

Sialic acid metabolism in gut microbes

Andrew Bell (BSc)

A thesis submitted for the degree of Doctor of Philosophy (PhD)
To the University of East Anglia

Quadram Institute Bioscience
Gut Microbes and Health
Norwich Research Park
Norwich
NR4 7UQ

September 2019

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK copyright law. In addition, any quotation or extract must include full attribution.

Abstract

Sialic acids are nine carbon monosaccharides that play a crucial role in many biological processes. In the gastrointestinal tract they are found at the terminal position of mucin glycans, and as such are a highly sought-after nutrient source for members of the gut microbiota. This is typically through the release of Neu5Ac and its subsequent metabolism. The gut symbiont *Ruminococcus gnavus* encodes an intramolecular *trans*-sialidase which releases 2,7-anhydro-Neu5Ac from mucin glycans instead of Neu5Ac. Here, we established protocols for the synthesis, detection and quantification of 2,7-anhydro-Neu5Ac, which has been a limiting factor in the study of this sialic acid derivative. We used this compound to demonstrate that the full 11 gene *nan* cluster the IT-sialidase is part of is dedicated to 2,7-anhydro-Neu5Ac metabolism. To investigate the metabolic pathway in more detail, the proteins encoded by the cluster were recombinantly expressed in *E. coli* and characterised using a range of biochemical and biophysical approaches. The predicted SAT2 transporter protein, *RgSBP*, was found to be uniquely specific to 2,7-anhydro-Neu5Ac, which once inside the cell is converted by a novel oxidoreductase enzyme, *RgNanOx*, to Neu5Ac. This is subsequently catabolised to ManNAc and pyruvate by the action of a sialic acid aldolase, *RgNanA*, which is structurally and biochemically similar to previously described sialic acid aldolases. Using gnotobiotic mice we confirmed that this cluster was important for *R. gnavus* colonisation and spatial location in the mucus layer *in vivo*, and our bioinformatic analysis showed that the cluster was found in very few species within the gut microbiota. This was supported by *in vitro* analysis using a simulated colon model showing that 2,7-anhydro-Neu5Ac supplementation promotes growth of *R. gnavus* in a complex microbial community. We also showed that the presence of *R. gnavus* can limit the outgrowth of *S. Typhimurium* post-antibiotic treatment by reducing the levels of free and bound Neu5Ac, by converting it to 2,7-anhydro-Neu5Ac which these pathogens cannot utilise. The data presented in this work unravel a unique metabolic pathway for the utilisation of 2,7-anhydro-Neu5Ac and highlight its potential in novel therapeutic strategies.

Acknowledgements

This PhD was funded by the BBSRC Norwich Research Park Doctoral Training Programme.

Thank you to Dr. Martin Fascione and Dr. Mark Webber for agreeing to read and discuss my work.

My biggest thanks go to my supervisor Prof. Nathalie Juge, who provided excellent scientific guidance, support and encouragement throughout the project. I would also like to thank all members of my supervisory committee, Prof. Rob Field, Dr. Gary Rowley, Dr. Emmanuelle Crost and Dr. Louise Tailford, for their support, suggestions and training throughout the project.

During the project I had the opportunity to interact with some excellent external collaborators: Prof. Jim Naismith (Research Complex at Harwell, Didcot, UK), Dr. Jesus Angulo (University of East Anglia, UK), Prof. Gavin Thomas (University of York, UK), Prof. Xi Chen (Department of chemistry, University of California, USA) and their lab groups as well as Dr. David Owen (Diamond Light Source, Didcot, UK), who all provided expert advice and analysis, and many aspects of the project would not have been possible without their assistance.

I would like to acknowledge the support of all IFR/QIB staff and students, past and present who trained me in various techniques or performed sample analysis. In particular, Dr. Jason Brunt for introducing me to the ClosTron technology, and Dr. Andrew Goldson and Arlaine Brion for their help with *in vivo* studies.

Within IFR/QIB a special thank you goes to each and every member of the Juge group who made the experience so enjoyable as well as providing excellent help and suggestions. I will be forever grateful to Dr. Marie Monestier for her dedication to producing our favourite sugar. I would also like to specifically mention Dr. Dimitris Latousakis for many helpful discussions and HPLC/MS expertise, Dr. Sandra Tribolo for introducing me to the joys of colon model and Dr. Laura Vaux for the amazing images from FISH analysis.

The PhD experience would not have been the same without the support of the wonderful friends I made along the way. Especially the coffee crew, Bhav, Mikhaela and Jess who were always there when I needed them / coffee. Bhav, in particular for her shameless brand of brutal honesty, and giving me a 'motivational' kick when required.

None of this journey would have been possible without the amazing support of my parents, Janine and Colin Bell. They have always encouraged me to do what I want to do and supported me throughout my life to the completion of my PhD, thank you for everything!!

Finally, a special thank you to Nic who has spent every day of this PhD journey with me. Thank you for putting up with me when I've been grumpy and cheering me up when I've been down, I truly appreciate everything you do for me and the fun times we spend together. And last but not least, Hammy, who can make us both smile no matter the situation.

Table of Contents

Abstract	2
Acknowledgements.....	3
Table of Contents	5
List of Abbreviations	8
List of Figures	11
List of Tables	14
List of Appendices	14
Chapter 1 Introduction	15
1.1 Structure and function of the gastrointestinal tract.....	16
1.2 The mucus layer.....	20
1.2.1 Mucin biosynthesis and secretion	22
1.2.2 Mucin glycosylation.....	25
1.3 Gut microbiota.....	27
1.3.1 Composition of the gut microbiota.....	29
1.3.2 Symbiosis with the gut microbiota.....	32
1.3.3 The gut microbiota in disease	34
1.3.4 <i>Ruminococcus gnavus</i> in health and disease.....	35
1.3.5 Modulation of the gut microbiota	36
1.4 Bacteria – Mucus interactions.....	38
1.4.1 Gut Commensal Bacteria – Mucus interactions.....	38
1.4.2 Enteric Pathogens – Mucus interactions	41
1.5 Sialic acid.....	43
1.5.1 Mucus associated sialic acid.....	43
1.5.1 Sialic acid metabolism in bacteria	44
1.6 Aims and objectives	49
Chapter 2 Materials and Methods	50
2.1 Bacterial strains and media	51
2.1.1 Media.....	51
2.1.2 Bacterial strains and culture conditions.....	52
2.2 Bioinformatics.....	53
2.2.1 Identification of <i>nan</i> genes in <i>Salmonella</i> and <i>C. difficile</i>	53
2.2.2 Sequence Similarity Networks	53
2.2.3 Multigene cluster BLAST analysis.....	54
2.3 Molecular Biology	54
2.3.1 Genomic DNA (gDNA) extractions	54
2.3.2 RNA extraction.....	54
2.3.3 cDNA synthesis	55
2.3.4 Polymerase chain reaction (PCR).....	55

2.3.5 Gene expression analysis by qRT-PCR	56
2.3.6 Bacterial quantification by qPCR.....	56
2.3.7 Sequencing.....	57
2.3.8 Heterologous expression and protein purification.....	59
2.3.9 X-ray crystallography	64
2.4 2,7-anhydro-Neu5Ac synthesis.....	65
2.5 2,7-anhydro-Neu5Ac detection and quantification	66
2.5.1 Charged Aerosol Detector (CAD).....	66
2.5.2 Mass Spectrometry	66
2.6 Binding assays	67
2.6.1 Differential Scanning Fluorimetry (DSF).....	67
2.6.2 Fluorescence spectroscopy	67
2.6.3 Isothermal titration calorimetry (ITC)	68
2.6.4 Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR) .	68
2.7 Enzyme activity assays	69
2.7.1 Thin Layer Chromatography (TLC)	69
2.7.2 HPLC analysis of 2-AB labelled products.....	69
2.7.3 HPLC analysis of Diamino-4,5-methylenedioxybenzene dihydrochloride (DMB) labelled reaction products.....	70
2.7.4 Kinetic enzymatic assays	70
2.7.5 NMR monitoring of the RUMGNA_02695 (<i>RgNanOx</i>) reaction	71
2.8 Clostron mutagenesis	71
2.9 Batch cultures	72
2.10 Mouse studies	72
2.10.1 Fluorescent in situ hybridisation (FISH).....	73
2.10.2 Glycan analysis of MUC2.....	74
Chapter 3 2,7-anhydro-Neu5Ac synthesis and detection methods.....	75
3.1 Introduction:	76
3.2 Results.....	79
3.2.1 Enzymatic synthesis of 2,7-anhydro-sialic acid derivatives:	79
3.2.2 Detection and quantification of 2,7-anhydro-Neu5Ac	84
3.3 Discussion.....	89
Chapter 4 Metabolism of 2,7-anhydro-Neu5Ac by <i>R. gnavus</i>	91
4.1 Introduction	92
4.2 Results	95
4.2.1 Growth of <i>R. gnavus</i> on sialylated substrates	95
4.2.2 The <i>R. gnavus nan</i> cluster is transcribed as one operon.....	95
4.2.3 <i>R. gnavus nan</i> operon is induced during growth on 2,7-anhydro-Neu5Ac or 3'SL	97
4.2.4 Bioinformatic analysis of the <i>R. gnavus nan</i> operon.....	99

4.2.5 The predicted transporter of <i>R. gnavus nan</i> operon is specific to 2,7-anhydro-Neu5Ac.....	106
4.2.6 Structural studies of RgSBP.....	110
4.2.7 Neu5Ac is the preferential substrate for RgNanA.....	111
4.2.8 RUMGNA_02695 catalyses the interconversion of 2,7-anhydro-Neu5Ac to Neu5Ac.....	117
4.2.9 Remaining 'uncharacterised' proteins from the <i>R. gnavus nan</i> cluster	124
4.3 Discussion.....	126
Chapter 5 Impact of the <i>R. gnavus nan</i> cluster in gut symbionts.....	130
5.1 Introduction	131
5.2 Results	132
5.2.1 2,7-anhydro-Neu5Ac preferentially promotes <i>R. gnavus</i> growth in an <i>in vitro</i> colon model	132
5.2.1 Importance of the <i>R. gnavus nan</i> cluster <i>in vitro</i> and <i>in vivo</i>	136
5.2.2 <i>In vivo</i> colonisation of germ-free mice by <i>R. gnavus</i> wild-type and <i>nan</i> mutants	140
5.3 Discussion	146
Chapter 6 Impact of the <i>R. gnavus nan</i> cluster on enteric pathogens	149
6.1 Introduction	150
6.2 Results	151
6.2.1 Bioinformatics of <i>S. Typhimurium</i> and <i>C. difficile</i> sialic acid utilisation pathway	151
6.2.2 <i>In vitro</i> mono-cultures of enteric pathogens	152
6.2.3 <i>In vitro</i> co-cultures of <i>S. Typhimurium</i> with <i>R. gnavus</i> or <i>B. thetaiotaomicron</i>	154
6.2.3 Impact of <i>R. gnavus</i> on the colonisation of <i>S. Typhimurium in vivo</i>	160
6.3 Discussion.....	166
Chapter 7 Conclusions and perspectives.....	168
References	176
Appendices.....	201

List of Abbreviations

Abbreviation	Full name
2-AB	2-aminobenzamide
2'FL	α 1,2 fucosyllactose
2-OMeNeu	N-acetyl-2-O-methyl- α -D-neuraminic acid
3'FL	α 1,3 fucosyllactose
3'SL	α 2,3 sialyllactose
6'SL	α 2,6 sialyllactose
AIM	Auto induction media
ATP	Adenosine triphosphate
BDM	Bacteroides defined media
BHI-YH	Brain heart infusion with yeast extract and hemin
C1GALT	<i>N</i> -acetylgalactosamine 3- β -galactosyltransferase
C3GnT	Core 3 beta1,3- <i>N</i> -acetylglucosaminyltransferase
CAD	Charged aerosol detector
CAZy	Carbohydrate active enzymes database
CBM	Carbohydrate binding module
cDNA	Complementary DNA
CFU	Colony forming unit
CMP	Cytidine monophosphate
Ct (qPCR)	Cycle threshold (qPCR)
CTP	Cytidine triphosphate
DEEP STD-NMR	Differential epitope saturation transfer difference nuclear magnetic resonance
DMB	Diamino-4,5-methylenedioxybenzene dihydrochloride
DSF	Differential scanning fluorimetry
EFI	Enzyme function initiative
EFI-ESN	Enzyme function initiative enzyme similarity network
ESI-MS	Electrospray ionisation mass spectrometry
FAD	Flavin adenine dinucleotide
FISH	Fluorescent <i>in situ</i> hybridisation
FMT	Faecal microbiota transplant
FUT	Fucosyltransferase
GalNAc	<i>N</i> -acetylgalactosamine
gDNA	Genomic DNA
GH(s)	Glycoside hydrolase(s)
GH33	CAZY family of sialidases
GI	Gastrointestinal
GlcNAc	<i>N</i> -acetylglucosamine
GnT	GalNAc transferase
HILIC	Hydrophobic interaction liquid chromatography
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation
IBD	Inflammatory bowel disease

IBS	Inflammatory bowel syndrome
IEX	Ion exchange chromatography
IMAC	Immobilised metal affinity chromatography
ITC	Isothermal titration calorimetry
IT-sialidase	Intramolecular <i>trans</i> -sialidase
K _{cat}	Catalytic constant
K _d	Dissociation constant
KDN	2-keto-3-deoxynononic-acid
K _M	Michaelis constant
LB	Luria Bertani
LC-MS	Liquid chromatography–mass spectrometry
LDH	Lactate dehydrogenase
LPS	Lipopolysaccharide
MACs	Microbiota-associated carbohydrates
ManNAc	<i>N</i> -acetylmannosamine
MEME	Membrane enclosed multienzyme
MRM	Multiple reaction monitoring
MUB	Mucus binding protein
MU-Neu5Ac	2'-(4-methylumbelliferyl)- α -D-N-acetylneuraminic acid
NAD ⁺ /NADH	Nicotinamide adenine dinucleotide
NagA	<i>N</i> -acetylglucosamine-6-phosphate deacetylase
NagB	<i>N</i> -acetylglucosamine-6-phosphate deaminase
nan cluster	Neuraminic acid (sialic acid) metabolism gene cluster
NanA	Sialic acid aldolase/lyase
NanE	<i>N</i> -acetylmannosamine-6-P-epimerase
NanH	Sialic acid hydrolase
NanK	ManNAc kinase
NanL	Sialic acid aldolase/lyase
NanS	Sialate-O-acetyltransferase
Neu	Neuraminic acid (sialic acid)
Neu5Ac	<i>N</i> -acetylneuraminic acid
Neu5Gc	<i>N</i> -glycolylneuraminic acid
NMR	Nuclear magnetic resonance
OPME	One-pot multienzyme
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
pPGM	Purified pig gastric mucins
qPCR	Quantitative PCR
qRT-PCR	Quantitative Reverse Transcription PCR
RgNanA	<i>Ruminococcus gnavus</i> sialic acid aldolase
RgnanH	<i>Ruminococcus gnavus</i> intramolecular <i>trans</i> -sialidase
RgNanOx	<i>Ruminococcus gnavus</i> sialic acid oxidoreductase
RgSBP	<i>Ruminococcus gnavus</i> sialic acid soluble binding protein
RokA	ManNAc kinase

-RT	No reverse transcriptase control
S. Typhimurium	<i>Salmonella enterica</i> serovar Typhimurium
SAT	Sialic acid transporter
SBP	Solute binding protein
SCFA	Short chain fatty acids
SDR	Short chain dehydrogenase/reductase
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEA	Sea-urchin-sperm protein-enterokinase-agrin
SSN	Sequence similarity network
STD-NMR	Saturation transfer difference nuclear magnetic resonance
sus	Starch utilisation system
TEMPOL	4-hydroxy-1-oxyl-2,2,6,6-tetramethylpiperidine
TLC	Thin layer chromatography
UHPLC	Ultra-high-performance liquid chromatography
VNTR	Variable number tandem repeat
YCFA	Yeast extract Casitone and Fatty Acid

List of Figures

Figure 1: Overview of the GI tract	17
Figure 2: Diagram showing the cross sectional layered structure of the GI tract	18
Figure 3: Diagram of the crypt-villus axis in the small intestine	19
Figure 4: Schematic representation of the mucus layers of the rat GI tract	22
Figure 5: MUC2 and MUC1 structures, synthesis and secretory pathways	23
Figure 6: Organisation of secreted mucins.....	24
Figure 7: Diversity of mucin glycan structures.....	27
Figure 8: Changes in physiology along the GI tract of Humans, and representation of changes in phyla.....	29
Figure 9: Structure of the microbiota across species.....	31
Figure 10: Representation of how the host species can influence the abundance of different bacterial phyla.....	32
Figure 11: Structure of the common sialic acid derivatives.....	44
Figure 12: Schematic representation of the five characterised bacterial sialic acid transporters	45
Figure 13: Pathways of sialic acid utilisation in <i>E. coli</i> (left; <i>nanA/K/E</i>) and <i>B. fragilis</i> (right; <i>nanL/E/T</i>).....	47
Figure 14: Products from the reactions of hydrolytic, <i>trans</i> - and IT <i>trans</i> -sialidases. ...	48
Figure 15: DMB labelling of sialic acid derivatives.....	78
Figure 16: Synthesis of 2,7-anhydro-Neu5Ac.....	80
Figure 17: Synthesis of 2,7-anhydro-Neu5Ac with addition of a sialic acid aldolase....	81
Figure 18: Schematic representation of the One-Pot Multienzyme (OPME) 2,7-anhydro-Neu5Ac synthesis.....	82
Figure 19: Assessment of Neu5Ac contamination in samples of 2,7-anhydro-Neu5Ac	83
Figure 20: HPLC purification of 2,7-anhydro-Neu5Ac	84
Figure 21: HPLC analysis of sialic acids using a Luna 5 μ HILIC-OH column (Phenomenex) and detection by CAD.....	85
Figure 22: Separation of sialic acids using LC-MS.....	87
Figure 23: Peak area of internal standard (2-OMeNeu) added to 30 caecal content samples.....	88
Figure 24: Growth curves of <i>R. gnavus</i> on sialic acids.....	95
Figure 25: Composition of the <i>R. gnavus nan</i> cluster.....	97
Figure 26: Expression of the <i>R. gnavus nan</i> cluster on 2,7-anhydro-Neu5Ac and 3'SL	98

