



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Functional screening of human faecal microbiome metagenome library for glycoside hydrolase enzymes
Author(s)	Mulualem, Daniel
Publication Date	2020-04-24
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/15909

Downloaded 2024-04-26T22:33:20Z

Some rights reserved. For more information, please see the item record link above.





NUI Galway
OÉ Gaillimh

Functional screening of human faecal microbiome metagenome library for glycoside hydrolase enzymes

Daniel Mehabie Muluaem, MSc.

Thesis submitted for the degree of Doctor of Philosophy (Microbiology) at
the National University of Ireland, Galway, Ireland

Research Supervisors:

Dr. Aoife Boyd

Prof. Conor O'Byrne

Dr. Michelle Kilcoyne

Discipline of Microbiology, College of Science, National University of
Ireland, Galway

April 2020

Table of Contents

Table of Contents.....	i-iv
Acknowledgments.....	v-vi
Abstract.....	vii
List of Figures.....	viii-xi
List of Tables.....	xii
List of Abbreviations.....	xiii-xiv
Chapter 1	1
Introduction	1
1.1. Anatomy of the human gastrointestinal tract.....	2
1.2. The mucus lining in GIT.....	4
1.2.1. Mucin O-glycosylation.....	9
1.3. Microbial adhesion to the GIT.....	11
1.4. Colonization of the GIT by microbiota.....	12
1.5. Diversity and composition of adult microbiota.....	15
1.5.1. Diet and the microbiota.....	17
1.6. Glycan foraging by the gut microbiota.....	19
1.7. Dysbiosis and its implications in disease.....	23
1.8. Probiotics, prebiotics and synbiotics.....	28
1.9. Metagenomics.....	33
1.9.1. Sequence based screening.....	34
1.9.2. Function based screening of metagenome libraries.....	35
1.10. Research goals.....	36
Chapter 2	39
Materials and Methods	39
2.1. Bacterial strains, plasmids and fosmids.....	40
2.2. Culture media and chemical supplements.....	41
2.3. Transfer of the human faecal metagenome library to a new <i>E. coli</i> EPI 300-T1 ^R host	41
2.3.1. Preparation of fosmids from the stock human faecal metagenome.....	42
library.....	42
2.3.2. Preparation of the heterologous host EPI 300-T1 ^R	42

2.3.3. Packaging the copy control fosmid DNA into MaxPlax Lambda.....	43
phages	43
2.3.4. Plating and selecting the packaged copy control fosmid library.....	43
2.4. Function based screening of the metagenome library.....	44
2.4.1. Functional Screening of metagenome library for β galactosidases.....	44
2.4.2. Functional Screening of metagenome library for α -fucosidases.....	45
2.4.3. Functional Screening of metagenome library for sialidases	46
2.5. Genetic techniques	46
2.5.1. Agarose gel electrophoresis	46
2.5.2. Purification of fosmids and plasmids.....	47
2.5.3. Restriction analysis of fosmids	47
2.5.4. Sequencing of fosmid inserts and bioinformatic analysis.....	48
2.5.5. Polymerase Chain Reaction	48
2.5.6. DNA purification using wizard SV Gel/ PCR clean up kit (Promega).....	50
2.5.7. Cloning of identified β -galactosidase gene (BAD_1582) of	51
<i>B. adolescentis</i>	51
2.5.8. Transformation of chemically competent, <i>E. coli</i> T7 express cells	51
2.5.9. Identification and Verification of cloned BAD_1582 into pET101.....	51
2.5.11. Preparation of chemically competent <i>E. coli</i> T7express (pDMg1a).....	52
2.5.12. Preparation and Transformation of Electrocompetent <i>E. coli</i> BL21.....	53
(DE3)	53
2.6. Optimisation of expression of His-tagged BAD_1582	54
2.7. Expression and preparation of cell lysates.....	54
2.8. Ni ²⁺ -nitrilotriacetate (Ni ²⁺ -NTA) affinity chromatography purification.....	55
2.9. Ultrafiltration with100K (MWCO) Amicon® Ultra-0.5 and 4 centrifugal filters	56
2.10. Protein analysis	56
2.10.1. Bradford assay	56
2.10.2. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis	57
(SDS PAGE).....	57
2.10.3. Staining SDS-PAGE gels.....	57
2.10.4. Immunoblotting	58
2.11. The β -Galactosidase (BAD_1582) enzyme characterisation	58
2.11.1. β -galactosidase activity assay	58

2.11.2. Effects of pH, temperature, divalent cations and denaturants /	59
detergents on the activity and /stability of BAD_1582	59
2.11.3. Substrate specificity and kinetic parameter	60
2.11.5. Degalactosylation activity of BAD_1582 on glycoproteins	61
2.11.6. Transgalactosylation activity assay	61
2.12. Immobilization of BAD_1582	62
2.12.1. The β -galactosidase activity and reuse of immobilised BAD_1582 and	63
<i>Aspergillus oryzae</i> β -galactosidase	63
2.12.2. Effect of pH, temperature and kinetic parameters of the	64
immobilised BAD_1582	64
2.13. Statistical analysis	64
Chapter 3	65
Validation of the human faecal microbiome metagenome library and functional screening for GH enzymes	65
3.1. Introduction	66
3.2. Validation of the human faecal microbiome metagenome library using	69
restriction pattern analyses	69
3.3. Validation of the human faecal microbiome metagenome library through end sequencing	72
3.4. Strategy for screening glycoside hydrolase enzymes from human faecal	74
metagenomic library	74
3.5. Screening of human faecal metagenome libraries for β -galactosidases	76
3.6. Screening of human faecal metagenome library for sialidases	79
3.7. Screening of human faecal metagenome libraries for α -fucosidases	83
3.8. Growth dependent assays for screening human faecal metagenome library	90
3.9. Discussion	93
Chapter 4	100
Biochemical characterization of the <i>Bifidobacterium adolescentis</i> β-galactosidase, BAD_1582	100
4.1. Introduction	101
4.2. Bioinformatic analysis to identify β -galactosidase gene (BAD_1582) from	104
4.3. Optimisation of expression of His-tagged BAD_1582	109

4.4. Expression and purification of the His-tagged β -galactosidase BAD_1582.....	114
4.5. Biochemical properties of BAD_1582.....	115
4.6. Substrate specificity and kinetic parameters.....	120
4.7. Inhibition of BAD_1582 by lactose, glucose and galactose	124
4.8. Degalactosylation activity of BAD_1582.....	127
4.9. Transgalactosylation activity and kinetics of GOS synthesis.....	128
4.10 Discussion.....	134
Chapter 5	146
Immobilisation of the <i>Bifidobacterium adolescentis</i> β-galactosidase BAD_1582	146
5.1. Introduction.....	147
5.2. Preparation of the immobilised <i>Bifidobacterium adolescentis</i> β -galactosidase.....	152
BAD_1582 and the <i>Aspergillus oryzae</i> β -galactosidase.....	152
5.3. Reuse of the immobilised BAD_1582 and <i>Aspergillus oryzae</i> β -galactosidase.....	154
5.4. Effect of pH and temperature on the activity of free and immobilised β -	156
5.5. Kinetic parameter of the immobilised β -galactosidase (BAD_1582).....	159
5.5. Discussion.....	161
Chapter 6	165
Discussion	165
6.1. Functional screening of human faecal microbiome metagenome library for	166
glycoside hydrolase enzymes.....	166
6.2. Biochemical properties, substrate specificity and kinetic parameter of.....	170
BAD_1582.....	170
6.3. Transgalactosylation activity of BAD_1582	174
6.4. Biological role of BAD_1582 to <i>B. adolescentis</i>	180
6.5. Conclusions.....	181
References	184

Acknowledgments

I am deeply indebted to my supervisors Dr. Aoife Boyd, Prof. Conor O'Byrne and Dr. Michelle Kilcoyne for providing me the opportunity to conduct this Ph.D. project. Besides, I would like to thank you for the constructive criticism you have provided throughout the 4-year period and correction of the thesis, without your support this wouldn't be possible.

I would like to pass my gratitude to my GRC members, Dr. Cyril Carroll, Dr. Florence Abram and Dr. Stephen Cunningham, for the evaluation and criticism you have provided about the progress of the project which I have been working on. Your suggestions have been helpful to accomplish the project in a positive way.

I would like to pass my gratitude to the technical assistants in the department of microbiology, NUI Galway; Ann Smith, Katrina Lacey, Mike Coughlan and Maurice Martins for their continuous support and competence throughout the 4 year period with cheerful personalities. I am also thankful to Caroline O'Connell, secretary of the microbiology department, NUI Galway, for her cooperation and assistance throughout the 4 years.

I would like to thank Dr. Maria Tuohy, a senior lecturer in Biochemistry at NUI Galway, for kindly providing me access to instrumentations in her laboratory. Besides I wanted to thank Rosalia Martinez a Ph.D. student in Dr. Maria Tuohy group.

I wish to thank Prof. Margo Moore and Dr. Mateusz Lezyk for kindly providing recombinant plasmids used in this thesis. I am grateful to Glycom Company for kindly providing some of human milk oligosaccharides used in this study.

I am thankful for the 4th years & MSc students, current and former Ph.D. candidates (Dr. Christy Agbavwe, Neyaz Khan and Laura Cutugno) which have been working in the pathogenic mechanism group in the past four years period, for their cooperativeness in the lab and for making the lab a good working atmosphere. Besides, I would like to thank all the lab members (particularly, Ph.D. and postdoctoral researchers) which were based in the pathogenic mechanism and bacterial stress response group for the active participation through lab talk and journal club presentations, in the joint lab meetings. I

would like to thank all the Ph.D. candidates, postdoctoral researchers, in the Department of Microbiology, NUI Galway, for your cooperativeness and for coordinating social events to bring all of us together.

I would like to thank a couple of friends who make my time in Galway a bit cheerful, Yonas Hubtemichael, Hirut, Dr. Corine Nzeteu, Dr. Mesenbet Assefa, Dr. Karlo Simic, Dr. Octavi Domingo, Dr. Enrico Tati, Dr. Paul McAteer, Dr. Yinka Somorin, Jacob Cesna, Nick Waters, Dr. Joe Gibbs, Simon Mills, Miky, Melat, Amha, Bamlak, Asamnew, Hawa, John Forde, Stephen Murray, Ratul Kumar and Justin Dsouza.

Finally, I would like to thank my father Mr. Mehabie Mulualem and my mother Mrs. Yeshe Yrsaw, for your kindness, and for being supportive throughout the years, I wish a long life to both of you. I am also deeply indebted to my brother Mr. Getachew Mehabie, and my sister Mrs. Tigist (Mareye) Mehabie for their unwavering support and courage in all circumstances I have been through. The telephone calls and the group chats we had were very supportive in many ways to keep me going.

Abstract

Members of the human gut microbiota use glycoside hydrolase (GH) enzymes, such as sialidases, α -fucosidases and β -galactosidases to forage on host mucin glycans and dietary fibers. A human faecal metagenomic library was constructed and screened to identify new glycoside hydrolase enzymes particularly α -fucosidases, sialidases and β -galactosidases. 33,000 clones were screened for α -fucosidases on LB agar supplemented with X-Fucose as substrate and 11,000 clones were screened on MacConkey base agar supplemented with 2'-Fucosyllactose, however α -fucosidase expressing clones were not detected. Similarly out of 16,128 clones screened for sialidase on MacConkey base agar supplemented with 3'-Sialyllactose and 6'-Sialyllactose, sialidase expressing clones were not detected. However, out of the 16,000 clones screened for β -galactosidases on MacConkey lactose agar, 30 β -galactosidase positive clones were identified. The β -galactosidase gene found in the majority of the clones was BAD_1582, from *Bifidobacterium adolescentis*. This gene was cloned with the incorporation of a hexa histidine tag and expressed in *Escherichia coli*. The His-tagged β -galactosidase enzyme was purified using Ni²⁺-NTA affinity chromatography and size filtration. The enzyme had optimal activity at pH 7.0 and 37°C, with a wide range of pH (4-10) and temperature (0-40°C) stability. The enzyme requires a divalent metal ion cofactor, with the highest activity detected in the presence of Mg²⁺ and activity was inhibited by the presence of Cu²⁺ and Mn²⁺. The kinetic parameters were determined using *ortho*-nitrophenyl- β -D-galactopyranoside (*o*NPG) as a substrate. The enzyme had a V_{max} of 51 μ mol/min/mg and a K_m of 2.5 mM. The enzyme was competitively inhibited by lactose with a K_i of 3.0 mM and exhibited low product inhibition by galactose with a K_i of 115 mM and a high tolerance for glucose (66% of its activity retained in the presence of 700 mM of glucose). In addition, this enzyme displayed transgalactosylation activity to produce galactooligosaccharides (GOS) using lactose as a substrate. BAD_1582 was immobilized in calcium alginate and gelatin blend crosslinked with glutaraldehyde. The immobilized BAD_1582 displayed enhanced pH and temperature stability compared to the free enzyme. The aforementioned properties of BAD_1582 makes it an ideal candidate to produce lactose free dairy products and manufacture of GOS.

List of figures

Fig 1.1 The major sections of the gastrointestinal tract.	2
Fig 1.2 Histologic structure of the GIT.....	4
Fig 1.3 The gastrointestinal mucosal barrier.	8
Fig 1.4 Mucin glycoprotein structure	10
Figure 1.5 A schematic representation of the function based and sequence-based screening of the human faecal metagenome, used in this study	38
Fig 2.1 The features of the pET101/D-TOPO® (5753 bp) vector.....	49
Fig 3.1 Restriction map of pCC1FOS vector, adapted from <i>Medicago truncatula</i> genomic resource.....	70
Fig 3.2 Restriction analysis of 10 randomly selected metagenome clones and the vector (pCC1FOS) before and after <i>Bam</i> HI (a , b) and <i>Sma</i> I (c) restriction digestion.....	71
Fig 3.3 Selection of β -galactosidase expressing clones from human faecal microbiome metagenome library.....	76
Fig. 3.4 Miller assay on distinct fosmid clones and the <i>lacZ</i> positive <i>E. coli</i> BL21 (DE3) (positive control) and <i>lacZ</i> <i>E. coli</i> EPI300(pCC1FOS) (negative control) of 1st round screening.....	77
Fig 3.5 Restriction analysis of 8 β -galactosidase positive metagenome clones before and after <i>Bam</i> HI restriction endonuclease enzyme digestion.....	78
Fig 3.6 <i>E. coli</i> EPI300 host strain streaked on MacConkey base agar supplemented with 0.5% sialic acid (a) and supplemented with 3'SL (b).....	80
Fig 3.7 Restriction pattern of pAf sialidase after double restriction endonuclease digestion (using <i>Not</i> I and <i>Nde</i> I). Lane 1, 1 kb molecular ladder (M), lane 2-4 pAf plasmid from three random colonies.....	81
Fig 3.8 <i>E. coli</i> BL21 (DE3) (pAf) and <i>E. coli</i> BL21 (DE3) (pET28b) on MacConkey agar supplemented with 3'SL	82

Fig 3.9 Screening strategy of human faecal metagenome library on MacConkey base agar.....	83
Fig 3.10 Restriction pattern analysis of putative α -fucosidase metagenome clones after <i>Sma</i> I digestion.. ..	84
Fig 3.11 Human faecal metagenome library screening on X-Fuc supplemented LB agar.....	90
Fig 3.12 Growth curves of <i>E. coli</i> EPI300 (pCC1FOS) and β gal clone #11 inoculated in M9 minimal media with or without carbon source under aerobic condition.....	92
Fig. 4.1 Gene map of fosmids which incorporated the nucleotide sequences of different regions of the genome of <i>B. adolescentis</i> and the six common genes shared between fosmids.....	105
Fig. 4.2 PCR of representative BAD-1582 harboring metagenome clones. PCR amplification products of the target gene (3.15 kb) of 17 clones (lanes 1-19)	105-106
Fig. 4.3 Homology based structure prediction of the <i>B. adolescentis</i> β -galactosidase, BAD_158.....	106
Fig. 4.4 Putative conserved domains of the protein sequence of BAD_1582, in the CDD and Pfam databases.....	107
Fig. 4.5 Evolutionary relationship of BAD_1582 with homologous proteins and landmark similar sequences. The evolutionary history was inferred using UPGMA method.....	108
Fig. 4.6 Restriction pattern analysis of recombinant clone pDMg1a and empty vector pET101 after double digestion using <i>Hind</i> III and <i>Sac</i> I.....	109
Fig. 4.7 The recombinant clone <i>E. coli</i> T7express (pDMg1a) and empty vector harboring clone <i>E. coli</i> T7express (pET101) streaked on LB agar supplemented with X-gal in the	

presence of 1 mM IPTG, indicating the expression of the recombinant β -Galactosidase (BAD_1582).....	110
Fig. 4.8 Miller assay on <i>E. coli</i> T7express (pDMg1a), <i>E. coli</i> T7 express (pET101) and <i>lacZ+</i> <i>E. coli</i> BL21 (DE3) with and without presence of 0.1 mM IPTG.....	111
Fig. 4.9 Miller assay on <i>E. coli</i> T7express (pDMg1a) (pLysS), <i>E. coli</i> T7 express (pET101) (pLysS) and <i>lacZ+</i> <i>E. coli</i> BL21 (DE3) with and without presence of 0.1 mM IPTG.....	112
Fig. 4.10 SDS PAGE analysis of expression of recombinant β -galactosidase (BAD_1582) after 4 h IPTG induction	113
Fig. 4.11 BAD_1582 Purification. Purification of the protein analysed by SDS-PAGE of purification step samples.....	114-115
Fig. 4.12 Effect of pH and temperature on BAD_1582.....	117
Fig. 4.13 Effect of detergents and denaturants on the activity of BAD_1582 in 50 mM sodium phosphate buffer.....	120
Fig. 4.14 Kinetic parameter determination of BAD_1582 using <i>o</i> NPG and lactose as a substrate.....	123
Fig. 4.15 Effect of lactose and its hydrolysis products on activity of BAD_1582 towards hydrolysis of <i>o</i> NPG after 30 min incubation.....	126
Fig. 4.16. Deagalactosylation activity of BAD_1582.....	127
Fig. 4.17 TLC analysis of transgalactosylation and hydrolysis activities of BAD_1582.	130-131
Fig. 4.18 HPLC analysis of transgalactosylation and hydrolysis activities of BAD_1582.....	133
Fig 5.1 Resuse of immobilised BAD_1582 and <i>Apsergillus oryzae</i> β -galactosidase.....	155

Fig 5.2 Effect of pH on activity and stability of free and immobilised BAD_1582.....157

Fig 5.3 Effect of temperature on activity and stability of free and immobilised BAD_1582158

Fig 5.4 Time course progress curve and non-linear regression fit of immobilised BAD_1582.....160

List of Tables

Table 1.1 Mucin types and distribution.....	6
Table 2.1 Bacterial strains, plasmids and fosmids used in this study.....	40
Table 3.1 Blast record of randomly picked metagenome clones.....	73-74
Table 3.3 Blast record of β -galactosidase positive distinct metagenome clones.....	79
Table 3.4 Blast record of putative α -fucosidase positive metagenome clones.....	85
Table 3.5 annotated protein details of putative fucosidase clone 7 sequence covered from 1307594-1339310 (31,716bp in length) of the matched strain (<i>B. adolescentis</i> ATCC 15703).....	87
Table 3.6 annotated protein details of putative fucosidase clone 9 sequence covered from 1339483-1376507 (37,024bp in length) - of the matched strain (<i>B. adolescentis</i> ATCC 15703).....	88
Table 3.7 annotated protein details of putative fucosidase clone 11 sequence covered from 1348975-1383124(34,149bp in length) - - of the matched strain (<i>B. adolescentis</i> ATCC 15703).....	89
Table 4.1 Effect of divalent cations and EDTA on the activity of BAD_1582 in 50 mM phosphate buffer/100 mM Tris-HCl buffer.....	119
Table 4.2 Substrate specificity of BAD_1582 tested on <i>p</i> NP based substrates.....	121
Table 4.3 The kinetic parameters of BAD_1582 using <i>o</i> NPG and lactose as substrate, and in the presence of lactose and D-galactose as competitive inhibitors of <i>o</i> NPG hydrolysis.....	124

List of abbreviations

Amp- Ampicillin

AmpR- Ampicilin resistance

CAZymes-Carbohydrate Active Enzymes

Cm- Chloroamphenicol

EFSA-European Food Safety Authority

FDA- Food and Drug Administration (U.S.)

FSAI- Food Safety Authority Ireland

FOS-fructooligosacchrides

2'-FL- 2-fucosyllactose

Gal-Galactose

GalNAc - N-acetyl-D-galactosamine

GIT-gastrointestinal tract

Glc-Glucose

GlcNAc -N-Acetyl-D-glucosamine

GOS- galactooligosacchrides

GRAS-Generally Recognized As Safe

HMOs-Human Milk Oligosaccharides

HPLC- High Performance Chromatography

HPAEC-PAD-High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection

IPTG- Isopropyl- β -D-thiogalactopyranoside

Kan-Kanamycin

NEB-New England Biolabs

OD- Optical density

*o*NPG- ortho nitrophenyl β -D-galactopyranoside

*p*NPG- para nitrophenyl β -D-galactopyranoside

3'SL- 3-Sialyllactose

6'SL- 6- Sialyllactose

SCFA- Short Chain FattyAcids

TLC-Thin Layer Chromatography

XOS- Xylooligosaccharides

WHO-World Health Organisation

V/V -Volume per volume

W/V-Weight per volume

W/W- Weight per weght

Chapter 1

Introduction

1.1. Anatomy of the human gastrointestinal tract

The human gastrointestinal tract (GIT) starts from the oral cavity and continues through the pharynx, oesophagus, stomach, small intestine (duodenum, ileum and jejunum) to the large intestine (cecum, colon, rectum and anus) (Fig 1.1) (Kararli 1995). Accessory organs, such as the salivary glands, liver, pancreas and the gall bladder assist the function of the digestive system by secretion of enzymes and /or emulsifying agents etc. The main function of the GIT is to breakdown ingested food through mechanical and digestive enzyme actions, which then absorb nutrients and water through circulatory or lymphatic systems, while excreting undigested material and secreted waste. The absorbed nutrients and water are utilised by the different cells for growth and differentiation and ultimately result to the healthy build-up of the human body (Van de Graaff 1986).

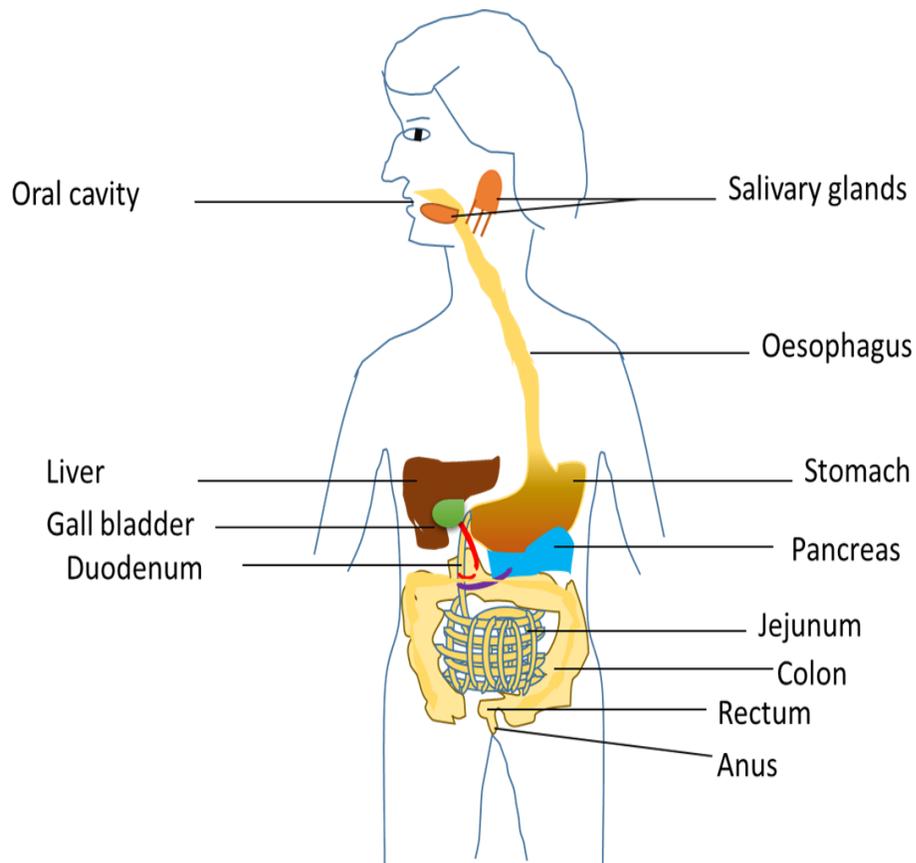


Fig 1.1 The major sections of the gastrointestinal tract. This picture is Adapted from (Walter and Ley 2011)

The histologic structure of the GIT consists of the mucosa, the submucosa, muscularis mucosae, and the serosa (Fig 1.2). The lumen of the GIT is surrounded by the mucosa, and contains a layer of cells called the epithelium supported by the lamina propria, a connective tissue which contains lymph nodules specialised in disease protection. The two thin layers of smooth muscles known as the muscularis mucosae are located outside of the lamina propria, and they are responsible for the folding of parts of the GIT. The folding increases the surface area for absorption of nutrients, as well as their ability to stretch and contract to enable the distention of the stomach to increase volume, increase the surface area for absorption of nutrients in the small intestine and to extract water in the large intestine. The specialised goblet cells in the mucosa secrete mucus throughout the GIT, however the thickness and density of the mucous layer varies across different parts of the GIT. The submucosa surrounds the mucosa muscles and consists of fats, fibrous connective tissues, blood vessels and nerves. At the outer side of the submucosa is a specialised nerve plexus called the submucosa plexus or Meissner plexus which innervates the muscularis mucosa. The muscularis externa (circular and longitudinal muscles) are responsible for churning of the bowel and peristaltic movement. These muscles are innervated by the myenteric plexus (plexus of Auerbach) located between the circular and longitudinal muscles. The outer layer of the GIT is the serosa (mesentery), which is the binding and protective layer comprised of loose fibrous connective tissue rich in simple squamous epithelium (Lowe and Anderson 2015; Van de Graaff 1986).

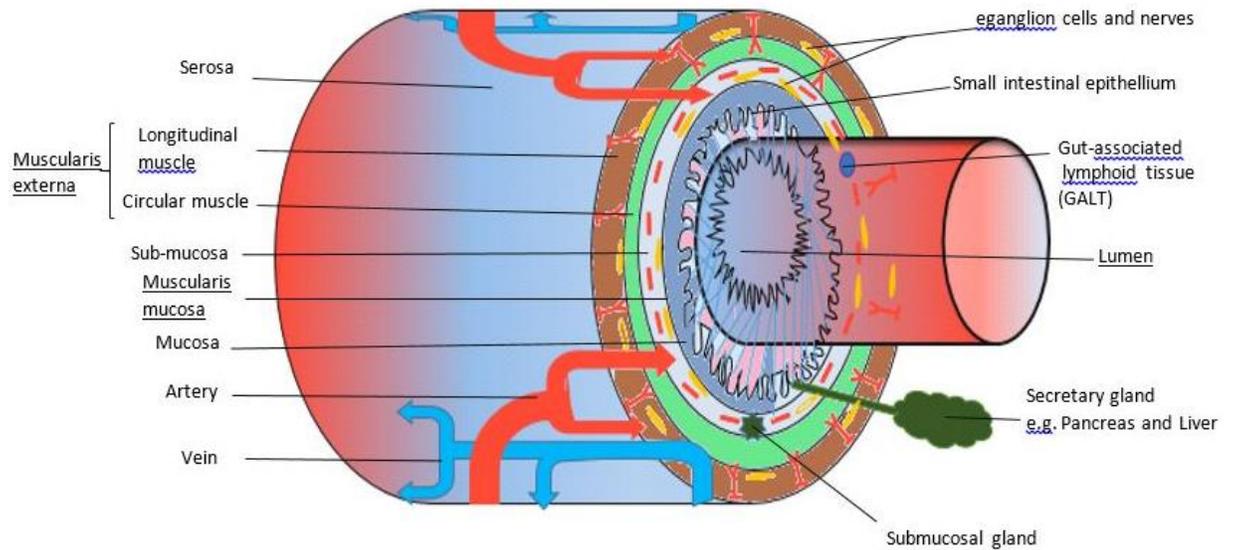


Fig 1.2 Histologic structure of the GIT. The GIT is a muscular tube lined internally by specialized epithelium which, for most of its length, has a combined secretory and absorptive function. The figure depicts the histologic structure of the GIT and the various secretory organs (e.g. pancreas and liver) which produce secretions which are vital for the tract to function normally; these secretions are emptied to the digestive tract lumen by one or more ducts. This figure is adapted from (Lowe and Anderson 2015).

1.2.The mucus lining in GIT

The gastrointestinal tract has exposure to exogenous substances and microorganisms in the form of ingested foods and drinks and endogenous chemicals such as HCl and proteolytic enzymes secreted in our body (Pelaseyed *et al.*, 2014). Despite these, it carries out all its function by maintaining homeostasis (Johansson *et al.*, 2013). The mucus layer serves as first line defence to proteolytic enzymes, acids, chemicals, microorganisms and their metabolites (Kim and Ho 2010). The mucus is a complex viscous adherent secretion synthesised primarily by goblet cells, in columnar epithelium of various mucosal surfaces such as the GIT, respiratory tract, reproductive tracts, and ocular and otorhinolaryngeal tracts (Bansil and Turner 2006). Mucus is comprised of primarily water which accounts for ~95%, lipids, salts, diverse antimicrobial molecules, proteins such as lysozyme, secreted immunoglobulins which target specific microbial or viral antigens, defensins,

growth factors and trefoil factors (Bansil and Turner 2006; McGuckin *et al.*, 2011b). However, the viscoelastic properties of mucus results from mucin, a glycoprotein which constitutes the major structural back bone of the mucus (Juge 2012). As a component of the body's innate immunity response the mucus layer and its various components serve as lubrication agents and act as a barrier to entry of pathogens and harmful substances to the epithelium but allow diffusion of gases and nutrients (Dhanisha *et al.*, 2018). The densely glycosylated mucins in the GIT are resistant to proteases secreted from the pancreas or intestinal epithelium as the glycosylation hinders the proteases from reaching the protein backbone (Arike and Hansson 2016).

To date over 21 MUC genes have been identified, each MUC genes are characterized by their particular expression depending on the physiological need in the cell (Table 1.1). There are two major types of mucins, the secreted mucins and membrane-associated mucins (Table 1.1). The predominant secreted mucins are the gel forming mucins. The secreted gel forming mucins consist of a very large core protein (~5000 amino acid residues) which is characterised by having cysteine rich N-terminal and C-terminal regions which enhance oligomerisation by disulphide bond formation (Hollingsworth and Swanson 2004). In addition, these mucins contain N and C-terminal D domains (dimerization domain), named after sequence similarity to the von Willebrand factor D. The D-domains of the C-terminal core protein form dimers while the N-terminal D-domains form trimers through disulphide bonding between cysteine residues (Lang *et al.*, 2007). The oligomers are usually hydrated and form viscoelastic gels, a characteristic feature of the protective mucus layer. The gel forming secreted mucins are expressed in epithelial cells of the respiratory and digestive system, where they function as major protective barrier to epithelial cells from adverse conditions such as food particles, viruses, bacteria, acidic secretions, enzymes etc. (Hollingsworth and Swanson 2004). In airways the secreted mucins play significant roles in the clearance of airways via mucocilliary transport (Ostedgaard *et al.*, 2017).

Table 1.1. Mucin types and distribution

Mucin type	Chromosome	Major distribution	References
Secreted –gel forming			
MUC2	11p15.5	Small intestine, colon, endometrium, respiratory tract, eye & middle ear epithelium	(Audie <i>et al.</i> , 1993; Buisine <i>et al.</i> , 1999; Buisine <i>et al.</i> , 2000; Copin <i>et al.</i> , 2001; Gum <i>et al.</i> , 1994)
MUC5AC	11p15.5	Respiratory tract, stomach, cervix, Eye & middle ear epithelium	(Buisine <i>et al.</i> , 1999; Gipson <i>et al.</i> , 2003; Porchet <i>et al.</i> , 1995)
MUC5B	11p15.5	Respiratory tract, salivary glands, cervix, gallbladder, seminal fluid, middle ear epithelium	(Audie <i>et al.</i> , 1993; Audie <i>et al.</i> , 1995; Balague <i>et al.</i> , 1994)
MUC6	11p15.5	Stomach, duodenum, gallbladder, pancreas, seminal fluid, cervix, middle ear epithelium	(Gipson <i>et al.</i> , 1997; Ho <i>et al.</i> , 1995; Toribara <i>et al.</i> , 1997)
MUC19	12q12	Sublingual gland, submandibular gland, respiratory tract, eye, middle ear epithelium	(Chen <i>et al.</i> , 2004)
Secreted –non gel forming			
MUC7	4q13.3	Salivary glands, respiratory tract, middle ear epithelium.	(Bobek and Situ 2003)
MUC8	12q24.33	Respiratory tract, uterus, endocervix, endometrium.	(D’Cruz <i>et al.</i> , 1996; Shankar <i>et al.</i> , 1994)
MUC9	1p13.2	Fallopian tubes	(Arias <i>et al.</i> , 1994)
Membrane associated			
MUC1	1q21	Stomach, breast, gallbladder, cervix, pancreas, respiratory tract, duodenum, colon, kidney, eye, B cells, T cells, dendritic cells, middle ear epithelium	(Andrianifahanana <i>et al.</i> , 2001; Arul <i>et al.</i> , 2000; Brayman <i>et al.</i> , 2004; Buisine <i>et al.</i> , 2000; Copin <i>et al.</i> , 2001; Gendler and Spicer 1995; Lan <i>et al.</i> , 1990)
MUC3A	7q22.1	Small intestine, colon, gall bladder, duodenum, middle ear epithelium	(Gum <i>et al.</i> , 1994)
MUC3B	7q22	Small intestine, colon, gall bladder, duodenum, middle ear epithelium	(Pratt <i>et al.</i> , 2000)
MUC4	3q29	Respiratory tract, colon, stomach, cervix, eye, middle ear epithelium	(Gipson <i>et al.</i> , 1997; Porchet <i>et al.</i> , 1995)
MUC12	7q22	Colon, small intestine, stomach, pancreas, lung, kidney, prostate, uterus	(Moehle <i>et al.</i> , 2006; Williams <i>et al.</i> , 1999)
MUC13	3q21.2	Colon, small intestine, trachea, kidney, appendix, stomach, middle ear epithelium.	(Williams <i>et al.</i> , 2001)
MUC14	4q24	Heart, kidney, lungs	(Liu <i>et al.</i> , 2001)
MUC15	11p14.3	Spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocyte, bone marrow, lymph node, tonsil, breast, fetal liver, lungs, middle ear epithelium	(Pallesen <i>et al.</i> , 2002; Shyu <i>et al.</i> , 2007)
MUC16	19p13.2	Peritoneal mesothelium, reproductive tract, respiratory tract, eye, middle ear epithelium	(Argüeso <i>et al.</i> , 2003; Matsuoka <i>et al.</i> , 1990)
MUC17	7q22	Small intestine, colon, duodenum, stomach, middle ear epithelium	(Gum <i>et al.</i> , 2002; Moehle <i>et al.</i> , 2006)
MUC20	3q29	Kidney, placenta, colon, lung, prostate, liver, middle ear epithelium	(Higuchi <i>et al.</i> , 2004)
MUC21	6p21	Respiratory tract, thymus, colon	(Itoh <i>et al.</i> , 2008)
MUC22	6p21.33	Lungs, placenta, testis	(Hijikata <i>et al.</i> , 2011)

As of now three secreted non gel forming mucins, MUC 7, 8, and 9 have been identified, each of them has specialised functions. For instance, MUC 7 is the major salivary mucin, secreted by salivary glands and involved in food mastication, speech, and clearing microbes from the oral cavity through its antimicrobial properties (Bobek and Situ 2003). MUC 9 is only produced in female fallopian tubes, and is presumably involved in protection of the fallopian tube and embryos (Lapensee *et al.*, 1997).

The membrane-associated (transmembrane) mucins, contain a membrane spanning hydrophobic domain and display similar properties with other membrane anchored proteins (Corfield 2015). The transmembrane mucins in humans contain an extracellular sea urchin sperm protein, with an enterokinase and agrin (SEA) domain which undertakes auto proteolysis cleavage to produce α and β mucin chains in the endoplasmic reticulum. After the cleavage, the chain is protruded extracellularly and the β -chain spans the membrane, yet both chains are connected by parallel β -sheets (Macao *et al.*, 2006). The cytoplasmic tails of the transmembrane mucins constitute the intracellular domain, which play significant role in intracellular signalling pathways through their phosphorylation sites (van Putten and Strijbis 2017). The transmembrane mucins have been implicated in prevention of adhesion of pathogenic bacteria such as *Helicobacter pylori* to the epithelial cells (Lindén *et al.*, 2009). Through immunomodulatory and various signalling pathways, the transmembrane mucins act with anti-inflammatory activity in healthy epithelial cells. Nevertheless, overexpression or aberrantly expressed transmembrane mucins, or changes in the glycosylation status contribute to various types of cancers such as colon cancer, breast cancer and pancreatic cancer (Balague *et al.*, 1994; van Putten and Strijbis 2017).

In the GIT, the secreted mucins are primarily synthesised by the goblet cells and the Paneth cells (Fig 1.3). The Paneth cells are specialised small intestinal epithelium cells which reside in the intestinal crypts adjacent to the stem cells (Clevers and Bevins 2013). The Paneth cells are specialised to secrete mucins and diverse array of antimicrobial proteins, whereby keep a sterile environment around the locality of the stem cells in the intestinal epithelium. In the Paneth cells, the expression of antimicrobial molecules such as RegIII γ and key α -defensin, Defcr-rs10 are regulated by NOD-like proteins and Toll-like receptors (TLRs) (Vaishnava *et al.*, 2008).

The mucus layer is thin in the small intestine and its thickness increases in the colon (Johansson *et al.*, 2011). In the small intestine the mucus layer is single and unattached to the epithelial cells, whereas in the colon it has two sub layers, an extended outer layer with a thickness of about 800 μm and an inner layer (50-100 μm thick) which is tightly bound to the epithelial cells (Fig 1.3) (Atuma *et al.*, 2001; Johansson *et al.*, 2008). The outer thick layer serves as point of contact with commensal microbial community (Fig 1.3), however the inner layer is presumed to be inaccessible for the microbiome and hence keeps the epithelial cells away from such interactions (Johansson and Hansson 2016).

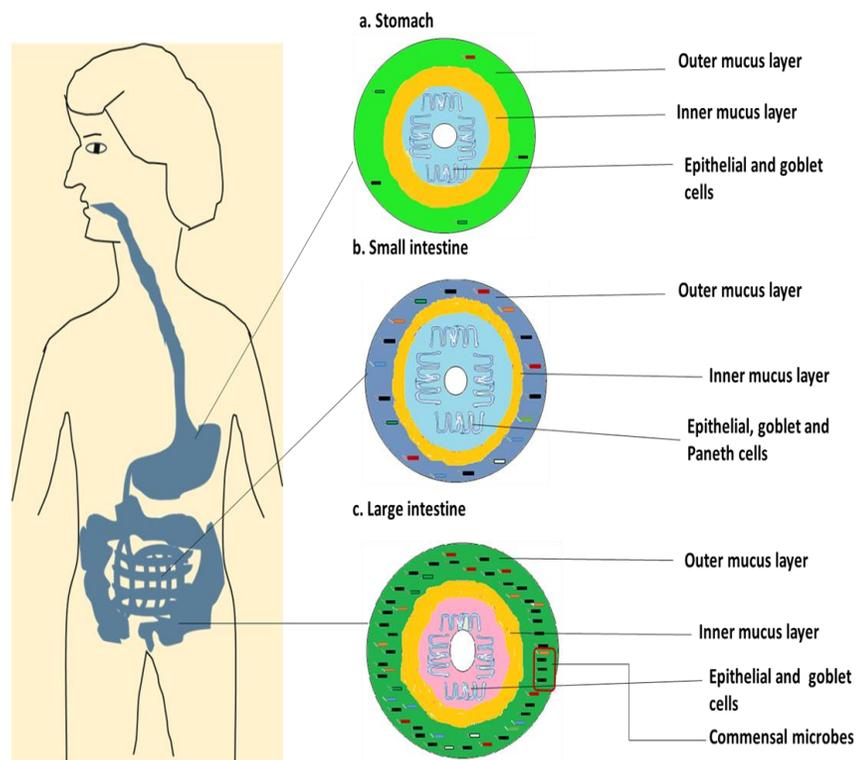


Fig 1.3 The gastrointestinal mucosal barrier. **a** The outer layer of the stomach measures $\sim 120 \mu\text{m}$ in rats and mice and contains low numbers of bacteria, whereas the inner layer is sterile and measures $\sim 100\text{--}150 \mu\text{m}$. **b** The mucus layer is thinner in the small intestine, with inner layer of $\sim 15\text{--}30 \mu\text{m}$ and an outer layer of $100\text{--}400 \mu\text{m}$; it is thickest in the ileum, $\sim 10^5\text{--}10^7$ bacteria present per gram of faeces in the lumen. **c** Mucus in the large intestine (mainly MUC2) is predominantly produced by goblet cells and consists of a sterile inner layer of $\sim 100 \mu\text{m}$ and a thick outer layer of $\sim 700 \mu\text{m}$. Numbers of bacteria

are greater in the large intestine 10^{10} – 10^{12} per gram of faeces. This picture is adapted from (McGuckin *et al.*, 2011b)

1.2.1. Mucin O-glycosylation

Glycosylation is one mode of post translational modification, whereby sugars (glycans) are enzymatically added to specific amino acid residues in the peptide backbone of a protein (Moran *et al.*, 2011). In N-linked glycosylation glycan repertoires are added on asparagine residues in an Asn-X-Ser sequences (where X-can be any amino acid other than proline) (Corfield *et al.*, 2001). On the other hand in O-linked glycosylation glycan repertoires are added on Serine or Threonine residues. Mucins are glycoproteins characterized by O-linked glycans linked to residues in the PTS (Proline, Threonine, Serine- rich) region, where the O-glycans constitute 50% or more by mass (Johansson and Hansson 2016; Juge 2012).

Mucin O-glycosylation is initiated by peptidyl-GalNAc transferases which utilize UDP-GalNAc to add GalNAc in an alpha-configuration to the hydroxyl group of serine or threonine residues in PTS region (Bennett *et al.*, 2012). Twenty GalNAc transferase genes have been identified from the human genome and classified as GT27 family members in the CAZy glycosyltransferase classification (Bennett *et al.*, 2012). After addition of GalNAc the glycan chain is elongated with core structures and two potential backbone regions containing N-acetyl-glucosamine (GlcNAc) and galactose (Gal) (Fig 1.4). In the core 1 structure, elongation is carried out by addition of galactose by β 1, 3- galactosyl transferase (C1GalT1 or T-synthase) to GalNAc in β 1, 3 –linkage. In the core 3 structure, elongation is carried out by β 1, 3- N-acetylglucosamine transferase 6, which adds GlcNAc to GalNAc in β 1, 3 –linkage. Core 1 and core 3 structures are further modified to generate Core 2 and Core 4 respectively, by addition of GlcNAc using β 1,6- N-acetylglucosamine transferases (Tran and Ten Hagen 2013). Termination usually occurs with addition of fucose (Fuc) using fucosyl transferases (FUT) and/or sialic acid (Neu5Ac) using sialyltransferases (ST) from an activated nucleotide donor (CMP-Neu5Ac) (Mehr and Withers 2016).

The structural diversity of mucins results from addition of several glycans such as galactose, hexosamines, fucose, sialic acid and sulfate groups (Macfarlane *et al.*, 2005). The particular sugar chain linked to the mucin protein depends on the glycosyl-transferases expressed at the moment of mucin production (Linden *et al.*, 2008). Both Gal and GlcNAc residues can be sulphated resulting in an even greater degree of diversification. Addition of sulfate and sialic acids to terminal mucin oligosaccharides confers resistance to digestion by glycosidases, nevertheless some faecal isolated microbes are enriched with glycosidases which can deconstruct mucin glycoproteins (Corfield *et al.*, 1992).

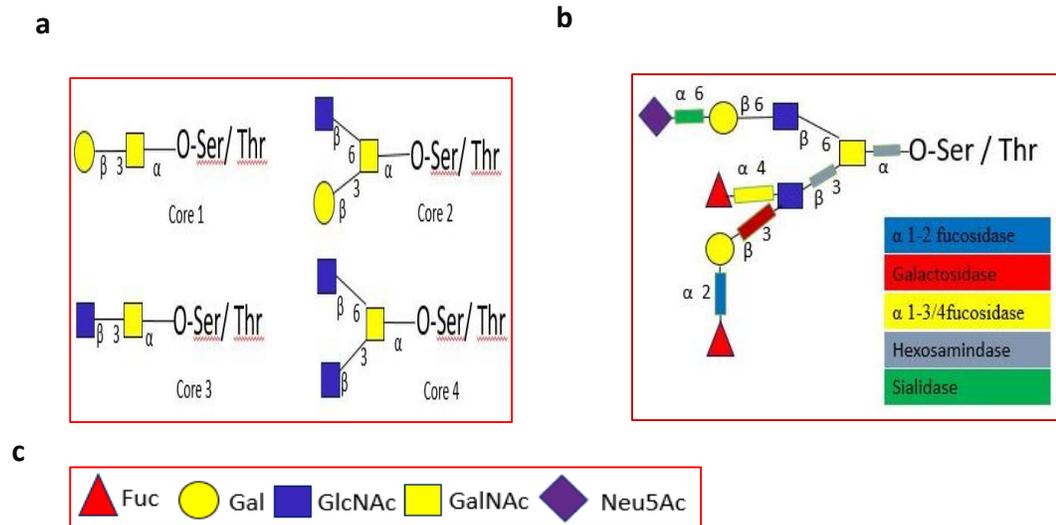


Fig 1.4 Mucin glycoprotein structure. **a** Core structures of mucin O- glycans. **b** extended core 4 mucin O-glycan structure and Glycoside hydrolase enzymes which cleave different linkages of mucin glycans to liberate individual monosaccharides. **c** schematic representation of monosaccharides. This figure is taken from (Marcobal *et al.*, 2013).

The O- linked glycan repertoire of the mucins serves as binding sites for commensal microbiota and some members of the microbiota even utilize these glycans as an energy source (Koropatkin *et al.*, 2012). The gut microbiota must synthesize a variety of hydrolytic enzymes (glycosidases, proteases, peptidases, and sulfatases) to completely degrade the mucus layer (Macfarlane *et al.*, 2005). In scenarios of scarcity or complete

depletion of dietary fibers rich in glycans, some members of the microbiota express glycoside hydrolases such as sialidases, α -fucosidases, β -galactosidases for deconstruction of mucin glycans and utilization of them for nutrient sources (Sonnenburg *et al.*, 2005). The depletion of the mucus layer as a result of microbial action causes the epithelial cells to be exposed for microbial colonization, which further contributes to the onset of pathogenesis. Furthermore, a direct contact of the microbiota with the epithelial cells trigger mucosal immunity and elicit overt immune responses (Johansson and Hansson 2016). In line with this, mice studies have shown that mice lacking MUC2 have no mucus layer which resulted in the development of inflammation and diarrhea, bleeding and epithelial cell proliferation which could lead to tumor formation.

1.3. Microbial adhesion to the GIT

The glycoconjugates or glycocalyx coat covering the cell surface of multicellular organisms is regarded as an interface where the microbes (commensal & symbiotic) or pathogens produce cell-cell contact (Imberty and Varrot 2008a; Ohtsubo and Marth 2006; Vasta 2009). Microbial adhesins play pivotal roles in such contact. Adhesion to cell surfaces is the primary step in colonization of the gut by commensal microbes and in biofilm formation and associated infections by pathogens (Dufrêne 2015). Throughout the years several surface bound adhesive proteins have been identified (Kline *et al.*, 2009). Most of these adhesins interact with specific ligands on host cells, which increases the chance of invasion and colonization. Often, the binding specificity involves fucosylated and sialylated glycoconjugates since these residues are exposed at the surface of epithelia (Imberty and Varrot 2008b). Pathogenesis of *P. aeruginosa* in cystic fibrosis patients is established by soluble fucose binding lectins which target fucosylated oligosaccharides (Mitchell *et al.*, 2002). Some of the adhesins from pathogenic *E. coli* have specific affinity to sialic acid (S adhesins), others bind to other glycan repertoires such as α -D-mannosides (MS adhesins) and glycophorin A (M adhesins) (Ott *et al.*, 1986). Protein-protein interactions also occur between adhesins and host proteins (Tuomanen 1992).

During infection, pathogenic bacteria use their surface attached adhesins to colonize different cells and tissues of the host. These adhesins bind to their cognate receptors on epithelial cells, leukocytes, endothelial cells or extracellular matrix (Tuomanen 1992).

Even though bacterial attachment to the host cells is essential for colonization, adhesion may have unprecedented cost, as it can stimulate infiltration of immune cells, phagocytosis and thereby clearing of bacteria from the host (Kline *et al.*, 2009). Many pathogenic bacteria have developed ways to escape the immune system, among these having an additional anti-phagocytic polysaccharide rich surface layer which is not recognized by the immune system and expression of adhesins as polymeric structures that extend from cell surfaces for establishing initial contact are highly recognized (Kline *et al.*, 2009).

Adherence of probiotic bacteria such as *Lactobacilli* to intestinal epithelial cells and/ or mucus lining is considered to be a desirable feature as it can promote the residence time and interaction with host epithelial and immune cells (Juge 2012; Kleerebezem *et al.*, 2010). In addition to specific bacterial adhesins, other cell surface proteins called S-layer proteins and extracellular appendages such as flagella, fimbriae and pili (Juge 2012) can also contribute to adhesion to host epithelial cells and mucus.

1.4. Colonization of the GIT by microbiota

It is believed that a healthy human foetus in the maternal womb is devoid of microorganisms. The first major colonisation of the new born infants begins during the delivery of the infant due to initial exposure with significant amount of microorganisms residing in the vaginal canal for vaginal born infants or skin for caesarean section (C-section) born infants. Thus, the gut of vaginal born infants is colonised by microorganisms that are common inhabitants of the vagina such as *Lactobacillus*, *Prevotella*, *Bifidobacterium* (such as *Bifidobacterium longum* and *Bifidobacterium breve*), *Escherichia* (*Escherichia coli*), *Bacteroides* (*Bacteroides vulgatus*) and *Parabacteroides* (*Parabacteroides distasonis*) (Shao *et al.*, 2019). However, the guts of C-section delivered infants are predominantly colonised by skin microbiota and hospital associated microbiota such as *Enterococcus* (*Enterococcus faecium*), *Enterobacter* (*Enterobacter cloacae*), *Staphylococcus epidermis* and *Klebsiella* species with low abundance of commensal microbes (Shao *et al.*, 2019). These observations are in line with the report that the microbiome transfer from mother to infant is compromised in C-section delivered neonates (< 4 weeks old) and 1 year old infants (Backhed *et al.*, 2015).

Besides the mode of delivery, there are several factors which contribute to the diversity and composition of the gut microbiota in infants such as, mode of feeding as in breast feeding or formula feeding, diseases, antibiotic treatment, and introduction of solid foods and environment (Marques *et al.*, 2010). It has been recommended that, infants should exclusively be breast fed for the first 6 months and should continue for up to 2 years age or beyond for the a healthy growth and development (WHO 2003). Human milk is comprised of different nutrients and bioactive active molecules which protect against infection, inflammation and immune maturation (Ballard and Morrow 2013). The macronutrients include proteins (0.9 to 1.2 gdL^{-1}), fats (3.2 to 3.6 gdL^{-1}) and 6.7 to 7.8 gdL^{-1} lactose (Ballard and Morrow 2013). Next to lactose the human milk oligosaccharides (HMOs) constitute the largest carbohydrate component in the human milk, typically found at concentrations $\geq 4 \text{gdL}^{-1}$ and as high as 15 gdL^{-1} depending on the lactation stage, feeding and maternal genetics (Kunz *et al.*, 2000). HMOs are composed of five monosaccharides: glucose (Glc), galactose (Gal), N-acetylglucosamine (GlcNAc), fucose (Fuc) and sialic acid (Cani *et al.*,) with N-acetylneuraminic acid (Neu5Ac) the predominant sialic acid (Bode 2012). The composition HMOs is dependent on expression of certain glycosyltransferases; thus, four groups of HMOs have been identified based on secretor (Se) and Lewis (Le) blood group system. Secretor (Se) individuals encode an active $\alpha 1, 2$ fructosyltransferase (*FUT2*) and hence the milk will be abundant in 2'fucosyllactose (2'FL) while in the non-secretor their HMO lack 2'FL due to lack of active *FUT2*. Individuals with an active *Le* locus, express an active $\alpha 1-3/4$ -fucosyltransferase (*FUT3*) which transfers Fuc in $\alpha 1-4$ linkage to sub terminal GlcNAc, and hence women who are Le negative lack $\alpha 1-4$ -fucosylated HMOs (Bode 2012). These patterns of HMO composition is genetic dependent where approximately 20% of population are non-secretors and this number increases to~ 40% in the African population (Ferrer-Admetlla *et al.*, 2009)

The HMOs of the human milk are considered to have potential prebiotic properties. As they are not digested by mammalian glycosidases, HMOs reach the colon unaltered where they selectively enhance the growth of selected commensal bacteria (De Leoz *et al.*, 2015). Several in vitro studies have shown that the HMOs have bifidogenic activity, by which they selectively promote the growth and proliferation of *Bifidobacterium* species.

Some species of *Bifidobacterium*, such as *B. bifidum* and *B. infantis*, that are predominantly found in breast fed infants, induce expression of genes specifically required for HMO utilization when they were grown in the presence of these oligosaccharides (Yoshida *et al.*, 2012). The other infant *Bifidobacterium* species, *B. longum* subsp *longum*, has been shown to express novel gene clusters which enable this species to utilize fucosylated HMOs, particularly two α -fucosidases which cleave terminally attached Fuc and abundant solute binding proteins for oligosaccharide transport (Garrido *et al.*, 2016).

Culture based studies have indicated that breast fed infants have high abundance of faecal *Bifidobacteria* and *Lactobacilli* and with less abundance of *Bacteroides*, *Clostridium coccoides* group, *Staphylococcus* and members of the *Enterobacteriaceae* family compared to that of formula-fed neonates (Fallani *et al.*, 2010). Sequence based studies have supported this observation as the gut microbiota of breast fed infants was dominated by *Bifidobacteria* (Milani *et al.*, 2015b). Sequence based analysis of the faecal microbiota of healthy infants (2-5 months of age) have indicated that the microbiota of infants was dominated by the Actinobacteria (88%) phyla and with less abundance of Firmicutes (11%) (Turroni *et al.*, 2012). The Bifidobacteriales was the most abundant class (80.6%) followed by Lactobacillales (7.2%) and Clostridiales (3.1%) class. The effect of diet in this study was less significant in terms of microbiota diversity in the infants (Turroni *et al.*, 2012). In summary, breast fed infants have a distinct gut microbiota compared to the formula fed infants and the feeding type is considered as one of the major factor which dictate the microbiota composition in infants. Nonetheless, the gut microbiota composition in infants change until a stable, more diverse microbiota is attained, and by the age of three and beyond, the toddler/child attains a more or less similar microbiota to the adults (Yatsunenکو *et al.*, 2012a).

1.5. Diversity and composition of adult microbiota

The adult human GIT harbours a highly diverse and enormous number of microorganisms which belong to the three domains of life, bacteria, archaea and eukarya, creating a complex community. Most of these microbes are either strict obligate anaerobes or facultative anaerobes (Zoetendal *et al.*, 2002). These endogenous microbes make up an estimated 100 trillion cells, about 10 times higher than the number of human cells in a body (Gill *et al.*, 2006). The genomes of these complex microbial communities together constitute about 100 times more genes than the human genome encodes (Backhed *et al.*, 2005). The gut microbiome endows us with numerous physiological functions which we ourselves couldn't be able to carry out.

The advances in sequencing technologies and bioinformatics tools has enabled scientists to investigate and study the genome of various organisms including human beings (Collins *et al.*, 2003). The results of the human genome project depicted ~20,000 protein coding genes. After the successful completion of the human genome project in 2003 (Collins *et al.*, 2003), the first human microbiome project was launched to understand and decipher the genetic resources of the symbiotic microbes that reside inside and on the human body, which encode for a large number of protein coding genes that are absent in the human genome (Turnbaugh *et al.*, 2007).

The human body harbours a large number of microbes in different body sites such as oral, skin, vaginal and the gut, however the largest number of microbes reside in the GIT (Turnbaugh *et al.*, 2007). The microbial diversity of these endogenous gut microbiota has been examined through culture independent approaches, primarily based on comparison of 16S rRNA sequences and phylogenetic analyses. Of note one of the leading studies of microbial diversity and composition of the human gut microbiota was conducted by Eckburg and colleagues (Eckburg *et al.*, 2005) where they analysed 13,355 rRNA prokaryotic sequences obtained from mucosal sites of the colon (from cecum to rectum) and faecal samples obtained from three healthy individuals. This study revealed the major Bacterial phylotypes are Firmicutes and Bacteroidetes, with less abundance of Proteobacteria, Actinobacteria, Fusobacteria and Verrucobacteria and one archaeal phylotype (*Methanobrevibacter smithii*) was detected. The majority of these microbes

were novel and were not cultured. Similarly phylogenetic analysis of 8,903 representative 16S rRNA gene sequences of human gut microbiota taken from genbank, revealed the dominant members the bacteria belong to three major phyla; the Firmicutes, Bacteroidetes and Proteobacteria (Backhed *et al.*, 2005).

With the aim to generate a full picture of the taxonomic diversity and functions of the microbes associated with the human body, the first NIH human microbiome project was launched in 2007 (Turnbaugh *et al.*, 2007). The HMP successfully produced about 2.3 tera bp 16S rRNA sequence data from samples collected from 15 body sites of 300 US individuals (Turnbaugh *et al.*, 2007). Furthermore, the advent of high throughput DNA sequencing technologies and advances in bioinformatics analysis tools has enabled to decipher the microbial diversity, composition and function of the human gut microbiota and microbiome. The metagenomics of human Intestinal tract (MetaHIT) project, via metagenome sequencing of faecal samples collected from 124 subjects (including healthy, overweight, obese and Inflammatory bowel disease (IBD) patients) has generated the first gene catalogue of the human gut microbiome (Qin *et al.*, 2010). This catalogue constitutes 3.3 million microbial genes, of which 99% belong to bacteria, the rest to archaea, while eukaryotes constitute 0.1% of the catalogue. The common bacterial core shared by each individual has also been revealed: 57 species were found in > 90% of the individuals even though there was inter individual variation, and the predominant members were the *Firmicutes* and *Bacteroidetes*. In line with this, the human microbiome project 2 (HMP2) using 16S rRNA and metagenome analysis of samples collected from 242 healthy individuals and different body sites (18 for women and 15 for men), has revealed a 4,768,112 gene catalogue of the gut microbiota, with more diversity and composition (The Human Microbiome Project *et al.*, 2012). The HMP project also indicated that there exists substantial inter personal and intra personal variation, such as different body sites having a characteristic microbial diversity and composition, even between healthy individuals (The Human Microbiome Project *et al.*, 2012). A report based on an expanded dataset of human microbiome project II (HMP II) consisting of whole metagenome sequence of new 1,635 samples making up a total of 2,355 metagenomes has also revealed that, there exists body wide strain diversity and composition and some stable subspecies unique to individuals were also discovered (Lloyd-Price *et al.*, 2017). Moreover, due to the

advances of new methodologies, new species-level taxa were revealed including eukaryotes, archaea, viruses and 54 new bacterial species in this metagenome compared to Human microbiome metagenome project I (HMPI) metagenome dataset (Lloyd-Price *et al.*, 2017). In addition, metabolic pathways which enhance microbiome colonisation to the human gut were also discovered.

1.5.1. Diet and the microbiota

The type of diet an individual consumes shapes the diversity and composition of the gut microbiota. A classical study on how diet shapes the microbiota composition was carried out by De Filippo and colleagues (De Filippo *et al.*, 2010) where they compared the microbiota of children from Burkina Faso (BF) and Italy (EU). The Burkina Faso children had a diet rich in starch, fibre and plant polysaccharides; such as cereals (millet, sorghum), legumes and vegetables while the Italian children's diet was mainly composed of Mediterranean and western foods. At the phylum level 94.2% sequences of both BF and EU belong to Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria. However, the proportion of the phyla in both samples were distinct. The microbiota composition of children from BF were enriched with Actinobacteria and Bacteroidetes, while the EU were enriched with Proteobacteria and Firmicutes. The BF constituted high abundance of the Bacteroidetes and with a depletion of Firmicutes, with some unique bacteria composition enriched with the genera *Prevotella* and *Xylanibacter* rich with sets of enzymes which hydrolyse dietary fibre. On the contrary the Firmicutes were twice as abundant in Italian children's. In addition, high microbial diversity and composition accompanied by high production of faecal SCFA was detected in the BF cohort, a trait related with high dietary fibre intake. A continuation of these studies has indicated that Burkina Faso children from rural villages and small towns constitute a gut microbiota dominated by Bacteroidetes (69 and 48% respectively), while children from the Capital city (Oguadugouu) and the Italian children harboured a relatively low abundance of gut Bacteroidetes (33% and 26% respectively) and a higher abundance of the Firmicutes (58 and 60% respectively) (De Filippo *et al.*, 2017). The researchers deduced that industrialisation and socioeconomic status has changed the way of life and diet in children from the cities of the Burkina Faso which ultimately contributed in the shift of microbiota

composition from rural type to western type. Similarly, comparative studies of faecal microbiota composition of rural Malawian and Venezuelan cohort with US cohort revealed that, there was similarity in microbiota composition within the rural cohort as compared to the US cohort. In addition, the rural cohort showed high production of SCFA, as compared to the western microbiota (US group) (Yatsunenko *et al.*, 2012b). Thus, the major factor for the distinct microbiota composition and diversity between the two cohorts was the type of diet.

A particular study linked to diet and microbiota composition was obtained in comparative studies of the Hadza hunter gatherer tribes of Tanzania with western microbiota. Faecal samples of 27 Hadza and 16 Italian adults showed a very distinct microbiota composition between the two groups (Schnorr *et al.*, 2014). The Hadza microbiota was dominated by Firmicutes (72%) and Bacteroidetes (17%). The remaining phyla consisted of Proteobacteria (6%), Spirochaetes (3%), and 2% of phylum level operational taxonomic units (OTUs) remained unclassified. At family level the Hadza microbiota was dominated by Ruminococcaceae (34%), Lachnospiraceae (10%), Prevotellaceae (6%), Clostridiales Incertae Sedis XIV (3%), Succinivibrionaceae (3%), Spirochetaceae (2%) and Eubacteriaceae (2%). A large number of taxa belonging to the Bacteroidetes, Clostridiales, Bacteroidales and Lachnospiraceae, were unassigned at the family and genus level. Although the Italian gut microbiota was also dominated by Firmicutes and Bacteroidetes, the Hadza microbiome has lower abundance of Firmicutes and higher abundance of Bacteroidetes compared to the Italian gut microbiome. Moreover, at the genus level a significant difference between the two groups were observed such as depletion of the *Bifidobacterium* in Hadza and high abundance of the *Treponema* genus specialised in cellulose and xylan degradation (Schnorr *et al.*, 2014). These differences were as a result of lifestyle and particularly attributed to the type of diet they consumed, the diet of the Hadza community consisted of wild foods such as meat, honey, berries and tubers, while the Mediterranean diet of the Italians consisted of plant based foods, fruit, pasta, bread, olive oil and low to moderate amount of dairy, poultry, fish and red meat. In summary, in addition to genetics, environmental factors, culture, diet is the most significant factor that contributes to the microbiota composition and diversity in the human gut.

1.6. Glycan foraging by the gut microbiota

The 100 trillion microbial cells in the human gut encode an enormous amount of genes which are absent in the human genome. The most pronounced difference is that the microbiota encode abundant carbohydrate active enzymes (CAZymes), such as glycosyl transferases which assemble the glycans or glycoside hydrolases, polysaccharide lyases and carbohydrate esterases which break down various glycans (Lombard *et al.*, 2014a). The human genome encodes only for 97 Glycoside hydrolase (GH) enzymes, and no polysaccharide lyases. Out of the 97 human GH enzymes, 8 enzymes belong to GH1, GH 13, GH 31 and GH 37 families directly participating in the hydrolysis of carbohydrates, while another 9 belong to the GH1, GH 9, GH 18, GH 31 and GH 35 families which have a possibility to be digestive. However the remaining 80 enzymes have other roles, such as regulatory function, tissue development and pathogen exclusion (Kaoutari *et al.*, 2013). Given that the human genome encodes less than 20 enzymes for breakdown of complex carbohydrates (Cantarel *et al.*, 2012), its apparent that the breakdown of complex glycans such as plant-based polysaccharides are only possible through the CAZymes of the gut microbiome in the human gut, mainly the distal gut (Backhed *et al.*, 2005). Besides, some members of the microbiota are adapted to utilise host-derived mucin glycoproteins and glycans derived from shed epithelial cells in the absence of key dietary complex glycans or nutrients (Kaoutari *et al.*, 2013). The ability to utilise complex dietary glycans and host-derived mucin glycans is crucial to the survival and proliferation of the key members of the microbiota, particularly when dietary glycans are absent (Sonnenburg *et al.*, 2005). However, the degradation of the mucin glycans could compromise the mucus layer and would expose the epithelial cells. This direct contact of microbiota to epithelial cells could elicit the onset of pathogenicity (McGuckin *et al.*, 2011a). For complete hydrolysis of mucin glycans, sulfatases, proteases and CAZymes are required (Tailford *et al.*, 2015). From the CAZymes, particular GH enzymes include the sialidases (GH33), α -fucosidases (GH 29 and GH 95), exo and endo N-acetyl-glucosaminidases (GH 84 and GH 85), β -galactosidases (GH 1, GH 2, GH 35, GH 42, GH 59 and GH 147), α -N-acetylglucosaminidases (GH89), and α -N-acetylgalactosaminidases (GH101, GH129) (Lombard *et al.*, 2014a).

Comparative genomic analysis of 397 human gut microbes for identification of CAZymes active in mucin glycan utilisation revealed that 86% of the analysed genomes encode for genes which cleave mucin glycans while 89% possessed genes related with catabolism of monosaccharides (Ravcheev and Thiele 2017). Sialidase (GH33) encoding genes were detected in 112 of the analysed genomes and in all phyla except Fusobacteria. α -fucosidase genes were detected in 131 genomes that belong to four phyla, Actinobacteria, Bacteroidetes, Firmicutes, and Verrucomicrobia. Glycoside hydrolases which release N-acetyl glucosamine (GlcNAc) from mucin glycans (hexosaminidases and α -N-acetylglucosaminidases) were detected in 257 genomes, with the maximum number detected in *Bacteriodes sp* (23 genes). Similarly α -galactosidases and β -galactosidases which release Gal from mucin glycans were found in 310 genomes (Ravcheev and Thiele 2017).

The *Bacteriodes* are the dominant member of the human gut microbiota. The genomes of this group of bacteria have been extensively studied throughout the past decades. Genome wide analysis has shown that these bacteria harbour genes that encode saccharolytic enzymes, and are clustered into polysaccharide utilisation loci (PULs) and the starch utilisation system (*sus*). PULs comprise a single genomic loci, harbouring genes which encode proteins to perform several functions, such as polysaccharide binding proteins, enzymes which undertake initial cleavage of polysaccharides to oligosaccharides, membrane transport proteins for importing oligosaccharides into the periplasmic space and enzymes which complete the hydrolysis of the oligosaccharides to monosaccharides, and regulatory proteins (Terrapon *et al.*, 2018). More than 100 PULs have been identified in some *Bacteriodes* species, constituting about 20% of their genome, clearly implicating them for the bacterium's success in human gut environment as primary glycan degraders. PULs from genomes of *Bacteriodes* can be found in the CAZY database (http://www.cazy.org/PULDB/index.php?sp_name=Bacteriodes&sp_ncbi) (Terrapon *et al.*, 2018). *In vivo* studies of humanised gnotobiotic mice with the intestinal symbiont *Bacteriodes thetaiotaomicron* showed that this bacteria up-regulates the expression of several CAZymes, such as glycoside hydrolase enzymes and polysaccharide lyases, to harvest dietary carbohydrates and host derived mucin glycans, depending on nutrient availability (Sonnenburg *et al.*, 2005). For instance in the absence of low dietary

carbohydrates, *B. thetaiotaomicron* isolated from cecum upregulates expression of enzymes involved in host mucin glycan utilisation such as sialidases and α -fucosidases.

