immobilization and applications of  $\beta$ -D-galactosidase**. *Journal of Chemical Technology and Biotechnology* 2006, **81**(4):530-543.
119. Martínez-Villaluenga C, Cardelle-Cobas A, Olano A, Corzo N, Villamiel M, Jimeno ML: **Enzymatic synthesis and identification of two trisaccharides produced from lactulose by transgalactosylation**. *Journal of Agricultural and Food Chemistry* 2008, **56**(2):557-563.
120. Cardelle-Cobas A, Corzo N, Villamiel M, Olano A: **Isomerization of lactose-derived oligosaccharides: A case study using sodium aluminate**. *Journal of Agricultural and Food Chemistry* 2008, **56**(22):10954-10959.
121. Park NH, Choi HJ, Oh DK: **Lactosucrose production by various microorganisms harboring levansucrase activity**. *Biotechnology Letters* 2005, **27**(7):495-497.
122. Han WC, Byun SH, Kim MH, Eun HS, Jung DL, Byung HU, Chul HK, Soon AK, Jang KH: **Production of lactosucrose from sucrose and lactose by a levansucrase from *Zymomonas mobilis***. *Journal of Microbiology and Biotechnology* 2009, **19**(10):1153-1160.
123. Lu L, Fu F, Zhao R, Jin L, He C, Xu L, Xiao M: **A recombinant levansucrase from *Bacillus licheniformis* 8-37-0-1 catalyzes versatile transfructosylation reactions**. *Process Biochemistry* 2014, **49**(9):1503-1510.
124. Wei L, Xiaoli X, Shufen T, Bing H, Lin T, Yi S, Hong Y, Xiaoxiong Z: **Effective enzymatic synthesis of lactosucrose and its analogues by  $\beta$ -d-galactosidase from *Bacillus circulans***. *Journal of Agricultural and Food Chemistry* 2009, **57**(9):3927-3933.
125. Pilgrim A, Kawase M, Ohashi M, Fujita K, Murakami K, Hashimoto K: **Reaction kinetics and modeling of the enzyme-catalyzed production of lactosucrose using  $\beta$ -fructofuranosidase from *Arthrobacter* sp. K-1**. *Bioscience, Biotechnology and Biochemistry* 2001, **65**(4):758-765.
126. Díez-Municio M, Herrero M, Jimeno ML, Olano A, Moreno FJ: **Efficient synthesis and characterization of lactulosucrose by *Leuconostoc mesenteroides* B-512F dextranucrase**. *Journal of Agricultural and Food Chemistry* 2012, **60**(42):10564-10571.
127. Díez-Municio M, Montilla A, Jimeno ML, Corzo N, Olano A, Moreno FJ: **Synthesis and characterization of a potential prebiotic trisaccharide from cheese whey permeate and sucrose by *Leuconostoc mesenteroides* dextranucrase**. *Journal of Agricultural and Food Chemistry* 2012, **60**(8):1945-1953.
128. Casci T, Rastall RA: **Manufacture of Prebiotic Oligosaccharides**. In.: John Wiley and Sons; 2012: 29-55.
129. Goffin D, Delzenne N, Blecker C, Hanon E, Deroanne C, Paquot M: **Will isomalto-oligosaccharides, a well-established functional food in Asia, break through the European and American market? The status of knowledge on these prebiotics**. *Crit Rev Food Sci Nutr* 2011, **51**(5):394-409.
130. Remaud-Simeon M, Willemot RM, Sarçabal P, Potocki De Montalk G, Monsan P: **Glucansucrases: Molecular engineering and oligosaccharide synthesis**. *J Mol Catal B Enzym* 2000, **10**(1-3):117-128.

131. Chung CH, Day DF: **Glucooligosaccharides from *Leuconostoc mesenteroides* B-742 (ATCC 13146): A potential prebiotic.** *Journal of Industrial Microbiology and Biotechnology* 2002, **29**(4):196-199.
132. Kim YM, Kang HK, Moon YH, Nguyen TTH, Day DF, Kim D: **Production and Bioactivity of Glucooligosaccharides and Glucosides Synthesized using Glucansucrases.** In: *Food Oligosaccharides: Production, Analysis and Bioactivity.* Wiley Blackwell; 2014: 168-183.
133. Fujimoto Y, Hattori T, Uno S, Murata T, Usui T: **Enzymatic synthesis of gentiooligosaccharides by transglycosylation with  $\beta$ -glycosidases from *Penicillium multicolor*.** *Carbohydrate Research* 2009, **344**(8):972-978.
134. Qin Y, Zhang Y, He H, Zhu J, Chen G, Li W, Liang Z: **Screening and identification of a fungal  $\beta$ -glucosidase and the enzymatic synthesis of gentiooligosaccharide.** *Applied Biochemistry and Biotechnology* 2011, **163**(8):1012-1019.
135. Spiro MD, Kates KA, Koller AL, O'Neill MA, Albersheim P, Darvill AG: **Purification and characterization of biologically active 1,4-linked  $\alpha$ -d-oligogalacturonides after partial digestion of polygalacturonic acid with endopolygalacturonase.** *Carbohydrate Research* 1993, **247**(C):9-20.
136. Olano-Martin E, Mountzouris KC, Gibson GR, Rastall RA: **Continuous production of pectic oligosaccharides in an enzyme membrane reactor.** *Journal of Food Science* 2001, **66**(7):966-971.
137. Canedo M, Jimenez-Estrada M, Cassani J, López-Munguía A: **Production of maltosylfructose (Erlöse) with levansucrase from *Bacillus subtilis*.** *Biocatalysis and Biotransformation* 1999, **16**(6):475-485.
138. Díez-Municio M, de las Rivas B, Jimeno ML, Muñoz R, Moreno FJ, Herrero M: **Enzymatic synthesis and characterization of fructooligosaccharides and novel maltosylfructosides by inulosucrase from *Lactobacillus gasseri* DSM 20604.** *Applied and Environmental Microbiology* 2013, **79**(13):4129-4140.
139. Irague R, Tarquis L, André I, Moulis C, Morel S, Monsan P, Potocki-Véronèse G, Remaud-Siméon M: **Combinatorial Engineering of Dextransucrase Specificity.** *PLoS ONE* 2013, **8**(10).
140. Di Bartolomeo F, Van den Ende W: **Fructose and Fructans: Opposite Effects on Health?** *Plant Foods for Human Nutrition* 2015, **70**(3):227-237.
141. Bremer AA, Stanhope KL, Graham JL, Cummings BP, Wang W, Saville BR, Havel PJ: **Fructose-Fed Rhesus Monkeys: A Nonhuman Primate Model of Insulin Resistance, Metabolic Syndrome, and Type 2 Diabetes.** *Clinical and Translational Science* 2011, **4**(4):243-252.
142. Valluru R, Van Den Ende W: **Plant fructans in stress environments: Emerging concepts and future prospects.** *Journal of Experimental Botany* 2008, **59**(11):2905-2916.
143. Phelps CF: **THE PHYSICAL PROPERTIES OF INULIN SOLUTIONS.** *Biochem J* 1965, **95**:41-47.
144. Visnapuu T, Mardo K, Alamäe T: **Levansucrases of a *Pseudomonas syringae* pathovar as catalysts for the synthesis of potentially prebiotic oligo- and polysaccharides.** *New Biotechnology* 2015, **32**(6):597-605.
145. Van Den Ende W, Peshev D, De Gara L: **Disease prevention by natural antioxidants and prebiotics acting as ROS scavengers in the gastrointestinal tract.** *Trends in Food Science and Technology* 2011, **22**(12):689-697.
146. Hendry GAF: **Evolutionary origins and natural functions of fructans - a climatological, biogeographic and mechanistic appraisal.** *New Phytologist* 1993, **123**(1):3-14.
147. Chassard C, Lacroix C: **Carbohydrates and the human gut microbiota.** *Current Opinion in Clinical Nutrition and Metabolic Care* 2013, **16**(4):453-460.
148. Shoaib M, Shehzad A, Omar M, Rakha A, Raza H, Sharif HR, Shakeel A, Ansari A, Niazi S: **Inulin: Properties, health benefits and food applications.** *Carbohydrate Polymers* 2016, **147**:444-454.
149. Stewart ML, Timm DA, Slavin JL: **Fructooligosaccharides exhibit more rapid fermentation than long-chain inulin in an in vitro fermentation system.** *Nutr Res* 2008, **28**(5):329-334.
150. Van Loo J: **The specificity of the interaction with intestinal bacterial fermentation by prebiotics determines their physiological efficacy.** *Nutrition Research Reviews* 2004, **17**(1):89-98.