Figure 27: Diagram depicting the possible scenarios for <i>R. gnavus</i> metabolism of 2,7-anhydro-Neu5Ac.....	99
Figure 28: Sequence Similarity Networks (SSN) of predicted proteins in the <i>R. gnavus nan</i> cluster	102
Figure 29: Gene organization in predicted homologs of the <i>R. gnavus nan</i> cluster ...	106
Figure 30: SDS-PAGE analysis of recombinant RUMGNA_02698 (<i>RgSBP</i>) purification	107
Figure 31: Differential scanning fluorimetry of <i>RgSBP</i> with Neu5Ac or 2,7-anhydro-Neu5Ac.....	108
Figure 32: Steady-state fluorescence analysis of ligand binding to <i>RgSBP</i>	109
Figure 33: ITC isotherms of <i>RgSBP</i> binding to sialic acid derivatives.....	110
Figure 34: STD NMR analysis of the interaction between <i>RgSBP</i> and 2,7-anhydro-Neu5Ac.....	111
Figure 35: Sequence alignments of <i>RgNanA</i> with sialic acid aldolases from known bacterial Neu5Ac utilisers	112
Figure 36: SDS-PAGE analysis of recombinant RUMGNA_02692 (<i>RgNanA</i>) purification	113
Figure 37: TLC analysis of sialic acid aldolase reaction products.....	113
Figure 38: HPLC analysis of 2-AB labelled sialic acid aldolase reactions.....	114
Figure 39: Enzymatic assay of <i>RgNanA</i> with Neu5Ac as substrate.....	115
Figure 40: Comparison of <i>RgNanA</i> crystal structure with characterised bacterial sialic acid aldolases.....	116
Figure 41: SDS-PAGE analysis of recombinant RUMGNA_02695 purification.....	117
Figure 42: HPLC analysis of recombinant RUMGNA_02695 reactions with 2,7-anhydro-Neu5Ac.....	118
Figure 43: Cofactor requirement and enzyme kinetics of RUMGNA_02695	120
Figure 44: NMR characterisation of the RUMGNA_02695 reaction.....	121
Figure 45: Predicted mechanism for the reversible conversion of 2,7-anhydro-Neu5Ac to Neu5Ac by RUMGNA_02695.....	122
Figure 46: Crystal structure of recombinant RUMGNA_02695	123
Figure 47: SDS-PAGE analysis of recombinant RUMGNA_02701 expression.....	125
Figure 48: SDS-PAGE analysis of recombinant RUMGNA_02700 expression.....	125
Figure 49: Proposed pathway of 2,7-anhydro-Neu5Ac metabolism by <i>R. gnavus</i>	128
Figure 50: Determination of <i>R. gnavus</i> growth in an <i>in vitro</i> colon model	133
Figure 51: Analysis of <i>R. gnavus</i> growth with vitamin B ₁₂ and/or histidine supplementation	135

Figure 52: Taxonomic profiles of batch culture experiments supplemented with sialic acids.....	136
Figure 53: Outline of ClosTron mutagenesis strategy	137
Figure 54: PCR analysis of <i>R. gnavus</i> mutants.....	138
Figure 55: In vitro growth assay of <i>R. gnavus nan</i> mutants on sialic acid and sialylated substrates.....	139
Figure 56: Gene expression analysis of wild-type and mutant <i>nan</i> operon.....	140
Figure 57: Colonisation of germ-free C57BL/6J mice with <i>R. gnavus</i> ATCC 29149 wild-type or <i>nan</i> mutant strains	141
Figure 58: Bacterial and host transcriptomic analysis of mice colonised by <i>R. gnavus</i> or <i>nan</i> mutant.....	142
Figure 59: In vitro growth assay of <i>R. gnavus nan</i> mutants on fucose and fucosylated substrates.....	144
Figure 60: Metabolite changes in germ-free mice colonised with <i>R. gnavus</i> wild-type or <i>nan</i> mutant.....	145
Figure 61: Domain structure of <i>R. gnavus</i> IT-sialidase.....	151
Figure 62: Growth curves of <i>S. Typhimurium</i> under aerobic and anaerobic conditions	153
Figure 63: Growth of <i>S. Typhimurium</i> and <i>C. difficile</i> on sialic acids	154
Figure 64: Growth curves of <i>S. Typhimurium</i> , <i>C. difficile</i> , <i>R. gnavus</i> and <i>B. thetaiotaomicron</i> on minimal media	155
Figure 65: Growth curves of <i>S. Typhimurium</i> co-cultures with <i>R. gnavus</i> or <i>B. thetaiotaomicron</i>	156
Figure 66: Growth curves of <i>R. gnavus</i> or <i>B. thetaiotaomicron</i> on 3'SL	157
Figure 67: Growth of <i>S. Typhimurium</i> in co-culture with <i>R. gnavus</i>	158
Figure 68: Co-cultures of <i>S. Typhimurium</i> in co-culture with <i>R. gnavus</i> or <i>B. thetaiotaomicron</i>	159
Figure 69: Growth curves of <i>S. Typhimurium</i> on Neu5Ac released from exogenous sialidase activity.....	160
Figure 70: Design of the in vivo study to assess the impact of <i>R. gnavus</i> colonisation on <i>S. Typhimurium</i> infection	161
Figure 71: <i>R. gnavus</i> colonisation of mice during the <i>S. Typhimurium</i> challenge study	162
Figure 72: Mouse welfare scores for in vivo <i>S. Typhimurium</i> challenge study.....	163
Figure 73: Impact of <i>R. gnavus</i> on <i>S. Typhimurium</i> infection <i>in vivo</i>	164
Figure 74: Taxonomic profiles from the <i>in vivo</i> study of <i>S. Typhimurium</i> colonisation	165

List of Tables

Table 1: Mucin expression in main tissues of the GI tract produced using information from Corfield, 2015.	21
Table 2: List of primers used in this work.	62
Table 3: Multiple reaction monitoring (MRM) transitions used in this work.	66
Table 4: Functions or predicted functions of <i>R. gnavus</i> nan operon proteins.	100
Table 5: Identification of species co-occurring with SSN analysis of <i>R. gnavus</i> proteins, grey shading denotes that the given species encodes a co-occurring protein.	103
Table 6: Kinetic parameters of characterised NanA proteins using figures taken from Wang <i>et al.</i> , 2018.....	115
Table 7: Identification of <i>R. gnavus</i> genes differentially expressed in nan mutant during colonisation, bold denotes genes in the <i>R. gnavus nan</i> cluster.	143

List of Appendices

Appendix 1: Map of pEHISTEV plasmid.....	201
Appendix 2: Map of pOPINF plasmid.	201
Appendix 3: Legend for taxonomic classifications made in Figure 52.....	202
Appendix 4: Legend for taxonomic classifications made in Figure 74.....	204
Appendix 5: Glycosylation pattern of MUC2 from mice in the <i>S. Typhimurium</i> challenge study.....	206

Chapter 1

Introduction

1.1 Structure and function of the gastrointestinal tract

The gastrointestinal (GI) tract is divided into upper and lower anatomical regions, fulfilling different biological roles. The upper region is comprised of the oesophagus, stomach and duodenum, whereas the lower region is subdivided into the small and large intestine. The duodenum links the upper sections of the GI tract to the lower regions comprising the jejunum and ileum in the small intestine, and the cecum, colon, and rectum in the large intestine (Figure 1). The major role of the GI tract is to facilitate the uptake of nutrients from foods and expel waste (Liao *et al.*, 2009). It also forms a critical barrier between the luminal space and underlying tissue to protect against pathogenic infection, while regulating immune response (Fasano and Shea-Donohue, 2005). The GI tract is the largest mucosal surface of the body covering some 400 m² and provides the greatest interface with the external environment. Pathogens and antigens derived from the diet are in constant contact with this surface making it a crucial site for mucosal immunity (Turner, 2009). The gut has a unique environment comprising of consistent temperature and large quantities of host-diet derived nutrient sources. This combination results in intense colonisation by commensal bacteria (microbiota), which must be limited to the luminal space by the epithelial layer and other defences (Peterson and Artis, 2014). The large influx of food and water from the external environment also brings with it a plethora of opportunistic and potentially pathogenic bacteria which must be expelled (Snoeck *et al.*, 2005). Due to the complex nature of the GI tract's functions, where nutrients need to be absorbed, antigens and pathogens excluded and microbiota tolerated, a complex membrane with selective permeability is required (Barker, 2014).

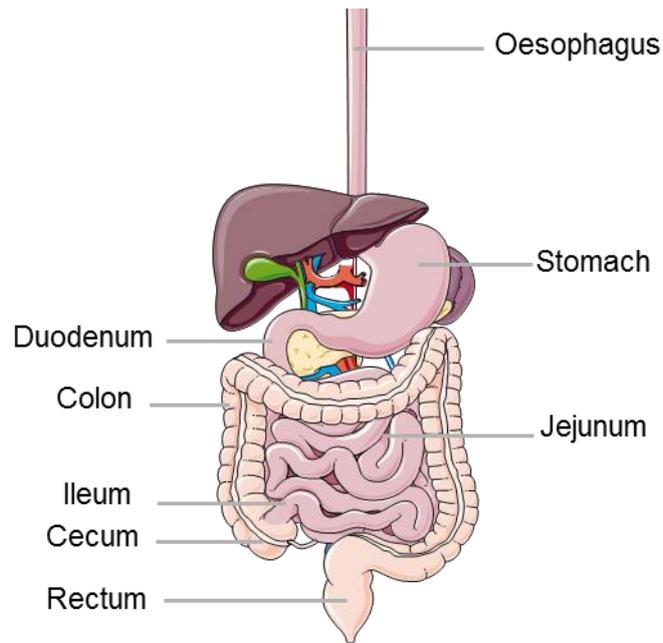


Figure 1: Overview of the GI tract

Overview of the GI tract, highlighting the key features of the Oesophagus, Stomach, Duodenum, Jejunum, Ileum, Cecum, Colon and Rectum. Adapted from Servier image bank. (<http://www.servier.com/Powerpoint-image-bank>) (*Servier Medical Art, 2016*).

In addition to longitudinal changes, the GI tract is organised into cross sectional layers (Liao *et al.*, 2009). The outermost layer of the GI tract is the adventia layer or serosa (Figure 2). These layers contain connective tissue, blood vessels, lymphatic tissue and nerves. The serosa secretes lubricative fluids that reduce friction at the peritoneal cavity. The adventitia on the other hand has a large collagen content used to bind and hold organs together. The layer responsible for peristalsis and digestive transit of food along the gut is the muscularis propria (or externa) (Figure 2). This is a layer containing smooth muscle arranged as two distinct layers, the inner layer is arranged circularly and the outer layer is longitudinal. The action of these muscles, initiated by the myenteric plexus, causes rhythmical waves of contraction and relaxation to move food through the gut. Next is the submucosa which continues the network of blood vessels and lymphatic tissue to allow nutrient transfer away from the gut (Figure 2). Furthermore, it contains submucosal glands that secrete mucus supporting the role of the goblet cells (Linden *et al.*, 2008).

The innermost layer is the mucosa (Figure 2) which processes nutrients and compacts waste products (Sancho *et al.*, 2003). It consists of a highly specialised epithelial lining, supported by the lamina propria – a network of lymphatic and connective tissues –

capillaries to transport digested products, and the intestinal subepithelial myofibroblasts (Leedham *et al.*, 2005, Peterson and Artis, 2014, Powell *et al.*, 2011).

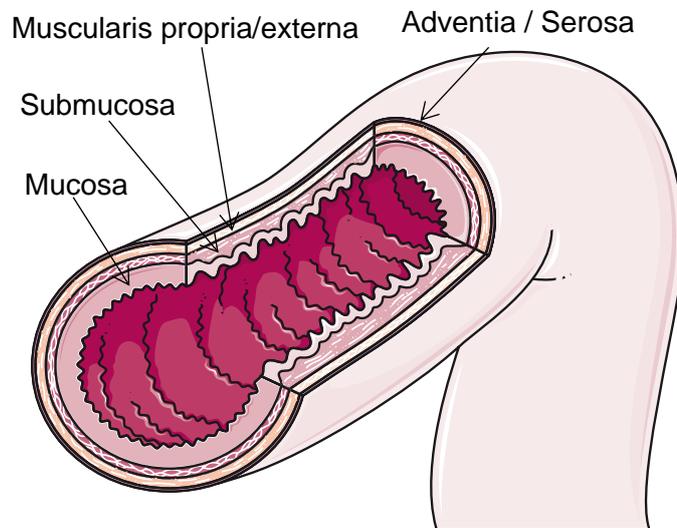


Figure 2: Diagram showing the cross sectional layered structure of the GI tract
Adapted from servier image bank.

The epithelial lining itself is arranged into crypts and villi and is held together by tight junctions to prevent leakage and maintain a continual barrier (Figure 3). At the base of the crypts lie intestinal stem cells which differentiate and divide to make all the cells and types of cell of the villi structures and epithelial layer (Cheng and Leblond, 1974). Cells become more specialised as they move along the crypt-villus axis. Above the stem cells are transit-amplifying cells (Leedham *et al.*, 2005), these cells continue to divide and amplify the number of progeny, however, their fate is to differentiate and not to remain as stem cells (Crosnier *et al.*, 2006). Dialogue with underlying mesenchyme cells provides a range of chemical signals that control cell fate and differentiation (Barker *et al.*, 2008, Powell *et al.*, 2011). To maintain size and organisation, epithelial cells are lost from the villus tip, in controlled and highly regulated manner, although the underlying mechanisms are still being elucidated (Patterson and Watson, 2017). The shedding, in combination with stem-cell derived renewal, gives rise to a barrier which is constantly being renewed (3-5 days in the small intestine and 5-7 days in the colon), and is crucial for maintaining barrier integrity and function (Barker, 2014).

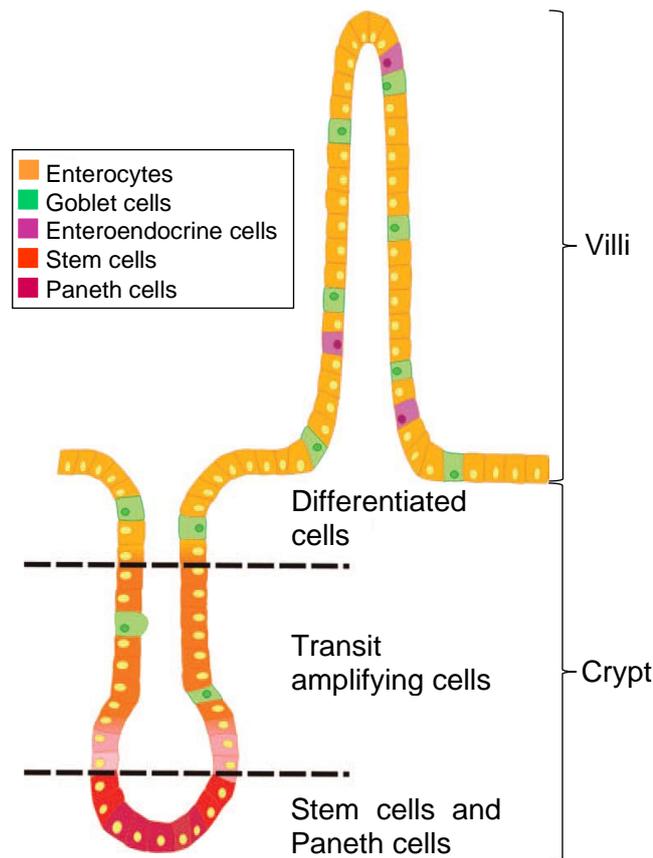


Figure 3: Diagram of the crypt-villus axis in the small intestine

The differentiation of stem cells and the types of cell found in the intestinal epithelial layer are shown. Adapted from (Dahan et al., 2007).

The majority of differentiated cells in the intestinal epithelium are absorptive enterocytes which are adapted for metabolic and digestive functions and secrete an array of hydrolytic enzymes (Barker *et al.*, 2008, Peterson and Artis, 2014). These cells are responsible for absorbing a range of dietary nutrients including salts, sugars, lipids, peptides, amino acids and water. Moreover, enterocytes are capable of taking up and processing antigens by lysozyme degradation, before presenting them directly to T-cells to elicit an immune response (Snoeck *et al.*, 2005). The other specialised cells of the epithelial layer include: enteroendocrine cells responsible for the release of hormones that control digestive function, mucin secreting goblet cells and Paneth cells which provide a physical and biochemical barrier to microbes (Barker *et al.*, 2008, Peterson and Artis, 2014). There are important structural changes from the small to large intestine most notably the lack of villi in the large intestine. Villi increase the surface area to allow maximal nutrient and water absorption in the small intestine. As nutrients and water are absorbed the left over material becomes drier, more compact and less nutritionally beneficial; therefore an increased surface area for absorption is not necessary in the

large intestine. The compact stool requires greater lubrication from mucin for a smooth transit so the number of goblet cells increases from the small to the large intestine, increasing the mucus thickness (Barker, 2014).

1.2 The mucus layer

The GI tract is lined by another protective barrier to luminal contents, a mucus layer formed from mucin secretions by goblet cells and submucosal glands (Corfield, 2015, Pelaseyed *et al.*, 2014). These secretions give protection against chemical and enzymatic attack of the epithelium as well as shielding from bacteria (Atuma *et al.*, 2001, Juge, 2012). The main structural components of mucus are mucins, the mucin family currently encompasses 21 proteins in humans, that are differentially expressed at different mucosal surfaces (Table 1), giving rise to diverse functions (Corfield, 2018). In the gut these mucins are divided into secretory gel forming and membrane bound (Bergstrom and Xia, 2013, Moniaux *et al.*, 2001). The membrane bound mucins form a glycocalyx around individual cells and provide key roles in signalling and attachment (Corfield, 2015, Johansson *et al.*, 2011a). The secretory mucins form the colonic mucus layer, with the thickness and properties of this layer vary along the GI tract. Colonic mucus is formed of a bi-layer, with the inner layer impermeable to bacteria and the outer layer heavily colonised by microbes. The colonised outer layer can reach up to 800 μm with the inner layer being less than 200 μm in rats (Etienne-Mesmin *et al.*, 2019, Juge, 2012). In the small intestine a thinner (150-300 μm) single layer of mucus fills the space between villi (Figure 4) (Atuma *et al.*, 2001). The stomach adopts a bi-layered mucus which protects the stomach from the harsh pH of the lumen and is colonised by very few, specialist bacteria (Juge, 2012, Atuma *et al.*, 2001). A recent study in rodents showed that the mucus layer also associated with faeces and trapped the bacteria to faecal content, suggesting that the luminal content plays a role in mucus organisation within the GI tract (Kamphuis *et al.*, 2017). Bacteria have been shown to be important for mucus organisation mediating changes in permeability, thickness and glycosylation (Arike *et al.*, 2017).

Table 1: Mucin expression in main tissues of the GI tract produced using information from Corfield, 2015.