In addition to *Bacteriodes*, *Bifidobacterium* encode CAZymes which enhance their foraging capability of glycans, such as host derived glycans. The genome analysis of 47 species of *Bifidobacterium* indicated that 14% of their genome encodes for genes related to carbohydrate metabolism, which provided them a competitive advantage in the human gut niche (Milani *et al.*, 2015a). In line with this the pan genome of the *Bifidobacterium* genus encodes for 57 glycoside hydrolases (GH), 13 glycosyl transferases (GT), 7 Carbohydrate esterases (CE), but no polysaccharide lyases (PL) were detected. The most abundant GH family enzyme in these *Bifidobacteria* was the GH 13 family, which accounts for 24% of their genome. This family of enzymes acts on a wide arrays of complex glycans such as plant based polysaccharides and starch (amylose, amylopectins and maltodextrins) (Milani *et al.*, 2015a). In line with this, the glycobioime of *Bifidobacterium* harbours GH enzymes that are specialised in the breakdown of host derived O-linked and N-linked glycans and HMOs. These include the exosialidases (GH33), α -fucosidases (GH 29 and GH 95), hexosaminidases (GH 20), lacto-N-biosidases (GH112), α -mannosidases (GH38 and GH 125) and α -N-acetylgalactosaminidases (GH101 and GH129) (Milani *et al.*, 2015a).

Although members of the gut microbiota which express GH enzymes which act on host mucin glycans gain a selective advantage to proliferate in the gut even in the absence of dietary fibres, the utilisation of host derived mucin glycans come with a severe health consequence to the host. Studies in mice have shown that MUC2 -deficient mice lack a mucus layer barrier and as a result the microbiota have access to the epithelial cells, which triggers inflammation and mucosal immunity which ultimately induces colitis. Mice with high inflammatory response have high abundance of Proteobacteria (Jakobsson *et al.*, 2015). Other murine models which have a defective mucus layer, are due to either defects in ion transport channel proteins resulting in poor formation of the outer mucus layer, or the outer mucus layer is composed of short length mucin glycans, which bacteria easily hydrolyse (Johansson *et al.*, 2014). *In vivo* experiments in mice were carried out by administering dextran sodium sulfate (DSS) (most commonly used colon inflammation

agent in rodents) in drinking water. Then the colon was removed and tissues were fixed and immunostained for MUC2 and bacterial 16 rRNA was stained using fluorescent in situ hybridisation (FISH). Bacteria were detected after 4 h near the inner mucus layer and after 24 h the inner mucus layer has disappeared which caused complete accessibility to the epithelial cells (Johansson *et al.*, 2010). These comprehensive studies demonstrated the importance of the mucus layer in protecting the accessibility of the epithelial cells by the resident microbiota and ultimately preventing the incidence of inflammatory response, colitis and colon cancer.

Most dietary polysaccharides are plant-based polysaccharides which are derived from fruits, vegetables, and resistant starch. The gut microbiota use a repertoire of CAZymes to breakdown the dietary fibres, through anaerobic fermentation to produce end products SCFA (acetate, butyrate and propionate), which are readily absorbed (Wong *et al.*, 2006). Butyrate serve as an energy source for colonocytes. Propionate is absorbed in the liver, while acetate enters the peripheral circulatory system for absorption in the peripheral tissues. After their absorption, SCFA take part in numerous physiological functions.

Butyrate produced after fermentation is taken up by the epithelial cells of the colon (colonocytes) via the monocarboxylate transporter (Tan *et al.*, 2014). Then it serves as main energy source (~ 70%) for colonocytes, through β -oxidation in the mitochondria (Koh *et al.*, 2016) thereby enhancing normal proliferation of colonocytes. On the other hand butyrate decreases the proliferation of cancerous cells either by acting as anti-tumorigenic agent or through induction of apoptosis of cancerous cells (Fung *et al.*, 2012). Butyrate, by acting as histone deacetylase inhibitor, (HDAC) regulates expression of cyclin-dependent kinase inhibitors cdk1a and cdk1c which subsequently inhibits proliferation of colorectal cancerous cells (Donohoe *et al.*, 2012). Recent studies have also shown that through inhibition of the pyruvate kinase M2 (PKM2) pathway and metabolic reprogramming, butyrate suppresses tumour proliferation in colorectal cancerous cells (Li *et al.*, 2018).

Acetate has numerous physiological functions. It is absorbed in the peripheral circulatory system and cross the blood brain barrier and serve as energy source for the glial cells (Wyss *et al.*, 2011). A study in mice has shown that acetate suppresses appetite via the

gamma amino butyric acid (GABA) neurological cycle in the hypothalamus, thereby prevent diet-induced weight gain (Frost *et al.*, 2014). In contrast to this, rodents fed with a high fat diet produced high levels of acetate compared to high fibre fed rodents. The high level of acetate stimulated the parasympathetic nervous system and enhanced secretion of the hunger hormone, ghrelin which stimulates food intake and this, subsequently conduces obesity (Perry *et al.*, 2016). In addition, this causes hypertriglyceridemia, adiposity in liver and skeletal muscles and insulin resistance in liver and muscles (Perry *et al.*, 2016). Although human studies are necessary to avoid inconsistency, these are important findings as consumption of dietary fibres could help in alleviating the growing global problem of obesity.

Propionate acts as a precursor for intestinal gluconeogenesis and together with butyrate, it induces the expression of enzymes of the gluconeogenesis pathway (Schroeder and Bäckhed 2016). On the other hand, propionate has been indicated in decreasing cholesterol biosynthesis. Mice fed with a high fat and high propionate diet had a reduced hepatic gene expression of lipogenic enzymes resulting in low concentration of hepatic triglycerides (up to 40% reduction) compared to mice which were fed a high fat diet without SCFA (Weitkunat *et al.*, 2016). This further increased insulin sensitivity in fat induced obesity.

Although some inconsistent data have been reported in terms of how acetate and propionate are involved in regulation of satiety, lipogenesis, and gluconeogenesis, overall the gut microbiota mediated synthesis of SCFA could enhance a healthy homeostasis development of the host.

1.7. Dysbiosis and its implications in disease

The vast human gut microbiota residing in the human gut has been considered as an inexhaustible source of protein coding genes which enable the human host to carry out metabolic functions such as degradation of dietary fibres and production of SCFA, synthesis of essential vitamins and amino acids, degradation of xenobiotics, regulation of immune system and prevention of pathogenic bacteria colonisation, to ultimately maintain homeostasis and health (Schuijt *et al.*, 2016). The microbiota composition can be altered by exogenous factors such as uptake of antibiotics (Willing *et al.*, 2011), diet (Albenberg

and Wu 2014), gastrointestinal pathogens (Gill *et al.*, 2011) and endogenous factors such as immune dysregulation or deficiencies (Willing *et al.*, 2010b). In cases where there is a dramatic shift of the microbiota diversity and composition called “dysbiosis”, the pathogens start to thrive and a huge number of gut microbiota linked diseases will result such as inflammatory bowel disease, obesity, type 2 diabetes, cardiovascular diseases (CDV), autism, colorectal cancer, and autoimmune disorders (Battson *et al.*, 2018; DeGruttola *et al.*, 2016; Hughes *et al.*, 2018; Lloyd-Price *et al.*, 2019). Herein, the major metabolic and auto immune disorders that resulted from a dysbiosis of the gut microbiota will be briefly discussed.

The perturbation of human gut microbiota have been associated in the development of inflammatory bowel disease (Lloyd-Price *et al.*, 2019); a chronic inflammatory complication of the GIT which includes Crohn’s disease (CD) and Ulcerative Colitis (UC) (Oligschlaeger *et al.*, 2019). IBD is a multifactorial chronic illness which may arise from aberrant immune response of the host to the commensal microbiota, which ultimately results in weak mucosal and intestinal barriers (Schirmer *et al.*, 2019). Some microbial metabolites such as SCFA, specifically butyrate, could enhance regulatory T cell (T_{reg}) development (Furusawa *et al.*, 2013) which has an immunomodulatory effect on the host cells and promote mucosal secretion from goblet cells to provide barriers of the epithelial cells (Chung *et al.*, 2012). In a multi-omics study using metagenomic analysis and metabolomic profiling of 220 stool samples cohorts of CD, UC, and non-IBD controls, it was indicated that there was significant differences in the microbial diversity and metabolome of IBD and non-IBD patients. Particularly, high abundance of sphingolipids, bile acids and depletion of triacylglycerols and tetrapyrroles were detected in IBD patients (Franzosa *et al.*, 2019). Various studies showed that IBD patients have depleted microbial diversity compared to non-IBD controls. The microbiota of IBD patients showed lower abundance of Firmicutes particularly *Ruminococcaceae* family, including *Faecalibacterium*, *Ruminococcaceae incertae sedis*, but high abundance of *Enterobacteriaceae* such as *E. coli* were detected (Willing *et al.*, 2010a). To fully understand the mechanism of host microbiome interaction for IBD development, the Inflammatory Bowel Disease Multi-omics Database project (<http://ibdmdb.org>) (Lloyd-Price *et al.*, 2019) has been established as part of the Human Microbiome Project 2

(HMP2). This project followed 132 individuals from five clinical centres for a course of one year (Lloyd-Price *et al.*, 2019). Stool, mucosal and blood samples were analysed. Significant perturbation of microbial composition was observed in the IBD patients within a week time, while no such significant shifts were observed in the controls. The study identified more than 2,900 host and microbial interactions which are related to the IBD. The associated microbial taxa in these interactions comprise *Faecalibacterium*, *Subdoligranulum*, *Roseburia*, *Alistipes*, and *Escherichia* and metabolites such as octanoyl carnitine, SCFA and various lipids. Regulators of interleukins are among the key contributors of these metabolic networks (Lloyd-Price *et al.*, 2019).

Obesity and type 2 diabetes are two associated metabolic disorders and have been a global problem in the past two or more decades. Major contributing factors are lack of physical exercise and unhealthy diet habits, such as an increased food intake and rich in fat diets greater than the amount of energy needed by the person (Ley *et al.*, 2005). However, the gut microbiota and its metabolites have been implicated to affect the host metabolism which triggers obesity and insulin resistance. Germ free mice colonised by microbiota isolated from caecum of a conventional grown mice showed a metabolic change and develop high adiposity linked to the microbiome colonisation despite reduction of food consumption (Backhed *et al.*, 2004). In line with this, the microbial ecology of genetically obese (*ob/ob*) and lean (+/+) mice was distinct: the former was characterised by 50 % low abundance of Bacteroidetes and high abundance of Firmicutes as compared to the lean mice (Turnbaugh *et al.*, 2006). The microbiota in *ob/ob* mice were efficient in extracting more calories than the microbiota of (+/+) mice which ultimately result in enhanced adiposity and obesity. Microbiota transplantation from *ob/ob* mice to germ free mice also resulted in obesity (Turnbaugh *et al.*, 2006). These observation were also observed in human studies, as the obese human subjects have low abundance of Bacteroidetes and high Firmicutes, which indicates that the microbial ecology plays significant roles in energy harvest and adiposity (Ley *et al.*, 2006).

Similar to obesity, the incidence of insulin resistance and type 2 diabetes mellitus (T2D) has been associated with the gut microbiota. An animal study conducted by Cani and colleagues (Cani *et al.*, 2007) indicated that metabolic endotoxemia (diagnosed by high

plasma concentration of gram negative lipopolysaccharides (LPS)) that result either from high fat diet or infusion caused insulin resistance and type 2 diabetes. The binding of LPS to the multifunctional phosphatidyl inositol receptor (mCD 14) and toll like receptors (TLR4) in immune cells such as macrophages and neutrophils triggers the secretion of pro inflammatory cytokines (Sweet and Hume 1996), which ultimately affects the action of insulin. mCD 14 mutant mice have high insulin sensitivity. These findings indicate that the LPS could affect insulin resistance through mCD14 based signal transduction (Cani *et al.*, 2007).

In parallel with this, the Type 2 diabetes mellitus (T2D) microbial gene catalogue was developed from faecal metagenome DNA sequencing of 145 Chinese individuals (71 T2D patients and 74 healthy controls) (Qin *et al.*, 2012) to generate an updated catalogue of 4,267,985 predicted genes. The T2D patients showed a moderate degree of gut microbiota dysbiosis compared to healthy individuals, such as reduction in abundance of butyrate producing bacteria (Qin *et al.*, 2012). This metagenome wide association study (MGWAS) revealed 60,000 T2D associated markers.

Other important implications of dysbiosis of the gut microbiome is cardiovascular disease (CDV) which is one of the leading cause of death in the world. The advent of meta-omic technologies has enabled identification of novel metabolic pathways of the gut microbiome linked to the development of CDV. One of the microbial metabolites that is linked to CDV is trimethylamine (TMA). TMA is an intermediate of dietary lipid phosphatidyl choline degradation by the gut microbiota and is hepatically converted into trimethylamine-N-oxide (TMAO) using flavin monooxygenase (FMOs) result in atherosclerosis (Wang *et al.*, 2011b). Supplementation of the phosphatidyl choline metabolites such as choline, TMAO, and betaine to germ free mice enhanced upregulation of multiple macrophage scavenger receptors linked to atherosclerosis development, while suppression of the microflora in atherosclerosis-prone mice inhibited dietary choline related atherosclerosis (Wang *et al.*, 2011b). Besides TMAO, metabolic products of amino acid catabolism by the gut microbiota and host liver produces uremic toxins such as indoxyl sulfate, indoxyl glucuronide and indoleacetic acid (Devlin *et al.*, 2016) which are linked with chronic kidney failure (CDK) and associated heart failure

(Lin *et al.*, 2014). In addition the gut microbiota affect cardiac health via secondary bile acids such as deoxycholic acid and lithocholic acid binding to G protein-coupled membrane receptor 5 (TGR5) (Duboc *et al.*, 2014).

The dysbiosis of the human gut microbiota has been associated with non-alcoholic fatty liver. A microbial gene catalogue of cirrhosis was generated from 98 patients and 83 healthy Chinese subjects (Qin *et al.*, 2014) generating 2,688,468 non-redundant open reading frames (ORFs). Merging of non-redundant ORFs from MetaHIT (Qin *et al.*, 2010) and T2D (Qin *et al.*, 2012) provided 5,382,817 genes. Phylogenetic analysis of the liver cirrhosis and the healthy individuals showed that the former constituted fewer Bacteroidetes and higher abundance of Proteobacteria and Fusobacteria than the latter, although the faecal microbiome of both group were dominated by Firmicutes and Bacteroidetes (Qin *et al.*, 2014).

Another notable effect of the microbiota is their positive effect in immune system development. This was first observed in germ free mice, where the intestinal immune system was underdeveloped compared to conventionally raised mice. Germ free mice exhibited a reduced number of inter-epithelial lymphocytes, reduced number of goblet cells with reduced mucus thickness, altered crypt structure, and reduced size and number of Peyer's patches (Deplancke and Gaskins 2001; Macpherson and McCoy 2015; Round and Mazmanian 2009). Besides GF mice were characterised by immature gut associated lymphoid tissues (GALT), lower number of antimicrobial molecules, and reduced amount of immunoglobulin A (IgA) (Chinen and Rudensky 2012). Most of these immune related cells and structures characteristics were reversed when the germ free mice were deliberately colonised by diverse microbiota (Bouskra *et al.*, 2008; Chinen and Rudensky 2012).

New studies have showed that perturbation in commensal gut microbiota can alter neural pathways and central nervous system (CNS) which can ultimately lead to anxiety, depression and autism spectrum disorders (Foster *et al.*, 2017). Studies which revealed a reduction of high stress reactivity by Germ free mice, after colonisation with gut microbiota enlightens the association of the microbiota and brain studies (Neufeld *et al.*, 2011).

1.8. Probiotics, prebiotics and synbiotics

In western societies, the prevalence of chronic illnesses, such as inflammatory diseases, colon cancer, and autoimmune disorders, which are all linked with gut microbiota dysbiosis, has increased (Sonnenburg and Sonnenburg 2019). Although antibiotic therapy has been the first line therapy in the treatment of many infectious diseases, in the case of chronic illnesses which have multifactorial causes, antibiotics have limited success in delivering the expected outcome. Besides, the inevitable antibiotic resistance exacerbates their use for such complex disorders (Boolchandani *et al.*, 2019). Thus the use of exogenous supplementations of dietary bacterial (probiotic) and food ingredients (prebiotics) could aid the indigenous gut microbiota (Rastall *et al.*, 2005) and prevent the occurrence of such chronic diseases.

The idea of replacing harmful microbes with useful microbes was first proposed by the Russian scientist Eli Metchnikoff (Metchnikoff 1907) and the French scientist Henry Tissier (Tissier 1907) independently. Metchnikoff suggested bacteria present in fermented milk products would prevent the bacterial fermentation in the intestine and may promote health to the consumers, he claimed the longevity of the Bulgarian peasants could be attributed to the consumption of fermented milk products (Metchnikoff 1907). Tissier observed the absence of the Y (“bifid”) shaped bacteria in infants with diarrhoea and proposed supplementation of this bacteria could alleviate the diarrheal problem (Tissier 1907).

The word probiotic was first used by (Lilly and Stillwell 1965) to describe the product of one protozoan promoting the other. later, it was modified by Fuller, to “A live microbial feed supplement which beneficially affects the host animal by improving its intestinal microbial balance” (Fuller 1989). The most recent definition of probiotics is ““Live microorganisms which when administered in adequate amounts confer a health benefit on the host” (FAO/WHO 2001). To be regarded as a probiotic a microorganism should fulfil the following criteria: be isolated from human GIT for safety, have a GRAS status by the FDA, be able to be produced in large scale in viable form and stay viable in the carrier and shelf lives, be resistant to gastric acid secretions in the GIT and bile acids, be able to bind to the intestinal mucus layer for colonisation and persistence, be able to produce

antimicrobial compounds to inhibit pathogens and to restore normal health microbiota, be safe in food or clinical use even for immunocompromised patients and finally their efficacy and safety must have been proven in randomized, double blind, placebo control trials in humans (Dunne *et al.*, 2001; Tuomola *et al.*, 2001).

The most widely used probiotics are developed from a few species of microorganisms belonging to the genera of *Lactobaccillus* (*L. casei*, *L. acidophilus*, *L. rhamnosus*, *L. johnsonii*, *L. reuteri*) and *Bifidobacterium* (*B. longum*, *B. breve* and *B. bifidum*) (Gibson 2004). The mechanism of action of probiotics are species and strain specific (Hill *et al.*, 2014). The exact mode of action is not known but could comprise production of antimicrobial compounds/peptides that kill pathogenic bacteria, blockage of adhesion of pathogenic bacteria on the epithelial cells and immunomodulation of host cells (Ng *et al.*, 2009).

Current progress of gut microbiome research has broadened the concept of probiotics, as previously uncharacterised microbial species have been studied for their potential probiotic properties, which are referred as next generation probiotics (O'Toole *et al.*, 2017). Beneficial commensal bacteria such as *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* are considered as next generation probiotics, because of their natural adaptation to the colon (Bermúdez-Humarán and Langella 2018). Animal model studies using these two commensals showed they help prevent obesity and inflammatory bowel diseases. Advances in human gut microbiome studies have revealed that butyrate-producing bacteria such as *Roseburia spp* and *Eubacterium hallii* help in prevention of dysbiosis related inflammation and immune modulation (Hill *et al.*, 2014). Faecal Microbiome Transplant (FMT) alleviated *C. difficile* infection (van Nood *et al.*, 2013) and associated ulcerative colitis. However, FMT is not regarded as a probiotic, unless the type of species and strains for FMT application are clearly defined (Hill *et al.*, 2014). Advances in technology has revolutionised the use of genetic modified derivatives of natural probiotic strains to produce “live bio-therapeutics” for alleviating targeted genetic defects. Recently, a genetically engineered phenylalanine degrading derivative of *E. coli* Nissle, SYN1618 strain was developed following FDA regulations to treat Phenylketonuria (PKU), a genetic defect where the patients lack the key enzyme,

phenylalanine hydroxylase (PAH) and are unable to metabolise phenylalanine (Isabella *et al.*, 2018).

Prebiotics are non-digestible ingredients that selectively promote the growth and activity of certain groups of microorganisms in the colon that positively affect the consumer (Gibson and Roberfroid 1995). The definition of prebiotics has been modified, throughout the years by different scientists, however the latest International Scientific Association for Probiotics and Prebiotics (ISAPP) panel has provided a consensus definition of prebiotics as “a substrate that is selectively utilized by host microorganisms conferring a health benefit” (Gibson and Hutkins 2017). The desirable criteria of a food ingredient to be regarded as a prebiotic comprise: resistant to the host digestive enzymes in the upper digestive tract and not absorbed in the GIT, fermented by the resident intestinal microbiota and to selectively stimulate the growth and proliferation of groups of bacteria associated with health and wellbeing (Gibson *et al.*, 2004). Prebiotics could be administered orally or directly to microbial colonised body sites such as skin and vagina. The ISAPP panel also provided the scope of the health impact of prebiotics, and it comprises promoting GIT health (for instance, via exclusion of pathogens and immune modulation), cardio metabolism (reduction of level of blood triglycerides and insulin resistance), and mental health (for instance metabolites that alter brain function). These health benefits of prebiotics should be confirmed in animal models through microbiota modulation (Gibson and Hutkins 2017). The selective fermentation of prebiotics by some group of microorganisms mainly *Lactobacilli* and *Bifidobacterium* has been well documented. Metagenome studies revealed that supplementation of prebiotics also enhances the growth of other members of the microbiota such as butyrate producing *Faecalibacterium prausnitzii* (Lopez-Siles *et al.*, 2017). Butyrate and the other SCFA have been clearly indicated in promoting normal health to the host. Thus, any ingredient which could be metabolised by some microorganisms in the gut but not by all and not by pathogens and enhance production of SCFA and ultimately promote overall health of the host can be regarded as prebiotic.

The past decades have resulted in the identification of carbohydrate prebiotics (listed below) with substantial experimental proof to enhance the selective growth of probiotic

bacteria, such as *Bifidobacteria* and *Lactobacilli*, and the inhibition of several enteric pathogens (Gibson *et al.*, 2004).

- i. Inulin constitutes an oligomer of either α -D-glucopyranosyl-[- β -D-fructofuranosyl]_{n-1}- β -D-fructofuranoside or β -D-fructopyranosyl-[β -D-fructofuranosyl]_{n-1}- β -D-fructofuranoside with β (2 \leftrightarrow 1) linkage between fructosyl - glucose always as in sucrose, but β (1 \leftarrow 2) linkage between fructosyl - fructose. Chicory inulin constitutes oligosaccharides of up to sixty degree of polymerisation (DP), and enzymatic hydrolysis of inulin using endo-inulinase produces fructooligosaccharides (FOS) from three to seven DP. The FOS can also be produced by transfructosylation activity of β -fructosidase.
- ii. Xylooligosaccharides (XOS) are produced via enzymatic hydrolysis of xylan from maize cobs.
- iii. Glucooligosaccharides can be produced by the enzymatic action of dextran sucrose from sucrose in the presence of maltose
- iv. Lactose-based prebiotics:
 - a) Galactooligosaccharides (GOS) are produced by the enzymatic transgalactosylation of lactose using β -galactosidases. These mixtures of oligosaccharides constitute tri to Penta saccharides with β (1 \rightarrow 6), β (1 \rightarrow 3) and β (1 \rightarrow 4) linkages.
 - b) Lactulose is a disaccharide produced via isomerisation of lactose and constitutes galactosyl β -(1 \rightarrow 4) fructose.
 - c) Lactosucrose is produced from lactose and sucrose by the enzymatic β -fructofuranosidase.
 - d) Lactitol, lactobionic acid, tagatose and gluconic acid are produced by chemical reaction such as oxidation or reduction from lactose.
- v. Soybean oligosaccharides constitute oligosaccharides with α galactosyl sucrose derivatives such as raffinose (tri saccharide,) and stachyose(a tetra saccharide; β -D-fructofuranosyl α -D-galactopyranosyl-(1 \rightarrow 6)- α -D-galactopyranosyl-(1 \rightarrow 6)- α -D-glucofuranoside.)

The selective fermentation of prebiotics in the human gut, mainly by *Bifidobacterium* and *Lactobacillus* species, has great health benefits (Rastall and Gibson 2015). Prebiotics and probiotics have potential in prevention of inflammatory bowel disease (Scaldaferri *et al.*, 2013). UC patients treated with probiotic VSL#3 mixture (*Bifidobacterium breve*, *Bifidobacterium longum*, *Bifidobacterium infantis*, *Lactobacillus acidophilus*, *Lactobacillus plantarum*, *Lactobacillus paracasei*, *Lactobacillus bulgaricus*, and *Streptococcus thermophilus*) showed remission (Bibiloni *et al.*, 2005). By promoting selective proliferation of commensals in the gut prebiotics could enhance competitive exclusion of pathogens via inhibition of adhesion or bacteriocin production (Aureli *et al.*, 2011). Moreover, fermentation of prebiotics such as GOS by bacteria to produce SCFA and lactic acid reduces the local pH in the GIT and pathogens are unable to thrive (Nath *et al.*, 2017). Although direct supplementation of FOS to Crohn's disease (CD) patients were not effective in alleviating symptoms, it induced the activity of immunoregulatory Dendritic Cells (Benjamin *et al.*, 2011). Lactulose has been implicated in decreasing hepatic encephalopathy, a condition caused by liver damage which ultimately causes high accumulation of ammonia in the blood stream (Weber 1996). The mechanism of its action is controversial but it's presumed the breakdown of lactulose by gut microbiota reduces absorption of ammonia in the blood stream and portal blood by decreasing pH (Clausen and Mortensen 1997). As a prebiotic lactulose selectively enhances growth of endogenous commensal microbiota such as *Lactobacilli* which results in reduction of urease producing bacteria (Riggio *et al.*, 1990). In line with this, it was reported that lactulose ingestion by volunteer healthy humans has increased faecal bifidobacterial counts (Bouhnik *et al.*, 2004). In addition, prebiotics could decrease the incidence of colon cancer by reducing the accumulation of gut microbiota metabolites such as secondary bile acid, N-nitroso compounds (NOCs), ammonia, polyamines etc. which have carcinogenic effects (Louis *et al.*, 2014). In addition, fermentation of prebiotics by bacteria results in the production of SCFA, such as butyrate, a nutrient to colonocytes and prevents colon cancer (Li *et al.*, 2018). The prebiotics and their effect on species such as *Bifidobacterium* and *Lactobacilli* have been implicated in modulation of immune system cells such as natural killer cells, macrophages, dendritic cells, T-lymphocytes etc. Prebiotics are implicated in regulation of energy intake, for instance, healthy adults fed with oligo fructose increased the

expression of gut hormones such as glucagon-like peptide 1 (GLP-1) and peptide YY3-36 (PYY), via SCFA based activation (Verhoef *et al.*, 2011).

Prebiotics and health-promoting bacterial species go hand in hand, and their positive health effect on the consumer relies on the presence of both. Abundant selective nutrient source in the form of prebiotics helps ensure colonisation of beneficial bacteria and their successful proliferation in the gut. Synbiotics are defined as “mixtures of probiotics and prebiotics that beneficially affect the host by improving the survival and implantation of live microbial dietary supplements in the gastrointestinal tract, by selectively stimulating the growth and/or by activating the metabolism of one or a limited number of health-promoting bacteria, thus improving host welfare” (Gibson and Roberfroid 1995). Synbiotic supplementation (for example *Bifidobacterium* strains and fructooligosaccharides) could ensure the viability of the probiotic strains in the gut as the specific substrate will be readily available for fermentation, thus provide the intended health benefits (Collins and Gibson 1999).

1.9 . Metagenomics

The vast majority of microorganisms are uncultured by current cultivation techniques. Culture-independent approaches circumvent this limitation. It has enabled the disclosure of the microbial diversity, ecological roles and prolific genetic resources of uncultivated microorganisms (Riesenfeld *et al.*, 2004). The term metagenome was first coined by Handelsman and his colleagues (Jo Handelsman 1998), to describe the collective genomes of microbial flora from a given environment. Thus, metagenomics describes the analysis of metagenome DNA.

Metagenomes are screened either based on activity (function) or sequence. In the function-based screening the metagenome clones which encode a particular gene would be identified by the desired function. One functional metagenomics method is cloning of fragments of metagenome DNA into vectors in a heterologous host, to create a metagenomics library for identification of clones that encode natural products. Though it's a straightforward approach, function-based screening is hindered by heterologous expression system of the host, mostly *E. coli*. On the other hand, sequence-based approach either use a direct sequencing of the genomes and look for homologues

sequences for a particular gene of interest in database, and /or PCR based or hybridisation-based approaches are followed to identify target gene. However, sequence-based approaches are limited by the inability to identify novel genes or novel sequences from metagenomes.

1.9.1. Sequence based screening

Recently sequence based screening approach are more widely used because of the advances of NGS sequencing (shot gun metagenomics) and easy accessibility of metagenomics sequencing data sets and computational tools that can be used to assemble and annotate short reads (Madhavan *et al.*, 2017). Protein- coding genes identified can be analysed with the help of protein data bank (NCBI, NR, Uniprot) or Pfam (Fajardo and Fiser 2013) and CAZy database (Lombard *et al.*, 2014b). Thus, one can infer the presence or absence of targeted genes (such as α -fucosidases and sialidases) based on sequence similarity searches and structural similarity with sequence data bases without the need of laborious screening procedures (Ngara and Zhang 2018). Once the necessary genes are identified, it's possible to PCR and clone the genes into expression host for functional studies. However, this approach has its own drawbacks: such as its effectiveness is limited by the accuracy of genome annotation, the quality of the metagenome sequence and bioinformatics skills. Therefore, novel sequences which lack significant similarity within the data bases, could possibly be missed (Ngara and Zhang 2018). Recently sequence-based screening has identified novel β -galactosidase enzyme sequences from metagenomic data using metagenomic rapid annotation subsystem technology (MG RAST) analysis. Thirty putative β -galactosidase genes were identified, the putative genes were PCR amplified, cloned and screened for β -galactosidase activity. Out of the 30 clones, 12 of them (40%) exhibited the predicted β -galactosidase activity (Liu *et al.*, 2019b). The CAZymes encoding profiles of the microbiota residing in the human body sites from HMP project were analysed, the highest abundance of CAZymes, mainly for complex polysaccharide utilisation were detected from microorganisms residing in the human GIT (Cantarel *et al.*, 2012).

1.9.2. Function based screening of metagenome libraries

Function based screening of metagenome libraries is one of the leading techniques implemented to investigate and identify novel genes from uncultured microorganisms. The chances of obtaining previously uncharacterised genes are dependent on many factors; such as the host-vector system, size of the gene, abundance in the metagenome library, the assay method and efficiency of expression of the target gene in the heterologous host, mainly *E.coli* (Uchiyama and Miyazaki 2009). The metagenome DNA can be cloned in various vectors which suit for desired size and expression in the surrogate host *E. coli*. Plasmids are preferred vectors for cloning DNA fragments smaller than 10kb size, cosmids (25-35 kb), fosmids (25-40kb) and bacterial artificial chromosomes (BAC) can accommodate 100-200 kb (Uchiyama and Miyazaki 2009). Many function-based screening for enzymes of the metagenome libraries have been conducted in agar based screening, usually based on the observation of a desired phenotype such as formation of halo or colour formation if chromogenic substrates are used. For example lipase and esterase enzymes from metagenome samples were screened in agar supplemented with tributyrin (Hardeman and Sjoling 2007) and positive clones that express functional lipase enzymes were identified by the formation of halo around the clones. Over the years function based screening of metagenomics library constructed from community DNA extracted from soil, sand, marine sediment, compost, hot springs, oceanic vents, acid drainage environments, soda lakes, human gut, termite gut etc. have been screened for identification of industrially and biotechnologically important enzymes (Lorenz and Eck 2005; Madhavan *et al.*, 2017). These screens have a lot of success in identification and characterisation of enzymes with novel properties that meet the demand of industries.

Although metagenome sequence based studies have the lions share in terms of understanding the microbial composition, diversity and function of the human gut microbiota, function based screening of human gut microbiome has identified enzymes from potentially uncultured microbes particularly with the complement of the genome annotation of the microbiome sequencing. In the past few years, function-based screens for various enzymes were conducted from metagenome libraries of human gut microbiota DNA extracted from a sample of either faecal, mucosa or combination of both. For

example, the function-based screening for identification of CAZymes which catabolise breakdown of five different dietary fibres, β -glucans, xylans, β -(1-4)-galactan, pectin, and amylose. Out of 156,000 clones ($\sim 5.46 \times 10^9$ bp DNA) screened 310 positive clones which express active β -glucanases, xylanases, galactanases, pectinases and amylases were identified (Tasse *et al.*, 2010). The metagenome inserts of 26 selected clones with active CAZymes were sequenced, and 622 protein coding genes were identified, which are related to carbohydrate metabolism. The taxonomic prediction of these inserts revealed >50% of the sequences have no similarity in databases. Three inserts were comprised *Bifidobacterium* genome sequences, 9 inserts were *Bacteroides*; and 2 clones *Eubacterium* (Tasse *et al.*, 2010). In line with other findings these data confirm that the *Bifidobacterium* and *Bacteroides* are the major glycan degraders in the human gut. Similarly function based screening identified CAZymes and novel metabolic pathways related to breakdown of prebiotic carbohydrates such as XOS, FOS, GOS, inulin and lactulose (Cecchini *et al.*, 2013b). Again, most of the selected clonal inserts were obtained from previously unidentified organisms. The identifiable inserts of faecal origin are predominated by *Bifidobacterium* species (50%) such as *B. longum* and *B. adolescentis*, followed by *Bacteroides* and Clostridiales. The ileum sourced inserts had high diversity of microbes such as Bacteroidales, Clostridiales, *Eubacterium rectale*, *Faecalibacterium prausnitzii*, *Streptococcus thermophiles*. The high abundance of *Bifidobacterium* and *Bacteroides* species is in parallel with the glycan foraging ability of these two groups, which would indicate that they possess a wide array of GH enzymes (Milani *et al.*, 2015a).

1.10. Research goals

In this PhD project, the major motivation is to understand the underlying mechanism of host–microbe interaction of the human gut microbiome through metagenomic and glycomic technologies. For this purpose, a human faecal microbiome metagenome library was used for function-based screening of glycoside hydrolase family enzymes particularly sialidases, α -fucosidases and β -galactosidases. The microbiome uses these enzymes to breakdown dietary fibres (mainly of plant based non-digestible polysaccharides), HMOs and deconstruction of mucin O-glycans. Members of the microbiome which encode these

repertoires of the glycoside hydrolase enzymes have a success in the colonisation and adaptation in the human gut environment.

The following were the hypotheses

Hypotheses

- How the glycoside hydrolases encoded by the members of the gut microbiome be identified through functional metagenomics approach for biotechnological applications?
- How to improve the enzymatic property of glycoside hydrolases through immobilisation for industrial application?
- Glycoside hydrolases encoded by human gut microbiome could impact host microbe interactions and provides clue about physiological status of the host.

Objectives

The main Objectives of this study were to:

- Identify glycoside hydrolases from human gut microbiome metagenome library particularly sialidases, α -fucosidases and β -galactosidases.
- Characterize the glycoside hydrolase properties of selected novel enzymes.
- Determine the biological significance and biotechnological applications of the Glycoside hydrolase enzymes.
- Characterise the property of an immobilized β -galactosidase from *B. adolescentis* and deduce its advantage over using free enzyme.

The schematic representation of the function-based screen procedures used in this study are indicated below (Fig. 1.5)

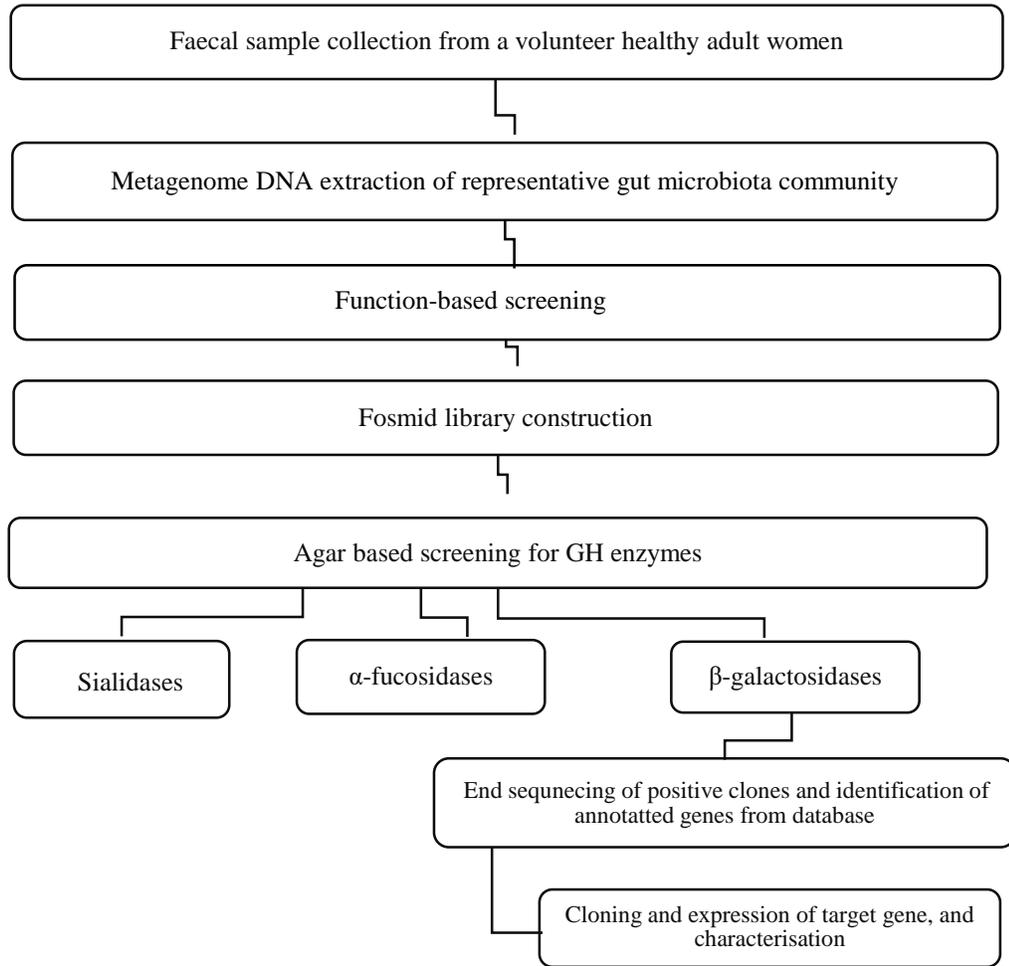


Fig 1.5. A schematic representation of the function based and sequence based screening of the human faecal metagenome, used in this study.

Chapter 2

Materials and Methods

2.1. Bacterial strains, plasmids and fosmids

The bacterial strains, plasmids and fosmids used for this study are listed in table 2.1.

Table 2.1 Bacterial strains, plasmids and fosmids used in this study

Bacterial strains, plasmids and fosmids	Genotype/ feature	Source
<i>E. coli</i>		
BL21 (DE3)	<i>E. coli</i> str. B F ⁻ <i>ompT gal dcm lon hsdS_B</i> (<i>r_B⁻m_B⁻</i>) λ(DE3 [<i>lacI lacUV5-T7p07</i> <i>ind1 sam7 nin5</i>]) [<i>malB⁺</i>] _{K-12} (λ ^S)	Novagen
EPI300	F ⁻ λ ⁻ <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) Φ80d <i>lacZΔM15</i> Δ(<i>lac</i>)X74 <i>recA1</i> <i>endA1 araD139</i> Δ(<i>ara, leu</i>)7697 <i>galU</i> <i>galK rpsL</i> (Str ^R) <i>nupG' trfA dhfr</i>	Epicentre
T7 express	<i>fhuA2 lacZ::T7 gene1 [lon] ompT gal</i> <i>sulA11 R(mcr- 73::miniTn10--Tet^S)2</i> <i>[dcm] R(zgb-210::Tn10--Tet^S endA1</i> Δ(<i>mcrC-mrr</i>)114::IS10	NEB (New England Biolabs)
<i>Plasmids and Fosmids</i>		
pET101	T7 promoter expression vector, Amp ^R	Invitrogen
pLysS	T7p20 <i>ori_{p15A}</i> , Cm ^R	Novagen
pCC1FOS	Fosmid, Cm ^R	Epicentre
pETM10	Expression vector harbouring the α-L-fucosidase gene of <i>Thermotoga maritima</i> , Chloroamphenicol ^R , Amp ^R	(Lezyk <i>et al.</i> , 2016).
pAf	Expression vector harbouring the sialidase gene of <i>Aspergillus fumigatus</i> Kan ^R	(Warwas <i>et al.</i> , 2010).
pDMg1a	pET101 harbouring BAD_1582 with C-terminal His-tag, Amp ^R	This study

2.2. Culture media and chemical supplements

All media and chemical supplements were obtained from Sigma Aldrich unless otherwise stated. Antibiotics were used at the following final concentrations: Chloramphenicol ($12.5 \mu\text{g ml}^{-1}$), Ampicillin (Amp) ($120 \mu\text{g ml}^{-1}$) and Kanamycin (Kan) ($50 \mu\text{g ml}^{-1}$).

Luria Bertani (LB) and MacConkey Media were prepared according to manufacturer's instructions: LB agar (Lennox), LB broth (Miller), MacConkey agar base (Difco). Desired carbohydrates such as lactose, 2-fucosyllactose, 3-sialyllactose, 6-sialyllactose were added as an individual carbon source to MacConkey Agar Base to a final concentration of 0.5-1%.

SOC medium constituted 20 g l^{-1} tryptone, 5 g l^{-1} yeast extract, 8.6 mM NaCl and 1 M glucose.

M9 salts (5X) $0.24 \text{ M Na}_2\text{HPO}_4$; $0.11 \text{ M KH}_2\text{PO}_4$; 43 mM NaCl ; $94 \text{ mM NH}_4\text{Cl}$; $0.09 \text{ mM FeSO}_4 \cdot 7\text{H}_2\text{O}$.

M9 minimal media was prepared with 1x M9 salts, 2 mM MgSO_4 , 0.1 mM CaCl_2 , and 0.3 mM thiamine and 0.061 mM leucine. The M9 minimal media were supplemented with 0.4% different carbon sources (lactose, glucose, 2-fucosyllactose, and fucose) were prepared.

2.3. Transfer of the human faecal metagenome library to a new *E. coli* EPI 300-T1^R host

A human faecal microbiome metagenome library that was previously constructed (Christy Agbavwe 2017) from a volunteer healthy adult human gut faecal sample using the Copy Control™ Fosmid Library Production Kit with pCC1FOS™ Vector (Epicentre) was used. The metagenome library was transferred into new heterologous host strain (EPI 300-T1R) as described below.

2.3.1. Preparation of fosmids from the stock human faecal metagenome library

Aliquots (7 ml) of stock human faecal metagenome library were taken out of the -80 °C freezer and thawed on ice. The aliquots were aseptically transferred to 75 ml LB broth containing Cm and 0.01% arabinose and incubated at 37°C with shaking at 150 rpm for 4 h. The culture was collected and centrifuged at 3000 x g for 15 min using Eppendorf centrifuge 5416. The pellets were pooled and washed with sterile water (2x). Fosmids were prepared using PureLink® HiPure Plasmid Filter Purification Kits (Invitrogen) according to manufacturer's instruction. The pellets were resuspended with 10 mL Resuspension Buffer (R3) with RNase A until cells suspension was homogeneous. Then 10 mL Lysis Buffer (L7) was added, mixed gently by inverting the capped tube and incubated for 5 min at room temperature. After, 10 mL Precipitation Buffer (N3) was added and mixed immediately by inverting the tube until the mixture was thoroughly homogeneous. The lysate mixture was transferred into equilibrated HiPure Filter Midi Column (by adding 15 mL Equilibration Buffer (EQ1) to the cartridge). The lysate was allowed to pass through the filter by gravity flow. Then the Midi column was washed with 20 ml Wash Buffer (W8). Fosmid DNA was eluted by adding 5 ml Elution buffer to the Midi column allowing the solution to drain by gravity. Thereafter, DNA was precipitated by adding 3.5 ml isopropanol to the eluate and incubated for 2 min at room temperature, centrifuged at 12000 x g for 30 min at 4°C (Sorvall GSA Rotor), the supernatant was discarded carefully. The pellet was resuspended in 3 ml 70% ethanol, centrifuged at 12000 x g for 5 min at 4°C (Sorvall GSA Rotor), supernatant discarded. The pelleted DNA was air dried for 10 min to remove residual ethanol and resuspended in 100 µl sterile water. The concentration of the purified fosmid DNA were determined using NanoDrop 2000 (ThermoFisher Scientific) and integrity of the DNA was checked using agarose gel electrophoresis. The fosmid DNA was kept in -20°C.

2.3.2. Preparation of the heterologous host EPI 300-T1^R

The EPI300-T1^R strain (supplied as a glycerol stock, by Epicentre) was streaked out on an LB plate without supplementation of antibiotic and incubated overnight at 37°C. The overnight grown strain was sealed with parafilm and stored at 4°C.

The day before the Lambda Packaging reaction, a single colony of EPI300-T1R cells were picked and inoculated into 50 ml LB broth + 10 mM MgSO₄ + 0.2% maltose. The flasks were incubated overnight at 37°C with shaking at 150 rpm. On the following day just prior to lambda phage packaging procedure, 0.5 ml of the overnight grown cells were inoculated to 50 ml LB broth + 10 mM MgSO₄ + 0.2% maltose and incubated at 37°C with shaking at 150 rpm till OD₆₀₀ of 0.8-1.0 was reached (~2 hours). The cells were stored at 4°C.

2.3.3. Packaging the copy control fosmid DNA into MaxPlax Lambda phages

A single tube of the MaxPlax Lambda Packaging Extracts was withdrawn from -80°C freezer and thawed on ice. 0.25 µg of midiprep fosmid DNA was transferred to 25 µl Lambda packaging extract and incubated at 30°C for 2 h. After that, an additional 25 µl of the packaging extract was added and mixed by pipetting and incubated for additional 2 h at 30°C. Afterwards, phage dilution buffer (PDB; 10 mM Tris HCl (pH 8.3), 100 mM NaCl, and 10 mM MgCl₂) was transferred to the packaging reaction to make a final volume of 1 ml and mixed gently by inverting the tube. The reaction was completed by addition of 25 µl of chloroform followed by gentle mixing by inverting the tubes. The titre of phage particles (packaged copy control fosmid clones) was done first by diluting ten-fold serial dilution (1:10¹-1:10³) in PDB. 10 µl of each dilution and undiluted phage particles were transferred to separate microfuges containing 100 µl EPI-300-T1^R and incubated for 1 h at 37°C to induce infection. All the infected EPI-300-T1^R cells were plated on LB agar containing C^m and incubated at 37°C overnight. The following day cfu/ml of each infection reaction were calculated, thereby the titre which gave the desired number of cfu/ml were selected.

2.3.4. Plating and selecting the packaged copy control fosmid library

The 1:10² dilution (1.0 E+07 cfu/ml) provided the desired number of colonies and 10 µl of this dilution was used to infect 100 µl of EPI-300-T1^R cells, this reaction was carried out in 100 separate microfuge and incubated for 1h at 37°C. After 1 h, 100 µl was withdrawn from each microfuge tubes and plated on separate LB agar, containing Cm and incubated overnight at 37°C. The overnight grew colonies from each plate were suspended

with LB broth. 2 ml LB broth was added to the 1st plate and colonies were detached from the plate with sterile spreader and suspension was transferred to a 2nd plate to resuspend colonies and this was repeated till the 5th plate, then detached colonies were pooled and transferred to a 50 ml falcon tube. Similarly, the colonies from the remaining 95 plates were detached and transferred to the 50 ml falcon tube. Then 20% glycerol was added to the pooled fosmid library, aliquots of 0.5 ml of this library were transferred to cryotubes and stored at -80°C.

2.4. Function based screening of the metagenome library

The human faecal metagenome library was screened for identification of glycoside hydrolase enzymes through agar-based screening. A vial of stock human faecal microbiome metagenome library was taken out from -80°C and thawed on ice, a serial dilution which provided ~500 colonies per plate were prepared in LB broth. 100µl was plated on LB agar supplemented with the chromogenic/fluorescent substrates at desired concentration and chloramphenicol. Positive clones were identified based on their hydrolytic activity of chromogenic and/or fluorescent substrates as the hydrolytic products either cause colour change in the agar or fluoresce under UV exposure respectively, while the negative clones were disregarded.

2.4.1. Functional Screening of metagenome library for β galactosidases

The human faecal microbiome metagenome library was screened for pink colonies expressing a functional β -galactosidase enzyme on MacConkey Agar supplemented with lactose containing Cm and arabinose. Clones that appeared pink on the MacConkey agar were picked and streaked on LB agar containing chloramphenicol, arabinose and the chromogenic substrate 5-bromo-4-chloro-3-indolyl- β -galactopyranoside (X-Gal) at final concentration of 40 $\mu\text{g ml}^{-1}$ for confirmatory test and positive clones identified by formation of blue colonies. The Miller assay (Miller 1972) was conducted on positive clones to determine β -galactosidase activity. The β -galactosidase positive clones, positive control *lacZ*⁺ *E.coli* B121 (DE3) and negative control *lacZ* *E. coli* EPI300(pCC1FOS) were grown overnight and OD 600 nm of the overnight culture was determined. 100 μl of the overnight grown cultures of each clone were transferred to 2 ml microcentrifuge tubes separately and permeabilised by adding 900 μl of Z buffer (60 mM Na₂HPO₄, 40 mM

NaH₂PO₄·2H₂O, 10 mM KCl, 1.0 mM MgSO₄·7H₂O, 50mM β-mercaptoethanol, pH 7.0) containing one drop of (0.1% SDS) and two drops of CHCl₃ (chloroform) and vortexed for 10 seconds. The tubes were incubated at 28 °C for 5 min then 200 µl of oNPG (4mg/ml stock, prepared in dH₂O) was added to initiate the reaction. The time of addition recorded, and the reaction was stopped after sufficient yellow colour was developed by addition of 0.5 ml 1M Na₂CO₃. Then Absorbance at 420 nm was recorded, Miller unit was calculated as follows:

$$\text{Miller Units} = \frac{1000}{1} \times \frac{\text{OD}_{420\text{nm}}}{t \times V \times \text{OD}_{600\text{nm}}}$$

Where; t= Duration of reaction in min

V= Volume of the culture added in ml.

2.4.2. Functional Screening of metagenome library for α-fucosidases

In the first approach the human faecal microbiome metagenome library was screened for pink colonies expressing a functional α-fucosidase enzyme on MacConkey base agar supplemented with 2-fucosyllactose containing chloramphenicol and arabinose. Clones that appeared pink on this agar were selected.

In the second approach the human faecal metagenome library was screened for α fucosidase enzyme encoding clones using LB agar supplemented with 100 µg ml⁻¹ of the chromogenic substrate 5-bromo-4-chloro-3-indolyl-α-L-fucopyranoside (X-fuc) (BIOSYNTH), positive clones were identified based on blue colony morphology. The strain *E. coli* BL21 (DE3) pLysS bearing an expression plasmid pETM10 with cloned recombinant codon optimised α-L- fucosidase gene of *Thermotoga martima* pETM10 (Lezyk *et al.*, 2016) were used as a positive control. A lawn of this strain streaked on LB agar supplemented with 100 µg ml⁻¹ X-fuc and 0.1 mM IPTG formed blue colonies.

The human gut microbiome metagenome library were screened for identifying α-L-fucosidase expressing clones using LB agar supplemented with 100 µg ml⁻¹ X-fuc , Cm and arabinose.

2.4.3. Functional Screening of metagenome library for sialidases

The human faecal microbiome metagenome library was screened for pink colonies expressing a functional sialidase enzyme on MacConkey base agar supplemented with 1.024% of 3-sialyllactose or 6-siallyl lactose (GLYCOM) containing Cm and arabinose. Sialidase expressing clones were identified by the pink colony formation. *E. coli* strain BL21 (DE3) clone expressing well-described *Aspergillus fumigatus* sialidase (Warwas *et al.*, 2010) was used as positive control, in the presence of 4-methylumbelliferyl-Neu5AC and IPTG, colonies fluoresce under UV light exposure. The *E. coli* BL21 (DE3) (pAf sialidase) and *E. coli* BL2 (DE3) (pET28b) were streaked on LB plates containing 0.025 mM IPTG and Kan overnight at 37°C. The following day, the culture was overlaid with 2.5 ml of solution containing (0.75% sterile agar, 1mg ml⁻¹ 2-(4-Methylumbelliferyl)- α -D-N- acetylneuraminic acid sodium salt (4-MU-NANA, Apollo Scientific), 50 mM sodium acetate (pH5.5) ; 150 mM sodium chloride ; 4 mM calcium chloride) and re-incubated for overnight at 37°C. The pAf sialidase positive clone and pET 28 clones were exposed to UV illumination. It was expected that pAf sialidase positive clone would appear fluorescent after hydrolyzing the 4MU-NANA substrate while pET28b clone would be non-fluorescent.

2.5. Genetic techniques

2.5.1. Agarose gel electrophoresis

Separation of DNA fragments were carried out by agarose gel electrophoresis (Biorad), in 1XTAE buffer (50X TAE stock was made; 242 g of Tris base, 57.1 ml of glacial acetic acid and 100 ml of 0.5 M EDTA (18.61 g of disodium EDTA•2H₂O to 80 ml of H₂O, pH 8.0), 1:10,000 SYBR Safe, (10,000X stock) (Invitrogen). HyperLadder™1 Kb (Bioline) was loaded on the first lane for reference and sample DNA was mixed with loading buffer blue (Bioline) and subsequently added to the agarose wells. After electrophoresis, DNA samples stained with SYBR safe were visualised with epi blue illumination using Azure™ gel documentation system, Azure Biosystems with an inbuilt C-series capture software.

2.5.2. Purification of fosmids and plasmids

For fosmid preparation a single colony of metagenome clone identified from function-based screening of human faecal microbiome metagenome library was inoculated in 15 ml LB supplemented with Cm and arabinose. The culture was grown for ~17 h at 37°C with constant shaking. For extraction of plasmid, a single colony of *E. coli* cells harbouring the target plasmid was inoculated to 15 ml LB broth supplemented with 50 µg ml⁻¹ kanamycin (pAf silalidase) and 34 µg ml⁻¹ chloramphenicol (pLys S) and culture was grown for ~17 h at 37°C with constant shaking. Plasmid pLysS was purified from *E. coli* Rosetta2 (DE3) pLysS (Novagen). The following day bacterial culture was centrifuged at 3000 x g (Eppendorf centrifuge, 5416) for 15 min at room temperature. The supernatant was decanted carefully, and the pellet was kept frozen until used. The pellet was resuspended in 600 µl of buffer P1 (QIAGEN Plasmid Mini Kit) by pipette action. Then 600 µl of buffer P2 (QIAGEN) was added, mixed by inverting the tubes 4-5 times until the lysis blue reagent in P1 is uniformly distributed and incubated for 5 min at room temperature to lyse the cells. At 5 min, 800 µl of buffer N3 was added and the cell lysis suspension was mixed well by inverting the tubes 4-5 times. The mixture was centrifuged (Eppendorf, 5415D) at 13,000 rpm for 10 minutes at room temperature. The supernatant (800µl at a time) was transferred to a single Miniprep column (3X spins, flow through removed after each spin). The column was washed with 500 µl buffer PB centrifuged at 13,000 rpm for 1 min, followed by a second wash with 750 µl buffer PE and centrifugation for 1 min at 13,000rpm. The residual PE buffer was removed by centrifugation at 13,000 rpm for 1 min. Fosmid/plasmid DNA was eluted from the mini prep column with 50µl of pre-heated (70°C) sterile water, elution was repeated (into the same tube) with 30µl of pre-heated sterile water. The concentration of fosmid/plasmid DNA was quantified using Nano drop.

2.5.3. Restriction analysis of fosmids

Fosmid DNA was digested with combination of either of the following restriction enzymes *Bam*HI, *Sac*I, *Sma*I, and *Hind*III (ThermoFisher scientific) and the restriction products visualised by agarose gel electrophoresis. The restriction reaction mix contained (4.5µl H₂O, 1 µl 10X FastDigestBuffer (ThermoFisher Scientific), 0.5 µl of restriction

endonuclease) 4 μ l fosmid DNA. The mixture was incubated at 37°C for 1 h and 30 min. Restriction products were analysed by the pattern of migration on agarose gel electrophoresis. Clones that had distinct restriction patterns were further analysed by end sequencing (Eurofins genomics).

2.5.4. Sequencing of fosmid inserts and bioinformatic analysis

Distinct fosmid clones which were identified through agar based screening were sequenced (Eurofins Genomics custom DNA sequencing service) using pCC1FOS sequencing primers (forward primer GGATGTGCTGCAAGGCGATTAAGTTGG and reverse primer CTCGTATGTTGTGTGGAATTGTGAGC) to generate DNA sequences of the flanking regions of insert sequences in the fosmids.

The forward and the reverse nucleotide sequences generated from end sequencing reactions of metagenome clones were analysed using NCBI nucleotide and protein Blast services (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Conserved domain databases (CDD), and Pfam were used to identify the conserved amino acid sequences in the target protein. Smart BLAST and Constraint-Based Multiple Alignment Tool COBALT (Papadopoulos and Agarwala 2007) were used to identify homologues of BAD_1582 with the landmark sequences. In addition PSORTb subcellular localization tool was implemented to predict subcellular protein localization (Yu *et al.*, 2010).

2.5.5. Polymerase Chain Reaction

To determine the presence of BAD_1582 in fosmid clones which showed β -galactosidase activity PCR was carried out following the Taq DNA polymerase kit (Bioline). The forward primer (ACGTATGCCTCGAATCG) and reverse primer (CATATTTGGATAGCTC) were designed to amplify BAD_1582 gene sequence of *B. adolescentis*. The PCR reaction mix contained (5x Mytaq DNA polymerase, 5 μ l; Forward primer (5 pmol μ l⁻¹), 2 μ l; reverse primer (5 pmol μ l⁻¹), 2 μ l; My Taq DNA polymerase, 0.125 μ l; Nuclease free water, 14.875 μ l) fosmid DNA from each clone, 10 ng μ l⁻¹ PCR was run at the following cycling condition (25 cycles: 30 s at 95°C, 30 s at 45°C, 90 s at 72°C) using thermocycler (Eppendorf mastercycler® gradient, Hamburg, Germany).

For the purpose of cloning, a blunt end PCR product of BAD_1582 was amplified using one of the β galactosidase positive fosmid clone DNA (Clone 31) as template. To enable directional cloning a forward primer was designed by incorporating four nucleotides, CACC, which base pair to the GTGG overhang of the pET 101/D-TOPO vector. On the other hand the reverse primer were designed in such a way that it didn't base pair with the overhang of the vector and by excluding the stop codon from the gene so as to incorporate the V5 epitope and C-terminal polyhistidine (6X His) tag from the vector as its depicted in Fig.2.1 below.

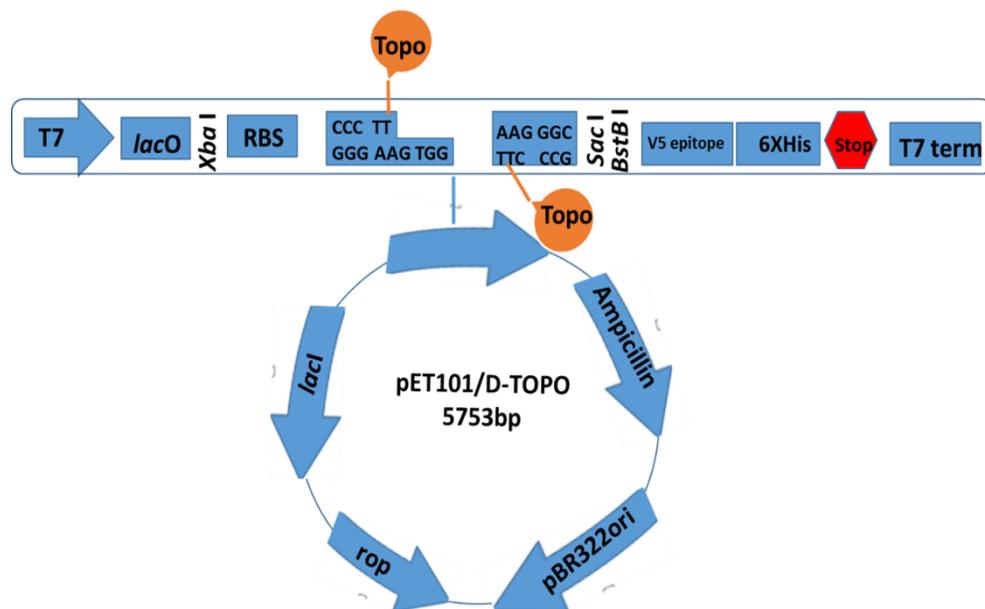


Fig 2.1: The features of the pET101/D-TOPO® (5753 bp) vector, adapted from CopyControl™ Fosmid LibraryProduction Kit with pCC1FOS™ Vector instruction manual (<http://www.epibio.com/applications/genomic-library-construction-kits-gene-pcr-cloning-kits/cloning-genomic/copycontrol-fosmid-library-production-kit>).

The forward primer (CACCATGGCAGATACAGCCGAACTC) and the reverse primer (GAACAGCTTGAGCTGAACGTTGAG) were used to amplify the target gene using Velocity DNA polymerase (Bioline) to generate a blunt end product. The PCR reaction contained (5x Hi-Fi Reaction Buffer, 10µl; 100mM dNTP Mix, 0.5µl; Primers (5pmol µl⁻¹ each, 1µl; Velocity polymerase, 1 µl, nuclease free water, 35.5 µl) 1 µl of template fosmid DNA Then sample and appropriate positive and negative controls were run in thermocycler (Eppendorf mastercycler® gradient, Hamburg, Germany) with cycling conditions (25 cycles: 30 s at 98°C, 30 s at 63°C, 1 min and 30 s at 72°C). The amplification of the target gene was confirmed using agarose gel electrophoresis and visualization with Azure Gel documentation system.

2.5.6. DNA purification using wizard SV Gel/ PCR clean up kit (Promega)

Purification of the PCR product was carried out using Promega- Wizard R SV gel and PCR clean up system following the manufacturer's instructions. First of all, the blunt end PCR product was carefully excised from the agarose gel with sterile scalpel under UV visualisation. 120µl of membrane binding solution was added to the 120 mg of gel slice. The mixture was vortexed briefly and incubated at 60°C for 10 min until the gel slice was completely dissolved. The dissolved gel mixture was transferred to the SV Mini column assembly and incubated for 1 min at room temperature. Then the SV Mini column assembly was centrifuged in Eppendorf micro centrifuge at 13,000 rpm for 1 min. the SV Mini column was washed with 700 µl membrane wash solution and centrifuged at 13,000 rpm for 1 min The wash was repeated with 500 µl membrane wash solution, followed by centrifugation at 13,000 rpm for 5 min. Centrifugation was repeated for 1 min with the lid of the centrifuge tube opened so as to allow evaporation of residual ethanol. Afterwards, 50 µl nuclease free water was added to the center of the column and incubated for 1 min at room temperature. The clean blunt end PCR product was eluted after centrifugation at 13,000 rpm and micro centrifuge containing the pure DNA was stored at 4°C.

2.5.7. Cloning of identified β -galactosidase gene (BAD_1582) of *B. adolescentis*

The amplified DNA was cloned into pET101 (TOPO D) vector (Invitrogen) following the directional cloning kit protocol. For successful cloning the molar ratio of PCR product to TOPO vector should be 2:1. Thus to determine the concentration of the PCR product, agarose electrophoresis was carried out (section 2.5.1) and HyperLadder™ 1kb was run as reference. The concentration of the PCR product was estimated by comparing the intensity of the band of the PCR product sample with the intensity of the bands in the HyperLadder™ 1kb. Then the molar ratio of the PCR product to TOPO vector was adjusted accordingly the cloning reaction contained (PCR product, 2 μ l; salt solution, 1 μ l; TOPO® vector, 1 μ l; 2 μ l of sterile water was added to bring the final volume to 6 μ l). Similarly, a negative control reaction was carried out incorporating all the components, except the PCR product. The cloning reaction and the control were mixed gently and incubated at room temperature for 30 min.

2.5.8. Transformation of chemically competent, *E. coli* T7 express cells

T7 Express Competent *E. coli* cells (C2566H) were thawed on ice for 10 minutes. 3 μ l of the TOPO® cloning reaction mixture was transferred to the thawed cell mixture. The control tube received an equal volume of negative control cloning reaction. After being kept on ice for 30 minutes, the cells were heat shocked at exactly 42°C for exactly 10 sec and immediately transferred on ice for 5 minutes. Finally, 950 μ l of room temperature SOC was transferred to the tubes and incubated at 37°C for 60 minutes with vigorous shaking (250 rpm). At the end of incubation, several 10-fold serial dilutions were made in SOC. 100 μ l of each dilution were spread onto a selection plate LB agar supplemented with Amp and incubated overnight at 37°C.

2.5.9. Identification and Verification of cloned BAD_1582 into pET101

A colony PCR was carried out to confirm the presence of inserted BAD_1582 gene in the transformed T7 Express *E. coli* colonies. For this purpose, portion of colonies from 37 individual colonies were picked and suspended in 20 μ l of sterile water separately. The suspension (1 μ l) was used as a template, PCR was carried out using My Taq DNA polymerase kit using the same primers and cycling conditions as used to generate the PCR

product for cloning. The PCR products were analyzed using 1% agarose gel electrophoresis for the presence of the correct size of BAD_1582 (3150 bp). The transformed colonies which contained pET101 vector harboring BAD_1582 gene were named as *E. coli* T7 express (pDMg1a) and the colonies which only contained the pET101 vector was named as *E. coli* T7express (pET101).

Furthermore, plasmid was prepared from *E. coli* T7 express (pDMg1a) and *E. coli* T7 express (pET 101) and the plasmids were digested using *Hind* III and *Sac* I and the restriction fragments analysed by agarose gel electrophoresis. The restriction enzymes were selected based on restriction map of pET 101D/Topo+BAD_1582 gene sequence built using pDraw software and the expected size of restriction products of pDMg1a (6058bp, 1576bp, 884bp, 374 bp) were predicted using this software.

The presence of an intact sequence of BAD_1582 in pDMg1a was confirmed by primer walk sequencing of the plasmid pDMg1a, using couple of primers the first sequences were generated using T7 forward primer (5'-TAATACGACTCACTATAGGG-3') and reverse primer (5'-TAGTTATTGCTCAGCGGTGG-3'). Subsequently using these sequences as template new primers were designed, forward primer (5'-ATGACTATCGAAAGCACGCTGC-3') reverse primer (5'-CGCCGAACGTATATGAAACCATAC-3') and additional forward primer (5'-GCGCTTCATTCGAAGTCGATTG-3') were also designed to cover the entire gene sequence (3150bp) of BAD_1582.