151. Gibson GR, Probert HM, Van Loo J, Rastall RA, Roberfroid MB: **Dietary modulation of the human colonic microbiota: Updating the concept of prebiotics.** *Nutrition Research Reviews* 2004, **17**(2):259-275.
152. Franken J, Brandt BA, Tai SL, Bauer FF: **Biosynthesis of Levan, a Bacterial Extracellular Polysaccharide, in the Yeast *Saccharomyces cerevisiae*.** *PLoS ONE* 2013, **8**(10).
153. Lammens W, Le Roy K, Schroeven L, Van Laere A, Rabijns A, Van Den Ende W: **Structural insights into glycoside hydrolase family 32 and 68 enzymes: Functional implications.** *Journal of Experimental Botany* 2009, **60**(3):727-740.
154. Belghith KS, Dahech I, Belghith H, Mejdoub H: **Microbial production of levansucrase for synthesis of fructooligosaccharides and levan.** *International Journal of Biological Macromolecules* 2012, **50**(2):451-458.
155. Li W, Yu S, Zhang T, Jiang B, Mu W: **Recent novel applications of levansucrases.** *Applied Microbiology and Biotechnology* 2015, **99**(17):6959-6969.
156. Ortiz-Soto ME, Seibel J: **Biotechnological synthesis and transformation of valuable sugars in the food and pharmaceutical industry.** *Curr Org Chem* 2014, **18**(8):964-986.
157. Bali V, Panesar PS, Bera MB, Panesar R: **Fructo-oligosaccharides: Production, Purification and Potential Applications.** *Crit Rev Food Sci Nutr* 2015, **55**(11):1475-1490.
158. Nguyen TH, Haltrich D: **Microbial production of prebiotic oligosaccharides.** In: *Microbial Production of Food Ingredients, Enzymes and Nutraceuticals.* Elsevier Ltd.; 2013: 494-530.
159. Dahech I, Belghith KS, Belghith H, Mejdoub H: **Partial purification of a *Bacillus licheniformis* levansucrase producing levan with antitumor activity.** *International Journal of Biological Macromolecules* 2012, **51**(3):329-335.
160. Srikanth R, Reddy CHSSS, Siddartha G, Ramaiah MJ, Uppuluri KB: **Review on production, characterization and applications of microbial levan.** *Carbohydrate Polymers* 2015, **120**:102-114.
161. Esawy MA, Ahmed EF, Helmy WA, Mansour NM, El-Senousy WM, El-Safty MM: **Production of levansucrase from novel honey *Bacillus subtilis* isolates capable of producing antiviral levans.** *Carbohydrate Polymers* 2011, **86**(2):823-830.
162. Abdel-Fattah AM, Gamal-Eldeen AM, Helmy WA, Esawy MA: **Antitumor and antioxidant activities of levan and its derivative from the isolate *Bacillus subtilis* NRC1aza.** *Carbohydrate Polymers* 2012, **89**(2):314-322.
163. Esawy MA, Abdel-Fattah AM, Ali MM, Helmy WA, Salama BM, Taie HAA, Hashem AM, Awad GEA: **Levansucrase optimization using solid state fermentation and levan biological activities studies.** *Carbohydrate Polymers* 2013, **96**(1):332-341.
164. Xu X, Gao C, Liu Z, Wu J, Han J, Yan M, Wu Z: **Characterization of the levan produced by *Paenibacillus bovis* sp. nov. BD3526 and its immunological activity.** *Carbohydrate Polymers* 2016, **144**:178-186.
165. Küçükaşık F, Kazak H, Güney D, Finore I, Poli A, Yenigün O, Nicolaus B, Öner ET: **Molasses as fermentation substrate for levan production by *Halomonas* sp.** *Applied Microbiology and Biotechnology* 2011, **89**(6):1729-1740.
166. Zhang T, Chi Z, Zhao CH, Chi ZM, Gong F: **Bioethanol production from hydrolysates of inulin and the tuber meal of Jerusalem artichoke by *Saccharomyces* sp. W0.** *Bioresource Technology* 2010, **101**(21):8166-8170.
167. Kim MJ, Park HE, Sung HK, Park TH, Cha J: **Action mechanism of transfructosylation catalyzed by *Microbacterium laevaniformans* levansucrase.** *Journal of Microbiology and Biotechnology* 2005, **15**(1):99-104.
168. Lu J, Lu L, Xiao M: **[Application of levansucrase in levan synthesis--a review].** *Wei Sheng Wu Xue Bao* 2014, **54**(6):601-607.
169. Porras-Domínguez JR, Ávila-Fernández Á, Rodríguez-Alegría ME, Miranda-Molina A, Escalante A, González-Cervantes R, Olvera C, López Munguía A: **Levan-type FOS production using a *Bacillus licheniformis* endolevanase.** *Process Biochemistry* 2014, **49**(5):783-790.

170. Jang KH, Kang SA, Cho Y, Kim YY, Lee YJ, Hong K, Seong KH, Kim SH, Kim CH, Rhree SK *et al*: **Prebiotic properties of levan in rats.** *Journal of Microbiology and Biotechnology* 2003, **13**(3):348-353.
171. Kilian S, Kritzinger S, Rycroft C, Gibson G, Du Preez J: **The effects of the novel bifidogenic trisaccharide, neokestose, on the human colonic microbiota.** *World Journal of Microbiology and Biotechnology* 2002, **18**(7):637-644.
172. Marx SP, Winkler S, Hartmeier W: **Metabolization of  $\beta$ -(2,6)-linked fructose-oligosaccharides by different bifidobacteria.** *FEMS Microbiology Letters* 2000, **182**(1):163-169.
173. Adamberg K, Tomson K, Talve T, Pudova K, Puurand M, Visnapuu T, Alamäe T, Adamberg S: **Levan enhances associated growth of Bacteroides, Escherichia, Streptococcus and Faecalibacterium in fecal microbiota.** *PLoS ONE* 2015, **10**(12).
174. Hamaker BR, Tuncil YE: **A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota.** *Journal of Molecular Biology* 2014, **426**(23):3838-3850.
175. Ramirez-Farias C, Slezak K, Fuller Z, Duncan A, Holtrop G, Louis P: **Effect of inulin on the human gut microbiota: Stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii.** *Br J Nutr* 2009, **101**(4):541-550.
176. Bujacz A, Jedrzejczak-Krzepkowska M, Bielecki S, Redzynia I, Bujacz G: **Crystal structures of the apo form of  $\beta$ -fructofuranosidase from Bifidobacterium longum and its complex with fructose.** *FEBS Journal* 2011, **278**(10):1728-1744.
177. Tian F, Karboune S, Hill A: **Synthesis of fructooligosaccharides and oligolevans by the combined use of levansucrase and endo-inulinase in one-step bi-enzymatic system.** *Innovative Food Sci Emerg Technol* 2014, **22**:230-238.
178. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL: **Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations.** *Cell* 2010, **141**(7):1241-1252.
179. Suarez FL, Savaiano DA: **Diet, genetics, and lactose intolerance.** *Food Technol* 1997, **51**(3):74-76.
180. Johnson AO, Semenya JG, Buchowski MS, Enwonwu CO, Scrimshaw NS: **Correlation of lactose maldigestion, lactose intolerance, and milk intolerance.** *American Journal of Clinical Nutrition* 1993, **57**(3):399-401.
181. Mahoney RR: **Galactosyl-oligosaccharide formation during lactose hydrolysis: A review.** *Food Chemistry* 1998, **63**(2):147-154.
182. Gänzle MG, Haase G, Jelen P: **Lactose: Crystallization, hydrolysis and value-added derivatives.** *International Dairy Journal* 2008, **18**(7):685-694.
183. Panesar PS, Kumari S, Panesar R: **Potential applications of immobilized  $\beta$ -galactosidase in food processing industries.** *Enzyme Res* 2010, **2010**.
184. Illanés A: **Whey upgrading by enzyme biocatalysis.** *Electronic Journal of Biotechnology* 2011, **14**(6).
185. Bakken AP, Hill Jr CG, Amundson CH: **Hydrolysis of lactose in skim milk by immobilized  $\beta$ -galactosidase (Bacillus circulans).** *Biotechnology and Bioengineering* 1992, **39**(4):408-417.
186. Feng Y, Chang X, Wang W, Ma R: **Stabilities of immobilized  $\beta$ -galactosidase of aspergillus sp. AF for the optimal production of galactooligosaccharides from lactose.** *Artificial Cells, Blood Substitutes, and Biotechnology* 2010, **38**(1):43-51.
187. Jacob F, Monod J: **Genetic regulatory mechanisms in the synthesis of proteins.** *Journal of molecular biology* 1961, **3**:318-356.
188. Welply JK, Fowler AV, Zabin I: **beta-Galactosidase alpha-complementation. Overlapping sequences.** *Journal of Biological Chemistry* 1981, **256**(13):6804-6810.
189. Lamsal BP: **Production, health aspects and potential food uses of dairy prebiotic galactooligosaccharides.** *Journal of the Science of Food and Agriculture* 2012, **92**(10):2020-2028.
190. Frenzel M, Zerge K, Clawin-Rädecker I, Lorenzen PC: **Comparison of the galacto-oligosaccharide forming activity of different  $\beta$ -galactosidases.** *LWT - Food Science and Technology* 2015, **60**(2):1068-1071.