Mucin	Main Tissue Expression
<i>secreted gel forming</i>	
MUC 2	Jejunum, Ileum, Colon
MUC 5AC	Stomach
MUC 6	Stomach, Ileum
<i>membrane bound</i>	
MUC 1	Duodenum, Ileum, Colon
MUC 3A/B	Small Intestine, Colon
MUC 4	Small Intestine, Colon
MUC 12	Colon
MUC 13	Small Intestine, Colon
MUC 15	Small Intestine, Colon
MUC 16	Respiratory tract
MUC 17	Stomach, Duodenum, Colon
MUC 20	Colon
MUC 21	Colon

As well as providing a physical barrier to bacteria, the mucus layer also hosts a number of non-mucin proteins, that help protect the intestinal barrier. These include immune regulatory molecules such as bactericidals, α -defensins and immunoglobulins (Etienne-Mesmin *et al.*, 2019, Johansson and Hansson, 2016). The mucus layer retains these important proteins in the mucus gel, allowing more time for the molecules to target bacteria, and slow their progression towards the epithelial barrier. Proteomic analysis of the mucus layer has also identified a range of proteins associated with the mucus layer whose functions are beginning to be elucidated. One example of this is ZG16. ZG16 knock-out mice showed a normal colonic mucus thickness, but bacteria were found associated with the inner mucus layer, indicating greater mucosal penetration. The recombinant ZG16 was found to bind to peptidoglycan in the bacterial cell wall and this mechanism is proposed to reduce bacterial motility (Bergstrom *et al.*, 2016).

A key feature of mucins is the variable number tandem repeat sequences (VNTR), rich in the amino acids serine, threonine and proline, which make up the majority of these proteins and vary in size between mucin types (Corfield, 2018). The threonine and serine hydroxyl groups allow for the attachment of glycan side chains which are comprised of a range of different monomeric sugars (see section 1.2.2). Mucin proteins are extremely large reaching thousands of amino acids in length. Secreted mucins are capable of disulphide bonding to each other through terminal D and CK domains to form oligomers and multimers (Perez-Vilar and Mabolo, 2007) (see section 1.2.1). The range and

variability of glycans added to the sheer size and multimeric forms of the mucin makes structural characterisation and analysis of chemical and physical properties challenging (Perez-Vilar and Mabolo, 2007).

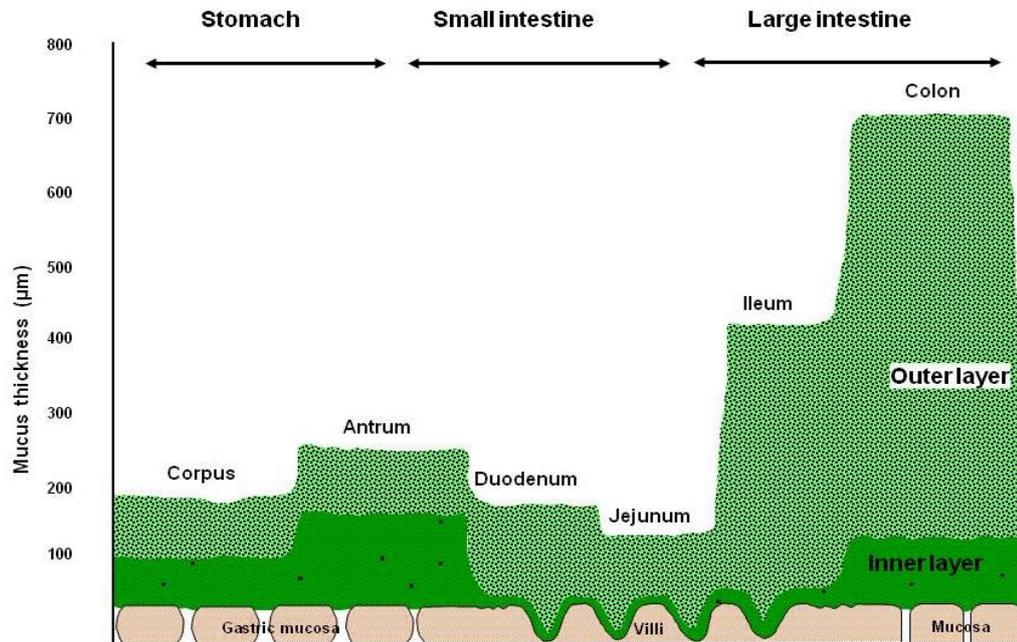


Figure 4: Schematic representation of the mucus layers of the rat GI tract taken from Juge, 2012.

1.2.1 Mucin biosynthesis and secretion

The gel forming secretory mucins are synthesised and secreted by goblet cells with MUC2 providing the structural basis of the mucus layer in the small and large intestine. MUC2 is comprised of heavily glycosylated VNTR domains and N and C-terminal von Willebrand D domains which are less glycosylated (Figure 5a). In contrast MUC1 contains a sea-urchin-sperm protein-enterokinase-agrin (SEA) domain, which is auto catalytic to form a non-covalent attachment to the transmembrane domain.

The dimerisation of secreted mucin proteins via the C-terminus has been shown to take place in the rough endoplasmic reticulum and is dependent on the *N*-glycosylation of the proteins (Asker *et al.*, 1998). The mucins then pass to the Golgi where further post-translational modification occurs; most notably *O*-glycosylation or mucin type glycosylation which can account for as much as 80 % of the glycoprotein mass (Arike and Hansson, 2016). These glycosylated mucin dimers then form oligomers and

multimers through disulphide bonds at the N-terminus before being stored in preparation for granule exocytosis (Figure 5b) (Johansson *et al.*, 2011a).

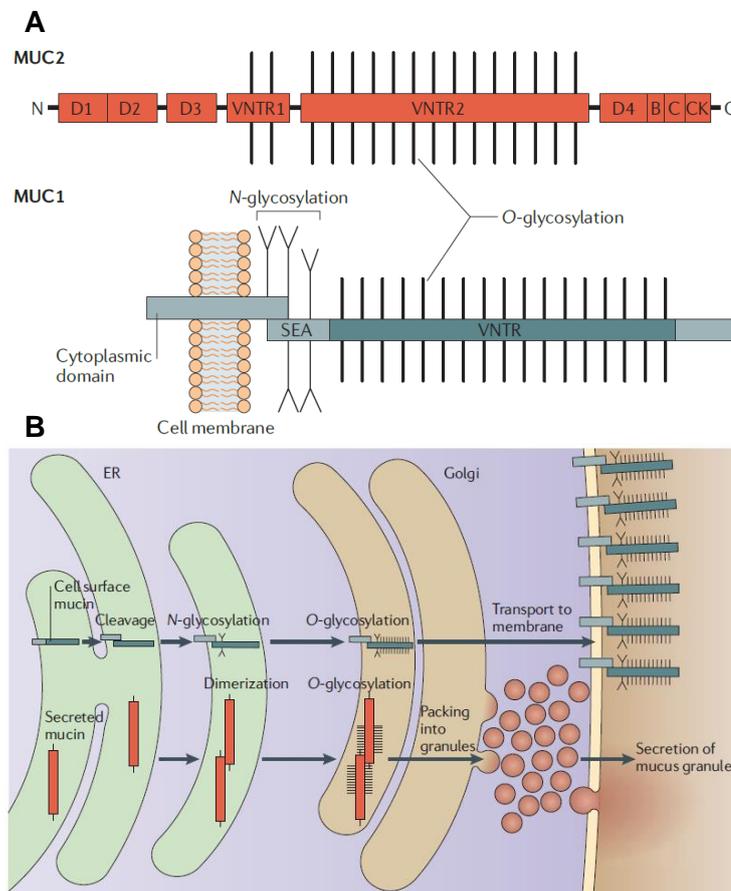


Figure 5: MUC2 and MUC1 structures, synthesis and secretory pathways

A) Structure of MUC2 and MUC1, the main secretory and non-secretory mucins in the small and large intestine, **B)** Diagrammatic representation of the synthesis, transport and secretion of non-secretory and secretory mucins. Taken from McGuckin *et al.*, 2011.

These interactions allow MUC2 to form (normally) six-sided ring structures (five and seven ring structures are also observed) which can be incorporated into a repeating pattern to form large unpacked sheets (Ambort *et al.*, 2012). The MUC2 polymers interact further by both covalent and non-covalent interactions to form an extremely well organised, stratified structure (Figure 6). In the colon, this inner layer remains as a well packed layer impenetrable to bacteria and resistant to a number a chemical processes (Johansson *et al.*, 2011a, Johansson *et al.*, 2008). The loose outer layer is similar in structure to the underlying inner layer, however, a series of proteolytic cleavage events dysregulate the close interactions between mucins causing a four-fold expansion of the mucus whilst not affecting the hexagonal polymeric network (Figure 6) (Johansson *et al.*, 2008). The addition of protease inhibitors to the firm inner mucus layer following the

removal of the outer layer leads to a reduced replenishment rate. Furthermore, this loosely adherent outer layer is shown to be present in germ-free mice, suggesting that its formation is due to the action of endogenous proteases (Johansson *et al.*, 2008).

The unfolding of secreted mucins to form the loose outer layer has been associated with pH levels and Ca^{2+} concentration (Ambort *et al.*, 2012). Increasing pH and decreasing Ca^{2+} was shown to weaken N-terminal interactions of MUC2, this allows water to enter and interact with the mucin glycans, forming the characteristic gel like nature of secreted mucins. In the small intestine this is related to bicarbonate release by the cystic fibrosis associated CFTR channel, whereas this process is not yet understood in the colon (Gustafsson *et al.*, 2012). The weakening of these interactions and expansion of the mucins then allows access to the merpin- β protease which releases the mucins for the epithelial layer (Arike and Hansson, 2016).

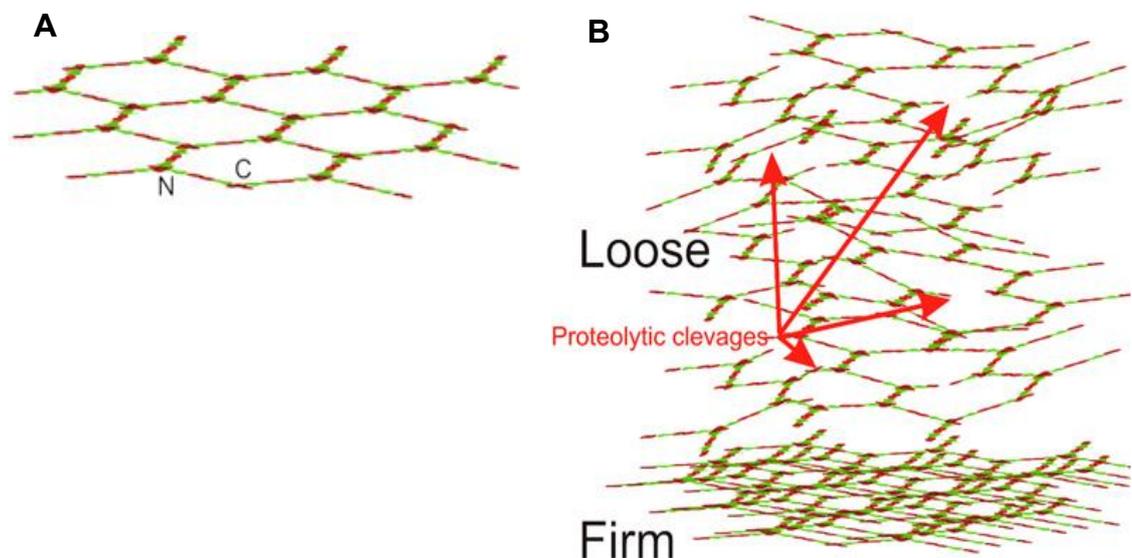


Figure 6: Organisation of secreted mucins

A) Hexagonal organisation of secreted mucins **B)** Stratified structure of the firmly attached colonic inner layer, which undergoes proteolytic cleavage to form the loosely adherent outer layer, adapted from Johansson *et al.*, 2008.

The intestinal glycocalyx is primarily made up of MUC3, MUC12 and MUC17, and like the secreted MUC2, these cell membrane mucins have large VNTR domains which are heavily glycosylated (Johansson *et al.*, 2011a). Membrane-bound mucins are synthesised in the endoplasmic reticulum and a cleavage event takes place splitting the protein into two subunits. Following *N*-glycosylation in the endoplasmic reticulum these mucins are transported into the golgi where, like the secreted mucins, *O*-glycosylation of the tandem repeats occurs before transport to the membrane (Figure 5) (McGuckin *et*

al., 2011). The size of these mucins and vast variability of the glycan structures means that the attached mucins can facilitate a wide range of interactions (Corfield, 2015). Membrane-bound mucins do not contribute to the inner or outer mucus layer themselves, however a study in mice encoding MUC1 showed an increased thickness of the mucus layer compared to MUC1^{-/-} mice (Malmberg *et al.*, 2006). This suggests that the make-up of the membrane bound glycocalyx can also impact upon the properties of the mucus layer.

1.2.2 Mucin glycosylation

It is proposed that the variability seen in mucin glycan chains can give rise to more coding capability than that of both DNA and proteins. This capacity is based on three elements of glycan structures, the first being the type of core structure of which there are 8 (Figure 7a) (Brockhausen *et al.*, 2009), the second being the flexible length (1 – 20 residues) and ability to form branches (Gunning *et al.*, 2013), and the third being the variability in the peripheral epitope which provides the greatest amount of variation with up to 16 distinct structures (Figure 7b) (Tailford *et al.*, 2015a).

The first stage of glycosylation of mucins is the attachment of *N*-acetylgalactosamine (GalNAc) to the hydroxyl group of serine and threonine residues by members of the GalNAc transferase (GnT) family forming the Tn antigen (Bennett *et al.*, 2012). Core transferases then act upon this *O*-GalNAc residue in a sequential fashion to form one of the 8 core structures, with structures 1 – 4 most commonly found in intestinal mucins (Tailford *et al.*, 2015a, Thomsson *et al.*, 2012). Following the formation of the Tn antigen, the core 1 T-antigen is synthesised by extension with galactose. A further addition of *N*-acetylglucosamine (GlcNAc) by the core 2 GnT enzyme produces the core 2 structure. Core 3 is produced from the T-antigen by addition of GlcNAc and addition of a second GlcNAc converts the core 3 to core 4 (Fu *et al.*, 2016).

The distribution of core structures has been shown to vary along the GI tract in humans and mice and is influenced by specific expression patterns of the core transferases (Arike *et al.*, 2017). Core 3 structures are spread throughout the intestinal and gastric mucus, with core 4 structures being present in the colonic mucus of humans. Core 1 and 2 glycans dominate in the gastric and duodenal mucus (Robbe *et al.*, 2004).

These core structures are then extended by the action of a range of other glycoside transferases to add galactose, GalNAc and/or GlcNAc residues. The regiospecific glycosylation is determined by the expression of glycoside transferases, which is

influenced by the presence of the gut microbiota. Compared to germ-free mice, conventionally raised animals show longer glycan chains, and this is proposed to give greater protection to the protein backbone from bacterial proteases (Arike *et al.*, 2017). In the same study, fewer enzymes responsible for O-glycan elongation were found in the small intestine, with shorter glycans structures also found in the small intestine of mice.

Following extension, glycans are terminated by one of the numerous epitopes which are, in many cases, fucosylated, sialylated or sulphated (Figure 7b). Like the core structures these terminal epitopes show species and regional specificity. For example, an increasing sialic acid gradient has been reported from the ileum to the colon in humans, with a reverse gradient for fucose (Robbe *et al.*, 2003). Interestingly, studies of the glycan patterns in mice showed opposite trends for both sialic acid and fucose making direct comparisons of murine model studies with humans more challenging (Larsson *et al.*, 2013). Furthermore, colonic mouse MUC2 contains more core 1 and core 2 structures as well as more neutral glycans than human colonic MUC2, demonstrating further that glycosylation patterns differ between species as well as along the GI tract within species (Thomsson *et al.*, 2012).

Abnormalities in O-glycosylation have been associated with diseases including inflammatory bowel disease (IBD), and mouse models deficient in glycosyltransferases have an to impaired barrier function and integrity, leading to the progression of disease (Etienne-Mesmin *et al.*, 2019). Mutations altering the function of *N*-acetylgalactosamine 3- β -galactosyltransferase 1 (C1GALT1) involved in the synthesis of core 1 and core 2 derived glycans has been associated with colorectal cancer, and mice deficient in C1GALT1 develop spontaneous colitis (Fu *et al.*, 2011). Core 3 β 1,3-*N*-acetylglucosaminyltransferase (C3GnT) mice deficient in core 3 production are more susceptible to protease degradation and have increased barrier permeability (An *et al.*, 2007). Double knockout core 1 and core 3 deficient mice develop a more severe colitis with earlier onset showing that core 1 and core 3 structures are crucial to barrier function and integrity (Bergstrom *et al.*, 2017). This relationship between glycosylation state and disease is not limited to large glycosylation changes resulting from mutations in core transferases but is also observed at the terminate glycan chains. For example, the lack of a functional copy of the fucosyltransferase gene *FUT2* leads to greater susceptibility to bacterial pathogens in both mice and humans, such as *Streptococcus pneumoniae*, *Neisseria meningitidis*, *Haemophilus influenzae* and *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*), while rates of viral infection were lowered (Taylor *et al.*, 2018).

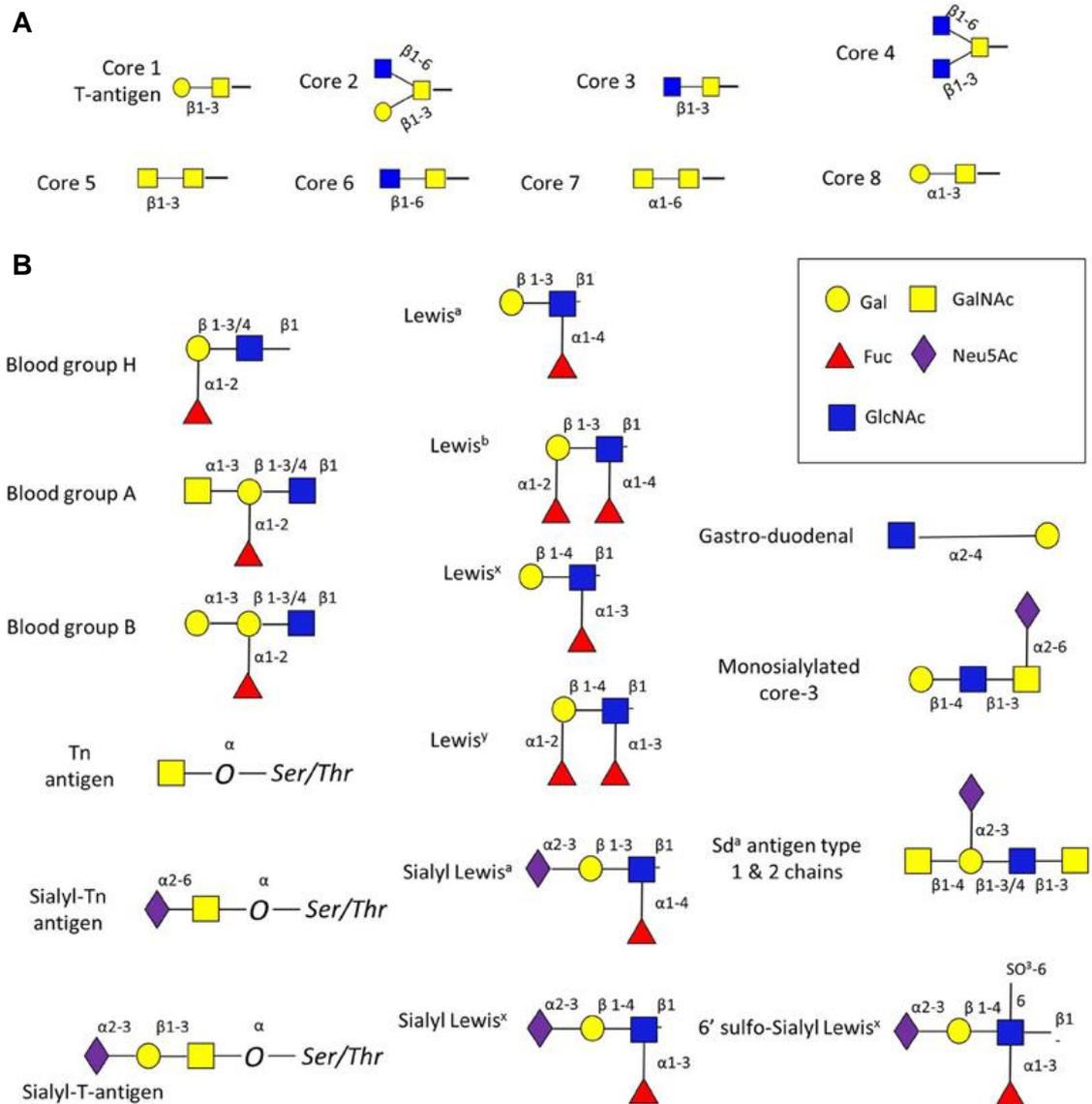


Figure 7: Diversity of mucin glycan structures

A) Representation of the 8 core structures of O-glycosylation B) Summary of the main epitope types found at the terminal position of O-glycan chains. Adapted from Tailford et al., 2015a.