2.5.11. Preparation of chemically competent *E. coli* T7express (pDMg1a) and transformation with plysS plasmid

Chemically competent *E. coli* T7 express (pDMg1a) and T7express (pET101) were prepared. A single colony of T7express (pDMg1a) and T7express (pET101) were picked and separately inoculated in 30 ml LB broth (Amp, 120 µg/ml) and incubated overnight at 37°C with shaking. The following day 0.1 OD_{600nm} of the overnight cultures were individually transferred to 30 ml of fresh LB broth in 250 ml flask, incubated with shaking at 37°C for 90 min until OD_{600nm} ~0.4-0.7 was reached. Cells were harvested by centrifugation at 2700 x g in a Sorvall GSA rotor for 10 minutes at 4°C. The cell pellets were resuspended in 30 ml of ice-cold sterile magnesium chloride and calcium chloride

solution (80 mM MgCl₂.6H₂O + 20 mM CaCl₂.2H₂O). Then the cells were recovered by centrifugation at 2700 x g for 10 minutes at 4°C, and the pellets were resuspended in 2 ml of ice cold 0.1M CaCl₂.2H₂O solution and became ready for chemical transformation and 50 µl of aliquots were dispensed in to 2 ml Eppendorf tubes for immediate use, while the remaining were stored in 15% glycerol at -80°C. 100 ng of pLys S plasmid DNA was transferred into these cells by chemical transformation as described above with selection of transformed colonies on Cm and Amp.

2.5.12. Preparation and Transformation of Electrocompetent *E. coli* BL21 (DE3)

Electrocompetent *E. coli* BL21(DE3) cells was prepared by using the method described by (Joseph Sambrook 2003). A single colony of *E. coli* BL21 (DE3) was inoculated in 30 ml LB broth and incubated overnight at 37°C with shaking at 150 rpm. The following day an equivalent of 0.1 OD₆₀₀ of the overnight grown cells were transferred to 50 ml LB broth. The culture was grown until the OD₆₀₀ was 0.4, and then it was transferred to ice bath for 15 min. cells were harvested by centrifugation at 2,500 rpm in a Sorvall GSA rotor for 15 minutes at 4°C. The cells were resuspended in 50 ml of ice-cold sterile H₂O and harvested by centrifugation for 20 min at 4°C. The cells were resuspended in 25 ml of ice cold 10% glycerol solution and cells were harvested by centrifugation for 20 min as above. The cells were resuspended in 1 ml of GYT medium (10% (v/v) glycerol, 0.125% (w/v) yeast extract, 0.25% (w/v) tryptone). The cell suspension was diluted in GYT medium to obtain an equivalent of 0.1 OD₆₀₀ (~ 2x10⁸ cells ml⁻¹), and aliquots of 40 µl were immediately stored at -80°C.

Electrocompetent *E. coli* BL21 (DE3) aliquots (40µl) were taken out from -80 freezer and thawed on ice. One µl of pAf sialidase plasmid DNA (~100 ng) (Warwas *et al.*, 2010) was transferred to the thawed *E. coli* BL21 (DE3) and to ice cold electroporation cuvettes. In parallel 1 µl (~100 ng) of empty pET28b vector was transferred to a separate tube containing thawed *E. coli* BL21 (DE3) for negative control. Both samples were electroporated using micro pulser (Biorad), using the pre-programmed setting for *E. coli* 0.2 cm cuvette (Ec2) and immediately 1 ml of room temperature SOC medium was added. These electroporated cells were transferred to bacterial culture tubes and incubated for 1h

at 37°C with shaking. 100 µl from each transformation were spread on selective LB plates containing kanamycin (50 µg ml⁻¹) and 0.1 mM Isopropyl-β-D- thiogalactopyranoside (IPTG).

2.6. Optimisation of expression of His-tagged BAD_1582

The following four strains T7 express *E. coli* (pDMg1a), T7 express *E. coli* (pDMg1a) (pLysS), T7 express *E. coli* (pET101), and T7 express *E. coli* (pET101) (pLysS) were used for optimisation of BAD_1582 expression for different parameters such as IPTG concentration and time of induction. A single colony of each of the four strains were grown overnight in 250 ml conical flask containing 30 ml LB broth (with added Amp for pDMg1a and pET101 harbouring strains and Amp⁺ Cm (34 µg ml⁻¹) for pLysS harbouring strains at 37°C with shaking. The following day 0.1 OD_{600nm} of the respective overnight grown strains were transferred to 250 ml conical flask containing 30 ml of LB broth supplemented with antibiotics incubated for 90 min at 37°C with shaking until OD_{600nm} (~0.4-0.7) in duplicates. Two concentrations of IPTG (0.1 mM & 1 mM) were transferred to either of the duplicates for each strain; and the flasks were incubated at 37°C with shaking for IPTG induction. Aliquots (50 µl) were withdrawn at 0 h, 1 h, 2 h, 3 h and 4 h time points after IPTG addition, centrifuged at 13000 rpm, pellets were stored at -20°C for SDS analysis. The pellets collected from each strain at those different conditions were resuspended in 40 µl of 1X Laemmli buffer and incubated at 85°C for 5 min. After, 10µl of each time point and IPTG induced cell lysates were analysed using SDS PAGE as described below (2.10.2).

2.7. Expression and preparation of cell lysates

E. coli T7 Express (pDMg1a) were grown to mid-log phase (OD_{600nm} between 0.4 and 0.7) in 1000 ml conical flask containing 150 ml LB broth at 37°C with shaking. 1 mM IPTG was added and the culture was incubated for 4 h at 37°C to induce expression. A 100 µl aliquot of the culture was then assayed for β-galactosidase activity using the Miller assay (Miller 1972). Cells were harvested by centrifugation at 3000g (Eppendorf, 5416 centrifuge), stored at -20°C. The pelleted cells were resuspended in 2.5 ml lysis solution (bug buster protein extraction reagent supplemented with lysozyme, 2 µl/ml of extraction buffer) (Merck, USA) incubated at room temperature on a shaker. A 30 µl aliquot of total

cell lysate was stored at -20°C for SDS PAGE analysis. The lysate was centrifuged at 4°C for 20 min at 16,000 g (Beckman Coulter™ Microfuge® 22R centrifuge), supernatant (soluble fraction) was collected and pooled in 2 ml microcentrifuge tubes and immediately used for Ni²⁺-nitrilotriacetate (Ni²⁺-NTA) affinity chromatography purification. The pellet (insoluble fraction) together with the total cell lysate and the soluble fraction were analyzed by SDS-PAGE (section 2.10.2) to confirm the location of the expressed enzyme (BAD_1582).

2.8. Ni²⁺-nitrilotriacetate (Ni²⁺-NTA) affinity chromatography purification

The soluble fraction collected after cell lysis were subjected to Ni²⁺-nitrilotriacetate (Ni²⁺-NTA) affinity chromatography to purify the His-tagged enzyme. HisPur™ Ni-NTA Spin Columns (0.2 ml) (ThermoFisher Scientific™) were used for purification of the His-tagged BAD_1582 enzyme from soluble protein extract according to manufacturer's instruction. For sample preparation, soluble protein extract obtained from one lysis tube (2.5 ml) was mixed with equal volume of equilibration buffer, 10 mM imidazole in PBS (20 mM sodium phosphate, 300 mM sodium chloride pH 7.4). The HisPur Ni-NTA Spin Column were equilibrated with 450 µl of equilibration buffer and allowing the buffer to enter the resin bed, the buffer was removed by centrifugation at 700xg for 2 min. 500 µl of the prepared protein extract sample were loaded to the resin, the columns were incubated on a rocker at room temperature for 30 min. and centrifuged at 700 xg for 2 min and the flow through was collected in new 2 ml microcentrifuge tube. The procedure was repeated until the whole sample was processed. Then the resin was washed 5 times with 500 µl 25 mM imidazole in PBS, thrice with 500 µl 50 mM imidazole in PBS and thrice with 500 µl 75 mM imidazole in PBS with centrifugation at 700 xg for 2 min after each wash to remove contaminant proteins. Finally, 250 µl of 250 mM imidazole in PBS was added to the resin bed, centrifuged at 700 xg for 2 min to elute the His-tagged BAD_1582. The elution was repeated 6X and eluates were collected in separate collection tubes. Aliquots 20 µl and 15 µl were withdrawn for Bradford assay and (2.10.1) and SDS PAGE analysis (2.10.2) respectively from each elution fractions, the remaining samples were stored at 4°C.

2.9. Ultrafiltration with 100K (MWCO) Amicon® Ultra-0.5 and 4 centrifugal filters

The eluted fractions were subjected to ultrafiltration using 100K MWCO Amicon Ultra-4 centrifugal filters (Merck, USA) according to manufacturer instruction.

When 0.5 centrifugal filters were used, first of all each filter device was inserted to the microcentrifuge tube provided. 500 µl of eluted fraction was added on the filter device and centrifuged at 14,000 x g for 10 min. Filtrate was kept for analysis, to the concentrated enzyme containing filter device 450 µl of PBS was added and centrifuged at 14,000 x g for 10 min, filtrate discarded and the concentrated enzyme of ~50 µl in PBS was retained in the filter device. To recover the concentrated protein, the filter device was placed upside down into a clean microcentrifuge tube, centrifuged at 1000 x g for 2 min. The filter device additionally rinsed with 50 µl of PBS by pipetting up and down, wash was collected by centrifugation at 1000 x g for 2 min.

A similar approach was undertaken for the 4 ml Amicon Ultra-4 centrifugal filter, using 4 ml of pooled eluates from the Ni²⁺-NTA column and centrifugal spins of 3000 x g for 20 min. Samples were stored at 4°C, or in 50% glycerol at -20°C for long term storage.

2.10. Protein analysis

2.10.1. Bradford assay

The concentration of proteins was determined using Bradford assay. Bovine serum albumin (BSA, 2 mg ml⁻¹ stock) was serially diluted in PBS (pH 7.4) to obtain final concentrations of 0.1-1.4 mg ml⁻¹. Then 5 µl each BSA protein standard concentration, PBS as blank and each protein sample was transferred to 96 well plates in triplicate. Then 250 µl Bradford reagent was transferred to the standards, samples and blank wells and incubated at room temperature for 15 min. After, absorbance at 595_{nm} was recorded using Tecan Sunrise™ and concentration of the samples was determined from the standard curve of BSA.

2.10.2. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS PAGE)

Analysis of proteins was carried out using SDS-PAGE under denaturing conditions (Laemmli 1970) using the Mini-PROTEAN tetra cell electrophoresis system (Biorad) according to manufacturer instruction.

For separation of the desired protein (~120KDa) a resolving gel of 10 % was selected (370 mM Tris-HCl, pH 8.8, 10% Acrylamide/Bis, 0.1% SDS, 0.1% ammonium persulfate and 0.1% N,N,N',N'-tetramethylethane-1,2-diamine (TEMED) . The stacking gels (4%) constituted (130 mM Tris-HCl, pH 6.8, 4% Acrylamide/Bis and 0.1% SDS, 0.07% ammonium persulfate and 0.14% TEMED. Then 1X running buffer (5X stock constitute, gl^{-1} , Tris base, 14.5; Glycine, 75; SDS, 5) was used to run the gel.

Samples, such as total cell lysates, soluble fractions, insoluble fractions, Ni-NTA eluted fractions and filtered fractions were prepared by mixing 20 μl of sample with 20 μl of 1x Laemmli sample buffer (diluted in dH_2O) and heated at 85°C in a heat block for 5 min. Color pre-stained protein ladder was loaded to the first well and subsequently, 8 μl of (total cell lysate, soluble and insoluble fractions) and 12 μl of elute and filtered fractions were loaded to the next wells and electrophoresed at 185 V for 1 h.

2.10.3. Staining SDS-PAGE gels

The SDS-PAGE gels were stained with Coomassie staining or silver staining. The former was used for routine staining and the gels were immersed in Coomassie staining solution (45% methanol, 10% acetic acid, 1 gl^{-1} Coomassie Blue R) for 45 min in a rocker. The gels were destained overnight with a destaining solution (5% ethanol and 7% acetic acid). The gels were visualized using the Azure gel doc system with white light imaging.

For more sensitive detection of purified proteins, SDS gels were stained with silver stain (Conway *et al.*, 2016) using the Thermo Scientific™ Pierce™ Silver Stain Kit according to manufacturer's instruction. The gel was washed with sterile water for 5 minutes (2 x) on an orbital shaker. Then the gel was incubated in fixing solution (30% ethanol: 10% acetic acid) for 15 minutes (2 x). Then washed with 10% ethanol for 5 min (2 x) and with sterile water for 5 min (2 x) successively. The gel was sensitized in a sensitizer solution (50 μl Sensitizer with 25mL water) for 1 min, after the sensitizer solution was removed,

the gel was washed with sterile water for 1 min (2 x). The gel was immersed in silver stain working solution (0.5 ml Enhancer with 25 ml Stain) and incubated on an orbital shaker for 30 min to be stained. The silver stain solution was removed, and the gel was washed with sterile water for 20 seconds (2x). Then, the gel was developed with developer working solution (0.5 ml Enhancer with 25 ml Developer) for 3 min. After the bands appear the developer solution was removed, washed briefly and replaced with stop solution (5% acetic acid) incubated for 10 minutes. Image was taken using CanoScan 9000F Mark II image scanner.

2.10.4. Immunoblotting

Western blotting of proteins were carried out according to the method developed by (Burnette 1981). The proteins from the SDS-PAGE gels were transferred to nitrocellulose membrane in transfer buffer (48 mM Trizma base, 39 mM glycine, 1.3 mM SDS, 20% methanol, pH 9.2). On a western blot apparatus (V20-SDB 20 x 20cm Semi-Dry Blotter, SCIE-PLAS). The transfer was carried out at 60 mA/ for 45 minutes. The nitrocellulose membrane was immersed in blocking solution (1.25 g skim milk in 1X TBS) and incubated for 1 h on an orbital shaker at room temperature. Then the nitrocellulose membrane was washed 3 times in 1xTBS for 5 min on an orbital shaker. Thereafter, antibody solution was added and incubated for 1 h on an orbital shaker. (Antibody solution; 15 ml 1xTBS, 0.15 g BSA (bovine serum albumin) and 15 µl of antibody (A7058-1VL monoclonal anti polyhistidine peroxidase antibody, Sigma Aldrich). The nitrocellulose membrane was washed 3 times with 1XTBS for 5 min. Finally, the blot was probed by adding 1 ml of Tetramethyl Benzidine (TMB) until sufficient color was developed. The membrane was rinsed with sterile water and the image of the blot was taken using Canon CanoScan 9000F Mark II scanner.

2.11. The β -Galactosidase (BAD_1582) enzyme characterisation

2.11.1. β -galactosidase activity assay

The β -galactosidase activity of BAD_1582 was determined using the chromogenic substrate *ortho*-nitrophenyl- β -D-galactopyranoside (*o*NPG). The assay was carried out in a microtiter plate containing 8.75 mg ml⁻¹ purified enzyme, 2 mM ONPG in 50 mM

sodium phosphate buffer in a final volume of 100 μ l, at pH 7.0 and 37°C for 30 min. The reaction was stopped by addition of 100 μ l 1M sodium carbonate (final conc. 500 mM). The hydrolysis product *ortho*-nitrophenol was detected by measuring absorbance at 420 nm (Biotek microplate spectrophotometer). A unit of an enzyme activity (1U) is defined as the amount of purified enzyme used to liberate 1 μ mol of *o*NP from the *o*NPG substrate per min in 50 mM sodium phosphate buffer (pH 7.0) at 37°C.

2.11.2. Effects of pH, temperature, divalent cations and denaturants / detergents on the activity and /stability of BAD_1582

The optimum pH for enzyme activity was determined by using *o*NPG as a substrate in sodium phosphate buffer (50 mM) and Z-buffer (Na₂HPO₄, 60 mM; NaH₂PO₄.2H₂O, 40 mM; KCl, 10 mM, MgSO₄.7H₂O, 1 mM; 2-mercaptoethanol (BME), 50 mM) set at pH ranges of 4.0 to 10.0. The pH stability of the enzyme was determined by pre-incubating the enzyme in 50 mM sodium phosphate buffer set at pH 4.0-10.0 for 24 h at 4°C. The pH of the pre-incubated enzyme was readjusted to pH 7.0 before the β -galactosidase activity were measured.

The effect of temperature on the enzyme activity was determined at different temperatures (0 -60°C) in 50 mM sodium phosphate buffer at pH 7.0. The temperature stability of the enzyme was determined by preincubating the enzyme solution in sodium phosphate buffer at those set temperatures for 1 h, after which the enzyme was brought to 37°C and β -galactosidase activity was determined.

The effect of metal ions on the activity of the enzyme was determined in 50 mM sodium phosphate buffer (pH 7) and Tris-HCl buffer (pH 7). Initially, the effect of Na⁺ (NaCl), K⁺ (KCl), Mg²⁺ (MgCl₂) on the activity of the enzyme were determined in sodium phosphate buffer at final concentrations of 10 mM. Secondly, as most of the divalent cations were insoluble in phosphate buffer, 100 mM Tris-HCl buffer was used to solubilize those cations and hence their effect determined. The following metal ions CaCl₂, MgCl₂, ZnCl₂, FeCl₃, CuSO₄, and MnCl₂ at a final concentration of 10 mM were dissolved in 100 mM Tris-HCl buffer (pH 7) and their effect on the activity of the enzyme determined. To determine the effect of EDTA on the activity of BAD_1582 the enzyme

was pre-incubated in the presence of EDTA (final concentration of 10 mM) in 50 mM sodium phosphate buffer (pH 7), for 3 h at 4°C. Controls of the same amount of enzyme incubated in sodium phosphate buffer containing 10 mM MgCl₂ and in sodium phosphate buffer with no additional divalent metal ions were similarly prepared. Then, the β -galactosidase activity was determined.

The effects of detergents and denaturants on the activity of BAD_1582 were determined by performing the β -galactosidase assay with the enzyme in 50 mM sodium phosphate buffer/10 mM MgCl₂ (pH 7) supplemented with the following reagents; sodium dodecyl sulfate (0.1, 0.5 and 1%), urea (0.1, 0.5, and 1 M), Triton X-100 (0.1, 0.5, and 1%), 2-mercaptoethanol (1, 10 and 50 mM).

2.11.3. Substrate specificity and kinetic parameter

For substrate specificity test determination chromogenic nitrophenyl based substrates were used at a final concentration of 2 mM in 50 mM sodium phosphate buffer/10 mM MgCl₂ (pH7), at 37°C. The nitrophenyl substrates used were *p*NP- β -D-galactopyranoside (*p*NPG), *p*NP- β -D-glucopyranoside (*p*NP- β -D-Glc), *p*NP- α -D-galactopyranoside (*p*NP- α -D-Gal), *p*NP- α -D-mannopyranoside (*p*NP- α -D-Man), *p*NP- α -L-fucopyranoside (*p*NP- α -L-Fuc) (Toronto Research Chemicals (TRC)), *p*NP- α -L-xylopyranoside (*p*NP- α -L-Xyl) (TRC), *p*NP-2-acetamido-2-deoxy- β -D-galactopyranoside (*p*NP-GalNAc) (TRC) and *p*NP-2-acetamido-2-deoxy- β -D-glucopyranoside (*p*NP-GlcNAc) (Carbosynth) and *o*NP- β -D-galactopyranoside (*o*NPG).

The kinetic parameters of BAD_1582 were determined by measuring the β -galactosidase activity at different concentrations of *o*NPG (0.25, 0.5, 0.75, 1, 2.5, 5, 10, 20 mM) in 50 mM sodium phosphate buffer (pH 7) at 37°C in a time course discontinuous enzyme assay monitored for 20 min. Aliquots were withdrawn at 2 min intervals and reaction was stopped by addition of 500 mM sodium carbonate and absorbance 420 nm were recorded. The V_{\max} and K_m were calculated using the enzyme kinetics module tool of Sigma Plot software version 14.0 (Systat Software, San Jose, CA) from initial velocity *versus* substrate concentration plot. Both Michaelis-Menten kinetics and Lineweaver-burke plot transformation were implemented to determine the kinetic parameters.

2.11.4. Inhibition assays using lactose, galactose and glucose

In order to investigate whether BAD_1582 is inhibited by lactose, galactose and glucose, β -galactosidase activity was determined in an assay containing 0.25-75 mM lactose, 5-700 mM galactose or 5-700 mM glucose in the presence of 1 mM colorimetric substrate *o*NPG and the enzyme BAD_1582.

A time course enzyme assay was carried out in the presence of lactose (0.75-30 mM) or galactose (10-700 mM) and 1mM *o*NPG to determine the K_i of lactose and galactose. The assays were monitored for 20 min, aliquots were withdrawn at 2 min intervals after reactions were stopped and absorbance at 420 nm recorded. The rate of the enzyme was calculated from progress curves of these time course assay. The K_i of lactose and galactose were calculated by applying competitive inhibition kinetics using sigma plot enzyme kinetic module.

2.11.5. Degalactosylation activity of BAD_1582 on glycoproteins

The degalactosylation activity of BAD_1582 was carried out using asialofetuin as a substrate, fetuin was used as a control. A final reaction volume of 10 μ l contained; 5 μ l of asialofetuin / or fetuin (1 mg/ml stock was prepared in water), 4 μ l of phosphate buffer supplemented with 20 mM $MgCl_2 \cdot 6H_2O$, and 1 μ l of enzyme (BAD_1582, 0.0875 μ g / μ l stock in PBS). A negative control without enzyme/ instead sterile water was prepared for both substrates. Two sets of triplicate reactions for both substrates were prepared. After 1 h and 8 h 37°C. an equal volume of 1x Laemmli loading buffer were added to each sample and incubated at 85°C in a heat block for 5 min. Samples were analyzed using SDS-PAGE (section 2.10.2).

2.11.6. Transgalactosylation activity assay

Transgalactosylation activity of BAD_1582 was analysed by three approaches in 50 mM sodium phosphate buffer/10 mM $MgCl_2$ with 1.5 units/ml enzyme in a volume of 0.2 ml at pH 7.0 and 37°C. The first approach was carried out using *p*NPG as galactosyl donor and various saccharides as acceptors such as fructose, L-arabinose, GlcNAc, galactose, glucose and lactose. The assay contained 100 mM *p*NPG (25 mM *p*NPG when lactose used as acceptor) and 200 mM acceptor sugar. In the second approach, lactose was used

both as galactosyl donor and acceptor in which the assay contained 234 mM lactose. In the third approach, lactose was used as a galactosyl donor and L-fucose and N-acetylneuraminic acid (Neu5Ac) were used as acceptors. The assay contained 200 mM lactose and 100 mM acceptor sugars. A control without enzyme was prepared for each of the transgalactosylation assays. All the aforementioned assays were carried out for 24 h. After 24 h, the enzyme was inactivated by heating at 95 °C for 5 min. Samples were stored at -20 °C, until they were analysed by thin layer chromatography (TLC).

TLC was carried out using silica gel 60G F₂₅₄ glass/ aluminium foil plates as stationary phase. Samples and standards (1 µl) were spotted on the plates and the plates were immersed on a mobile phase of n-propanol: distilled water: ethyl acetate (7:2:1 v/v/v). The standard sugars were prepared in 50 mM sodium phosphate buffer (pH 7.0) with same concentration as transgalactosylation assays. Standard sugars used were commercial GOS (8%, Carbosynth), lactose, glucose, galactose, fructose, L-arabinose, GlcNAc, L-fucose, Neu5Ac and pNPG (75 mM). After separation, the TLC plates were sprayed with 10% sulfuric acid in ethanol, allowed to completely dry and then charred at 105°C for 5 min to visualise migration. Retention factor (R_f) values for each spot of transgalactosylation products were calculated by taking the ratio of distance travelled by each spot to distance travelled to solvent front and compared to the R_f values of the standard sugars.

The synthesis of GOS using BAD_1582 were monitored in a time dependent assay with 234 mM lactose. An aliquot of 6 µl were withdrawn from each triplicate at 4 h, 8 h, 24 h, 48 h and 72 h and kept frozen. The aliquots from each time point were analysed using TLC, a commercial GOS (Carbosynth), lactose, glucose and galactose were used as a standard.

2.12. Immobilization of BAD_1582

Immobilization of Ni²⁺-NTA column purified BAD_1582 was carried out in sodium alginate and gelatin beads cross linked with glutaraldehyde (Freitas *et al.*, 2012). First 6.6% of sodium alginate and 4.05% of gelatin (gelatin from bovine skin, Type B powder) were weighed out. Sterile water (7.935 ml) was added to bring the volume to 9 ml, the solution was mixed by vortexing and heating in a water bath to 80°C, clumps were removed with continuous stirring, until the solution was homogenous and maintained a

jelly consistency. The solution was brought back to a temperature below 40°C, kept in ice for 60 seconds and the purified enzyme (1ml, 37 units ml⁻¹) were transferred to the solution and gently mixed. At room temperature the enzyme containing mixture was pumped at constant speed 8 rpm using Watson Marlow peristaltic pump, with 0.85mm diameter peristaltic tube (Watson Marlow), in to a 10 ml of 3.64% glutaraldehyde and 0.05M CaCl₂ solution placed in a magnetic stirrer for the formation of immobilized enzyme beads. The beads were maintained in the glutaraldehyde and CaCl₂ solution for 12 h at 4°C. After the beads were washed with 50 mM phosphate buffer (pH 7) and maintained in this buffer at 4°C until used for the enzymatic assay. Commercial β -galactosidase from *Aspergillus oryzae* (sigma) solution were prepared in sterile water to bring the specific activity 37 units ml⁻¹ and then immobilized in a similar fashion as that of BAD_1582, and the beads were washed with 50 mM acetate buffer (pH 4.5) . Negative control beads without any added enzyme were prepared in a final volume of 10 ml.

2.12.1. The β -galactosidase activity and reuse of immobilised BAD_1582 and *Aspergillus oryzae* β -galactosidase

The β -galactosidase activity of immobilized BAD_1582 was determined in an assay containing a final volume of 0.5 ml, 0.225 ml of 50 mM sodium phosphate buffer and 2 mM *o*NPG (0.025ml of 20mM, stock prepared in distilled water) at pH 7 and at 37°C for 30 min. The assay was stopped with 500 mM sodium carbonate and absorbance at 420 nm was measured using Biotek spectrophotometer and specific activity of the enzyme was determined. The β -galactosidase activity was also determined using lactose as a substrate, at pH 7 and at 37°C for 30 min the assay was stopped by deactivating the enzyme at 95°C for 5 min and the amount of released glucose was quantitated using glucose-oxidase/peroxidase assay kit.

The reuse of immobilised BAD_1582 and commercial β -galactosidase were conducted in 50 mM phosphate buffer/10mM MgCl₂ (pH 7) and acetate buffer (pH 4.5) respectively. The reuse activity of immobilised enzyme was conducted for 12 rounds with the same amount of beads and substrate except, the same beads were reused following stringent washes (4x) with buffer (50 mM sodium phosphate buffer for BAD_1582 and acetate buffer for β -galactosidase) to remove residual *o*NP from each assay.

2.12.2. Effect of pH, temperature and kinetic parameters of the immobilised BAD_1582

The effect of pH on β -galactosidase activity of the immobilized BAD_1582 was determined using *o*NPG as a substrate in 50 mM sodium phosphate buffer adjusted from pH ranges 4.5-10. To determine the pH stability of the immobilised BAD_1582, the beads were pre-incubated in 50 mM sodium phosphate buffer adjusted in pH ranges 4.5-10 for 24 h at 4°C, then, the residual activity was determined by bringing back the beads to pH 7 at 37°C.

The effect of temperature on the activity of immobilised enzyme was determined by incubating the assay at different temperature (0-60°C), for temperature stability immobilized beads were pre- incubated at temperatures (0-60°C) in 50mM sodium phosphate buffer pH 7 for 1 h, then the residual activity of immobilised enzyme was determined bringing back the beads at 37°C at pH 7. For comparison, the effect of pH and temperature on the β -galactosidase activity of free enzyme (equivalent concentration with the immobilised enzyme) was also determined alongside with the immobilised enzyme, in a final volume 0.1 ml, containing 2 mM *o*NPG.

The kinetic parameter of the immobilised BAD_1582 was determined from time course enzyme assay containing 2 mM *o*NPG in 50 mM sodium phosphate buffer at pH 7 and at 37°C monitored for 20 min. The kinetics parameters were calculated using Sigma plot.

2.13. Statistical analysis

Three independent experiments were conducted in triplicate and the graphs were plotted from the mean of the triplicates. The error bars in the graphs represent the standard deviation (SD) (mean \pm SD). Statistical significance was determined by using Student's t-test, significance was represented with *p*-values $< 0.001 = ***$; $< 0.01 = **$ and $< 0.05 = *$.

Chapter 3

Validation of the human faecal microbiome metagenome library and functional screening for GH enzymes

3.1. Introduction

One of the major hurdles presented in microbiological studies was the difficulty to cultivate representative species of microbes using the existing cultivation techniques. Studies have indicated that only <1% of the microorganisms in any habitat are cultivable (Amann *et al.*, 1995). This was clearly described by the great plate count anomaly, a phenomenon in which the number of cultivated microbial cells is inconsistent with the number of cells counted under a microscope (Staley and Konopka 1985). Different cultivation techniques have been implemented to enhance the cultivation of previously uncultivated microorganisms. Most of these cultivation methods mimic the natural environment of the microorganisms. Nevertheless, some natural conditions are difficult to mimic under laboratory settings, for instance, the chemistry of the microorganism's natural habitat, diversity of microbes in a particular niche/habitat, seasonal changes, interactions of the biotic and abiotic factors (Pham and Kim 2012). Hence through cultivation methods, an insufficient number of microbial species are isolated, and the majority of the microbial species are either underrepresented or absent.

The advancement of molecular techniques such as 16S rDNA sequencing, whole genome sequencing and functional metagenomics played a significant role in accessing genomes of uncultivated microbes from a wide range of habitats. The term functional metagenomics was first coined by Handelsman in 1998 (Jo Handelsman 1998). One methodology of functional metagenomics entails the isolation of representative microbial community DNA from a given habitat, cloning sections of this DNA into large vector constructs such as BACs and expression in a heterologous host for identification of novel functionalities. Functional metagenomics had become a well-established method in the identification of novel genes and enzymes or natural products from microorganisms residing in a range of habitats, such as soil, marine sediments and the human gut (Calderon *et al.*, 2019; Cowan *et al.*, 2005; Hardeman and Sjoling 2007; Langer *et al.*, 2006; Matsuzawa and Yaoi 2017). Functional metagenomics enabled access to novel genes and novel enzymes of industrial importance from uncultivable bacteria and archaea that reside in unexploited environments.

In the past two decades a considerable amount of research has been carried out to understand the microbial diversity and the metabolic function of the microbiota residing in the human gut. 16S rDNA sequencing and subsequent phylogenetic analyses of samples collected from mucosal sites of different regions of the colon and faecal samples of healthy human subjects has revealed that the Bacteroidetes and Firmicutes are the dominant bacterial phyla residing in the human gut, with a minority of Actinobacteria while archaea was only represented by a single phylotype; *Methanobrevibacter smithii* (Eckburg *et al.*, 2005; Gill *et al.*, 2006). Deep sequencing of stool sample from 98 individuals have revealed the predominant archaea was *Methanobrevibacter* genus (present in 30 of samples), followed by *Nitrososphaera* spp (detected in 16 samples) and several other less abundant archaea were identified, besides 66 genera of fungi were also identified predominated by *Candida* genus (Hoffmann *et al.*, 2013). Three enterotypes (which belong to the genera of *Bacteriodes*, *Prevotella* and *Ruminococcus*) of human gut microbiota were revealed after phylogenetic analysis of 39 human gut microbiome metagenome samples collected from human subjects in Europe and America (Arumugam *et al.*, 2011). Enterotypes are defined as “densely populated areas in a multi-dimensional space of community composition” and should not be seen as discrete clusters, but as a way of stratifying samples to reduce complexity (Arumugam *et al.*, 2011). Each individual human being harbours its own unique microbial species and strains, however diversity and composition of the microbiome also varies from each ecological niches in the human body, with maximum diversity detected in the human gut (Backhed *et al.*, 2005).

The most intriguing human gut microbiome gene catalogue was obtained after Illumina metagenome sequencing of faecal samples collected from 124 human subjects (healthy adults, overweight, obese and Inflammatory bowel disease (IBD) patients) (Qin *et al.*, 2010). The result demonstrated that the human gut microbiome gene catalogue contained 3.3 million microbial genes. 99.1% of the genes were of bacterial origin, 0.8% were archaeal and 0.1% of eukaryote and viral origin. An enormous number of genes encoded by the microbiota are lacking in the host, and are involved in the synthesis of essential amino acids and vitamins, hydrolysis of non-digestible complex carbohydrates, breakdown of xenobiotics, regulation of host immune system etc. (Jourova *et al.*, 2019;

Kaoutari *et al.*, 2013). Therefore, the gut microbiome is considered to be the second genome of the human host (Zhu *et al.*, 2010). The human gut microbiota plays an enormous role in the maintaining homeostasis and health of the human host through affecting metabolism, modulation of immune system development and inhibition of pathogen colonization (Bull and Plummer 2014; Wardwell *et al.*, 2011). It has been demonstrated that an imbalance in the composition and diversity of the microbiota (dysbiosis) is linked to several metabolic disorders, such as type II diabetes (Patterson *et al.*, 2016), Inflammatory bowel disease (Lloyd-Price *et al.*, 2019; Oligschlaeger *et al.*, 2019), asthma (Abrahamsson *et al.*, 2014), and autism spectrum disorders (Liu *et al.*, 2019a). Besides, a shift in the composition of the microbiota that resulted in a high ratio of Firmicutes to Bacteroidetes phyla has been associated with obesity and *vice versa* (Ley *et al.*, 2006).

The microbiota in the gut obtain their energy sources from dietary non-digestible carbohydrates, that pass the upper digestive tract such as plant material rich cellulose and lignocelluloses, resistant starch and short chain oligosaccharides (Flint *et al.*, 2008). The microbiota encode a repertoire of carbohydrate active enzymes (CAZymes) to deconstruct the non-digestible carbohydrates. The CAZymes which break down carbohydrates are classified into four classes; Glycoside hydrolases (GHs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs) and very new class of CAZymes called auxiliary activities (AAs) are also created which function in breakdown of lignocellulosic polysaccharides (Lombard *et al.*, 2014a). Besides the dietary carbohydrates, some members of the gut microbiota source, their energy from host-derived O- linked and N-linked glycans, which are principal components of the mucus secretions (Marcobal *et al.*, 2013) and part of shed epithelial cells respectively (Koropatkin *et al.*, 2012).

The fermentation products of the non-digestible carbohydrates produce short-chain fatty acids (SCFA) such as acetate, butyrate, and propionate, which serve as nutrients for colonocytes and epithelial cells (Koropatkin *et al.*, 2012). Besides the SCFAs serve as effector molecules in a signal transduction cascade related to metabolism (Trent and Blaser 2016), neural function through (gut microbiota-brain axis) (Dinan and Cryan 2017).

Having a repertoire of glycoside hydrolase enzymes enable members of the microbiome to succeed in the human gut environment where the major carbon sources are plant derived dietary oligo and polysaccharides that transit the upper digestive tract, and host derived glycans. The members of the human gut microbiome such as *Bifidobacterium* and *Bacteriodes* encode wide array of glycoside hydrolase enzymes (Pokusaeva *et al.*, 2011; Wexler and Goodman 2017) Herein a human faecal metagenomic library was constructed and used for screening clones which encode Glycoside hydrolase enzymes, particular interest was given to sialidases, α -fucosidases and β -galactosidases.

3.2. Validation of the human faecal microbiome metagenome library using restriction pattern analyses

The human faecal microbiome metagenome library that was previously constructed was used as a starting material for this study (Christy Agbavwe 2017). Fosmids from this library were extracted and transferred to a new heterologous EPI-300T1^R host. The initial library was calculated to contain 42,000 clones with each clone incorporating an insert size of 25-45 kb (Christy Agbavwe 2017). Hence an equivalent amount of clones with these insert sizes were generated with the new heterologous host. Twenty random clones were selected to predict the diversity generated in the metagenome clones. Fosmids were extracted and subjected to restriction endonuclease digestion (*Bam* HI and *Sma* I) and the diversity among the clones was examined by the restriction pattern generated on 1% agarose gel electrophoresis. The pCC1FOS vector (8,139 bp size) has two restriction sites for *Bam* HI (at positions 354 and 366 bp on the vector map) and three restriction sites for *Sma* I (at positions 351, 652 and 3492 bp) (Fig 3.1). The cloned insert in the pCC1FOS vector is located at position 361, and the bands of the pCC1FOS vector (8,139 bp) after restriction enzyme digestion (8127 bp and 12 bp for *Bam* HI and 351 bp, 301 bp, 2,840 bp and 4,647 bp for *Sma*I) were used as a reference and the presence of insert DNA in the metagenome clones were revealed by the presence of different restriction fragments other than the vector bands.

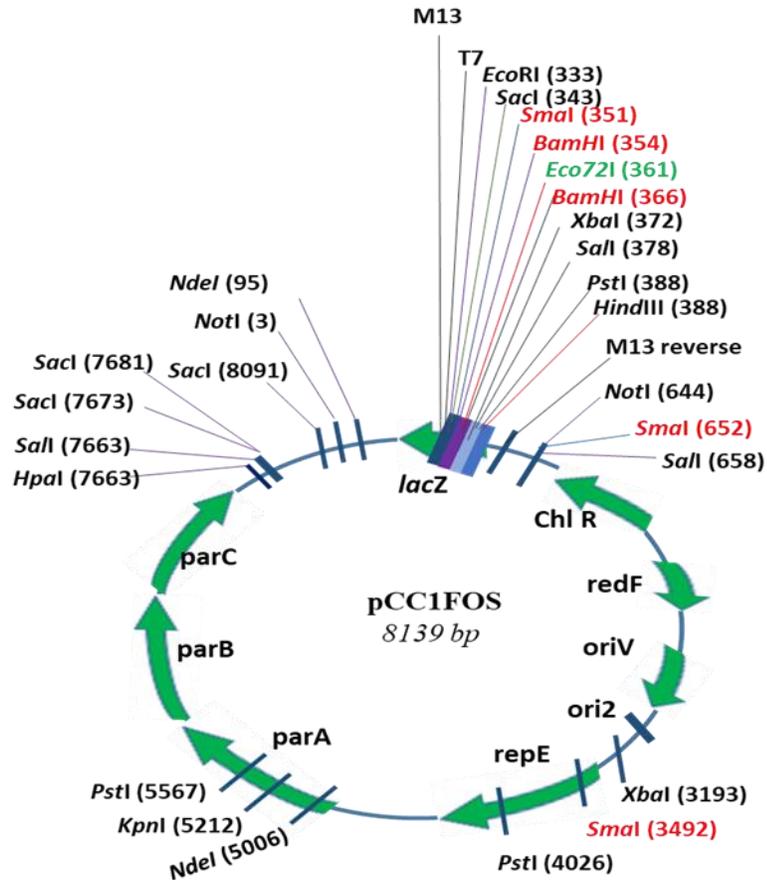


Fig 3.1 Restriction map of pCC1FOS vector, the restriction enzymes used indicated in red. This picture was adapted from *Medicago truncatula* genomic resource.

It was expected that metagenome clones which harboured different inserts of the faecal microbiome DNA could show different restriction fragments. Hence the more diverse the metagenomic library the more diverse restriction fragments that would be obtained from the fosmid inserts after restriction endonuclease enzyme digestion. However, the absence of restriction sites within the fosmid inserts for the target restriction enzymes would result in exactly similar restriction patterns showing only the vector bands and the undigested insert DNA. The restriction pattern of the twenty clones revealed that clone 2 and 3 had the same restriction pattern after digestion by each of the two restriction endonucleases which indicates ~10 % of the clones could presumably be identical clones (Fig 3.2). Clone 14 had only vector bands and hence ruled out from the further analysis (data not shown). The remaining clones have distinct restriction patterns with either one or both of the

restriction enzymes, and presumably, these are distinct metagenome clones (Fig 3.2). Therefore, the restriction analysis implies that 85% of the metagenome library contains distinct clones and hence the metagenome library is diverse. To confirm the diversity observed via restriction pattern analysis, the nineteen clones were end-sequenced and bioinformatic analyses were carried out to uncover the underlining diversity of the metagenome library.

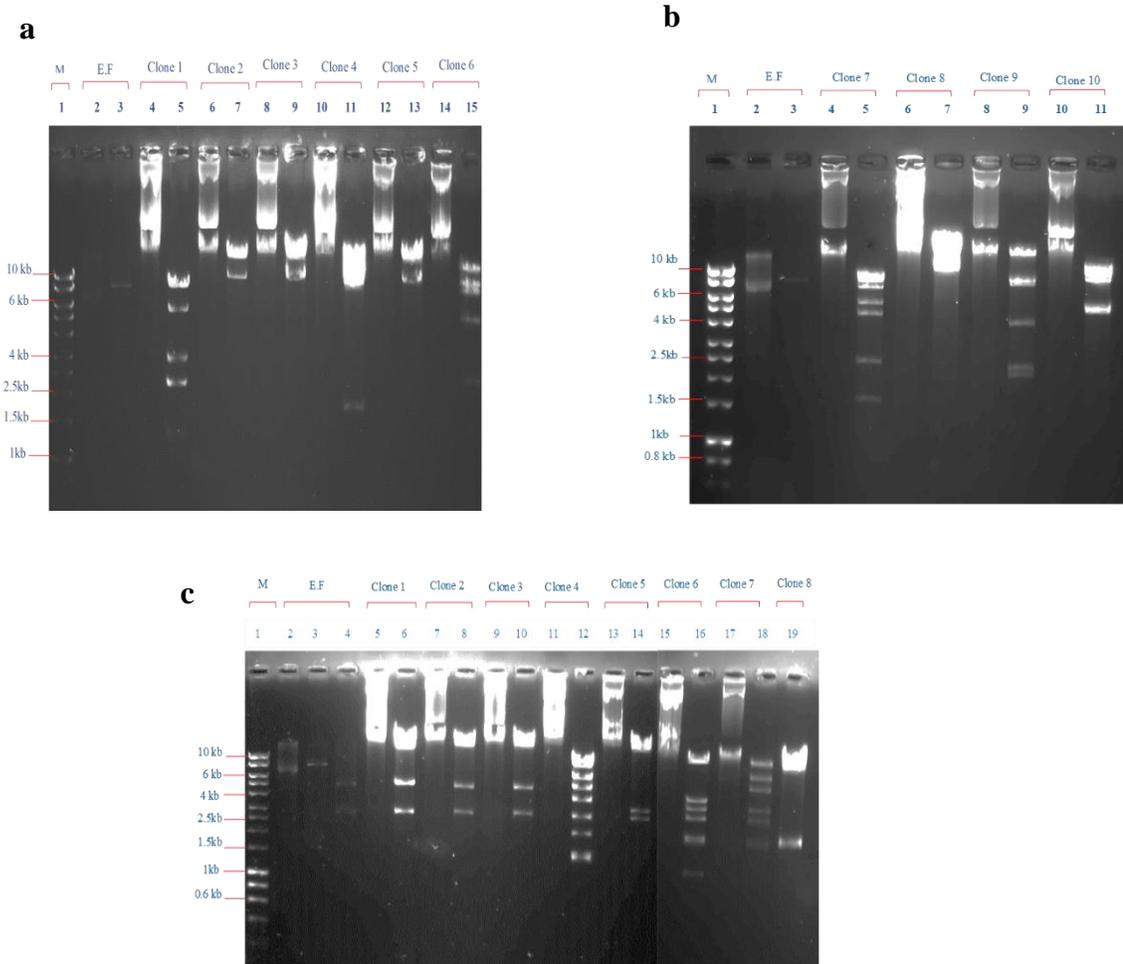


Fig 3.2 Restriction analysis of 10 randomly selected metagenome clones and the vector (pCC1FOS) before and after *Bam*HI (**a, b**) and *Sma* I (**c**) restriction digestion. Lane 1: 1kb molecular ladder (M), even-numbered lanes: undigested fosmid, odd-numbered lanes: digested fosmid.

3.3. Validation of the human faecal microbiome metagenome library through end sequencing

The diversity of the metagenome library is a key determining factor for functional metagenomics studies. As long as the heterologous host has the expression machinery for expression of the target genes, a more diverse metagenome library would enhance the likelihood of discovering novel genes. To confirm the diversity of the constructed human faecal metagenome library, the nineteen random clones were end-sequenced using pCC1FOS forward and reverse primers (Epicentre). Bioinformatics analysis of the sequenced clones was carried out to identify homologue sequences in the Genbank database. From the sequencing reactions of 19 random clones, the sequences generated for four clones (2, 3, 5, and 8) were either too short or matched with *E. coli* and the sequence for pCC1FOS was not found. Thus they were regarded as failed sequencing reactions and disregarded from further analysis. The blast record of the successfully sequenced clones indicated, of the 15 clones, ten clones (67%) matched with *Bifidobacterium adolescentis*. Eight clones (2, 6, 7, 15, 16, 17, 19, and 20) had high identity with *B. adolescentis* at both the 5' and 3' ends of the inserts (Table 3.1). The forward sequence of clone 18 matched with *B. adolescentis* (99%) while no significant similarity was detected on the reverse sequence. The forward sequence of clone 11 matched with *Oscillibacter* sp. (96% identity) while the reverse sequence matched with *B. adolescentis* (99% identity). The inserts in these clones may have originated from *B. adolescentis* strains that differ from the sequenced reference strain by the acquisition of additional DNA by horizontal transfer. Each of these 10 inserts mapped to a different chromosomal location. One clone matched with *Bacteriodes fragilis* (clone 12) with 99% and 100% identity for forward and reverse sequences respectively (Table 3.1).

Two clones (13%) matched with *Gordonibacter pamelaee* with 80% identity (clone 10) and with *Collinsella aerofaciens* with 94% identity (clone 13) on reverse and forward sequences respectively; but no significant similarity was found on the forward and reverse sequences respectively of these two clones. (Table 3.1). On the other hand clone 9 had similarity with a different bacterial species at each end of the insert. The forward sequence of clone 9 had similarity with *Bifidobacterium bifidum* (83% identity) while the reverse sequence had a match with *Olsenella uli* (73% identity). Clone 1 matched *B. adolescentis*

however the % identity was only 77% and 95% for the forward and reverse sequences, respectively. The % identity detected in these 4 clones was relatively low and the absence of consistent matches at both end of the inserts presumably indicates the inserts originate from bacterial species that have not been sequenced (Table 3.1).

Table 3.1. Blast record of randomly picked metagenome clones

Clones/ Sequence	Species	% identity	BP length of similarity
C#1 forward	<i>Bifidobacterium adolescentis</i>	77%	846
C#1 reverse	<i>Bifidobacterium adolescentis</i>	95%	844
C# 4 forward	<i>Bifidobacterium adolescentis</i>	99%	795
C# 4 reverse	<i>Bifidobacterium adolescentis</i>	100%	520
C# 6 forward	<i>Bifidobacterium adolescentis</i>	98%	816
C# 6 reverse	<i>Bifidobacterium adolescentis</i>	99%	825
C#7 Forward	<i>Bifidobacterium adolescentis</i>	99%	885
C# 7 reverse	<i>Bifidobacterium adolescentis</i>	100%	793
C# 9 Forward	<i>Bifidobacterium bifidum</i>	83%	605
C# 9 Reverse	<i>Olsenella uli</i>	75%	579
C# 10 Forward	<i>No significant similarity found</i>		
C# 10 Reverse	<i>Gordonibacter pamelaee</i>	80%	440
C# 11 Forward	<i>Oscillibacter sp.</i>	96%	722
C#11 Reverse	<i>Bifidobacterium adolescentis</i>	99%	769
C#12 Forward	<i>Bacteroides fragilis</i>	99%	758
C#12 Reverse	<i>Bacteroides fragilis</i>	100%	744
C#13 Forward	<i>Collinsella aerofaciens</i>	94%	773
C#13 Reverse	<i>No significant similarity found</i>		
C#15 Forward	<i>Bifidobacterium adolescentis</i>	99%	799
C#15 Reverse	<i>Bifidobacterium adolescentis</i>	99%	778
C#16 Forward	<i>Bifidobacterium adolescentis</i>	99%	763
C#16 Reverse	<i>Bifidobacterium adolescentis</i>	99%	479
C#17 Forward	<i>Bifidobacterium adolescentis</i>	100%	678
C#17 Reverse	<i>Bifidobacterium adolescentis</i>	92%	589
C#18 Forward	<i>Bifidobacterium adolescentis</i>	99%	709
C#18 Reverse	<i>short sequence</i>		

C#19 Forward	<i>Bifidobacterium adolescentis</i>	100%	783
C#19 Reverse	<i>Bifidobacterium adolescentis</i>	99%	727
C#20 Forward	<i>Bifidobacterium adolescentis</i>	99%	728
C#20 Reverse	<i>Bifidobacterium adolescentis</i>	99%	764

In summary, the bioinformatics analyses of the sequenced clones indicated that the human faecal metagenomic library contained DNA sequences which belong to the major phylotypes of the human gut microbiota. The predominant members of the gut microbiota were detected from a small number of randomly picked clones of the library. The library is dominated by Actinobacteria, represented by genus *Bifidobacterium* (67%), in particular *B. adolescentis*, and possibly *Gordonibacter* and *Collinsella* ; followed by Bacteroidetes represented by *Bacteriodes Fragillis* whereas the Firmicutes phyla was only represented by *Oscillibacter*. Given that the human faecal microbiome metagenomic library was dominated by the Actinobacteria, followed by Bacteroidetes, with proven ability of both phyla to ferment wide range of carbohydrates, this library was suitable for functional screening studies targeting identification of novel Glycoside hydrolase enzymes (GHs).

3.4. Strategy for screening glycoside hydrolase enzymes from human faecal metagenomic library

The gut microbiome is a rich source of CAZymes which metabolise complex glycans. In this study a new approach to screen a human faecal metagenomic library for CAZymes was developed. This approach used MacConkey agar-based screening for glycoside hydrolase (GHs) enzymes. The fundamental principles of MacConkey agar screening were applied with slight modification for screening different GH enzymes, by supplementing with different carbohydrates as a sole carbon source. Positive metagenome clones with active GH enzymes undertook fermentation of the supplemented carbohydrates which subsequently result in reduction of the pH. The reduction of pH to lower than 6.8 caused the colour indicator (neutral red) to change to red/ pink, and thus positive clones would have a red/pink morphology around the colonies. Negative clones

which cannot ferment the supplemented carbohydrates would appear colourless due to the fact that neutral red remains colourless at higher pH.

The human faecal metagenome library was screened for three GH enzymes; β -galactosidases, sialidases, and α -fucosidases using MacConkey agar supplemented with lactose, 3'-sialyllactose (3'-SL) and/or 6'-sialyllactose (6'-SL) and 2'-fucosyllactose (2'-FL) respectively. The latter two are components of the human milk oligosaccharides (HMOs), which specifically promote growth of certain members of the infant gut microbiota, mainly *Bifidobacteria* and *Lactobacilli* (Thongaram *et al.*, 2017a). Besides, the O-linked mucin glycans which have structural similarity with HMOs comprise fucose and sialic acid. Interestingly both the composition of HMOs and mucosal glycans are affected by many factors, such as genetics of the individual. For instance people that have a functional *FUT2* (gene encoding α 1-2 fucosyl transferase) or *FUT3* (gene encoding α 1-3 fucosyl transferase) called “secretors” possess L-fucose both in their mucin glycans and HMOs (Ferrer-Admetlla *et al.*, 2009). On the other hand, mutants for these transferases called “non-secretors” possess more sialylated glycans, presumably due to lack of competition for glycosylation availability for glycosyltransferases (Bennett *et al.*, 2012). Mucin adapted members of the microbiota such as *Bacteroides* express a repertoire of CAZymes which hydrolyse these mucin glycans and HMOs via mucin utilisation pathways (Marcobal *et al.*, 2011).

In vitro studies have implicated that prominent members of the infant gut microbiota such as *Bifidobacterium longum*, *Bifidobacterium kashiwanohense* and *Bifidobacterium bifidum* utilise HMOs such as 2'-FL, 3'-SL and 6'-SL as carbon sources (Bunesova *et al.*, 2016). Thus the notion of screening the human faecal microbiome library for HMO utilisation was supported by such previous findings and hence we designed agar based and growth based strategy for screening of the library for clones which encode the aforementioned GH enzymes. Positive clones from each enzymatic screening were confirmed using LB agar supplemented with chromogenic substrates, X-Gal and X-Fuc for active β -galactosidases and α -fucosidases, respectively. Fluorescent substrate 4MU-NANA supplemented agar media was used for confirmation of sialidase activity.

3.5. Screening of human faecal metagenome libraries for β -galactosidases

A vial of the stock metagenome library was diluted and 100 μ l containing ~500 colonies per plate were seeded on MacConkey lactose agar and incubated overnight at 37°C. This number was chosen to screen large proportion of the metagenome library with small amount of substrates as the latter were expensive for large scale usage. Through this agar-based screening 16,000 clones were screened, and 32 positive clones (~0.2%) were identified by their red colony morphology (Fig 3.3a). The 32 clones were streaked on LB agar supplemented with X-Gal to confirm an active β -galactosidase. 30 of the clones had a blue morphology, which confirm these clones possess an active β -galactosidase enzyme. The two other clones were ruled out from further characterisation. Fig 3.3 b and c depicts two representative β -galactosidase clones (clone 11 and 31) streaked on MacConkey lactose agar and LB agar supplemented with X-Gal together with the *lacZ* positive *E. coli* strain BL21(DE3) and *lacZ* *E. coli* host strain EPI300 (pCC1FOS). The two clones showed strong β -galactosidase activity revealed by the high intensity of pink and blue colony morphology on MacConkey agar and X-Gal LB agar media, respectively.

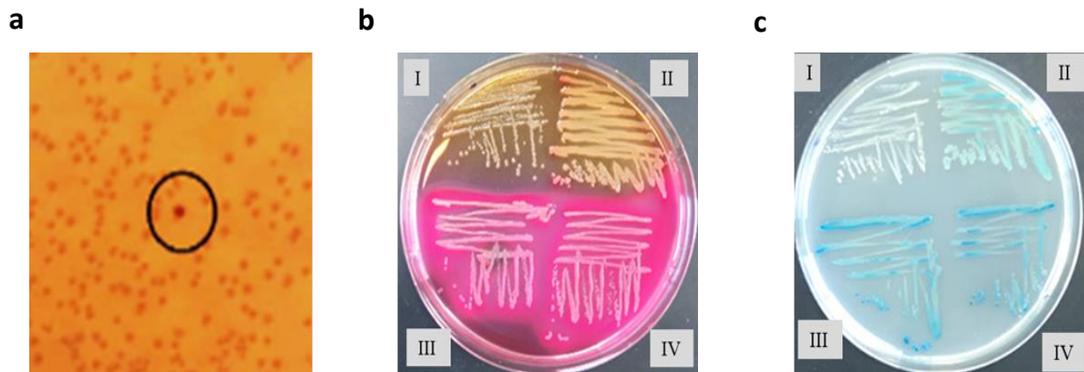


Fig 3.3 Selection of β -galactosidase expressing clones from human faecal microbiome metagenome library. **a** β -galactosidase positive expressing clone with pink colony morphology. **b** EPI300(pCC1FOS) (I), BL21 (DE3) (II), clone#11 (III), and clone#31 (IV) on MacConkey lactose agar. **c** EPI300(pCC1FOS) (I), BL21 (DE3) (II), clone#11 (III), and clone#31 (IV) on X-gal supplemented LB agar.

The Miller assay was conducted on the β -galactosidase positive metagenome clones and appropriate positive control (*lacZ* positive *E. coli* BL21 (DE3) and negative control (*lacZ* *E. coli* EPI300 (pCC1FOS)) strains. High β -galactosidase activity was detected for the β -gal clones compared to *lacZ* positive *E. coli* BL21 (DE3) and the negative control *lacZ* *E. coli* host strain EPI300 (pCC1FOS). The Miller units of 9 sequenced β -galactosidase positive metagenome clones ranged from 618-5795 Miller units while it was 1093 for *lacZ* positive *E. coli* and 36 miller units for the negative control *lacZ* host strain (Fig 3.4).

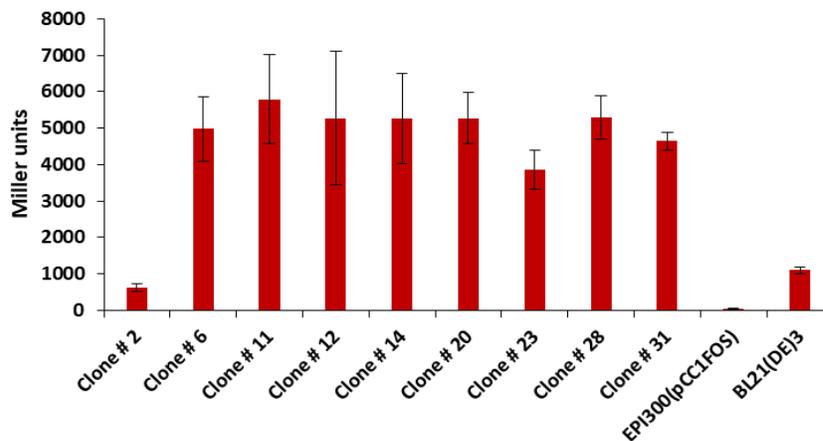


Fig. 3.4 Miller assay on distinct fosmid clones and the *lacZ* positive *E. coli* BL21 (DE3) (positive control) and *lacZ* *E. coli* EPI300(pCC1FOS) (negative control) of 1st round screening.

To determine whether the thirty β -galactosidase positive metagenome clones were distinct clones or not, fosmids were extracted from each clone and characterised by their restriction pattern. Fig 3.5 depicts the restriction pattern after *Bam*HI digestion. The thirty clones fall in to 16 different restriction patterns. The first group comprises clones 4, 7, 11, 15, 18, 21 and 29, the second group includes clones 5, 9 and 14, the third group includes clones 17 and 28, the fourth group includes clones 6 and 10, and the fifth group comprises clones 1 and 13. The remaining 11 clones (2, 3, 12, 16, 19, 20, 23, 27, 30, 31 and 32) had distinct restriction patterns (Fig 3.5).

Representative clones from 9 distinct restriction patterns such as clone 2, 6 (group 4), 11 (group 1), 12, 14 (group 2), 20, 23, 28 (group 3) and 31 were further characterised by end sequencing and bioinformatic analysis (Fig 3.5). The BLAST record of these sequenced clones (except clone 2 and clone 12) had high similarity with *Bifidobacterium adolescentis* (Table 3.2). Clone 2 had no significant similarity in the database, presumably the sequence came from previously uncharacterized bacterial taxa, while the reverse sequence of clone 12 matched with extrachromosomal DNA. The clones which had identical restriction pattern to sequenced clones presumably also contain *B. adolescentis* DNA inserts.

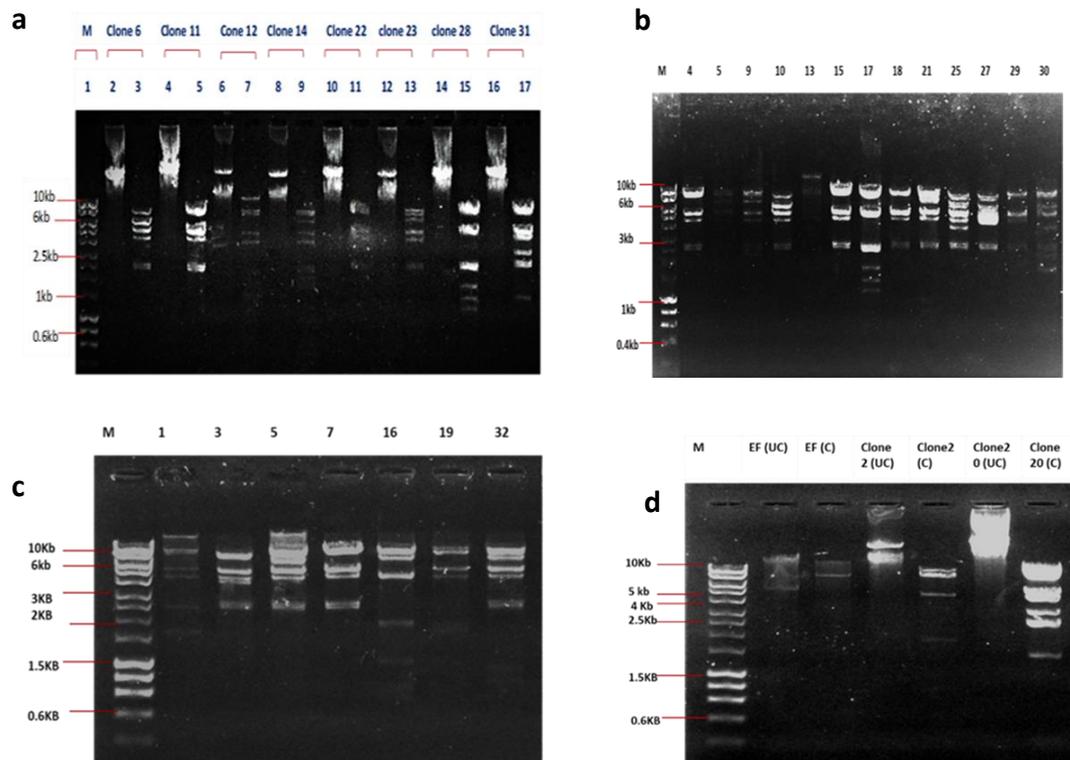


Fig 3.5 Restriction analysis of 8 β -galactosidase positive metagenome clones before and after *Bam*HI restriction endonuclease enzyme digestion. **a** Lane 1, 1 kb molecular ladder (M); Odd numbered lanes, undigested fosmid; even-numbered lanes, digested fosmid. **b** Lane M (1Kb molecular ladder) Lanes 2-13 *Bam* HI digested β -gal positive clones (4, 5, 9, 10, 13, 15, 17, 18, 21, 25, 27, 29 and 30). **c** Lane M (1Kb molecular ladder) Lanes 2-7

Bam HI digested β -gal positive clones (1, 3, 5, 7, 16, 19 and 32) respectively. **d** Lane M (1Kb molecular ladder), Lanes 2, 3 (Empty fosmid uncut and cut), Lanes 4, 5 (clone 2 uncut and cut), Lanes 6 and 7 (clone 20 uncut and cut) respectively.

Table 3.3 Blast record of β -galactosidase positive distinct metagenome clones

Clones/ Sequence	Species	% identity	BP length of similarity
β gal C#6 forward	<i>Bifidobacterium adolescentis</i>	99%	475
β gal C#6 reverse	<i>Bifidobacterium adolescentis</i>	97%	946
β gal C#11 forward	<i>Bifidobacterium adolescentis</i>	99%	734
β gal C#11 reverse	<i>Bifidobacterium adolescentis</i>	97%	844
β gal C#12 forward	<i>Bifidobacterium adolescentis</i>	88%	382
β gal C#12 reverse	<i>Uncultured bacterium extrachromosomal DNA RGI00148</i>	86%	603
β gal C#14 forward	<i>Bifidobacterium adolescentis</i>	93%	208
β gal C#14 reverse	<i>Bifidobacterium adolescentis</i>	91%	891
β gal C#20 forward	<i>Bifidobacterium adolescentis</i>	98%	701
β gal C#20 Reverse	<i>Bifidobacterium adolescentis</i>	86%	908
β gal C#23 forward	<i>Bifidobacterium adolescentis</i>	93%	599
β gal C#23 reverse	<i>Bifidobacterium adolescentis</i>	99%	918
β gal C#28 forward	<i>Bifidobacterium adolescentis</i>	99%	771
β gal C#28 reverse	<i>Bifidobacterium adolescentis</i>	99%	876
β gal C#31 forward	<i>Bifidobacterium adolescentis</i>	98%	696
β gal C#31 reverse	<i>Bifidobacterium adolescentis</i>	98%	696

3.6. Screening of human faecal metagenome library for sialidases

The screening of the human faecal metagenome library for clones which encode sialidase enzymes was carried out on MacConkey base agar supplemented with 3-sialyllactose (3'-SL) and 6'-sialyllactose (6'-SL) as sole carbon sources. First the ability of the host strain *E. coli* EPI300 (*lacZ*) to ferment sialic acid, but not 3'-SL, was determined by cultivating the host strain on two separate plates containing MacConkey base agar supplemented with 0.5% sialic acid or ~1.0% of 3'-SL (molar equivalent of 0.5% sialic acid, ~0.0162M) and incubated overnight at 37°C. Since the addition of sialic acid resulted in a drop of the pH below 7, the pH of the MacConkey agar media supplemented with sialic acid was adjusted to pH 7. The host strain was able to grow on MacConkey base agar supplemented with sialic acid and fermented sialic acid, thus the colonies were pink (Fig 3.6.a). However the host strain was able to grow on MacConkey base agar supplemented with 3'-SL, but did

not ferment 3'-SL and the colonies were colourless (Fig 3.6.b). Thus the MacConkey base agar supplemented with 3'-SL or 6'-SL could be used for screening metagenome clones which encode active exo sialidase enzymes, as positive clones will appear pink after hydrolysis and fermentation of the terminally attached sialic acid of 3'-SL or 6'-SL.

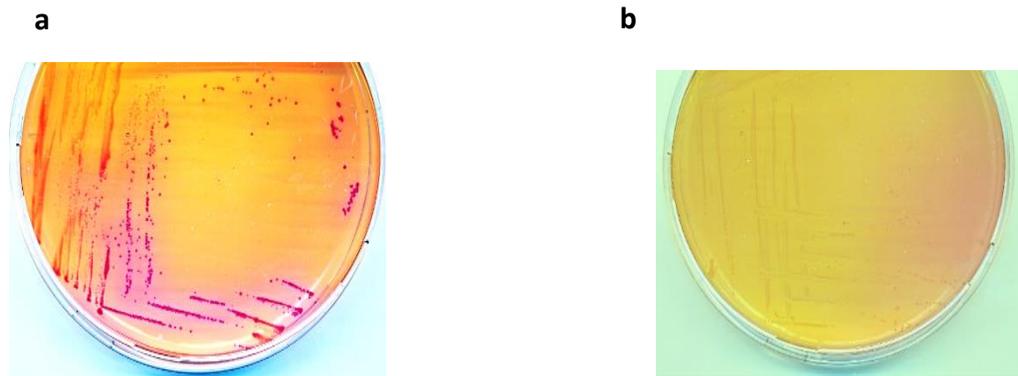


Fig 3.6 *E. coli* EPI300 host strain streaked on MacConkey base agar supplemented with 0.5% sialic acid (a) and supplemented with 3'SL (b).

A total of 16,128 metagenome clones were screened for sialidase (11,000 on MacConkey base supplemented with 3'-SL and 5,128 on MacConkey supplemented with 6'-SL), none of these clones showed sialidase activity demonstrated by absence of pink morphology around the colonies similar to the negative control *E. coli* EPI300 host strain. In summary, the tested clones either did not harbor a sialidase gene, the *E. coli* host did not express the sialidase gene or perhaps the sialidase gene was present and expressed but had not been secreted.

To investigate the validity of our approach for screening of the human faecal metagenome library for sialidase expressing clones using sialyllactoses, a positive control which expresses recombinant sialidase was utilised. The sialidase gene of *Aspergillus fumigatus* had been previously cloned in the pET28a vector to generate the plasmid pAf (Warwas *et al.*, 2010). The enzyme has a predicted mass of 42 kDa, besides protein sequence analysis revealed this enzyme has an extracellular localization. The pAf was used to transform electrocompetent *E. coli* BL21 (DE3) for expression of the sialidase using the host T7 RNA polymerase with IPTG induction. First, the presence of the sialidase gene (~1221bp) in the pAf plasmid was confirmed by restriction endonuclease enzymes *NotI* and *NdeI*.

The restriction products belonging to the pAf sialidase gene (~1221 bp) and plasmid DNA (5369 bp) was obtained, which confirm the pAf harboured the putative sialidase gene of *Aspergillus fumigatu* (Fig 3.7).

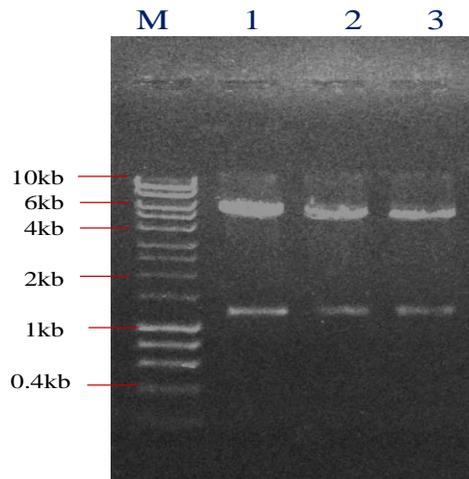


Fig 3.7 Restriction pattern of pAf sialidase after double restriction endonuclease digestion (using *NotI* and *NdeI*). Lane 1, 1 kb molecular ladder (M), lane 2-4 pAf plasmid from three random colonies.