191. Rodriguez-Colinas B, Fernandez-Arrojo L, Ballesteros AO, Plou FJ: **Galactooligosaccharides formation during enzymatic hydrolysis of lactose: Towards a prebiotic-enriched milk.** *Food Chemistry* 2014, **145**:388-394.
192. Gosling A, Stevens GW, Barber AR, Kentish SE, Gras SL: **Recent advances refining galactooligosaccharide production from lactose.** *Food Chemistry* 2010, **121**(2):307-318.
193. Sangwan V, Tomar SK, Singh RRB, Singh AK, Ali B: **Galactooligosaccharides: Novel Components of Designer Foods.** *Journal of Food Science* 2011, **76**(4):R103-R111.
194. Verheijden KAT, Willemsen LEM, Braber S, Leusink-Muis T, Jeurink PV, Garssen J, Kraneveld AD, Folkerts G: **The development of allergic inflammation in a murine house dust mite asthma model is suppressed by synbiotic mixtures of non-digestible oligosaccharides and Bifidobacterium breve M-16V.** *Eur J Nutr* 2016, **55**(3):1141-1151.
195. Verheijden KAT, Braber S, Leusink-Muis T, Thijssen S, Boon L, Kraneveld AD, Garssen J, Folkerts G, Willemsen LEM: **Regulatory T cell depletion abolishes the protective effect of dietary galacto-oligosaccharides on eosinophilic airway inflammation in house dust mite-induced asthma in mice.** *Journal of Nutrition* 2016, **146**(4):831-837.
196. Verheijden KAT, Akbari P, Willemsen LEM, Kraneveld AD, Folkerts G, Garssen J, Fink-Gremmels J, Braber S: **Inflammation-Induced Expression of the Alarmin Interleukin 33 Can Be Suppressed by Galacto-Oligosaccharides.** *Int Arch Allergy Immunol* 2015, **167**:127-136.
197. Hashmi A, Naeem N, Farooq Z, Masood S, Iqbal S, Naseer R: **Effect of Prebiotic Galacto-Oligosaccharides on Serum Lipid Profile of Hypercholesterolemics.** *Probiotics Antimicrob Proteins* 2016, **8**(1):19-30.
198. Bruno-Barcena JM, Azcarate-Peril MA: **Galacto-oligosaccharides and colorectal cancer: Feeding our intestinal probiome.** *J Funct Foods* 2015, **12**:92-108.
199. Vandenplas Y, Zakharova I, Dmitrieva Y: **Oligosaccharides in infant formula: More evidence to validate the role of prebiotics.** *Br J Nutr* 2015, **113**(9):1339-1344.
200. Henrissat B, Sulzenbacher G, Bourne Y: **Glycosyltransferases, glycoside hydrolases: surprise, surprise!** *Current Opinion in Structural Biology* 2008, **18**(5):527-533.
201. Nakao M, Harada M, Kodama Y, Nakayama T, Shibano Y, Amachi T: **Purification and characterization of a thermostable  $\beta$ -galactosidase with high transgalactosylation activity from *Saccharopolyspora rectivirgula*.** *Applied Microbiology and Biotechnology* 1994, **40**(5):657-663.
202. Rabiou BA, Jay AJ, Gibson GR, Rastall RA: **Synthesis and Fermentation Properties of Novel Galacto-Oligosaccharides by  $\beta$ -Galactosidases from Bifidobacterium Species.** *Applied and Environmental Microbiology* 2001, **67**(6):2526-2530.
203. Zhang H, Li W, Rui X, Sun X, Dong M: **Lactobacillus plantarum 70810 from Chinese paocai as a potential source of  $\beta$ -galactosidase for prebiotic galactooligosaccharides synthesis.** *European Food Research and Technology* 2013, **236**(5):817-826.
204. Tzortzis G, Goulas AK, Gibson GR: **Synthesis of prebiotic galactooligosaccharides using whole cells of a novel strain, Bifidobacterium bifidum NCIMB 41171.** *Applied Microbiology and Biotechnology* 2005, **68**(3):412-416.
205. Goulas TK, Goulas AK, Tzortzis G, Gibson GR: **Molecular cloning and comparative analysis of four  $\beta$ -galactosidase genes from Bifidobacterium bifidum NCIMB41171.** *Applied Microbiology and Biotechnology* 2007, **76**(6):1365-1372.
206. Goulas A, Tzortzis G, Gibson GR: **Development of a process for the production and purification of  $\alpha$ - and  $\beta$ -galactooligosaccharides from Bifidobacterium bifidum NCIMB 41171.** *International Dairy Journal* 2007, **17**(6):648-656.
207. Jung SJ, Houde R, Baurhoo B, Zhao X, Lee BH: **Effects of galacto-oligosaccharides and a Bifidobacteria lactis-based probiotic strain on the growth performance and fecal microflora of broiler chickens.** *Poultry Science* 2008, **87**(9):1694-1699.
208. Hung MN, Lee B: **Purification and characterization of a recombinant  $\beta$ -galactosidase with transgalactosylation activity from Bifidobacterium infantis HL96.** *Applied Microbiology and Biotechnology* 2002, **58**(4):439-445.

209. Hung MN, Xia Z, Hu NT, Lee BH: **Molecular and Biochemical Analysis of Two  $\beta$ -Galactosidases from *Bifidobacterium infantis* HL96.** *Applied and Environmental Microbiology* 2001, **67**(9):4256-4263.
210. Roy D, Daoudi L, Azaola A: **Optimization of galacto-oligosaccharide production by *Bifidobacterium infantis* RW-8120 using response surface methodology.** *Journal of Industrial Microbiology and Biotechnology* 2002, **29**(5):281-285.
211. Hsu CA, Lee SL, Chou CC: **Enzymatic production of galactooligosaccharides by  $\beta$ -galactosidase from *Bifidobacterium longum* BCRC 15708.** *Journal of Agricultural and Food Chemistry* 2007, **55**(6):2225-2230.
212. Akiyama K, Takase M, Horikoshi K, Okonogi S: **Production of galactooligosaccharides from lactose using a  $\beta$ -glucosidase from *Thermus* sp. Z-1.** *Bioscience, Biotechnology and Biochemistry* 2001, **65**(2):438-441.
213. Choi JJ, Oh EJ, Lee YJ, Suh DS, Lee JH, Lee SW, Shin HT, Kwon ST: **Enhanced expression of the gene for  $\beta$ -glycosidase of *Thermus caldophilus* GK24 and synthesis of galacto-oligosaccharides by the enzyme.** *Biotechnology and Applied Biochemistry* 2003, **38**(2):131-136.
214. Kim CS, Ji ES, Oh DK: **Characterization of a thermostable recombinant  $\beta$ -galactosidase from *Thermotoga maritima*.** *Journal of Applied Microbiology* 2004, **97**(5):1006-1014.
215. Mozaffar Z, Nakanishi K, Matsuno R, Kamikubo T: **Purification and properties of  $\beta$ -galactosidases from *Bacillus circulans*.** *Agricultural and Biological Chemistry* 1984, **48**(12):3053-3061.
216. Mozaffar Z, Nakanishi K, Matsuno R: **Continuous production of galacto-oligosaccharides from lactose using immobilized  $\beta$ -galactosidase from *Bacillus circulans*.** *Applied Microbiology and Biotechnology* 1986, **25**(3):224-228.
217. Mozaffar Z, Nakanishi K, Matsuno R: **Effect of glutaraldehyde on oligosaccharide production by  $\beta$ -galactosidase from *Bacillus circulans*.** *Applied Microbiology and Biotechnology* 1987, **25**(5):426-429.
218. Boon MA, Janssen AEM, Van Der Padt A: **Modelling and parameter estimation of the enzymatic synthesis of oligosaccharides by  $\beta$ -galactosidase from *Bacillus circulans*.** *Biotechnology and Bioengineering* 1999, **64**(5):558-567.
219. Cheng CC, Yu MC, Cheng TC, Sheu DC, Duan KJ, Tai WL: **Production of high-content galacto-oligosaccharide by enzyme catalysis and fermentation with *Kluyveromyces marxianus*.** *Biotechnology Letters* 2006, **28**(11):793-797.
220. Cheng TC, Duan KJ, Sheu DC: **Application of tris(hydroxymethyl)phosphine as a coupling agent for  $\beta$ -galactosidase immobilized on chitosan to produce galactooligosaccharides.** *Journal of Chemical Technology and Biotechnology* 2006, **81**(2):233-236.
221. Placier G, Watzlawick H, Rabiller C, Mattes R: **Evolved  $\beta$ -galactosidases from *Geobacillus stearothermophilus* with improved transgalactosylation yield for galacto-oligosaccharide production.** *Applied and Environmental Microbiology* 2009, **75**(19):6312-6321.
222. Stevenson DE, Stanley RA, Furneaux RH: **Oligosaccharide and alkyl  $\beta$ -galactopyranoside synthesis from lactose with *Caldocellum saccharolyticum*  $\beta$ -glycosidase.** *Enzyme and Microbial Technology* 1996, **18**(8):544-549.
223. Nguyen TH, Splechtna B, Krasteva S, Kneifel W, Kulbe KD, Divne C, Haltrich D: **Characterization and molecular cloning of a heterodimeric  $\beta$ -galactosidase from the probiotic strain *Lactobacillus acidophilus* R22.** *FEMS Microbiology Letters* 2007, **269**(1):136-144.
224. Maischberger T, Nguyen TH, Sukyai P, Kittl R, Riva S, Ludwig R, Haltrich D: **Production of lactose-free galacto-oligosaccharide mixtures: comparison of two cellobiose dehydrogenases for the selective oxidation of lactose to lactobionic acid.** *Carbohydrate Research* 2008, **343**(12):2140-2147.
225. Splechtna B, Nguyen TH, Steinböck M, Kulbe KD, Lorenz W, Haltrich D: **Production of prebiotic galacto-oligosaccharides from lactose using  $\beta$ -galactosidases from *Lactobacillus reuteri*.** *Journal of Agricultural and Food Chemistry* 2006, **54**(14):4999-5006.
226. Greenberg NA, Mahoney RR: **Formation of oligosaccharides by  $\beta$ -galactosidase from *Streptococcus thermophilus*.** *Food Chemistry* 1983, **10**(3):195-204.