1.3 Gut microbiota

Trillions of microbes reside within the lumen of the gut, a population that is relatively stable in adults, but become perturbed in diseases. This community is dominated by bacteria but also includes archaea, fungi and viruses. The density of microbes changes throughout the gut culminating in the highest abundance of $10^{11} - 10^{12}$ cells per gram of tissue in the large intestine (Kleerebezem and Vaughan, 2009). The composition of the gut microbiome is highly dynamic and varies depending on genetic and environmental factors (Schmidt *et al.*, 2018).

During the early years of life the microbiota can vary substantially, especially at important stages such as weaning where new complex foods such as cereals, fruits and vegetables are introduced (Koropatkin *et al.*, 2012). The microbiota matures towards an adult like microbiota by around the age of 3 years old and then continues to mature throughout ageing (Yatsunenکو *et al.*, 2012). The average adult microbiota has been shown to contain an estimated 10^{14} bacteria made up of between 150 and 250 bacterial strains from around 100 different species (Faith *et al.*, 2013). Data from the MetaHit (Li *et al.*, 2014) and Human Microbiome Project (The Human Microbiome Project *et al.*, 2012) identified more than 2000 species isolated from humans that classify into 12 phyla with Firmicutes and Bacteroidetes dominating in humans (Thursby and Juge, 2017, Lozupone *et al.*, 2012). Such detailed analysis of the microbiota has historically been restricted by technological limitations; however, the increasing power of next-generation sequencing strategies has opened the door to more highly detailed studies. Recently released databases have greatly increased the percentage of mapped reads from large sequencing projects, giving greater accuracy and confidence in the data analysed (Zhou *et al.*, 2018). Generally, genes and genetic function encoded by the microbes have been ignored in large sequencing projects but arguably have the greatest impact on the host. Species level identification does not account for the large genetic diversity found between different strains of the same species. Strain level identification and metagenome sequencing gives a better representation of the coding capacity, with the human microbiome project cataloguing nearly 10,000,000 genes (The Human Microbiome Project *et al.*, 2012). Recent research has developed strain level identification of bacteria in the microbiome further by refining and improving metagenomic pipelines to be able to demonstrate vertical transmission of strains from mother to infant (Asnicar *et al.*, 2017). This type of strain level identification has also led to the discovery of new microbial species that are represented at low abundance in the microbiome (Hildebrand *et al.*, 2019). These lower abundance species and strains often represent the greatest inter-individual differences as the core microbiome is relatively stable.

Due to functional redundancy, changes in microbial profile may not correspond to changes in protein or metabolite profiles (Thursby and Juge, 2017, Moya and Ferrer, 2016). Therefore, recent research has been geared toward profiling metabolites as differences here would have the greatest impact on host health and are controlled by the cooperation of gene products from both the host and the microbiome. Monitoring changes in concentration of well-studied metabolites such as Short Chain Fatty Acids

(SCFAs) and identification of new metabolites with unknown functions will allow for a deeper understanding into host-microbiome interactions (Rooks and Garrett, 2016).

1.3.1 Composition of the gut microbiota

The gut microbiota varies within an individual as well as between individuals and between species. To date most studies of the microbiome have focused on sequencing from faecal samples giving a snapshot of the microbial composition. However, variation is also seen throughout the GI tract with differences in the types of microbial species found from the ileum to colon. A whole host of physiological factors throughout the GI tract impact upon microbe colonisation, including the glycan composition and mucus thickness (see sections 1.2.1 / 1.2.2) (Donaldson *et al.*, 2016). These include an increase in pH and reduction in antimicrobials and oxygen content from proximal to distal regions (Figure 8). These changes are associated with a much increased bacterial load and a change in dominant phyla from *Lactobacillaceae* and *Enterobacteriaceae* to *Bacteroidaceae*, *Prevotellaceae*, *Rikenellaceae*, *Lachnospiraceae*, *Ruminococcaceae* (Donaldson *et al.*, 2016).

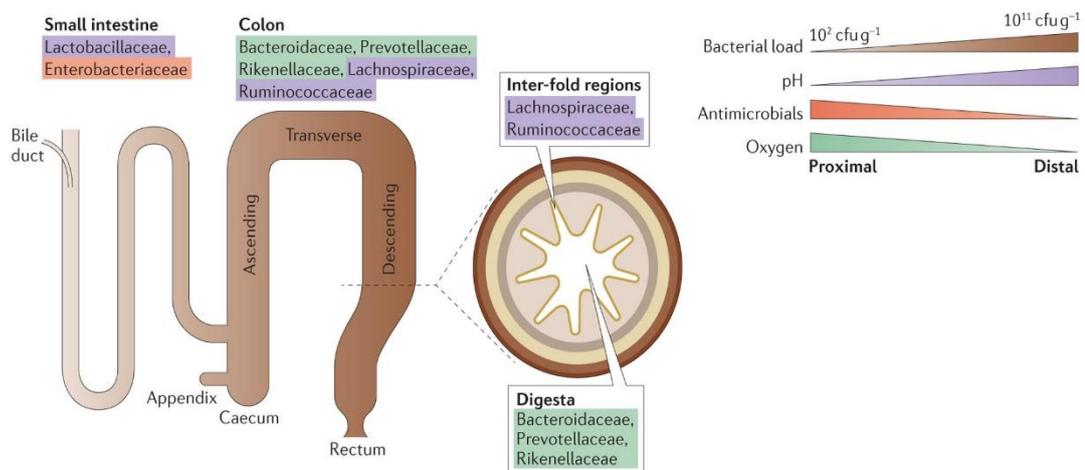


Figure 8: Changes in physiology along the GI tract of Humans, and representation of changes in phyla

Adapted from Donaldson *et al.*, 2016.

As well as differences in the longitudinal direction, cross sectional variations also exist, specifically when comparing microbes that closely associate with the mucus layer compared to those that reside in the lumen (Tropini *et al.*, 2017, Li *et al.*, 2015). This distinct microbial community has often been overlooked when studying the microbiome due to difficulties in obtaining samples in comparison to faecal samples. Recent studies, particularly those focussing on IBD patients where mucosal biopsies are taken, have

investigated the differences in the microbiota associating with mucosa compared to microbial species in the lumen, with greater variation in microbial composition found among mucus associating bacteria than luminal (Lavelle *et al.*, 2015). Methods to target this spatial microbial biogeography are being developed to allow 16S sequencing from small samples taken from specific sites (Sheth *et al.*, 2019).

Age plays a critical role in shaping the gut microbiome, especially at key developmental stages of early life such as weaning. Early colonisers such as *Bacteroides* and *Bifidobacterium*, have an essential role in determining the profile of the microbial community throughout life (Collado *et al.*, 2016, Backhed *et al.*, 2015). These are generally believed to be established during birth, though some studies suggest colonisation can occur *in utero* (Jimenez *et al.*, 2008), although this was recently disputed (de Goffau *et al.*, 2019). Delivery method is also a key determinant in establishing this early microbiome. Children delivered by C-section have been shown to have lower diversity and lack *Bacteroidetes* species, which dominate the healthy adult gut, compared to vaginally delivered babies (Jakobsson *et al.*, 2014). After birth, the gut microbiome shifts towards species that can utilise lactose and human milk oligosaccharides (HMOs) such as *Bifidobacterium* (Backhed *et al.*, 2015, Zivkovic *et al.*, 2011). Children that are not breast fed also lack microbial species that babies are exposed to through lactation (Hunt *et al.*, 2011).

Following weaning and the introduction of solid foods the microbiome undergoes another shift towards bacteria that can degrade dietary substrates (Koenig *et al.*, 2011). At this stage the microbiome is more stable, and changes are not as drastic as in earlier life stages. However, the dynamic community can still adapt to environmental factors, in particular, changes in diet. In the colon, strict anaerobes generally utilise host-glycans or fibre that cannot be digested by the host. Therefore, diets rich in fibre benefit the growth of *Bacteroidetes* species that thrive on these nutrient sources (Desai *et al.*, 2016). Bacteria in the small intestine thrive on simple short sugars that are more abundant in the small intestine (Zoetendal *et al.*, 2012). However, these short sugars rarely reach the large microbial population of the large intestine as they are mostly absorbed by the host.

Host genetics also play a role in shaping the microbial composition (Schmidt *et al.*, 2018). One of the important contributors is mucin glycosylation, which can provide nutrients and/or binding sites that can preferentially benefit specific microbes. This has been demonstrated in transgenic murine models. For example, deletion of *FUT2*, leads to microbiome composition changes in mice (Kashyap *et al.*, 2013). Host derived microRNAs can also control aspects of the gut microbiota, throughout co-evolution of the

host with the gut microbiota, it would appear that the host has developed these microRNA molecules to target specific genes of species in the microbiota controlling bacterial growth and replication (Liu *et al.*, 2016). In addition bacteria have developed a wide range of lectins to target specific glycan residues in mucins (Juge, 2012). In parallel the host has derived its own defence lectins against bacterial glycans which have immunomodulating properties (Matsushita *et al.*, 2012), providing support for a model of co-evolution between host and microbiome.

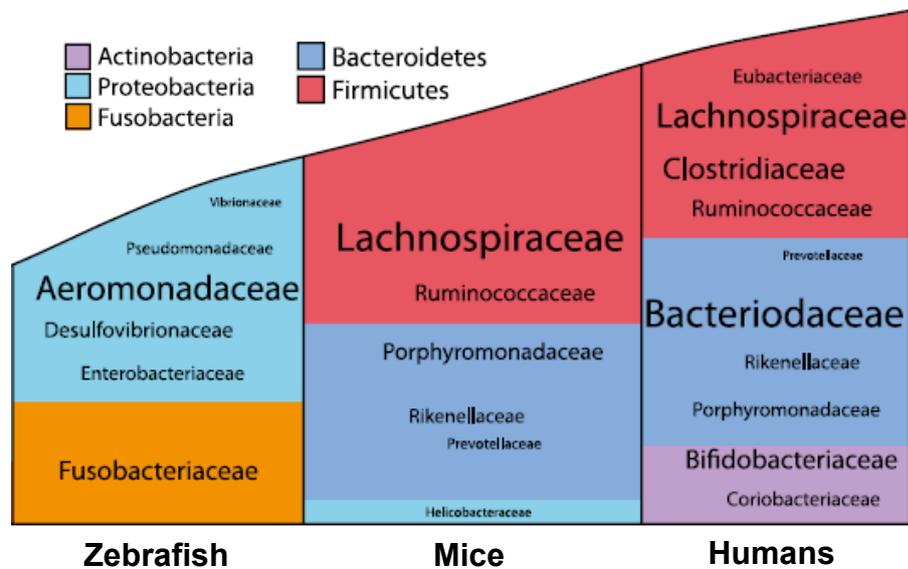


Figure 9: Structure of the microbiota across species

Phyla are represented by colour, and the relative abundance indicated by font size, adapted from Kostic *et al.*, 2013.

Evidence of host genetic effects on the establishment and maintenance of microbial communities can be found in a study where the microbiota from zebrafish was transplanted into germ-free mice and vice-versa. Overtime the relative abundances of each bacterial phyla shifted to be more representative of a typical microbiota found in a wild-type representative of the recipient (Figure 10) (McFall-Ngai, 2006, Rawls *et al.*, 2006), supporting the role of host genetics in microbiota establishment and maintenance. Furthermore, as discussed above, changes in mucin glycosylation has shown shifts in the microbial communities, highlighting the importance of interactions with specific glycan structures (Arike *et al.*, 2017, Sommer *et al.*, 2014, Thomsson *et al.*, 2012).

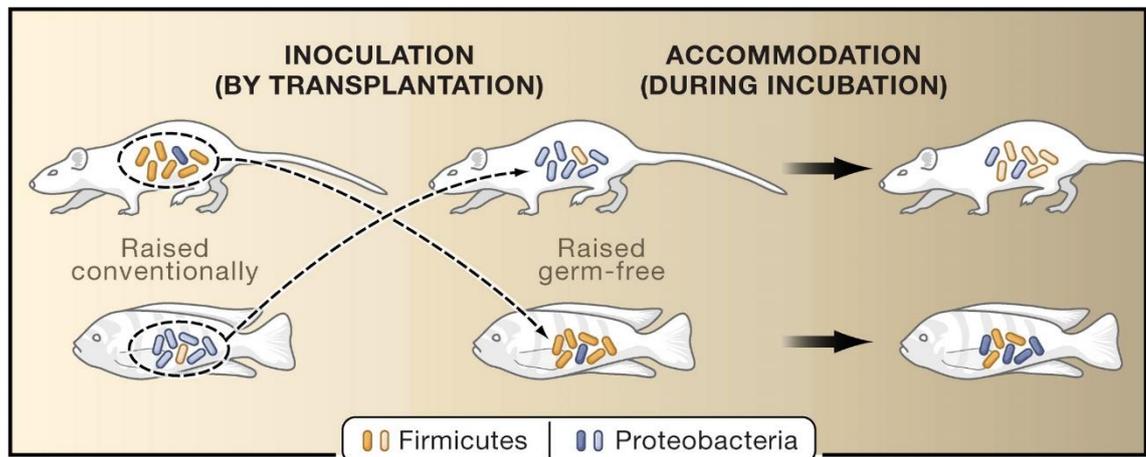


Figure 10: Representation of how the host species can influence the abundance of different bacterial phyla

When the microbiota from a mouse is transplanted into a germ-free zebrafish and vice-versa, the host reshapes the microbiota to represent that of a conventionally raised animal, taken from McFall-Ngai, 2006.

1.3.2 Symbiosis with the gut microbiota

The host can benefit from the microbes inhabiting the gut in a number of ways including the release of nutrients from complex carbohydrates, production of vitamins, pathogen exclusion and development of the immune system. The major advantage of symbiosis comes from the digestion of complex carbohydrates undigested by the host. It has been shown that germ-free rats require a 30 % greater calorie intake to maintain weight than wild-type rats (Wostmann *et al.*, 1983). Humans can consume a range of dietary compounds, many of which cannot be degraded owing to a lack of host glycoside hydrolases (GHs) required for their breakdown. In order to overcome this shortfall, humans utilise the symbiotic microbes within the distal intestine which encode for thousands of such enzymes opposed to the ~ 17 encoded for by humans (El Kaoutari *et al.*, 2013). The microbes residing in the small and large intestine encode 150 times more genes than the human genome (Zimmermann *et al.*, 2019). Tolerating a large community of microbes and their associated genetic information means that the host can rely on this community to fulfil a range of functions and produce numerous molecules that reduce the number of genes the host has to encode for. The dense microbial population here is well adapted to ferment starch and complex polysaccharides not degraded in the small intestine (Hooper *et al.*, 2002). One of the most dominant bacteria in the colon are *Bacteroides* which are known to degrade a wide range of plant polysaccharides and starch (Salyers *et al.*, 1977). *Bacteroides thetaiotaomicron* degrades starch by the coordinated action of enzymes encoded by the starch utilisation system (*sus*). The

degradation of these products also releases by-products including SCFAs, vitamins and nutrients that can be used by the host (Koropatkin *et al.*, 2012). SCFAs are particularly important, for example butyrate is the preferred energy source of colonocytes, and has also been suggested to suppress tumour growth and inflammation as well as impact satiety (Hamer *et al.*, 2008). Acetate and propionate are absorbed and transported to the liver where they are incorporated into lipid and glucose metabolism (den Besten *et al.*, 2013). Furthermore it has been reported that the presence of acetate may supplement butyrate production and inhibit some enteric pathogen colonisation (Koropatkin *et al.*, 2012).

In addition to the fermentation of undigested foods and production of SCFAs, the gut microbiome harbours a large repertoire of genes with the capacity to produce a range of bioactive molecules which could be beneficial to the host. This extra coding capacity provides humans with a source of essential vitamins, in particular, vitamin B₁₂, which is exclusively sought from bacteria, folate, riboflavin and vitamin K (LeBlanc *et al.*, 2013).

The gut microbiota can also significantly affect the development of the host immune system and has been extensively reviewed (e.g. Pickard *et al.*, 2017). Early experiments with germ-free mice showed that the gut microbiota is vital for optimal organisation and development of lymphoid tissue. These microbe-depleted animals show impaired healing of damaged epithelial cells, which could be reversed by gavaging the mice with lipopolysaccharide (LPS) from bacteria (Hooper *et al.*, 2012). Furthermore, the microbiota affects parts of the immune system away from the intestinal lamina propria, including the production of T-cells and other immune regulatory cells. This alteration of T-cell balance can have far-reaching consequences including autoimmune and allergic diseases (Pickard *et al.*, 2017, Hirota *et al.*, 2012).

Another key benefit of the gut microbiota is the defence against pathogens; the presence of microbial species can stimulate the production of mucus and impact on dynamics including the production of an impenetrable inner layer (Johansson *et al.*, 2008). Sentinel goblet cells react to Toll-like receptor signals, activated by the presence of microbes, to stimulate the production of MUC2 and also signal to adjacent goblet cells to produce MUC2, leading to bacterial exclusion from epithelial layer (Birchenough *et al.*, 2016). Furthermore, the degradation of the mucus layer by commensal bacteria leads to new mucus production which brings with it a host of new antimicrobial compounds against pathogenic bacteria (McGuckin *et al.*, 2011). Finally, the presence of commensal microbes means that many niche environments are already inhabited, and the microbes present there will compete with pathogens for binding sites or nutrients, a process

referred to as pathogen exclusion, which contributes to preventing pathogenic colonisation and infection (Canny and McCormick, 2008). For example, colonisation with a non-toxic *Bacteroides fragilis* strain led to the competitive exclusion of toxigenic *B. fragilis* in mice, along with a subsequent lowering of toxin exposure (Hecht *et al.*, 2016). In contrast, by pre-treating the mice with antibiotics, the microbial community is diminished opening up novel niches and nutrient sources, allowing the outgrowth of opportunistic pathogens. The normally non-infectious bacteria (patho-symbionts) then gain an advantage and can cause infection (Cornick *et al.*, 2015, Ng *et al.*, 2013).