E. coli BL21 (DE3) (pAf), was used as a positive control for sialidase activity. *E. coli* BL21 (DE3) harbouring pAf or the empty pET28b (negative control) were streaked on MacConkey base agar supplemented with 3'SL, in the presence of IPTG (0.05 mM and 0.1 mM) and Kan and incubated overnight at 37°C. In terms of colony morphology the pAf sialidase clone and the pET28b clone were identical, as both appeared pink however the pAf sialidase clone seem to have stunted colony growth, which presumably resulted from over expression of the eukaryotic sialidase upon IPTG induction (Fig 3.8a). In this assay both the positive and negative control displayed pink morphology because the substrates may have small amount of free sialic acid/ lactose which could potentially be hydrolysed by parent strain β -galactosidase.

In another sialidase activity assay a culture of *E. coli* BL2 (DE3) (pAf) and *E. coli* BL21 (DE3) (pET28b) on LB plates was overlaid with 4MU-NANA and illuminated with UV radiation. The pAf clone appeared fluorescent as a result of the hydrolysis of the

terminally attached sialic acid to the fluorescent chemical 4-methyl-umbelliferyl by the expressed sialidase resulting in fluorescence, while the negative control showed no fluorescence as it doesn't express active sialidase (Fig 3.8b). However plating pAf clone and pET28b clone in mixed culture (at a ratio of 1:65) it was not possible to clearly differentiate the pAf clones from the pET28b clones as all the colonies appeared fluorescent. These was presumably the extracellular sialidase enzyme expressed by the pAf clones could eventually hydrolysed the 4MU-NANA which contributed for the fluorescence of the agar, which result in an experimental artefact (Fig 3.8c). Hence screening the faecal metagenome library for extracellular sialidase enzymes with 4MU-NANA was not feasible.

The screening of the faecal microbiome metagenome library for sialidase enzymes, was not successful using MacConkey base agar supplemented with 3'SL/6'SL. The pAf clones were able to express extracellular sialidase and hydrolysed the Neu5AC from the 4 MU-NANA substrates. Nevertheless, it was impossible to carry out the screen of the metagenome library using 4 MU-NANA as substrate, unless separate plates for each clone were used which is cost intensive.

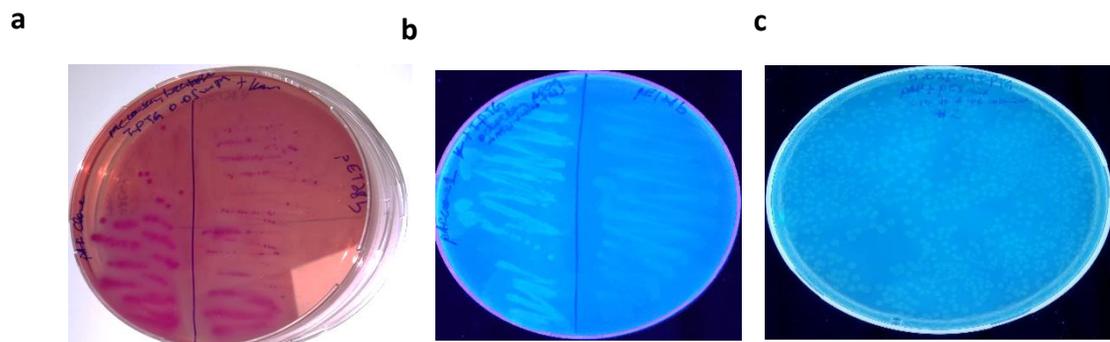


Fig 3.8 *E. coli* BL21(DE3)(pAf)and *E. coli* BL21(DE3) (pET28b)on MacConkey agar supplemented with 3'SL(a) and on LB agar supplemented with 4MU_NANA(b), the positive and negative control mixed (10 to 646) in 4MU-NANA supplemented LB agar(c).

3.7. Screening of human faecal metagenome libraries for α -fucosidases

The screening of the human faecal metagenome library for clones which encode α -fucosidase enzymes were carried out on MacConkey base agar supplemented with 2-fucosyllylactose (2'-FL / Fuc α 1-2Gal β 1-4Glc) as sole carbon source. First the ability of the heterologous host strain *E. coli* EPI300 (*lacZ*⁻) to ferment L-fucose was determined by cultivating the host strain in MacConkey base agar supplemented with 0.5% fucose or 1.48% of 2'-FL (~molar equivalent of 0.5% fucose; 0.0304 M);. The host strain streaked on this plates was incubated overnight at 37°C. It was observed that the host strain streaked on MacConkey agar supplemented with fucose appeared pink and while the host strain streaked on 2'-FL was colourless, which indicates the host strain can ferment fucose but not 2'-FL (Fig 3.9 a and b). Thus the human faecal metagenome library were screened using MacConkey base agar supplemented with 1.48% 2'-FL, about ~500 colonies were lawned per plate and incubated overnight at 37°C. About ~11,000 clones were screened and 15 potential positive clones were selected based on pale pink colony morphology. Seven clones (, 2, 3, 6, 7, 9, 10 and 11) maintained pink morphology in the confirmation experiment (Fig 3.9 c).



Fig 3.9 Screening strategy of human faecal metagenome library on MacConkey base agar. **a.** *E. coli* EPI300 (pCC1FOS) host strain streaked on MacConkey base agar supplemented with 0.5% fucose. **b.** *E. coli* EPI300 (pCC1FOS) host strain streaked on MacConkey base agar supplemented with 2'-FL. **c.** *E. coli* EPI300 (pCC1FOS) host strain(-ve control) and

15 putative α -fucosidase expressing metagenome clones streaked on MacConkey base agar supplemented with 2'FL.

After *Sma*I endonuclease digestion, it was found that all of the eight clones were distinct clones as they displayed different restriction patterns (Fig 3.10). Thus the fosmid inserts of these clones were end sequenced and characterised through bioinformatic approach.

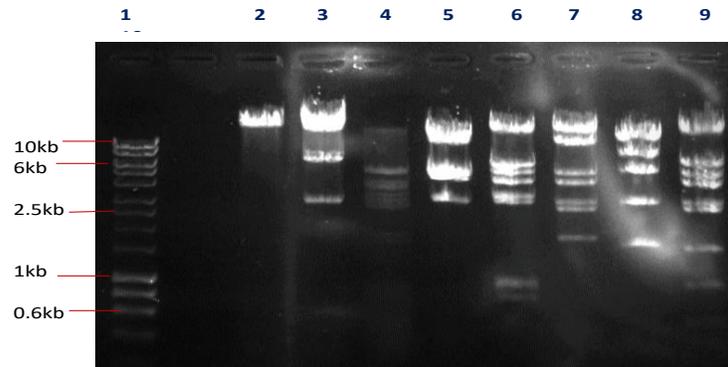


Figure 3.10 Restriction pattern analysis of putative α -fucosidase metagenome clones after *Sma*I digestion. Lane 1, 1 kb molecular marker (M), lanes 2-9 *Sma*I digested putative α -fucosidase clones (2, 3, 5, 6, 7, 9, 10, and 11) respectively.

To identify α -fucosidase genes encoded within the fosmid inserts of the metagenome clones, nucleotide blast searches of the end sequenced fosmid inserts of each clone were carried out. The BLAST record of the sequenced fucosidase clones indicated that three clones (7, 9, and 11) have a significant similarity with *B. adolescentis* whereas clone 6 had no significant similarity in the database. Clone 10 have a forward sequence matched with *Olsenella sp.* (78% identity) while no significant similarity on the reverse sequence. Likewise, forward sequence of clone 1 had matched with *Sus scrofa* (wild boar) genome assembly (99%) while the reverse sequence had no significant similarity in the database (Table 3.4). The metagenomic DNA fragments in the three clones 1, 6 and 10 could presumably be derived from previously un-sequenced novel bacteria in the human faecal microbiome.

Table 3.4 Blast record of putative α -fucosidase positive metagenome clones

Clones/ Sequence	Species	% identity	BP length of similarity
Fuc C#1 forward	<i>Sus scrofa</i> genome assembly, chromosome: Y	99%	297
Fuc C#1 reverse	No significant similarity found		
Fuc C# 6 forward	No significant similarity found		
Fuc C# 6 reverse	Short sequence		
Fuc C#7 forward	<i>Bifidobacterium adolescentis</i>	97%	532
Fuc C# 7 reverse	<i>Bifidobacterium adolescentis</i>	98%	367
Fuc C# 9 forward	<i>Bifidobacterium adolescentis</i>	99%	475
Fuc C# 9 reverse	<i>Bifidobacterium adolescentis</i>	99%	366
Fuc C# 10 forward	<i>Olsenella sp. Marseille-P3237</i>	78%	600
Fuc C# 10 reverse	No significant similarity found		
Fuc C# 11 forward	<i>Bifidobacterium adolescentis</i>	98 %	536
Fuc C#11 reverse	<i>Bifidobacterium adolescentis</i>	96 %	718

Of the three clones whose sequences matched with *B. adolescentis*, presence of α -fucosidase gene in these clones were searched using *B. adolescentis* ATCC15703 genome annotation in NCBI database. Clone 9 and 11 contain overlapping inserts and both contain the 21 genes BAD_RS05785 to BAD_RS05885. Clone 7 contains DNA adjacent to this locus but no shared genes. The three clones didn't harbour annotated α -fucosidase genes in the inserts (Tables 3.5, 3.6 and 3.7). Furthermore, a BLASTX search of the 3 clones illustrated that none of the clones harboured a known α -fucosidase-encoding gene.

As is indicated in Table 3.5, the annotated genes for the insert sequence of clone 7 indicated several genes for an ABC transporter and permease system, which could contribute in the transportation of sugars and other molecules. The conserved 21 genes of clone 9 and clone 11 indicated enzymes involved in peptidoglycan synthesis for instance UDP-N-acetylmuramoyl-L-alanine-D-glutamate ligase and a UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyro phosphoryl-undecaprenol N-acetylglucosamine transferase (Table 3.6) as well as an ABC transporter system (Table 3.7). Each insert contained hypothetical genes. These hypothetical genes were not annotated as GH on CAZymes. Together this analysis suggested that it therefore seemed unlikely that the inserts contained a fucosidase enzyme. The metabolism of 2'FL by these 3 clones could

be due to the permease systems importing 2'-FL, or other molecules, and subsequent intracellular metabolism, or alternatively non-specific metabolism by the peptidoglycan synthesis proteins or other enzymes. Furthermore, these clones did not show a colour change on MacConkey lactose agar, thus the colour change observed on MacConkey+2'-FL agar presumably is induced by the presence of 2'-FL.

Table 3.5; annotated protein details of putative fucosidase clone 7 sequence covered from 1307594-1339310 (31,716bp in length) of the matched strain (*B. adolescentis* ATCC 15703, accession number NC_008618.1)

Locus tag	Protein product	Length (aa)	Protein name
BAD_RS05605	WP_011743410.1	197	hypothetical protein
BAD_RS05610	WP_011743411.1	421	RNA methyltransferase
BAD_RS05615	WP_011743412.1	246	hypothetical protein
BAD_RS05620	WP_011743413.1	282	hypothetical protein
BAD_RS05625	WP_003810666.1	286	exodeoxyribonuclease III
BAD_RS05630	WP_011743414.1	669	ABC transporter ATP-binding protein
BAD_RS05635	WP_003810669.1	322	peptide ABC transporter permease
BAD_RS05640	WP_011743415.1	308	ABC transporter permease
BAD_RS05645	WP_035011216.1	550	ABC transporter substrate-binding protein
BAD_RS05650	WP_041777355.1	308	recombinase XerC
BAD_RS05655	WP_011743418.1	281	ABC transporter ATP-binding protein
BAD_RS05660	WP_011743419.1	948	ABC transporter permease
BAD_RS05665	WP_021913854.1	85	hypothetical protein
BAD_RS05670	WP_011743420.1	346	prephenate dehydrogenase
BAD_RS05675	WP_011743421.1	326	chorismate mutase
BAD_RS05680	WP_011743422.1	186	hypothetical protein
BAD_RS05685	WP_003810688.1	643	MULTISPECIES: GTP-binding protein TypA
BAD_RS05690	WP_011743423.1	451	MFS transporter
BAD_RS05695	WP_011743424.1	267	ADP-ribose pyrophosphatase
BAD_RS05700	WP_011743425.1	279	MULTISPECIES: chromosome partitioning ATPase
BAD_RS05705	WP_003810696.1	317	site-specific tyrosine recombinase XerD
BAD_RS05710	WP_003810697.1	127	50S ribosomal protein L20
BAD_RS05715	WP_003810699.1	64	MULTISPECIES: 50S ribosomal protein L35
BAD_RS05720	WP_035010673.1	232	translation initiation factor IF-3
BAD_RS05725	WP_011743427.1	628	asparagine synthetase B
BAD_RS05730	WP_011743428.1	248	thiamine pyrophosphokinase

Table 3.6: annotated protein details of putative fucosidase clone 9 sequence covered from 1339483-1376507(37,024bp in length) - of the matched strain (*B. adolescentis* ATCC 15703). Conserved regions with clone 11 shown in Bold.

Locus tag	Protein product	Length	Protein name
BAD_RS05740	WP_033499760.1	168	aminoacyl-tRNA deacylase
BAD_RS05745	WP_041777356.1	334	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
BAD_RS05750	WP_003810714.1	316	aldose 1-epimerase
BAD_RS05755	WP_003810716.1	83	hypothetical protein
BAD_RS05760	WP_041777357.1	201	RNA polymerase sigma factor RpoE
BAD_RS05765	WP_011743431.1	517	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase
BAD_RS05770	WP_011743432.1	414	carboxylate--amine ligase
BAD_RS05775	WP_011743433.1	241	aspartate racemase
BAD_RS05780	WP_011743434.1	532	dihydrofolate synthase
BAD_RS05785	WP_041777358.1	533	peptidase M24 family protein
BAD_RS05790	WP_011743436.1	173	ADP-ribose pyrophosphatase
BAD_RS05795	WP_011743437.1	570	ABC transporter ATP-binding protein
BAD_RS05800	WP_003810731.1	391	peptide ABC transporter permease
BAD_RS05805	WP_011743438.1	357	peptide ABC transporter permease
BAD_RS05810	WP_011743439.1	541	peptide ABC transporter substrate-binding protein
BAD_RS05815	WP_011743440.1	400	NagC family transcriptional regulator
BAD_RS05820	WP_003810740.1	299	2-dehydro-3-deoxygluconokinase
BAD_RS05825	WP_003810744.1	134	flavin-nucleotide-binding protein
BAD_RS05830	WP_003810745.1	329	cell division protein FtsQ
BAD_RS05835	WP_011743442.1	521	UDP-N-acetylmuramate--L-alanine ligase
BAD_RS05840	WP_011743443.1	393	undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase
BAD_RS05845	WP_033499762.1	363	cell division protein FtsW
BAD_RS05850	WP_041777512.1	477	UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase
BAD_RS05855	WP_003810755.1	365	phospho-N-acetylmuramoyl-pentapeptide-transferase
BAD_RS05860	WP_011743446.1	489	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase
BAD_RS05865	WP_033499725.1	292	hypothetical protein
BAD_RS05870	WP_011743448.1	600	cell division protein
BAD_RS05875	WP_003810762.1	153	hypothetical protein
BAD_RS05880	WP_011743449.1	343	ribosomal RNA small subunit methyltransferase H
BAD_RS05885	WP_011743450.1	171	division/cell wall cluster transcriptional repressor MraZ

Table 3.7: annotated protein details of putative fucosidase clone 11 sequence covered from 1348975-1383124(34,149bp in length) - of the matched strain (*B. adolescentis* ATCC 15703). Conserved regions with clone 9 shown in Bold.

Locus tag	Protein product	Length	Protein name
BAD_RS05785	WP_041777358.1	533	peptidase M24 family protein
BAD_RS05790	WP_011743436.1	173	ADP-ribose pyrophosphatase
BAD_RS05795	WP_011743437.1	570	ABC transporter ATP-binding protein
BAD_RS05800	WP_003810731.1	391	peptide ABC transporter permease
BAD_RS05805	WP_011743438.1	357	peptide ABC transporter permease
BAD_RS05810	WP_011743439.1	541	peptide ABC transporter substrate-binding protein
BAD_RS05815	WP_011743440.1	400	NagC family transcriptional regulator
BAD_RS05820	WP_003810740.1	299	2-dehydro-3-deoxygluconokinase
BAD_RS05825	WP_003810744.1	134	flavin-nucleotide-binding protein
BAD_RS05830	WP_003810745.1	329	cell division protein FtsQ
BAD_RS05835	WP_011743442.1	521	UDP-N-acetylmuramate--L-alanine ligase
BAD_RS05840	WP_011743443.1	393	undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase
BAD_RS05845	WP_033499762.1	363	cell division protein FtsW
BAD_RS05850	WP_041777512.1	477	UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase
BAD_RS05855	WP_003810755.1	365	phospho-N-acetylmuramoyl-pentapeptide-transferase
BAD_RS05860	WP_011743446.1	489	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase
BAD_RS05865	WP_033499725.1	292	hypothetical protein
BAD_RS05870	WP_011743448.1	600	cell division protein
BAD_RS05875	WP_003810762.1	153	hypothetical protein
BAD_RS05880	WP_011743449.1	343	ribosomal RNA small subunit methyltransferase H
BAD_RS05885	WP_011743450.1	171	division/cell wall cluster transcriptional repressor MraZ
BAD_RS05890	WP_011743451.1	765	ATP-dependent DNA helicase
BAD_RS05895	WP_003810774.1	399	D-3-phosphoglycerate dehydrogenase
BAD_RS05900	WP_011743453.1	170	transcriptional regulator NrdR
BAD_RS05905	WP_050731481.1	85	peptidoglycan-binding protein
BAD_RS05910	WP_003810781.1	257	ArsR family transcriptional regulator
BAD_RS05915	WP_011743455.1	313	cation transporter

The second approach used to screen for α - fucosidase encoding clones from the human faecal metagenome library was LB agar supplemented with the chromogenic substrate X-Fuc. A positive control *E. coli* BL21 (DE3) pLysS bearing a plasmid pETM10 with cloned α -L-fucosidase gene of *Thermotoga maritima* displayed a blue colony morphology on this agar (Fig 3.11a), while the negative control *E.coli* EPI300(pCC1FOS) appeared colourless. ~33,000 clones of the human faecal microbiome metagenome library were screened and none of the clones displayed α -fucosidase activity (Fig 3.11b). A confirmatory test of the 15 putative α -fucosidase clones identified on MacConkey base agar supplemented with 2'-FL had turned out colourless on LB supplemented with X-Fuc. This provided further evidence that these clones did not contain a fucosidase enzyme and they were ruled out from further characterisation.

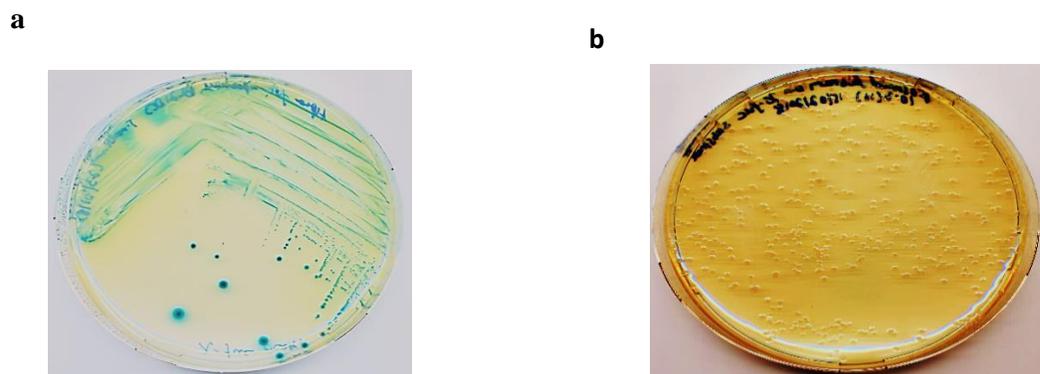


Fig 3.11 Human faecal metagenome library screening on X-Fuc supplemented LB agar. **a.** The positive control α -fucosidase expressing *E.coli* BL21 (DE3) clone streaked on LB agar supplemented with X-Fuc. **b.** Human faecal metagenome clones seeded on LB agar supplemented with X-Fuc, showing a negative result.

3.8. Growth dependent assays for screening human faecal metagenome library

One of the approaches used for screening of human faecal metagenome library for target genes was based on a growth-dependent assay. In the phenotypic approach, chromogenic substrates were supplemented in the agar media and its hydrolysis by the positive clones result in colour change. The advantage is it is straight forward to detect the expression of the target enzyme. Whereas in the growth approach substrate will be supplemented as a

sole carbon source in a minimal media, and positive clones are selected by their ability to metabolise the substrate and grow. This approach is more comprehensive in identifying not only the expression of the target enzyme but also possibly -the metabolic pathways involved in transport and metabolism of the supplemented carbohydrates and/or their metabolites.

Thus, to optimise the growth dependent screening of the metagenome library, M9 minimal media supplemented with three different carbon sources (glucose, lactose, or fucose) at 0.4% final concentration and M9 minimal media without carbon source were used. To optimise the assay two control strains were first assessed: *E. coli* EPI 300(pCC1FOS) and a β -galactosidase positive fosmid clone selected from our library (β -gal clone #11). *E. coli* EPI 300(pCC1FOS) and β -gal clone #11 grew on M9+glucose (Figure 3.12 a), while neither strain grew on M9 minimal media without carbon source (Figure 3.12 b). Since the host strain is *lacZ* *E. coli* EPI 300(pCC1FOS) and did not grow on M9+lactose. However, the β gal clone #11, grew sufficiently on M9 + lactose, which proves the heterologously expressed β -galactosidase BAD_1582 enabled the *lacZ* host strain to utilize the lactose as a sole carbon source (Fig 3.12c).

Both *E. coli* EPI 300(pCC1FOS) and β gal clone #11 were unable to grow on the M9+fucose (Fig 3.12d), demonstrating that the M9 minimal media with fucose a sole carbon source was not suitable to support growth of the host *E. coli*, even though our studies demonstrated the ability of this strain to ferment fucose in MacConkey agar. From this investigation, the inability of the host strain to grow on fucose as a sole carbon source hindered us to continue selection of fosmid clones expressing α -fucosidase through supplementing fucosyl substrates such as 2-fucosyllactose (2'-FL).

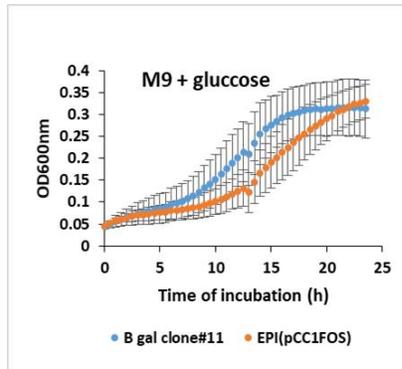
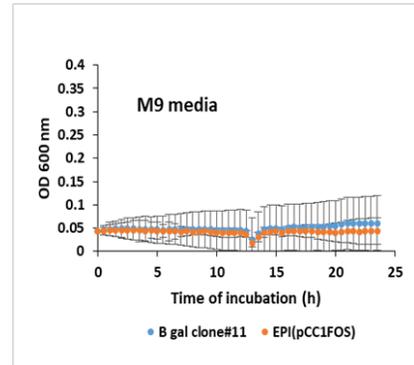
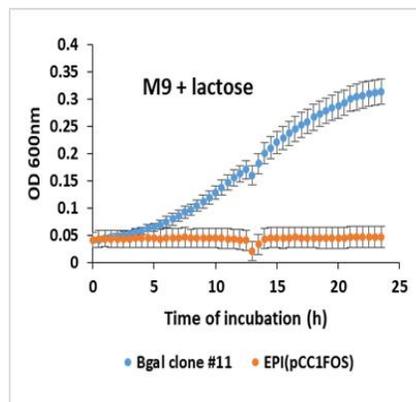
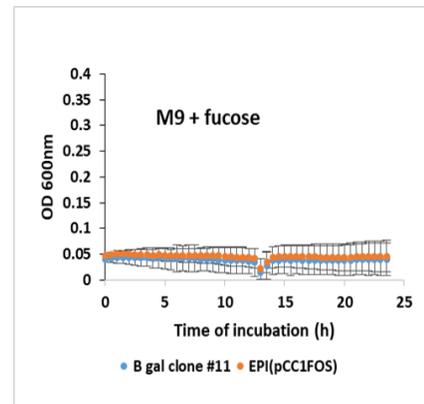
a**b****c****d**

Fig 3.12 Growth curves of *E. coli* EPI300 (pCC1FOS) and β gal clone #11 inoculated in M9 minimal media with or without carbon source under aerobic condition. **a.** M9 minimal media without carbon source. **b.** M9 + glucose. **c.** M9+lactose. **d.** M9+fucose. The data shows averages of three independent experiments. The values represent the mean of multiple technical replicates.

3.9. Discussion

There is a strong interest in studying the human gut microbiome to understand the mechanisms of host microbiome interaction in health and diseases states including development of colon cancer. Only then there could be a way of developing therapeutic interventions to halt the inevitable health issues which are associated with the modern life style and diet.

Herein a functional metagenomics approach was used to exploit the genetic potential of the human faecal microbiome to identify genes which encode GH enzymes, particularly sialidases, α -fucosidases and β -galactosidases. These enzymes are utilised by some members of the human gut microbiome not only to digest dietary carbohydrates but also they are involved in the deconstruction of complex O-linked mucin glycans which are terminally capped by sialic acid and fucose (Tailford *et al.*, 2015). Thus, our interest was to screen human faecal microbiome metagenome library and identify genes which encode for these enzymes and further study to understand their roles in host microbiome interactions.

A human faecal microbiome metagenome library was developed using the Epicentre copy control fosmid library production kit. The diversity of the metagenome library was determined by restriction pattern analysis and bioinformatics analysis. Our faecal metagenome library was dominated by Actinobacteria at the phylum level, specifically the species *B. adolescentis*, and Bacteroidetes and Firmicutes were also represented. 16S rDNA sequencing and metagenome sequence data sets of the human gut microbiome has revealed phylogenetic abundance at the phylum level of Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Synergistetes, Verrucomicrobia, Fusobacteria, Euryarchaeota in descending order (Arumugam *et al.*, 2011). However in a few studies, they did not detect Bacteroidetes (Gill *et al.*, 2006) and it was presumed that differences in faecal lysis and DNA extraction protocols contributed to this variation. The high abundance of the Actinobacteria in our metagenome library could be the suitability of our faecal lysis and DNA extraction procedure for these phyla, and/or due to the high abundance of *B. adolescentis* in the donor faecal sample.

Functional metagenomics provides an opportunity to discover novel genes based on their function through transcription and translation of cloned DNA in a surrogate host. The most widely used host strain, *E. coli* can express about 40% of genes from randomly cloned environmental DNA (Gabor *et al.*, 2004). Therefore, one major limitation of functional metagenomics would be absence of transcription factors and signalling systems to enhance expression of novel genes in the heterologous host, in our case *E. coli*. On top of this, codon usage bias in *E. coli*, i.e. synonymous codons that are used by heterologous genes and are rarely used in *E. coli*, would contribute to heterologous proteins being expressed inefficiently (Boël *et al.*, 2016), because of the low abundance of cognate tRNA in *E. coli* for the rare codons, genes (Lipinszki *et al.*, 2018a). Another factor that could contribute to poor detection of extracellular enzyme activities could be the limitation of *E. coli* secretion systems to recognise the heterologous proteins as substrates, so that the protein remains intracellular. The enzyme may therefore not be functional, or its activity may not be detected if its substrate is only located outside the cell. Nonetheless, novel target protein coding genes from the human faecal microbiome that are efficiently expressed by *E. coli* could be identified.

The human faecal microbiome metagenome library was screened for β -galactosidase expressing metagenome clones using MacConkey lactose agar-based screening and a confirmatory experiment with X-Gal supplemented LB agar was performed. Thirty clones were identified, and distinct clones were selected by their restriction pattern and characterised by end sequencing and Miller assay. High β -galactosidase activity was detected in most of the tested fosmid clones, which indicated there was high level of expression of the β -galactosidase gene using the heterologous host (*E. coli*) expression system and/or that the enzyme(s) was highly efficient.

A total of 30 β -galactosidase positive clones were obtained from screening of ~16,000 clones of the human faecal microbiome metagenome library using MacConkey lactose agar. The frequency of β -galactosidase expressing clone obtained from our human faecal metagenome library was quite high, compared to other metagenome library screens for β -galactosidases from different environments (Cheng *et al.*, 2017a; Erich *et al.*, 2015). Soil metagenomes screened for β -galactosidases using X-Gal as a substrate identified 3 β -

galactosidase expressing clones out of 12,000 clones in one study (Wang *et al.*, 2010). In another soil metagenome screen one positive clone was identified from metgenome 8,000 clones (Zhang *et al.*, 2013). Similarly, from a soil metagenome screen of $\sim 10^6$ clones using X-Gal as substrate ~ 354 positive clones were obtained, of which only 97 (27.4%) were positive for lactose hydrolysis (Erich *et al.*, 2015), presumably the difference in conformation of the indoxyl group of X-Gal compared to natural hexoses may alter binding to the active site of β -galactosidases and was not hydrolysed by these enzymes (Erich *et al.*, 2015). Therefore, it can be deduced that the human gut environment is a good source for identification of β -galactosidase enzymes. The prominent members of the human gut microbiome encode wide arrays of CAZymes, for instance the genome of *B. thetaiotaomicron* includes 236 glycoside hydrolases and 15 polysaccharides lyases and the carbohydrate-responsive regulatory network contains 290 genes, 30 transcriptional factors (TFs), and a global Crp (Cyclic AMP receptor protein)-like regulator (Ravcheev *et al.*, 2013). From the Firmicutes phylum, 11 representative strains of Roseburia and *E. rectale* (the ‘Roseburia/*E. rectale* group’) encode a mean number of 85 GHs per genome (O Sheridan *et al.*, 2016). Similarly from the Actinobacteria phylum, the pangenome of 18 *B. adolescentis* strains encode 36 glycosyl hydrolases (GH), 12 glycosyl transferases (GT) and four carbohydrate esterases (CE) (Duranti *et al.*, 2016). The *B. adolescentis* genome contains 9 annotated β -galactosidase genes. The high percentage of β -galactosidase positive clones detected in our screen could be attributed to the high abundance of *B. adolescentis* genomic DNA in our library. The other β -galactosidases of *B. adolescentis* or β -galactosidases of other microbiome sources were not detected from our metagenome library, presumably due to inefficient expression of the genes in *E. coli*, such as improper folding of the proteins, codon biasing, perhaps these proteins require post translational modification to be functional such as disulphide bond formation between subunits (Boel *et al.*, 2016; Fakruddin *et al.*, 2013; Kaur *et al.*, 2018). Enhanced expression of genes from metagenome libraries have been context specific (Martinez and Osburne 2013) and some genes are highly expressed when clones are incubated at low temperature (30°C) for 3 days (Rondon *et al.*, 2000). Metagenome library constructed from human gut faecal microbiome was also screened for identification of α -fucosidase expressing clones. From the two-approach screening $\sim 33,000$ clones were screened on

LB agar supplemented with X-Fuc and ~11000 clones were screened using MacConkey base agar supplemented with 2'FL. The X-Fuc screening methodology was confirmed by blue colony morphology of the positive control *T. martima* α -fucosidase expressing clone, while negative clones were colourless. A similar approach using X-Fuc as a substrate was used for identification of α -fucosidase expressing clones from soil metagenome library, in which 7 α -fucosidase expressing clones were identified from 100, 000 clones. Six clones belonged to the GH 29 family (α -fucosidases) (Lezyk *et al.*, 2016). MacConkey base supplemented with 2'FL screening was a novel approach for screening metagenome libraries for α -fucosidase expressing clones based on the hydrolysis of fucose and subsequent red colour formation by the positive clones. From the later screening 7 putative clones were identified, though further confirmation using X-Fuc and bioinformatic analyses revealed that these clones neither express active α -fucosidase enzyme nor contain annotated α -fucosidase genes. From the X-Fuc screening none of the clones hydrolysed X-Fuc which confirmed the absence of α -fucosidase expressing clones in our metagenome library. Therefore, the putative clones appeared red presumably from hydrolysing residual fucose in the 2'FL substrate.

Genome analysis of the gut microbiome revealed at least one α -fucosidase in each of 131 Genomes belonging to the Actinobacteria, Bacteroidetes, Firmicutes, and Verrucomicrobia phyla. However, the largest number of α -fucosidases genes are encoded by representative members of the Firmicutes (*Lachnospiraceae* bacterium 3_1_57FAA_CT1, 16 genes) and Bacteroidetes (*Bacteroides coprophilus* DSM18228, 14 genes) (Ravcheev and Thiele 2017). The predominant *Bifidobacterium* species in the adult human gut include *B. adolescentis* and *B. longum* subsp. *longum* that are well equipped to degrade plant-derived polysaccharides rather than HMOs, manifested by absence of annotated α -L-fucosidase and exo sialidase genes in the genomes of both species. However, infant gut adapted *Bifidobacterium* species, such as *B. longum* subsp. *infantis* and *B. bifidum*, encode repertoire of CAZymes which enabled them to utilise HMOs. Hence the occurrence of α -fucosidase enzymes is high in infant gut adapted *Bifidobacterium* species, which was shown to grow on fucosyllactose (HMOs) as a sole carbon source (Bunesova *et al.*, 2016). Interestingly, members of the *B. scardovii*, *B. longum* subsp. *infantis* and *B. bifidum* species have been reported to possess the most

extensive set of host-glycan-degrading glycoside hydrolases (GH), such as GH33, exo-sialidases, GH29 and GH95, which represent fucosidases, and GH20, which include hexosaminidase and lacto-N-biosidase activities (Milani *et al.*, 2015a). Even though *B. bifidum* has been predominantly detected from infant faecal microbiome, low number of this bacteria has been detected in the faecal microbiome of adult gut (Turroni *et al.*, 2012). In adult faeces *B. adolescentis* and *B. catenulatum* are among the *Bifobacterium* species which are mostly detected (Turroni *et al.*, 2014), and there are no annotated fucosidase genes in the genomes of these *Bifidobacterial* species. The absence of α -fucosidase expressing clones from our human faecal microbiome metagenome library could be because the library was dominated by adult gut adapted *Bifidobacterium* species, mainly *B. adolescentis*.

~16,128 metagenome clones were screened for exo sialidase using MacConkey base agar supplemented with 3'-SL/6'-SL substrates activity, none of which expressed the target enzyme. This could be due to inefficient expression and secretion system of *E. coli*. In addition, the dominant members of these library, *B. adolescentis* have no annotated exo sialidase gene in their genome. Genome analysis of human gut microbiome detected sialidases in 112 genomes with the maximum number detected in *B. fragilis* 638R and *Bacteroides* sp. D22 genomes (7 genes per genome) but it was absent in the genomes of Fusobacteria (Ravcheev and Thiele 2017).

The positive control expressing the sialidase gene of *A. fumigatus* showed hydrolysis of the fluorescent substrate 4 MU-NANA, which would have been a good start for screening the metagenome library for exo sialidase genes. However, exo sialidase screening using this substrate was not feasible as the extracellular enzyme diffuse through the agar and hydrolyse the substrate and make it impossible to distinguish positive clones from negative clones. Thus, using 4MU-NANA as a substrate in agar based screening with plate spreading of the library would not be ideal. Instead, a spot assay of individual clones can be used (Moncla and Braham 1989), however, substantially high amount of substrate would be required when screening forty thousand clones due to the low number of clones screened per plate. For confirmation of sialidase expression, culture supernatants of putative positive clones could be used in 96 well plate assay containing 4MU-NANA as

a substrate and liberation of 4-methylumbelliferone quantified by measuring the absorbance at 360 nm (Vilei *et al.*, 2011).

The growth studies showed that heterologously expressed β -galactosidase gene in β -galactosidase positive clone (β -gal Clone 11) promote growth of the *lac*⁻ host *E. coli* strain (EPI 300) in the presence of M9 media containing lactose (0.4%) as sole carbon source. However, M9 media supplemented with fucose (0.4%) as sole carbon source did not support growth. Though *E. coli* K12 strain was shown to metabolise and grow on fucose as a sole carbon source under aerobic and anaerobic condition (Hacking and Lin 1976). In anaerobic condition, the enzyme propanediol oxidoreductase is induced for conversion of lactaldehyde to 1, 2 propanediol which ultimately released into the culture medium. Under aerobic condition, intermediate level of this enzyme was detected (Boronat and Aguilar 1981; Hacking and Lin 1976) and lactaldehyde would rather be catabolized by lactaldehyde dehydrogenase, to lactate which is converted to pyruvate by lactate dehydrogenase (Hacking and Lin 1976). *E. coli* cultured on M9 media supplemented with 1% fucose showed a slow growth rate compared to M9 media supplemented with 1% glucose and two main fermentation products (acetic acid and 1,2 propanediol) were produced on fucose culture while acetic acid was the main fermentation product on glucose culture (Kim *et al.*, 2019). The metabolomic and metatranscriptomic analysis of the two cultures revealed that the intracellular metabolic intermediates of *E. coli* cultured on M9+fucose were distinct from and *E. coli* cultured M9+glucose, which bring the conclusion of fucose induced distinct metabolic networks (pentose phosphate pathway related metabolites) that are not induced by glucose (glycolysis and hexose metabolic pathways related metabolites) (Kim *et al.*, 2019). The absence of detectable growth in our host *E. coli*, cultured in M9 media supplemented with 0.4% fucose could be presumably the low concentration of fucose added compared to the aforementioned studies, as only higher concentrations of fucose (1%) support growth yet at slow rate. Thus about 2.96% 2'-FL (equivalent of 1% fucose) must be supplemented to support growth, however this substrate is very expensive to use at high concentrations for screening purpose of metagenome libraries. In conclusion, the hypothesis for screening the metagenome library on M9 minimal media supplemented with 2'-FL, whereby α -fucosidase expressing clones could grow by breaking down 2'-FL and metabolising fucose was disregarded.

The functional metagenomics approach that was used to identify glycoside hydrolase enzymes from human faecal microbiome metagenome library was new by its approach. Despite, our screening of human faecal microbiome for sialidase and α -fucosidase enzymes were not successful, to our knowledge this was the first functional metagenomics study that has attempted to identify these two enzymes from the human gut microbiome. In addition, the screening strategy using MacConkey base agar supplemented with fermentable carbohydrates such as 2'-FL, 3'-SL/6'-SL was new by its approach. It could be applicable in functional screening of metagenomics libraries that would be constructed from uncultured microbes inhabiting wide range of habitats, with a potential to identify industrially and pharmaceutically important enzymes.

Through function-based screening of human faecal microbiome metagenomic library, more than 30 β -galactosidase expressing clones were identified. A new approach of screening glycoside hydrolase enzymes such as sialidases and α -fucosidases were optimised and developed. The MacConkey base agar supplemented with 2'-FL and 3'-SL could be a potential approach for screening of metagenome libraries from different environments for identification of sialidases and α -fucosidases. The *B. adolescentis* β -galactosidase (BAD_1582) was the predominant β -galactosidase identified from the human faecal microbiome metagenome library screening, the properties of this enzyme have not been experimentally characterised previously. The next chapter of this thesis deals with the cloning, purification and biochemical characterisation of this enzyme.

Chapter 4

Biochemical characterization of the *Bifidobacterium adolescentis* β -galactosidase, BAD_1582

4.1. Introduction

β -galactosidases (E.C.3.2.1.23) are enzymes which catalyse the hydrolytic cleavage of galactose residues from the non-reducing end of β -galactosides. β -galactosidases belong to the six glycoside hydrolase families of GH 1, GH 2, GH 35, GH 42, GH 59 and GH 147 within the Carbohydrate Active enZymes (CAZy) database (<http://www.cazy.org/>) (Lombard *et al.*, 2014a). They have been widely used in dairy industries for the production of lactose free dairy products (Johnson *et al.*, 1993). These products are required by the large group of lactose intolerant people world-wide which account for 2-70% (Europeans and USA) or ~100% (Asians) of the population resulting from a gradual decline in lactase production (Vandenplas 2015). However in some rare exceptions infants are congenitally deficient for lactase, which makes the condition even worse (Vandenplas 2015). Moreover, the hydrolysis of lactose in dairy products decreases lactose crystallisation problems resulting from its low solubility and the hydrolysis products increase the sweetness (Gänzle *et al.*, 2008). Besides, these enzymes have been used for reduction of lactose contents in whey, which is a yellowish green liquid by product of cheese and casein production (Ryan and Walsh 2016). Worldwide about 190×10^6 ton/year whey is produced, and only about 50 % of these whey is processed for animal feed, functional food productions for human consumption such as whey protein powders, and ethanol production; while the rest is released as dairy waste (Baldasso *et al.*, 2011). Lactose is the major solid constituent of whey (~70-72%), followed by proteins (~8-10%) and minerals (10-15%) (Venetsaneas *et al.*, 2009). Bioconversion of lactose in whey decreases about 75% of water pollution (Mawson 1994).

In addition, several β -galactosidases can perform transgalactosylation reactions. Transgalactosylation, is a biochemical reaction carried out by many β -galactosidases and some GH enzymes (Torres *et al.*, 2010b), in which these enzymes cleave and retain the terminally attached galactose from lactose, then transfer the galactose moiety to an acceptor nucleophile (for instance lactose at start of the reaction, Glc and Gal formed by hydrolysis of lactose or GOS formed by transgalactosylation reaction could be acceptor sugars) rather than the hydroxyl group of water (Iqbal *et al.*, 2010). Transgalactosylation involves both intra and intermolecular reactions (Torres *et al.*, 2010b); In intramolecular

transgalactosylation a direct transfer of the galactose moiety to glucose in β -1, 6 linkage generate allolactose, a disaccharide and regioisomer of lactose. While in intermolecular transgalactosylation, the enzyme transfers the galactose moiety to lactose, Glc, Gal or GOS molecule to generate mixture of galactooligosaccharides (GOS) (Huber *et al.*, 1976; Torres *et al.*, 2010b). These various mixture of oligosaccharides constitute various degree of polymerisation (DP) of 2-10 with terminal Glc and attached with β -(1 \rightarrow 2, 3, 4, or 6) linkages (Coulier *et al.*, 2009). These products such as galactooligosaccharides (GOS) and several other oligosaccharides form lactose-based prebiotics (Reuter *et al.*, 1999) which are widely used in infant formula. The benefits of prebiotics such as GOS come about through enhancing the colonisation and proliferation of both exogenous and endogenous beneficial bacterial species in the gut. These species play enormous role in prevention of infectious diseases such as viral and bacterial diarrhoea, mainly through competition for nutrients and binding sites and synthesis of antimicrobial metabolites (Rastall *et al.*, 2005). In addition supplementation of probiotics and Faecal Microbiota Transplantation (FMT) to restore microbial-host balance are the most recent approaches being used and have been implicated in prevention of chronic inflammatory diseases through modulation of host immune system (Sun and Chang 2014), although efficacy and safety issues remain in question for the later (Mills *et al.*, 2018). Several studies have shown that GOS administered together with probiotic bacteria such as *Bifidobacterium* (synbiotics) (Gibson and Roberfroid 1995) could be a valuable way to promote proliferation of exogenously supplemented probiotic strains in humans (Kolida and Gibson 2011). A synbiotic treatment of ulcerative colitis patients with *B. breve* yalkut and GOS showed improved clinical symptoms (Ishikawa *et al.*, 2011). Several *in vitro* and *in vivo* studies have shown that GOS selectively stimulate the growth of *Bifidobacteria* (Torres *et al.*, 2010b; Walton *et al.*, 2012). Supplementation of GOS to healthy elderly human subjects showed an increase in abundance of faecal *Bifidobacteria* and a significant increase in phagocytosis, natural killer (NK) cell activity, and the production of anti-inflammatory cytokine interleukin-10 (IL-10) (Davis *et al.*, 2010; Vulevic *et al.*, 2008). *In vitro* studies showed addition of GOS in growth medium enhanced expression of genes required for metabolising GOS by some members of the *Bifidobacterium* genus such as *B. adolescentis* (Akiyama *et al.*, 2015). Faecal short chain fatty acids (SCFA)

level of neonates fed with a formula supplemented with GOS as prebiotics are comparable to that of breast fed neonates (Salli *et al.*, 2019). Moreover, GOS and sialyllactose positively regulate cell cycle pathways involved in epithelial cell proliferation and differentiation in vitro, a mechanism which enhance repair of damaged epithelial cells (Perdijk *et al.*, 2019). In addition to their prebiotic effect, GOSs have low sweetness, low cariogenicity and low caloric values and thus undesirable effects associated with GOS supplementation is minimal (Maischberger *et al.*, 2008).

Considerable work has been done to obtain novel β -galactosidases for their invaluable biotechnological and pharmaceutical applications. Function-based metagenomics and bioinformatics-guided gene identification are the most widely used techniques to identify novel β -galactosidases from different environments. They have been identified, and subsequently characterised, through analysis of metagenome sequences retrieved from Metagenomic Rapid Annotations using Subsystems Technology (MG RAST) website (<http://www.mg-rast.org/>) (Liu *et al.*, 2019b). On the other hand, metagenome libraries constructed from bacterial communities in various environments, mainly soil samples, have been utilised to isolate and characterise novel β -galactosidases without the necessity for traditional cultivation approaches (Erich *et al.*, 2015; Wang *et al.*, 2010; Zhang *et al.*, 2013). The most characterised and commercially used β -galactosidases are obtained from the bacterium *Bacillus circulans*, the yeast *Kluyveromyces lactis* and the fungus *Aspergillus oryzae* (Kim *et al.*, 2003; Neri *et al.*, 2009; Park and Oh 2010). These commercial β -galactosidases have some undesirable properties such as a low affinity for lactose and product inhibition by galactose at low K_i (Erich *et al.*, 2015). Thus there is a demand for discovery of β -galactosidases with improved properties such as low product inhibition, a higher affinity for lactose and a wide range of thermal and pH stability which suits the demand of downstream process in dairy industries. In addition, there is a demand for novel β -galactosidases that can catalyse transgalactosylation reactions to synthesise GOS with various linkages and degree of polymerisation (DP).

This chapter describes the purification and biochemical characterisation of the *B. adolsecentis* β -galactosidase (BAD_1582) that was identified through function-based screening of our human faecal microbiome metagenome library (Chapter 3). A new β -

galactosidase enzyme from *B. adolescentis* (BAD_1582) was identified from the positive metagenome clones. The enzyme was recombinantly expressed with a hexa histidine tag, purified and the properties of the expressed enzymes were characterised. BAD_1582 exhibited efficient hydrolysis of lactose and transgalactosylation activities at low concentration of lactose (6-8%) to produce GOS with different DP and linkages. These properties make this enzyme an ideal candidate for large scale enzymatic hydrolysis of lactose and synthesis of potential prebiotic oligosaccharides in dairy and pharmaceutical industries.

4.2. Bioinformatic analysis to identify β -galactosidase gene (BAD_1582) from human faecal microbiome metagenome clones

The end sequences of the metagenomic inserts in the fosmids showing putative β -galactosidase enzyme activity were analysed bioinformatically to establish the origin of the DNA. Nucleotide blast searches indicated that seven of the metagenome fosmid sequences had high identity (98-99%) with genomic DNA of *B. adolescentis*, a common inhabitant of an adult human gut (Table 3.3). The genome annotation of *B. adolescentis* ATCC15703 (NC_008618.1, NCBI) was used to identify possible β -galactosidase genes encompassing the sequences of the fosmid inserts. The gene map of the clones shows the different regions of *B. adolescentis* genomic DNA contained within each fosmid (Fig. 4.1) All the seven clones contained the β -galactosidase gene BAD_1582 and an additional five genes (BAD_1583 to BAD_1587). These include the HdeD family acid-resistance protein (BAD_1583), LacI family transcriptional regulator (BAD_1584), carbohydrate ABC transporter substrate-binding protein (BAD_1585), sugar ABC transporter permease (BAD_1586), and carbohydrate ABC transporter permease (BAD_1587). Most of these genes encode sugar transporting permeases. Those proteins are required by the bacteria for adaptation in the gut environment by accessing the sugars. Clone 12 had similarity with the *B. adolescentis* genome at one end but matched with uncultured bacterium extrachromosomal DNA (85% identity) at the other end (Table 3.3). Such insertion sequences could be horizontally transferred to *B. adolescentis* within the normal gut environment.

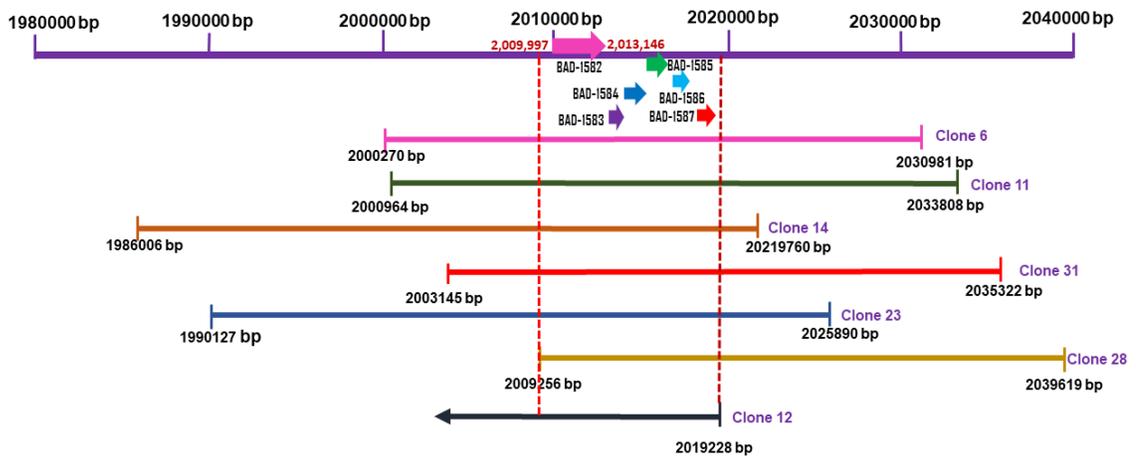
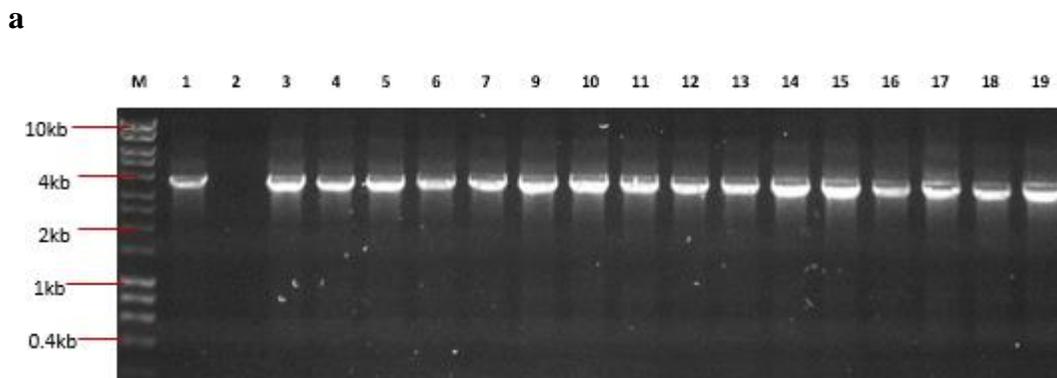


Fig. 4.1 Gene map of fosmids which incorporated the nucleotide sequences of different regions of the genome of *B. adolescentis* and the six common genes shared between fosmids. These includes the β -galactosidase gene (BAD_1582), HdeD family acid-resistance protein (BAD-1583), LacI family transcriptional regulator (BAD_1584), carbohydrate ABC transporter substrate-binding protein (BAD_1585), sugar ABC transporter permease (BAD_1586), and carbohydrate ABC transporter permease (BAD_1587). The numbers on top and end of each clone show location of the nucleotide sequences on the ATCC15703 genome.

PCR to amplify BAD_1582 was carried out using the fosmids from the thirty β -galactosidase positive clones (Chapter 3.5) as template. Apart from clone 2, all the fosmids contain the gene sequence of BAD_1582 (Fig. 4.2).



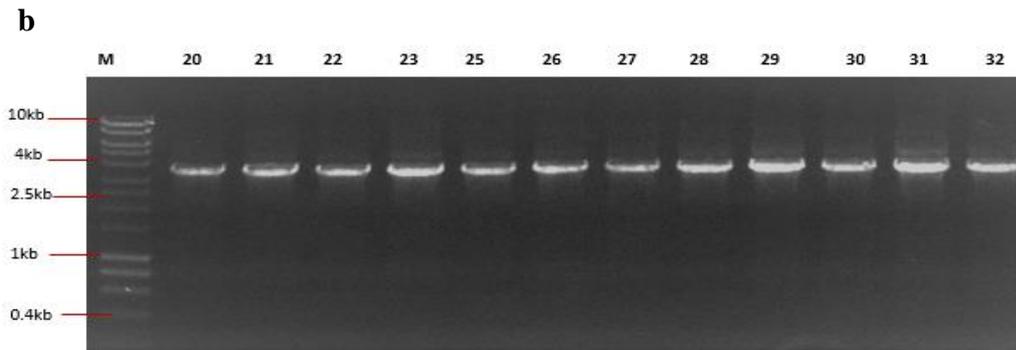


Fig. 4.2 PCR of representative BAD_1582-harboring metagenome clones. **a** PCR amplification product of the target gene (3.15 kb) of 17 clones (β -gal clones 1-19). Clone 2 does not contain this gene. **b** PCR amplification products of the target gene (3.15 kb) of 12 clones (β -gal clones 20-32).

Protein subcellular prediction tool (PSORTb) revealed that BAD_1582 is an intracellular protein. The homology-based structure of the enzyme was predicted using SWISS-MODEL, and BAD_1582 has a homo dimer structure, based on the highest Global Model Quality Estimation (GMQE) score (0.66) (Fig 4.3).

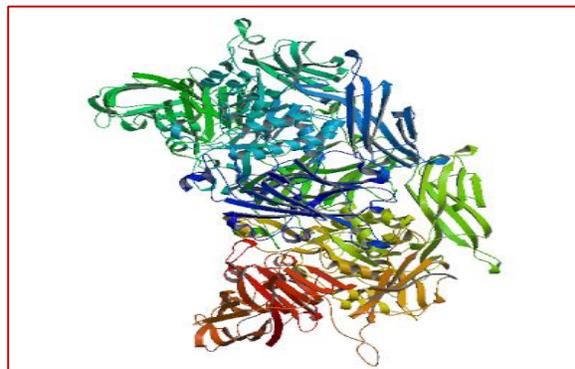


Fig. 4.3 Homology based structure prediction of the *B. adolescentis* β -galactosidase, BAD_1582. The model was constructed using SWISS-MODEL based on homology of the amino acid sequences of this enzyme with other known sequences in database, resulting in a homodimer (matching prediction) with a highest GMQE score.

The putative conserved domains with the *Bifidobacterium* β -galactosidase, BAD_1582 were identified from conserved domain database (CDD) and Pfam database. As is indicated in Fig 4.4 the enzyme has specific hits with *lacZ*, Glycoside hydrolase family 2 and β -galactosidase small chain with E-values of $1.88e^{-143}$, $2.82e^{-138}$ and $6.85e^{-97}$ respectively, which infers BAD_1582 is a β -galactosidase enzyme. This enzyme has non-specific hits with evolved β -galactosidase (EbgA), cryptic β -D-galactosidase subunit α with E-value of 0. Thus, based on sequence identity on conserved domain database (CDD) and Pfam, BAD_1582 belongs to the glycoside hydrolase family 2 (GH2).

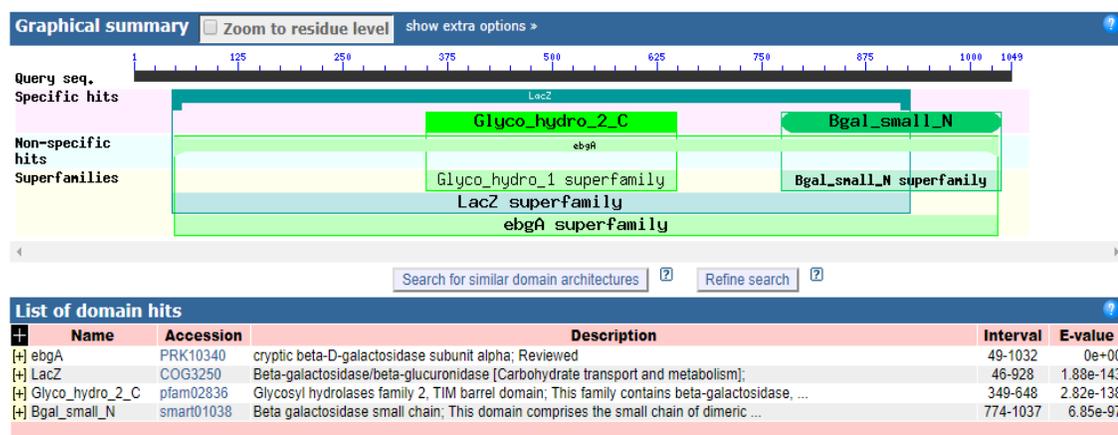


Fig. 4.4 Putative conserved domains of the protein sequence of BAD_1582, in the CDD and Pfam databases.

A phylogenetic tree was constructed using *Bifidobacterium* species protein sequences similar to BAD_1582 (E-value = 0, % identity >68%) and the 5 SmartBLAST reference landmark protein sequences for known GH2 β -galactosidases (E-value < e^{-113} , % identity >30%). Phylogenetic analysis revealed that BAD_1582 is more distantly related to other known GH2 β -galactosidases identified from different species of bacteria and plants such as *Arabidopsis thaliana* (30% identity), soybean (32% identity) *Streptomyces coelicolor* (40% identity), *Thermotoga maritima* (33% identity), *E. coli* (32% identity). However, BAD_1582 is closely related to β -galactosidases identified from some *Bifidobacterium* species such as *B. ruminatum* (88% identity), *B. pseudocatenulatum* (77% identity), *Bifidobacterium* sp. N4GO5 (68% identity), *B. kashiwanohense* (68% identity), *B. tsurumiense* (68% identity), *B. catenulatum* (68% identity) and *B. longum* (68% identity), which can indicate these enzymes might be descended from the same ancestral protein

(Fig. 4.5) On the contrary the enzyme has less homology (<68% identity) to β -galactosidases from *B. animalis*, *B. infantis*, *B. bifidum*, *B. breve*, *B. dentium*, *B. reuteri* and other members of the *Bifidobacterium* taxa.

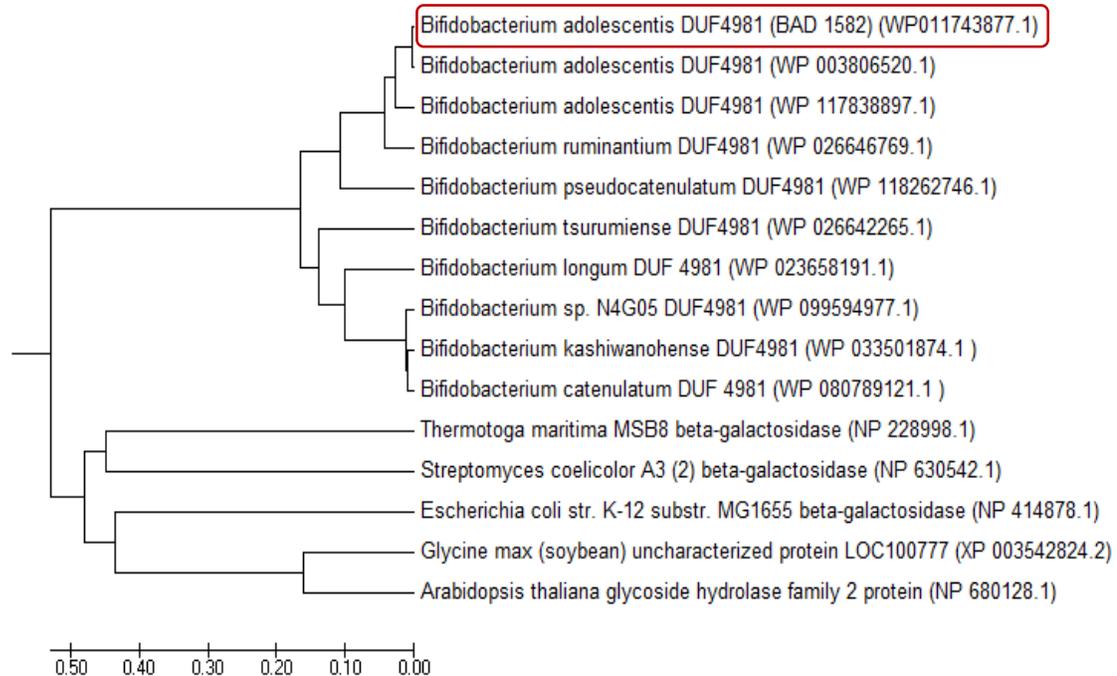


Fig. 4.5 Evolutionary relationship of BAD_1582 with homologous proteins and landmark similar sequences. The evolutionary history was inferred using UPGMA method. The optimal tree with the sum of branch length=3.18 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 15 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 891 positions in the final data set. Evolutionary analyses were conducted in Mega 7. The top three *B. adolescentis* refer BAD_1582 and two β -galactosidases with high identities with BAD_1582 ($\geq 99\%$ and $\geq 93\%$) and the top identity matches for each *Bifidobacterium* ($\geq 68\%$) were incorporated for this analysis..

4.3. Optimisation of expression of His-tagged BAD_1582

BAD_1582 was amplified by PCR and cloned in the expression plasmid pET101 with a hexa histidine-tag to generate plasmid pDMg1a. The plasmids pDMg1a and the empty vector pET101 were transformed into *lacZ* expression host *E. coli* T7 express, to generate *E. coli* T7 Express (pDMg1a), and *E. coli* T7 Express (pET101), respectively. The *E. coli* T7express (pDMg1a) was used for expression of the recombinant *Bifidobacterium* β -galactosidase BAD_1582. *E. coli* T7express (pET101) was used as negative control.

To confirm the cloned gene in pDMg1a was intact and in correct orientation, the plasmids pDMg1a and pET101 were purified and then subjected to double restriction enzyme digestion using *Hind*III and *Sac*I. The restriction pattern analysis indicated that the correct size bands (6058 bp, 1576 bp, 884 bp, 374 bp) for pDMg1a harbouring intact BAD_1582 in a correct orientation were obtained, while in the case of pET101 (5,753 bp) only vector bands were detected (Fig. 4.6). Furthermore, the gene sequence obtained from primer walk sequencing revealed that an intact gene (3,150 bp) was cloned in correct orientation in pDMg1a.

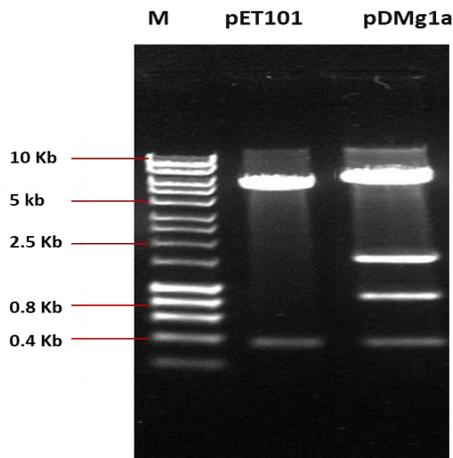


Fig. 4.6 Restriction pattern analysis of recombinant clone pDMg1a and empty vector pET101 after double digestion using *Hind* III and *Sac*I. Lane 1 (M), 1 Kb molecular ladder; Lane 2, pET101 and Lane 3, pDMg1a.

E. coli T7 express (pDMg1a) and *E. coli* T7 express (pET101) were cultured on LB agar plates supplemented with X-gal in the presence of 1 mM IPTG. The β -Galactosidase activity of recombinant BAD_1582 was confirmed by the characteristic blue colonies while *E. coli* T7 express (pET101) appeared white (Fig. 4.7).

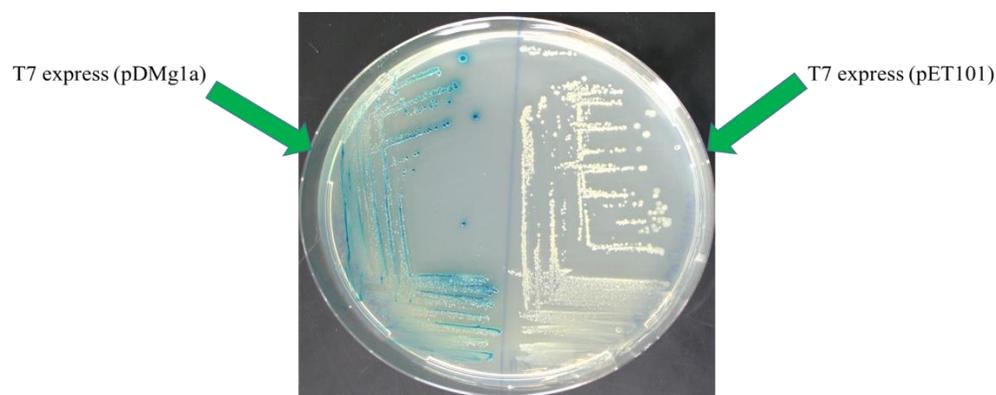


Fig. 4.7 The recombinant clone *E. coli* T7express (pDMg1a) and empty vector harboring clone *E. coli* T7express (pET101) streaked on LB agar supplemented with X-gal in the presence of 1 mM IPTG, indicating the expression of the recombinant β -galactosidase, BAD_1582.

After the X-Gal confirmatory test, the Miller assay was conducted on the *E. coli* T7 express (pDMg1a) and *E. coli* T7 express (pET101) to analyse the level of β -galactosidase activity. A fresh single colony of *E. coli* T7 express (pDMg1a), *E. coli* T7 express (pET101) and a positive control *E. coli* BL21 (DE3) were inoculated in LB broth (with added Amp for the former 2 strains) overnight at 37°C. The following day 0.1 OD₆₀₀ of the overnight cultures was transferred to fresh LB broth and incubated at 37°C with shaking until OD₆₀₀~0.4-07 was reached. Thereafter, 0.1 mM IPTG was added to the experimental group to induce expression of the recombinant β -galactosidase and no IPTG was added to the control strains and incubated for 3 h at 37°C with shaking. After that, 100 μ l aliquots from each culture was used for Miller assay. The level of β -galactosidase activity for *E. coli* BL21 (DE3) in the presence of IPTG was (7,088 Miller Units) and in the absence of IPTG (2,246 Miller Units) as expected higher activity was detected in IPTG

induced culture than the un-induced culture. On the other hand, the β -galactosidase activity of *E. coli* T7express (pDMg1a) was relatively less compared to the positive control BL21 (DE3), with IPTG (361 Miller Units) and without IPTG (1,787 Miller Units) were detected. This result confirmed the enzyme activity was unaffected by the incorporation of the His-tag. However, the level of β -galactosidase activity was five-fold higher in un-induced *E. coli* T7express (pDMg1a) compared to the induced culture. The later could be due to most likely a leaky T7 promoter which allowed the expression of the recombinant BAD_1582 without IPTG induction, besides clump of cells could have been pipetted during the assay (experimental error) which contributed for the high Miller Units by the culture without IPTG (Fig. 4.8). The negative control *E. coli* T7 express (pET101) had very low β -galactosidase activity (3 Miller Units).