227. Lu L, Xiao M, Xu X, Li Z, Li Y: **A novel  $\beta$ -galactosidase capable of glycosyl transfer from *Enterobacter agglomerans* B1.** *Biochemical and Biophysical Research Communications* 2007, **356**(1):78-84.
228. Lu LI, Xiao M, Li Zy, Li Ym, Wang Fs: **A novel transglycosylating  $\beta$ -galactosidase from *Enterobacter cloacae* B5.** *Process Biochemistry* 2009, **44**(2):232-236.
229. Huber RE, Kurz G, Wallenfels K: **A quantitation of the factors which affect the hydrolase and transgalactosylase activities of  $\beta$ -galactosidase (*E. coli*) on lactose.** *Biochemistry* 1976, **15**(9):1994-2001.
230. Chen SX, Wei DZ, Hu ZH: **Synthesis of galacto-oligosaccharides in AOT/isooctane reverse micelles by  $\beta$ -galactosidase.** *J Mol Catal B Enzym* 2001, **16**(2):109-114.
231. Beloqui A, de María PD, Golyshin PN, Ferrer M: **Recent trends in industrial microbiology.** *Current Opinion in Microbiology* 2008, **11**(3):240-248.
232. Warnecke F, Hess M: **A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts.** *Journal of Biotechnology* 2009, **142**(1):91-95.
233. Fernández-Arrojo L, Guazzaroni ME, López-Cortés N, Beloqui A, Ferrer M: **Metagenomic era for biocatalyst identification.** *Current Opinion in Biotechnology* 2010, **21**(6):725-733.
234. Uchiyama T, Miyazaki K: **Functional metagenomics for enzyme discovery: challenges to efficient screening.** *Current Opinion in Biotechnology* 2009, **20**(6):616-622.
235. Carninci P, Waki K, Shiraki T, Konno H, Shibata K, Itoh M, Aizawa K, Arakawa T, Ishii Y, Sasaki D *et al*: **Targeting a complex transcriptome: The construction of the mouse full-length cDNA encyclopedia.** *Genome Research* 2003, **13**(6 B):1273-1289.
236. Chauthaiwale VM, Therwath A, Deshpande VV: **Bacteriophage lambda as a cloning vector.** *Microbiological Reviews* 1992, **56**(4):577-591.
237. Verma D, Satyanarayana T: **An improved protocol for DNA extraction from alkaline soil and sediment samples for constructing metagenomic libraries.** *Applied Biochemistry and Biotechnology* 2011, **165**(2):454-464.
238. Dos Reis Falcao V, Pedroso Tonon A, Cabral Oliveira M, Colepicolo P: **RNA isolation method for polysaccharide rich algae: Agar producing *Gracilaria tenuistipitata* (Rhodophyta).** *Journal of Applied Phycology* 2008, **20**(1):9-12.
239. Short JM, Fernandez JM, Sorge JA, Huse WD: **Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties.** *Nucleic Acids Research* 1988, **16**(15):7583-7600.
240. Sambrook J, Russell D: **Molecular Cloning: A Laboratory Manual:** Cold Spring Harbor Laboratory Press; 2001.
241. Datsenko KA, Wanner BL: **One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products.** *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(12):6640-6645.
242. Tu Z, He G, Li KX, Chen MJ, Chang J, Chen L, Yao Q, Liu DP, Ye H, Shi J *et al*: **An improved system for competent cell preparation and high efficiency plasmid transformation using different *Escherichia coli* strains.** *Electronic Journal of Biotechnology* 2005, **8**(1):113-120.
243. Studier FW: **Stable expression clones and auto-induction for protein production in *E. Coli*.** In: *Methods in Molecular Biology*. vol. 1091; 2014: 17-32.
244. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10**:421.
245. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.** *Analytical Biochemistry* 1976, **72**(1-2):248-254.
246. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**(5259):680-685.
247. Wang SD, Guo GS, Li L, Cao LC, Tong L, Ren GH, Liu YH: **Identification and characterization of an unusual glycosyltransferase-like enzyme with  $\beta$ -galactosidase activity from a soil metagenomic library.** *Enzyme and Microbial Technology* 2014, **57**:26-35.
248. Culligan EP, Sleator RD, Marchesi JR, Hill C: **Metagenomics and novel gene discovery: Promise and potential for novel therapeutics.** *Virulence* 2014, **5**(3).

249. Simon C, Daniel R: **Achievements and new knowledge unraveled by metagenomic approaches.** *Applied Microbiology and Biotechnology* 2009, **85**(2):265-276.
250. Varma A, Padh H, Shrivastava N: **Plant genomic DNA isolation: An art or a science.** *Biotechnology Journal* 2007, **2**(3):386-392.
251. Davies G, Henrissat B: **Structures and mechanisms of glycosyl hydrolases.** *Structure* 1995, **3**(9):853-859.
252. Davies GJ, Gloster TM, Henrissat B: **Recent structural insights into the expanding world of carbohydrate-active enzymes.** *Current Opinion in Structural Biology* 2005, **15**(6):637-645.
253. Ramos CRR, Abreu PAE, Nascimento ALTO, Ho PL: **A high-copy T7 Escherichia coli expression vector for the production of recombinant proteins with a minimal N-terminal his-tagged fusion peptide.** *Brazilian Journal of Medical and Biological Research* 2004, **37**(8):1103-1109.
254. Schoepfer R: **The pRSET family of T7 promoter expression vectors for Escherichia coli.** *Gene* 1993, **124**(1):83-85.
255. Fischer C, Kleinschmidt T: **Synthesis of galactooligosaccharides using sweet and acid whey as a substrate.** *International Dairy Journal* 2015.
256. Ansari SA, Satar R: **Recombinant  $\beta$ -galactosidases - Past, present and future: A mini review.** *Journal of Molecular Catalysis B: Enzymatic* 2012, **81**:1-6.
257. Di Lauro B, Strazzulli A, Perugino G, La Cara F, Bedini E, Corsaro MM, Rossi M, Moracci M: **Isolation and characterization of a new family 42 beta-galactosidase from the thermoacidophilic bacterium Alicyclobacillus acidocaldarius: identification of the active site residues.** *Biochimica et biophysica acta* 2008, **1784**(2):292-301.
258. Wierzbicka-Woś A, Bartasun P, Cieśliński H, Kur J: **Cloning and characterization of a novel cold-active glycoside hydrolase family 1 enzyme with  $\beta$ -glucosidase,  $\beta$ -fucosidase and  $\beta$ -galactosidase activities.** *BMC Biotechnology* 2013, **13**.
259. Vester JK, Glaring MA, Stougaard P: **Discovery of novel enzymes with industrial potential from a cold and alkaline environment by a combination of functional metagenomics and culturing.** *Microbial Cell Factories* 2014, **13**(1).
260. Ghosh M, Pulicherla KK, Rekha VPB, Raja PK, Sambasiva Rao KRS: **Cold active  $\beta$ -galactosidase from Thalassospira sp. 3SC-21 to use in milk lactose hydrolysis: A novel source from deep waters of Bay-of-Bengal.** *World Journal of Microbiology and Biotechnology* 2012, **28**(9):2859-2869.
261. Chockchaisawasdee S, Athanasopoulos VI, Niranjan K, Rastall RA: **Synthesis of galactooligosaccharide from lactose using  $\beta$ -galactosidase from kluveromyces lactis: Studies on batch and continuous UF membrane-fitted bioreactors.** *Biotechnology and Bioengineering* 2005, **89**(4):434-443.
262. Byfield J, Cardenas S, Alméciga-Díaz CJ, Sánchez OF:  **$\beta$ -galactosidase and galactooligosaccharides production and applications.** *Recent Patents on Chemical Engineering* 2010, **3**(1):17-29.
263. Braga ARC, Manera AP, Ores JC, Sala L, Maugeri F, Kalil SJ: **Kinetics and thermal properties of crude and purified  $\beta$ -galactosidase with potential for the production of galactooligosaccharides.** *Food Technology and Biotechnology* 2013, **51**(1):45-52.
264. Arreola SL, Intanon M, Suljic J, Kittl R, Pham NH, Kosma P, Haltrich D, Nguyen TH: **Two  $\beta$ -galactosidases from the human isolate Bifidobacterium breve DSM 20213: Molecular cloning and expression, biochemical characterization and synthesis of galacto-oligosaccharides.** *PLoS ONE* 2014, **9**(8).
265. Hernández O, Ruiz-Matute AI, Olano A, Moreno FJ, Sanz ML: **Comparison of fractionation techniques to obtain prebiotic galactooligosaccharides.** *International Dairy Journal* 2009, **19**(9):531-536.
266. Bindels LB, Delzenne NM, Cani PD, Walter J: **Opinion: Towards a more comprehensive concept for prebiotics.** *Nature Reviews Gastroenterology and Hepatology* 2015, **12**(5):303-310.
267. Rastall RA, Gibson GR: **Recent developments in prebiotics to selectively impact beneficial microbes and promote intestinal health.** *Current Opinion in Biotechnology* 2015, **32**:42-46.