1.3.3 The gut microbiota in disease

With the many essential roles the gut microbiota play towards host health, it is not surprising that changes in this microbial community can have major physiological consequences. Alterations of the microbial composition is described as a dysbiosis and can be caused by a number of factors including antibiotics, dietary changes or influxes of harmful microorganisms (Koropatkin *et al.*, 2012). The mechanisms underpinning dysbiosis have not been fully elucidated and it is difficult to determine whether changes in the bacterial population are causative or a result of the development of a disease (Schmidt *et al.*, 2018). Indeed, dysbiosis of the gut microbiota has been associated with a number of diseases including IBD, cancer, colitis, obesity and non-gut related diseases such as liver ailments, asthma or arthritis (Rooks and Garrett, 2016, Power *et al.*, 2014, Keeney *et al.*, 2014, Koropatkin *et al.*, 2012). More recently, microbiome changes have also been associated with brain disorders including depression, autism, Parkinson's and Alzheimer's disease, leading to an increased effort to understand the gut-brain axis (Sherwin *et al.*, 2018). In some diseases the changes are associated with an individual taxa, where in other diseases, the microbiome changes are more general, such as a reduction in microbial diversity (Schmidt *et al.*, 2018). Although causality remains to be demonstrated, these changes in the microbial composition have been proposed as a biomarker for several diseases.

A major cause of dysbiosis is due to the overuse of antibiotics commonly used to treat infections. Whilst treating potentially harmful bacteria, many antibiotics will have unwanted effects on commensal bacteria. Keeney *et al.* (2014) have summarised the effects of some common antibiotics and antibiotic cocktails on the microbiota (Keeney *et al.*, 2014). In general, antibiotic use is associated with a decrease in microbial diversity. Usually the community will recover within a matter of days post-treatment, but in cases of sustained antibiotic use, some microbes can fail to return or return at much reduced levels, resulting in dysbiosis which can lead to long-term health problems (Power *et al.*,

2014). Furthermore, it has been demonstrated that repeated antibiotic treatment can result the mono-dominance of species, that go undetected in the healthy microbiome (Hildebrand *et al.*, 2019). This impact on microbial communities and the associated health concerns, combined with the ever-increasing antibiotic resistance is leading to research directed at finding novel antimicrobial strategies.

Another way in which the microbiome is being indirectly linked to the progression and treatment of disease is the varied way in which the microbiome metabolises a broad spectrum of drugs. A recent study of 271 drugs showed that 176 were modified by at least one strain found in the human gut microbiota (Zimmermann *et al.*, 2019). The modifications were mostly found to be oxidation, reduction, deacetylation, hydrogenation, hydroxylation, acetylation and propionylation. A small number of previous reports have shown that these modifications by the microbiome can inactivate or lead to the toxification of drugs used to treat a range of conditions.

1.3.4 *Ruminococcus gnavus* in health and disease

Ruminococcus gnavus is a gram-positive anaerobic bacterium of the Firmicutes phylum. It is recognised as a mucin glycan foraging bacterium and is present in more than 90% of individuals (Qin *et al.*, 2010). The *Ruminococcus* genus is split across two bacterial families, the *Ruminococcaceae* and *Lachnospiraceae* with *R. gnavus* being part of the latter (Lawson and Finegold, 2015). The classification of *R. gnavus* has been recently challenged following the discovery of *Mediterraneibacter massiliensis*, which identified the key phenotypic and genotypic features of the *Lachnospiraceae* associated *Ruminococcus* species. As a result a new genus '*Mediterraneibacter*' incorporating *R. gnavus*, *R. faecis*, *R. lactaris* and *R. torques* has been proposed (Togo *et al.*, 2018).

R. gnavus is a common member of the human gut microbiota and an early coloniser present in both formula and breast-fed infants (Sagheddu *et al.*, 2016, Tannock *et al.*, 2013). *R. gnavus* mostly produces propionate which is an important SCFA in gluconeogenesis, appetite regulation and blood pressure control (Chambers *et al.*, 2018). *R. gnavus* has also been associated with a number of disease states, most commonly in IBD where *R. gnavus* is shown to have an increased abundance (Olbjorn *et al.*, 2019, Henke *et al.*, 2019, Hall *et al.*, 2017). However, it remains unclear whether this is causative or a result of changes in the mucosal environment. Indeed, mucin glycans have been shown to be altered in IBD with shorter and more sialylated glycans which could provide *R. gnavus* with an ecological advantage (Croft *et al.*, 2016, Theodoratou *et al.*, 2014, Larsson *et al.*, 2011). *R. gnavus* has also been shown to have

a causal relationship in ameliorating malnourishment symptoms and promoting weight gain in mice colonised with microbiota from undernourished children (Chu and Aagaard, 2016, Blanton *et al.*, 2016). Changes in *R. gnavus* abundance is increasingly being linked to non-gut related disease, including spondyloarthritis and in the gut-brain cross-talk where it is implicated in autism and stress response in high intensity sport (Breban *et al.*, 2017, Toh and Allen-Vercoe, 2015, Allen *et al.*, 2015).

1.3.5 Modulation of the gut microbiota

The gut microbiota continues to be linked to a number of diseases and to conferring a symbiotic relationship with the host. The mechanism behind some of these interactions, and the specific strains impacting these interactions are starting to be unravelled, leading to novel therapeutic targets (Cani, 2018, Langdon *et al.*, 2016). Targeted and untargeted approaches have both been used to move the gut microbiome to a more 'healthy' state.

As previously mentioned (see section 1.3.3), antibiotic treatments are routinely used to combat pathogens in the gut but oral antibiotics are also used to treat or prevent bacterial infection across body sites. These antibiotics are seldom specific to the pathogen and are usually broad-spectrum, especially in cases where the pathogen has not been identified. This means that gut symbionts are often adversely affected by antibiotic treatment leading to changes in microbial composition and these can be long-lasting changes. For example ciprofloxacin, which has been reported to have a lower effect on the microbial composition than other common antibiotics such as clindamycin or amoxicillin, was shown to affect the abundance of 30% of the gut taxa identified using a 16S approach in humans (Dethlefsen *et al.*, 2008). Furthermore, a follow up study revealed that the community in some participants had not reverted back to the original state (Dethlefsen and Relman, 2011). A recent review discussing a range of human and mouse studies, reported that specific antibiotics target different populations within the gut microbiome. Fluoroquinolones and β -lactams reduced diversity by 25% and increased the ratio of *Bacteroidetes* to Firmicutes, moxifloxacin decreased *Faecalibacterium* and *Bacteroides*, cefazolin increased *Bacteroides* levels as did amoxicillin, and vancomycin led to a bloom in the vancomycin resistant *Enterococcus*, as well as diminishing important immune factors (Dudek-Wicher *et al.*, 2018, Panda *et al.*, 2014, Perez-Cobas *et al.*, 2013, Ubeda and Pamer, 2012).

Probiotic treatments use live organisms to confer a health benefit to the host (Hill *et al.*, 2014). However, the usefulness of probiotics is limited by the lack of persistence of these organisms in the gut. To confer long-term health benefits probiotics need to be taken on

a regular basis to maintain a constant population (Kristensen *et al.*, 2016). Therefore, there is a need to develop next generation probiotics. These organisms will need to target the ecological niche, as well as encode factors that confer the health benefit to the host. This is underlined by strain-specific properties within gut bacteria species. Next generation probiotics are likely to come from novel species within the diversity of organisms found within our gut microbiome and move away from the microbes traditionally used as probiotics (O'Toole *et al.*, 2017). One such example is that of *Akkermansia muciniphila*. This bacterial strain has been shown to confer health benefits to the host, in particular a reduction in body weight and fat mass in mice supplemented with *A. muciniphila* (Zhao *et al.*, 2017). The probiotic effect of *A. muciniphila* was shown to be enhanced when the mucus layer in mice was thinned by a high fat diet, this is proposed to lead to an enhanced mucin production and restoration of mucus layer (Shin *et al.*, 2019). Furthermore, *A. muciniphila* was recently shown to reduce fat mass as well as improve a number of metabolic processes in an exploratory study using human volunteers (Depommier *et al.*, 2019).

Prebiotic treatments are designed to preferentially boost host organisms that confer a health benefit to the host (Gibson *et al.*, 2017). Numerous prebiotics have been reported to selectively promote organisms, however, the potential impact of these substrates on the whole microbiota is less well studied. Prebiotics could also be used in combination with probiotics in symbiotic approaches, to improve the chances of prolonged probiotic colonisation by providing a targeted competitive advantage.

Changes in diet can also be used as a means of regulating the microbial community (Flint *et al.*, 2017). An important component of the diet in shaping the gut microbiota is microbiota-associated carbohydrates (MACs). The primary source of MACs in the diet comes from plant fibres, which are lacking in western diet leading to the increase in prevalence of diseases such as asthma and IBD (Daïen *et al.*, 2017). This is because MACs are not digested by the host and pass through to the diverse community of bacteria present in the colon, acting as their primary nutrient source. Bacteria capable of degrading these MACs produce many products beneficial to the host, notably SCFAs. Depletion of MACs from the diet has been shown to lead to a thinner mucus layer and in turn a closer association of microbes to the epithelial layer, together with an increase in inflammatory markers (Desai *et al.*, 2016, Earle *et al.*, 2015). The direct impact of MACs in boosting colonic microbes was also shown to increase SCFAs which helped to ameliorate persistence of *Clostridium difficile* (Hryckowian *et al.*, 2018). Studies in humans have shown that altering the level of resistant starch can promote certain

microbes within days of changing diet (Walker *et al.*, 2011). Another study comparing 'animal based' and 'plant based' diets showed much more wide-ranging changes to the microbial composition (David *et al.*, 2014). Gaining a further understanding of these microbial community changes as a response to diet could give rise to targeted diets sustaining a 'healthy' microbial community (Shanahan *et al.*, 2017).

Faecal microbiota transplantation (FMT) is a more drastic microbiome-based strategy which aims to replace the entire microbial community of the patient. It has proven effective in treating recurring *Clostridium difficile* infection (van Nood *et al.*, 2013), although in other diseases the success rate has been much reduced. FMT is now being investigated as a potential treatment option for a range of conditions (Schmidt *et al.*, 2018). As of June 2019, clinicaltrials.gov list 300 registered clinical trials of FMT treatment for conditions including melanoma, depression and irritable bowel syndrome (IBS). FMT relies on donated samples from 'healthy' donors, but a more defined cocktail of microbial strains specifically tailored to patients and specific diseases may result in an improved outcome.

1.4 Bacteria – Mucus interactions

1.4.1 Gut Commensal Bacteria – Mucus interactions

The mucus layer itself should be viewed as a dynamic structure rather than a static one due to its ability to react and change based on external stimuli, in particular bacteria (McGuckin *et al.*, 2011). Despite the number of host defence mechanisms, members of the gut microbiota are able to permanently colonise the outer-layer of the gut mucus (Costello *et al.*, 2009). These bacteria do not penetrate the inner layer, apart from when mucus and/or barrier integrity is lost (Johansson *et al.*, 2008, McGuckin *et al.*, 2011). The mutually beneficial relationship with these commensal organisms is the culmination of appropriate barrier function that has resulted from long term co-evolution with the host (Neish, 2009).

At the ecological level, bacteria that colonise the gut thrive on the host glycans found in the mucus layer (Marcobal *et al.*, 2013). The ability to utilise host glycans has been studied in a few commensal bacteria. For example, *B. thetaiotaomicron* is able to utilise host glycans as well as host dietary components (Sonnenburg *et al.*, 2005). The ability to adapt and use host glycans as a carbon source gives bacteria a sustainable and consistent nutrient supply and in turn an advantage over many other, less well adapted species. The presence of mucins in *in vitro* fermentation models increases the

prevalence of certain strains of bacteria, including *Bacteroidetes*, *Akkermansia* and *Lachnospiraceae* species that are known mucin degraders, whilst levels of *Lactobacillus* and *Bifidobacterium* decrease (Tran *et al.*, 2016). Furthermore, analysis of the human gut microbiota revealed that as many as 86% of the microbiota encode genes for cleavage of mucin glycans, with 89% of genomes capable of metabolising the monosaccharides released (Ravcheev and Thiele, 2017).

Co-evolution of bacteria and host has led to a complex environment containing bacterial species that, in combination, can fully degrade glycan chains and the mucin backbone (McGuckin *et al.*, 2011). In germ-free mice, the mucus layer is thinner than in wild-type mice. This effect can be reversed by stimulation with bacterial compounds such as peptidoglycan or LPS, demonstrating that the presence of bacteria and bacterial compounds stimulate mucin production and secretion (Birchenough *et al.*, 2016, Petersson *et al.*, 2011). These findings suggest that the presence of the microbiota is important for renewal as well as degradation of mucin and is therefore essential for maintenance and turnover of the mucus layer, sustaining a functional barrier against pathogens. This was demonstrated in mice gavaged with *A. muciniphila*, which was shown to positively modulate barrier function in mice (Everard *et al.*, 2013). However, balance is important, a study using a humanised mice model showed that the absence of fibre in the diet led to a dramatic increase in mucus degrading microbes and increased susceptibility to pathogens by depletion of the mucus barrier (Desai *et al.*, 2016).

The degradation of mucins requires the sequential action of a number of enzymes in order to access this plentiful nutrient source. *B. thetaiotaomicron* and other gut commensal species have acquired a range of GHs, sulfatases and proteases to liberate utilisable sugar monomers (Etienne-Mesmin *et al.*, 2019). GHs are classified in the carbohydrate active enzymes database (CAZy; www.cazy.org) by functional domains of enzymes that cleave, modify or create glycosidic bonds (Lombard *et al.*, 2014). Although many bacterial GH genes have been predicted by transcriptomic analyses, relatively few mucin glycan degrading GHs have been functionally and structurally characterised (Tailford *et al.*, 2015a). Microbes most adept at mucin degradation often encode sulfatases, deacetylases, sialidases and fucosidases to remove terminal structures and grant greater accessibility to the extended core structures (Etienne-Mesmin *et al.*, 2019, Ndeh and Gilbert, 2018). The monomers freed by the action of these enzymes may be utilised by the bacteria themselves or released in the local environment for other, scavenging bacteria to use (Marcobal *et al.*, 2013). Sulfatases are of increasing interest in the regulation of the gut microbiota. Sulphate residues terminate mucin glycans and

have been proposed to prevent GHs from removing terminal sugars. The release of sulphate residues has also been proposed to increase the levels of sulphate-reducing bacteria in the gut, this leads to the production of H₂S which can disrupt the mucus network and also lead to epithelial damage (Praharaj *et al.*, 2018, Ijssennagger *et al.*, 2016, Rey *et al.*, 2013).

A suite of GHs including galactosidases, *N*-acetylglucosaminidases and *N*-acetylgalactosaminidases can then degrade the extended core structures and remove the core structure from the mucin backbone, again releasing free monomers (Marcobal *et al.*, 2013, Tailford *et al.*, 2015a). Following removal of the glycan chains, the mucin backbone can then be more easily targeted by proteases. These proteases have been studied in pathogenic and commensal *Escherichia coli* strains and have been proposed to be involved in infection and degradation of the inner mucus layer (see section 1.4.2) (Hews *et al.*, 2017).

In addition to providing a reliable food source, mucin glycans also provide attachment sites for bacteria which can give advantages in transient colonisation. As discussed above, glycans vary greatly between species and different species have contrasting core microbiomes (Figure 9) (Kostic *et al.*, 2013). Microbes are able to bind to mucin glycans of their given host using lectins or other adhesion factors (Etienne-Mesmin *et al.*, 2019, Juge, 2012, Johansson *et al.*, 2011a), also supporting a model of co-evolution (Marcobal *et al.*, 2013). The majority of mechanisms underlying the attachment of microbes to the gut have primarily been carried out using pathogenic models; however, recent work has demonstrated that commensals can use similar mechanisms (Donaldson *et al.*, 2016, Etzold *et al.*, 2014). Bacteria can adhere to the gut via the mucus and its glycan structures by utilising a range of lectins for different sugar specifications allowing binding to mucins and other glycoproteins (Etzold and Juge, 2014). Despite the multitude of lectins present throughout the microbiome, a lack of biochemical evidence for their specific binding patterns to oligosaccharides limits useful manipulation of their mechanisms for probiotics (Juge, 2012). *Lactobacillus* and *Lactococcus* of the Lactobacillaceae family have been used as models in studying commensal adhesion to the gut with exopolysaccharides, pili and cell wall anchored mucin binding proteins (MUBs) shown to be involved, and host specificity determined by diversity in these factors (Donaldson *et al.*, 2016, Frese *et al.*, 2013, MacKenzie *et al.*, 2010).

Furthermore, probiotic strains have been shown to alter oligosaccharide and mucus composition where *Lactobacillus plantarum* colonisation increases the production and secretion of MUC2 and MUC3 compounds, and certain probiotic mixtures can impact

upon oligosaccharide composition (Cornick *et al.*, 2015, Tsirtsikos *et al.*, 2012). Similar findings have been shown for members of the gut microbiota, for example, *B. thetaiotaomicron* can up-regulate host fucosyltransferases, increasing fucose availability and changing the mucus composition to suit its needs (Marcobal *et al.*, 2013). Further evidence for bacterial manipulation of the mucus layer comes from studies in germ-free mice where distinct colonies of genetically identical germ-free mice colonised by different microbial communities showed major differences in barrier dynamics and in mucus permeability (Jakobsson *et al.*, 2015). Comparisons of germ-free and conventionally raised mice showed large changes in glycosyltransferase expression throughout the GI tract, in particular those that allow for extension of the glycan chains. This was supported by the detection of shorter MUC2 glycans in the germ-free mice (Arike *et al.*, 2017). Taken together, these studies show that there is strong evidence of a cross-communication between the host mucosal surface and microbiota to reach symbiosis. A detailed understanding of these processes at the molecular level could lead to the development of novel therapeutics targeting the mucus-associated microbiota.

1.4.2 Enteric Pathogens – Mucus interactions

Pathogenic bacteria have the capacity to migrate through mucus and reach the epithelial layer via flagella and chemotactic stimuli (Josenhans and Suerbaum, 2002, Haiko and Westerlund-Wikstrom, 2013). In an attempt to slow down this migration, the host immune system responds by releasing stored mucin granules, and with it, antimicrobial compounds (Birchenough *et al.*, 2016, McGuckin *et al.*, 2011). Such interactions with the mucins can result in shedding of the mucus layer and thus exclude pathogens (Ribet and Cossart, 2015). If bacteria are able to penetrate through the outer and inner mucus layers, the next challenge is to overcome the glycocalyx and membrane-bound mucins, to reach the epithelial layer.