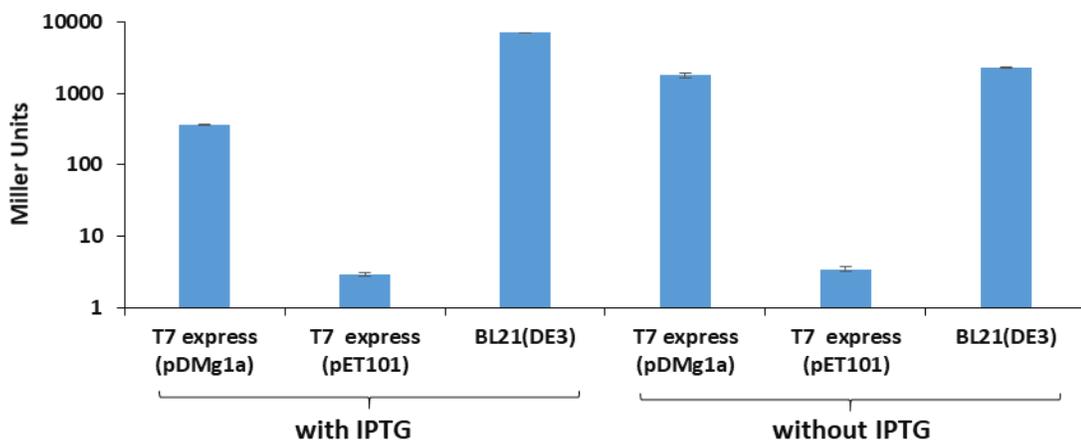


Fig. 4.8 Miller assay on *E. coli* T7 express (pDMg1a), *E. coli* T7 express (pET101) and *lacZ*⁺ *E. coli* BL21DE (3) with and without presence of 0.1 mM IPTG. The values represent the means of three biological replicates, each independent biological replicate had three technical replicates. Error bars represent standard deviations.

Hence to decrease the basal expression of BAD_1582, a pLysS plasmid was transformed into *E. coli* T7express (pDMg1a) and *E. coli* T7 express (pET101). The pLysS plasmid is used in DE3 lysogenic hosts to suppress basal expression from the T7 promoter by producing T7 lysozyme, a natural inhibitor of T7 RNA polymerase (Novagen). The strains containing pLysS and the *E. coli* BL21 (DE3) were incubated in the presence or absence

of 0.1 mM IPTG at 37°C with shaking. Thereafter, the Miller assay was conducted. The result showed the presence of pLysS inhibited the basal expression of recombinant β -galactosidase from *E. coli* T7 express (pDMg1a) in the absence of IPTG, as a higher β -galactosidase activity (238 Miller Units) was detected after IPTG induction, than without IPTG (37 Miller Units) (Fig 4.9). The β -galactosidase activity of *E. coli* T7 express (pDMg1a) in the presence of IPTG with or without pLysS was similar, while the β -galactosidase activity of the uninduced strain + pLysS was considerable lower than in the absence of pLysS, and was similar to the negative control (compare Fig 4.8 and Fig 4.9).

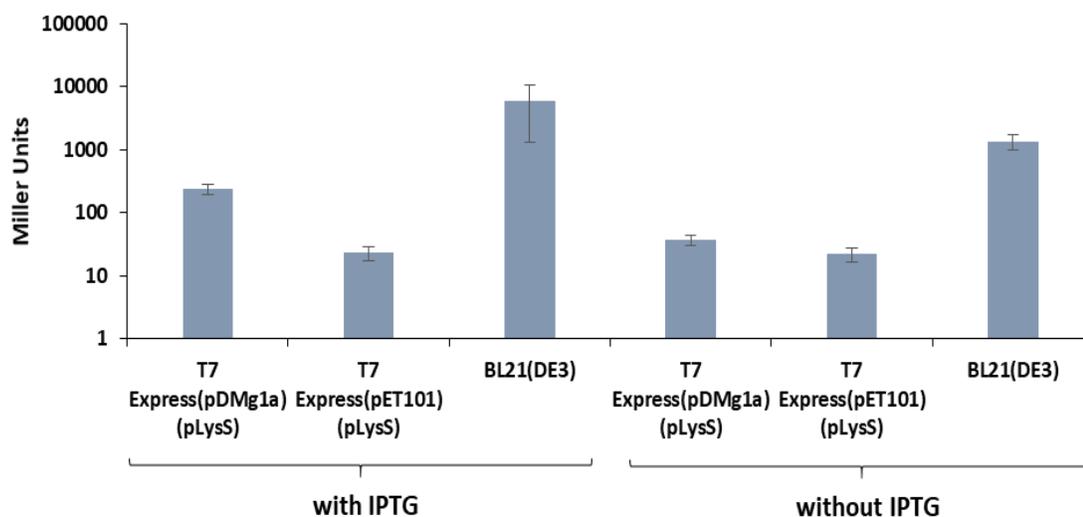


Fig. 4.9 Miller assay on *E. coli* T7 express (pDMg1a) (pLysS), *E. coli* T7 express (pET101) (pLysS) and *lacZ*⁺ *E. coli* BL21 (DE3) with and without presence of 0.1 mM IPTG. The values represent the means of three biological replicates and each independent biological replicate had three technical replicates. Error bars represent standard deviations.

For optimisation of expression of BAD_1582, the four strains T7 express *E. coli* (pDMg1a), T7 express *E. coli* (pDMg1a) (pLysS), T7 express *E. coli* (pET101), and T7 express *E. coli* (pET101) (pLysS) were used. The parameters used were two different concentrations of IPTG (0.1 and 1 mM) and four time points of induction (0, 1, 2, 3 and 4 h) with the culturing protocol described above. Aliquots (50 µl) were withdrawn and analysed with SDS-PAGE. The enzyme has a predicted molecular size of 117 kDa by amino acid content. Strongest expression of BAD_1582 was detected at 4 h (data not shown). At this time point the protein was not detected in either of the negative control T7 express *E. coli* (pET101) strains. The protein band of BAD_1582 was not clearly seen in the T7 express *E. coli* (pDMg1a) without pLysS, presumably due to poor expression. Nevertheless, the band of BAD_1582 was clearly seen in the T7 express (pDMg1a) (pLysS) sample, which indicates expression had happened, besides expression was similar at both concentrations of IPTG in this sample (Fig. 4.10).

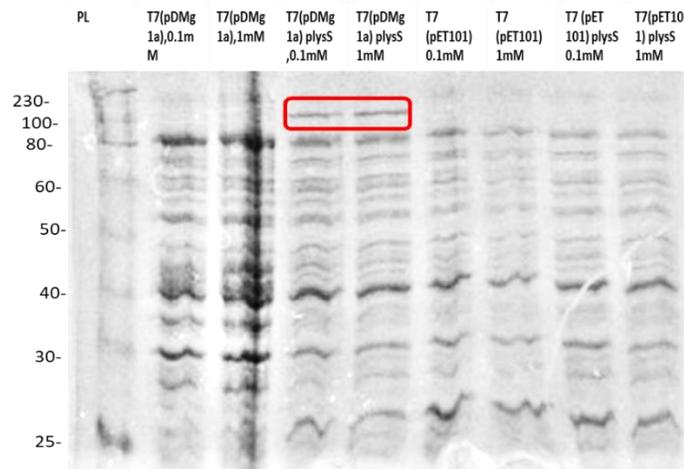
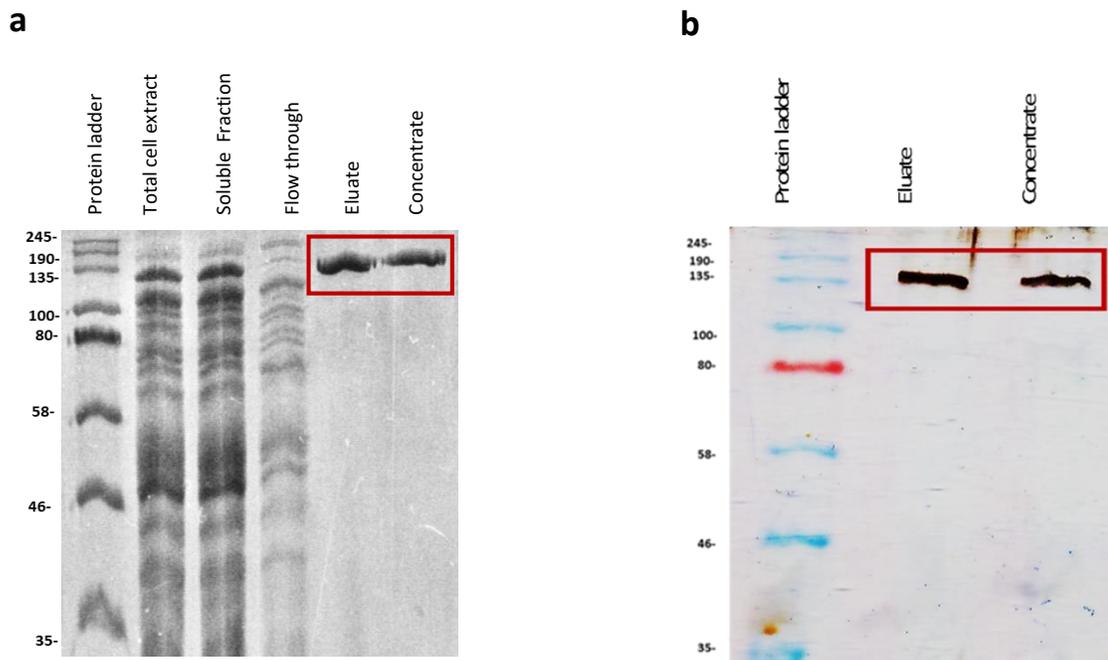


Fig. 4.10 SDS-PAGE analysis of expression of BAD_1582 after 4 h IPTG induction. Red marker box indicates enzyme with predicted size of 117kDa. Strains assessed were *E. coli* T7 express (pDMg1a) (*E. coli* T7 express (pDMg1a) (pLysS) *E. coli* T7 express (pET 101) and *E. coli* T7 express (pET 101) induced at 0.1 and 1 mM IPTG PL, Protein molecular weight marker.

4.4. Expression and purification of the His-tagged β -galactosidase BAD_1582

In order to express the recombinant BAD_1582, the expression host *E. coli* T7 express (pDMg1a) was cultured to exponential phase, 1 mM IPTG was added and induction of protein expression was continued for 4 h. PSORTb subcellular prediction tool predicted that BAD_1582 is localized in the cytoplasm and hence for purification of this protein IPTG induced cells were lysed to extract the expressed proteins from the cytoplasm. IPTG induced *E. coli* (pDMg1a) cells were harvested, lysed, and the His-tagged enzyme was purified from the soluble fraction by Ni²⁺-NTA column chromatography (12 μ g ml⁻¹ culture was purified). Remaining low molecular mass co-purified contaminant proteins were removed by filtration through a 100 kDa amicon filter and the resulting 117 kDa protein was the only protein present by SDS-PAGE analysis with Coomassie and silver staining (Fig. 4.11 a, b). Immunoblotting to detect the His-tag confirmed the presence of His-tagged BAD_1582 (Fig. 11c).



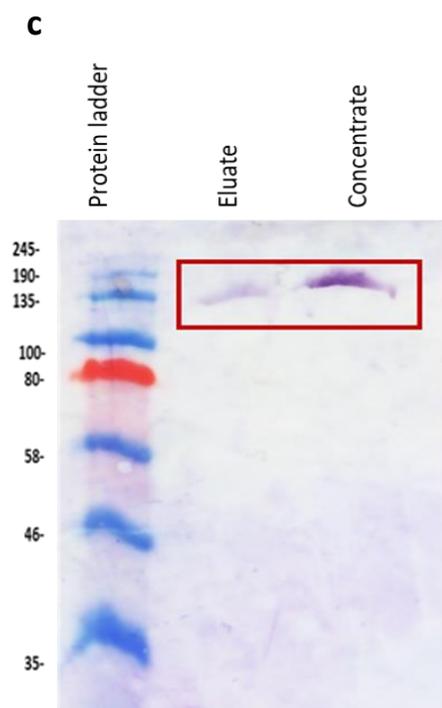


Fig. 4.11 BAD_1582 Purification. Purification of the protein analysed by SDS-PAGE of purification step samples. **a** Coomassie stained gel, **b** Silver stained gel, **c** western blot image using monoclonal anti-polyhistidine antibody conjugated to peroxidase and Tetramethyl Benzidine (TMB) for staining. The flow through and eluates are samples obtained from Ni²⁺-NTA column purification and concentrate is concentrated sample using amicon centrifugal fitters. The protein molecular weight marker used was Color Prestained Protein Standard, Broad Range (11–245 kDa).

4.5. Biochemical properties of BAD_1582

Every enzymatic reaction is affected by several external factors, such as pH, temperature, ionic strength, enzyme concentration, substrate concentration, metal ions, detergents, presence of inhibitors or activators etc. (Bisswanger 2014). Thus, it is necessary to determine the optimal conditions that an enzyme works best. The optimal conditions of BAD_1582, were determined by measuring the β -galactosidase activity using a colorimetric enzyme assay. The assay was carried out in the presence of the chromogenic

substrate *o*NPG and the liberation of the product (*o*-nitrophenol) by the enzyme was measured using spectrophotometer (stopped assay at a defined time). All characterization of the enzyme was carried out in 50 mM sodium phosphate buffer, pH 7 (prepared by adding 39 parts sodium monobasic and 61 parts sodium dibasic by volume, both at 50 mM stock) containing 10 mM MgCl₂. The effect of pH on the β -galactosidase activity of BAD_1582 was tested by incorporating 8.75 $\mu\text{g ml}^{-1}$ of purified enzyme in 50 mM sodium phosphate buffer containing 10 mM MgCl₂ over the pH ranges of 4.0 to 10.0 in the presence of 2 mM *o*NPG for 30 min at 37°C. After that the reaction was stopped using 500 mM sodium carbonate and the released product (*o*-nitrophenol) was measured using BioTek Spectrophotometer at 420_{nm}. This enzyme had an optimum pH of 7.0 and retained 60% of its activity between pH 6.0 and 8.0 (Fig. 4.12a), however its activity were declined at acidic and basic pH. The pH stability of BAD_1582 was determined by preincubating the enzyme (8.75 $\mu\text{g ml}^{-1}$) in the aforementioned buffer over pH ranges of 4.0-10.0 for 24 h at 4°C, then the enzyme from each pH ranges were brought back to pH 7.0 and enzyme assay was conducted in the presence of 2 mM *o*NPG for 30 min at 37°C. BAD_1582 was stable at all pH tested with 87% of its activity retained at pH 4 (Fig. 4.12a).

The effect of temperature on the β -galactosidase activity of BAD_1582 was tested by incubating the enzyme (8.75 $\mu\text{g ml}^{-1}$) between 0-60°C in 50 mM sodium phosphate buffer containing 10 mM MgCl₂ and 2 mM *o*NPG for 30 min. The optimum temperature of this enzyme was determined to be 37°C, but it was active over a wide range of temperatures, retaining 60% of its activity at reaction temperatures between 20 to 45°C (Fig. 4.12b). The residual enzyme activity was measured in a reaction performed at 37°C after 1 h preincubation of the enzyme (8.75 $\mu\text{g ml}^{-1}$) in 50 mM sodium phosphate buffer/10 mM MgCl₂ set at various temperatures (0-60°C). These results showed that the enzyme was stable for 1 h up to 40°C (65% of its activity was retained), and its stability declined as temperatures increased (at 50°C less than 20% activity retained) (Fig. 4.12b). The effect of storage temperatures on the activity of BAD_1582 was determine by preincubation of the enzyme (8.75 $\mu\text{g ml}^{-1}$) in 50 mM sodium phosphate buffer/10 mM MgCl₂ (pH 7.0) at 4°C, -20°C and RT (20°C) for 24 h and after the stored enzyme was brought back to 37°C and β -galactosidase activity was determined by adding 2 mM *o*NPG for 30 min. Storage at 4°C and -20°C did not decrease the activity of the enzyme. On the other hand, storage

at room temperature for 24 h decreased the activity of this enzyme by 35% (Fig. 4.12c). The enzyme was stored at 4°C for more than five weeks and its activity was not decreased (data not shown), but for long term storage -20°C in a 50% glycerol stock was used and the activity was unaffected.

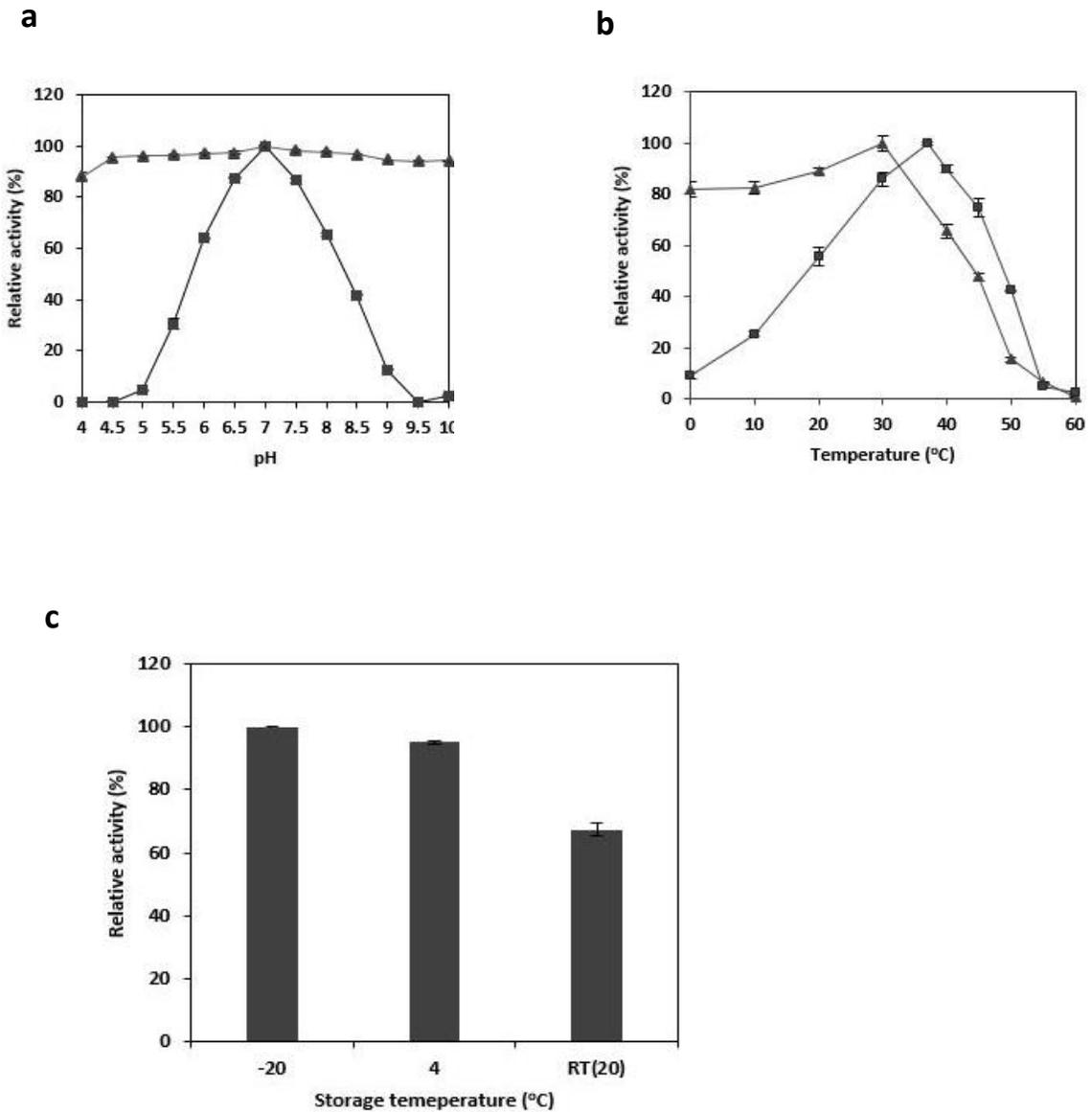


Fig. 4.12 Effect of pH and temperature on BAD_1582. **a and b** Effect of pH (**a**) and temperature (**b**) on the β -galactosidase activity (squares) and stability (triangles) of BAD_1582 in 50 mM sodium phosphate buffer. **c** Effect of storage temperature on the

activity of BAD_1582, each data compared to 4°C storage. The values represent the mean of three independent experiments and the error bars represent standard deviation.

Some enzymes require the presence of a cofactor or prosthetic group for their activity. To determine the metal ion requirement of BAD_1582, β -galactosidase activity assay was carried out in 50 mM sodium phosphate buffer (pH 7) or 100 mM Tris-HCl buffer (pH 7) supplemented with monovalent and divalent metal ions at a final concentration of 10 mM. The assay contained, enzyme (8.75 $\mu\text{g ml}^{-1}$), substrate *o*NPG (2 mM) and metal ions (10 mM) incubated at 37°C for 30 min. Two buffers were used because the divalent ions (Ca^{2+} , Mn^{2+} , Zn^{2+} and Cu^{2+}) could not dissolve in sodium phosphate buffer, thus 100 mM tris HCl buffer was used for this metal ions. Moreover, the effect of EDTA on the activity of BAD_1582 was determined by preincubating the enzyme (8.75 $\mu\text{g ml}^{-1}$) in 50 mM sodium phosphate buffer (pH 7.0) in the presence of 10 mM EDTA for 3 h at 4°C. After that, the β -galactosidase activity was determined by addition of 2 mM *o*NPG at 37°C for 30min. The addition of Mg^{2+} enhanced the activity of BAD_1582 by 30% to 70% (depending on the buffer) followed by Ca^{2+} and Zn^{2+} , respectively. On the other hand, the enzyme activity was abolished by the addition of Cu^{2+} and showed 59% reduction in the presence of Mn^{2+} ions. The addition of EDTA inhibited the activity of the enzyme compared to the control which indicated a reliance of the enzyme on a divalent cation. However, the EDTA did not abolished the enzyme activity, presumably some of the metal ions were tightly bound to the enzyme. The control without added metal ions showed activity due to a carryover of the bound metal ion in the enzyme and were not lost because of the purification procedure (Table 4.1).

Table 4.1 Effect of divalent cations and EDTA on the activity of BAD_1582 in 50mM phosphate buffer/100mM Tris-HCl buffer.

Substances (10mM)	Relative activity (%) in phosphate buffer	Relative activity (%) in 100 mM Tris HCl
None (control)	100±0.4	100±0.3
NaCl	93±6	n.d
KCl	105±5.6	n.d
MgCl ₂	130±2.4	174 ±3.6
CaCl ₂	n.d	141 ±1.8
ZnCl ₂	n.d	133±5.1
MnCl ₂	n.d	41 ± 2.6
CuSO ₄	n.d	0±0.1
EDTA	36 ± 0.7	n.d

* n.d- the experiment in the specified buffer was not determined. The values represent the mean ±Standard deviation of three biological replicates, each with three technical replicates.

The activity of enzymes is affected by the exposure to detergents and denaturing agents, as these factors alter the native conformation of the enzyme which ultimately result in the loss of activity either reversibly or irreversibly. Herein, the effect of different detergents or denaturing agents on BAD_1582 were determined, by performing an enzyme assay containing BAD_1582 (8.75 µg ml⁻¹), *o*NPG (2 mM) in the presence of different concentrations of SDS, Triton X-100, urea and β-mercaptoethanol and incubated at 37°C for 30 min. As shown in Fig 4.13, the enzyme was severely inhibited by the addition of SDS at a concentration of 0.5% and above whereas addition of urea at 0.1 mM enhanced the activity of BAD_1582 while no inhibition or enhancement were observed at concentrations of at 0.5 mM and 1 mM. On the other hand, Triton X-100 enhanced activity

at 0.5% and 1%, and no inhibition was observed at 0.1%. However, the presence of β -mercaptoethanol did not have an effect on the activity of BAD_1582 at the concentrations tested. This is a desirable feature for biotechnological application as this enzyme did not require a reducing agent for its activity.

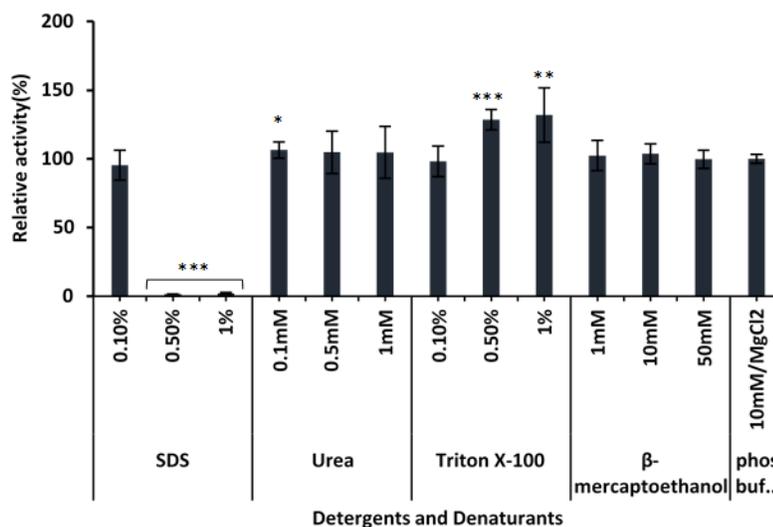


Fig. 4.13 Effect of detergents and denaturants on the activity of BAD_1582 in 50 mM sodium phosphate buffer. The data are relative to the enzyme incubated in phosphate buffer/10 mM MgCl₂ (positive control) without any added detergent or denaturants. The values represent the mean of three biological replicates, each with three technical replicates. The error bars represent the standard deviation.

4.6. Substrate specificity and kinetic parameters

Enzymes are extraordinarily selective catalysts, endowed with the ability to select a single substrate from a range of similar compounds. The specificity of the enzymes arises from the three dimensional conformation of the active site (Hedstrom 2001). On the other hand, substrate specificity of enzymes cannot be absolute as many metabolic enzymes lack this and may act on alternative substrates (Peracchi 2018). In this case substrates are compared based on the ratio of the K_{cat}/K_m values for their reactions (Tawfik 2014), which is a discrimination factor when two or more alternative substrates of the enzymes are presented, the enzyme selects one over the other (Peracchi 2018).

To determine the substrate specific of BAD_1582, different chromogenic substrates with α and β -linkages were used. The assay contained BAD_1582 ($8.75 \mu\text{g ml}^{-1}$) and the substrates (2 mM) in 50 mM sodium phosphate buffer (pH 7) and incubated at 37°C for 30 min. This enzyme has high substrate specificity on β -linked galactose residues, as shown by its high specific activity on *p*NPG ($112 \mu\text{mol/min/mg}$) and *o*NPG ($18 \mu\text{mol/min/mg}$), whereas no activity was detected on α -linked galactose as in *p*Np- α -Gal. The enzyme doesn't have any activity on the other tested *p*NP based substrates (*p*NP- β -D-Glc, *p*NP- α -D-Man, *p*NP- α -L-Fuc, *p*NP- α -L-Xyl, *p*NP-GalNAc and *p*NP-GlcNAc) which proves this enzyme has a very narrow range of substrates, and acted selectively only on β -linked galactose residues (Table 4.2). A higher specific activity of this enzyme towards *p*NPG was detected compared to *o*NPG during the substrate specificity test. These could be due to the difference in presentation of nitro group on the 4th and 2nd carbon of phenol in *p*NPG and *o*NPG providing a distinct interaction of these substrates with the active site of this enzyme.

Table 4.2. Substrate specificity of BAD_1582 tested on *p*NP based substrates

Substrate (2mM)	Activity ($\mu\text{mol/min/mg}$)
<i>p</i> NP- β -D-galactopyranoside	112
<i>p</i> NP- β -D-glucopyranoside	0
<i>p</i> NP- α -D-galactopyranoside	0
<i>p</i> NP- β -D-GalNAc	0
<i>p</i> NP- β -D-GlcNAc	0
<i>p</i> NP- α -L-fucopyranoside	0
<i>p</i> NP- α -D-xylopyranoside	0
<i>p</i> NP- α -D-mannopyranoside	0
<i>o</i> NP- β -D-galactopyranoside	18.1

To determine the kinetic parameter of BAD_1582 a time course enzyme assay was carried out using *o*NPG and lactose as a substrate. Time course β -galactosidase activity was monitored for 18 min using different concentrations of the chromogenic substrate *o*NPG (0.25-20 mM) and the substrate lactose (5-200 mM), relatively high initial concentrations of lactose (≥ 5 mM) was used to make sure products (glucose) is detected from indirect assay, besides to examine if there exists substrate inhibition by lactose. The release of *o*NP from *o*NPG was directly measured using Biotech spectrophotometer at 420 nm while the release of glucose from lactose was quantified using glucose oxidase/oxidase assay kit (Sigma). The liberated glucose is oxidized to gluconic acid and hydrogen peroxide by glucose oxidase. Hydrogen peroxide reacts with *o*-dianisidine (reduced and colourless) in the presence of peroxidase to form a brown colored product (Oxidized *o*-dianisidine). The oxidized *o*-dianisidine reacts with sulfuric acid to form a more stable pink colored product (Oxidized *o*-dianisidine). The intensity of the pink colour is measured at 540 nm and it is proportional to the concentration of glucose (Fig. 4.14a and c). The enzyme showed substrate inhibitions at concentrations ≥ 2.5 mM *o*NPG (Fig. 4.14a). In the case of lactose, substrate inhibition was not observed when lactose was used as a substrate even as concentration increased to 200 mM (Fig. 4.14c). The enzyme rate ($\mu\text{mol min}^{-1} \text{mg}^{-1}$) was determined from non-linear regression fit model of substrate concentration vs enzyme rate using Michaelis-Menten kinetics for *o*NPG (Fig. 4.14b) and lactose (Fig. 4.14d).

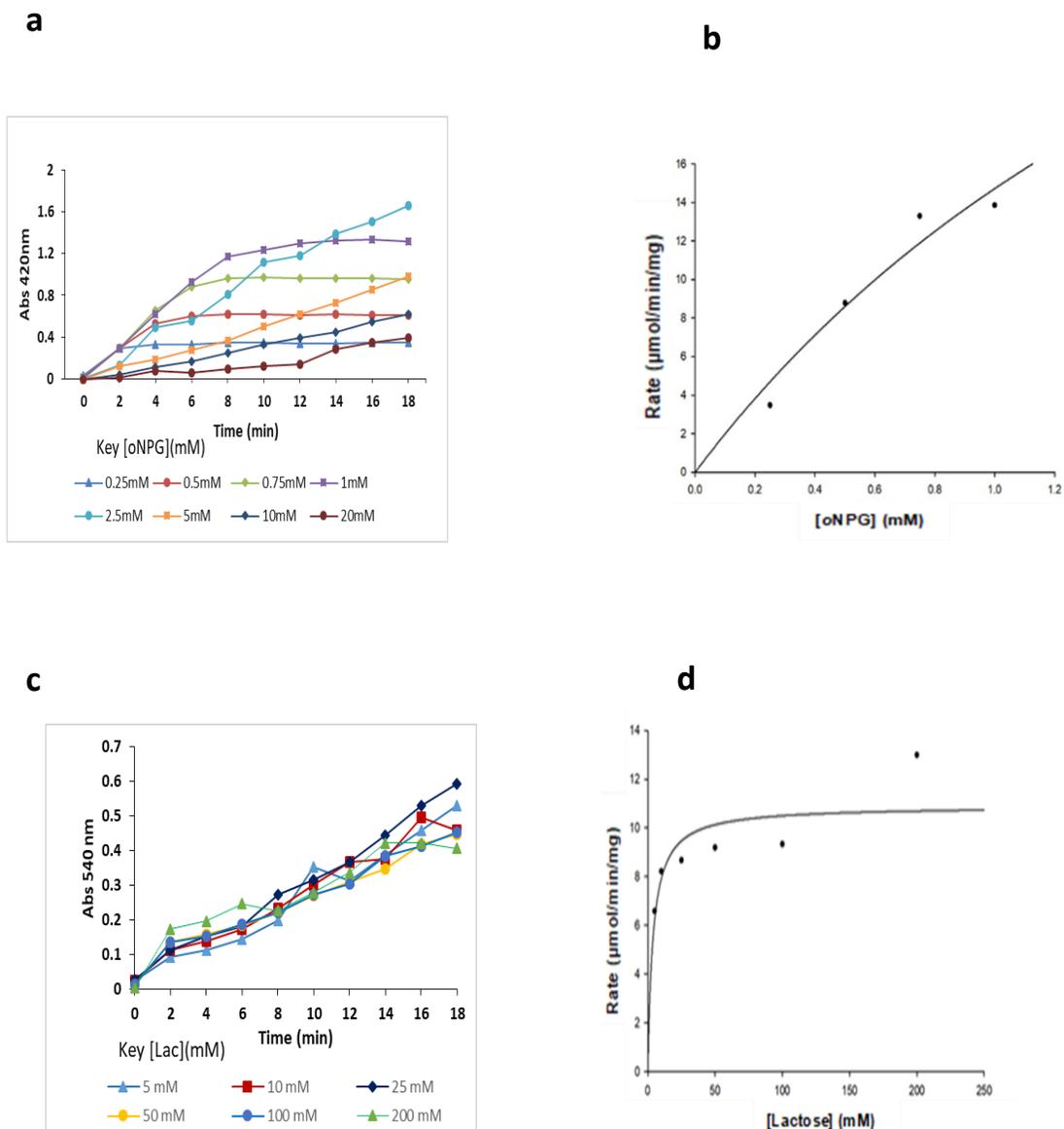


Fig. 4.14 Kinetic parameter determination of BAD_1582 using *o*NPG and lactose as a substrate. **a** Time course assay of BAD_1582 activity in the presence of different concentration of *o*NPG. **b** Non-linear regression fit for estimation of the kinetic parameters of BAD_1582 using *o*NPG as a substrate (0.25-1.00 mM) between 0- 6 min. **c** Time course assay of BAD_1582 in the presence of different concentration of lactose as a substrate. **d** Non-linear regression fit for estimation of the kinetic parameters of BAD_1582 using lactose as a substrate (5-200 mM) between 0-6 min.

The enzyme had a K_m of 2.5 mM and V_{max} of 51 $\mu\text{molmin}^{-1}\text{mg}^{-1}$ for *o*NPG and K_m of 3.8 mM and V_{max} of 11 $\mu\text{molmin}^{-1}\text{mg}^{-1}$ for lactose, Besides, the catalytic efficiency; K_{cat}/K_m of this enzyme for *o*NPG was 39.6 $\text{s}^{-1}\text{mM}^{-1}$ which was 7 fold higher than of lactose which was 6.0 $\text{s}^{-1}\text{mM}^{-1}$ (Table 4.3). Nevertheless, the rate of this enzyme for *o*NPG showed a substantial decrease with increasing concentration of the chromogenic substrate *o*NPG (≥ 2.5 mM) (Fig. 4.14 a) but the enzyme did not display substrate inhibition when lactose was used as substrate. The K_i of the enzyme for *o*NPG was $6.2 \times 10^{-4} \pm 0.82$ mM.

Table 4.3 The kinetic parameters of BAD_1582 using *o*NPG and lactose as substrate, and in the presence of lactose and D-galactose as competitive inhibitors of *o*NPG hydrolysis.

Substrate/Inhibitor	K_m (mM)	V_{max} ($\mu\text{mol/ min/ mg}$)	K_{cat} (s^{-1})	K_{cat}/K_m ($\text{s}^{-1}\text{mM}^{-1}$)	K_i (mM)
<i>o</i> NPG	2.5 \pm 3.1	51.0 \pm 48.14	99.1	39.6	6.2 $\times 10^{-4} \pm 0.82$
Lactose	3.8 \pm 2.0	11.0 \pm 0.96	22.5	6.0	n.d
<i>o</i> NPG +Lactose	1.6 \pm 3.4 $\times 10^7$	21.0 \pm 2.7 $\times 10^8$	40.8	25.5	3.0 \pm 2.5 $\times 10^7$
<i>o</i> NPG+D-galactose	1.0 \pm 4.8 $\times 10^6$	20.4 \pm 4.8 $\times 10^7$	39.7	38.4	115.4 \pm 2.7 $\times 10^8$

* n.d- no substrate inhibition observed and hence K_i was not determined for lactose, *o*NPG (1 mM) was used as a substrate, lactose (0.75-30 mM) and galactose (10-700 mM) were used as inhibitors.

4.7. Inhibition of BAD_1582 by lactose, glucose and galactose

The effect of different concentrations of the substrate lactose and its hydrolysis products, glucose and galactose, on the activity of BAD_1582 were examined in 50 mM sodium phosphate buffer/10 mM MgCl_2 (pH 7) in the presence of *o*NPG (1 mM) at 37°C for 30 min. The β - galactosidase activity towards the synthetic substrate *o*NPG declined markedly in the presence of increasing concentrations of lactose (Fig. 4.15a). This suggested that at higher concentrations lactose competed for the active site and decreased the availability of the enzyme active sites for *o*NPG to bind and hence caused a decrease in *o*NPG hydrolysis. The kinetic parameters of the enzyme towards *o*NPG in the presence of different concentrations of lactose were determined from the time course assay (Fig 4.15b). The K_i of lactose as a competitive inhibitor was 3.0 mM. In addition, K_m , V_{max} and K_{cat} for *o*NPG in the presence of lactose were 1.6 mM, 21 $\mu\text{mol min}^{-1}\text{mg}^{-1}$ and 25 s^{-1}

¹, respectively (Table 4.3). Thus, the catalytic efficiency K_{cat}/K_m towards *o*NPG was low in the presence of lactose.

The effects of the hydrolytic products of lactose on the activity of the enzyme towards *o*NPG was also determined. The enzyme was highly tolerant to glucose, with less than 20% reduction in activity up to 400 mM and 66% of the enzyme activity was retained even in the presence of 700 mM glucose (Fig. 4.15c). However, it was observed that the enzyme was less tolerant to galactose and the activity of the enzyme decreased as the concentration of D-galactose increased above 100 mM (Fig. 4.15c). About 44% of its relative activity was retained at 700 mM galactose. The kinetic parameters of the enzyme towards *o*NPG in the presence of galactose as an inhibitor were determined from time course assay using *o*NPG (1mM) as substrate and in the presence of different concentrations galactose (Fig. 4.15d). The β -galactosidase activity of BAD_1582 towards *o*NPG declined correspondingly with increasing concentrations of galactose. The K_i , K_m , V_{max} , and K_{cat} of BAD_1582 using *o*NPG as a substrate and galactose as inhibitor were 115 mM, 1.0 mM, 20.4 $\mu\text{mol min}^{-1}\text{mg}^{-1}$ and 39.7 s^{-1} , respectively (Table 4.3). Having a relatively high K_i for galactose makes this enzyme an ideal candidate for its application in the enzymatic hydrolysis of lactose, due to its limited inhibition by its hydrolysis products.

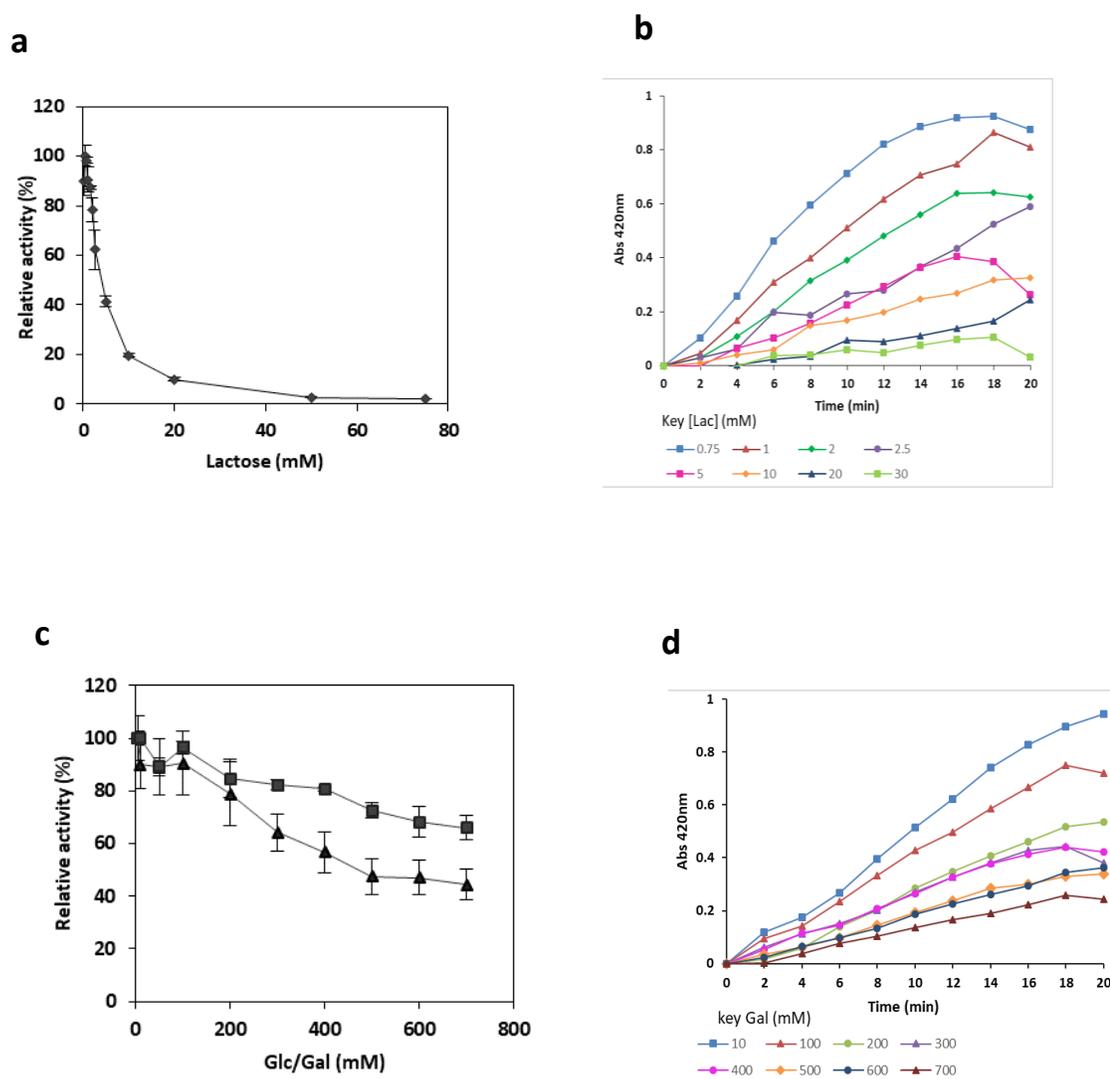


Fig. 4.15 Effect of lactose and its hydrolysis products on activity of BAD_1582 towards hydrolysis of *o*NPG after 30 min incubation. **a** Effect of lactose on the activity of BAD_1582. **b** time course assay of BAD_1582 in the presence of different concentration of lactose and 1mM *o*NPG. **c** Effect of D-glucose (squares) and D-Galactose (triangles) on the activity of BAD_1582 after 30 min incubation. **d** time course assay of BAD_1582 in the presence of different concentration of galactose and 1mM *o*NPG.

4.8. Degalactosylation activity of BAD_1582

The degalactosylation activity of BAD_1582 was determined using asialofetuin as substrate and fetuin as a control. The asialofetuin contains terminal Gal residues (Fig 4.16a), which could be available for β -galactosidases, while in the case of fetuin the terminal sialic acids prevent the accessibility of Gal residues and hence no degalactosylation could be possible.

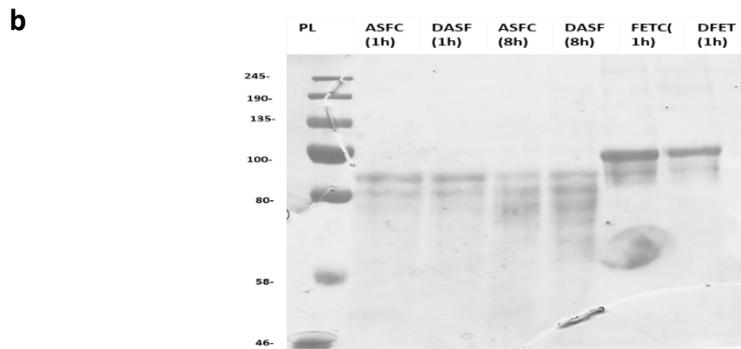
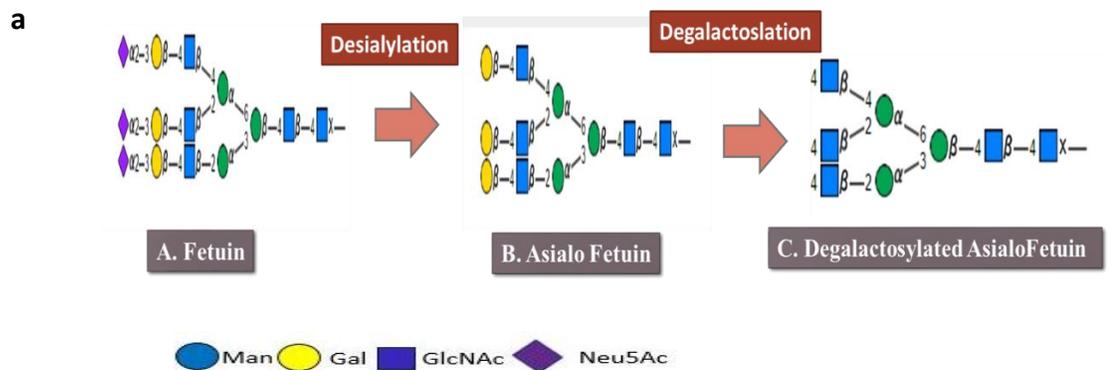


Fig. 4.16. Deagalactosylation activity of BAD_1582. **a** schematic representation of fetuin, asialofetuin and degalactosylated asialofetuin. **b** SDS PAGE analysis for detection of degalactosylation activity of BAD_1582 using asialofetuin as a substrate. PL, Protein ladder, ASFC asialofetuin control without enzyme; DASFC Degalactosylated asialofetuin with BAD_1582. FETC, Fetuin control without enzyme. DFET, Degalactosylated Fetuin with BAD_1582.

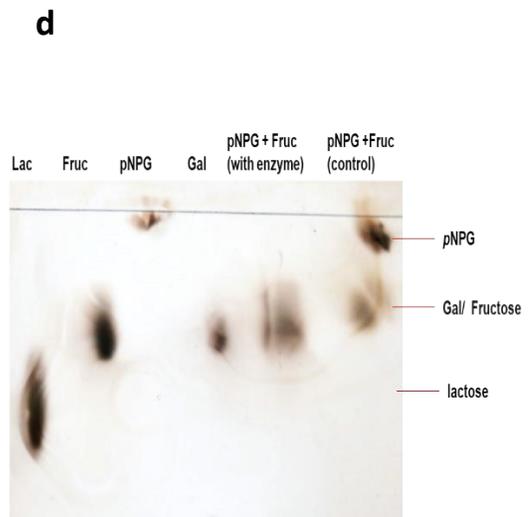
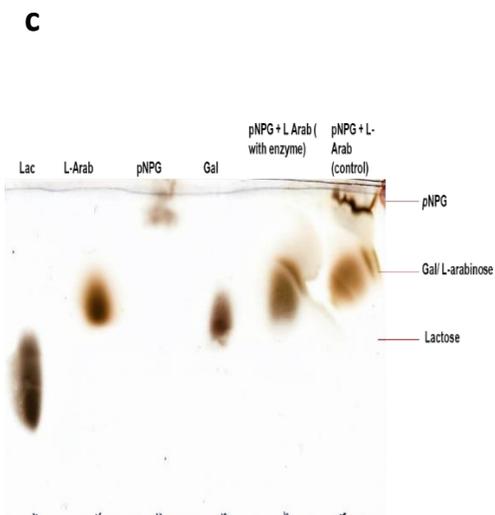
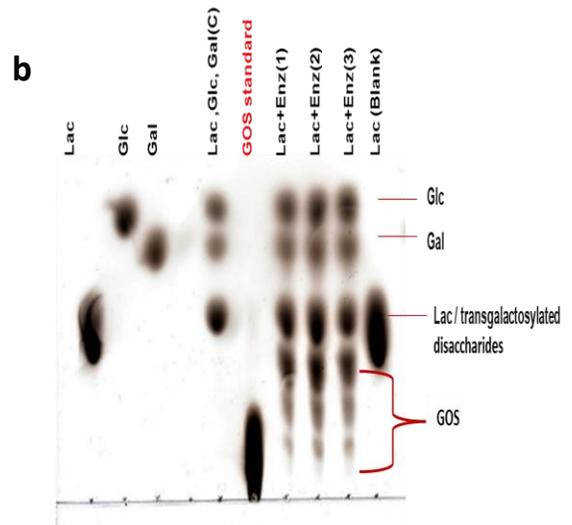
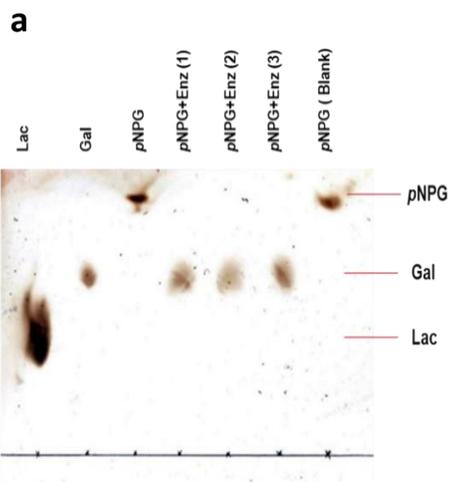
The enzymatic treatment of asialofetuin and fetuin with BAD_1582, after 1h and 8h incubation showed no size difference compared to the asialofetuin and fetuin controls incubated without enzyme. This indicates that BAD_1582 did not undertake degalactosylation of asialofetuin (Fig 4.16b).

4.9. Transgalactosylation activity and kinetics of GOS synthesis

β -galactosidases carry out transgalactosylation reaction during lactose hydrolysis, when initial lactose concentration is high (Guerrero *et al.*, 2015b). The transgalactosylation activity result in the synthesis of disaccharides (other than lactose) and higher DP oligosaccharides, known as GOS (Torres *et al.*, 2010a). To determine the transgalactosylation activity of BAD_1582, different sugars as galactosyl donor and acceptor were examined. In the first approach *p*NPG (100 mM) as galactosyl donor and various sugars such as (arabinose, glucose, galactose, fructose, GlcNAc and lactose) at 200 mM concentrations as acceptors; lactose (234 mM) were used both as galactosyl donor and acceptor; lactose (200 mM) as galactosyl donor, fucose and sialic acid (100 mM) as acceptors were examined. All the assays were carried out in 50 mM sodium phosphate buffer and contained 8.75 $\mu\text{g ml}^{-1}$ purified BAD_1582, incubated at 37°C for 24 h. After that, the enzyme was inactivated by heating at 95°C for 5 min. Thin layer chromatography (TLC) was conducted to examine the hydrolysis and transgalactosylation activity of BAD_1582. The GOS and the hydrolytic products of lactose i.e. glucose and galactose were distinguished by their retention factor (*R_f*) values compared to the standard sugars that were run together on the TLC plates. The enzyme hydrolysed both *p*NPG and lactose, since the released galactose and glucose (from lactose hydrolysis) and galactose but not *p*NP (from *p*NPG hydrolysis) were detected on the TLC plates (Fig. 4.17a and b). This confirms the enzyme acted on β -linked galactose. *p*NP was not visualized on the TLC because the visualization reagent (10% sulfuric acid in ethanol) used could not react with phenol and hence not detected, 1% FeCl_3 in water and ethanol (50% each) is visualizing reagent for phenols. Synthesis of GOS, occurred only in the assays containing lactose as galactosyl donor and acceptor but not in other assays (*p*NPG as Gal donor; Glc, Gal, fructose, arabinose and GlcNAc as acceptor) (Fig.4.17c and d). The three bottom spots indicated in Fig. 4.17b are GOS with different DP (with the higher DP at the

bottom), and disaccharides (presumably allolactose) in line with the lactose spot were also detected. The very top two spots belong to Glc and Gal respectively. The GOS produced using BAD_1582 had equivalent R_f values to the standard GOS used.

With lactose as a substrate, kinetic studies showed that the transgalactosylation activity of BAD_1582 precedes its hydrolytic activity. The GOS with a low degree of polymerization (DP) started to accumulate after 4 h of incubation but low hydrolytic products were detected on the TLC at this time point (Fig. 4.17e). However, the DP of the GOS started to increase in a time dependent manner, with higher DP detected at 24 h incubation and greater amounts of Gal and Glc (hydrolysis products) were also detected at this time point and the amount of lactose is considerably decreased. After 72 h incubation a very low amount of residual lactose was left (Fig. 4.17e). In addition, at 72 h the accumulation of GOS markedly declined compared to the preceding time of incubation which suggests the enzyme can also hydrolyse GOS. Overall this enzyme could enable this *Bifidobacterium* species to sequester the carbon from lactose by converting it to GOS and prevent it from accessibility to other lactose hydrolysing bacteria in the human gut. Thereby confer a selective advantage and enable these bacteria to excel in the competitive human adult gut environment.



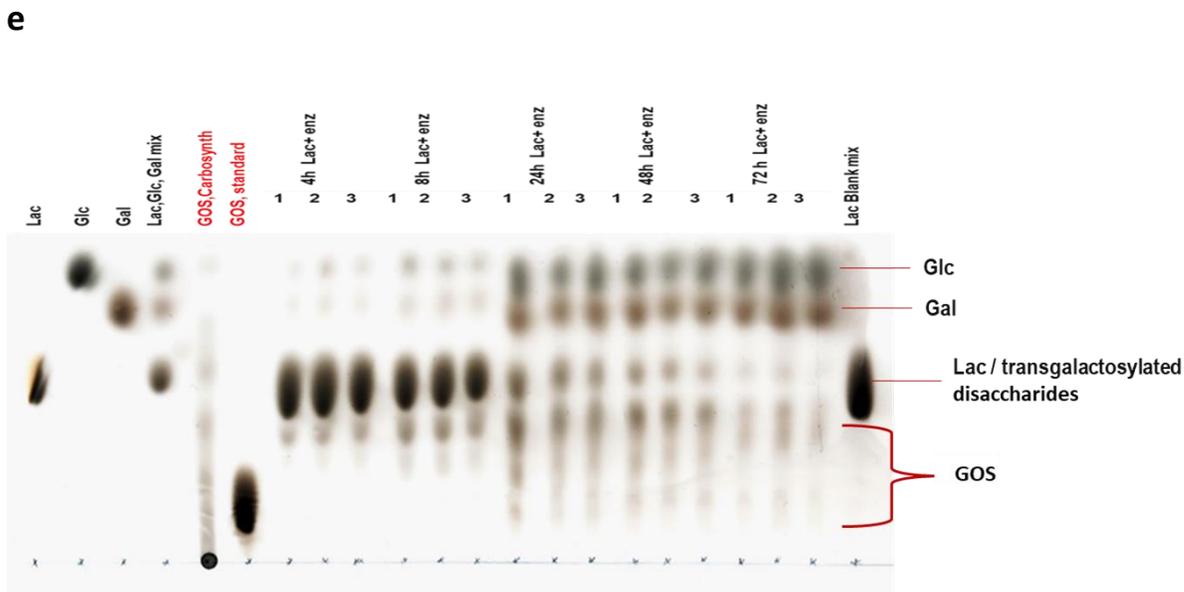


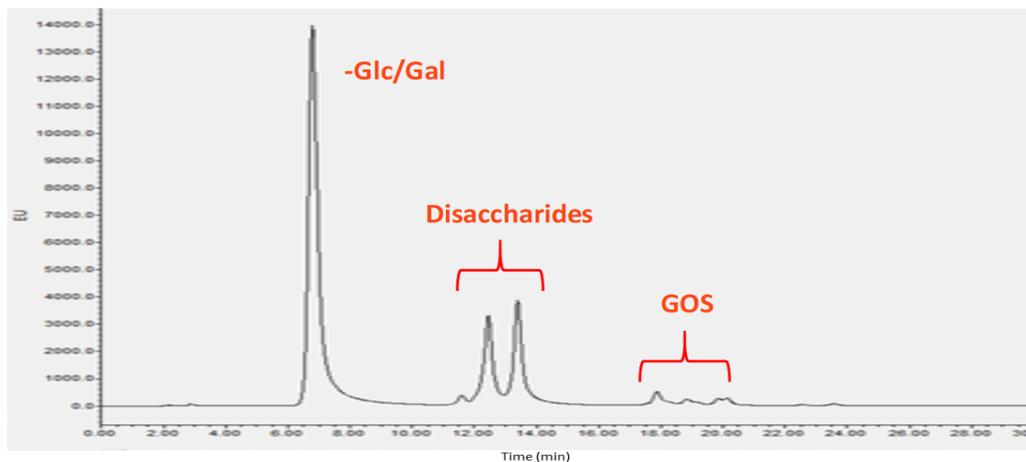
Fig. 4.17 TLC analysis of transgalactosylation and hydrolysis activities of BAD_1582. Hydrolysis of *p*NPG after 24 h incubation in the presence of 1.5 unit ml⁻¹ enzyme and control without enzyme at pH 7 and at 37°C **(a)** Transgalactosylation reaction products with lactose (234 mM) both as donor and acceptor after 24 h incubation with 1.5 unit ml⁻¹ enzyme and a control without enzyme at pH 7 and at 37°C **(b)** transgalactosylation reaction using *p*NPG as Gal donor, L-arabinose **(c)** and fructose **(d)** as acceptor. Kinetics of GOS synthesis of BAD_1582 using lactose as donor and acceptor, monitored between 4 h-72 h incubation at pH 7 and at 37°C, the numbers 1, 2, and 3 represent technical replicates for each time point **(e)**

HPLC analysis of the GOS synthesized using BAD_1582 was carried out using Waters HPLC system (USA). Samples (GOS synthesised with BAD_1582 and standard GOS) were labelled with a labeling reagent (comprised of 2-aminobenzamide and sodium cyanoborohydride (NaCNBH₃)). The labeled samples were purified using GlycoCleanTMS cartridges to remove non-glycan contaminants following manufacturers instruction. Eluted glycan samples were freeze dried and used for HPLC analysis. These samples were dissolved in HPLC grade water, 1 in 10 diluted sample and undiluted sample were transferred to HPLC vials and placed into auto sample unit of the HPLC. The run was

carried out for 24 h, the eluted glycans were detected using the fluorescent detector unit. Individual sugars were identified by comparing the retention units with the standard sugars.

HPLC analysis confirmed that GOS with various DP were synthesised through the transgalactosylation activity of BAD_1582. The column separates 2-AB labelled glycans based on their molecular size Fig 4.18a indicates the HPLC chromatograms of transgalactosylation and hydrolytic products. The first peak with elution time of 7 min corresponds to Glc and Gal as they were eluted first because of their smaller molecular size. The following two distinct peaks with elution time of 12-14 min, presumably represent disaccharides with different linkages eg. β -D-Galp-(1-6)-D-Glc (allolactose). The last three peaks with elution time of 18-21 min correspond to GOS with different DP and align with the peaks of the standard GOS (Carbosynth) run used as a control (Fig 4.18b). The molecular size of the GOS are higher and hence eluted last. Thus, the HPLC analysis is in parallel with the TLC data, which confirm BAD_1582 carry out efficient hydrolysis of lactose. Moreover, its transgalactosylation activity produces disaccharides with various linkages and GOS with various DP and linkages. Thus, the *B. adolescentis* derived β -galactosidase (BAD_1582) could be a potential candidate to produce lactose-free dairy products and prebiotic manufacture.

a



b

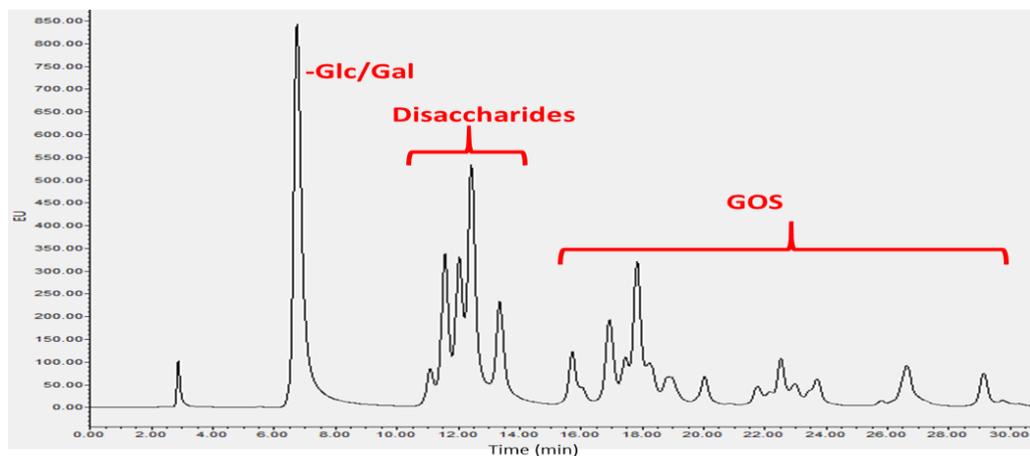


Fig. 4.18 HPLC analysis of GOS produced using BAD_1582 and GOS standard (Carbosynth). **a** HPLC analysis of transgalactosylation and hydrolysis activities of BAD_1582, with lactose (234 mM) both as donor and acceptor after 24 h incubation with 1.5 unit ml^{-1} enzyme at pH 7 and at 37°C. The first peak correspond to Glc/Gal, with elution time (7 min) the second set of peaks correspond to disaccharides with different linkages with elution time (12-14 min), and the third group of peaks correspond to GOS (DP) with elution time (18-21 min). **b** HPLC analysis of GOS standard (Carbosynth) showing the monosaccharides, disaccharides and GOS with various DP.

4.10 Discussion

In this study through function-based screening of a human faecal microbiome metagenome library, a new β -galactosidase enzyme from *B. adolescentis* was identified. Although several β -galactosidases have been identified and characterized, only a few are used for food industry applications, as the source microbes should be generally recognized as safe (GRAS) by the FDA (Adam *et al.*, 2004). Also, food enzymes should strictly fulfill the requirements of the food enzyme (FE) regulation of the European Food Safety Authority (EFSA).

Therefore, the most widely used β -galactosidases are obtained from *Kluyveromyces* spp. and *Aspergillus* spp. Even though these β -galactosidases have been widely utilized by dairy industries for removal of lactose and synthesis of GOS based prebiotics, these enzyme showed some limitations such as low affinity for lactose and product inhibition by Gal (Erich *et al.*, 2015). Thus, there is a demand for isolating new β -galactosidases which can surpass the aforementioned limitations. In this regard, we screened a human faecal microbiome metagenomic library for identification of β -galactosidases with industrially desirable properties such as higher affinity for lactose and lower inhibition by Gal.

While the gene sequence of BAD_1582 was identified during the genome sequencing of *B. adolescentis* (Suzuki 2006), it was not previously characterized. 29 fosmids which had incorporated different regions of the genome of *B. adolescentis* harboring BAD_1582 were selected in agar-based functional screening. However, β -gal clone 2 did not harbor the BAD_1582 gene, the sequences for this clone have no significant similarity in databases, and presumably it contained sequences from previously uncharacterized microbiome. The high abundance of this bacterium in the adult human gut microbiome likely contributed to the high number of clones containing this gene being selected (Milani *et al.*, 2015a). Indeed, analysis of the metagenomic library indicated that 47% of clones contained *B. adolescentis* DNA (Chapter 3). The method of DNA purification during metagenomic library construction might have favored the lysis of this *Bifidobacterium* strain. Some of the β -galactosidase encoding clones were not expressed by heterologous host, *E. coli* due to several factors such as codon bias, poor expression of molecular

chaperones, toxicity of proteins etc. (Kaur *et al.*, 2018; Lipinszki *et al.*, 2018b; Rosano and Ceccarelli 2014)

The genome analysis of *Bifidobacterium* species had revealed that their genome comprise numerous genes involved in carbohydrate metabolism (Milani *et al.*, 2015a). β -galactosidases with novel properties from *Bifidobacterium* species have been isolated and characterised through cloning and expression of the recombinant enzymes in *L. lactis* and *E. coli*. Among these, a few displayed transgalactosylation activities and hence synthesis of GOS. The *B. longum* BCRC 15708 β -galactosidase transgalactosylation activity produced two types of GOS; tri and tetra saccharides (with trisaccharide the dominant GOS) from 40% initial lactose at 45°C and pH 6.8 (Hsu *et al.*, 2007). The *Bifidobacterium breve* β -Gal I and β -Gal II produce GOS with β -D-Galp-(1-6)-D-Glc (allolactose) and β -D-Galp-(1-3)-D-Lac the predominant GOS respectively, at pH of 7.0 and 50°C (β -Gal I) and pH of 6.5 and 55°C (β -Gal II) (Arreola *et al.*, 2014b). *B. longum* subsp. *longum* RD47 had a maximum production of GOS at pH 8.5 and 45°C (Oh *et al.*, 2017) while *B. breve* B24 β -galactosidase showed 42% GOS yield from 1M lactose at pH 7.0 and 45°C (Yi *et al.*, 2011).

Phylogenetic analysis demonstrated that BAD_1582 had strong evolutionary linkages with β -galactosidases from other species of *Bifidobacterium*. The genome of *B. adolescentis* ATCC15703 contains eight galactosidase genes (https://www.ncbi.nlm.nih.gov/genome/proteins/683?genome_assembly_id=300284), of which two are α -galactosidases (BAD_RS08025/1528 (GH36) & BAD_RS08295/1576 (GH36)) and the remaining six are β -galactosidases (BAD_RS06400/1211 (GH42), BAD_RS07395/1401(GH42) (Hinz *et al.*, 2004; Van Laere 2000), BAD_RS07400/1402, BAD_RS08435/1603 (GH42), BAD_RS08325/1582 (GH 2) and BAD_RS08455/1605 (GH2).

A protein alignment of the six *B. adolescentis* β -galactosidases was created using MEGAX software with MUSCLE algorithm. The alignment is indicated below, conserved sequences are shaded.