268. Singh R, Sharma PK, Malviya R: **Prebiotics: Future trends in health care.** *Mediterr J Nutr Metab* 2012, **5**(2):81-90.
269. Tan H, O'Toole PW: **Impact of diet on the human intestinal microbiota.** *Current Opinion in Food Science* 2015, **2**:71-77.
270. Reddy BL, Saier MH: **Autism and our intestinal microbiota.** *Journal of Molecular Microbiology and Biotechnology* 2015, **25**(1):51-55.
271. Rastall RA: **Galacto-Oligosaccharides as Prebiotic Food Ingredients.** In.: John Wiley and Sons; 2012: 101-109.
272. Nath A, Verasztó B, Basak S, Koris A, Kovács Z, Vatai G: **Synthesis of Lactose-Derived Nutraceuticals from Dairy Waste Whey—a Review.** *Food Bioprocess Technol* 2016, **9**(1):16-48.
273. Van Leeuwen SS, Kuipers BJH, Dijkhuizen L, Kamerling JP: **Comparative structural characterization of 7 commercial galacto-oligosaccharide (GOS) products.** *Carbohydrate Research* 2016, **425**:48-58.
274. Chung CH: **Human milk oligosaccharides and prebiotic oligosaccharides in infant formula.** *Korean Journal of Microbiology and Biotechnology* 2010, **38**(1):1-6.
275. Guerrero C, Vera C, Conejeros R, Illanes A: **Transgalactosylation and hydrolytic activities of commercial preparations of  $\beta$ -galactosidase for the synthesis of prebiotic carbohydrates.** *Enzyme and Microbial Technology* 2015, **70**:9-17.
276. Otieno DO: **Synthesis of  $\beta$ -Galactooligosaccharides from Lactose Using Microbial  $\beta$ -Galactosidases.** *Comprehensive Reviews in Food Science and Food Safety* 2010, **9**(5):471-482.
277. Zárate G, Saez G, Chaia AP: **Microbial transformation of lactose: Potential of  $\beta$ -galactosidases for probiotic and prebiotic purposes.** In.: Nova Science Publishers, Inc.; 2013: 1-50.
278. Park AR, Oh DK: **Galacto-oligosaccharide production using microbial  $\beta$ -galactosidase: Current state and perspectives.** *Applied Microbiology and Biotechnology* 2010, **85**(5):1279-1286.
279. Michlmayr H, Kneifel W:  **$\beta$ -Glucosidase activities of lactic acid bacteria: Mechanisms, impact on fermented food and human health.** *FEMS Microbiology Letters* 2014, **352**(1):1-10.
280. Hassan N, Nguyen TH, Intanon M, Kori LD, Patel BKC, Haltrich D, Divne C, Tan TC: **Biochemical and structural characterization of a thermostable  $\beta$ -glucosidase from *Halothermothrix orenii* for galacto-oligosaccharide synthesis.** *Applied Microbiology and Biotechnology* 2014.
281. Nakkharat P, Haltrich D: **Purification and characterisation of an intracellular enzyme with  $\beta$ -glucosidase and  $\beta$ -galactosidase activity from the thermophilic fungus *Talaromyces thermophilus* CBS 236.58.** *Journal of Biotechnology* 2006, **123**(3):304-313.
282. Studier FW: **Protein production by auto-induction in high density shaking cultures.** *Protein Expression and Purification* 2005, **41**(1):207-234.
283. Hill AD, Reilly PJ: **Computational analysis of glycoside hydrolase family 1 specificities.** *Biopolymers* 2008, **89**(11):1021-1031.
284. Zechel DL, Withers SG: **Dissection of nucleophilic and acid-base catalysis in glycosidases.** *Current Opinion in Chemical Biology* 2001, **5**(6):643-649.
285. Gabor EM, Alkema WBL, Janssen DB: **Quantifying the accessibility of the metagenome by random expression cloning techniques.** *Environmental Microbiology* 2004, **6**(9):879-886.
286. Matthews BW: **The structure of *E. coli*  $\beta$ -galactosidase.** *Comptes Rendus - Biologies* 2005, **328**(6 SPEC. ISS.):549-556.
287. Juers DH, Matthews BW, Huber RE: **LacZ  $\beta$ -galactosidase: Structure and function of an enzyme of historical and molecular biological importance.** *Protein Science* 2012, **21**(12):1792-1807.
288. Kim JH, Lee DH, Lee JS: **Production of galactooligosaccharide by  $\beta$ -galactosidase from *Kluyveromyces maxianus* var *lactis* OE-20.** *Biotechnology and Bioprocess Engineering* 2001, **6**(5):337-340.
289. Martínez-Villaluenga C, Cardelle-Cobas A, Corzo N, Olano A, Villamiel M: **Optimization of conditions for galactooligosaccharide synthesis during lactose hydrolysis by  $\beta$ -galactosidase from *Kluyveromyces lactis* (Lactozym 3000 L HP G).** *Food Chemistry* 2008, **107**(1):258-264.
290. Zhang ZG: **Structure analysis and homology modeling of thermostable  $\beta$ -galactosidase.** *Modern Food Science and Technology* 2013, **29**(4):706-709+748.

291. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P: **PROSITE: a documented database using patterns and profiles as motif descriptors.** *Brief Bioinform* 2002, **3**(3):265-274.
292. Sigrist CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I: **New and continuing developments at PROSITE.** *Nucleic Acids Research* 2013, **41**(D1):D344-D347.
293. Petersen TN, Brunak S, Von Heijne G, Nielsen H: **SignalP 4.0: Discriminating signal peptides from transmembrane regions.** *Nature Methods* 2011, **8**(10):785-786.
294. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: Towards a more sustainable future.** *Nucleic Acids Research* 2016, **44**(D1):D279-D285.
295. Liebl W, Gabelsberger J, Schleifer KH: **Comparative amino acid sequence analysis of *Thermotoga maritima*  $\beta$ -glucosidase (BglA) deduced from the nucleotide sequence of the gene indicates distant relationship between  $\beta$ -glucosidases of the BGA family and other families of  $\beta$ -1,4-glycosyl hydrolases.** *Molec Gen Genet* 1994, **242**(1):111-115.
296. Wutor VC, Togo CA, Pletschke BI: **The effect of physico-chemical parameters and chemical compounds on the activity of beta-d-galactosidase (B-GAL), a marker enzyme for indicator microorganisms in water.** *Chemosphere* 2007, **68**(4):622-627.
297. Gänzle MG: **Enzymatic synthesis of galacto-oligosaccharides and other lactose derivatives (hetero-oligosaccharides) from lactose.** *International Dairy Journal* 2012, **22**(2):116-122.
298. Ghatak A, Guha AK, Ray L: **Beta-D-galactosidase from *Enterobacter cloacae*: production and some physicochemical properties.** *Appl Biochem Biotechnol* 2010, **162**(6):1678-1688.
299. Kim CS, Ji ES, Oh DK: **Expression and characterization of *Kluyveromyces lactis* beta-galactosidase in *Escherichia coli*.** *Biotechnol Lett* 2003, **25**(20):1769-1774.
300. Yang S, Jiang Z, Yan Q, Zhu H: **Characterization of a thermostable extracellular  $\beta$ -glucosidase with activities of exoglucanase and transglycosylation from *Paecilomyces thermophila*.** *Journal of Agricultural and Food Chemistry* 2008, **56**(2):602-608.
301. Choi JY, Park AR, Kim YJ, Kim JJ, Cha CJ, Yoon JJ: **Purification and characterization of an extracellular beta-glucosidase produced by *Phoma* sp. KCTC11825BP isolated from rotten mandarin peel.** *J Microbiol Biotechnol* 2011, **21**(5):503-508.
302. Fang Z, Fang W, Liu J, Hong Y, Peng H, Zhang X, Sun B, Xiao Y: **Cloning and characterization of a  $\beta$ -glucosidase from marine microbial metagenome with excellent glucose tolerance.** *Journal of Microbiology and Biotechnology* 2010, **20**(9):1351-1358.
303. Wallenfels K: **Enzymatische Synthese von Oligosacchariden aus Disacchariden.** *Die Naturwissenschaften* 1951, **38**(13):306-307.
304. Park AR, Kim HJ, Lee JK, Oh DK: **Hydrolysis and transglycosylation activity of a thermostable recombinant  $\beta$ -glycosidase from *Sulfolobus acidocaldarius*.** *Applied Biochemistry and Biotechnology* 2010, **160**(8):2236-2247.
305. Nguyen TH, Splechna B, Steinböck M, Kneifel W, Lettner HP, Kulbe KD, Haltrich D: **Purification and characterization of two novel  $\beta$ -galactosidases from *Lactobacillus reuteri*.** *Journal of Agricultural and Food Chemistry* 2006, **54**(14):4989-4998.
306. Pisani FM, Rella R, Raia CA, Rozzo C, Nucci R, Gambacorta A, De Rosa M, Rossi M: **Thermostable  $\beta$ -galactosidase from the archaeobacterium *Sulfolobus solfataricus*. Purification and properties.** *European Journal of Biochemistry* 1990, **187**(2):321-328.
307. Rhimi M, Aghajari N, Jaouadi B, Juy M, Boudebouze S, Maguin E, Haser R, Bejar S: **Exploring the acidotolerance of  $\beta$ -galactosidase from *Lactobacillus delbrueckii* subsp. *bulgaricus*: an attractive enzyme for lactose bioconversion.** *Research in Microbiology* 2009, **160**(10):775-784.
308. Coulier L, Timmermans J, Richard B, Van Den Dool R, Haaksman I, Klarenbeek B, Slaghek T, Van Dongen W: **In-depth characterization of prebiotic galactooligosaccharides by a combination of analytical techniques.** *Journal of Agricultural and Food Chemistry* 2009, **57**(18):8488-8495.
309. Kailemia MJ, Ruhaak LR, Lebrilla CB, Amster IJ: **Oligosaccharide analysis by mass spectrometry: A review of recent developments.** *Analytical Chemistry* 2014, **86**(1):196-212.