To circumvent these processes, pathogens have evolved a range of virulence factors designed to evade host defences. In relation to the mucus layer, one mechanism pathogens have developed is the secretion of proteases capable of MUC2 cleavage to accelerate and enhance colonisation (Cornick *et al.*, 2015). These cleavages dysregulate the mucus layer, resulting in diminished viscosity and a dilution of antimicrobial compounds (Martens *et al.*, 2018, McGuckin *et al.*, 2011). For example, the StcE *E. coli* metalloprotease can cleave mucins, helping the bacteria to access the epithelial layer where it can cause disease (Hews *et al.*, 2017).

A second strategy is to release toxins that can diffuse through the mucus layer. These toxins have the capacity to disrupt tight junctions and block epithelial growth, disrupting the intestinal epithelial layer (Awad *et al.*, 2017). Furthermore, toxins can be used to impair mucin secretion pathways, causing a decrease in the viscosity and abundance of trapped antimicrobial compounds leading to greater colonisation (McGuckin *et al.*, 2011). Pathogens may also trigger inflammation of the gut which leads to an increase in the production of antimicrobial compounds. As pathogens have a greater tolerance to these compounds than commensal bacteria, this response provides pathogens a competitive advantage to overcome colonisation resistance (Ribet and Cossart, 2015).

Pathogens can also interact with mucin glycans by using these as a nutrient source, either by releasing nutrients from glycan chains using their GHs, or by scavenging nutrients released by commensal organisms (see section 1.4.1). These cross-feeding interactions by pathogens have been described for *B. thetaiotaomicron* (Martens *et al.*, 2018). The terminal sugars of mucin glycans, fucose and sialic acid, are released from the mucin glycans by *B. thetaiotaomicron* as well as other commensal organisms. Free sialic acid has been shown to promote growth of the pathogens *S. Typhimurium* (*S. Typhimurium*) and *C. difficile* (Ng *et al.*, 2013), whereas fucose promotes enterohaemorrhagic *E. coli* infection (Pacheco *et al.*, 2012). Some pathogens encode their own GHs, these include *S. Typhimurium* LT2 which is a *Salmonella* strain encoding its own sialidase, though this is uncommon among *Salmonella* strains (Arabyan *et al.*, 2017, Arabyan *et al.*, 2016, Minami *et al.*, 2013), and *Vibrio cholerae* which releases and utilises GlcNAc from mucin glycans (Mondal *et al.*, 2014).

As well as disturbing the mucus layer and disrupting intestinal cell layer integrity, some pathogenic species actively avoid the mucus layer to cause infection. M-cells are specialised epithelial cells located above Peyer's patches of the gut which are abundant in cells related to immunity such as macrophages. M-cells regularly sample the luminal environment for bacteria and present antigens directly to those underlying immune system cells initiating a secretory IgA immune response (Martens *et al.*, 2018, McGuckin *et al.*, 2011, Corr *et al.*, 2008). In order to do this, these areas of the GI tract are devoid of goblet cells leaving gaps in the mucus layer (Ermund *et al.*, 2013). *S. Typhimurium* is one example of a bacterium that utilises these M-cells as an entry point to invade the host tissue. Once transported across the epithelial layer and taken up by macrophages, *S. Typhimurium* uses a number of effector proteins encoded by pathogenicity islands to inhibit the activity of lysozymes by forming a protective vacuole where the bacteria can grow and divide (Fabrega and Vila, 2013).

1.5 Sialic acid

1.5.1 Mucus associated sialic acid

Sialic acid is a major epitope of mucins, and due to its terminal location, is a target for many gut pathogenic and commensal species. The complex nature of the 'sialome' leads to the diversity and specificity of these interactions (Cohen and Varki, 2010). Two major forms of sialic acid exist, 2-keto-3-deoxynononic-acid (KDN) and neuraminic acid (Neu). The latter can be modified to *N*-acetylneuraminic acid (Neu5Ac) and then onto *N*-glycolylneuraminic acid (Neu5Gc) which are the most common forms found on mammalian cells, with humans known to lack Neu5Gc (Cohen and Varki, 2010) (Figure 11). Further modifications of the hydroxyl groups add diversity to sialic acids giving the first level of complexity. Evidence suggests that these sialic acid modifications are differentially expressed in different tissue types (Langereis *et al.*, 2015), however, the biological significance of these sialic acid modifications remains largely underexplored. Acetylation is abundant in the human GI tract, except for individuals that do not express *O*-acetylated sialic acids, interestingly this does not appear to be associated with an increased prevalence of disease (Corfield, 2018). This acetylation is believed to have a negative impact on sialic acid recognition by blocking binding, particularly for sialidases. Loss in *O*-acetylation has been shown in cases of IBD and colorectal cancer, using immunohistochemistry approaches, this research also showed changes in *O*-acetylation along the GI tract of individuals (Corfield *et al.*, 1999, Smithson *et al.*, 1997, Reid *et al.*, 1984, Culling *et al.*, 1979). More recently nidovirus virolectins have been used to explore *O*-acetylation of sialic acids *in situ* and have been implicated in aspects of cell development, homeostasis, and other functions (Langereis *et al.*, 2015).

The second level of diversity in the sialome comes from the type of linkage to the underlying sugar. The most common linkages are α 2-3 and α 2-6 to other sugars as well as α 2-8 linkages to other sialic acids. Rarer occurrences of α 2-4 and α 2-9 linkages also exist. The underlying sugar composition and glycan chains (discussed above) that sialic acids are terminally attached to provide a third level of complexity. The glycan and mucin types provide the fourth level of complexity with the final level of diversity provided by the spatial distribution and organisation of sialic acids at the cell surface (Cohen and Varki, 2010).

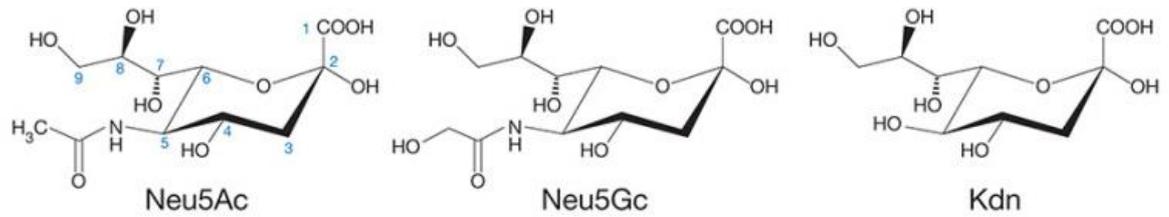


Figure 11: Structure of the common sialic acid derivatives

Further diversity in sialic acids is given by substitutions at the hydroxyl groups.

Sialic acid can be utilised by gut microbes as a nutrient source, but also provides a key target site for attachment and toxin release. Commonly this is associated with viral infections, particularly influenza where specificity for different sialic acid linkages is used in virus typing (Gamblin and Skehel, 2010). Binding to sialic acids has also been demonstrated in the gut for parasitic infection (e.g. *Plasmodium falciparum*) and pathogenic bacteria (Varki, 2008). *Helicobacter pylori* utilises the SabA adhesin which specifically binds to sialyl-lewis-x blood groups in the stomach mucosa (Yamaoka, 2008). In *S. Typhimurium*, the recognition of sialic acids is an important mechanism by which infection takes place. Depletion of Neu5Ac from epithelial Caco-2 cells led to a reduction in *S. Typhimurium* association (Arabyan *et al.*, 2016).

1.5.1 Sialic acid metabolism in bacteria

The metabolic pathway required for bacteria to utilise sialic acid as a nutrient source is based on two distinct metabolic pathways, originally demonstrated in *E. coli* and *B. fragilis* (Brigham *et al.*, 2009). Some bacteria have incomplete gene clusters for sialic acid utilisation, either being able to release sialic acid from glycans but not utilise it e.g. *B. thetaiotaomicron* and *A. muciniphila*, or utilise but not release sialic acid e.g. *E. coli* (Lewis and Lewis, 2012). The latter has also been demonstrated for the enteric pathogens *S. Typhimurium* and *C. difficile* which benefit from the availability of free sialic acid in the mucosal environment (Ng *et al.*, 2013). This leads to cross-feeding between species and symbiotic intra-species relationships, for example *Bifidobacterium bifidum* releases sialic acid that can then be utilised by *Bifidobacterium breve* to sustain its growth (Nishiyama *et al.*, 2018).

In order to utilise sialic acid as a carbon and nitrogen source, it must first be liberated from glycan chains (Almagro-Moreno and Boyd, 2009). This is performed by sialidases of the GH33 CAZy family (www.cazy.org), which are released by gut bacteria and cleave the terminal sialic acid residues. Gut bacteria that express sialidases include multiple

species of *Clostridia* and *Bacteroides*, as well as specific strains of *Bifidobacterium*, *R. gnavus* and *A. muciniphila* (Juge *et al.*, 2016). Strain specificity of sialidase expression is a particularly important factor to consider, for example, only one sialidase from *S. Typhimurium* has to date been functionally characterised (Minami *et al.*, 2013). Bacteria also encode O-acetyl esterases to remove the acetyl groups, often present on sialic acids, facilitating its release by sialidases (Robinson *et al.*, 2017).

To then utilise this nutrient source, sialic acids must be transported into the bacteria. Five different systems of sialic acid transport have been described to date in bacteria, as recently reviewed (Figure 12) (Thomas, 2016). The *nanT* type transporter described in *E. coli* K-12 is a major facilitator superfamily like transporter acting as a H⁺-coupled secondary transporter (Martinez *et al.*, 1995). Other bacteria utilise a tripartite ATP-independent periplasmic transporter, the SiaP protein, which binds free sialic acid and delivers it to the membrane associated SiaQM for transport (Allen *et al.*, 2005). The transmembrane QM domains can either be encoded by different genes or by a single gene. *S. Typhimurium* encodes two transporter systems, the first one being homologous to the *E. coli* NanT transporter and the second one being a sodium solute symporter (SSS) (Severi *et al.*, 2010, Almagro-Moreno and Boyd, 2009). Interestingly the SSS like transporter of sialic acid has been shown to be important for sialic acid uptake and colonisation in *C. difficile* (Ng *et al.*, 2013).

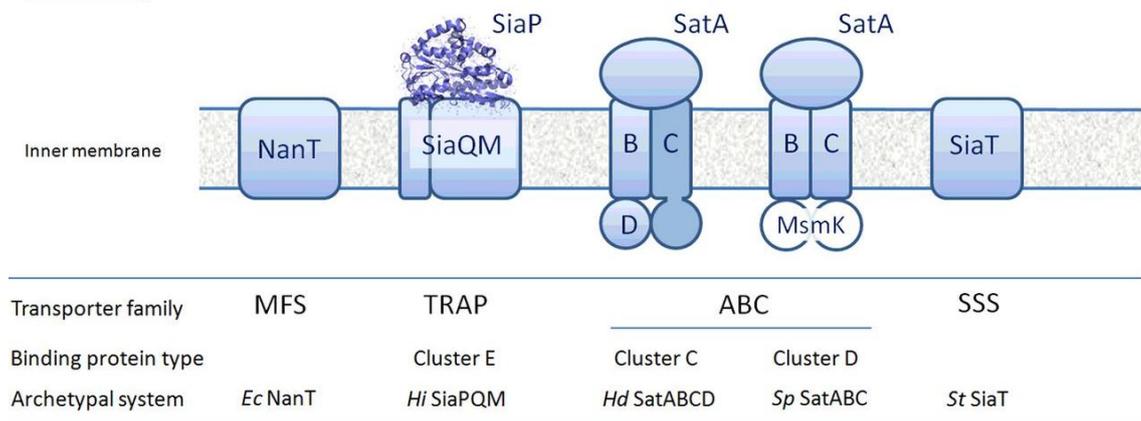


Figure 12: Schematic representation of the five characterised bacterial sialic acid transporters Adapted from Thomas, 2016.

The final class of transporters is the ABC transporter system. Two types of sialic acid ABC transporters have been reported to date, a traditional type found in *B. breve* (Egan *et al.*, 2014) and other species, and a second type that has been named *satABC* but is structurally different, lacking its own dedicated nucleotide-binding protein (Thomas,

2016). This CUT1 ABC transporter of sialic acid transporter has been described in *Streptococcus pneumoniae* and contains two permeases and a substrate binding lipoprotein (Bidossi *et al.*, 2012).

The most recognised pathway of utilising sialic acid is the *nanA/K/E* cluster first described in *E. coli* (Figure 13) (Plumbridge and Vimr, 1999). Following transport of sialic acid residues into the cell by NanT, a sialic aldolase or lyase (NanA), cleaves sialic acid into pyruvate and *N*-acetylmannosamine (ManNAc), which is then phosphorylated by a ManNAc kinase (NanK). The phosphorylated ManNAc is then converted to a phosphorylated GlcNAc (GlcNAc-6-P) by an epimerase (NanE). It is then successively deacetylated and deaminated by NagA (*N*-acetylglucosamine-6-phosphate deacetylase) and NagB (*N*-acetylglucosamine-6-phosphate deaminase), respectively, two enzymes not encoded within the *nan* operon but elsewhere in the genome. This results in the production of fructose-6-phosphate which then enters the glycolysis pathway, while the by-products from the enzyme reactions can be used as precursors in a number of important cellular processes (Vimr, 2013). An alternative pathway has later been discovered in *B. fragilis* where the *nanLET* cluster is utilised (Figure 13). Here NanT transports sialic acid into the cell before a sialic acid lyase (NanL) cleaves it into pyruvate and ManNAc. Then, NanE – which is dissimilar to the NanE found in the *nanA/K/E* cluster – directly epimerises ManNAc to GlcNAc without the need for prior phosphorylation. An alternative ManNAc kinase (RokA) then performs the phosphorylation reaction before conversion to fructose-6-phosphate by NagA and NagB (Brigham *et al.*, 2009).

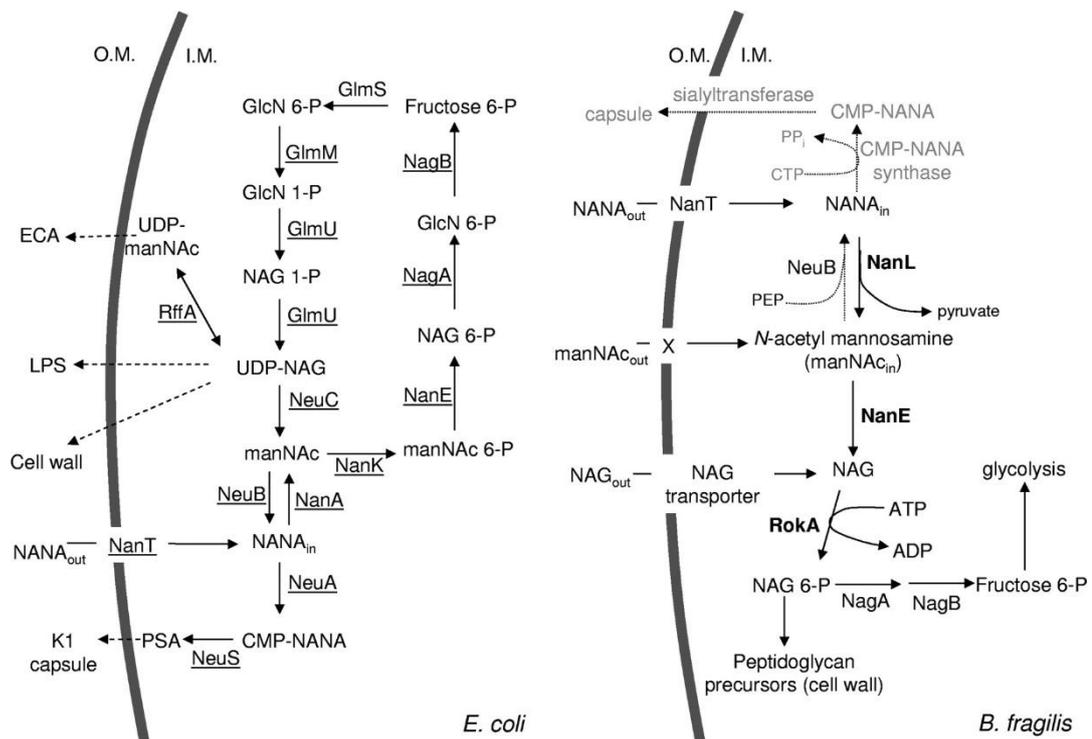


Figure 13: Pathways of sialic acid utilisation in *E. coli* (left; *nanA/K/E*) and *B. fragilis* (right; *nanL/E/T*)

Adapted from Brigham *et al.*, 2009.

Previous research in the group identified an example of an intramolecular *trans*-sialidase (IT-sialidase) in the gut symbiont *R. gnavus* (Crosth *et al.*, 2013). This IT-sialidase – encoded by the *nanH* gene – releases a modified form of sialic acid termed 2,7-anhydro-sialic acid (2,7-anhydro-Neu5Ac) upon cleavage (Figure 14). This is the first example of an IT-sialidase found in a gut bacterial species; however, meta-analysis suggests their occurrence in other bacterial species (Tailford *et al.*, 2015b). In contrast to hydrolytic sialidases which have a broad specificity and can cleave α 2-3-, α 2-6- and α 2-8-linked sialic acid, IT-sialidases can only cleave α 2-3 linkages (Juge *et al.*, 2016).

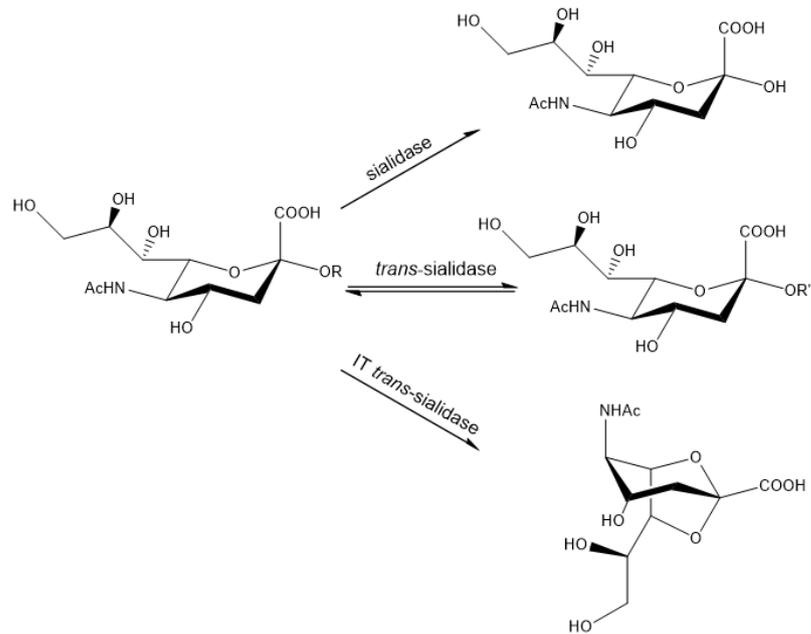


Figure 14: Products from the reactions of hydrolytic, *trans*- and IT *trans*-sialidases.