	10	20	30	40	50	60	70	80	90	100
BAD_1582	--MADTAELAIVHATTASASW---	LTDPTVFAANRKP	AHSSHRYVIGETR	-EPK---	QSLDGEWKVRIEQARNV	VESAPFAAVDFEDGDFGAIEVPGH				
BAD_1211	-MSRNEVDHIL-----	FGAAYY-----	DEY---	LMMKGIDR-----	IEMKMMKDAGLNVIRIA-----					
BAD_1401	MSARRNFEWPELLTADGRGIA-----	FGGDYNP-----	DQW-----	SEDI-----	WIDIRLMKQAGVNTVALA-----					
BAD_1402	--MRYEVDLSEASKFARTPAYGIPVIRAAAACGPDGGTLGATNRFLTRDG--	SPW---	IPIAGEMHYSRIAERW	TELAKMAAAGI	VVSTY-----					
BAD_1603	MNQRREHRWRPRLGRRARIW-----	YGGDYNP-----	DQW-----	PEEV-----	WELVRLMVKAGVNLVSVG-----					
BAD_1605	--MANETRIEHASE-----	TW---	LADSTVFEVNRVPAHSDHKCYAHD	PQTNEWSDLRQSLDGEWRVEVVQASDIE	FNEEPFVRENF	DSAFERIQVPGH				

	110	120	130	140	150	160	170	180	190	200
BAD_1582	LQMAGYLKKNKYVNIQYFWDGHE	DPQAPNIPENN	HVAIYRRRE	FALDAQLARTLE	NDGTVSLTFHGAATAIYVW	DGTFVGYGEDG	FTPSEFDVTEALRNGN			
BAD_1211	-----	ESTWSTCEPQPG	-----	VFDET	-----	YVDRALDAAQRAGI	-----			
BAD_1401	-----	IFSWDRICPTED	-----	RWDFG	-----	WLDRIIDKLGAGI	-----			
BAD_1402	-----	VF-WNHHEEREG	-----	EWDFSGNR	-----	NVRRFVELCARHGLHV	-----			
BAD_1603	-----	IFSWAKIEPRED	-----	VYDFG	-----	WLDRIIDNLGKAGI	-----			
BAD_1605	LQMAGLMNNKYVNIQYFWDGHE	NPAPENIPENN	HVALYRKTET	TMANRLADTKNAGGTVSIVFHGMATAIYVW	NGMFVGYGEDG	FTPNEFDITEMLHDG-				

	210	220	230	240	250	260	270	280	290	300
BAD_1582	GNAADSPEAEHTT	ACYEYSS-----	ASWLEDQDFWRLHGLFRTVELAAQPHTHV	ETVQLEADYTAADTAGTADTAEL	NAALTLRNPADAMTIE	STLR				
BAD_1211	-----	DVIIGTPTYAV	-----	PSWLVKLDP	-----	SVLAVTPNGQ-GKYGARQIMDIVNATYRFG	ERVIRKLISHVA			
BAD_1401	-----	VWIASATATA	-----	PLWYESHP	-----	EVLPRDKYGHFVNAGSRQSWSPVVFKEYALT	CRKLAERYG			
BAD_1402	-----	IRGPFCHGEVRNGGLPDMLYGKSY	-----	-----	-----	EVRSLDAGF	-----	LDAVRGLYAHIAQQIRGLYFK		
BAD_1603	-----	AVDASATASP	-----	PMWLTQAHF	-----	EVLWKDYRGDVCQPGARQHW	RPTSPVFCEYALKICRAMAEHYK			
BAD_1605	-----	ENVVACYEYSS	-----	ASWLEDQDFWRLHGLFRSVELAAQPHVHIENMQIESDWDPE	SGSAS	-----	LDAAALTVRNAADAATISATLK			

	310	320	330	340	350	360	370	380	390	400	
BAD_1582	DGDGNVWVW	ESTQ--ACNGEIALNSGKMTNIA	EW-----	SAESPTLYTLTVRVVGH	DAI	ETVTQKIGFRTFRI	ENGIMTINGKRIVFKGADRHE	FD	AKR		
BAD_1211	DHPAVIGYQVDNETKYYD--SVSEDMQRLFKLREK	FHGD	LNELNHFGLDYWSNRIDSWED	-----	F-----	PDVTATINESLGGEF	---	DKFR	RQVR		
BAD_1401	TNPYVTAWHMGNEYGWNREDYSDNALEAFRAWCRRKY	-GTID	ALNQA	WTFWGQEMNGFDEVLIPRF	----	MGADSMVNP	GKLD	---	ERFG	NMMLL	
BAD_1402	DGGPIIAAQVDNEY-----	MASSAPW	-----	EMTTGISREWVPS--	GHDG	AGYLERLRDIA	----	IEEGI	-----	DPPMF	
BAD_1603	DNPYVVAWHVHGNEYGCHNRFDYSEDAERAFCQDWC	CEERY	-GTIE	AVND	AWGTA	FWA	QLNDFSEIVPPRF	----	IGDGN	FMPGKLLDF---KRESSDALK	
BAD_1605	DSDGNVWVWETAN--CADPDT	SISTGL	LN	GIREW	-----	SAEDP	VLYEF	EVTVIDHAGNIAE	VA	VCKVGFRRFR	IEDGIMTINGKRIVFKGADRHEFDPKR

	410	420	430	440	450	460	470	480	490	500
BAD_1582	RGRAITREDMLSDVVFCKRHNINAI	R	TSHYPNQ	EYWY	LCDEYGLYLIDETNMETH	GTW-VANN	ERP	EDGIPGSREPEWEGACVDRINSMMRRDYNHPS		
BAD_1211	RAFLQWADIVREYA-----	H	DQFITHNFDFEWRGYSYGVQPAVDHFKA	--	ATA	DIT	GV	-----	IY	HPT
BAD_1401	LDYKAERDAIAEIC-----	E	KPFTT-NFMVST	DQCCM	----	DYAAW--	AKENFVSND	-----	HY	FHEG
BAD_1402	FTCTGWCS	SPV-----	F	-----	L	PLWGGY	----	AYRPWLFYDGVGAEGMT	-----	E--HPA
BAD_1603	KSFYVAERDALAEIT-----	PEK	PLTT-NFMVS	AGGSVL	----	DYDDW--	GGENDFVSND	-----	HY	FIPG
BAD_1605	RGRAITEQDMIDDVVFCKRHNLN	AI	R	TSHYPNQERWY-ELCDEYGIYLIDETNLETH	GSW	CLPGD	VLTEETAVPGSKAHWEGACVDRVNSMVRDYNHPS			

	510	520	530	540	550	560	570	580	590	600
BAD_1582	VLIWSLGNE--S	SAGEVFRAMYRHA	HTIDPNRPVHYEGSVH	REFEDVTDIESRMYAH	ADEIERYLNDGSPAHTDGP	KKPYIS	CYMHAMGN	SCGNMDEY		
BAD_1211	EDDLTGKEI--A	FGDMTR	STK-----	N--	GQNYL	VI	ET	-----	EA	QGHW
BAD_1401	E--SHLDEL--A	CSDALM	DSL	A-----	L--	GKFWY	VM	EH	----	STSAVQ-W
BAD_1402	TDEYR	YRRLHGS	STSDG	FDP	PY-----	DPDSR	FYAC	EM	----	GGGMFNSY
BAD_1603	E--AHLDEL--A	FSASL	VDGIS	-----	R--	KDFW	FL	MEH	----	STSAVN-W
BAD_1605	VLIWSLGNE--S	YTG	DVFRAMYKRVHDIDPNRPVHYEGVTHNRDYN	DVTDIETRMYAHADAIEEYLK	----	NDPQ	KPYIS	CYMHAMGN	SCGNMDEY	

	610	620	630	640	650	660	670	680	690	700
BAD_1582	TALERYPMYQGGFIWDFIDQAIETKLPDGTTRM	CYGGDFG	---	DRPSDYEFSGDGLLFAD	---	RTPSPKACQ	VKQLYAN	VKIAVSVDEARITNDNLFVS		
BAD_1211	VPF-----PGQ-----LRLQAYSHLASGADM	NEY-WHH	---	SIHNSFETYWKGLLSHD	---	LEPNPTYR	HAGVFGREIA	--KPEV-GERLVHLKKN		
BAD_1401	KPL-NTRKRKGE-----TVRDSLAVHVMGADAN	NE-FQWR	---	ASAFGAEFHSAMVPHA	---	GEDTKLFRQV	CELGASLH	--TLAD-AGVQGT	ELAHS	
BAD_1402	DYRFVLPKRS-----VDAMANIKLGS	GDFELGY-YME	HGGTNP	KGAGFLNEGQTPKRSYDFQAPL	GKFGQVRESYRRLR	--	TLHEFVKA	FGPELAGM		
BAD_1603	RPI-NYRKEPGQ-----LVRDSLAVHVMGADAN	CY-FQWR	---	QSRSGAEKFSAMLPHA	---	GEDSQTFRD	VCELGRDLG	--TLAD-EGLLGT	KLAKS	
BAD_1605	TALERYPKYQGGFIWDFIDQAIYATCPDGTTS	RYGGDFG	---	DRPSDYEFSGNGLVFAD	---	RKPTPKACQ	VKQLYSN	VHIDVAEDSVTIKNDNLF	TS	

	710	720	730	740	750	760	770	780	790	800
BAD_1582	TGDYRFVLRILADGKPVWSTTRRF	DVAAGESASFEV	W	PVDDYRSNAE-ELVLE	V	SQQLGNACDWAPAGYELAF	GQC	VVAGAKTTADAV	DAAGAPADGTV	
BAD_1211	KVAIMVSNESLTALD--WFLIEAG-F	PPGGTLKYN	Q	VVRNVYDALFELNVECT	F	IPSDAPAERL--ASYEMIV	TPALYCAPQ	ETTDRL-REFV	SNGGHI	
BAD_1401	DTAILFSAES---E--WATRSQT-L	PSMKLNHWH	Q	VRDWYRAFLDAGSR	A	IVPL--AYDW--SSYKTV	VVLP	TVLILSAADTQRL-ADFAA	AGGR	
BAD_1402	PVSLQDGD	DRIDPSD--SETLRYAVRSD	GERGFVSI	INFQDHARVPSRHGEQ	I	VTTGTGERIRF--TGIGL	AEGENCIL	PFNMDLDG	V-RLV	RANLQFV
BAD_1603	SVAIVFDYES---E--WASEHTA-T	P	TQNVHHID-E-PLAWFRAL	ADVGVTA	V	VPI---RSNW--DEYD	VAILPSV	IILSEENTRRV	R	RDYVANGGKI
BAD_1605	TGEYTFVLRVLADG	EPVWQSERRFDVPAGSTEKLDV	W	PLDLYRDGAS-ELVLE	V	SQRLAKATNWAVAGYELAF	GQTVVAGS	SKKASAPV	----	KPVDGI

	810	820	830	840	850	860	870	880	890	900	
BAD_1582	TLGRWNAG	RGQGREALFSRTQG	---	GMVSYTFGEREFVLR	RRPS-ITTFRPLTDNDRG	AGHAFERA	AWAVAGKYARC	VDCAIANRGENA	-EATYT	----	
BAD_1211	VSTMRS-F	TDDEVTVWHDRAPH	---	N-LTDVFGMTYNQ	FTRPN-----	GHVSVEFAGALAE	---	TASTDAQSLIE	LNA-DADTE	VLASIGH	
BAD_1401	VVGYAT-G	IDEHFHTWLG	GGYPGAGDGLLR	SMLGVRGEEFNILG	--AEAGEPEGEIRL	SSADDSAA	---	LDGTTTRLWQ	N	NVTGEHAQLATYAGEEA	
BAD_1402	TVLR---	PGRPFRTFV	FLKPD---GMDELTFH	FEGDIQDVTVG	KADFERFM	VRDAAEIVCVTRA	---	LAGRMT-VLD	GG	LA	FVDDDNAIL--YED--G
BAD_1603	IATYYT-G	SDERDHWV	LG	GGYPG---S-IRDV	GVRIEEFAPMG	--SDWPGVPDHL	LDL	-----	NGAVAH	DIVDV	GSIGKDAKVLASEKD
BAD_1605	TVGRWNV	GGSGREVLLSRTQG	---	GLVSYTFNNREFVLR	RPA-VTTFRALTDNDRG	AGHGFERA	QWL	GAGRYARCIGNE	EQIDENT	-VKASYT	----

	910	920	930	940	950	960	970	980	990	1000																						
BAD_1582	YELAI	PQR	KVT	VRY	VAD	TAG	LVS	LDV	EYP	GKNG	DLPT	PAFG	IEW	ALP	VEY	ANL	RFY	GAG	PEE	TYA	RRH	AKL	GVW	STT	AGD	DCAP	YLL	LPQ	ETG	NHED	V	
BAD_1211	YAWK	---	DYAA	VTR	HAF	GKG	----	DAEW	VATL	-----	LDAD	SIRA	VMR	-----	EAVE	HAG	VEG	AGT	AL	-----	AGQ	AVR	QGV	NAL	GEI							
BAD_1401	DEWEL	-DG	TAA	VTR	NPY	GSG	----	EAYF	VGCD	-----	LDVAD	LTKL	NR	-----	AYLA	ASS	QEN	-----	ADVL	H--	TV	RAS	ADA									
BAD_1402	DGWT	LES	PRA	ENGI	AA	PEGL	IGS	PRP	VEP	VS	YARM	RRH	ADRY	EIAL	P-----	AGM	GRL	FRQ	RQV	-----	ADV	NL--	NI	AY	QGD	I						
BAD_1603	DPWT	GMD	GRPA	IVS	NPY	GEG	----	RSVY	VGAR	-----	LGRD	GIAR	SLP	-----	MILE	TG	VEV	KDSS	D-----	PDL	R	IER	VDE	STGA								
BAD_1605	YELAT	PQR	KVT	VSY	TAD	TTG	RVN	LHV	EYP	GEP	-GDL	PT	PAFG	IEW	TLP	VQY	SNL	RFY	GAG	PEE	TYA	RRH	AKL	GVW	STT	AGD	DCAP	YLL	LPQ	ETG	NHED	V

	1010	1020	1030	1040	1050	1060	1070	1080	1090	1100																							
BAD_1582	RWAE	ITDD	SGHG	VVR	KRG	GAG	PFAM	SLL	PPYS	STM	LEEA	LHQ	DELP	KPR	HMFL	R-----	LAA	QMG	VGG	D---	DSW	MSP	VHE	QY	QL	PAD	QPL	SLN	VQL	KL			
BAD_1211	NVTY	LLN	YSAD	EVTV	ASPI	AGE	VVV	VAP	VVI	ATD	GSID	EVA	NAE	VVLK	-----	EGAA	VKQ	GD	PLT	IGR	WN	VAV	IAG	-----									
BAD_1401	TFDF	YLP	-RG	KKT	VEL	QGIE	GE	PVIL	F-----																								
BAD_1402	GWLF	----	CGD	VLI	ADN	FCN	GET	WQV	GLRD	YADA	IDA	AAGG	KLTL	VVTP	LRR	GV	RKVT	SSMA	ARLE	QSD	ACV	ADV	TDV	YAT	PVY	HMP	VGE	-----					
BAD_1603	RFTF	LFN	R	TK	EP	V	SML	--VE	GRP	V	M	SLAD	CA-----																				
BAD_1605	RWAE	ITDE	KGH	GLR	ISRA	E	GAEP	FAM	SLQ	PYS	SFM	LEEA	LHQ	DELP	KPR	HMFL	RV-----	LAE	QMG	VGG	D---	DSW	MSP	VHP	QY	HI	PAD	QPI	SLD	V	LDL	DL	

The blast record indicated that BAD_1582 was distantly related to the five β -galactosidases, it has the highest similarity with BAD_1605 with 67.53% identity and E-value 0.0 followed by BAD_1211 with 60.0% identity and E-value 3.5. However, its similarity was less significant to the remaining β -galactosidases, with 33.33% and E-value of 1.8 and 1.7 for BAD_1401 & BAD_1402 respectively. Whereas, it was least similar with BAD_1603, with 26% identity and E-value 2.1.

The β -galactosidases BgaB/ BGal II with locus tag of BAD_1401 has been extensively studied for lactose hydrolysis and transgalactosylation activity, and had an optimal activity at pH 6.0 and 50°C showed low GOS production in the presence of 10 mg ml⁻¹ initial lactose concentration (Hinz *et al.*, 2004). Its expression was induced when *B. adolescentis* was grown on GOS supplemented media as sole carbon source (Akiyama *et al.*, 2015; Van Laere *et al.*, 2000). This enzyme was highly abundant only when the GOS is available in the media was active towards GOS (including from soy sources), but did not act on fructooligosaccharides (Van Laere *et al.*, 2000). On the other hand the other Bgal I β -galactosidase which was also expressed in the media containing GOS (with less extent) did not act on GOS rather it was active only on lactose (Van Laere *et al.*, 2000). When *B. adolescentis* was grown in GOS containing media high level expression of BAD_1605 was obtained (50 times higher than when *B. adolescentis* was grown in medium supplemented with glucose), BAD_1582, BAD_1401 were also expressed but at lower level (Akiyama *et al.*, 2015). Thus, the presence of several galactosidase enzymes in the genome of this bacterium supports the existence of possible specialization among these enzymes for various glycans as substrates, such as the host mucin oligosaccharides and dietary fibers within the gut. Bifidobacterial β -galactosidases with wide range of substrate specificities have been reported. The *B. breve* UCC2003 β -galactosidase Bbr_0010 hydrolyze lactose and lacto-N-neotetraose (LNnT) (Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc) while Bbr_1552 hydrolyze GOS, lacto-N-tetraose (LNT) (Gal β 1-3GlcNAc β 1-3Gal β 1-4Glc) and LNnT (James *et al.*, 2016). β -galactosidases of *B. breve* specialized on hydrolysis of plant derived galactans have also been reported (O'Connell Motherway *et al.*, 2011). 20 β -galactosidases identified from four infant derived *Bifidobacterium* species; *B. breve*, *B. bifidum*, *B. longum* subsp. *longum* and *B. longum* subsp. *infantis* were tested for their ability to hydrolyse various β -linked galactose containing substrates such

as galactobiose (Gal β 1-6Gal), galactobiose (Gal β 1-4Gal), LNT, LNnT, 2'-FL and 3'-fucosyllactose (3'-FL) (Ambrogi *et al.*, 2019). The *B. bifidum* β -galactosidases are among the glycoside hydrolases involved in the deconstruction of mucin O-glycans (Turrone *et al.*, 2010).

Besides the galactosidases, the genome of *B. adolescentis* ATCC15703 encodes two *lacS* (Lactose and galactose permease, GPH translocator family) genes (BAD_1188 and BAD_1604), which are galactoside symporters for translocating sugars and cations. The LacS symporter (BAD_1604) (a galactoside pentose-hexuronide permease) transports lactose and cations simultaneously using the potential energy of electrochemical gradients caused by cation intake (Akiyama *et al.*, 2015). These the *lacS* symporters have an equivalent function with the lactose permease of *E. coli*. In addition BAD_1582 gene is surrounded by the ABC transporter substrate-binding protein (BAD_1585), sugar ABC transporter permease (BAD_1586), and carbohydrate ABC transporter permease (BAD_1587) may have a role in the transport of sugars (lactose, and other oligosaccharides) in and out of the cell. However their expression was not induced during GOS supplementation, presumably these cluster of genes may not be necessary for GOS uptake (Akiyama *et al.*, 2015). However, the role of this transporters in the uptake of other kinds of glycans in the human gut environment needs further research. The LacI family transcription regulators (BAD_1584) and LacI-type transcriptional regulator (BAD_1189) as other known *lacI* families presumably encode for *lacI* regulatory protein (*lac* repressor) which controls the transcription of the *lac* operon by binding to the operator in the absence of lactose (Oehler 2009). BAD_1582 was sufficiently expressed using the transcriptional and translational machinery of the *E. coli* EPI300 host.

To examine the properties and application of BAD_1582, the gene was cloned with a His-tag for optimal induction and expression in *E. coli*. The His-tagged enzyme with a molecular mass of 117 kDa was obtained using Ni²⁺-NTA column chromatography and size filtration. BAD_1582 has optimum hydrolytic activity at pH 7.0 and 37°C using oNPG as a substrate. On the other hand, β -gal II (BAD_RS07395, GH 42) from *B. adolescentis*, has optimal activity at pH 6.0 and 35°C using pNPG as a substrate (Hinz *et*

al., 2004). Thus, different β -galactosidases display different properties even if they are isolated from the same strain of bacteria.

As a normal inhabitant of the human gut, the hydrolytic activity of this enzyme at normal physiological pH and temperature is compatible with the bacterium's environment. However temporary fluctuations from a fruit rich diet, acidic secretions from stomach, bile salts and SCFA resulting from fermentation of dietary fibers by resident microbiota might contribute to the change of the pH within the colon (Tomasello *et al.*, 2016; Tomasello *et al.*, 2014). Having a wide range of pH stability (pH 4-10) might contribute to better adaptation of this beneficial bacteria than the other members of the microbiota with β -galactosidases that exhibit a narrower range of pH stability. Most enzymes have a limited range of pH stability (usually close to their optimal pH) due to the irreversible loss of their tertiary structure. Nevertheless a few enzymes have been shown to have wide pH stability, such as trypsin with an optimal pH activity of 9.5, which is stored at acidic pH (pH 3) to prevent autolysis (Bisswanger 2014). On the other hand, the enzyme has low stability at temperatures above 40°C, which is also a feature that result from adaptation to the physiological temperature in the colon. In line with this, several β -galactosidases has shown a low stability above their optimal temperature (Liu *et al.*, 2017b). However, thermostable β -galactosidase (BbgIV) from *B. bifidum* displayed high transgalactosylation activity at 65°C (Osman *et al.*, 2012) and functional screening of metagenome DNA isolated from various environments uncovered thermostable and thermotolerant β -galactosidases (Maruthamuthu and van Elsas 2017; Zhang *et al.*, 2013).

BAD_1582 was reliant on divalent cations and displayed high activity in the presence of Mg^{2+} and Ca^{2+} ions but was strongly inhibited by Cu^{2+} and Mn^{2+} . Having higher activity in the presence of Ca^{2+} would be a desirable feature as natural milk is rich in this metal ion (Šimun Zamberlin 2012). The GH 2 β -galactosidases β -gal I and β -gal II of *B. breve* DSM 20213 have improved thermal stability in the presence of 1 mM Mg^{2+} (Arreola *et al.*, 2014c), which showed the Mg^{2+} is involved in stability of the enzyme rather than catalytic role. The GH 2 β -galactosidases from *Arthrobacter sp. 32cB* have highest activity in the presence of Mg^{2+} followed by Ca^{2+} but was inhibited in the presence of either Cu^{2+} and Mn^{2+} (Han *et al.*, 2014; Pawlak-Szukalska *et al.*, 2014). In contrast the

GH 35 β -galactosidase from *Akkermansia muciniphila* exhibit strong inhibition in the presence of Mg^{2+} (Guo *et al.*, 2018). There seems a similar pattern of divalent cation requirement between the β -galactosidases which belong in the same GH family of the CAZy database.

Similarly to other known β -galactosidases (Guo *et al.*, 2018), SDS strongly inhibited the activity of BAD_1582, however BAD_1582 was unaffected by β -mercaptoethanol, urea, and Triton-X-100. β -mercaptoethanol is added in many enzymatic reactions to enhance protein stability by inhibition of oxidation of sulfhydryl groups in the enzyme. BAD_1582 is unaffected by β -mercaptoethanol, which could indicate the inaccessibility of free sulfhydryl group which could potentially oxidise and de-stabilize the enzyme. This property could be useful for food industry application as addition of reducing agents such as β -mercaptoethanol could be toxic to consumers.

BAD_1582 displayed a narrow range of substrate specificity as it was only active on two of the chromogenic substrates tested. It cleaves only galactose, and furthermore only β -linked galactose, as evidenced by the high specific activity of this enzyme to *p*NPG and *o*NPG, but the absence of activity on α -linked *p*NP substrates. The enzyme has a strong preference for *p*NPG rather than *o*NPG. Different β -galactosidases showed different activity towards *p*NPG and *o*NPG. the four β -galactosidases from *Bacillus circulans* all display higher specific activity to *p*NPG than *o*NPG (Jingyuan *et al.*, 2011). On the other hand, a large number of β -galactosidases displayed higher activity to *o*NPG than *p*NPG (Wang *et al.*, 2010; Zhang *et al.*, 2013). The presence of distinct flanking motifs in the active site of three GH 42 β -galactosidase of *B. longum* have been shown to affect the distinct substrate specificities of these enzymes (Viborg *et al.*, 2014). Hence the structure of the active site contributes for the narrow range of substrate specificity of BAD_1582.

BAD_1582 has a V_{max} of $51 \mu\text{mol min}^{-1}\text{mg}^{-1}$ and a K_m of 2.5 mM towards *o*NPG. Nevertheless, as the concentration of *o*NPG increased, its hydrolysis was correspondingly decreased with a K_i of 62 μM . The inhibition of BAD_1582 by *o*NPG is in accordance with the heterodimer GH 2 β -galactosidase isolated from *Lactobacillus acidophilus* which displayed inhibition in the presence of high concentration of *o*NPG but with a relatively high K_i of 31.7 mM (Nguyen *et al.*, 2007). The catalytic efficiency (K_{cat}/K_m) of

BAD_1582 towards *o*NPG was $39.6 \text{ s}^{-1}\text{mM}^{-1}$ was lower than other β -galactosidases, for instance β gal I and β gal II of *B. breve* had a $(K_{\text{cat}}/K_{\text{m}})$ 722 ± 66 and $543 \pm 65 \text{ s}^{-1}\text{mM}^{-1}$ respectively (Arreola *et al.*, 2014a). Besides the soil metagenome derived β -galactosidase (ZD410) had a $(K_{\text{cat}}/K_{\text{m}})$ of $226 \text{ s}^{-1}\text{mM}^{-1}$ (Wang *et al.*, 2010). The low $(K_{\text{cat}}/K_{\text{m}})$ of BAD_1582 possibly resulting from the inhibitory effect of *o*NPG at higher concentration presumably due to steric hindrance developed near the active site of the enzyme.

One of the properties required in industrial enzymes is their tolerance to product inhibition. BAD_1582 displayed high tolerance to galactose ($K_{\text{i}}=115 \pm 2.7 \times 10^8 \text{ mM}$) and glucose. A couple of metagenome-derived β -galactosidases which showed higher tolerance to galactose with a K_{i} value of $197 \pm 56 \text{ mM}$ (Erich *et al.*, 2015) and K_{i} value of 238 mM (Zhang *et al.*, 2013) have been reported. On the contrary, the hydrolytic activity of some β -galactosidases toward *o*NPG showed strong inhibition by galactose with low K_{i} , for instance, β -gal I and II of *B. breve* have a K_{i} of 15 mM and 34 mM for galactose, respectively (Arreola *et al.*, 2014c). The β -galactosidase from *Aspergillus oryzae* has a K_{i} of 25 mM for galactose (Vera *et al.*, 2011). Having a high K_{i} for galactose makes BAD_1582 an ideal candidate to undertake enzymatic hydrolysis of lactose with less risk of product inhibition.

BAD_1582 can hydrolyse lactose efficiently even though the catalytic efficiency towards lactose was lower than *o*NPG. A couple of GH 2 and GH 42 β -galactosidases isolated from *Bifidobacterium* species and metagenome clones showed a higher catalytic activity towards *o*NPG than lactose (Arreola *et al.*, 2014c; Yi *et al.*, 2011; Zhang *et al.*, 2013). There is substantial evidence that lactose is a natural substrate for some GH1 and GH2 β -galactosidases (Adam *et al.*, 2004) while some GH 35 and GH 42 β -galactosidases act on different β - linked galactose containing glycosides. *B. adolescentis* may possess more than six β -galactosidases for specialization for different substrates. BAD_1582 did not degalactosylate asialofetuin, a glycoprotein with terminal Gal residues. This suggests that intracellular BAD_1582 may act preferentially on lactose and other small β -galactosides as a natural substrate but not extracellularly on proteoglycans. Transgalactosylation activity of BAD_1582 using lactose as donor and acceptor produced disaccharides and GOS with different DP. However, this enzyme did not transfer Gal from *p*NPG to different sugars (such as Glc, Gal, GlcNAc, L-arabinose and fructose) as acceptor.

Presumably the acceptor sugars might not interact into the active site of BAD_1582, due to steric hindrance created by the active site of this enzyme to discriminate its substrates and hence the enzyme could not catalyse the transfer of Gal to these sugars.

BAD_1582 synthesise GOS at relatively low concentration (234 mM) compared to known β -galactosidases which require the presence of high concentration of lactose (40% w/v, or 1.17 M) to undertake efficient transgalactosylation (Guerrero *et al.*, 2015a; Hsu *et al.*, 2007; Oh *et al.*, 2017; Reuter *et al.*, 1999). Hence production cost of GOS using BAD_1582 would be less. The ability of this enzyme to synthesis GOS preceding its hydrolysis presumably is a feature required for carbon sequestration under physiological condition in the human gut. This provides a selective advantage in interspecific competition over non-GOS fermenters (commensals) or pathogens. On the other hand if synthesized GOS are exported outside of the cell of this bacteria to the gut environment, it might enhance cross-feeding of GOS fermenting commensals mainly *Bifidobacterium* species (Turrone *et al.*, 2018). Most of the health benefits of GOS arose from being selectively fermented by the beneficial microorganisms such as *Bifidobacterium* and *Lactobacilli* (Macfarlane *et al.*, 2006). These bacteria in turn protect the host against enteric infections, increased mineral absorption, and immune system modulation (Gibson 2004). Moreover, the fermentation products of GOS by these bacteria result in SCFA synthesis which have various significance to the host, such as energy transduction in the colonocytes, regulate lipid and carbohydrate metabolism and provide energy to kidney, muscles and brain (Torres *et al.*, 2010b).

In the present work, the novel β -galactosidase enzyme from *B. adolescentis*, BAD_1582, was discovered in a human faecal microbiome metagenomic library. Its wide pH stability, stability at refrigeration storage conditions and high tolerance for Gal make it a desirable enzyme for industrial usage of removing lactose from dairy products. Furthermore, the transgalactosylation activity of the enzyme demonstrates the potential for further utility of prebiotic manufacture. The successful discovery and expression of a functional β -galactosidase with valuable properties from a human faecal microbiome metagenomic library demonstrates that human *in vivo* microenvironments are rich and important sources of industrially useful enzymes.

Chapter 5

Immobilisation of the *Bifidobacterium adolescentis* β -galactosidase BAD_1582

5.1. Introduction

Because of the advances in technologies and expansion of biotechnological industries there is a growing demand for enzymes which can operate at industrial scale and settings. Various industries such as biofuel industries, paper manufacturing industries, detergent industries, leather industries, textile industries, food industries, operate at extreme conditions of pH and temperature (Satyanarayana *et al.*, 2005). Enzymes which are resistant to extremes of pH, temperature, salt concentrations and pressure (extremozymes) (Hough and Danson 1999) have been isolated from extremophiles (microorganisms that inhabit extreme environments, such as hot springs, deep-sea hydrothermal vents, alkaline lakes) for their application in detergent industries, leather industries, pulp and paper industries, food and dairy industries (Li *et al.*, 2012; Satyanarayana *et al.*, 2005). Functional metagenomics approaches has also been used for identification of extremozymes from screening clone libraries of metagenome DNA (Mirete *et al.*, 2016). Although nature has been the major source of extremozymes, throughout the years research has provided different techniques to enhance the activity and stability of enzymes for their application in various industries (Li *et al.*, 2012). These techniques include protein engineering, immobilisation, and chemical modification (such as phosphorylation) (Silva *et al.*, 2018). In medium engineering, chemical substances are incorporated in the operational medium to enhance stability and activity of industrial enzymes (Laane 1987). Some of the substances which are used as enzyme stabilisers include polyols, saccharides, surfactants, water soluble polymers such as poly vinyl alcohol, poly ethylene glycol (PEG), and poly ethylene imine (Balcão and Vila 2015). In protein engineering, the genes encoding for the target enzymes are manipulated through rational design or by directed evolution to improve the properties of the enzymes (Porter *et al.*, 2016). The rational design approach uses computer modelling to predict structure and/or functional domains of proteins and uses this model to alter amino acids by site directed mutagenesis to obtain desired properties in enzymes. In directed evolution, mutant libraries are created by random mutagenesis of the gene encoding the target enzyme and screened for desired properties (Jacques *et al.*, 2017).

The most widely used chemical modification is crosslinking of enzymes with glutaraldehyde (Migneault *et al.*, 2004). Glutaraldehyde is a linear five carbon di-aldehyde, which is soluble both in water, alcohol and organic solvents. The carbonyl group reacts with the amine group of the amino acids in the peptides or proteins generating thermally and chemically stable cross links (Migneault *et al.*, 2004). Glutaraldehyde cross linking is used together with immobilisation procedures for enhancing the stability of enzymes. Pegylation of enzymes is another chemical modification where the peptides or proteins are modified by linking one or more methoxy polyethylene glycol chains, resulting in increased stability (Veronese 2001).

One of the primarily used technique to stabilise and enhance the activity of enzymes for industrial application is enzyme immobilisation (Mateo *et al.*, 2007). The term immobilised enzymes refers to “enzymes confined or restricted in a certain region or space with retaining its catalytic activity that can be use used repeatedly or continuously” (Katchalski-Katzir 1993). Since enzymes dissolved in a liquid medium are hard to retain or recycle, immobilisation of enzymes is the best strategy to circumvent these problem and enabled to reuse or recycle enzymes for repeated industrial processes which ultimately save excessive cost of enzyme preparation (Kasche 1999). Generally, immobilisation procedures result in restriction of enzyme mobility and consequently could also result in restriction of diffusion of substrates to the active site, a phenomena termed as mass-transfer effects (Kasche 1999). In such cases a reduced catalytic efficiency/reaction rate will be exhibited by the immobilised enzyme as compared to the soluble/free enzyme.

Immobilisation techniques can be broadly categorised in to two major classes; carrier bound and carrier free immobilisation, each with its own advantages and disadvantages (Cao *et al.*, 2003). In carrier bound immobilisation the enzymes are confined in an insoluble carriers for ease of separation and recyclability of the enzymes from the operational media (Cao 2005). The interaction between the carrier/ support and enzyme dictates the attainment of optimum properties of the enzyme, thus an ideal support should consider physical and chemical properties such as resistance to compression, chemical stability, pore size, hydrophobic/hydrophilic properties, surface chemistry to match the

enzyme surface properties, biodegradability, readily accessible with low cost (Hanefeld *et al.*, 2009; Mohamad *et al.*, 2015)

There are various techniques of immobilising enzymes on carrier matrix including non-covalent adsorption and deposition, immobilisation using ionic interactions, covalent attachment, and entrapment in polymeric gel (Hanefeld *et al.*, 2009). Adsorption utilises the possible interactions between the enzyme and the surface of carrier through weak noncovalent interactions such as van der Waals, ionic and hydrogen bonding. Since these interactions are weak the enzyme active site are usually unaltered (Jesionowski *et al.*, 2014). The first immobilisation was performed in 1916, in which invertase (sucrase) was adsorbed on solid support of charcoal and aluminium hydroxide and retained same activity with the free enzyme (Homaei *et al.*, 2013). In non-covalent adsorption, enzymes with large hydrophobic surface area bound to the hydrophobic carrier through hydrophobic interactions and the interactions are further strengthened by van der Waals interaction (Messing 1976). Enzymes with large hydrophilic surface, such as glycosylated surfaces bound to a hydrophilic carrier such as celite, cellulose, silica gel, porous glass through hydrogen bonding (Datta *et al.*, 2013; Mateo *et al.*, 2007). The advantage of non-covalent adsorption is that it does not require pre-treatment or chemical modification of enzyme and the disadvantage is leakage/desorption of enzyme would happen in aqueous media, thus better to use this immobilisation in non-aqueous media catalysis (Hanefeld *et al.*, 2009). Most prominent example are immobilisation of lipases on hydrophobic support by adsorption, due to the abundance of surface exposed hydrophobic amino acid side chains on these enzyme allow hydrophobic interactions (Fernandez-Lafuente *et al.*, 1998). In non-covalent deposition, the carrier matrix is immersed in enzyme solution and the enzyme is deposited to the carrier through co-precipitation of the carrier and enzymes or separation of the aqueous phase via evaporation (Hanefeld *et al.*, 2009). Immobilisation of enzymes through ionic interactions relies on the surface charge of the enzymes mostly due to the isoelectric point of surface exposed charged amino acids and pH of the solution. Based on the predicted charge of the enzymes, the type of ion exchange carrier matrix will be selected, for instance negatively charged matrix will be used for positively charged surface enzyme or positively charged carriers for negatively charged enzymes (Cao 2005). On the other hand enzymes that contain high amount of imidazole ring or perhaps His-

tagged enzymes can be immobilised on a carrier matrix containing metal ions, particularly Co^{2+} , Ni^{2+} and Cu^{2+} (Nahalka *et al.*, 2003).

Covalent immobilisation of enzymes is among the most widely used technique and involves formation of covalent bond between the protein and the carrier, compared to non covalent adsorption the covalent linkages provide strongest linkage between the enzyme and the carrier (Cao 2005). These linkages usually involves the reaction between amino acid residues on the surface of the protein and functionalities on the carrier, for instance the amino group of the enzymes are utilised to make covalent bonding to the aldehyde or epoxide functional groups of the carrier (Mateo *et al.*, 2007). The advantage of covalent immobilisation is the leaching of enzyme is minimised (Brena *et al.*, 2013). Thus it can be applied for both aqueous and non-aqueous medium industrial processes, however the disadvantage of this techniques is, and enzymes are chemically modified and might result in reduction of activity (Hanefeld *et al.*, 2009).

The other techniques of carrier-bound immobilisation is entrapment, and refers to enzymes enclosed within carrier matrix without chemically modifying the enzyme (Trevan 1988). These techniques has been regarded as safe in terms of protection of the enzyme from the adverse effects of media, denaturing agents, or chemical modifications (Trevan 1988). Most widely used entrapment technique was encapsulation of enzymes in solo-gels, porous silica gels, agarose, polyacrylamide gels and calcium alginate, blend of gelatine and calcium alginate; by allowing the polymerisation of the gels in the presence of the enzyme (Braun *et al.*, 1990; Fadnavis *et al.*, 2003; Tanriseven and Doğan 2002; Trevan 1988).

Carrier free immobilisation doesn't require the presence of a carrier matrix to immobilise enzymes, rather enzymes are crosslinked with a cross linking agent such as glutaraldehyde to produce an insoluble cross-linked enzyme crystals (CLECs), cross-linked enzymes dissolved enzymes (CLEs), cross-linked enzyme aggregates (CLEAs) (Cao *et al.*, 2003; Roessl *et al.*, 2010). CLEs are made by crosslinking enzymes in solution, although it enhanced thermal stability of enzymes, CLEs were difficult to handle and displayed poor mechanical stability and significant loss of activity (Roessl *et al.*, 2010). The CLECs, are characterised by their higher stability in extreme pH, temperature and display efficient

catalytic activity in non-aqueous reaction medium compared to CLEs, free enzyme and some conventional immobilised enzymes (Sheldon 2007). However the primary disadvantage of CLECs is the need for crystallisation of enzymes which is extremely laborious and requires high level of enzyme purity (Sheldon 2007). The use of simple protein precipitation procedures such as ammonium sulphate precipitation produce protein aggregates, cross linking of the protein aggregates result a permanently insoluble enzyme called CLEAs (Sheldon 2007). To produce CLEAs first aggregate of enzyme proteins are prepared by changing the surface of the proteins such as hydration of the proteins or by precipitation of proteins with salts, organic solvents or ionic polymers, thereafter the aggregates formed are cross-linked with a cross linking agent (Wang *et al.*, 2015). Despite CLEAs are less costly to prepare than CLECs, CLEAs had some disadvantages such as precipitation of enzyme has to be optimised for each enzyme as there is no defined protocol that works for all enzymes, moreover CLEAs have mass transfer problem which resulted in poor accessibility of macromolecular substrates (such as are proteins and polysaccharides) to the aggregates and less catalytic activity (Wang *et al.*, 2011a).

Although there are various immobilisation protocols, there is no one immobilisation procedure that's ideal for every application. Rational combination of different immobilisation techniques has improved drawbacks that would result from using a single techniques, for instance enzyme leakage in entrapment technique have been overcome by crosslinking the enzyme and carrier matrix (Cao 2005).

β -galactosidases are very important enzymes particularly for production of lactose free dairy products, reduction of lactose content of whey and GOS through transgalactosylation BAD_1582 has demonstrated efficient hydrolysis of lactose, desirable for production of lactose free dairy products or whey and carry out transgalactosylation of lactose to produce GOS, which is one of the primary lactose derived prebiotics. To utilise this enzyme in industry setting for the industrial applications, it is desired to enhance the stability and recyclability of the enzyme which can considerably reduce the production cost of enzymes. Because most industrial enzymes are immobilised enzymes, immobilisation of BAD_1582 could be the best approach for

using this enzyme for industrial application. Herein, BAD_1582 was immobilised by entrapment in gelatin and calcium alginate carrier matrix, followed by cross-linking with glutaraldehyde and the properties of the immobilised enzyme was characterised. This method was chosen based on a literature survey and had been indicated to enhance the stability of enzymes without reduction of activity (Fadnavis *et al.*, 2003). Similarly, Freitas and colleagues (Freitas *et al.*, 2012) demonstrated that immobilised β -galactosidase from *Aspergillus oryzae* has an enhanced stability compared to free enzyme and with minimal enzyme leakage. Enzyme leakage has been reported when enzymes are immobilised single-handedly on calcium alginate matrix (Blandino *et al.*, 2001). In addition, this immobilisation procedure is cost effective, can be carried out in reasonable amount of time, the major constituents of the matrix (gelatin and calcium alginate) are non-toxic, can be used in food industry and are eco-friendly (Tumturk *et al.*, 2008).

5.2. Preparation of the immobilised *Bifidobacterium adolescentis* β -galactosidase

BAD_1582 and the *Aspergillus oryzae* β -galactosidase

The *B. adolescentis* β -galactosidase BAD_1582 was purified using Ni²⁺-NTA affinity chromatography, as described in Chapter 4. The purity of the enzyme was analysed using SDS-PAGE and Coomassie Brilliant Blue staining. The concentration and specific activity of the purified protein were quantified. In a similar fashion β -galactosidase from *Aspergillus oryzae* obtained from sigma (8 units/mg solid) were prepared in distilled water to obtain an equivalent specific activity as the Ni²⁺-NTA purified BAD_1582.

Purified BAD_1582, *Aspergillus oryzae* β -galactosidase and a control (sterile water) were mixed separately with 6.6% sodium alginate and 4.05% gelatin solution. The enzyme solutions in the gelatin and sodium alginate carrier matrix were pumped through 0.85 mm peristaltic tubing into an equal volume of 3.64% glutaraldehyde and 0.05% CaCl₂. The resulting beads of immobilised BAD_1582, *Aspergillus* β -galactosidase and the control were stored for 12 h at 4°C in glutaraldehyde and 0.05% CaCl₂ solution to increase crosslinking and hardening of the beads. After, the immobilised beads of the *Bifidobacterium* β -galactosidase were washed in 50 mM sodium phosphate buffer (pH 7) and stored in the same buffer until used. While the immobilised beads of the *Aspergillus* β -galactosidase were washed in acetate buffer (pH 4.5) and stored in the same buffer at

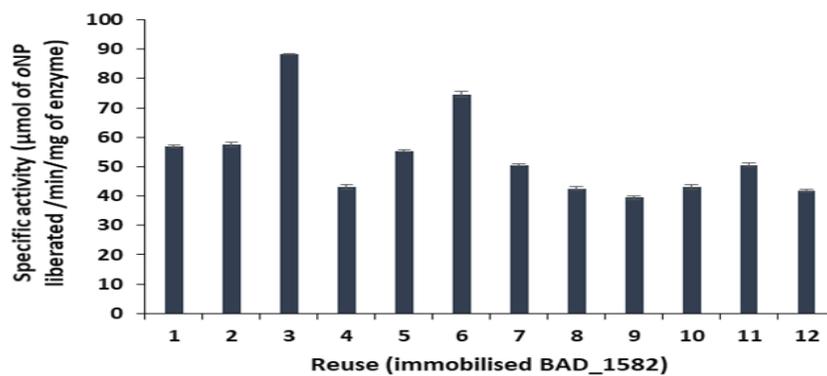
4°C till used. These buffers were selected based on the optimal activity of the enzymes, optimal activity of BAD_1582 was obtained in 50 mM sodium phosphate buffer (noted in Chapter 4), and while the acetate buffer were used for *Aspergillus* β -galactosidase based on manufacturers recommendations. The control beads were washed with water and stored at 4°C. The average size of the beads immediately after immobilisation range between ~ 65-75 mg, however a gradual increase in size were observed after wash steps.

The specific activity of the immobilised enzymes were determined in 50 mM sodium phosphate buffer (pH 7) (for BAD_1582) or 50 mM acetate buffer (pH4.5) (for *Aspergillus* β -galactosidase), 2 mM *o*NPG and 250 μ l of beads (measured by volume displacement) to make final volume of 500 μ l. The enzymes were incubated at 37°C for BAD_1582 and at 30°C for *Aspergillus* β -galactosidase for 30 min. The temperature and pH for BAD_1582 were selected based on its optimal activity obtained in this study (Chapter 4) while for the *Aspergillus* β -galactosidase manufacturer's instructions were followed. After 30 min incubation, the assay was stopped with 500 mM sodium carbonate, the amount of *o*NP released was measured at 420 nm and specific activity of the immobilised enzymes were determined. The specific activity of the free enzyme using *o*NPG as substrate was 37 μ mol /min/mg while the specific activity of immobilised BAD_1582 for *o*NPG was 57 μ mol /min/mg. The specific activity of free and immobilised enzyme were also determined using lactose as a substrate, for free enzyme a final reaction volume of 100 μ l containing 5 mM lactose, in 50 mM sodium phosphate buffer (pH 7), enzyme was added to initiate the reaction and incubated at 37°C. For the immobilised enzymes, the reaction volume of 500 μ l constitute 5 mM lactose and 250 μ l immobilised beads, in 50 mM sodium phosphate buffer (pH 7) and incubated at 37°C for 30 min. After that, the reaction was stopped by heating at 95°C for 5 min and the glucose liberated was measured using glucose-oxidase/peroxidase assay kit (described in Chapter 2). The specific activity of free BAD_1582 and immobilised BAD_1582 for lactose hydrolysis were 38.5 μ mol/min/mg and 89.5 μ mol/min/mg respectively. The immobilised enzyme showed ~57% increase in specific activity compared to the free enzyme when lactose was used as a substrate. These indicates the immobilisation not only enhanced stability rather it also enhanced its hydrolytic activity towards lactose without limitation of mass transfer.

5.3. Reuse of the immobilised BAD_1582 and *Aspergillus oryzae* β -galactosidase

After the first activity test, the beads were washed in phosphate buffer or acetate buffer, as appropriate, to remove residual *o*NPG, galactose and *o*NP, and the activity of the immobilised enzymes was measured again in fresh complete β -galactosidase assay solution with *o*NPG as substrate. This was repeated for 12 rounds and the specific activity of the beads after each round were measured. Both the BAD_1582 and the *Aspergillus* β -galactosidase showed minimal reduction in activity when the first use was compared to the successive 12 rounds (from 57 to 42 and from 61 to 46 $\mu\text{mol}/\text{min}/\text{mg}$ of enzyme added respectively) (Fig 5.1 a and b). Although a linear decline in activity was not detected, the activity of the immobilised enzyme showed a slight gradual decrement after 8th round of reuse for BAD_1582 and after 9th round for the *Aspergillus* β -galactosidase. After each rounds of reuse the beads gradually grew in size as more water from the wash buffer was absorbed in to the inner matrix which eventually led to some of the beads breaking apart and compromised the continuation of the activity assay after the 12th round. Overall, the results showed that, the immobilisation procedure using calcium alginate and gelatine cross linked with glutaraldehyde enhanced the recyclability of the enzyme for 12 rounds of catalysis, which confirm these procedures could potentially be utilised at industrial scale for application of BAD_1582 for lactose hydrolysis and GOS synthesis.

a



b

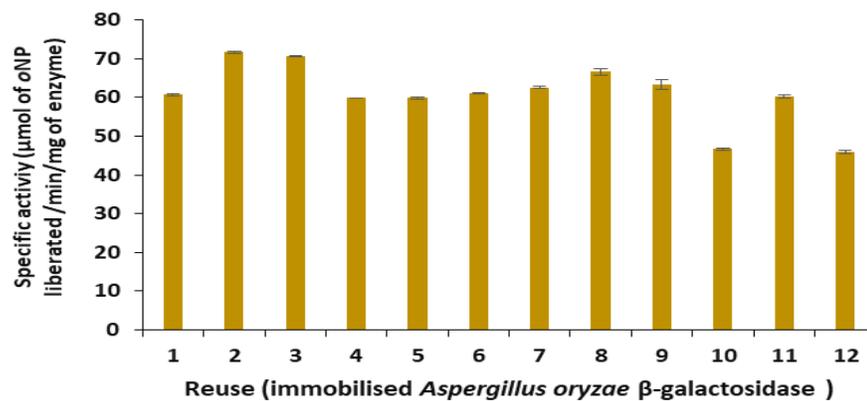


Fig 5.1 Resuse of immobilised BAD_1582 and *Aspergillus oryzae* β-galactosidase. **a** The specific activity of immobilised β-galactosidase from *B. adolescentis* (BAD_1582) and **b** the immobilised *Aspergillus oryzae* β-galactosidase determined for for the 12 rounds of reuse assay using oNPG as a substrate. The specific activity values represent the mean ± Standard deviations of three independent experiments/ from three batches of immobilised bead preparations.

5.4. Effect of pH and temperature on the activity of free and immobilised β -galactosidase enzyme BAD_1582

Since most immobilisation procedures change the properties of the enzyme and potentially enhance enzymes stability to industrial operational conditions, such as wide ranges of pH and temperature, thus it was necessary to examine the properties of the immobilised BAD_1582 by comparing its activity with the free enzyme. To determine how the immobilisation of BAD_1582 affect the property of this enzyme, the activity of free and immobilised BAD_1582 were examined on different pH and temperature. The effect of pH on the activity of the free and immobilised enzyme were determined in 50 mM sodium phosphate buffer adjusted to different pH using *o*NPG as a substrate as described above. The amount of free enzyme added per assay was 1 μ g (for pH activity) and 0.1 μ g (for pH stability), and for the immobilised enzyme beads which provide equivalent amount enzyme with the free enzyme were added. The optimum pH for free enzyme was pH 7 (in parallel with the data in Chapter 4) and the enzyme only retained 58% and 52% of its activity at pH 4 and at pH 8.5 respectively, however at pH 10 nearly the activity was abolished. On the other hand, the immobilised enzyme was fully active at all the tested pH ranges, and retained 93% and 81% activity at pH 4.5 and pH 10 respectively (Fig 5.2a). Therefore, the immobilisation procedure enhanced the enzymes pH resistance. Overall the immobilised BAD_1582 showed slightly higher activity than the free enzyme even though same amount of enzyme was added which confirm the immobilisation procedure used didn't restrict the movement of substrates and buffer in to the carrier matrix, which happens in most immobilisation procedure (Fig. 5.2 a and b).

In the case of pH stability, the free enzyme and the immobilised enzyme were incubated in 50 mM sodium phosphate buffer adjusted to different pH ranges and incubated for 24 h at 4°C. The following day all the pH were brought to pH 7 and the residual activity of the enzyme was assayed in the presence of 2 mM *o*NPG at 37°C for 30 min. The residual activity of the immobilised enzyme indicated that immobilised BAD_1582 was stable at all the tested pH (Fig 5.2b). The free enzyme showed slightly different residual activity, it retained $\geq 66\%$ of residual activity at acidic pH (pH 4-6.5) and showed very high residual activity at alkaline pH (pH 7.5-10) compared to pH 7. The free enzyme stability data

obtained here was different compared to the stability data obtained in chapter 4, the major contributing factor for this variability could be due to adjustment of sodium phosphate buffer at alkaline pH in this occasion could have accumulated a large amount of salt which possibly have affected the ionic interactions in the active site and yet unreversed when the pH was brought back to pH 7.

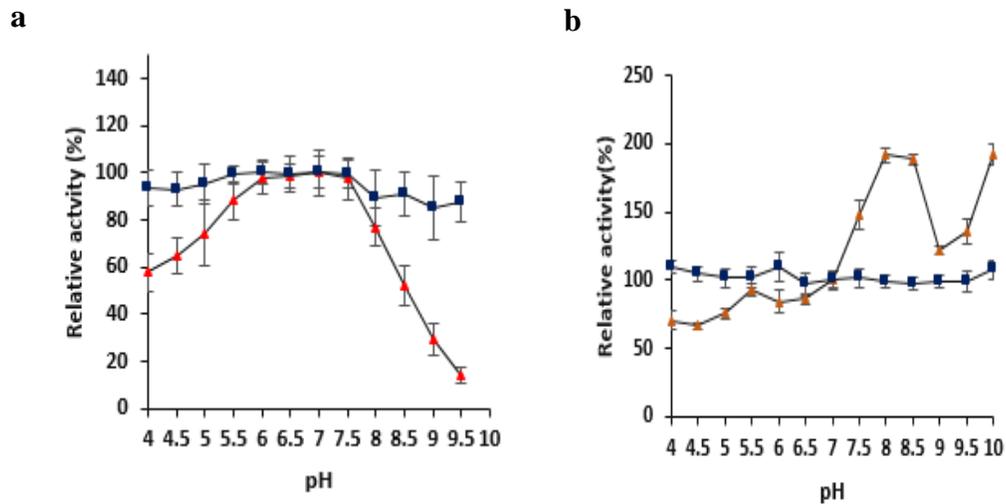


Fig 5.2 Effect of pH on activity and stability of free and immobilised BAD_1582 **a** Effect of pH on the activity of free (triangles) and immobilised (squares) BAD_1582. **b** pH stability of free (triangles) and immobilised BAD_1582 (squares). Values represent the mean of three independent experiments/ three bathes of immobilised beads and error bars are calculated from the averages of standard deviation of the three independent experiments.

The effect of temperature on the activity of the immobilised β -galactosidase (BAD_1582) enzyme and free β -galactosidase (BAD_1582) was carried out at a temperature ranges of 0-60°C. The highest activity for the immobilised enzyme was detected at 40°C while the free enzyme has highest activity at 37°C (Fig 5.3a). The immobilisation shifted the maximum temperature activity toward 40°C and retained 81% of its activity at 50°C but showed relatively low activity between temperature of 0-20°C. The highest activity of the free enzyme was obtained at 37°C and retained 68% and 41% of its activity at 40°C and 50°C respectively it was in line with the activity obtained at chapter 4 the enzyme

retained 60% between 20- 45°C The free enzyme showed high activity at lower temperatures between 0-20°C (38-92% of activity was retained), this was contrary to the data obtained at chapter 4.

Effect of temperature stability was determined by measuring the residual activity of the free enzyme and immobilised enzyme which were incubated at temperatures of 0-60°C for 1 h and were brought back to 37°C and incubated for 30 min. After that, the assay was stopped by addition of 500 mM sodium carbonate. The maximum temperature stability of both the immobilised and free enzyme was 37°C, yet about 48% and 74% of residual activity were retained at 40°C by the free and immobilised enzyme respectively (Fig 5.3b). Both the free enzyme and the immobilised enzyme were unstable at 45°C and above, however $\geq 61\%$ and $\geq 75\%$ residual activity was retained at temperature ranges of 0-37°C for the free enzyme and immobilised enzyme respectively. The residual activity of the free enzyme was in line with the data obtained in chapter 4 which retained 65% of activity at 40°C (nevertheless 37°C stability was not assayed, chapter 4). Overall, the immobilisation did not enhance temperature stability compared to the free enzyme except higher residual activity was detected at each temperature.

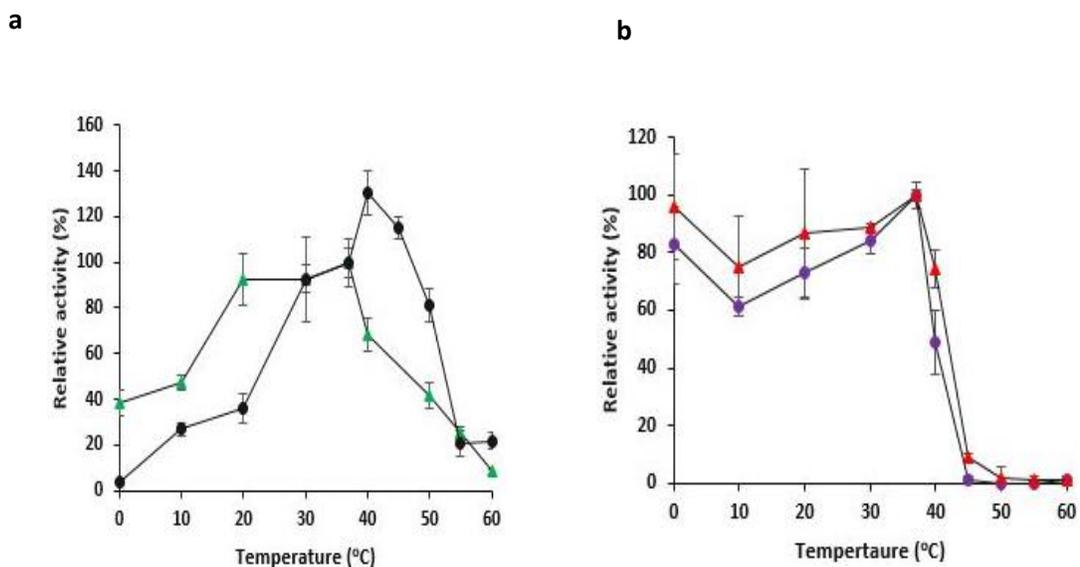


Fig 5.3 Effect of temperature on activity and stability of free and immobilised BAD_1582 **a** Effect of temperature on the activity free (triangles) and immobilised (circles) β -galactosidase (BAD_1582). **b** Effect of temperature on stability of free (triangles) and immobilised enzyme (circles) β -galactosidase (BAD_1582). The values represent the

mean of three independent experiments (three bathes of beads for immobilised enzyme) and the error bars are calculated from the average of the standards deviation of these experiments.

5.5. Kinetic parameter of the immobilised β -galactosidase (BAD_1582)

The kinetic parameters of the immobilised BAD_1582 was determined in 50 mM sodium phosphate buffer at pH 7 using different concentrations of *o*NPG (0.25-10 mM) as substrate in a time course enzyme assay. From the progress curve of the time course assay, no substrate or product inhibition was observed at the tested concentrations of *o*NPG within the first 18 min (Fig 5.4a). The K_m and V_{max} of the immobilised enzyme towards the *o*NPG were 0.81 ± 0.22 mM and 7.4 ± 0.63 $\mu\text{mol}/\text{min}/\text{mg}$ of enzyme respectively as calculated from the linear part of the progress curve (between 0-6 min) of the time course assay followed by non-linear regression fit using the Michaelis-Menten plot (Fig 5.4b). The immobilisation of the BAD_1582 has improved its tolerance to substrate inhibition, which was observed in free enzyme where the hydrolytic activity of free BAD_1582 markedly declined at *o*NPG concentrations higher than 2.5 mM, with K_m of 2.5 mM and V_{max} of 51 $\mu\text{mol}/\text{min}/\text{mg}$ of enzyme (Chapter 4). The K_m of the immobilised BAD_1582 towards *o*NPG has reduced three fold compared to the K_m of the free enzyme for the same substrate, which indicates the immobilisation has increased the affinity of the enzyme towards *o*NPG. Therefore, the immobilisation of the BAD_1582 have generated a desirable property for industrial application.

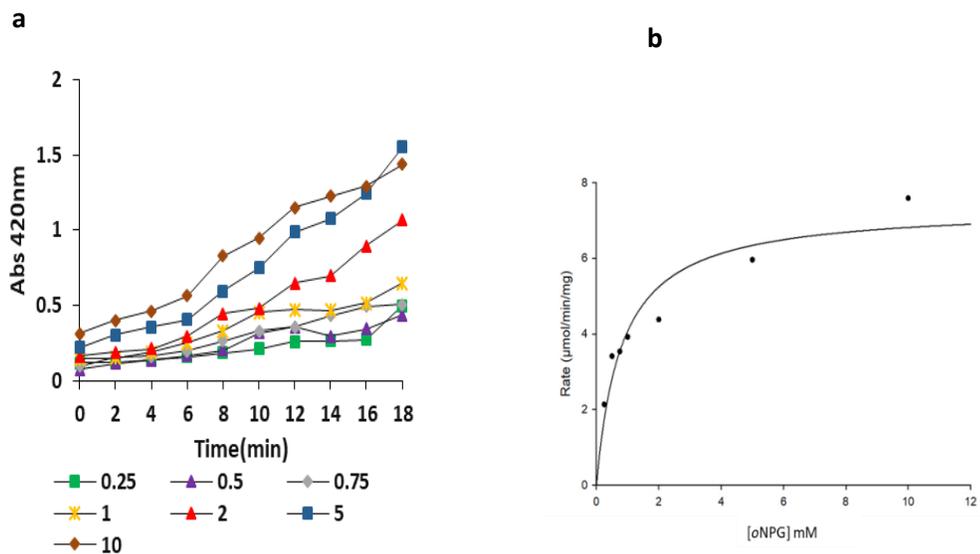


Fig 5.4 Time course progress curve and non-linear regression fit of immobilised BAD_1582 **a** time course progress curve of the immobilised β -galactosidase, BAD_1582 monitored at 2 min interval for 18 min, in the presence of different concentrations of *o*NPG (0.25-10 mM). **b** Non-linear regression fit model of Michaelis-Menten plot for determination of the kinetic parameters of immobilised β -galactosidase (BAD_1582).

5.5. Discussion

Immobilised enzymes have long history of success for their application in various industries as it enhanced stabilisation and recyclability of enzymes without compromising the efficiency of catalysis. In this study, the *B. adolescentis* β -galactosidase (BAD_1582) and the *A. oryzae* β -galactosidase has been immobilised in calcium alginate and gelatin beads in the presence of cross-linking agent glutaraldehyde. The immobilised BAD_1582 and *A. oryzae* β -galactosidase maintained 69% and 75% activity respectively after the 12th rounds of reuse. On the other hand Freitas and colleagues (Freitas *et al.*, 2012) reported that an immobilised *Aspergillus oryzae* β -galactosidase, using the same procedure maintained 80% of initial activity after 25 reuse. Although, in our case we have observed that, the washing steps have caused the carrier matrix to absorb more water/buffer which ultimately cause the beads to break apart after 12th round, and hence we could not continue the reuse assay as it affects the specific activity of the immobilised enzyme. However, the *A. oryzae* β -galactosidase immobilised beads of were slightly stronger and less fragile after the 12th round reuse. It was indicated that the pH of the medium dictates the mechanical strength of calcium-alginate- gelatin beads. The structure of alginate contains (1-4) linked β -D-manuronic acid and α -L-gluronic acid, which explains for their high stability at acidic pH and a low stability and degradation at basic pH (Tumturk *et al.*, 2008). Thus, at industrial scale the immobilised enzyme can work without decreasing efficiency at least to 12th rounds or more providing that the stringency of the wash steps could affect the mechanical shearing of the beads. Overall, a very low leakage of free enzyme was attained and hence the immobilisation procedure was efficient and, could be applied to industrial application. Enzyme leakage has been reported in cobalt-alginate immobilised β -galactosidase, which decreased the activity to 67.5% after first usage, while the treatment of these immobilised beads with 1% glutaraldehyde improved the relative activity to 83% and enzyme was stable up to 8 rounds of reuse (Mehmetoglu 1997). Enzyme leakage in calcium alginate beads are a common phenomenon due to the large pore size of the polymer and weak mechanical strength (Tumturk *et al.*, 2008). To overcome this, several techniques were developed, among this a blend of gelatine with calcium alginate and cross linked with glutaraldehyde provided beads with high comprehensive strength which can be reused several times without significant loss of

activity via enzyme leakage (Fadnavis *et al.*, 2003). Glutaraldehyde cross links the enzyme and gelatine and produces an insoluble structures, besides it stabilises alginate gel and hence leakage of enzyme is prevented (Mehmetoglu 1997; Tanriseven and Doğan 2002).

The immobilised BAD_1582 were equivalently active at all tested pH (4-10), however the free enzyme has an optimal pH of 7 was sensitive at acidic pH (pH 4-4.5) and alkaline pH (8.5- 10). Therefore, the immobilisation of this enzyme enhanced its activity to wide pH ranges, compared to the free enzyme. Contrary to this, immobilised β -galactosidase using calcium alginate-gelatine cross linked with glutaraldehyde did not change the optimum pH, but only slightly improved its activity at alkaline pH compared to the free enzyme (Tanriseven and Doğan 2002). Similarly, immobilised glucose isomerase in calcium alginate beads cross linked with glutaraldehyde demonstrated same pH optima with the free enzyme (pH 7.5), although slightly higher activity at alkaline pH was detected by the immobilised enzyme (Tumturk *et al.*, 2008). In line with this, immobilised xylanase in calcium alginate beads cross linked with glutaraldehyde have showed similar acidic pH optima with the free enzyme, but the immobilised enzymes were on the higher side (Bhushan *et al.*, 2015)

The pH stability, measured by residual activity demonstrated that the immobilised enzyme was stable after 24 h incubation at all tested pH ranges (pH 4-10), whereas the free enzyme have showed less stability at acidic pH and showed enhanced residual activity at alkaline pH. However, these data contradict the stability data obtained in Chapter 4, as the free enzyme was stable in all tested pH ranges. The predicted contributing factor for this variability could be ionic strength generated during pH adjustments which alter the tertiary structure of the enzyme via ionic interactions with key charged amino acids in the active site. On the other hand, the residual activity revealed that, the immobilised enzyme was active at all tested pH ranges which confirm the stability of the immobilised enzymes in all tested pH ranges. Overall the immobilisation enhanced the pH stability and activity of this enzyme.

The immobilised BAD_1582 had optimal temperature at 40°C while the free enzyme had optimal temperature at 37°C. The cross linking of the enzyme with gelatin in the presence

of glutaraldehyde could have altered the native structure of the enzyme and henceforth shifted its optimal temperature. In line with this immobilised β -galactosidase using cobalt-alginate cross linked with glutaraldehyde had shifted its temperature optima to 60°C, which was around 50°C for free enzyme (Mehmetoglu 1997). Both the free and the immobilised BAD_1582 had same optimum residual activity at 37°C. The free enzyme retained 48% residual activity at 40°C, its activity was completely lost at 45°C, on the other hand the immobilised BAD_1582 retained 74% and 18% activity at 40 °C and 45°C respectively, which implicate immobilisation increased a slight temperature tolerance compared to the free enzyme.

The kinetic parameters of the immobilised BAD_1582 revealed that the K_m was 0.81 ± 0.22 mM, which was three fold lower compared to the free enzyme data, which indicates the immobilisation had enhanced the affinity of the enzyme towards the synthetic substrate, *o*NPG. Being a two phase system, immobilised enzymes usually face an unavoidable mass transfer problem, which result in decrement of overall catalytic performances (Krajewska 2004). However, it was indicated that some immobilisation procedures may decrease the K_m of the enzyme, as a result of electrostatic interactions between the substrate and the carrier matrix used and diffusion effects (Demirel *et al.*, 2006). Besides, immobilisation of enzymes enhanced the activity of some enzymes, due to a phenomena called microenvironment effect (Cao 2005). For instance, immobilised lipase–surfactant complexes entrapped in *n*-vinyl-2-pyrrolidone gel matrix was 51-fold more active than the native enzyme (Goto *et al.*, 2005). In line with this the specific activity of the immobilised BAD_1582 has showed 2.3-fold increase in lactose hydrolysis compared to the free enzyme, which also approves the immobilisation enhanced the activity of this enzyme.

In summary, the immobilisation of BAD_1582 in calcium alginate-gelatin beads crosslinked with glutaraldehyde shifted the optimal temperature and optimal pH. The immobilisation seemingly avoided the mass transfer effect problem of most immobilisation procedures. In addition, the immobilisation procedure is simple and less costly. Thus, immobilised BAD_1582 using the aforementioned procedure, could suit the

operational conditions in industry for production of lactose free dairy products and also for manufacture of lactose-based prebiotics.

Chapter 6

Discussion

The human gut microbiome has been the centre of attention for the past decades (Cho and Blaser 2012; Gill *et al.*, 2006). Advances in meta o'mic technologies have revolutionised the study of human gut microbiome and expand the centre of research beyond taxonomic identification of microbial communities in the GIT to the metabolic networks that exist between the host and the microbiome (Cui *et al.*, 2018). Moreover, the availability of metagenome datasets have increased significantly in recent years as compared to the metagenome data sets available a decade ago. For example, in 2017 1003 reference genomes of bacterial and archaeal had been identified through genomic encyclopaedia of bacteria and archaea (GEBA) sequencing and phylogenetic identification (Mukherjee *et al.*, 2017). These data sets enhanced the availability of reference genomes for the identification of unidentified or unclassified taxa. Moreover prediction of metabolic networks associated with the host-microbiome interactions which could impact the development of various diseases have started to emerge due to such advancements (Proctor *et al.*, 2019). In spite of these developments, meta o'mic technologies need to be complemented by experimental data such as targeted gene characterisation through functional genomics studies. In addition, the predicted function of metabolic networks needs *in vitro* and *in vivo* confirmation. This PhD study has unveiled the potential of human gut environment for identification of GH enzymes through functional screening of a human faecal microbiome metagenome library.

6.1. Functional screening of human faecal microbiome metagenome library for glycoside hydrolase enzymes

Functional metagenomics has a proven record for the identification and characterisation of novel enzymes from uncultured microorganism that reside in various habitats (Heath *et al.*, 2009; Ilmberger *et al.*, 2012). Function-based screening of metagenomic libraries from soil samples has enabled the identification of various glycoside hydrolase enzymes, particularly novel β -galactosidases (Cheng *et al.*, 2017b; Wang *et al.*, 2010; Zhang *et al.*, 2013). However only a couple of function based screenings of human gut microbiome metagenome library has been carried out where the aim is the identification of novel CAZymes (Cecchini *et al.*, 2013a; Tasse *et al.*, 2010). These studies revealed enzymes involved in prebiotic and dietary polysaccharide metabolism from the human gut

microbiome, one particular example was the GOS hydrolysing β -galactosidases, which belong to GH 2 and GH 42 and showed significant similarity with enzymes from the *Bifidobacterium* genus (Cecchini *et al.*, 2013a).

In this PhD study, new strategies to identify glycoside hydrolase enzymes, particularly β -galactosidases, α -fucosidases and sialidases have been optimised and developed. The approach used was MacConkey base agar supplemented with different sugars including lactose and different HMOs (2'-FL, 3'-SL and 6'-SL) which were used for identification of metagenome clones expressing β -galactosidases, α -fucosidases and sialidases respectively. Although sialidase and α -fucosidase expressing clones were not identified from our metagenome library, with these screening strategies it was possible to identify a previously uncharacterised β -galactosidase (BAD_1582), from *B. adolescentis*.