310. Hernández-Hernández O, Calvillo I, Lebrón-Aguilar R, Moreno FJ, Sanz ML: **Hydrophilic interaction liquid chromatography coupled to mass spectrometry for the characterization of prebiotic galactooligosaccharides.** *Journal of Chromatography A* 2012, **1220**:57-67.
311. Bruins ME, Strubel M, Van Lieshout JFT, Janssen AEM, Boom RM: **Oligosaccharide synthesis by the hyperthermostable  $\beta$ -glucosidase from *Pyrococcus furiosus*: Kinetics and modelling.** *Enzyme and Microbial Technology* 2003, **33**(1):3-11.
312. Park HY, Kim HJ, Lee JK, Kim D, Oh DK: **Galactooligosaccharide production by a thermostable  $\beta$ -galactosidase from *Sulfolobus solfataricus*.** *World Journal of Microbiology and Biotechnology* 2008, **24**(8):1553-1558.
313. Gonçalves BCM, Baldo C, Celligoi MAPC: **Levan and Levansucrase—a Mini Review.**
314. Bekers M, Laukevics J, Upite D, Kaminska E, Vigants A, Viesturs U, Pankova L, Danilevics A: **Fructooligosaccharide and levan producing activity of *Zymomonas mobilis* extracellular levansucrase.** *Process Biochemistry* 2002, **38**(5):701-706.
315. Biedendieck R, Beine R, Gamer M, Jordan E, Buchholz K, Seibel J, Dijkhuizen L, Malten M, Jahn D: **Export, purification, and activities of affinity tagged *Lactobacillus reuteri* levansucrase produced by *Bacillus megaterium*.** *Applied Microbiology and Biotechnology* 2007, **74**(5):1062-1073.
316. Beine R, Moraru R, Nimtz M, Na'amnieh S, Pawlowski A, Buchholz K, Seibel J: **Synthesis of novel fructooligosaccharides by substrate and enzyme engineering.** *Journal of Biotechnology* 2008, **138**(1-2):33-41.
317. Semjonovs P, Shakirova L, Treimane R, Shvirksts K, Auzina L, Cleenwerck I, Zikmanis P: **Production of extracellular fructans by *Gluconobacter nephelii* P1464.** *Letters in Applied Microbiology* 2016, **62**(2):145-152.
318. Patel S, Goyal A: **Functional oligosaccharides: Production, properties and applications.** *World Journal of Microbiology and Biotechnology* 2011, **27**(5):1119-1128.
319. Grossman TH, Kawasaki ES, Punreddy SR, Osburne MS: **Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability.** *Gene* 1998, **209**(1-2):95-103.
320. Studier FW: **Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system.** *Journal of Molecular Biology* 1991, **219**(1):37-44.
321. Dubendorf JW, Studier FW: **Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor.** *Journal of Molecular Biology* 1991, **219**(1):45-59.
322. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF: **Understanding the Differences between Genome Sequences of *Escherichia coli* B Strains REL606 and BL21(DE3) and Comparison of the *E. coli* B and K-12 Genomes.** *Journal of Molecular Biology* 2009, **394**(4):653-680.
323. Xu J, Banerjee A, Pan SH, Li ZJ: **Galactose can be an inducer for production of therapeutic proteins by auto-induction using *E. coli* BL21 strains.** *Protein Expression and Purification* 2012, **83**(1):30-36.
324. Adhya S, Echols H: **Glucose effect and the galactose enzymes of *Escherichia coli*: correlation between glucose inhibition of induction and inducer transport.** *Journal of Bacteriology* 1966, **92**(3):601-608.
325. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics.** *Nucleic Acids Research* 2009, **37**(SUPPL. 1):D233-D238.
326. Jang KH, Kang SA, Kim CH, Lee JC, Kim MH, Son E, Rhee SK: **Characterization of levan hydrolysis activity of levansucrase from *Zymomonas mobilis* ATCC 10988 and *Rahnella aquatilis* ATCC 33071.** *Food Science and Biotechnology* 2007, **16**(3):482-484.
327. Velázquez-Hernández ML, Baizabal-Aguirre VM, Bravo-Patiño A, Cajero-Juárez M, Chávez-Moctezuma MP, Valdez-Alarcón JJ: **Microbial fructosyltransferases and the role of fructans.** *Journal of Applied Microbiology* 2009, **106**(6):1763-1778.
328. Meng G, Fütterer K: **Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase.** *Nature Structural Biology* 2003, **10**(11):935-941.

329. Martínez-Fleites C, Ortiz-Lombardía M, Pons T, Tarbouriech N, Taylor EJ, Arrieta JG, Hernández L, Davies GJ: **Crystal structure of levansucrase from the Gram-negative bacterium *Gluconacetobacter diazotrophicus***. *Biochemical Journal* 2005, **390**(1):19-27.
330. Van Hijum SAFT, Kralj S, Ozimek LK, Dijkhuizen L, Van Geel-Schutten IGH: **Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria**. *Microbiology and Molecular Biology Reviews* 2006, **70**(1):157-176.
331. Gassem MA, Sims KA, Frank JF: **Extracellular polysaccharide production by *Lactobacillus delbrueckii* subsp. *bulgaricus* RR in a continuous fermentor**. *LWT - Food Science and Technology* 1997, **30**(3):273-278.
332. Bekers M, Upite D, Kaminska E, Laukevics J, Ionina R, Vigants A: **Catalytic activity of *Zymomonas mobilis* extracellular "levan-levansucrase" complex in sucrose medium**. *Communications in agricultural and applied biological sciences* 2003, **68**(2 Pt A):321-324.
333. Le Gorrec K, Connes C, Guibert A, Uribelarrea JL, Combes D: **Identification of three inducible and extracellular enzymatic activities working on sucrose in *Bacillus subtilis* NCIMB 11871 and 11872 supernatant**. *Enzyme and Microbial Technology* 2002, **31**(1-2):44-52.
334. Tian F, Inthanavong L, Karboune S: **Purification and characterization of levansucrases from *Bacillus amyloliquefaciens* in intra- and extracellular forms useful for the synthesis of levan and fructooligosaccharides**. *Bioscience, Biotechnology and Biochemistry* 2011, **75**(10):1929-1938.
335. Tajima K, Tanio T, Kobayashi Y, Kohno H, Fujiwara M, Shiba T, Erata T, Munekata M, Takai M: **Cloning and sequencing of the levansucrase gene from *Acetobacter xylinum* NCI 1005**. *DNA Research* 2000, **7**(4):237-242.
336. Martínez-Fleites C, Tarbouriech N, Ortiz-Lombardía M, Taylor E, Rodríguez A, Ramírez R, Hernández L, Davies GJ: **Crystallization and preliminary X-ray diffraction analysis of levansucrase (LsdA) from *Gluconacetobacter diazotrophicus* SRT4**. *Acta crystallographica Section D, Biological crystallography* 2004, **60**(Pt 1):181-183.
337. Hettwer U, Gross M, Rudolph K: **Purification and characterization of an extracellular levansucrase from *Pseudomonas syringae* pv. *phaseolicola***. *J Bacteriol* 1995, **177**(10):2834-2839.
338. Sangiliyandi G, Gunasekaran P: **Polymerase and hydrolase activities of *Zymomonas mobilis* levansucrase separately modulated by in vitro mutagenesis and elevated temperature**. *Process Biochemistry* 2001, **36**(6):543-548.
339. El-Refai HA, Abdel- Fattah AF, Mostafa FA: **Immobilization and properties of *Bacillus circulans* levansucrase**. *Acta Pharmaceutica Scientia* 2009, **51**(2):149-156.
340. Yanase H, Iwata M, Nakahigashi R, Kita K, Kato N, Tonomura K, Iwata M, Nakahigashi R, Tonomura K: **Purification, crystallization, and properties of the extracellular levansucrase from *Zymomonas mobilis***. *Bioscience, Biotechnology and Biochemistry* 1992, **56**(8):1335-1337.
341. Szwengiel A. CM, Czarnecki Z.: **Levan synthesis during associated action of levansucrase and *Candida cacaoi* DSM 2226 yeast**. *Polish Journal of Food and Nutrition Sciences* 2007, **4**(57):433-440.
342. Ben Ammar Y, Matsubara T, Ito K, Iizuka M, Limpaseni T, Pongsawasdi P, Minamiura N: **Characterization of a thermostable levansucrase from *Bacillus* sp. TH4-2 capable of producing high molecular weight levan at high temperature**. *Journal of Biotechnology* 2002, **99**(2):111-119.
343. Hernandez L, Arrieta J, Menendez C, Vazquez R, Coego A, Suarez V, Selman G, Petit-Glatron MF, Chambert R: **Isolation and enzymic properties of levansucrase secreted by *Acetobacter diazotrophicus* SRT4, a bacterium associated with sugar cane**. *Biochemical Journal* 1995, **309**(1):113-118.
344. Ozimek LK, Euverink GJW, Van Der Maarel MJEC, Dijkhuizen L: **Mutational analysis of the role of calcium ions in the *Lactobacillus reuteri* strain 121 fructosyltransferase (levansucrase and inulosucrase) enzymes**. *FEBS Letters* 2005, **579**(5):1124-1128.
345. Chambert R, Petit-Glatron M-F: **Reversible thermal unfolding of *Bacillus subtilis* levansucrase is modulated by Fe<sup>3+</sup> and Ca<sup>2+</sup>**. *FEBS Letters* 1990, **275**(1-2):61-64.