Whilst the metabolism of Neu5Ac is understood, how 2,7-anhydro-Neu5Ac is utilised by bacteria is yet to be elucidated. In *R. gnavus* ATCC 29149 and ATCC 35913, the IT-sialidase is part of a larger gene cluster which also encodes homologs of the *nanA/K/E* cluster. This gene cluster is only present in mucin glycan foraging strains of *R. gnavus* (Croft *et al.*, 2013). The genes of the *nan* cluster are induced when *R. gnavus* is grown on sialyl-conjugates. These data suggest a similar metabolic pathway as the one described above. However, *R. gnavus* does not grow on Neu5Ac, suggesting a different transport mechanism. It is also unknown whether 2,7-anhydro-Neu5Ac is converted into Neu5Ac inside the cell (and which enzyme may be responsible for this) or if it is directly metabolised by enzymes from the canonical *nan* cluster. This mechanism could greatly impact bacterial species that rely on scavenging free Neu5Ac, by reducing Neu5Ac availability, such as *S. Typhimurium* and *C. difficile* strains which encode the *nan* cluster for Neu5Ac utilisation but lack the sialidase to remove it from glycan chains (Ng *et al.*, 2013).

1.6 Aims and objectives

The overall aim of this work is to investigate the role of 2,7-anhydro-Neu5Ac in gut symbiosis.

Specific objectives include:

- (i) to develop methods allowing the synthesis and analysis of 2,7-anhydro-Neu5Ac,
- (ii) to unravel the metabolic pathway of 2,7-anhydro-Neu5Ac in *R. gnavus*
- (iii) to determine the role of 2,7-anhydro-Neu5Ac metabolism on *R. gnavus* spatial adaptation to the gut
- (iv) to investigate the impact of 2,7-anhydro-Neu5Ac on enteric pathogens and gut microbiome.

Chapter 2

Materials and Methods

2.1 Bacterial strains and media

2.1.1 Media

2.1.1.1 Yeast extract Casitone and Fatty Acid (YCFA)

1 L YCFA contained 10 g casitone, 2.5 g yeast extract, 4 g NaHCO₃, 1 g L-cysteine hydrochloride, 450 mg K₂HPO₄, 450 mg KH₂PO₄, 900 mg NaCl, 90 mg MgSO₄·7H₂O, 90 mg CaCl₂, 1 mg resazurin, 10 mg hemin, 10 mg biotin, 10 mg cobalamin, 30 mg p-aminobenzoic acid, 50 mg folic acid and 150 mg pyridoxamine. Final concentrations of short-chain fatty acids (SCFA) in the medium were 33 mM acetate, 9 mM propionate and 1 mM each of isobutyrate, isovalerate and valerate. The pH was adjusted to 6.5. The medium was prepared under a headspace of 85% N₂, 10% H₂ and 5% CO₂ gas mix. Thiamine and riboflavin were added anaerobically to the medium to give a final concentration of 50 mg/L each and then the medium was autoclaved. Media was supplemented with 11.1 mM carbon source (see section 2.1.1.7)

Brain Heart Infusion with Yeast extract and Hemin (BHI-YH)

BHI-YH contained Brain heart infusion broth (Oxoid LTD, UK) supplemented with 5 g/L of Bacto™ yeast extract (Becton, Dickinson and Company, Sparks, MD) and 5 mg/L of hemin (Sigma-Aldrich, USA). BHI-YH agar was made by adding 1.5 % agar.

2.1.1.2 Bacteroides Defined Media (BDM)

BDM contained 100 mM KH₂PO₄ pH 7.4, 15 mM NaCl, 8.5 mM (NH₄)₂SO₄, 0.1 mM MgCl₂, 50 μM CaCl₂, 0.37 nM vitamin B₁₂, 1.9 μM hemin (Sigma-Aldrich, USA), 0.2 mM histidine and 4 mM L-cysteine.

2.1.1.3 Luria Bertani (LB)

1 L LB broth contained 10 g tryptone, 5 g yeast extract and 10 g NaCl. LB agar was made by adding 1.5 % agar.

2.1.1.4 Auto-induction media (AIM)

AIM is based on 'Terrific Broth Base with Trace Elements' autoinduction media (ForMedium, UK)

2.1.1.5 M9

M9 contained 48 mM Na₂HPO₄, 22 mM KH₂PO₄, 85 mM NaCl, 19 mM NH₄Cl, 2 mM MgSO₄, 0.1 mM CaCl₂, 0.04 mg/ml histidine

2.1.1.6 Batch culture media

1 L batch culture medium contained 2 g peptone water, 2 g Yeast extract, 0.1 g NaCl, 0.04 g K₂HPO₄, 0.04 g KH₂PO₄, 0.01 g MgSO₄.6H₂O, 0.01 g CaCl₂.2H₂O, 2 g NaHCO₃, 0.5 g Cysteine HCl, 0.5 g bile salts, 2 ml Tween 80, 0.02 g hemin, 10 µl 0.5% v/v vitamin K. Sugars were added at a concentration 11.1 mM and vitamin B₁₂ supplementation at 0.37 nm where indicated.

2.1.1.7 Carbon sources

The carbon sources used were glucose (Sigma-Aldrich, USA), lactose (Sigma-Aldrich, USA), Neu5Ac (Sigma-Aldrich, USA), Neu5Gc (Carbosynth, UK), fucose (Glycom, Denmark), α2,3 sialyllactose (3'SL; Carbosynth, UK), α2,6 sialyllactose (6'SL; Carbosynth, UK), α1,2 fucosyllactose (2'FL; Glycom, Denmark), α1,3 fucosyllactose (3'FL; Glycom, Denmark), Pig gastric mucin (PGM; Sigma-Aldrich, USA) previously purified using methods published (Gunning *et al.*, 2013) to give pPGM, 2,7-anhydro-Neu5Ac (see section 2.4).

2.1.2 Bacterial strains and culture conditions

The bacterial strains used in this work are as follows, *Escherichia coli* BL21, DH5-α, CA434, *Ruminococcus gnavus* E1, ATCC 29149 and ATCC 35913, *Salmonella enterica* serovar Typhimurium LT2 and SL1344, *Clostridium difficile* NCTC12726, *Bacteroides thetaiotaomicron* ATCC 29148. The growth of strains was monitored using an ultraspec 10 cell density meter (Biochrom, UK)

R. gnavus ATCC 29149, ATCC 35913, E1, *B. thetaiotaomicron* ATCC 29148 and *C. difficile* NCTC 12726 were routinely grown in an anaerobic cabinet (Don Whitley, UK) at 37 °C.

S. Typhimurium strains were grown in aerobic conditions in a static 37 °C incubator and anaerobic conditions by either limiting air flow with a layer of mineral oil or placing inside an anaerobic cabinet. *E. coli* strains were grown at 37 °C in aerobic conditions with 180 – 250 rpm shaking.

Growth curves were obtained using 200 µl cultures in 96-well microtiter plates. The OD₅₉₅ was measured every 30 or 60 min for 10 - 48 hours in an infinite F50 plate reader (Tecan, UK) housed within an anaerobic cabinet connected to Magellan V7.0 software. Growth of *S. Typhimurium* strains was also monitored using spectramax M5 (Molecular Devices, USA), a temperature-controlled plate reader for aerobic and mineral oil assays

Co-cultures of *S. Typhimurium* SL1344 and *B. thetaiotaomicron* or *R. gnavus* were grown in either YCFA or BDM medium supplemented with 11.1 mM 3'SL as carbon source under anaerobic conditions at 37 °C. The medium was inoculated with 10⁸ of each bacterial strain. Co-culture growth was monitored using an ultraspec 10 cell density meter (Biochrom, UK). To assess *S. Typhimurium* growth, 10-fold serial dilutions of the co-cultures were plated on LB containing 100 µg/ml streptomycin and grown under aerobic conditions overnight before counting. Alternatively, DNA was extracted (see section 2.3.1.1) and bacterial growth quantified by qPCR (see section 2.3.6)

2.2 Bioinformatics

2.2.1 Identification of *nan* genes in *Salmonella* and *C. difficile*

Bioinformatics searches of NanA/K/E cluster proteins and NanH (sialidase) from *R. gnavus* ATCC 29149 were performed against a range of *Salmonella* and *C. difficile* strains using the TBLASTN function on the Ensembl Bacteria website with default settings (<http://bacteria.ensembl.org/>). The NanH was also split by domain into CBM40 (carbohydrate binding module, www.cazy.org), IT (uncharacterised domain within the catalytic domain of *RgNanH* not present in bacterial hydrolytic sialidases) and catalytic domain (Glycoside Hydrolase family 33, www.cazy.org) in an attempt to distinguish IT-sialidases from hydrolytic sialidases. The 146 *Salmonella* strains and 71 *C. difficile* strains were selected to represent the range and variety of strains available within the database.

2.2.2 Sequence Similarity Networks

To carry out Sequence Similarity Network (SSN) analysis of proteins encoded by the *R. gnavus nan* cluster, the InterPro families for *RgNanH* (Glycoside Hydrolase, family 33; IPR001860) and *RgNanA* (N-acetylneuraminatase lyase; IPR005264) were identified using the UniProt database, the family identifier was used to extract protein sequences using Enzyme Function Initiative (EFI) Enzyme Similarity tool (Gerlt *et al.*, 2015). For the other proteins, the families found in the InterPro database were too large to be analysed, so the sequence BLAST tool was used with a maximum of 2500 protein sequences

extracted. From this SSNs were generated and viewed in Cytoscape version 3.6 (Shannon *et al.*, 2003).

2.2.3 Multigene cluster BLAST analysis

Homologous gene clusters were identified for the *R. gnavus* ATCC 29149 *nan* cluster using MultiGeneBlast (Medema *et al.*, 2013). The BCT (Bacteria) GenBank subdivision was queried with the *R. gnavus nan* cluster sequence spanning RUMGNA_02701 – RUMGNA_02691. The data were manually curated, excluding all clusters that do not contain a predicted sialidase or are homologous to the functionally characterized *S. pneumoniae* NanC cluster (Owen *et al.*, 2015, Xu *et al.*, 2011).

2.3 Molecular Biology

2.3.1 Genomic DNA (gDNA) extractions

2.3.1.1 In vitro anaerobic cultures

Genomic DNA was extracted from bacterial mono- or co-cultures using on 1 ml of liquid cultures grown to exponential phase using GeneJET Genomic DNA Purification Kit (ThermoFisher, UK), according to the manufacturer's instructions.

2.3.1.2 Faecal samples and in vitro colon model

Genomic DNA was extracted from faecal samples and *in vitro* colon model samples using the MP Biomedicals Fast DNA™ SPIN kit for Soil DNA extraction with the following modifications. The samples were resuspended in 978 µl of sodium phosphate buffer (provided) before being incubated at 4°C for one hour following addition of 122 µl MT Buffer. The samples were then transferred to the lysing tubes and homogenised in a FastPrep® Instrument (MP Biomedicals) 3 times for 40 sec at a speed setting of 6.0 with 5 min on ice between each bead-beating step. The protocol was then followed as recommended by the supplier.

2.3.2 RNA extraction

2.3.2.1 In vitro anaerobic cultures

RNA was extracted from 3 ml of liquid cultures of *R. gnavus* grown to exponential phase using RNA protect (Qiagen, Germany) and RNeasy Mini Kit (protocol 2: Enzymatic Lysis and Mechanical Disruption of Bacteria and protocol 7: Purification of Total RNA from Bacterial Lysate Using the RNeasy Mini Kit), according to the manufacturer's

instructions. DNase treatment was performed using the TURBO DNA-free kit (Invitrogen) or RNeasy Mini Kit (Qiagen) according to manufacturer's instructions.

Alternatively, RNA extraction was performed using the Monarch Total RNA Miniprep Kit (New England Biolabs) extraction kit, according to the manufacturer's instructions followed by DNase treatment with RQ1 RNase-Free DNase (Promega).

2.3.2.2 Caecal content

Caecal content was resuspended in RNAlater (Sigma-Aldrich, USA) and stored at -80°C until extraction. RNA extraction was performed using RNeasy PowerMicrobiome Kit (Qiagen) according to the manufacturer's instructions.

2.3.2.3 Mouse tissue

Mouse colonic tissue was gently washed and stored in RNAlater (Sigma-Aldrich, USA) at -80 °C until RNA extraction. Homogenisation was achieved with acid washed glass beads using the FastPrep®-24 (MP Biomedicals, USA) by 3 intermittent runs of 30 s at 6 m/s speed every 5 min, at room temperature. RNA extraction was performed using the RNeasy mini kit (QIAGEN) following the manufacturer's instructions for purification of total RNA from animal tissues, including the on-column DNase digestion. Elution was performed as recommended with 50 µl RNase-free water. The quality and concentration of the RNA samples was assessed using the NanoDrop 2000 Spectrophotometer Nanodrop, the Qubit RNA HS assay on Qubit® 2.0 fluorometer (Life Technologies) or Agilent RNA 600 Nano kit on Agilent 2100 Bioanalyzer (Agilent Technologies, Stockport, UK).

2.3.3 cDNA synthesis

Following RNA extraction, cDNA synthesis was carried out using QuantiTect Reverse Transcriptase (Qiagen) according to the manufacturer's instructions with control samples without addition of the transcriptase to test for DNA contamination (-RT).

2.3.4 Polymerase chain reaction (PCR)

PCR was carried out using the HotStarTaq Master Mix Kit (Qiagen, Germany) for both genomic DNA and cDNA amplifications, or Go Taq® Green (Promega), with primers detailed in Table 2. For HotStarTaq PCR, the conditions consisted of an initial 10 min 95 °C denaturation step followed by 40 cycles of 95 °C 15 sec, 60 °C 30 sec, 72 °C 1 min (variable depending on length of amplicon; 1 min per 1 kb). For Go Taq® Green PCR,

the conditions consisted of an initial 2 min 95 °C denaturation step followed by 35 cycles of 95 °C 30 sec, 60 °C 30 sec, 72 °C 1 min (variable depending on length of amplicon; 1 min per 1 kb).

The PCR products were analysed on 1.5 % agarose gels using 1 kb or 100 bp DNA markers (New England Biolabs, USA), run for 45 – 60 min at 80 – 100 volts. Gels were stained with ethidium bromide and imaged under UV light using an Alphaimager, alternatively 1 µl of Midori green direct DNA stain (Geneflow, UK) was added to 9 µl of PCR product prior to loading on the gel. The sequence of the PCR products was confirmed by sequencing using Mix2Seq (Eurofins genomics) where required.

2.3.5 Gene expression analysis by qRT-PCR

To assess to expression of the *nan* operon in *R. gnavus* ATCC 29149 during growth on different carbon sources, qPCR was carried out in an Applied Biosystems 7500 Real-Time PCR system (Life Technologies Ltd). Briefly, *R. gnavus* was grown on glucose, 3'SL or 2,7-anhydro-Neu5Ac in triplicate with cDNA samples derived from each (see section 2.3.2.1 / 2.3.3). One pair of primers was designed for each target gene using ProbeFinder version 2.45 (Roche Applied Science, Penzberg, Germany). Calibration curves were prepared for each primer pair using 2.5-fold serial dilutions of *R. gnavus* ATCC 29149 gDNA (see section 2.3.1.1). The standard curves showed a linear relationship of log input DNA vs. the threshold cycle (Ct), with acceptable values for the slopes and the regression coefficients (R^2). The dissociation curves were also performed to check the specificity of the amplicons. Each qPCR reaction (10 µl) was carried out in triplicate with 1 ng sample (cDNA or -RT) and 0.2 µM of each primer (Table 2), using the QuantiFast SYBR Green PCR kit (Qiagen) according to supplier's advice (with the exception of that the combined annealing/extension step which was extended to 35 sec instead of 30 sec).

Data obtained with cDNA were analysed only when Ct values above 36 were obtained for the corresponding -RT (see section 2.3.3). For each cDNA sample, the 3 Ct values obtained for each gene were averaged. The data were then analysed using the $2^{-\Delta\Delta CT}$ method relative to the housekeeping *gyrB* (RUMGNA_00867) gene as a reference gene and glucose as a reference condition.

2.3.6 Bacterial quantification by qPCR

Colonisation was quantified using qPCR carried out in an Applied Biosystems 7500 Real-Time PCR system (Life Technologies Ltd) or Applied Biosystems StepOnePlus Real-

Time PCR system (Life Technologies Ltd). One pair of primers was designed to specifically target each bacterial species or strain. The primers (Table 2) were between 18 and 23 nt-long, with a T_M of 58–62°C. Standard curves were prepared in triplicate for all primer pairs using a 10-fold serial dilution of DNA corresponding to 10^7 copies / 2 μ l to 10^2 copies / 2 μ l diluted in 5 μ g/ml Herring sperm DNA to reduce non-specific amplification. The standard curves showed a linear relationship of log input DNA vs. the threshold cycle (Ct), with acceptable values for the slopes and the regression coefficients (R^2). The dissociation curves were also performed to verify the specificity of the amplicons. Each qPCR reaction (10 μ l) was then carried out in triplicate with 2 μ l of 1 ng/ μ l DNA (diluted in 5 μ g/ml Herring sperm DNA) and 0.2 μ M of each primer, using the QuantiFast SYBR Green PCR kit (Qiagen) according to the manufacturer's instructions (except that the combined annealing/extension step was extended to 35 sec instead of 30 sec). Data obtained were analysed using the prepared standard curves.

2.3.7 Sequencing

2.3.7.1 16S sequencing

Following extraction of gDNA (see section 2.3.1.2) the concentration and quality of gDNA was assessed by Qubit and Nanodrop. DNA was normalised to 5 ng/ μ l and sequenced in house by David Baker as follows. For PCR each well contained 4 μ l kapa2G buffer, 0.4 μ l dNTPs, 0.08 μ l kapa2G polymerase, 0.4 μ l 10 μ M forward tailed specific primer, 0.4 μ l 10 μ M reverse tailed specific primer and 13.72 μ l PCR grade water and 1 μ l normalised DNA. The PCR conditions were 95°C for 5 min followed by 30 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 30 sec with a final 72°C for 5 min. Following PCR, a 0.7X SPRI using KAPA Pure Beads (Roche) was performed eluting the DNA in 20 μ l of EB (10mM Tris-HCl).