Given that the metagenome library was dominated by Actinobacteria at the phylum level (discussed in Chapter 3), the absence of α -fucosidase and sialidase expressing metagenome clones from our metagenome library could be attributed to the low abundance of the *Bifidobacterium* species which are adapted in utilisation of HMOs and host mucin glycan degradation such as *B. bifidum* and *B. longum* subsp. *infantis*. The latter two *Bifidobacterium* species encode a repertoire of CAZymes which are active in mucin glycan degradation and utilisation of HMOs such as sialidases, α -fucosidases, and hexosaminidases, α -mannosidases and lacto-N-biosidase activities (Milani *et al.*, 2015a). *In vitro* growth studies using HMOs as a sole carbon source has indicated that *B. longum* subsp. *infantis* expresses two β -galactosidases (Bga42A and Bga2A) which are active in metabolism of type-1 (Gal β 1-3GlcNAc) and type-2 (Gal β 1-4GlcNAc) isomers of HMOs (Yoshida *et al.*, 2012). Thus, function-based screening for α -fucosidases and sialidases should be carried out with growth dependent approach using HMOs as sole carbon source thereby the expression of these genes could be induced to metabolise HMOs. Besides, it is important to consider the use of different expression hosts other than *E. coli*, as the absence of expression machinery and /or secretory systems in the heterologous host could compromise the identification of extracellular enzymes (Fakruddin *et al.*, 2013; Rosano and Ceccarelli 2014).

On the other hand, the metagenome library we used for this study was dominated by *B. adolescentis* which accounted for (~60%) of the sequenced clones. The glyco-biome of 18 sequenced *B. adolescentis* strains have revealed that, this species encodes a repertoire of CAZymes which comprised 36 glycoside hydrolases (GH), 12 glycosyl transferases (GT) and 4 carbohydrate esterases (CE) particularly active in uptake and utilisation of plant derived polysaccharides, such as starch, amylose, maltodextrin, pullulan etc. (Duranti *et al.*, 2016). In all of the 18 species, the GH 13 families which include range of activities (α -amylase, pullulanase, cyclomaltodextrin glucanotransferase, cyclomaltodextrinase, α -glucosidase etc.) (Lombard *et al.*, 2014a) was represented by high frequency. On the other hand *in silico* analysis of the 18 species uncovered that GH families which belong to GH 10 (endo-1, 4- β -xylanase), GH 29 (α -L-fucosidase and α -1,3/1,4-L-fucosidase) and GH 59 (β -galactosidase and galactocerebrosidase) were represented only by a single strain of *B. adolescentis* species namely, *B. adolescentis* 703B, *B. adolescentis* AD2-8 and *B. adolescentis* 70B respectively (Duranti *et al.*, 2016). The apparent absence of sialidases (GH 33) and α -L-fucosidases (GH 29 and GH 95) from most *B. adolescentis* strains which predominantly obtained in the metagenome library, was in line with this analysis. Thus, for future research focusing on isolation of α -fucosidases and sialidases from human gut microbiome, the following considerations should be taken: screening of metagenome libraries could be done on agar-based media containing mucin as a sole carbon source to induce the expression of the genes responsible for mucin glycan hydrolysis. For instance transcriptome analysis of *B. bifidum* PRL2010, isolated from infant stool, that was grown in mucin containing media had induced the expression of genes responsible for utilization of host derived mucin glycans, such as exo- α -sialidases, 1,2- α -L-fucosidase and 1,3/4- α -L-fucosidase, extracellular enzymes α -L-arabinofuranosidase and α -1,3-galactosidase (Turrone *et al.*, 2010). Secondly, having amplicon and whole genome sequence data of the metagenome DNA before construction of the metagenome library would provide knowledge on the extent of microbial diversity in the library. The more diverse the metagenome library the greater the chance of discovering novel sequences and novel functions. However, the function-based approach has also its own draw backs (discussed in chapter 4). Therefore, it is best to use both sequence based and function-based

approaches for the identification of the target genes as the approaches complement each other.

The glyco biome of the 18 *B. adolescentis* strains revealed that all strains constitute abundant number of β -galactosidases which belong to GH 2 and GH 42 families (Duranti *et al.*, 2016). Of the six or more β -galactosidases encoded by *B. adolescentis*, BAD_1582 was predominantly identified in this study from the function-based screening of the human faecal microbiome metagenome library. There are possible reasons for this, first of all, the promoter region of some of these β -galactosidases may not have been incorporated in the metagenome clones and hence the RNA polymerase of *E. coli* could not bind which would hamper expression and hence they were not identified. In addition, *E. coli* may not have specific molecular chaperones to enhance the correct folding of specific proteins from *B. adolescentis*, thus other β -galactosidases were unable to fold correctly and hence were not functional thus were not identified in function-based screening. The bioinformatics analysis using CDD, Pfam, of the protein sequence of BAD_1582, revealed that BAD_1582 belongs to the GH 2 enzyme families. Different reports showed that, lactose is presumably the natural substrate for most GH 2 β -galactosidases rather than other terminally attached Gal constituting β -galactosides (Husain 2010; Rodriguez-Colinas *et al.*, 2011). On the contrary, the GH 42 family act on wide range of substrates, for instance β -gal II of *B. adolescentis* hydrolyse Gal(β 1–4)Gal and Gal(β 1–4)Gal-containing oligosaccharides; with low activity towards Gal(β 1–3)Gal, lactose, and Gal(β 1–3)GalOMe (Hinz *et al.*, 2004). β -galactosidases have been widely used in food industries for their ability to hydrolyse lactose and for their transgalactosylation activity, allowing lactose-based prebiotics including GOS could be produced. β -galactosidases isolated from bacterial and yeast are mostly intracellular enzymes while fungal β -galactosidases are extracellular (Selvarajan and Mohanasrinivasan 2015). The PSORTb prediction, revealed that BAD_1582 is predicted to be an intracellular enzyme. In addition, the other five β -galactosidases encoded by *B. adolescentis* (BAD_RS06400/1211, BAD_RS07395/1401, BAD_RS07400/1402 and BAD_RS08435/1603, BAD_RS08455/1605) are predicted to be intracellular. Whereas, the localization score of BAD_RS08060/1534 is as yet unknown. The intracellular enzymes have access to various glycans through sugar transporting permeases, for

instance the LacS symporter, a galactoside-pentose-hexuronide permease transports lactose and cations simultaneously using the potential energy generated by electrochemical gradients of cations (Akiyama *et al.*, 2015). The expression of this permease has been increased in *B. adolescentis* grown in the presence of GOS, presumably the bacterium uses this permease for uptake of one or more components of GOS (Akiyama *et al.*, 2015).

6.2. Biochemical properties, substrate specificity and kinetic parameter of

BAD_1582

Enzymes are efficient bio catalysts and have been widely used in various industries such as baking, brewing, detergents, pharmaceuticals, leather processing, paper and pulp industries, biofuel production etc. (Singh *et al.*, 2016). The use of enzymes for industrial purpose have been widely researched, because enzymes offer several advantages such as they operate under mild operational conditions, have extreme product selectivity, low physiological toxicity and are eco-friendly (Chapman *et al.*, 2018).

Thus, purification of BAD_1582 were conducted on the cell lysate of expression clone. BAD_1582 had a wide range of pH stability (pH 4-10) while it had an optimal activity at pH 7, and had optimal temperature at 37°C and was stable up to 40°C, which correlates to the adaptation of the *B. adolescentis* under human physiological condition in the GIT environment. An interesting feature of this enzyme was its storage temperature stability, although its activity declined at lower temperatures, its activity was unaltered after six weeks storage at 4°C in a buffer without addition of cryoprotectant such glycerol. Hence, the stability of this enzyme at 4°C, would decrease the continuous production cost for industrial application, providing that the proper storage and handling procedures are followed.

BAD_1582 has a requirement for divalent cations particularly Mg²⁺ (74%) and Ca²⁺ (41%), which was confirmed by the decline of its activity by EDTA (74%), a chelating agent which chelates divalent metal ions bound to the enzyme active site. Structural prediction of amino acid sequences of BAD_1582 from Pfam indicate this enzyme as a GH2 family enzyme containing TIM barrel domain (which is a catalytic domain, with

catalytic residues). Most GH2 enzymes harbour the catalytic nucleophile/base and catalytic proton donor amino acid, Glutamic acid (E.461/ E.537) in their active site which was demonstrated for an *E. coli* β -galactosidase (Cupples *et al.*, 1990; Gebler *et al.*, 1992). Presumably BAD_1582 could also use either of the glutamic acid residues as a nucleophile, because of the conserved sequences of GH2 families. Most GH2 β -galactosidases have a requirement for Mg^{2+} , as the binding of this cation in the active site assists to appropriately place acidic /basic residues in the active site. Furthermore, homodimer structural predictions using SWISS MODEL showed that (Proline, P.947), (Leucine, L.992), (Glutamic acid, E.993), and (alanine, A.995) bind K^+ . The second model prediction showed (Glutamic acid, E. 430 and E.483), and (Histidine, H.432) can bind Mg^{2+} . Many GH2 β -galactosidases, that require Mg^{2+} have been reported, in contrast a GH 35 β -galactosidase that was inhibited in the presence of Mg^{2+} at concentrations of 2 mM has also been reported (Guo *et al.*, 2018). Divalent cations could be required for either stability of β -galactosidase enzymes or for catalysis.

Having high activity in the presence of Ca^{2+} , could be a desirable property of β -galactosidase enzymes for dairy industry application to hydrolyse lactose, as cow milk comprises high concentrations of Ca^{2+} (107-133 mg per 100 g) the enzyme benefit of using the Ca^{2+} ions reduces the cost of addition of exogenous cofactors in the bioreactors.

The biochemical properties of BAD_1582 have been examined using *o*NPG as a substrate, however the hydrolytic activity and transgalactosylation activity of BAD_1582 were further characterised using lactose as a substrate. BAD_1582, as a member of the GH2 families, displayed an efficient catalytic activity towards lactose and synthetic substrate *o*NPG. Besides, the substrate specificity of the enzyme also revealed that it has a very narrow range of substrates, primarily targeting β -inked galactose residues including *o*NPG and *p*NPG. However, the enzyme has barely any activity on other tested chromogenic substrates with α and β -linkages, and no activity was detected on asialofetuin which constitute one Gal (β 1–4) Mannose and two Gal (β 1–2) Mannose. In contrast, the β -gal II enzyme from *B. adolescentis* was able to hydrolyse Gal (β 1–4)Gal and Gal (β 1–3) Gal into Gal, but not act on Gal(β 1–6)Gal (Hinz *et al.*, 2004). Three GH2 β -galactosidases from *B. bifidum* (BIF1, BIF2 and BIF3) and single GH 42 β -

galactosidases from *B. infantis* (INF1) also showed a narrow range substrate specificity where all demonstrate hydrolysis of *o*NPG but exhibited less than 10% activity to other chromogenic substrates such as *o*NP- β -D-glucose, *o*NP- β -D-xylose, *o*NP- β -D-fucose, and *o*NP- β -D-galactose-6-phosphate (Moller *et al.*, 2001).

BAD_1582 has a very low K_m for lactose (3.8 ± 2.0) as compared to other GH2 β -galactosidase from *Bifidobacterium* species such as Bgal I and II of *B. breve* which had 15.3 ± 3.2 and 7.5 ± 0.9 (mM) respectively (Arreola *et al.*, 2014c). In addition, the novel β -galactosidase, *galA* obtained from human isolate *B. breve* B24 has K_m of 95.58 mM for lactose (Yi *et al.*, 2011). The low K_m for BAD_1582 indicates that this enzyme has high affinity for lactose and could hydrolyse lactose with greater efficiency than the other enzymes which have high K_m . This is desirable feature in terms of dairy industry application. In addition, BAD_1582 demonstrated less product inhibition with galactose and showed high tolerance to glucose. Kinetically the K_i for galactose was 115 mM which is higher than commercially available β -galactosidase from *A. oryzae* with K_i of 25 mM, thus BAD_1582 could be suitable for treatment of milk and produce lactose-free dairy products with less product inhibition.

BAD_1582 immobilised using calcium alginate and gelatin blend crosslinked with glutaraldehyde can be reused without significant loss of activity for 12 rounds and showed compression strength. Similarly, these procedures have been implicated to minimise leakage of enzymes, which result from porosity of calcium alginate immobilisation. The immobilised enzyme improved the temperature and pH optima of the enzyme. The immobilised BAD_1582 has a temperature optimum at 40°C. The immobilised BAD_1582 was active at all tested pHs, thus immobilisation enhanced its pH resistance both at acidic and alkaline pH. Thus, this immobilisation procedure, being simple and non-toxic, could be applied for industrial scale production of lactose free dairy products and GOS manufacture.

Although immobilisation is a way of improving the properties of enzymes as it indicated in Chapter 5, BAD_1582 exhibited better pH stability than the native enzyme. The properties of this enzyme could further be improved using protein engineering. Some of desirable properties include enhancing the catalytic efficiency of the enzyme, as high

efficiency could be required in industrial setting. Sequence based rational enzyme design provided a platform to modify the properties of enzyme, for instance a mutation of Cys176Tyr of haloalkane dehalogenase (DhaA) which act on 1,2,3-Trichloropropane (TCP) modified the protein access and export routes without directly affecting the active site and enhanced three fold conversion of substrate compared to the native enzyme (Banáš *et al.*, 2006). The catalytic efficiency of BAD_1582 could be improved with directed evolution methodologies. In addition, since natural milk contains different enzymes and at least some resident bacteria, that produce proteases which can act on this enzyme, a mutation which can provide protease resistance could be a desirable feature for dairy industry application of BAD_1582.

Currently the β -Galactosidases that are used in the dairy industry for GOS synthesis come from the bacterium *Bacillus circulans*, the yeast *Kluveromyces lactis* and the fungus *Aspergillus oryzae*. These enzymes are selected because of their high hydrolysis and high transgalactosylation activity that produce different ranges of product with a variety of degree of polymerisation (DP) and linkages (Neri *et al.*, 2009; Warmerdam *et al.*, 2014; Zhou and Dong Chen 2001).

The purified *A. oryzae* β -Galactosidase had an optimum pH of 5.0 and an optimum temperature of 50°C. The pH stability of this enzyme range between 3.5 and 8.0, at room temperature for overnight. The K_m of this enzyme was 0.77 mM and 50 mM for *o*NPG and lactose respectively. The V_{max} of this enzyme were 55.6 μ g/min/mg and 2.4 μ g/min/mg of protein for *o*NPG and for lactose respectively. The activity of the enzyme was not affected in the presence of metal ions (in the range 0.01-1 mM) and sulfhydryl reagent (0.01–0.1 mM of p-chloromercuribenzoate). However, the activity of this enzyme was inhibited competitively by galactose but it was not inhibited by glucose (Park *et al.*, 1979). Transgalactosylation activity of this enzyme (Enzeco® fungal lactase) was favored at 40% (w/w) initial lactose at pH 4.5 and 40°C. This enzyme was inhibited by lactose with a K_i of 25 mM and had a K_m of 93.8 mM for lactose, which is a disadvantage (Vera *et al.*, 2011).

The purified *K. lactis* β -galactosidase, had an optimum pH of 7. This enzyme had an optimal activity at 37°C and 40°C for *o*NPG and lactose respectively. The K_m and V_{max}

values of the purified enzyme for *o*NPG were 1.5 mM and 560 $\mu\text{mol}/\text{min}/\text{mg}$, and for lactose 20 mM and 570 $\mu\text{mol}/\text{min}/\text{mg}$. The purified enzyme required Mn^{2+} for activity and was inactivated irreversibly by imidazole above 50 mM respectively (Kim *et al.*, 2003). The GOS yield using the two soluble enzymes *K. lactis* (Lactozym 3000 LHPG and Maxilact LGX5000) in the presence of 400 g/l lactose at pH 6.8 and 40°C were, 160 g/l and 154 g/l respectively (Rodriguez-Colinas *et al.*, 2011).

The commercial β -galactosidase of *B. circulans* (Biolacta N5) has four isoforms (β -Gal-A, B, C and D) with molecular weights of 189, 154, 135, and 92 kDa respectively (Warmerdam *et al.*, 2014). β -Gal-A has a low transgalactosylation activity while β -Gal-B, β -Gal-c and β -Gal-D produce high amounts of tri- and tetra saccharides. A commercial product comprising all the isoforms of *B. circulans* β -galactosidases (Biocon NTL 3000) had an optimal pH 6.0 and 60°C, this enzyme showed a lower K_m values (~ 15 mM) to lactose compared to *A. oryzae* (Enzeco® fungal lactase) (122 mM) and *K. lactis* (Lactozym) β -galactosidases (90 mM) owing to the advantages of having the four isoforms together. However, three of the most widely used β -galactosidases favor transgalactosylation over hydrolysis when the initial lactose concentration was 40% w/w lactose (Guerrero *et al.*, 2015a).

BAD_1582 demonstrated some desirable properties such as having a low K_m (3.8 mM) for lactose and having a high K_i (115 mM) for galactose make it a very good candidate for commercial application.

6.3. Transgalactosylation activity of BAD_1582

β -galactosidases have the ability to undertake transgalactosylation reactions in which they transfer Gal moiety to another acceptor sugar molecule. β -galactosidases have received much attention for their transgalactosylation activity for the production of prebiotics to modulate the gut microbiota (Liu *et al.*, 2017a; Perdijk *et al.*, 2019; Van den Broek *et al.*, 2005). The type of linkages and the DP of oligosaccharides produced by β -galactosidases varies depending on the source of parental strain and reaction kinetics (Iqbal *et al.*, 2010). GOS is considered as a novel food and fall within the novel food regulation in EFSA and Food Safety Authority Ireland (FSAI). Thus GOS have been on the market in EU mainly

infant formula and food supplements. GOS, having a resistance to digestive enzymes has allowed their availability in the distal GIT for microbial fermentation (Sako *et al.*, 1999). Several studies have indicated that the supplementation of infant formula with GOS improved stool consistency and frequency, increased faecal *Bifidobacterium* and *Lactobacilli* and the GOS fermentation by these probiotics strains are accompanied by acetate and lactate production which results in a drop of the faecal pH, similar to that of human milk fed infants (Ben *et al.*, 2008). However, the effect of GOS on the selective growth of *Bifidobacterium* and *Lactobacilli* is strain dependent. Recent studies showed that variation of different strains of *Bifidobacterium* and *Lactobacilli* in utilisation of different commercial GOS, for instance *B. longum* BB536 (BL_BB536) and *L. fermentum* CECT5716 (LF_5716) were capable of utilizing Purimune GOS (GOS-PP) (GTC Nutrition, Golden, Colorado, USA) but not Vivinal GOS (GOS-VP) (Friesland Campina Domo, Amersfoort, Netherlands) which implicates the type of linkages in some GOS preparation could not be fermented by some probiotics (Thongaram *et al.*, 2017b). Growth studies on commercial GOS (Purimune GOS, GTC Nutrition, Golden Colorado, USA) conducted on twenty two strains of *B. longum* subsp. *infantis*, with the exception of one strain (SC97) all strains showed from moderate to vigorous growth, furthermore, the TLC analysis confirm that these strains metabolised all the monosaccharides and depletion of DP3 GOS was detected, which was in correlation with high OD 600_{nm} by the metabolising strains (Garrido *et al.*, 2013). In line with this, *B. longum* subsp. *infantis* ATCC 15697 that was grown on GOS containing media has induced two genes that encode solute binding proteins (SBPs), which are families of ABC transporters (Garrido *et al.*, 2011), presumably required for transportation of GOS to the cytosol. Recent studies on human intervention using FOS and GOS on 35 volunteer young healthy human subjects (10 male and 25 female) revealed that supplementation of these prebiotics on the first 14 days didn't change BMI, or body weight gain or loss, however composition of the gut microbiota was significantly altered after the intervention with these two prebiotics. FOS intervention was accompanied by an increase in *Bifidobacterium* and a reduction in *Phascolarctobacterium*, *Enterobacter*, *Turicibacter*, *Coprococcus* and *Salmonella*; similarly after GOS intervention the abundance of *Bifidobacterium* was high and lower level of *Ruminococcus*, *Dehalobacterium*, *Synergistes* and *Holdemania* (Liu *et al.*,

2017a). These data revealed an increase *Bifidobacterium* was accompanied by a decrease in butyrate producing bacteria and decrease in faecal SCFA which refutes the other reports which revealed prebiotic supplementations increase SCFA production (Perdijk *et al.*, 2019). For future research, the GOS synthesised using BAD_1582 could be further purified and tested *in vitro* for its prebiotic properties. The growth of different probiotics (such as *Bifidobacterium* and *Lactobacilli* species) will be conducted with the strain and species specific growth medium and condition (can be obtained from the source of the strain, such as DSMZ) supplemented with GOS (synthesised using BAD_1582) as a sole carbon source. Then OD 600_{nm} of the strains would be compared to a control (medium supplemented with glucose). Furthermore, comparative studies could be carried out on animal models such as mice/rats fed with purified GOS and placebo fed mice/rats could be characterised by sequencing of faecal DNA samples, to determine variation in microbial composition resulting from GOS supplemented diet. β -galactosidases that have retaining capacity have received attention for their transgalactosylation activity and ability to produce prebiotics which have shown *in vitro* and *in vivo* experimental evidence for modulation of the gut microbiota (Liu *et al.*, 2017a; Perdijk *et al.*, 2019; Van den Broek *et al.*, 2005). However, not all β -galactosidases from *Bifidobacterium* are able to carry out transgalactosylation, for instance β -gal II (GH 42) of *B. adolescentis* ATCC 15703 had a very low transgalactosylation activity (Hinz *et al.*, 2004). The transgalactosylation activity of BAD_1582 has been determined using 8% (234 mM) lactose as a substrate. This enzyme readily undertook transgalactosylation activity at low concentration of lactose (8%) unlike most known β -galactosidases that require a substantial quantity of lactose ($\geq 1M$) for the transgalactosylation reaction. Most commercially available β -galactosidases from the yeast *K. lactis* and the fungi *A. oryzae* undertook transgalactosylation at 40% lactose (Vera *et al.*, 2011; Yin *et al.*, 2017). The lactose content of natural bovine milk is 5% (W/V) and synthesis of GOS during natural milk treatment using commercial β -galactosidases was negligible (Rodriguez-Colinas *et al.*, 2014), presumably because transgalactosylation is favoured by these enzymes when lactose content is as high as 15-40% (W/V). Complete solubility of lactose at higher concentrations is not possible at room temperature (20°C) and hence heating to 80-90°C is required to enhance solubility of lactose ($\geq 20\%$ W/V) in buffer (Hunziker and Nissen

1926). Moreover, as BAD_1582 β -galactosidase require less initial substrate to carryout transgalactosylation, its application will be eco-friendly in terms of saving energy and cost in production of GOS, as compared to the commercial enzymes.

Kinetically the transgalactosylation activity precedes the hydrolytic activity of these enzyme, when lactose was used as a substrate. This can be translated to the role of this enzyme to the adaptation of *B. adolescentis* in the competitive human GIT. Dairy products including milk, yoghurt and whey constitute substantial dietary component in the western type diet, hence as an adult the undigested lactose would reach the human GIT unaltered where it can serve as carbon source for the commensal microbiota. The ability of *B. adolescentis* to immediately convert lactose into GOS rather than hydrolysing it, could be a metabolic adaptation to conserve carbon for particular cross feeding of *Bifidobacterium* species in the GIT, and at the same time outcompete other pathogens, which lack the repertoire of enzymes to breakdown complex oligosaccharides such as GOS. TLC and HPLC analysis of the transgalactosylation products of BAD_1582 revealed that disaccharides with different linkages and GOS with different DP and linkages were produced, nevertheless the type of linkages were not characterised in this study. By comparing the elution time of the corresponding standard sugars and standard GOS, the BAD_1582 transgalactosylation products, were characterised. It is presumed that, BAD_1582 produces a mixture of GOS including disaccharides, tri, and tetra and pentasaccharides. The synthesis of the disaccharide (presumably, allolactose) at first could possibly indicate, the bacteria uses this product for regulation of gene expression in lactose metabolism, in similar fashion to *lac* operon of *E. coli* (Wheatley *et al.*, 2013). Several β -galactosidases from *Bifidobacterium* origin have been indicated to carry out transgalactosylation activity, and produce GOS, for instance the human isolate *B. breve* undertook transgalactosylation reaction in the presence of lactose at conc. of (100-1000 mM), maximum GOS production obtained at 1M of lactose (Yi *et al.*, 2011). Three GH2 β -galactosidases from *B. bifidum* (BIF1, BIF2 and BIF3) and single GH 42 β -galactosidases from *B. infantis* (INF1) have demonstrated transgalactosylation activity in the presence of 400 mM lactose and produce different GOS mixtures, with higher activity detected by BIF1 and BIF2 (Moller *et al.*, 2001). In addition the four β -galactosidases BbgI, BbgII, BbgIII and BbgIV isolated from *B. bifidum* NCIMB41171, showed

transgalactosylation activity using lactose (30-40%) as a substrate, higher activity detected using BbgI and BbgIV with 42% to 47% GOS yields respectively (Goulas *et al.*, 2009).

The transgalactosylation activity of BAD_1582 could be further tested for synthesis of hydroquinone- galactosides (HQ-Gal). HQ-Gal have received considerable attention in the cosmetic industry as a skin whitening agent. BAD_1582 could be tested for transgalactosylation based synthesis of HQ-Gal by using lactose as Gal donor and hydroquinone as an acceptor. If the transgalactosylation reaction produces HQ-Gal, using TLC transgalactosylation products would be first analysed, which can be further confirmed using HPLC. Although, BAD_1582 looks to undertake transgalactosylation reaction only in the presence of lactose as donor and acceptor, however it was not tested on hydroquinone which is a completely different substrate. The HQ-Gal produced could first be purified and its cell viability, inhibitory effect on tyrosinase enzyme and its inhibitory effect on synthesis of melanin on melanoma cells could be studied *in vitro*. Tyrosinase is the enzyme that catalyses the rate limiting step in melanin synthesis, thus most skin whitening industries look for agents that inhibit this enzyme (Pillaiyar *et al.*, 2017). The most widely used skin whitening agents which target tyrosinase inhibitors include hydroquinone (HQ), arbutin, kojic acid, azelaic acid, L-ascorbic acid, ellagic acid, tranexamic acid each with some known side effects (Pillaiyar *et al.*, 2017). Although regulatory bodies such as FDA has placed a concern and has proposed to ban the use of some of these skin whitening agent, particularly HQ containing products, because of their unprecedented side effect on the customers (Levitt 2007). Nevertheless, HQ containing skin whitening agents which exceed the limit of FDA regulation have been on the market (<https://www.fda.gov/cosmetics/cosmetics-laws-regulations/prohibited-restricted-ingredients-cosmetics>) mostly in sub-Saharan west Africa and ethnic shops in Canada (Gbetoh and Amyot 2016). To be on the market all FDA limits should be respected, and all the ethical and efficacy issues should be fulfilled for topical usage by humans.

Although the TLC and HPLC analysis indicated that BAD_1582 synthesises GOS with a range of DPs, further characterisation to determine the type of linkages, DP and structure of the GOS should be determined using nuclear magnetic resonance (NMR) studies.

Despite this, we have studied a range of substrates as galactosyl donor and acceptor to synthesise different lactose based oligosaccharides (such as using *p*NPG as Gal donor; Glc, Gal, Lac, GlcNAC, arabinose and fructose as Gal acceptor or in the second scheme, Lac as Gal donor and acceptor and in the third scheme Lac as Gal donor, Fucose and sialic acid acceptor). Only the second scheme where Lac was used both as a donor and acceptor of galactosyl moiety was successful in production of GOS. Thus, it's highly recommended to optimise variety of monosaccharides and disaccharides to enhance synthesis of lactose derived oligosaccharides. If synthesis of oligosaccharides is successful, High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAE-PAD) analysis should be carried out to determine the monosaccharide composition of the GOS; further characterisation using NMR should be carried out to determine the linkage and structure of the individual oligosaccharide components. Furthermore, the prebiotic properties of these oligosaccharides could be examined using *in vitro* growth studies by culturing different probiotics strains (*Bifidobacterium* and *Lactobacillus* spp.) in the presence of these oligosaccharides as sole carbon source. Moreover HPAE-PAD analysis should be carried out on post fermentation broths to determine utilisation of GOS by the probiotic strains.

Commercially available GOS are also synthesised through enzymatic transgalactosylation of lactose. For instance the novasep passion and smart processes (<https://www.novasep.com/home/about-novasep.html>) produce GOS after enzymatic conversion of lactose, which result in mixtures of oligosaccharides with different DP and monosaccharides. This is followed by initial purification using Applexion® ion exchange technology, to obtain pre-purified GOS. Subsequently, further purification using Novasep's Applexion® SSMB (Sequential Simulated Moving Bed) and fine tuning with evaporation and spray drying, premium grade GOS-97 is produced. Therefore, for commercial scale production of GOS using BAD_1582 for different applications such as infant formula production or as prebiotic supplement, the purification techniques to increase the purity of GOS need optimisation. To enhance the transgalactosylation activity of BAD_1582 and increase yield of GOS, perhaps it's important to optimise the GOS yield by increasing the initial concentrations of lactose used in the reactions to higher concentrations 40% (W/V), although preliminary work was carried out but insolubility of

lactose at room temperature was observed and at temperatures higher than 37°C where lactose solubility enhanced the enzyme was unstable.

The β -galactosidase BAD_1582 has a transgalactosylation activity using lactose as a substrate. In the first 4h of reaction kinetics, this enzyme favours transgalactosylation activity over hydrolysis of lactose, presumably an evolutionary adaptation by this bacteria to contain carbon sources for future use in the competitive gut environment.

Protein engineering methods such as sequence based rational redesign and machine learning algorithms (Lutz and Iamurri 2018), could be implemented to enhance the substrates specificity of BAD_1582. The transgalactosylation activity of this enzyme is an important property to synthesise various oligosaccharides that exhibit prebiotic properties. Improving the substrate specificity of this enzyme in such a way that it can bind various acceptor monosaccharides (such fucose/sialic acid as an acceptor) could enable to synthesise HMOs and various lactose-based prebiotics such as lactulose from using Lactose as Gal donor. Studies have shown that enzyme engineering using random or site directed mutations on glycosidases has significantly enhanced the transglycosylation activity over hydrolysis activity (Arab-Jaziri *et al.*, 2015; Feng *et al.*, 2005). Moreover enzyme engineering to enhance the transgalactosylation activity of BAD_1582 over hydrolysis could be possible.

6.4. Biological role of BAD_1582 to *B. adolescentis*

Bifidobacterium species are among the dominant members of the human gut microbiome of breast fed infants. A key to their success is they have the ability to utilise wide arrays of carbohydrates. Among the six β -galactosidase enzymes encoded by the genome of *B. adolescentis*, BAD_1582 is one of them. This enzyme is ~117 kDa and is predicted to be intracellular. This enzyme had a transgalactosylation activity and produced disaccharides and GOS with different DP. It can efficiently hydrolyse lactose and GOS (Fig.17e). Thus, the role of this enzyme for the *B. adolescentis* is to utilise lactose and lactose derived GOS in the human gut environment. In addition, the transgalactosylation activity of BAD_1582 was favoured over hydrolysis even at a low concentration of lactose (~8%), while most β -galactosidase favour transgalactosylation when initial lactose concentration is above 20%. These suggest that *B. adolescentis* presumably uses this enzyme to convert lactose

to GOS based oligosaccharides and perhaps exports extracellularly through wide array of galactoside symporters to promote cross-feeding of other *Bifidobacterium* species. The resource sharing behaviour between members of the *Bifidobacterium* has been the key to their success in colonisation and adaptation in the human gut (Turrone *et al.*, 2018). To better understand the biological role of BAD_1582, future research could be carried out on the hydrolytic activity of this enzyme with a wide array of substrates such as HMOs, mucin O-glycans and plant-derived polysaccharides.

6.5. Conclusions

These PhD project aimed to identify and characterise glycoside hydrolase enzymes through function-based screening of a human faecal microbiome metagenomic library. Emphasis was given to the identification of GH enzymes which could have activity in host mucin glycan degradation and metabolism of complex dietary polysaccharides. The human gut environment seems neglected in terms of identification of industrially relevant enzymes, however with lots of efforts it's possible to identify and characterise industrially relevant enzymes from the genetic resources of the human microbiota. This research project developed methodology for screening glycoside hydrolases using MacConkey base agar supplemented with various glycans including HMOs. The human faecal microbiome metagenome library that was used for this study didn't harbour metagenome clones which express active sialidases or α -fucosidases, presumably due to the absence of representative strains which encode for such enzymes. The function-based screening of the human faecal microbiome enabled the identification of new GH 2 β -galactosidase (BAD_1582) from *B. adolescentis*, which was not previously characterised. The biochemical characterisation of BAD_1582 showed that, this enzyme has wide pH stability, a temperature stability up to 40°C. It has divalent cation requirement, for Mg^{2+} and Ca^{2+} , especially having high activity with Ca^{2+} makes it an ideal candidate for dairy industry application as it benefits from the availability of high Ca^{2+} from milk and other dairy products. Moreover, BAD_1582 acted up on narrow ranges of substrates, such as *o*NPG, *p*NPG, lactose and GOS (synthesised by this enzyme) but it was not active on other tested chromogenic substrates and didn't act on asialofetuin, a glycoprotein which had terminally linked galactose residues. Thus, it suggests this enzyme particularly act on

lactose and lactose-based oligosaccharides (GOS). This leads to the conclusion that, the natural substrate for BAD_1582 could be lactose. However, its activity towards various other β -galactosides in the human GIT such as host mucin glycans were not studied so this would need further investigation. The β -galactosidase activity of BAD_1582 towards the chromogenic synthetic substrate *o*NPG was inhibited by Gal and lactose but was unaltered by Glc. The inhibition of this enzyme by Gal but not by Glc is in line with many β -galactosidases, as hydrolytic products cause feedback inhibition which is one of the many ways of regulation of enzymatic reactions in biochemical pathway. BAD_1582 has lower K_m for lactose, which is a desirable feature as it can confirm the high affinity of this enzyme to its natural substrate. BAD_1582 immobilised on calcium alginate and gelatin beads complemented with cross-linking using glutaraldehyde shifted the optimum temperature of the enzyme to 40°C but the maximum stability of both the immobilised and the free enzyme were at 37°C, with slightly higher residual activity obtained by the immobilised enzyme at 45°C. Immobilisation of BAD_1582 abolished the substrate inhibition presented by *o*NPG (≥ 2.5 mM) on the free enzyme. The immobilised enzyme reduced the K_m compared to the free enzyme. The immobilised BAD_1582 had high hydrolytic activity towards lactose. In summary the properties of free and the immobilised BAD_1582, make it an ideal candidate for dairy industry application and for lactose based prebiotic manufacture, particularly GOS.

The enzyme not only hydrolysed lactose but also displayed transgalactosylation activity at lower concentration of lactose (8%), while most commercially available β -galactosidases require presence of high concentration of lactose (up to ~40%). Moreover, the kinetics of hydrolytic and transgalactosylation activity, showed that the later precedes the former, which presumably is an evolutionary conserved trait by the *B. adolescentis* to scavenge and keep carbon source readily accessible form to these species presumably enhancing cross-feeding between different *Bifidobacterium* species while competitively excluding other bacteria particularly pathogens from thriving in the human GIT. TLC and HPLC analysis of the transgalactosylation products revealed that, disaccharides (presumably allolactose) were formed first, followed by GOS, the DP of which increased with time of incubation presumably to produce tri, tetra and pentasaccharides.

Overall in this thesis, a functional metagenomics study of human faecal gut microbiome metagenome library was conducted to identify GH enzymes such as α -fucosidases, sialidases and β -galactosidases. MacConkey base agar supplemented with different HMOs and lactose were used for screening the metagenome library for the three enzymes. Although, α -fucosidases and sialidases were not detected, a previously uncharacterised β -galactosidase from the *B. adolescentis* BAD_1582 were identified. The *B. adolescentis* β -galactosidase BAD_1582, has been identified and its properties were characterised for lactose hydrolysis and prebiotic manufacture. Given that the parent strain is a probiotic strain the β -galactosidase enzyme (BAD_1582) could be used for food industry application as GRAS status has been granted for *Bifidobacterium* species. Based on the findings and promising results obtained using this enzyme, we recommend the following additional works in future. The enzyme could be further studied for industrial application for creation of lactose free dairy products and manufacture of GOS.

References

- Abrahamsson TR, Jakobsson HE, Andersson AF, Bjorksten B, Engstrand L, Jenmalm MC (2014) Low gut microbiota diversity in early infancy precedes asthma at school age. *Clinical and Experimental Allergy* 44(6):842-50
doi:10.1111/cea.12253
- Adam AC, Rubio-Teixeira M, Polaina J (2004) Lactose: the milk sugar from a biotechnological perspective. *Critical reviews in food science and nutrition* 44(7-8):553-7
- Akiyama T, Kimura K, Hatano H (2015) Diverse galactooligosaccharides consumption by *Bifidobacteria*: implications of β -galactosidase--LacS operon. *Bioscience, biotechnology, and biochemistry* 79(4):664-72
doi:10.1080/09168451.2014.987204
- Albenberg LG, Wu GD (2014) Diet and the intestinal microbiome: associations, functions, and implications for health and disease. *Gastroenterology* 146(6):1564-72 doi:10.1053/j.gastro.2014.01.058
- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological reviews* 59(1):143-69
- Ambrogi V, Bottacini F, O'Sullivan J, O'Connell Motherway M, Linqu C, Schoemaker B, Schoterman M, van Sinderen D (2019) Characterization of GH2 and GH42 β -galactosidases derived from bifidobacterial infant isolates. *AMB Express* 9(1):9
doi:10.1186/s13568-019-0735-3
- Andrianifahanana M, Moniaux N, Schmied BM, Ringel J, Friess H, Hollingsworth MA, Buchler MW, Aubert JP, Batra SK (2001) Mucin (MUC) gene expression in human pancreatic adenocarcinoma and chronic pancreatitis: a potential role of MUC4 as a tumor marker of diagnostic significance. *Clinical Cancer Research* 7(12):4033-40
- Arab-Jaziri F, Bissaro B, Tellier C, Dion M, Faure R, O'Donohue MJ (2015) Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α -L-arabinofuranosidase. *Carbohydr Res* 401:64-72
doi:10.1016/j.carres.2014.10.029
- Argüeso P, Spurr-Michaud S, Russo CL, Tisdale A, Gipson IK (2003) MUC16 Mucin is expressed by the human ocular surface epithelia and carries the H185 carbohydrate epitope. *Invest Ophth Vis Sci* 44(6):2487-2495
doi:10.1167/iovs.02-0862
- Arias EB, Verhage HG, Jaffe RC (1994) Complementary deoxyribonucleic acid cloning and molecular characterization of an estrogen-dependent human oviductal glycoprotein. *Biol Reprod* 51(4):685-94 doi:10.1095/biolreprod51.4.685

- Arike L, Hansson GC (2016) The densely O-glycosylated MUC2 mucin protects the intestine and provides food for the commensal bacteria. *Journal of Molecular Biology* 428(16):3221-9 doi:10.1016/j.jmb.2016.02.010
- Arreola SL, Intanon M, Suljic J, Kittl R, Pham NH, Kosma P, Haltrich D, Nguyen T-H (2014a) Two β -galactosidases from the human isolate *Bifidobacterium breve* DSM 20213: molecular cloning and expression, biochemical characterization and synthesis of galacto-oligosaccharides. *PloS one* 9(8):e104056-e104056 doi:10.1371/journal.pone.0104056
- Arreola SL, Intanon M, Suljic J, Kittl R, Pham NH, Kosma P, Haltrich D, Nguyen T-H (2014b) Two β -Galactosidases from the Human Isolate *Bifidobacterium breve* DSM 20213: Molecular Cloning and Expression, Biochemical Characterization and Synthesis of Galacto-Oligosaccharides. *PloS one* 9(8):e104056 doi:10.1371/journal.pone.0104056
- Arreola SL, Intanon M, Suljic J, Kittl R, Pham NH, Kosma P, Haltrich D, Nguyen T-H (2014c) Two β -Galactosidases from the human isolate *Bifidobacterium breve* DSM 20213: molecular cloning and expression, biochemical characterization and synthesis of galactooligosaccharides. *PloS one* 9(8):1-13 doi:10.1371/journal.pone.0104056
- Arul GS, Moorghen M, Myerscough N, Alderson DA, Spicer RD, Corfield AP (2000) Mucin gene expression in Barrett's oesophagus: an in situ hybridisation and immunohistochemical study. *Gut* 47(6):753-61 doi:10.1136/gut.47.6.753
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Meta HITC, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariáz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Kristiansen K, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P (2011) Enterotypes of the human gut microbiome. *Nature* 473:174 doi:10.1038/nature09944 <https://www.nature.com/articles/nature09944#supplementary-information>
- Atuma C, Strugala V, Allen A, Holm L (2001) The adherent gastrointestinal mucus gel layer: thickness and physical state *in vivo*. *Am J Physiol-Gastr L* 280(5):G922-G929
- Audie JP, Janin A, Porchet N, Copin MC, Gosselin B, Aubert JP (1993) Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained

by *in situ* hybridization. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society 41(10):1479-85
doi:10.1177/41.10.8245407

- Audie JP, Tetaert D, Pigny P, Buisine MP, Janin A, Aubert JP, Porchet N, Boersma A (1995) Mucin gene expression in the human endocervix. Human reproduction (Oxford, England) 10(1):98-102 doi:10.1093/humrep/10.1.98
- Aureli P, Capurso L, Castellazzi AM, Clerici M, Giovannini M, Morelli L, Poli A, Pregliasco F, Salvini F, Zuccotti GV (2011) Probiotics and health: an evidence-based review. Pharmacological research 63(5):366-76
doi:10.1016/j.phrs.2011.02.006
- Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI (2004) The gut microbiota as an environmental factor that regulates fat storage. Proceedings of National Academy of Sciences of the U S A 101(44):15718-23 doi:10.1073/pnas.0407076101
- Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) Host-bacterial mutualism in the human intestine. Science 307(5717):1915-20
doi:10.1126/science.1104816
- Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J (2015) Dynamics and stabilization of the human gut microbiome during the first year of life. Cell host & microbe 17(5):690-703 doi:10.1016/j.chom.2015.04.004
- Balague C, Gambus G, Carrato C, Porchet N, Aubert JP, Kim YS, Real FX (1994) Altered expression of MUC2, MUC4, and MUC5 mucin genes in pancreas tissues and cancer cell lines. Gastroenterology 106(4):1054-61
doi:10.1016/0016-5085(94)90767-6
- Balcão VM, Vila MMDC (2015) Structural and functional stabilization of protein entities: state-of-the-art. Advanced Drug Delivery Reviews 93:25-41
doi:<https://doi.org/10.1016/j.addr.2014.10.005>
- Baldasso C, Barros T, Tessaro I (2011) Concentration and purification of whey proteins by ultrafiltration. Desalination 278(1-3):381-386
- Ballard O, Morrow AL (2013) Human milk composition: nutrients and bioactive factors. Pediatr Clin North Am 60(1):49-74 doi:10.1016/j.pcl.2012.10.002
- Banáš P, Otyepka M, Jeřábek P, Petřek M, Damborský J (2006) Mechanism of enhanced conversion of 1, 2, 3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling. Journal of computer-aided molecular design 20(6):375-383

- Bansil R, Turner BS (2006) Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid & Interface Science* 11(2):164-170 doi:<https://doi.org/10.1016/j.cocis.2005.11.001>
- Battson ML, Lee DM, Jarrell DK, Hou S, Ecton KE, Weir TL, Gentile CL (2018) Suppression of gut dysbiosis reverses western diet-induced vascular dysfunction. *American Journal of Physiology: Endocrinology & Metabolism* 314(5):E468-e477 doi:10.1152/ajpendo.00187.2017
- Ben XM, Li J, Feng ZT, Shi SY, Lu YD, Chen R, Zhou XY (2008) Low level of galacto-oligosaccharide in infant formula stimulates growth of intestinal Bifidobacteria and Lactobacilli. *World journal of gastroenterology* 14(42):6564-8 doi:10.3748/wjg.14.6564
- Benjamin JL, Hedin CRH, Koutsoumpas A, Ng SC, McCarthy NE, Hart AL, Kamm MA, Sanderson JD, Knight SC, Forbes A, Stagg AJ, Whelan K, Lindsay JO (2011) Randomised, double-blind, placebo-controlled trial of fructo-oligosaccharides in active Crohn's disease. *Gut* 60(7):923 doi:10.1136/gut.2010.232025
- Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA (2012) Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22(6):736-756 doi:10.1093/glycob/cwr182
- Bermúdez-Humarán LG, Langella P (2018) Live bacterial biotherapeutics in the clinic. *Nature biotechnology* 36(9):816-818 doi:10.1038/nbt.4248
- Bhushan B, Pal A, Jain V (2015) Improved enzyme catalytic characteristics upon glutaraldehyde cross-linking of alginate entrapped xylanase isolated from *Aspergillus flavus* MTCC 9390. *Enzyme research* 2015:210784-210784 doi:10.1155/2015/210784
- Bibiloni R, Fedorak RN, Tannock GW, Madsen KL, Gionchetti P, Campieri M, De Simone C, Sartor RB (2005) VSL#3 probiotic-mixture induces remission in patients with active ulcerative colitis. *American Journal of Gastroenterology* 100(7):1539-46 doi:10.1111/j.1572-0241.2005.41794.x
- Bisswanger H (2014) Enzyme assays. *Perspectives in Science* 1(1):41-55 doi:<https://doi.org/10.1016/j.pisc.2014.02.005>
- Blandino A, Macías M, Cantero D (2001) Immobilization of glucose oxidase within calcium alginate gel capsules. *Process Biochemistry* 36(7):601-606 doi:[https://doi.org/10.1016/S0032-9592\(00\)00240-5](https://doi.org/10.1016/S0032-9592(00)00240-5)
- Bobek LA, Situ H (2003) MUC7 20-Mer: investigation of antimicrobial activity, secondary structure, and possible mechanism of antifungal action. *Antimicrobial agents and chemotherapy* 47(2):643-652 doi:10.1128/aac.47.2.643-652.2003
- Bode L (2012) Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* 22(9):1147-62 doi:10.1093/glycob/cws074

- Boël G, Letso R, Neely H, Price WN, Wong K-H, Su M, Luff JD, Valecha M, Everett JK, Acton TB, Xiao R, Montelione GT, Aalberts DP, Hunt JF (2016) Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529:358 doi:10.1038/nature16509
<https://www.nature.com/articles/nature16509#supplementary-information>
- Boel G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff J, Valecha M, Everett JK, Acton TB, Xiao R, Montelione GT, Aalberts DP, Hunt JF (2016) Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529(7586):358-363 doi:10.1038/nature16509
- Boelchandani M, D'Souza AW, Dantas G (2019) Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics* 20(6):356-370 doi:10.1038/s41576-019-0108-4
- Boronat A, Aguilar J (1981) Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: differences in induction of propanediol oxidoreductase. *Journal of bacteriology* 147(1):181-185
- Bouhnik Y, Attar A, Joly FA, Riottot M, Dyard F, Flourié B (2004) Lactulose ingestion increases faecal bifidobacterial counts: a randomised double-blind study in healthy humans. *European journal of clinical nutrition* 58(3):462-466 doi:10.1038/sj.ejcn.1601829
- Bouskra D, Brezillon C, Berard M, Werts C, Varona R, Boneca IG, Eberl G (2008) Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* 456(7221):507-10 doi:10.1038/nature07450
- Braun S, Rappoport S, Zusman R, Avnir D, Ottolenghi M (1990) Biochemically active sol-gel glasses: the trapping of enzymes. *Materials Letters* 10(1):1-5 doi:[https://doi.org/10.1016/0167-577X\(90\)90002-4](https://doi.org/10.1016/0167-577X(90)90002-4)
- Brayman M, Thathiah A, Carson DD (2004) MUC1: a multifunctional cell surface component of reproductive tissue epithelia. *Reprod Biol Endocrinol* 2:4-4 doi:10.1186/1477-7827-2-4
- Brena B, Gonzalez-Pombo P, Batista-Viera F (2013) Immobilization of enzymes: a literature survey. *Methods in molecular biology* 1051:15-31 doi:10.1007/978-1-62703-550-7_2
- Buisine MP, Desreumaux P, Debailleul V, Gambiez L, Geboes K, Ectors N, Delescaut MP, Degand P, Aubert JP, Colombel JF, Porchet N (1999) Abnormalities in mucin gene expression in Crohn's disease. *Inflamm Bowel Dis* 5(1):24-32 doi:10.1097/00054725-199902000-00004
- Buisine MP, Devisme L, Degand P, Dieu MC, Gosselin B, Copin MC, Aubert JP, Porchet N (2000) Developmental mucin gene expression in the gastroduodenal tract and accessory digestive glands. II. duodenum and liver, gallbladder, and pancreas. *J Histochem Cytochem* 48(12):1667-76 doi:10.1177/002215540004801210

- Bull MJ, Plummer NT (2014) Part 1: The Human gut microbiome in health and disease. *Integr Med (Encinitas)* 13(6):17-22
- Bunesova V, Lacroix C, Schwab C (2016) Fucosyllactose and L-fucose utilization of infant *Bifidobacterium longum* and *Bifidobacterium kashiwanohense*. *BMC microbiology* 16(1):248 doi:10.1186/s12866-016-0867-4
- Burnette WN (1981) “Western Blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* 112(2):195-203 doi:[https://doi.org/10.1016/0003-2697\(81\)90281-5](https://doi.org/10.1016/0003-2697(81)90281-5)
- Calderon D, Peña L, Suarez A, Villamil C, Ramirez-Rojas A, Anzola JM, García-Betancur JC, Cepeda ML, Uribe D, Del Portillo P, Mongui A (2019) Recovery and functional validation of hidden soil enzymes in metagenomic libraries. *MicrobiologyOpen* 8(4):e00572 doi:10.1002/mbo3.572
- Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, Neyrinck AM, Fava F, Tuohy KM, Chabo C, Waget A, Delmée E, Cousin B, Sulpice T, Chamontin B, Ferrières J, Tanti J-F, Gibson GR, Casteilla L, Delzenne NM, Alessi MC, Burcelin R (2007) Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* 56(7):1761 doi:10.2337/db06-1491
- Cantarel BL, Lombard V, Henrissat B (2012) Complex carbohydrate utilization by the healthy human microbiome. *PloS one* 7(6):e28742 doi:10.1371/journal.pone.0028742
- Cao L (2005) Carrier-bound immobilized enzymes principles, applications and design. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
- Cao L, Langen Lv, Sheldon RA (2003) Immobilised enzymes: carrier-bound or carrier-free? *Current opinion in biotechnology* 14(4):387-394 doi:10.1016/s0958-1669(03)00096-x
- Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Dore J, Henrissat B, Remaud-Simeon M, Monsan P, Potocki-Veronese G (2013a) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PloS one* 8(9):e72766 doi:10.1371/journal.pone.0072766
- Cecchini DA, Laville E, Laguerre S, Robe P, Leclerc M, Doré J, Henrissat B, Remaud-Siméon M, Monsan P, Potocki-Véronèse G (2013b) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PloS one* 8(9):e72766 doi:10.1371/journal.pone.0072766
- Chapman J, Ismail AE, Dinu CZ (2018) Industrial applications of enzymes: recent advances, techniques, and outlooks. *Catalysts* 8(6):238
- Chen Y, Zhao YH, Kalaslavadi TB, Hamati E, Nehrke K, Le AD, Ann DK, Wu R (2004) Genome-wide search and identification of a novel gel-forming mucin MUC19/Muc19 in glandular tissues. *Am J Resp Cell Mol* 30(2):155-65 doi:10.1165/rcmb.2003-0103OC

- Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD, Charles TC (2017a) Functional metagenomics reveals novel β -galactosidases not predictable from gene sequences. *PloS one* 12(3):e0172545 doi:10.1371/journal.pone.0172545
- Cheng J, Romantsov T, Engel K, Doxey AC, Rose DR, Neufeld JD, Charles TC (2017b) Functional metagenomics reveals novel β -galactosidases not predictable from gene sequences. *PloS one* 12(3):e0172545-e0172545 doi:10.1371/journal.pone.0172545
- Chinen T, Rudensky AY (2012) The effects of commensal microbiota on immune cell subsets and inflammatory responses. *Immunological Reviews* 245(1):45-55 doi:10.1111/j.1600-065X.2011.01083.x
- Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* 13(4):260-270 doi:10.1038/nrg3182
- Christy Agbavwe MS (2017) Identification and characterisation of novel glycan binding bacterial adhesins encoded by the human gut microbial metagenome. PhD thesis, National University of Ireland, Galway
- Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, Reading NC, Villablanca EJ, Wang S, Mora JR, Umesaki Y, Mathis D, Benoist C, Relman DA, Kasper DL (2012) Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* 149(7):1578-93 doi:10.1016/j.cell.2012.04.037
- Clausen MR, Mortensen PB (1997) Lactulose, disaccharides and colonic flora. Clinical consequences. *Drugs* 53(6):930-42 doi:10.2165/00003495-199753060-00003
- Clevers HC, Bevins CL (2013) Paneth Cells: maestros of the small intestinal crypts. *Annual Review of Physiology* 75(1):289-311 doi:10.1146/annurev-physiol-030212-183744
- Collins FS, Morgan M, Patrinos A (2003) The human genome project: lessons from large-scale biology. *Science* 300(5617):286 doi:10.1126/science.1084564
- Collins MD, Gibson GR (1999) Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. *American Journal of Clinical Nutrition* 69 doi:10.1093/ajcn/69.5.1052s
- Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuetz SE, Kelly RM (2016) Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by caldicellulosiruptor species. *The Journal of biological chemistry* 291(13):6732-6747 doi:10.1074/jbc.M115.707810
- Copin MC, Buisine MP, Devisme L, Leroy X, Escande F, Gosselin B, Aubert JP, Porchet N (2001) Normal respiratory mucosa, precursor lesions and lung carcinomas: differential expression of human mucin genes. *Frontiers in Bioscience* 6:D1264-75 doi:10.2741/copin

- Corfield AP (2015) Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochimica et biophysica acta* 1850(1):236-52
doi:10.1016/j.bbagen.2014.05.003
- Corfield AP, Carroll D, Myerscough N, Probert CS (2001) Mucins in the gastrointestinal tract in health and disease. *Frontiers in Bioscience* 6:D1321-57
doi:10.2741/corfield
- Corfield AP, Wagner SA, Clamp JR, Kriaris MS, Hoskins LC (1992) Mucin degradation in the human colon: production of sialidase, sialate O-acetyltransferase, N-acetylneuraminidase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infection and immunity* 60(10):3971-8
- Coulier L, Timmermans J, Bas R, Van Den Dool R, Haaksman I, Klarenbeek B, Slaghek T, Van Dongen W (2009) In-depth characterization of prebiotic galactooligosaccharides by a combination of analytical techniques. *Journal of agricultural and food chemistry* 57(18):8488-8495 doi:10.1021/jf902549e
- Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P (2005) Metagenomic gene discovery: past, present and future. *Trends in Biotechnology* 23(6):321-9 doi:10.1016/j.tibtech.2005.04.001
- Cui X, Ye L, Li J, Jin L, Wang W, Li S, Bao M, Wu S, Li L, Geng B, Zhou X, Zhang J, Cai J (2018) Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients. *Scientific reports* 8(1):635
doi:10.1038/s41598-017-18756-2
- Cupples CG, Cabrera M, Cruz C, Miller JH (1990) A set of lacZ mutations in *Escherichia coli* that allow rapid detection of specific frameshift mutations. *Genetics* 125(2):275-80
- D'Cruz OJ, Dunn TS, Pichan P, Haas GG, Jr., Sachdev GP (1996) Antigenic cross-reactivity of human tracheal mucin with human sperm and trophoblasts correlates with the expression of mucin 8 gene messenger ribonucleic acid in reproductive tract tissues *. *Fertility and Sterility* 66(2):316-326
doi:10.1016/S0015-0282(16)58460-X
- Datta S, Christena LR, Rajaram YRS (2013) Enzyme immobilization: an overview on techniques and support materials. *3 Biotech* 3(1):1-9 doi:10.1007/s13205-012-0071-7
- Davis LM, Martinez I, Walter J, Hutkins R (2010) A dose dependent impact of prebiotic galactooligosaccharides on the intestinal microbiota of healthy adults. *International journal of food microbiology* 144(2):285-92
doi:10.1016/j.ijfoodmicro.2010.10.007
- De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences* 107(33):14691 doi:10.1073/pnas.1005963107

- De Filippo C, Di Paola M, Ramazzotti M, Albanese D, Pieraccini G, Banci E, Miglietta F, Cavalieri D, Lionetti P (2017) Diet, environments, and gut microbiota. A preliminary investigation in children living in rural and urban Burkina Faso and Italy. *Frontiers in microbiology* 8:1979 doi:10.3389/fmicb.2017.01979
- De Leoz ML, Kalanetra KM, Bokulich NA, Strum JS, Underwood MA, German JB, Mills DA, Lebrilla CB (2015) Human milk glycomics and gut microbial genomics in infant feces show a correlation between human milk oligosaccharides and gut microbiota: a proof-of-concept study. *Journal of proteome research* 14(1):491-502 doi:10.1021/pr500759e
- DeGruttola AK, Low D, Mizoguchi A, Mizoguchi E (2016) Current understanding of dysbiosis in disease in human and animal models. *Inflamm Bowel Dis* 22(5):1137-50 doi:10.1097/mib.0000000000000750
- Demirel G, Özçetin G, Şahin F, Tümtürk H, Aksoy S, Hasirci N (2006) Semi-interpenetrating polymer networks (IPNs) for entrapment of glucose isomerase. *Reactive and Functional Polymers* 66(4):389-394 doi:<https://doi.org/10.1016/j.reactfunctpolym.2005.08.015>
- Deplancke B, Gaskins HR (2001) Microbial modulation of innate defense: goblet cells and the intestinal mucus layer. *American Journal of Clinical Nutrition* 73(6):1131s-1141s doi:10.1093/ajcn/73.6.1131S
- Devlin AS, Marcobal A, Dodd D, Nayfach S, Plummer N, Meyer T, Pollard KS, Sonnenburg JL, Fischbach MA (2016) Modulation of a circulating uremic solute via rational genetic manipulation of the gut microbiota. *Cell host & microbe* 20(6):709-715 doi:10.1016/j.chom.2016.10.021
- Dhanisha SS, Guruvayoorappan C, Drishya S, Abeesh P (2018) Mucins: structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. *Critical reviews in oncology/hematology* 122:98-122 doi:10.1016/j.critrevonc.2017.12.006
- Dinan TG, Cryan JF (2017) Brain-gut-microbiota axis and mental health. *Psychosomatic medicine* 79(8):920-926 doi:10.1097/psy.0000000000000519
- Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ (2012) The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Molecular cell* 48(4):612-26 doi:10.1016/j.molcel.2012.08.033
- Duboc H, Tache Y, Hofmann AF (2014) The bile acid TGR5 membrane receptor: from basic research to clinical application. *Digestive and Liver Disease* 46(4):302-12 doi:10.1016/j.dld.2013.10.021
- Dufrêne YF (2015) Sticky microbes: forces in microbial cell adhesion. *Trends in microbiology* 23(6):376-382 doi:<http://dx.doi.org/10.1016/j.tim.2015.01.011>
- Dunne C, O'Mahony L, Murphy L, Thornton G, Morrissey D, O'Halloran S, Feeney M, Flynn S, Fitzgerald G, Daly C, Kiely B, O'Sullivan GC, Shanahan F, Collins JK (2001) *In vitro* selection criteria for probiotic bacteria of human origin:

- correlation with *in vivo* findings. American Journal of Clinical Nutrition 73(2 Suppl):386s-392s doi:10.1093/ajcn/73.2.386s
- Duranti S, Milani C, Lugli GA, Mancabelli L, Turrone F, Ferrario C, Mangifesta M, Viappiani A, Sanchez B, Margolles A, van Sinderen D, Ventura M (2016) Evaluation of genetic diversity among strains of the human gut commensal *Bifidobacterium adolescentis*. Scientific reports 6:23971 doi:10.1038/srep23971
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) Diversity of the human intestinal microbial flora. Science 308(5728):1635 doi:10.1126/science.1110591
- Erich S, Kuschel B, Schwarz T, Ewert J, Bohmer N, Niehaus F, Eck J, Lutz-Wahl S, Stressler T, Fischer L (2015) Novel high-performance metagenome β -galactosidases for lactose hydrolysis in the dairy industry. J Biotechnol 210:27-37 doi:10.1016/j.jbiotec.2015.06.411
- Fadnavis NW, Sheelu G, Kumar BM, Bhalerao MU, Deshpande AA (2003) Gelatin blends with alginate: gels for lipase immobilization and purification. Biotechnology progress 19(2):557-64 doi:10.1021/bp010172f
- Fajardo JE, Fiser A (2013) Protein structure based prediction of catalytic residues. BMC Bioinformatics 14(1):63 doi:10.1186/1471-2105-14-63
- Fakruddin M, Mohammad Mazumdar R, Bin Mannan KS, Chowdhury A, Hossain MN (2013) Critical factors affecting the success of cloning, expression, and mass production of enzymes by recombinant *E. coli*. ISRN biotechnology 2013:590587 doi:10.5402/2013/590587
- Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, Aguilera M, Khanna S, Gil A, Edwards CA, Dore J (2010) Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. Journal of pediatric gastroenterology and nutrition 51(1):77-84 doi:10.1097/MPG.0b013e3181d1b11e
- FAO/WHO (2001) Expert consultation on evaluation of health and nutrition properties of probiotics in food including powder milk with live lactic acid bacteria. 85
- Feng H-Y, Drone J, Hoffmann L, Tran V, Tellier C, Rabiller C, Dion M (2005) Converting a β -glycosidase into a β -transglycosidase by directed evolution. Journal of Biological Chemistry 280(44):37088-37097
- Fernandez-Lafuente R, Armisén P, Sabuquillo P, Fernández-Lorente G, M. Guisán J (1998) Immobilization of lipases by selective adsorption on hydrophobic supports. Chemistry and physics of lipids 93(1):185-197 doi:[https://doi.org/10.1016/S0009-3084\(98\)00042-5](https://doi.org/10.1016/S0009-3084(98)00042-5)
- Ferrer-Admetlla A, Sikora M, Laayouni H, Esteve A, Roubinet F, Blancher A, Calafell F, Bertranpetit J, Casals F (2009) A natural history of FUT2 polymorphism in humans. Molecular biology and evolution 26(9):1993-2003 doi:10.1093/molbev/msp108

- Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Reviews Microbiology* 6(2):121-31 doi:10.1038/nrmicro1817
- Foster JA, Rinaman L, Cryan JF (2017) Stress & the gut-brain axis: regulation by the microbiome. *Neurobiol Stress* 7:124-136 doi:10.1016/j.ynstr.2017.03.001
- Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, Vatanen T, Hall AB, Mallick H, McIver LJ, Sauk JS, Wilson RG, Stevens BW, Scott JM, Pierce K, Deik AA, Bullock K, Imhann F, Porter JA, Zhernakova A, Fu J, Weersma RK, Wijmenga C, Clish CB, Vlamakis H, Huttenhower C, Xavier RJ (2019) Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology* 4(2):293-305 doi:10.1038/s41564-018-0306-4
- Freitas FF, Marquez LDS, Ribeiro GP, Brandão GC, Cardoso VL, Ribeiro EJ (2012) Optimization of the immobilization process of β -galactosidase by combined entrapment-cross-linking and the kinetics of lactose hydrolysis. *Brazilian Journal of Chemical Engineering* 29:15-24
- Frost G, Sleeth ML, Sahuri-Arisoylu M, Lizarbe B, Cerdan S, Brody L, Anastasovska J, Ghourab S, Hankir M, Zhang S, Carling D, Swann JR, Gibson G, Viardot A, Morrison D, Louise Thomas E, Bell JD (2014) The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nature communications* 5:3611 doi:10.1038/ncomms4611
- Fuller R (1989) Probiotics in man and animal. *Journal of Applied Bacteriology* 66:365-378
- Fung KY, Ooi CC, Lewanowitsch T, Tan S, Tan HT, Lim TK, Lin Q, Williams DB, Lockett TJ, Cosgrove LJ, Chung MC, Head RJ (2012) Identification of potential pathways involved in induction of apoptosis by butyrate and 4-benzoylbutyrate in HT29 colorectal cancer cells. *Journal of proteome research* 11(12):6019-29 doi:10.1021/pr3007107
- Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, Nakanishi Y, Uetake C, Kato K, Kato T, Takahashi M, Fukuda NN, Murakami S, Miyauchi E, Hino S, Atarashi K, Onawa S, Fujimura Y, Lockett T, Clarke JM, Topping DL, Tomita M, Hori S, Ohara O, Morita T, Koseki H, Kikuchi J, Honda K, Hase K, Ohno H (2013) Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 504(7480):446-450 doi:10.1038/nature12721
- Gabor EM, de Vries EJ, Janssen DB (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environmental microbiology* 6(9):948-58 doi:10.1111/j.1462-2920.2004.00643.x

- Gänzle MG, Haase G, Jelen P (2008) Lactose: crystallization, hydrolysis and value-added derivatives. *Int Dairy J* 18(7):685-694
doi:<https://doi.org/10.1016/j.idairyj.2008.03.003>
- Garrido D, Kim JH, German JB, Raybould HE, Mills DA (2011) Oligosaccharide binding proteins from *Bifidobacterium longum* subsp. *infantis* reveal a preference for host glycans. *PloS one* 6(3):e17315
doi:10.1371/journal.pone.0017315
- Garrido D, Ruiz-Moyano S, Jimenez-Espinoza R, Eom HJ, Block DE, Mills DA (2013) Utilization of galactooligosaccharides by *Bifidobacterium longum* subsp. *infantis* isolates. *Food microbiology* 33(2):262-70 doi:10.1016/j.fm.2012.10.003
- Garrido D, Ruiz-Moyano S, Kirmiz N, Davis JC, Totten SM, Lemay DG, Ugalde JA, German JB, Lebrilla CB, Mills DA (2016) A novel gene cluster allows preferential utilization of fucosylated milk oligosaccharides in *Bifidobacterium longum* subsp. *longum* SC596. *Scientific reports* 6:35045 doi:10.1038/srep35045
- Gbetoh MH, Amyot M (2016) Mercury, hydroquinone and clobetasol propionate in skin lightening products in West Africa and Canada. *Environmental Research* 150:403-410 doi:<https://doi.org/10.1016/j.envres.2016.06.030>
- Gebler JC, Aebersold R, Withers SG (1992) Glu-537, not Glu-461, is the nucleophile in the active site of (lac Z) β -galactosidase from *Escherichia coli*. *Journal of Biological Chemistry* 267(16):11126-30
- Gendler SJ, Spicer AP (1995) Epithelial mucin genes. *Annual Review of Physiology* 57:607-34 doi:10.1146/annurev.ph.57.030195.003135
- Gibson GR (2004) From probiotics to prebiotics and a healthy digestive system. *Journal of Food Science* 69(5):141-143
- Gibson GR, Hutkins R (2017) The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nature Reviews Gastroenterology & Hepatology* 14 doi:10.1038/nrgastro.2017.75
- Gibson GR, Probert HM, Loo J (2004) Dietary modulation of the human colonic microbiota: updating the concept of the prebiotics. *Nutrition Research Reviews* 17 doi:10.1079/nrr200479
- Gibson GR, Roberfroid MB (1995) Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *Journal of Nutrition* 125(6):1401-12 doi:10.1093/jn/125.6.1401
- Gill N, Wlodarska M, Finlay BB (2011) Roadblocks in the gut: barriers to enteric infection. *Cellular microbiology* 13(5):660-669 doi:10.1111/j.1462-5822.2011.01578.x
- Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE (2006) Metagenomic analysis of the

human distal gut microbiome. *Science* 312(5778):1355
doi:10.1126/science.1124234