346. Homann A, Biedendieck R, Götze S, Jahn D, Seibel J: **Insights into polymer versus oligosaccharide synthesis: Mutagenesis and mechanistic studies of a novel levansucrase from *Bacillus megaterium***. *Biochemical Journal* 2007, **407**(2):189-198.
347. Rairakhwada D, Seo JW, Seo MY, Kwon O, Rhee SK, Kim CH: **Gene cloning, characterization, and heterologous expression of levansucrase from *Bacillus amyloliquefaciens***. *Journal of Industrial Microbiology and Biotechnology* 2010, **37**(2):195-204.
348. Hernández L, Sotolongo M, Rosabal Y, Menéndez C, Ramírez R, Caballero-Mellado J, Arrieta J: **Structural levansucrase gene (*IsdA*) constitutes a functional locus conserved in the species *Gluconacetobacter diazotrophicus***. *Archives of Microbiology* 2000, **174**(1-2):120-124.
349. Waldherr FW, Meissner D, Vogel RF: **Genetic and functional characterization of *Lactobacillus panis* levansucrase**. *Arch Microbiol* 2008, **190**(4):497-505.
350. Kang HK, Seo MY, Seo ES, Kim D, Chung SY, Kimura A, Day DF, Robyt JF: **Cloning and expression of levansucrase from *Leuconostoc mesenteroides* B-512 FMC in *Escherichia coli***. *Biochimica et Biophysica Acta - Gene Structure and Expression* 2005, **1727**(1):5-15.
351. Cote GL, Imam SH: **Purification and properties of an extracellular levansucrase from *Erwinia herbicola* NRRL B-1678**. *Carbohydrate Research* 1989, **190**(2):299-307.
352. Ohtsuka K, Hino S, Fukushima T, Ozawa O, Kanematsu T, Uchida T: **Characterization of Levansucrase from *Rahnella aquatilis* JCM-1683**. *Bioscience, Biotechnology, and Biochemistry* 1992, **56**(9):1373-1377.
353. Goldman D, Lavid N, Schwartz A, Shoham G, Danino D, Shoham Y: **Two active forms of *Zymomonas mobilis* levansucrase. An ordered microfibril structure of the enzyme promotes levan polymerization**. *The Journal of biological chemistry* 2008, **283**(47):32209-32217.
354. Yanase H, Maeda M, Hagiwara E, Yagi H, Taniguchi K, Okamoto K: **Identification of functionally important amino acid residues in *Zymomonas mobilis* levansucrase**. *Journal of Biochemistry* 2002, **132**(4):565-572.
355. Shih IL, Yu YT: **Simultaneous and selective production of levan and poly( $\gamma$ -glutamic acid) by *Bacillus subtilis***. *Biotechnology Letters* 2005, **27**(2):103-106.
356. Sims IM, Frese SA, Walter J, Loach D, Wilson M, Appleyard K, Eason J, Livingston M, Baird M, Cook G *et al*: **Structure and functions of exopolysaccharide produced by gut commensal *Lactobacillus reuteri* 100-23**. *ISME Journal* 2011, **5**(7):1115-1124.
357. Jathore NR, Bule MV, Tilay AV, Annapure US: **Microbial levan from *Pseudomonas fluorescens*: Characterization and medium optimization for enhanced production**. *Food Science and Biotechnology* 2012, **21**(4):1045-1053.
358. Han YW, Clarke MA: **Production and characterization of microbial levan**. *Journal of Agricultural and Food Chemistry* 1990, **38**(2):393-396.
359. Marshall K, Weigel H: **Relative molecular masses and structures of some levans elaborated by strains of *Streptococcus salivarius***. *Carbohydrate Research* 1980, **80**(2):375-377.
360. Vega R, Zuniga-Hansen ME: **A new mechanism and kinetic model for the enzymatic synthesis of short-chain fructooligosaccharides from sucrose**. *Biochemical Engineering Journal* 2014, **82**:158-165.
361. Tian F, Khodadadi M, Karboune S: **Optimization of levansucrase/endo-inulinase bi-enzymatic system for the production of fructooligosaccharides and oligolevans from sucrose**. *Journal of Molecular Catalysis B: Enzymatic* 2014, **109**:85-93.
362. Escobar-Zepeda A, De León AV, Sanchez-Flores A: **The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics**. *Front Genet* 2015, **6**(DEC).
363. Woese CR: **Bacterial evolution**. *Microbiological Reviews* 1987, **51**(2):221-271.
364. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors**. *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(12):5463-5467.
365. Bartlett JM, Stirling D: **A short history of the polymerase chain reaction**. *Methods in molecular biology (Clifton, NJ)* 2003, **226**:3-6.
366. Torsvik V, Goksoyr J, Daae FL: **High diversity in DNA of soil bacteria**. *Applied and Environmental Microbiology* 1990, **56**(3):782-787.

367. Amann RI, Ludwig W, Schleifer KH: **Phylogenetic identification and in situ detection of individual microbial cells without cultivation.** *Microbiological Reviews* 1995, **59**(1):143-169.
368. Daniel R: **The metagenomics of soil.** *Nature Reviews Microbiology* 2005, **3**(6):470-478.
369. Torsvik V, Øvreås L, Thingstad TF: **Prokaryotic diversity - Magnitude, dynamics, and controlling factors.** *Science* 2002, **296**(5570):1064-1066.
370. Giovannoni SJ, DeLong EF, Schmidt TM, Pace NR: **Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton.** *Applied and Environmental Microbiology* 1990, **56**(8):2572-2575.
371. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM: **Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products.** *Chemistry and Biology* 1998, **5**(10):R245-R249.
372. Wang GYS, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ, Davies J: **Novel natural products from soil DNA libraries in a streptomycete host.** *Org Lett* 2000, **2**(16):2401-2404.
373. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J: **Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA.** *Applied and Environmental Microbiology* 2002, **68**(9):4301-4306.
374. Kowalchuk GA, Speksnijder AGCL, Zhang K, Goodman RM, Van Veen JA: **Finding the needles in the metagenome haystack.** *Microb Ecol* 2007, **53**(3):475-485.
375. Delavat F, Phalip V, Forster A, Plewniak F, Lett MC, Liévreumont D: **Amylases without known homologues discovered in an acid mine drainage: Significance and impact.** *Scientific Reports* 2012, **2**.
376. Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala HA, Rowland SE, Blanch HW, Clark DS, Robb FT: **Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment.** *Nature Communications* 2011, **2**(1).
377. Taupp M, Mewis K, Hallam SJ: **The art and design of functional metagenomic screens.** *Current Opinion in Biotechnology* 2011, **22**(3):465-472.
378. Ferrer M, Ghazi A, Beloqui A, Vieites JM, López-Cortés N, Marín-Navarro J, Nechitaylo TY, Guazzaroni ME, Polaina J, Waliczek A *et al*: **Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen.** *PLoS ONE* 2012, **7**(6).
379. Nimchua T, Thongaram T, Uengwetwanit T, Pongpattanakitshote S, Eurwilaichitr L: **Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes.** *Journal of Microbiology and Biotechnology* 2012, **22**(4):462-469.
380. Wang Y, Hayatsu M, Fujii T: **Extraction of bacterial RNA from soil: Challenges and solutions.** *Microbes Environ* 2012, **27**(2):111-121.
381. Ogram A, Sayler GS, Barkay T: **The extraction and purification of microbial DNA from sediments.** *Journal of Microbiological Methods* 1987, **7**(2-3):57-66.
382. Berry AE, Chiocchini C, Selby T, Sosio M, Wellington EMH: **Isolation of high molecular weight DNA from soil for cloning into BAC vectors.** *FEMS Microbiology Letters* 2003, **223**(1):15-20.
383. Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS *et al*: **Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products.** *Applied and Environmental Microbiology* 2003, **69**(1):49-55.
384. Courtois S, Frostegård Å, Göransson P, Depret G, Jeannin P, Simonet P: **Quantification of bacterial subgroups in soil: Comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation.** *Environmental Microbiology* 2001, **3**(7):431-439.
385. Kim UJ, Shizuya H, De Jong PJ, Birren B, Simon MI: **Stable propagation of cosmid sized human DNA inserts in an F factor based vector.** *Nucleic Acids Research* 1992, **20**(5):1083-1085.
386. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M: **Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-**

- factor-based vector.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(18):8794-8797.
387. Kakirde KS, Parsley LC, Liles MR: **Size does matter: Application-driven approaches for soil metagenomics.** *Soil Biology and Biochemistry* 2010, **42**(11):1911-1923.
388. Wild J, Hradecna Z, Szygbalski W: **Conditionally amplifiable BACs: Switching from single-copy to high-copy vectors and genomic clones.** *Genome Research* 2002, **12**(9):1434-1444.
389. Krüger A, Elleuche S, Sahm K, Antranikian G: **Robust biocatalysts - routes to new diversity.** In: *Applied Biocatalysis: From Fundamental Science to Industrial Applications.* Wiley-VCH Verlag; 2016: 31-51.
390. Liebl W: **Metagenomics.** *Encyclopedia of geobiology* 2011:553-558.
391. Steele HL, Jaeger KE, Daniel R, Streit WR: **Advances in recovery of novel biocatalysts from metagenomes.** *Journal of Molecular Microbiology and Biotechnology* 2008, **16**(1-2):25-37.
392. Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD, Charles TC: **Functional metagenomics reveals novel  $\beta$ -galactosidases not predictable from gene sequences.** *PLoS ONE* 2017, **12**(3).
393. Namiki T, Hachiya T, Tanaka H, Sakakibara Y: **MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads.** *Nucleic Acids Research* 2012, **40**(20).
394. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Research* 2000, **28**(1):45-48.
395. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al*: **The Pfam protein families database.** *Nucleic Acids Research* 2012, **40**(D1):D290-D301.
396. Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White O: **TIGRFAMs and Genome Properties: Tools for the assignment of molecular function and biological process in prokaryotic genomes.** *Nucleic Acids Research* 2007, **35**(SUPPL. 1):D260-D264.
397. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S *et al*: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Research* 2012, **40**(D1):D13-D25.
398. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S *et al*: **InterPro in 2011: New developments in the family and domain prediction database.** *Nucleic Acids Research* 2012, **40**(D1):D306-D312.
399. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: New developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Research* 2001, **29**(1):22-28.
400. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: Tool for the unification of biology.** *Nature Genetics* 2000, **25**(1):25-29.
401. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T *et al*: **KEGG for linking genomes to life and the environment.** *Nucleic Acids Research* 2008, **36**(SUPPL. 1):D480-D484.
402. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Söhngen C, Stelzer M, Thiele J, Schomburg D: **BRENDA, the enzyme information system in 2011.** *Nucleic Acids Research* 2011, **39**(SUPPL. 1):D670-D676.
403. Cummings JH, Macfarlane GT, Macfarlane S: **Intestinal bacteria and ulcerative colitis.** *Current Issues in Intestinal Microbiology* 2003, **4**(1):9-20.
404. Bersaneti GT, Pan NC, Baldo C, Celligoi MAPC: **Co-production of Fructooligosaccharides and Levan by Levansucrase from *Bacillus subtilis* natto with Potential Application in the Food Industry.** *Applied Biochemistry and Biotechnology* 2017:1-14.
405. Jelen P: **Industrial Whey Processing Technology : An Overview.** *Journal of Agricultural and Food Chemistry* 1979, **27**(4):658-661.

406. Wilson B, Whelan K: **Prebiotic inulin-type fructans and galacto-oligosaccharides: definition, specificity, function, and application in gastrointestinal disorders.** *J Gastroenterol Hepatol* 2017, **32**:64-68.
407. Vulevic J, Rastall RA, Gibson GR: **Developing a quantitative approach for determining the in vitro prebiotic potential of dietary oligosaccharides.** *FEMS Microbiology Letters* 2004, **236**(1):153-159.
408. Harmsen HJM, Wildeboer-Veloo ACM, Raangs GC, Wagendorp AA, Klijn N, Bindels JG, Welling GW: **Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods.** *Journal of Pediatric Gastroenterology and Nutrition* 2000, **30**(1):61-67.
409. Macfarlane S, Macfarlane GT, Cummings JH: **Review article: Prebiotics in the gastrointestinal tract.** *Alimentary Pharmacology and Therapeutics* 2006, **24**(5):701-714.
410. Newburg DS: **Oligosaccharides in human milk and bacterial colonization.** *Journal of Pediatric Gastroenterology and Nutrition* 2000, **30**(SUPPL. 2):S8-S17.
411. Boehm G, Lidestri M, Casetta P, Jelinek J, Negretti F, Stahl B, Marini A: **Supplementation of a bovine milk formula with an oligosaccharide mixture increases counts of faecal bifidobacteria in preterm infants.** *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2002, **86**(3):F178-F181.
412. Moro GE, Stahl B, Fanaro S, Jelinek J, Boehm G, Coppa GV: **Dietary prebiotic oligosaccharides are detectable in the faeces of formula-fed infants.** *Acta Paediatrica, International Journal of Paediatrics, Supplement* 2005, **94**(449):27-30.
413. Bode L: **Human milk oligosaccharides: Every baby needs a sugar mama.** *Glycobiology* 2012, **22**(9):1147-1162.
414. Kunz C, Rudloff S, Baier W, Klein N, Strobel S: **Oligosaccharides in human milk: Structural, functional, and metabolic aspects.** In: *Annual Review of Nutrition*. vol. 20; 2000: 699-722.
415. Tanaka R, Takayama H, Morotomi M, Kuroshima T, Ueyama S, Matsumoto K, Kuroda A, Mutai M: **Effects of administration of TOS and Bifidobacterium breve 4006 on the human fecal flora.** *Bifidobacteria Microflora* 1983, **2**(1):17-24.
416. Bartosch S, Woodmansey EJ, Paterson JCM, McMurdo MET, Macfarlane GT: **Microbiological effects of consuming a synbiotic containing Bifidobacterium bifidum, Bifidobacterium lactis, and oligofructose in elderly persons, determined by real-time polymerase chain reaction and counting of viable bacteria.** *Clin Infect Dis* 2005, **40**(1):28-37.
417. Ito M, Deguchi Y, Matsumoto K, Kimura M, Onodera N, Yajima T: **Influence of galactooligosaccharides on the human fecal microflora.** *Journal of Nutritional Science and Vitaminology* 1993, **39**(6):635-640.
418. Ito M, Deguchi Y, Miyamori A, Matsumoto K, Kikuchi H, Matsumoto K, Kobayashi Y, Yajima T, Kan T: **Effects of administration of galactooligosaccharides on the human faecal microflora, stool weight and abdominal sensation.** *Microbial Ecology in Health and Disease* 1990, **3**(6):285-292.
419. Gibson GR, Roberfroid MB: **Dietary modulation of the human colonic microbiota: Introducing the concept of prebiotics.** *Journal of Nutrition* 1995, **125**(6):1401-1412.
420. Hosono A, Kitazawa H, Yamaguchi T: **Antimutagenic and antitumour activities of lactic acid bacteria.** *Probiotics 2* 1997:89-132.
421. Tzortzis G, Vulevic J: **Galacto-oligosaccharide prebiotics.** *Prebiotics and Probiotics Science and Technology* 2009:207-244.
422. Tzortzis G, Goulas AK, Gee JM, Gibson GR: **A novel galactooligosaccharide mixture increases the bifidobacterial population numbers in a continuous in vitro fermentation system and in the proximal colonic contents of pigs in vivo.** *Journal of Nutrition* 2005, **135**(7):1726-1731.
423. Depeint F, Tzortzis G, Vulevic J, l'Anson K, Gibson GR: **Prebiotic evaluation of a novel galactooligosaccharide mixture produced by the enzymatic activity of Bifidobacterium bifidum NCIMB 41171, in healthy humans: A randomized, double-blind, crossover, placebo-controlled intervention study.** *American Journal of Clinical Nutrition* 2008, **87**(3):785-791.

424. Hamdy AA, Elattal NA, Amin MA, Ali AE, Mansour NM, Awad GEA, Awad HM, Esawy MA: **Possible correlation between levansucrase production and probiotic activity of *Bacillus* sp. isolated from honey and honey bee.** *World Journal of Microbiology and Biotechnology* 2017, **33**(4).
425. Chen XY, Gänzle MG: **Lactose and lactose-derived oligosaccharides: More than prebiotics?** *International Dairy Journal* 2017, **67**:61-72.
426. Hassan N, Geiger B, Gandini R, Patel BKC, Kittl R, Haltrich D, Nguyen TH, Divne C, Tan TC: **Engineering a thermostable *Halothermothrix orenii*  $\beta$ -glucosidase for improved galactooligosaccharide synthesis.** *Applied Microbiology and Biotechnology* 2016, **100**(8):3533-3543.
427. Yang J, Gao R, Zhou Y, Anankanbil S, Li J, Xie G, Guo Z:  **$\beta$ -Glucosidase from *Thermotoga naphthophila* RKU-10 for exclusive synthesis of galactotrisaccharides: Kinetics and thermodynamics insight into reaction mechanism.** *Food Chemistry* 2018, **240**:422-429.
428. Yang J, Wang Q, Zhou Y, Li J, Gao R, Guo Z: **Engineering *T. naphthophila* B-glucosidase for enhanced synthesis of galactooligosaccharides by site-directed mutagenesis.** *Biochemical Engineering Journal* 2017, **127**:1-8.
429. Mlichová Z, Rosenberg M: **Current trends of  $\beta$ -galactosidase application in food technology.** *Journal of Food and Nutrition Research* 2006, **45**(2):47-54.
430. Khatami S, Zokaee Ashtiani F, Bonakdarpour B, Mehrdad M: **The enzymatic production of lactulose via transglycosylation in conventional and non-conventional media.** *International Dairy Journal* 2014, **34**(1):74-79.
431. Shen Q, Yang R, Hua X, Ye F, Wang H, Zhao W, Wang K: **Enzymatic synthesis and identification of oligosaccharides obtained by transgalactosylation of lactose in the presence of fructose using  $\beta$ -galactosidase from *Kluyveromyces lactis*.** *Food Chemistry* 2012, **135**(3):1547-1554.
432. Guerrero C, Vera C, Plou F, Illanes A: **Influence of reaction conditions on the selectivity of the synthesis of lactulose with microbial  $\beta$ -galactosidases.** *Journal of Molecular Catalysis B: Enzymatic* 2011, **72**(3-4):206-212.
433. Rastall RA: **Functional oligosaccharides: Application and manufacture.** *Annual Review of Food Science and Technology* 2010, **1**(1):305-339.
434. Schüle S, Schulz-Fademrecht T, Garidel P, Bechtold-Peters K, Frieß W: **Stabilization of IgG1 in spray-dried powders for inhalation.** *Eur J Pharm Biopharm* 2008, **69**(3):793-807.
435. Klewicki R: **Transglycosylation of a  $\beta$ -galactosyl radical, in the course of enzymic hydrolysis of lactose, in the presence of selected polyhydroxyalcohols.** *Biotechnology Letters* 2000, **22**(13):1063-1066.
436. Klewicki R: **Effect of selected parameters of lactose hydrolysis in the presence of  $\beta$ -galactosidase from various sources on the synthesis of galactosyl-polyol derivatives.** *Engineering in Life Sciences* 2007, **7**(3):268-274.
437. Kuusisto J, Tokarev AV, Murzina EV, Roslund MU, Mikkola JP, Murzin DY, Salmi T: **From renewable raw materials to high value-added fine chemicals-Catalytic hydrogenation and oxidation of d-lactose.** *Catal Today* 2007, **121**(1-2):92-99.
438. Playne MJ, Crittenden RG: **Galacto-oligosaccharides and other products derived from lactose.** *Advanced Dairy Chemistry* 2009, **3**:121-201.