- Gipson IK, Ho SB, Spurr-Michaud SJ, Tisdale AS, Zhan Q, Torlakovic E, Pudney J, Anderson DJ, Toribara NW, Hill JA, 3rd (1997) Mucin genes expressed by human female reproductive tract epithelia. *Biol Reprod* 56(4):999-1011
doi:10.1095/biolreprod56.4.999
- Gipson IK, Spurr-Michaud S, Argüeso P, Tisdale A, Ng TF, Russo CL (2003) Mucin gene expression in immortalized human corneal–limbal and conjunctival epithelial cell lines. *Invest Ophth Vis Sci* 44(6):2496-2506 doi:10.1167/iovs.02-0851
- Goto M, Hatanaka C, Goto M (2005) Immobilization of surfactant–lipase complexes and their high heat resistance in organic media. *Biochemical Engineering Journal* 24(1):91-94 doi:<https://doi.org/10.1016/j.bej.2005.01.027>
- Goulas T, Goulas A, Tzortzis G, Gibson GR (2009) Expression of four β -galactosidases from *Bifidobacterium bifidum* NCIMB41171 and their contribution on the hydrolysis and synthesis of galactooligosaccharides. *Appl Microbiol Biot* 84(5):899-907 doi:10.1007/s00253-009-2009-5
- Guerrero C, Vera C, Conejeros R, Illanes A (2015a) Transgalactosylation and hydrolytic activities of commercial preparations of β -galactosidase for the synthesis of prebiotic carbohydrates. *Enzyme and Microbial Technology* 70:9-17
doi:10.1016/j.enzmictec.2014.12.006
- Guerrero C, Vera C, Conejeros R, Illanes A (2015b) Transgalactosylation and hydrolytic activities of commercial preparations of β -galactosidase for the synthesis of prebiotic carbohydrates. *Enzyme and Microbial Technology* 70:9-17
- Gum JR, Jr., Crawley SC, Hicks JW, Szymkowski DE, Kim YS (2002) MUC17, a novel membrane-tethered mucin. *Biochemical and biophysical research communications* 291(3):466-75 doi:10.1006/bbrc.2002.6475
- Gum JR, Jr., Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) Molecular cloning of human intestinal mucin (MUC2) cDNA. identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *Journal of Biological Chemistry* 269(4):2440-6
- Guo BS, Zheng F, Crouch L, Cai ZP, Wang M, Bolam DN, Liu L, Voglmeir J (2018) Cloning, purification and biochemical characterisation of a GH35 β -1,3/beta-1,6-galactosidase from the mucin-degrading gut bacterium *Akkermansia muciniphila*. *Glycoconjugate J* doi:10.1007/s10719-018-9824-9
- Hacking AJ, Lin EC (1976) Disruption of the fucose pathway as a consequence of genetic adaptation to propanediol as a carbon source in *Escherichia coli*. *Journal of bacteriology* 126(3):1166

- Han YR, Youn SY, Ji GE, Park MS (2014) Production of α - and β -galactosidases from *Bifidobacterium longum* subsp. *longum* RD47. *Journal of microbiology and biotechnology* 24(5):675-82
- Hanefeld U, Gardossi L, Magner E (2009) Understanding enzyme immobilisation. *Chem Soc Rev* 38(2):453-68 doi:10.1039/b711564b
- Hardeman F, Sjolting S (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS microbiology ecology* 59(2):524-34 doi:10.1111/j.1574-6941.2006.00206.x
- Heath C, Hu XP, Cary SC, Cowan D (2009) Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. *Applied and Environmental Microbiology* 75(13):4657-9 doi:10.1128/AEM.02597-08
- Hedstrom L (2001) Enzyme specificity and selectivity. e LS
- Higuchi T, Orita T, Nakanishi S, Katsuya K, Watanabe H, Yamasaki Y, Waga I, Nanayama T, Yamamoto Y, Munger W, Sun HW, Falk RJ, Jennette JC, Alcorta DA, Li H, Yamamoto T, Saito Y, Nakamura M (2004) Molecular cloning, genomic structure, and expression analysis of MUC20, a novel mucin protein, up-regulated in injured kidney. *Journal of Biological Chemistry* 279(3):1968-79 doi:10.1074/jbc.M304558200
- Hijikata M, Matsushita I, Tanaka G, Tsuchiya T, Ito H, Tokunaga K, Ohashi J, Homma S, Kobashi Y, Taguchi Y, Azuma A, Kudoh S, Keicho N (2011) Molecular cloning of two novel mucin-like genes in the disease-susceptibility locus for diffuse panbronchiolitis. *Hum Genet* 129(2):117-28 doi:10.1007/s00439-010-0906-4
- Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, Morelli L, Canani RB, Flint HJ, Salminen S, Calder PC, Sanders ME (2014) Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nature Reviews Gastroenterology & Hepatology* 11(8):506-14 doi:10.1038/nrgastro.2014.66
- Hinz SW, van den Brock LA, Beldman G, Vincken JP, Voragen AG (2004) β -galactosidase from *Bifidobacterium adolescentis* DSM20083 prefers β (1,4)-galactosides over lactose. *Appl Microbiol Biot* 66(3):276-84 doi:10.1007/s00253-004-1745-9
- Ho SB, Robertson AM, Shekels LL, Lyftogt CT, Niehans GA, Toribara NW (1995) Expression cloning of gastric mucin complementary DNA and localization of mucin gene expression. *Gastroenterology* 109(3):735-47 doi:10.1016/0016-5085(95)90380-1
- Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD (2013) Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PloS one* 8(6):e66019 doi:10.1371/journal.pone.0066019

- Hollingsworth MA, Swanson BJ (2004) Mucins in cancer: protection and control of the cell surface. *Nature reviews Cancer* 4(1):45-60 doi:10.1038/nrc1251
- Homaei AA, Sariri R, Vianello F, Stevanato R (2013) Enzyme immobilization: an update. *Journal of chemical biology* 6(4):185-205 doi:10.1007/s12154-013-0102-9
- Hough DW, Danson MJ (1999) Extremozymes. *Current opinion in chemical biology* 3(1):39-46 doi:10.1016/s1367-5931(99)80008-8
- Hsu C-A, Lee S-L, Chou C-C (2007) Enzymatic production of galactooligosaccharides by β -galactosidase from *Bifidobacterium longum* BCRC 15708. *Journal of agricultural and food chemistry* 55(6):2225-2230
- Huber RE, Kurz G, Wallenfels K (1976) A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (*E. coli*) on lactose. *Biochemistry-U.S.* 15(9):1994-2001 doi:10.1021/bi00654a029
- Hughes HK, Rose D, Ashwood P (2018) The gut microbiota and dysbiosis in autism spectrum disorders. *Current neurology and neuroscience reports* 18(11):81 doi:10.1007/s11910-018-0887-6
- Hunziker O, Nissen B (1926) Lactose solubility and lactose crystal formation: I. Lactose solubility. *Journal of dairy science* 9(6):517-537
- Husain Q (2010) β -galactosidases and their potential applications: a review. *Critical reviews in biotechnology* 30(1):41-62 doi:10.3109/07388550903330497
- Ilmberger N, Meske D, Juergensen J, Schulte M, Barthen P, Rabausch U, Angelov A, Mientus M, Liebl W, Schmitz RA, Streit WR (2012) Metagenomic cellulases highly tolerant towards the presence of ionic liquids-linking thermostability and halotolerance. *Appl Microbiol Biot* 95(1):135-46 doi:10.1007/s00253-011-3732-2
- Imberty A, Varrot A (2008a) Microbial recognition of human cell surface glycoconjugates. *Current opinion in structural biology* 18(5):567-76 doi:10.1016/j.sbi.2008.08.001
- Imberty A, Varrot A (2008b) Microbial recognition of human cell surface glycoconjugates. *Current opinion in structural biology* 18(5):567-576 doi:<http://dx.doi.org/10.1016/j.sbi.2008.08.001>
- Iqbal S, Nguyen T-H, Nguyen TT, Maischberger T, Haltrich D (2010) β -Galactosidase from *Lactobacillus plantarum* WCFS1: biochemical characterization and formation of prebiotic galacto-oligosaccharides. *Carbohydr Res* 345(10):1408-1416 doi:<https://doi.org/10.1016/j.carres.2010.03.028>
- Isabella VM, Ha BN, Castillo MJ, Lubkowitz DJ, Rowe SE, Millet YA, Anderson CL, Li N, Fisher AB, West KA, Reeder PJ, Momin MM, Bergeron CG, Guilmain SE, Miller PF, Kurtz CB, Falb D (2018) Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nature biotechnology* 36(9):857-864 doi:10.1038/nbt.4222

- Ishikawa H, Matsumoto S, Ohashi Y, Imaoka A, Setoyama H, Umesaki Y, Tanaka R, Otani T (2011) Beneficial effects of probiotic *Bifidobacterium* and galacto-oligosaccharide in patients with ulcerative colitis: a randomized controlled study. *Digestion* 84(2):128-33 doi:10.1159/000322977
- Itoh Y, Kamata-Sakurai M, Denda-Nagai K, Nagai S, Tsuiji M, Ishii-Schrade K, Okada K, Goto A, Fukayama M, Irimura T (2008) Identification and expression of human epiglycanin/MUC21: a novel transmembrane mucin. *Glycobiology* 18(1):74-83 doi:10.1093/glycob/cwm118
- Jacques P, Bechet M, Bigan M, Caly D, Chataigne G, Coutte F, Flahaut C, Heuson E, Leclere V, Lecouturier D, Phalip V, Ravallec R, Dhulster P, Froidevaux R (2017) High-throughput strategies for the discovery and engineering of enzymes for biocatalysis. *Bioprocess and biosystems engineering* 40(2):161-180 doi:10.1007/s00449-016-1690-x
- Jakobsson HE, Rodriguez-Pineiro AM, Schutte A, Ermund A, Boysen P, Bemark M, Sommer F, Backhed F, Hansson GC, Johansson ME (2015) The composition of the gut microbiota shapes the colon mucus barrier. *EMBO reports* 16(2):164-77 doi:10.15252/embr.201439263
- James K, Motherway MOC, Bottacini F, van Sinderen D (2016) *Bifidobacterium breve* UCC2003 metabolises the human milk oligosaccharides lacto-N-tetraose and lacto-N-neo-tetraose through overlapping, yet distinct pathways. *Scientific reports* 6(1):38560 doi:10.1038/srep38560
- Jesionowski T, Zdarta J, Krajewska B (2014) Enzyme immobilization by adsorption: a review. *Adsorption* 20(5):801-821 doi:10.1007/s10450-014-9623-y
- Jingyuan S, Abe K, Imanaka H, Imamura K, Minoda M, Yamaguchi S, Nakanishi K (2011) Causes of the production of multiple forms of β -galactosidase by *Bacillus circulans*. *Bioscience, biotechnology, and biochemistry* 75(2):268-278 doi:10.1271/bbb.100574
- Jo Handelsman MRRSFB, Jon Clardy and Robert M Goodman (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5(10):245-249
- Johansson ME, Gustafsson JK, Holmen-Larsson J, Jabbar KS, Xia L, Xu H, Ghishan FK, Carvalho FA, Gewirtz AT, Sjovall H, Hansson GC (2014) Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* 63(2):281-91 doi:10.1136/gutjnl-2012-303207
- Johansson ME, Hansson GC (2016) Immunological aspects of intestinal mucus and mucins. *Nature Reviews Immunology* 16(10):639-49 doi:10.1038/nri.2016.88
- Johansson ME, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proceedings of the National Academy of Sciences* 105(39):15064-15069

- Johansson MEV, Gustafsson JK, Sjöberg KE, Petersson J, Holm L, Sjövall H, Hansson GC (2010) Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS one* 5(8):e12238-e12238 doi:10.1371/journal.pone.0012238
- Johansson MEV, Sjövall H, Hansson GC (2013) The gastrointestinal mucus system in health and disease. *Nature Reviews Gastroenterology & Hepatology* 10(6):352-361 doi:10.1038/nrgastro.2013.35
- Johnson AO, Semanya JG, Buchowski MS, Enwonwu CO, Scrimshaw NS (1993) Correlation of lactose maldigestion, lactose intolerance, and milk intolerance. *American Journal of Clinical Nutrition* 57(3):399-401 doi:10.1093/ajcn/57.3.399
- Joseph Sambrook EFFaTM (2003) *Molecular cloning _ a laboratory manual*-Cold Spring Harbor Laboratory 1-1448
- Jourova L, Anzenbacher P, Matuskova Z, Vecera R, Strojil J, Kolar M, Nobilis M, Hermanova P, Hudcovic T, Kozakova H, Kverka M, Anzenbacherova E (2019) Gut microbiota metabolizes nabumetone *in vitro*: consequences for its bioavailability *in vivo* in the rodents with altered gut microbiome. *Xenobiotica; the fate of foreign compounds in biological systems*:1-7 doi:10.1080/00498254.2018.1558310
- Juge N (2012) Microbial adhesins to gastrointestinal mucus. *Trends in microbiology* 20(1):30-9 doi:10.1016/j.tim.2011.10.001
- Kaoutari AE, Armougom F, Gordon JJ, Raoult D, Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology* 11:497 doi:10.1038/nrmicro3050 <https://www.nature.com/articles/nrmicro3050#supplementary-information>
- Kararli TT (1995) Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. *Biopharmaceutics & Drug Disposition* 16(5):351-380 doi:10.1002/bdd.2510160502
- Kasche WTaV (1999) Immobilized enzymes crystals or carriers. *Trends in Biotechnology* 17:326-335
- Katchalski-Katzir E (1993) Immobilized enzymes--learning from past successes and failures. *Trends in Biotechnology* 11(11):471-8 doi:10.1016/0167-7799(93)90080-s
- Kaur J, Kumar A, Kaur J (2018) Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *International Journal of Biological Macromolecules* 106:803-822 doi:10.1016/j.ijbiomac.2017.08.080
- Kim CS, Ji ES, Oh DK (2003) Expression and characterization of *Kluyveromyces lactis* β -galactosidase in *Escherichia coli*. *Biotechnology letters* 25(20):1769-74

- Kim J, Cheong EY, Jung I, Kim HK (2019) Metabolomic and transcriptomic analyses of *Escherichia coli* for efficient fermentation of L-Fucose. *Marine drugs* 17(2) doi:10.3390/md17020082
- Kim YS, Ho SB (2010) Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Current gastroenterology reports* 12(5):319-30 doi:10.1007/s11894-010-0131-2
- Kleerebezem M, Hols P, Bernard E, Rolain T, Zhou M, Siezen RJ, Bron PA (2010) The extracellular biology of the lactobacilli. *FEMS microbiology reviews* 34(2):199-230 doi:10.1111/j.1574-6976.2009.00208.x
- Kline KA, Falker S, Dahlberg S, Normark S, Henriques-Normark B (2009) Bacterial adhesins in host-microbe interactions. *Cell host & microbe* 5(6):580-92 doi:10.1016/j.chom.2009.05.011
- Koh A, De Vadder F, Kovatcheva-Datchary P, Backhed F (2016) From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* 165(6):1332-1345 doi:10.1016/j.cell.2016.05.041
- Kolida S, Gibson GR (2011) Synbiotics in health and disease. *Annual review of food science and technology* 2:373-93 doi:10.1146/annurev-food-022510-133739
- Koropatkin NM, Cameron EA, Martens EC (2012) How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology* 10(5):323-35 doi:10.1038/nrmicro2746
- Krajewska B (2004) Application of chitin-and chitosan-based materials for enzyme immobilizations: a review. *Enzyme and Microbial Technology* 35(2):126-139 doi:<https://doi.org/10.1016/j.enzmictec.2003.12.013>
- Kunz C, Rudloff S, Baier W, Klein N, Strobel S (2000) Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annual review of nutrition* 20(1):699-722 doi:10.1146/annurev.nutr.20.1.699
- Laane C (1987) Medium-engineering for bio-organic synthesis. *Biocatalysis* 1(1):17-22 doi:10.3109/10242428709040127
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259):680
- Lan MS, Batra SK, Qi WN, Metzgar RS, Hollingsworth MA (1990) Cloning and sequencing of a human pancreatic tumor mucin cDNA. *Journal of Biological Chemistry* 265(25):15294-9
- Lang T, Hansson GC, Samuelsson T (2007) Gel-forming mucins appeared early in metazoan evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104(41):16209-14 doi:10.1073/pnas.0705984104
- Langer M, Gabor EM, Liebeton K, Meurer G, Niehaus F, Schulze R, Eck J, Lorenz P (2006) Metagenomics: an inexhaustible access to nature's diversity. *Biotechnology journal* 1(7-8):815-21 doi:10.1002/biot.200600111

- Lapensee L, Paquette Y, Bleau G (1997) Allelic polymorphism and chromosomal localization of the human oviductin gene (MUC9). *Fertility and Sterility* 68(4):702-8 doi:10.1016/s0015-0282(97)00317-8
- Levitt J (2007) The safety of hydroquinone: a dermatologist's response to the 2006 federal register. *Journal of the American Academy of Dermatology* 57(5):854-72 doi:10.1016/j.jaad.2007.02.020
- Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI (2005) Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America* 102(31):11070-5 doi:10.1073/pnas.0504978102
- Ley RE, Turnbaugh PJ, Klein S, Gordon JI (2006) Human gut microbes associated with obesity. *Nature* 444(7122):1022-1023 doi:10.1038/4441022a
- Lezyk M, Jers C, Kjaerulff L, Gotfredsen CH, Mikkelsen MD, Mikkelsen JD (2016) Novel alpha-L-Fucosidases from a soil metagenome for production of fucosylated human milk oligosaccharides. *PloS one* 11(1):e0147438 doi:10.1371/journal.pone.0147438
- Li Q, Cao L, Tian Y, Zhang P, Ding C, Lu W (2018) Butyrate suppresses the proliferation of colorectal cancer cells via targeting pyruvate kinase M2 and metabolic reprogramming. *PLoS one* 17(8):1531-1545 doi:10.1074/mcp.RA118.000752
- Li S, Yang X, Yang S, Zhu M, Wang X (2012) Technology prospecting on enzymes: application, marketing and engineering. *Comput Struct Biotechnol J* 2:e201209017-e201209017 doi:10.5936/csbj.201209017
- Lilly DM, Stillwell RH (1965) Probiotics: growth promoting factors produced by microorganisms. *Science* 147 doi:10.1126/science.147.3659.747
- Lin CJ, Pan CF, Chuang CK, Sun FJ, Wang DJ, Chen HH, Liu HL, Wu CJ (2014) P-cresyl sulfate is a valuable predictor of clinical outcomes in pre-ESRD patients. *BioMed research international* 2014:526932 doi:10.1155/2014/526932
- Lindén SK, Sheng YH, Every AL, Miles KM, Skoog EC, Florin THJ, Sutton P, McGuckin MA (2009) MUC1 limits *Helicobacter pylori* infection both by steric hindrance and by acting as a releasable decoy. *PLoS pathogens* 5(10):e1000617 doi:10.1371/journal.ppat.1000617
- Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA (2008) Mucins in the mucosal barrier to infection. *Mucosal immunology* 1(3):183-197
- Lipinszki Z, Vernyik V, Farago N, Sari T, Puskas LG, Blattner FR, Posfai G, Gyorfy Z (2018a) Enhancing the translational capacity of *E. coli* by resolving the codon bias. *ACS Synthetic Biology* 7(11):2656-2664 doi:10.1021/acssynbio.8b00332
- Lipinszki Z, Vernyik V, Farago N, Sari T, Puskas LG, Blattner FR, Posfai G, Gyorfy Z (2018b) Enhancing the Translational Capacity of *E. coli* by Resolving the Codon Bias. *ACS Synth Biol* 7(11):2656-2664 doi:10.1021/acssynbio.8b00332

- Liu C, Shao ZM, Zhang L, Beatty P, Sartippour M, Lane T, Livingston E, Nguyen M (2001) Human endomucin is an endothelial marker. *Biochemical and biophysical research communications* 288(1):129-36 doi:10.1006/bbrc.2001.5737
- Liu F, Li J, Wu F, Zheng H, Peng Q, Zhou H (2019a) Altered composition and function of intestinal microbiota in autism spectrum disorders: a systematic review. *Translational psychiatry* 9(1):43 doi:10.1038/s41398-019-0389-6
- Liu F, Li P, Chen M, Luo Y, Prabhakar M, Zheng H, He Y, Qi Q, Long H, Zhang Y, Sheng H, Zhou H (2017a) Fructooligosaccharide (FOS) and galactooligosaccharide (GOS) increase *Bifidobacterium* but reduce butyrate producing bacteria with adverse glycemic metabolism in healthy young population. *Scientific reports* 7(1):11789 doi:10.1038/s41598-017-10722-2
- Liu P, Wang W, Zhao J, Wei D (2019b) Screening novel β -galactosidases from a sequence-based metagenome and characterization of an alkaline β -galactosidase for the enzymatic synthesis of galactooligosaccharides. *Protein Expres Purif* 155:104-111 doi:<https://doi.org/10.1016/j.pep.2018.12.001>
- Liu Y, Chen Z, Jiang Z, Yan Q, Yang S (2017b) Biochemical characterization of a novel β -galactosidase from *Paenibacillus barengoltzii* suitable for lactose hydrolysis and galactooligosaccharides synthesis. *International Journal of Biological Macromolecules* 104(Pt A):1055-1063 doi:10.1016/j.ijbiomac.2017.06.073
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, Casero D, Courtney H, Gonzalez A, Graeber TG, Hall AB, Lake K, Landers CJ, Mallick H, Plichta DR, Prasad M, Rahnavard G, Sauk J, Shungin D, Vazquez-Baeza Y, White RA, 3rd, Investigators I, Braun J, Denson LA, Jansson JK, Knight R, Kugathasan S, McGovern DPB, Petrosino JF, Stappenbeck TS, Winter HS, Clish CB, Franzosa EA, Vlamakis H, Xavier RJ, Huttenhower C (2019) Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569(7758):655-662 doi:10.1038/s41586-019-1237-9
- Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, McDonald D, Franzosa EA, Knight R, White O, Huttenhower C (2017) Strains, functions and dynamics in the expanded human microbiome project. *Nature* 550(7674):61-66 doi:10.1038/nature23889
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014a) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* 42(Database issue):D490-5 doi:10.1093/nar/gkt1178
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014b) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* 42(Database issue):D490-D495 doi:10.1093/nar/gkt1178

- Lopez-Siles M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M (2017) *Faecalibacterium prausnitzii*: from microbiology to diagnostics and prognostics. The ISME journal 11(4):841-852 doi:10.1038/ismej.2016.176
- Lorenz P, Eck J (2005) Metagenomics and industrial applications. Nature Reviews Microbiology 3(6):510-516 doi:10.1038/nrmicro1161
- Louis P, Hold GL, Flint HJ (2014) The gut microbiota, bacterial metabolites and colorectal cancer. Nature Reviews Microbiology 12(10):661-72 doi:10.1038/nrmicro3344
- Lowe JS, Anderson PG (2015) Chapter 11 - Alimentary tract. In: Lowe JS, Anderson PG (eds) Stevens & Lowe's Human Histology (Fourth Edition) (Fourth Edition). Mosby, Philadelphia, pp 186-224
- Lutz S, Iamurri SM (2018) Protein engineering: past, present, and future Protein engineering. Springer, pp 1-12
- Macao B, Johansson DG, Hansson GC, Hard T (2006) Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. Nature structural & molecular biology 13(1):71-6 doi:10.1038/nsmb1035
- Macfarlane S, Macfarlane GT, Cummings JH (2006) Review article: prebiotics in the gastrointestinal tract. Alimentary pharmacology & therapeutics 24(5):701-14 doi:10.1111/j.1365-2036.2006.03042.x
- Macfarlane S, Woodmansey EJ, Macfarlane GT (2005) Colonization of mucin by human intestinal bacteria and establishment of biofilm communities in a two-stage continuous culture system. Applied and environmental microbiology 71(11):7483-92 doi:10.1128/AEM.71.11.7483-7492.2005
- Macpherson AJ, McCoy KD (2015) Standardised animal models of host microbial mutualism. Mucosal Immunology 8(3):476-486 doi:10.1038/mi.2014.113
- Madhavan A, Sindhu R, Parameswaran B, Sukumaran RK, Pandey A (2017) Metagenome Analysis: a powerful tool for enzyme bioprospecting. Applied Biochemistry & Biotechnology 183(2):636-651 doi:10.1007/s12010-017-2568-3
- Maischberger T, Nguyen TH, Sukyai P, Kittl R, Riva S, Ludwig R, Haltrich D (2008) Production of lactose-free galacto-oligosaccharide mixtures: comparison of two cellobiose dehydrogenases for the selective oxidation of lactose to lactobionic acid. Carbohydr Res 343(12):2140-7 doi:10.1016/j.carres.2008.01.040
- Marcobal A, Barboza M, Sonnenburg Erica D, Pudlo N, Martens Eric C, Desai P, Lebrilla Carlito B, Weimer Bart C, Mills David A, German JB, Sonnenburg Justin L (2011) *Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell host & microbe 10(5):507-514 doi:<https://doi.org/10.1016/j.chom.2011.10.007>
- Marcobal A, Southwick AM, Earle KA, Sonnenburg JL (2013) A refined palate: bacterial consumption of host glycans in the gut. Glycobiology 23(9):1038-1046 doi:10.1093/glycob/cwt040

- Marques TM, Wall R, Ross RP, Fitzgerald GF, Ryan CA, Stanton C (2010) Programming infant gut microbiota: influence of dietary and environmental factors. *Current opinion in biotechnology* 21(2):149-56 doi:10.1016/j.copbio.2010.03.020
- Martinez A, Osburne MS (2013) Preparation of fosmid libraries and functional metagenomic analysis of microbial community DNA. *Methods in enzymology* 531:123-42 doi:10.1016/b978-0-12-407863-5.00007-1
- Maruthamuthu M, van Elsas JD (2017) Molecular cloning, expression, and characterization of four novel thermo-alkaliphilic enzymes retrieved from a metagenomic library. *Biotechnology for biofuels* 10:142 doi:10.1186/s13068-017-0808-y
- Mateo C, Palomo JM, Fernandez-Lorente G, Guisan JM, Fernandez-Lafuente R (2007) Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enzyme and Microbial Technology* 40(6):1451-1463 doi:<https://doi.org/10.1016/j.enzmictec.2007.01.018>
- Matsuoka Y, Endo K, Kawamura Y, Yoshida T, Saga T, Watanabe Y, Koizumi M, Nakashima T, Konishi J, Yamaguchi N, *et al.*, (1990) Normal bronchial mucus contains high levels of cancer-associated antigens, CA125, CA19-9, and carcinoembryonic antigen. *Cancer* 65(3):506-10 doi:10.1002/1097-0142(19900201)65:3<506::aid-cnrcr2820650322>3.0.co;2-0
- Matsuzawa T, Yaoi K (2017) Screening, identification, and characterization of a novel saccharide-stimulated β -glycosidase from a soil metagenomic library. *Appl Microbiol Biot* 101(2):633-646 doi:10.1007/s00253-016-7803-2
- Mawson AJ (1994) Bioconversions for whey utilization and waste abatement. *Bioresource technology* 47(3):195-203 doi:[https://doi.org/10.1016/0960-8524\(94\)90180-5](https://doi.org/10.1016/0960-8524(94)90180-5)
- McGuckin MA, Linden SK, Sutton P, Florin TH (2011a) Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* 9(4):265-78 doi:10.1038/nrmicro2538
- McGuckin MA, Lindén SK, Sutton P, Florin TH (2011b) Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* 9(4):265-278
- Mehmetoglu SAaU (1997) A new method for immobilization of β -galactosidase and its utilization in a plug flow reactor. *Process Biochemistry*, 32(5): 433-436
- Mehr K, Withers SG (2016) Mechanisms of the sialidase and trans-sialidase activities of bacterial sialyltransferases from glycosyltransferase family 80. *Glycobiology* 26(4):353-9 doi:10.1093/glycob/cwv105
- Messing RA (1976) [11] Adsorption and inorganic bridge formations *Methods in enzymology*. vol 44. Academic Press, pp 148-169
- Metchnikoff II (1907) Lactic Acid as Inhibiting Intestinal Putrefaction. . In *The Prolongation of Life_ Optimistic Studies (Classics in Longevity and Aging)*:116-129

- Migneault I, Dartiguenave C, Bertrand MJ, Waldron KC (2004) Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *BioTechniques* 37(5):790-802 doi:10.2144/04375RV01
- Milani C, Lugli GA, Duranti S, Turrone F, Mancabelli L, Ferrario C, Mangifesta M, Hevia A, Viappiani A, Scholz M, Arioli S, Sanchez B, Lane J, Ward DV, Hickey R, Mora D, Segata N, Margolles A, van Sinderen D, Ventura M (2015a) *Bifidobacteria* exhibit social behavior through carbohydrate resource sharing in the gut. *Scientific reports* 5:15782 doi:10.1038/srep15782
- Milani C, Mancabelli L, Lugli GA, Duranti S, Turrone F, Ferrario C, Mangifesta M, Viappiani A, Ferretti P, Gorfer V, Tett A, Segata N, van Sinderen D, Ventura M (2015b) Exploring vertical transmission of *Bifidobacteria* from mother to child. *Applied and Environmental Microbiology* 81(20):7078-87 doi:10.1128/aem.02037-15
- Miller J (1972) Miller JH.. *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Mills JP, Rao K, Young VB (2018) Probiotics for prevention of *Clostridium difficile* infection. *Curr Opin Gastroen* 34(1):3
- Mirete S, Morgante V, González-Pastor JE (2016) Functional metagenomics of extreme environments. *Current opinion in biotechnology* 38:143-149 doi:<https://doi.org/10.1016/j.copbio.2016.01.017>
- Mitchell E, Houles C, Sudakevitz D, Wimmerova M, Gautier C, Perez S, Wu AM, Gilboa-Garber N, Imberty A (2002) Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nature structural & molecular biology* 9(12):918-921
- Moehle C, Ackermann N, Langmann T, Aslanidis C, Kel A, Kel-Margoulis O, Schmitz-Madry A, Zahn A, Stremmel W, Schmitz G (2006) Aberrant intestinal expression and allelic variants of mucin genes associated with inflammatory bowel disease. *Journal of Molecular Medicine* 84(12):1055-66 doi:10.1007/s00109-006-0100-2
- Mohamad NR, Marzuki NHC, Buang NA, Huyop F, Wahab RA (2015) An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotechnol Biotechnol Equip* 29(2):205-220 doi:10.1080/13102818.2015.1008192
- Moller PL, Jorgensen F, Hansen OC, Madsen SM, Stougaard P (2001) Intra- and extracellular β -galactosidases from *Bifidobacterium bifidum* and *B. infantis*: molecular cloning, heterologous expression, and comparative characterization. *Applied and Environmental Microbiology* 67(5):2276-83 doi:10.1128/AEM.67.5.2276-2283.2001
- Moncla BJ, Braham P (1989) Detection of sialidase (neuraminidase) activity in *Actinomyces* species by using 2'-(4-methylumbelliferyl) α -D-N-acetylneuraminic acid in a filter paper spot test. *Journal of Clinical Microbiology* 27(1):182

- Moran A, Gupta A, Joshi L (2011) Sweet-talk: role of host glycosylation in bacterial pathogenesis of the gastrointestinal tract. *Gut:gut*. 2010.212704
- Mukherjee S, Seshadri R, Varghese NJ, Eloie-Fadrosch EA, Meier-Kolthoff JP, Göker M, Coates RC, Hadjithomas M, Pavlopoulos GA, Paez-Espino D, Yoshikuni Y, Visel A, Whitman WB, Garrity GM, Eisen JA, Hugenholtz P, Pati A, Ivanova NN, Woyke T, Klenk H-P, Kyrpides NC (2017) 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nature biotechnology* 35(7):676-683 doi:10.1038/nbt.3886
- Nahalka J, Liu Z, Chen X, Wang PG (2003) Superbeads: immobilization in “sweet” chemistry. *Chemistry – A European Journal* 9(2):372-377 doi:10.1002/chem.200390038
- Nath A, Mondal S, Csighy A, Molnár MA, Pásztorné-Huszár K, Kovács Z, Koris A, Vatai G (2017) Biochemical activities of lactose-derived prebiotics — a review. *Acta Alimentaria* 46(4):449-456 doi:10.1556/066.2017.46.4.7
- Neri DFM, Balcão VM, Costa RS, Rocha ICAP, Ferreira EMFC, Torres DPM, Rodrigues LRM, Carvalho LB, Teixeira JA (2009) Galacto-oligosaccharides production during lactose hydrolysis by free *Aspergillus oryzae* β -galactosidase and immobilized on magnetic polysiloxane-polyvinyl alcohol. *Food Chemistry* 115(1):92-99 doi:10.1016/j.foodchem.2008.11.068
- Neufeld KM, Kang N, Bienenstock J, Foster JA (2011) Reduced anxiety-like behavior and central neurochemical change in germ-free mice. *Neurogastroenterology & Motility* 23(3):255-64, e119 doi:10.1111/j.1365-2982.2010.01620.x
- Ng SC, Hart AL, Kamm MA, Stagg AJ, Knight SC (2009) Mechanisms of action of probiotics: recent advances. *Inflamm Bowel Dis* 15 doi:10.1002/ibd.20602
- Ngara TR, Zhang H (2018) Recent advances in function-based metagenomic screening. *Genomics Proteomics Bioinformatics* 16(6):405-415 doi:10.1016/j.gpb.2018.01.002
- Nguyen T-H, Splechna B, Krasteva S, Kneifel W, Kulbe KD, Divne C, Haltrich D (2007) Characterization and molecular cloning of a heterodimeric β -galactosidase from the probiotic strain *Lactobacillus acidophilus* R22. *FEMS microbiology letters* 269(1):136-144 doi:doi:10.1111/j.1574-6968.2006.00614.x
- O'Connell Motherway M, Fitzgerald GF, van Sinderen D (2011) Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microbial biotechnology* 4(3):403-416 doi:10.1111/j.1751-7915.2010.00218.x
- O Sheridan P, Martin JC, Lawley TD, Browne HP, Harris HMB, Bernalier-Donadille A, Duncan SH, O'Toole PW, P Scott K, J Flint H (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microbial genomics* 2(2):e000043-e000043 doi:10.1099/mgen.0.000043

- O'Toole PW, Marchesi JR, Hill C (2017) Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nature Microbiology* 2(5):17057 doi:10.1038/nmicrobiol.2017.57
- Oehler S (2009) Feedback regulation of Lac repressor expression in *Escherichia coli*. *Journal of bacteriology* 191(16):5301-5303 doi:10.1128/JB.00427-09
- Oh SY, Youn SY, Park MS, Kim HG, Baek NI, Li Z, Ji GE (2017) Synthesis of β -Galactooligosaccharide using *Bifidobacterial* β -Galactosidase purified from recombinant *Escherichia coli*. *Journal of Microbiology and Biotechnology* 27(8):1392-1400 doi:10.4014/jmb.1702.02058
- Ohtsubo K, Marth JD (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* 126(5):855-67 doi:10.1016/j.cell.2006.08.019
- Oligschlaeger Y, Yadati T, Houben T (2019) Inflammatory bowel disease: a stressed "gut/feeling". 8(7) doi:10.3390/cells8070659
- Osman A, Tzortzis G, Rastall RA, Charalampopoulos D (2012) BbgIV is an important *Bifidobacterium* β -galactosidase for the synthesis of prebiotic galactooligosaccharides at high temperatures. *Journal of Agriculture and Food Chemistry* 60(3):740-8 doi:10.1021/jf204719w
- Ostedgaard LS, Moninger TO, McMenimen JD, Sawin NM, Parker CP, Thornell IM, Powers LS, Gansemer ND, Bouzek DC, Cook DP, Meyerholz DK, Abou Alaiwa MH, Stoltz DA, Welsh MJ (2017) Gel-forming mucins form distinct morphologic structures in airways. *Proceedings of the National Academy of Sciences of the United States of America* 114(26):6842-6847 doi:10.1073/pnas.1703228114
- Ott M, Hacker J, Schmoll T, Jarchau T, Korhonen TK, Goebel W (1986) Analysis of the genetic determinants coding for the S-fimbrial adhesin (sfa) in different *Escherichia coli* strains causing meningitis or urinary tract infections. *Infection and immunity* 54(3):646-653
- Pallesen LT, Berglund L, Rasmussen LK, Petersen TE, Rasmussen JT (2002) Isolation and characterization of MUC15, a novel cell membrane-associated mucin. *European Journal of Biochemistry* 269(11):2755-63 doi:10.1046/j.1432-1033.2002.02949.x
- Papadopoulos JS, Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics (Oxford, England)* 23(9):1073-9 doi:10.1093/bioinformatics/btm076
- Park A-R, Oh D-K (2010) Galacto-oligosaccharide production using microbial β -galactosidase: current state and perspectives. *Appl Microbiol Biot* 85(5):1279-1286 doi:10.1007/s00253-009-2356-2
- Park YK, De Santi MSS, Pastore GM (1979) Production and characterization of β -Galactosidase from *Aspergillus oryzae*. *Journal of Food Science* 44(1):100-103 doi:10.1111/j.1365-2621.1979.tb10016.x

- Patterson E, Ryan PM, Cryan JF, Dinan TG, Ross RP, Fitzgerald GF, Stanton C (2016) Gut microbiota, obesity and diabetes. *Postgraduate medical journal* 92(1087):286-300 doi:10.1136/postgradmedj-2015-133285
- Paul B. Eckburg EMB, 2 Charles N. Bernstein,3, Elizabeth Purdom LD, 2 Michael Sargent,3, Steven R. Gill KEN, 5 David A. Relman1,2,6* (2005) <Diversity of the Human Intestinal Microbial Flora.pdf>. *Science* 308:5
- Pawlak-Szukalska A, Wanarska M, Popinigis AT, Kur J (2014) A novel cold-active β -d-galactosidase with transglycosylation activity from the Antarctic *Arthrobacter* sp. 32cB – gene cloning, purification and characterization. *Process Biochemistry* 49(12):2122-2133 doi:10.1016/j.procbio.2014.09.018
- Pelaseyed T, Bergstro JH, ¨m, Gustafsson JK, Ermund A, Birchenough GMH, Andre, Schu, ¨tte, Post Svd, Svensson F, Rodri´guez-Pin AM, Nystro EEL, Wising C, Johansson MEV, Hansson GC (2014) The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunological Reviews* 260:8–20
- Peracchi A (2018) The limits of enzyme specificity and the evolution of metabolism. *Trends in biochemical sciences* 43(12):984-996 doi:<https://doi.org/10.1016/j.tibs.2018.09.015>
- Perdijk O, van Baarlen P, Fernandez-Gutierrez MM, van den Brink E, Schuren FHJ, Brugman S, Savelkoul HFJ, Kleerebezem M, van Neerven RJJ (2019) Sialyllactose and galactooligosaccharides promote epithelial barrier functioning and distinctly modulate microbiota composition and short chain fatty acid production *in vitro*. *Frontiers in Immunology* 10(94) doi:10.3389/fimmu.2019.00094
- Perry RJ, Peng L, Barry NA, Cline GW, Zhang D, Cardone RL, Petersen KF, Kibbey RG, Goodman AL, Shulman GI (2016) Acetate mediates a microbiome–brain– β -cell axis to promote metabolic syndrome. *Nature* 534(7606):213-217 doi:10.1038/nature18309
- Pham VHT, Kim J (2012) Cultivation of unculturable soil bacteria. *Trends in Biotechnology* 30(9):475-484 doi:<https://doi.org/10.1016/j.tibtech.2012.05.007>
- Pillaiyar T, Manickam M, Namasivayam V (2017) Skin whitening agents: medicinal chemistry perspective of tyrosinase inhibitors. *J Enzyme Inhib Med Chem* 32(1):403-425 doi:10.1080/14756366.2016.1256882
- Pokusaeva K, Fitzgerald GF, van Sinderen D (2011) Carbohydrate metabolism in *Bifidobacteria*. *Genes & nutrition* 6(3):285-306 doi:10.1007/s12263-010-0206-6
- Porchet N, Pigny P, Buisine MP, Debailleul V, Degand P, Laine A, Aubert JP (1995) Human mucin genes: genomic organization and expression of MUC4, MUC5AC and MUC5B. *Biochemical Society transactions* 23(4):800-5 doi:10.1042/bst0230800

- Porter JL, Rusli RA, Ollis DL (2016) Directed evolution of enzymes for industrial biocatalysis. *Chembiochem* 17(3):197-203 doi:10.1002/cbic.201500280
- Pratt WS, Crawley S, Hicks J, Ho J, Nash M, Kim YS, Gum JR, Swallow DM (2000) Multiple transcripts of MUC3: evidence for two genes, MUC3A and MUC3B. *Biochemical and biophysical research communications* 275(3):916-23 doi:10.1006/bbrc.2000.3406
- Proctor LM, Creasy HH, Fettweis JM, Lloyd-Price J, Mahurkar A, Zhou W, Buck GA, Snyder MP, Strauss JF, Weinstock GM, White O, Huttenhower C, The Integrative HMPRNC (2019) The integrative human microbiome project. *Nature* 569(7758):641-648 doi:10.1038/s41586-019-1238-8
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59-65 doi:10.1038/nature08821
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490(7418):55-60 doi:10.1038/nature11450
- Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, Zhou J, Ni S, Liu L, Pons N, Batto JM, Kennedy SP, Leonard P, Yuan C, Ding W, Chen Y, Hu X, Zheng B, Qian G, Xu W, Ehrlich SD, Zheng S, Li L (2014) Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513:59 doi:10.1038/nature13568
<https://www.nature.com/articles/nature13568#supplementary-information>
- Rastall RA, Gibson GR (2015) Recent developments in prebiotics to selectively impact beneficial microbes and promote intestinal health. *Current opinion in biotechnology* 32 doi:10.1016/j.copbio.2014.11.002
- Rastall RA, Gibson GR, Gill HS, Guarner F, Klaenhammer TR, Pot B, Reid G, Rowland IR, Sanders ME (2005) Modulation of the microbial ecology of the human colon by probiotics, prebiotics and synbiotics to enhance human health: an overview of enabling science and potential applications. *FEMS microbiology ecology* 52(2):145-52 doi:10.1016/j.femsec.2005.01.003

- Ravcheev DA, Godzik A, Osterman AL, Rodionov DA (2013) Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. *BMC genomics* 14(1):873 doi:10.1186/1471-2164-14-873
- Ravcheev DA, Thiele I (2017) Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host-synthesized mucin glycans and utilization of mucin-derived monosaccharides. *Frontiers in genetics* 8:111-111 doi:10.3389/fgene.2017.00111
- Reuter S, Rusborg Nygaard A, Zimmermann W (1999) β -Galactooligosaccharide synthesis with β -galactosidases from *Sulfolobus solfataricus*, *Aspergillus oryzae*, and *Escherichia coli*. *Enzyme and Microbial Technology* 25(6):509-516 doi:[https://doi.org/10.1016/S0141-0229\(99\)00074-5](https://doi.org/10.1016/S0141-0229(99)00074-5)
- Riesenfeld CS, Schloss PD, Handelsman J (2004) Metagenomics: genomic analysis of microbial communities. *Annual review of genetics* 38:525-52 doi:10.1146/annurev.genet.38.072902.091216
- Riggio O, Varriale M, Testore GP, Di Rosa R, Di Rosa E, Merli M, Romiti A, Candiani C, Capocaccia L (1990) Effect of lactitol and lactulose administration on the fecal flora in cirrhotic patients. *Journal of clinical gastroenterology* 12(4):433-6 doi:10.1097/00004836-199008000-00016
- Rodriguez-Colinas B, de Abreu MA, Fernandez-Arrojo L, de Beer R, Poveda A, Jimenez-Barbero J, Haltrich D, Ballesteros Olmo AO, Fernandez-Lobato M, Plou FJ (2011) Production of galacto-oligosaccharides by the β -galactosidase from *Kluyveromyces lactis*: comparative analysis of permeabilized cells versus soluble enzyme. *Journal of agricultural and food chemistry* 59(19):10477-10484
- Rodriguez-Colinas B, Fernandez-Arrojo L, Ballesteros AO, Plou FJ (2014) Galactooligosaccharides formation during enzymatic hydrolysis of lactose: towards a prebiotic-enriched milk. *Food Chemistry* 145:388-94 doi:10.1016/j.foodchem.2013.08.060
- Roessl U, Nahálka J, Nidetzky B (2010) Carrier-free immobilized enzymes for biocatalysis. *Biotechnology letters* 32(3):341-350 doi:10.1007/s10529-009-0173-4
- Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and environmental microbiology* 66(6):2541-7 doi:10.1128/aem.66.6.2541-2547.2000
- Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in microbiology* 5:172-172 doi:10.3389/fmicb.2014.00172

- Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology* 9(5):313-323 doi:10.1038/nri2515
- Ryan MP, Walsh G (2016) The biotechnological potential of whey. *Reviews in Environmental Science and Bio/Technology* 15(3):479-498 doi:10.1007/s11157-016-9402-1
- Sako T, Matsumoto K, Tanaka R (1999) Recent progress on research and applications of non-digestible galacto-oligosaccharides. *Int Dairy J* 9(1):69-80 doi:[https://doi.org/10.1016/S0958-6946\(99\)00046-1](https://doi.org/10.1016/S0958-6946(99)00046-1)
- Salli K, Anglenius H, Hirvonen J, Hibberd AA, Ahonen I, Saarinen MT, Tiihonen K, Maukonen J, Ouwehand AC (2019) The effect of 2'-fucosyllactose on simulated infant gut microbiome and metabolites; a pilot study in comparison to GOS and lactose. *Scientific reports* 9(1):13232 doi:10.1038/s41598-019-49497-z
- Satyanarayana T, Raghukumar C, Shivaji S (2005) Extremophilic microbes: diversity and perspectives. *Current Science* 89(1):78-90
- Scaldaferri F, Gerardi V, Lopetuso LR, Del Zompo F, Mangiola F, Boskoski I, Bruno G, Petito V, Laterza L, Cammarota G, Gaetani E, Sgambato A, Gasbarrini A (2013) Gut microbial flora, prebiotics, and probiotics in IBD: their current usage and utility. *BioMed research international* 2013:435268 doi:10.1155/2013/435268
- Schirmer M, Garner A, Vlamakis H, Xavier RJ (2019) Microbial genes and pathways in inflammatory bowel disease. *Nature Reviews Microbiology* 17(8):497-511 doi:10.1038/s41579-019-0213-6
- Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turrone S, Biagi E, Peano C, Severgnini M, Fiori J, Gotti R, De Bellis G, Luiselli D, Brigidi P, Mabulla A, Marlowe F, Henry AG, Crittenden AN (2014) Gut microbiome of the Hadza hunter-gatherers. *Nature communications* 5(1):3654 doi:10.1038/ncomms4654
- Schroeder BO, Bäckhed F (2016) Signals from the gut microbiota to distant organs in physiology and disease. *Nature Medicine* 22:1079 doi:10.1038/nm.4185
- Schuijt TJ, Lankelma JM, Scicluna BP, de Sousa e Melo F, Roelofs JJ, de Boer JD, Hoogendijk AJ, de Beer R, de Vos A, Belzer C, de Vos WM, van der Poll T, Wiersinga WJ (2016) The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut microbiota* 65(4):575-83 doi:10.1136/gutjnl-2015-309728
- Selvarajan E, Mohanasrinivasan V (2015) Kinetic studies on exploring lactose hydrolysis potential of β galactosidase extracted from *Lactobacillus plantarum* HF571129. *J Food Sci Technol* 52(10):6206-6217 doi:10.1007/s13197-015-1729-z

- Shankar V, Gilmore MS, Elkins RC, Sachdev GP (1994) A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization. *The Biochemical journal* 300 (Pt 2)(Pt 2):295-298 doi:10.1042/bj3000295
- Shao Y, Forster SC, Tsaliki E, Vervier K, Strang A, Simpson N, Kumar N, Stares MD, Rodger A, Brocklehurst P, Field N, Lawley TD (2019) Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* 574(7776):117-121 doi:10.1038/s41586-019-1560-1
- Sheldon RA (2007) Enzyme immobilization: the quest for optimum performance. *Advanced Synthesis & Catalysis* 349(8-9):1289-1307 doi:10.1002/adsc.200700082
- Shyu MK, Lin MC, Shih JC, Lee CN, Huang J, Liao CH, Huang IF, Chen HY, Huang MC, Hsieh FJ (2007) Mucin 15 is expressed in human placenta and suppresses invasion of trophoblast-like cells *in vitro*. *Human Reproduction* 22(10):2723-32 doi:10.1093/humrep/dem249
- Silva C, Martins M, Jing S, Fu J, Cavaco-Paulo A (2018) Practical insights on enzyme stabilization. *Critical reviews in biotechnology* 38(3):335-350 doi:10.1080/07388551.2017.1355294
- Šimun Zamberlin NA, Jasmina Havranek, and Dubravka Samaržija (2012) Mineral elements in milk and dairy products. *Mljekarstvo* 62((2)):111-125
- Singh R, Kumar M, Mittal A, Mehta PK (2016) Microbial enzymes: industrial progress in 21st century. *3 Biotech* 6(2):174 doi:10.1007/s13205-016-0485-8
- Sonnenburg JL, Sonnenburg ED (2019) Vulnerability of the industrialized microbiota. *Science* 366(6464):eaaw9255 doi:10.1126/science.aaw9255
- Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307(5717):1955-9 doi:10.1126/science.1109051
- Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiol* 39:321-46 doi:10.1146/annurev.mi.39.100185.001541
- Sun J, Chang EB (2014) Exploring gut microbes in human health and disease: pushing the envelope. *Genes & Diseases* 1(2):132-139 doi:<http://dx.doi.org/10.1016/j.gendis.2014.08.001>
- Bifidobacterium adolescentis* complete genome sequence (2006) https://www.ncbi.nlm.nih.gov/nuccore/NC_008618.1?report=genbank&from=2009997&to=2013146.
- Sweet MJ, Hume DA (1996) Endotoxin signal transduction in macrophages. *J Leukocyte Biol* 60(1):8-26 doi:10.1002/jlb.60.1.8
- Tailford LE, Crost EH, Kavanaugh D, Juge N (2015) Mucin glycan foraging in the human gut microbiome. *Frontiers Genetics* 6:81 doi:10.3389/fgene.2015.00081

- Tan J, McKenzie C, Potamitis M, Thorburn AN, Mackay CR, Macia L (2014) The role of short-chain fatty acids in health and disease. *Advances in immunology* 121:91-119 doi:10.1016/b978-0-12-800100-4.00003-9
- Tanriseven A, Doğan Ş (2002) A novel method for the immobilization of β -galactosidase. *Process Biochemistry* 38(1):27-30 doi:[https://doi.org/10.1016/S0032-9592\(02\)00049-3](https://doi.org/10.1016/S0032-9592(02)00049-3)
- Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, Cantarel BL, Coutinho PM, Henrissat B, Leclerc M, Dore J, Monsan P, Remaud-Simeon M, Potocki-Veronese G (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome research* 20(11):1605-12 doi:10.1101/gr.108332.110
- Tawfik DS (2014) Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Current opinion in chemical biology* 21:73-80 doi:10.1016/j.cbpa.2014.05.008
- Terrapon N, Lombard V, Drula E, Lapebie P, Al-Masaudi S, Gilbert HJ, Henrissat B (2018) PULDB: the expanded database of polysaccharide utilization loci. *Nucleic acids research* 46(D1):D677-d683 doi:10.1093/nar/gkx1022
- The Human Microbiome Project C, Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PSG, Chen IMA, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Michael Dunne W, Scott Durkin A, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney LJ, Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Griggs A, Gujja S, Kinder Haake S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, King NB, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, La Rosa PS, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu Y, Lo C-C, Lozupone CA, Dwayne Lunsford R, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavromatis K, McCorrison JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Pop M, Priest ME,

Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers Y-H, Ross MC, Russ C, Sanka RK, Sankar P, Fah Sathirapongsasuti J, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Warren W, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu Y, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Methé BA, Nelson KE, Petrosino JF, Weinstock GM, Wilson RK, White O (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486:207 doi:10.1038/nature11234
<https://www.nature.com/articles/nature11234#supplementary-information>

Thongaram T, Hoeflinger JL, Chow J, Miller MJ (2017a) Human milk oligosaccharide consumption by probiotic and human-associated *Bifidobacteria* and *Lactobacilli*. *Journal of dairy science* 100(10):7825-7833 doi:10.3168/jds.2017-12753

Thongaram T, Hoeflinger JL, Chow J, Miller MJ (2017b) Prebiotic galactooligosaccharide metabolism by probiotic *Lactobacilli* and *Bifidobacteria*. *Journal of agricultural and food chemistry* 65(20):4184-4192 doi:10.1021/acs.jafc.7b00851

Tissier H (1907) Traitement des infections intestinales par la methode de transformation de la flore bacterienne de l'intestin. [s.n.], [S.l.]

Tomasello G, Mazzola M, Leone A, Sinagra E, Zummo G, Farina F, Damiani P, Cappello F, Gerges Geagea A, Jurjus A, Bou Assi T, Messina M, Carini F (2016) Nutrition, oxidative stress and intestinal dysbiosis: Influence of diet on gut microbiota in inflammatory bowel diseases. *Biomedical Papers* 160(4):461-466 doi:10.5507/bp.2016.052

Tomasello G, Tralongo P, Damiani P, Sinagra E, Di Trapani B, Zeenny MN, Hussein IH, Jurjus A, Leone A (2014) Dismicrobism in inflammatory bowel disease and colorectal cancer: changes in response of colocytes. *World Journal of Gastroenterology* 20(48):18121-30 doi:10.3748/wjg.v20.i48.18121

Toribara NW, Ho SB, Gum E, Gum JR, Jr., Lau P, Kim YS (1997) The carboxyl-terminal sequence of the human secretory mucin, MUC6. Analysis Of the primary amino acid sequence. *Journal of Biological Chemistry* 272(26):16398-403 doi:10.1074/jbc.272.26.16398

Torres DPM, Gonçalves MdPF, Teixeira JA, Rodrigues LR (2010a) Galacto-Oligosaccharides: Production, Properties, Applications, and Significance as Prebiotics. *Comprehensive Reviews in Food Science and Food Safety* 9(5):438-454 doi:10.1111/j.1541-4337.2010.00119.x

- Torres DPM, Gonçalves MdPF, Teixeira JA, Rodrigues LR (2010b) Galactooligosaccharides: production, properties, applications, and significance as prebiotics. *Comprehensive Reviews in Food Science and Food Safety* 9(5):438-454 doi:10.1111/j.1541-4337.2010.00119.x
- Tran DT, Ten Hagen KG (2013) Mucin-type O-glycosylation during development. *Journal of Biological Chemistry* 288(10):6921-9 doi:10.1074/jbc.R112.418558
- Trent CM, Blaser MJ (2016) Microbially produced acetate: a "missing link" in understanding obesity? *Cell metabolism* 24(1):9-10 doi:10.1016/j.cmet.2016.06.023
- Trevan MD (1988) Enzyme immobilization by entrapment. *Methods in molecular biology* 3:491-4 doi:10.1385/0-89603-126-8:491
- Tumturk H, Gokhan D, LTINOK H, AKSOY S, HASIRCI N (2008) Immobilization of glucose isomerase in surface-modified alginate gel beads. *Journal of Food Biochemistry* 32(2):234-246 doi:10.1111/j.1745-4514.2008.00171.x
- Tuomanen AIMHaEI (1992) Consequences of microbial attachment: directing host cell functions with adhesins. *Infection and immunity* 60, (5): 1729-1733
- Tuomola E, Crittenden R, Playne M, Isolauri E, Salminen S (2001) Quality assurance criteria for probiotic bacteria. *The American Journal of Clinical Nutrition* 73(2):393s-398s doi:10.1093/ajcn/73.2.393s
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. *Nature* 449(7164):804-10 doi:10.1038/nature06244
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444(7122):1027-31 doi:10.1038/nature05414
- Turroni F, Bottacini F, Foroni E, Mulder I, Kim J-H, Zomer A, Sánchez B, Bidossi A, Ferrarini A, Giubellini V (2010) Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proceedings of the National Academy of Sciences* 107(45):19514-19519
- Turroni F, Duranti S, Bottacini F, Guglielmetti S, Van Sinderen D, Ventura M (2014) *Bifidobacterium bifidum* as an example of a specialized human gut commensal. *Frontiers in microbiology* 5:437 doi:10.3389/fmicb.2014.00437
- Turroni F, Milani C, Duranti S, Mahony J, van Sinderen D, Ventura M (2018) Glycan Utilization and cross-feeding activities by *Bifidobacteria*. *Trends in microbiology* 26(4):339-350 doi:10.1016/j.tim.2017.10.001
- Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M (2012) Diversity of *Bifidobacteria* within the infant gut microbiota. *PloS one* 7(5):e36957 doi:10.1371/journal.pone.0036957

- Uchiyama T, Miyazaki K (2009) Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current opinion in biotechnology* 20(6):616-22 doi:10.1016/j.copbio.2009.09.010
- Vaishnava S, Behrendt CL, Ismail AS, Eckmann L, Hooper LV (2008) Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proceedings of the National Academy of Sciences* 105(52):20858 doi:10.1073/pnas.0808723105
- Van de Graaff KM (1986) Anatomy and physiology of the gastrointestinal tract. *Pediatric infectious disease* 5(1 Suppl):S11-6
- Van den Broek LAM, Hinz SWA, Beldman G, Doeswijk-Voragen CHL, Vincken J-P, Voragen AGJ (2005) Glycosyl hydrolases from *Bifidobacterium adolescentis* DSM20083. an overview. *Le Lait* 85(1-2):125-133 doi:10.1051/lait:2005004
- Van Laere KM, Abee T, Schols HA, Beldman G, Voragen AG (2000) Characterization of a novel β -galactosidase from *Bifidobacterium adolescentis* DSM 20083 active towards transgalactooligosaccharides. *Applied and environmental microbiology* 66(4):1379-1384 doi:10.1128/aem.66.4.1379-1384.2000
- Van Laere KMA, T. Schols, H. A. Beldman, G. Voragen, A. G. (2000) Characterization of a novel β -galactosidase from *Bifidobacterium adolescentis* DSM 20083 active towards transgalactooligosaccharides. *Applied and environmental microbiology* 66(4):1379-84
- van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JFWM, Tijssen JGP, Speelman P, Dijkgraaf MGW, Keller JJ (2013) Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *New England Journal of Medicine* 368(5):407-415 doi:10.1056/NEJMoa1205037
- van Putten JPM, Strijbis K (2017) Transmembrane mucins: signaling receptors at the intersection of inflammation and cancer. *Journal of innate immunity* 9(3):281-299 doi:10.1159/000453594
- Vandenplas Y (2015) Lactose intolerance. *Asia Pacific journal of clinical nutrition* 24 Suppl 1:S9-13 doi:10.6133/apjcn.2015.24.s1.02
- Vasta GR (2009) Roles of galectins in infection. *Nature Reviews Microbiology* 7(6):424-38 doi:10.1038/nrmicro2146
- Venetsaneas N, Antonopoulou G, Stamatelatou K, Kornaros M, Lyberatos G (2009) Using cheese whey for hydrogen and methane generation in a two-stage continuous process with alternative pH controlling approaches. *Bioresource technology* 100(15):3713-7 doi:10.1016/j.biortech.2009.01.025
- Vera C, Guerrero C, Illanes A (2011) Determination of the transgalactosylation activity of *Aspergillus oryzae* β -galactosidase: effect of pH, temperature, and galactose and glucose concentrations. *Carbohydr Res* 346(6):745-52 doi:10.1016/j.carres.2011.01.030

- Verhoef SP, Meyer D, Westerterp KR (2011) Effects of oligofructose on appetite profile, glucagon-like peptide 1 and peptide YY3-36 concentrations and energy intake. *British Journal of Nutrition* 106(11):1757-62
doi:10.1017/s0007114511002194
- Veronese FM (2001) Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials* 22(5):405-417 doi:[https://doi.org/10.1016/S0142-9612\(00\)00193-9](https://doi.org/10.1016/S0142-9612(00)00193-9)
- Viborg AH, Katayama T, Abou Hachem M, Andersen MC, Nishimoto M, Clausen MH, Urashima T, Svensson B, Kitaoka M (2014) Distinct substrate specificities of three glycoside hydrolase family 42 β -galactosidases from *Bifidobacterium longum* subsp. *infantis* ATCC 15697. *Glycobiology* 24(2):208-16
doi:10.1093/glycob/cwt104
- Vilei EM, Johansson A, Schlatter Y, Redhead K, Frey J (2011) Genetic and functional characterization of the NanA sialidase from *Clostridium chauvoei*. *Veterinary Research* 42(1):2 doi:10.1186/1297-9716-42-2
- Vulevic J, Drakoularakou A, Yaqoob P, Tzortzis G, Gibson GR (2008) Modulation of the fecal microflora profile and immune function by a novel trans-galactooligosaccharide mixture (B-GOS) in healthy elderly volunteers. *American Journal of Clinical Nutrition* 88(5):1438-46
doi:10.3945/ajcn.2008.26242
- Walter J, Ley R (2011) The human gut microbiome: ecology and recent evolutionary changes. *Annual review of microbiology* 65(1):411-429 doi:10.1146/annurev-micro-090110-102830
- Walton GE, van den Heuvel EG, Kusters MH, Rastall RA, Tuohy KM, Gibson GR (2012) A randomised crossover study investigating the effects of galactooligosaccharides on the faecal microbiota in men and women over 50 years of age. *British Journal of Nutrition* 107(10):1466-75
doi:10.1017/s0007114511004697
- Wang K, Li G, Yu SQ, Zhang CT, Liu YH (2010) A novel metagenome-derived β -galactosidase: gene cloning, overexpression, purification and characterization. *Appl Microbiol Biot* 88(1):155-65 doi:10.1007/s00253-010-2744-7
- Wang M, Jia C, Qi W, Yu Q, Peng X, Su R, He Z (2011a) Porous-CLEAs of papain: application to enzymatic hydrolysis of macromolecules. *Bioresource technology* 102(3):3541-3545 doi:<https://doi.org/10.1016/j.biortech.2010.08.120>
- Wang M, Qi W, Su R, He Z (2015) Advances in carrier-bound and carrier-free immobilized nanobiocatalysts. *Chemical Engineering Science* 135:21-32
doi:10.1016/j.ces.2015.03.051
- Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, DuGar B, Feldstein AE, Britt EB, Fu X, Chung Y-M, Wu Y, Schauer P, Smith JD, Allayee H, Tang WHW, DiDonato JA, Lusis AJ, Hazen SL (2011b) Gut flora metabolism of

- phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57-63
doi:10.1038/nature09922
- Wardwell LH, Huttenhower C, Garrett WS (2011) Current concepts of the intestinal microbiota and the pathogenesis of Infection. *Current Infectious Disease Reports* 13(1):28-34 doi:10.1007/s11908-010-0147-7
- Warmerdam A, Zisopoulos FK, Boom RM, Janssen AE (2014) Kinetic characterization of galacto-oligosaccharide (GOS) synthesis by three commercially important β -galactosidases. *Biotechnology progress* 30(1):38-47 doi:10.1002/btpr.1828
- Warwas ML, Yeung JH, Indurugalla D, Mooers AO, Bennet AJ, Moore MM (2010) Cloning and characterization of a sialidase from the filamentous fungus, *Aspergillus fumigatus*. *Glycoconjugate J* 27(5):533-48 doi:10.1007/s10719-010-9299-9
- Weber FL, Jr. (1996) Lactulose and combination therapy of hepatic encephalopathy: the role of the intestinal microflora. *Digestive Diseases* 14 Suppl 1:53-63
doi:10.1159/000171583
- Weitkunat K, Schumann S, Nickel D, Kappo KA, Petzke KJ, Kipp AP, Blaut M, Klaus S (2016) Importance of propionate for the repression of hepatic lipogenesis and improvement of insulin sensitivity in high-fat diet-induced obesity. *Molecular nutrition & food research* 60(12):2611-2621 doi:10.1002/mnfr.201600305
- Wexler AG, Goodman AL (2017) An insider's perspective: *Bacteroides* as a window into the microbiome. *Nature Microbiology* 2:17026
doi:10.1038/nmicrobiol.2017.26
- Wheatley RW, Lo S, Jancewicz LJ, Dugdale ML, Huber RE (2013) Structural explanation for allolactose (lac operon inducer) synthesis by lacZ β -galactosidase and the evolutionary relationship between allolactose synthesis and the lac repressor. *Journal of Biological Chemistry* 288(18):12993-3005
doi:10.1074/jbc.M113.455436
- WHO G (2003) Global strategy for infant and young child feeding.
- Williams SJ, McGuckin MA, Gotley DC, Eyre HJ, Sutherland GR, Antalis TM (1999) Two novel mucin genes down-regulated in colorectal cancer identified by differential display. *Cancer Res* 59(16):4083-9
- Williams SJ, Wreschner DH, Tran M, Eyre HJ, Sutherland GR, McGuckin MA (2001) Muc13, a novel human cell surface mucin expressed by epithelial and hemopoietic cells. *Journal of Biological Chemistry* 276(21):18327-36
doi:10.1074/jbc.M008850200
- Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, Järnerot G, Tysk C, Jansson JK, Engstrand L (2010a) A Pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139(6):1844-1854.e1
doi:<https://doi.org/10.1053/j.gastro.2010.08.049>

- Willing BP, Gill N, Finlay BB (2010b) The role of the immune system in regulating the microbiota. *Gut microbes* 1(4):213-223 doi:10.4161/gmic.1.4.12520
- Willing BP, Russell SL, Finlay BB (2011) Shifting the balance: antibiotic effects on host-microbiota mutualism. *Nature Reviews Microbiology* 9(4):233-43 doi:10.1038/nrmicro2536
- Wong JM, de Souza R, Kendall CW, Emam A, Jenkins DJ (2006) Colonic health: fermentation and short chain fatty acids. *Journal of clinical gastroenterology* 40(3):235-43 doi:10.1097/00004836-200603000-00015
- Wyss MT, Magistretti PJ, Buck A, Weber B (2011) Labeled acetate as a marker of astrocytic metabolism. *Journal of Cerebral Blood Flow & Metabolism* 31(8):1668-74 doi:10.1038/jcbfm.2011.84
- Yatsunenکو T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI (2012a) Human gut microbiome viewed across age and geography. *Nature* 486(7402):222-227 doi:10.1038/nature11053
- Yatsunenکو T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI (2012b) Human gut microbiome viewed across age and geography. *Nature* 486:222 doi:10.1038/nature11053
<https://www.nature.com/articles/nature11053#supplementary-information>
- Yi SH, Alli I, Park KH, Lee B (2011) Overexpression and characterization of a novel transgalactosylase and hydrolytic β -galactosidase from a human isolate *Bifidobacterium breve* B24. *New Biotechnology* 28(6):806-13 doi:10.1016/j.nbt.2011.07.006
- Yin H, Bultema JB, Dijkhuizen L, van Leeuwen SS (2017) Reaction kinetics and galactooligosaccharide product profiles of the β -galactosidases from *Bacillus circulans*, *Kluyveromyces lactis* and *Aspergillus oryzae*. *Food Chemistry* 225:230-238 doi:10.1016/j.foodchem.2017.01.030
- Yoshida E, Sakurama H, Kiyohara M, Nakajima M, Kitaoka M, Ashida H, Hirose J, Katayama T, Yamamoto K, Kumagai H (2012) *Bifidobacterium longum* subsp. *infantis* uses two different β -galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology* 22(3):361-8 doi:10.1093/glycob/cwr116
- Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics (Oxford, England)* 26(13):1608-15 doi:10.1093/bioinformatics/btq249

- Zhang X, Li H, Li CJ, Ma T, Li G, Liu YH (2013) Metagenomic approach for the isolation of a thermostable β -galactosidase with high tolerance of galactose and glucose from soil samples of Turpan Basin. *BMC microbiology* 13:237
doi:10.1186/1471-2180-13-237
- Zhou QZK, Dong Chen X (2001) Immobilization of β -galactosidase on graphite surface by glutaraldehyde. *Journal of Food Engineering* 48(1):69-74
doi:[https://doi.org/10.1016/S0260-8774\(00\)00147-3](https://doi.org/10.1016/S0260-8774(00)00147-3)
- Zhu B, Wang X, Li L (2010) Human gut microbiome: the second genome of human body. *Protein & cell* 1(8):718-25 doi:10.1007/s13238-010-0093-z
- Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans ADL, de Vos WM (2002) Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Applied and environmental microbiology* 68(7):3401
doi:10.1128/AEM.68.7.3401-3407.2002