Following the first PCR and clean-up, a second PCR was performed using 5 μ l of the clean PCR product, 4 μ l kapa2G buffer, 0.4 μ l dNTPs, 0.08 μ l kapa2G polymerase, 2 μ l of each P7 and P5 of Nextera XT Index Kit v2 index primers (Illumina) and 6.52 μ l. The PCR was run using 95°C for 5 min, 10 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 30 sec followed by a final 72°C for 5 min. Following the PCR reaction, the libraries were quantified using the Quant-iT dsDNA Assay Kit, high sensitivity kit (Fisher Scientific) and run on a FLUOstar Optima plate reader. Libraries were pooled following quantification in equal quantities. The final pool was cleaned with 0.7X SPRI using KAPA Pure Beads. The final pool was quantified on a Qubit 3.0 instrument and run on a High

Sensitivity D1000 ScreenTape (Agilent) using the Agilent TapeStation 4200 to calculate the final library pool molarity.

The pool was run at a final concentration of 8 pM on an Illumina MiSeq instrument using MiSeq® Reagent Kit v3 (600 cycle, Illumina) following the Illumina recommended denaturation and loading recommendations which included a 20% PhiX spike in (PhiX Control v3 Illumina). The raw data were analysed locally on the MiSeq using MiSeq reporter.

Taxonomic classifications were made using the QIIME2 program (Bolyen *et al.*, 2019) and Greengenes database (Werner *et al.*, 2012, McDonald *et al.*, 2012, DeSantis *et al.*, 2006) by Tristan Seecharran in the group.

2.3.7.2 RNA seq

RNAseq was carried out by Novogene (HK) (Hong Kong). Briefly, mRNA was enriched using oligo(dT) beads, fragmented randomly in fragmentation buffer, followed by cDNA synthesis using random hexamers and reverse transcriptase. After first-strand synthesis, a custom second-strand synthesis buffer (Illumina) was added with dNTPs, RNase H and *E. coli* polymerase I to generate the second strand by nick-translation. The final cDNA library was obtained after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment. Library concentration was first quantified using a Qubit® 2.0 fluorometer (Life Technologies), and then diluted to 1 ng/µl before checking insert size on an Agilent 2100 and quantifying to greater accuracy by qPCR (library activity >2 nM). Sequencing of the library was carried out on Illumina HiSeq platform and 125/150 bp paired-end reads were generated.

FASTQ files containing base calls and quality information for all reads that passed quality filtering were generated. Reads were mapped to the mouse reference genome using TopHat2 (Kim *et al.*, 2013). The mismatch parameter was set to two, and other parameters were set to default. Appropriate parameters were also set, such as the longest intron length. Filtered reads were used to analyse the mapping status of RNA-seq data to the reference genome. The HTSeq software was used to analyse the gene expression levels, using the union mode (Anders and Huber, 2010). In order for the gene expression levels estimated from different genes and experiments to be comparable, the FPKM (Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) was used to take into account the effects of both sequencing depth and gene length. The differential gene expression analysis was carried out using the DESeq

package (Anders and Huber, 2010) and the readcounts from gene expression level analysis as input data. An adjusted p value (padj) cut-off of 0.05 was used to determine differential expressed transcripts.

2.3.8 Heterologous expression and protein purification

2.3.8.1 Cloning

The target gene sequences were amplified from *R. gnavus* ATCC 29149 genomic DNA (obtained as described in section 2.3.1.1) by PCR using Phusion high fidelity polymerase (New England Biolabs) and specific primers designed to add *Bam*HI or *Hind*III restriction sites (Table 2). PCR fragments were cloned into pEHISTEV vector (Appendix 1), in frame with the sequence encoding an N-terminal HIS-tag for purification, following restriction enzyme digestions with *Bam*HI, *Hind*III (Roche) and ligation with T4-DNA ligase (Promega).

For transformation into *E. coli*, 100 ng of DNA was added to 50 µl of chemically competent DH5-α cells, incubated on ice for 30 min, heated to 42°C for 45 sec and cooled on ice. 300 µl LB was added to the suspension and cells were incubated at 37°C with shaking at 180 rpm for 1 hour. Cells were subsequently plated on LB supplemented with 30 µg/ml kanamycin (Sigma) plates and grown at 37 °C overnight. Insertion of the target sequence in the plasmid was confirmed by colony PCR using GoTaq G2 polymerase (Promega) with vector specific T7 primers and products analysed using 1% agarose gel. Cells containing recombinant plasmids with a correct size construct were grown in 10 ml LB containing kanamycin overnight and the plasmid purified using either Qiagen or NEB miniprep kits and sequenced with T7 primers using Mix2Seq (Eurofins genomics).

Purified recombinant plasmids with the correct sequence were transformed into chemically competent BL21 cells using the above methodology, and successful transformation and verification of sequence also confirmed as above, before glycerol stocks were made by mixing 250 µl of overnight culture and 250 µl 80 % glycerol. Note, the full-length *R. gnavus* IT-sialidase (*Rg*NanH) and catalytic domain (GH33) had previously been cloned into the pOPINF (Appendix 2) expression system which allows the fusion of the gene of interest to an artificial HIS-tag sequence at the N-terminus and encodes chloramphenicol resistance (Tailford *et al.*, 2015b).

2.3.8.2 Mutagenesis

The *RgNanA* active site mutant, K167A, was generated using the QuikChange Lightning mutagenesis kit (Agilent) from purified plasmid using primers listed in Table 2.

2.3.8.3 Protein expression

For protein expression, *E. coli* BL21 recombinant cells were grown overnight in 10 ml LB supplemented with kanamycin (or chloramphenicol (Sigma) for *RgNanH* and GH33) at 37 °C with shaking at 180 rpm. This overnight culture was then used to inoculate 800 ml AIM containing appropriate antibiotic selection and grown at 37°C with shaking at 250 rpm for 3 hours before the temperature was reduced to 16°C for a further 72 hours. Pellets were harvested by centrifugation at 8,000 rpm for 15 min and lysed using BugBuster, according to the suppliers advice (Merck, Germany).

2.3.8.4 Protein purification by Immobilised Metal Affinity Chromatography (IMAC)

The bacterial cell extracts were dialysed in binding buffer (0.5 M NaCl, 20 mM Tris-HCl pH 7.9) for IMAC using 5 ml HisTrap High Performance column (GE Healthcare), or directly loaded onto HiTrap Fast Flow Crude columns (GE Healthcare). The crude protein mix was loaded onto the IMAC column using the AKTApurify (GE Healthcare) washed with binding buffer until the UV level was stable and the protein eluted using binding buffer containing 500 mM imidazole. Proteins were immediately desalted into gel filtration (150 mM NaCl, 20 mM Tris-HCl pH 7.9) or Ion exchange buffer using a HiPrep 26/10 desalting column using the AKTApurify.

2.3.8.5 Ion exchange using the Resource Q column

The RUMGNA_02695 (*RgNanOx*) protein required further purification prior to size exclusion chromatography. Following IMAC, the protein was desalted into 20 mM TRIS-HCl pH 6.5 and loaded onto a Resource Q anion exchange column (GE healthcare). The protein was eluted using an increasing concentration of Buffer B (20 mM TRIS-HCl, 1M NaCl pH 6.5) over 20 column volumes, before the protein was desalted into gel filtration buffer as above.

2.3.8.6 Size exclusion chromatography

Partially purified proteins were concentrated using a 10-kDa MWCO Vivaspinn column (Vivaspin, Germany) prior to gel filtration with the AKTApurify in gel filtration buffer using a HiLoad 16/600 Superdex 75 or 200 pg column (GE Healthcare). Note that for

RUMGNA_02695 (*RgNanOx*) 2 mM Tris(2-carboxyethyl) phosphine hydrochloride (TCEP) was added to the gel filtration buffer.

2.3.8.7 Analysis of recombinant proteins

Collected fractions were analysed at each stage of purification using NuPAGE Novex 4-12 % Bis-Tris gels (Life Technologies, UK). Fractions were pooled and concentrated using a 10-kDa MWCO Vivaspin column (Vivaspin, Germany). Protein concentration was determined by NanoDrop (Thermo Scientific, USA) using the extinction coefficient calculated by ProtParam (ExPASy-Artimo, 2012) from the peptide sequence.

2.3.8.8 List of primers used

Primers were supplied by Eurofins MWG (Germany)

Table 2: List of primers used in this work.

Primer Name	Sequence	Function	
RUMGNA_02690_F	ACGTCCGGTGGTATTCGTAA		qPCR - gene specific primer
RUMGNA_02690_R	CATACACAGGGCGAGTTCAA		
RUMGNA_02691_F	TGGGATGGATACCATATCTTCG		
RUMGNA_02691_R	CAATGGCACGATTACAAATCA		
RUMGNA_02692_F	CTGGAACTTGGTGCAGTAAGAG		
RUMGNA_02692_R	AATCTCCATGTCCTCATCTACCA		
RUMGNA_02693_F	TGCAGGAGTCAAACACAAGG		
RUMGNA_02693_R	CCTTGCCTTTTGGGGTGTA		
RUMGNA_02694_F	ATAGACTGGCCACGGGATT		
RUMGNA_02694_R	TGTAGGAAGCACTCCCTTGATC		
RUMGNA_02695_F	GTGATTGTGGCGGTACTG		
RUMGNA_02695_R	CTGTGACTCATGCACAAGGAA		
RUMGNA_02696_F	GGTTTCCGTGGCATTAAAATC		
RUMGNA_02696_R	CCAATCATATCTCCCCAGTTCA		
RUMGNA_02697_F	TATGCCCGTTGGATTTTTGT		
RUMGNA_02697_R	CAGGCTGAAAATATTGGGTAT		
RUMGNA_02698_F	TCGGAAGGTGAAAAAGATGG	Quantification of <i>RgSBP</i>	
RUMGNA_02698_R	CCACATAACCTGTGCGTGAG		
RUMGNA_02699_F	AAGTGAAAAGTACACGCCACAA		
RUMGNA_02699_R	AAATAGGCTGCTGAATCTCTGTT TA		
RUMGNA_02700_F	GGATGAATCAGGCGGTTATC		
RUMGNA_02700_R	TATGTGCATCGCTTGAAAG		
RUMGNA_02701_F	GTATCCTGGGAATAGGAATTAATG AT		
RUMGNA_02701_R	TATACGGTGTCCGGCTCCAAT		
RUMGNA_02702_F	CGAATCCAACACTCAAAGCA		
RUMGNA_02702_R	GCACATAGCGCAATCCTTTC		
RUMGNA_00867_Fc	GGAGCAGACCAGATCCAAAT	qPCR - housekeeping	
RUMGNA_00867_Rc	CCAATATACATTCCCGGTCTTT		
RUMGNA_02692_pEHISTEV_F	CGTCATGGATCCATGAGAAATCT TGAGAAGTATAAAGG		Cloning Recombinant Proteins
RUMGNA_02692_pEHISTEV_R	ACTGGTAAGCTTTTATAGAAATT TCTTCTTCGC		
RUMGNA_02694_pOPINF_F	AAGTTCTGTTTCAGGGCCCGCA AGAGGCCAGACAGAT		
RUMGNA_02694_pOPINF_R	ATGGTCTAGAAAGCTTTATGGT TGAACTTTCAGTTCATC		
RUMGNA_02695_pEHISTEV_F	CGTCATGGATCCATGAAAACAG TAGGATATGCAATTG		

RUMGNA_02695_ pEHISTEV_R	ACTGGTAAGCTTTCATATCGTAC TTCCCTCTCC		
RUMGNA_02698_ pEHISTEV_F	CGTCATGGATCCGCAAATCAAG TGGCATCC	st. curve of <i>Rg</i> SBP	
RUMGNA_02698_ pEHISTEV_R	ACTGGTAAGCTTATCATTGCTG CACGGTACTG		
RUMGNA_02700_ pEHISTEV_F	CGTCATGGATCCATGATATTTGA TTCGATTCAG		
RUMGNA_02700_ pEHISTEV_R	CGTTAAAATTAAAGTTTAAAAGC TTACCAGT		
RUMGNA_02701_ pEHISTEV_F	CGTCATGGATCCATGAGTTTTG AGGAGATATTTTATTTTG		
RUMGNA_02701_ pEHISTEV_R	CGTTAAAAAATATGAGATTCTA ATTTGAAAGCTTACCAGT		
RUMGNA_02692_ K167A_F	CCCAACGTCATTGGTGTGCGG AATTCCTCTATGCCGG		
RUMGNA_02692_ K167A_R	CCGGCATAGAGGAATTCGCGAC ACCAATGACGTTGGG		
RUMGNA_02694_ F2	TGTTTCTATAGCAGCAACTGC	Primers to check ClosTron mutagenesis	
RUMGNA_02694_ R2	CCAGCGTATATTTAAACTCGC		
16S_Full_27F	AGAGTTTGATCMTGGCTCAG	Universal 16S gene	Preparation of st. curve for quantification by qPCR
16S_Full_RP2	ACGGCTACCTTGTACGACTT		
InvA_Full_F	CTATGTTTCGTCATTCCATTACC	InvA gene in <i>Salmonella</i>	
InvA_Full_R	CGTCAGACCTCTGGCAGTAC		
Rg_16S_5F	TGGCGGCGTGCTTAACA	16S of <i>R.</i> <i>gnavus</i>	
Rg_16S_5rev	TCCGAAGAAATCCGTCAAGGT		
UNI_16S_F	GTGSTGCAYGGYYGTCGTCA	Universal 16S gene	qPCR primers specific to <i>R.</i> <i>gnavus</i> Wild type or mutant
UNI_16S_R	ACGTCRTCCMCNCCTTCCTC		
InvA_qPCR_F	GATTTGAAGGCCGGTATTATTG	InvA gene in <i>Salmonella</i>	
InvA_qPCR_R	TGACGGTGCGATGAAGTTTAT		
RUMGNA_WT/ MUT_F	GAAATCTCGGAAGGTACAGG	Primers based on insertion in <i>Rg</i> NanH to distinguish Wild type and mutant	
RUMGNA_WT_R	TTCCTTTTGTACCATTTCCC		
RUMGNA_MUT_R	AGGGGTACGTACGGTTCC		

2.3.9 X-ray crystallography

For X-ray crystallography purposes, the recombinant proteins were purified as discussed above followed by the removal of the N-terminal HIS-tag using proTEV plus (promega) protease cleavage according to manufacturers' instructions in 10 ml gel filtration buffer with concentration of the protease reduced to 25 U/ml and cleavage performed at 4 °C overnight. The reaction was dialysed in binding buffer and loaded onto an IMAC column using the AKTExpress system. The cleaved protein does not bind to the column and passes into the flow through, while the proTEV protease contains an internal HIS-tag which binds and remains on the IMAC column. The flow through containing the cleaved target protein was concentrated and purified by gel filtration as described above.

2.3.9.1 RgNanA crystallisation and 3D structure solving

RgNanA crystallisation was performed by David Owen (Diamond Light Source, Didcot, UK). Sitting drop vapour diffusion crystallisation experiments of *RgNanA* wild-type were set up at a concentration of 20 mg/ml and monitored using the VMXi beamline at Diamond Light Source (Sanchez-Weatherby *et al.*, 2019). The described *RgNanA* wild-type crystal structure was acquired from a crystal grown in the Morpheus screen (Molecular Dimensions), 0.2 M 1,6-hexandiol, 0.2 M 1-butanol, 0.2 M 1,2-propanediol, 0.2 M 2-propanol, 0.2 M 1,4-butanediol, 0.2 M 1,3-propanediol, 0.1 M Hepes/MOPS pH 6.5, 20% ethylene glycol, 10% PEG 8000. The diffraction experiment was performed on beamline I24 beamline at Diamond Light Source Ltd at 100K using a wavelength of 0.9686 Å. The data were processed with Xia2 making use of aimless, dials, and pointless. The structure was phased using MrBump through CCP4 online and Molrep (Krissinel *et al.*, 2018, Vagin and Teplyakov, 2010, Keegan and Winn, 2008), by CdNal from *C. difficile* (PDB 4woq) prepared using Chainsaw. Refinement was carried out using Refmac, Buster, and PDB redo (van Beusekom *et al.*, 2018, Emsley, 2017, Smart *et al.*, 2012, Langer *et al.*, 2008, Winn *et al.*, 2003). Coot and ArpWarp were used for model building and Molprobity for structure validation (Williams *et al.*, 2018). It was not possible to crystallise *RgNanA* wild-type in the presence of Neu5Ac as it caused protein precipitation and Neu5Ac soaking experiments dissolved the crystals. Therefore, experiments with *RgNanA* K167A mutant were set up at 25 mg/ml. Diffracting crystals grew in 0.1 M Tris/BICINE pH 8.5, 20% ethylene glycol, 100 mM MgCl₂, 10% PEG 8000 and diffraction experiments performed on beamline I04 at Diamond Light Source using a wavelength of 0.9795 Å. The crystal structure was phased with PHASER (McCoy *et al.*, 2018) using the *RgNanA* wild-type crystal structure. A 60 s 5 mM Neu5Ac soak prior to freezing generated the *RgNanA* K167A-Neu5Ac complex. Due to data anisotropy, the

data were processed in autoPROC (McCoy *et al.*, 2018, Vonrhein *et al.*, 2011) with the STARANISO option (Tickle and Vonrhein, 2018) and refined with Buster using the previously obtained models of RgNanA wild-type and K167A-Neu5Ac complex.

2.3.9.2 RgNanOX crystallisation and 3D structure solving

RgNanOx crystallisation was performed by Micah Lee (Research Complex at Harwell, Didcot, UK). Sitting drop vapour diffusion crystallisation experiments of RgNanOx were set up at a concentration of 20 mg/ml. The described crystal structure was acquired from a crystal grown in the JCSG Plus screen (100 mM sodium citrate pH 5.5, 20% PEG 3000). The diffraction experiment was performed on the I04 beamline at Diamond Light Source Ltd at 100K using a wavelength of 0.9795 Å. The data were processed with Xia2 making use of aimless, dials, and pointless. The structure was phased using MrBump through CCP4 online, by PDB 5UI9_B prepared using Sculptor (Krissinel *et al.*, 2018, Bunkóczi and Read, 2011, Vagin and Teplyakov, 2010, Keegan and Winn, 2008). Refinement was carried out using Phenix AutoBuild, Refmac, Buccaneer, TLSNMD, and PDB redo (van Beusekom *et al.*, 2018, Adams *et al.*, 2010, Painter and Merritt, 2006, Cowtan, 2006, Winn *et al.*, 2003). Coot was used for model building and Molprobit for structure validation (Williams *et al.*, 2018, Emsley, 2017). It was not possible to crystallise the protein with substrate due to the protein crashing in presence of 2,7-anhydro-Neu5Ac, likely due to pH changes.

2.4 2,7-anhydro-Neu5Ac synthesis

2,7-anhydro-Neu5Ac was synthesised from 2'-(4-methylumbelliferyl)- α -D-N-acetylneuraminic acid (MU-NANA) (Toronto research Chemicals, Canada), or fetuin (Sigma-Aldrich, USA) pre-dialysed in 100 μ M ammonium formate. The starting material was dissolved in 2.5 ml (MU-NANA) or 20 ml (fetuin) 100 μ M ammonium formate and 33 nM RgNanH or GH33 (see section 2.3.8) was added. The reactions were carried out at 37 °C statically or with gentle shaking at 100 rpm. The product was separated from MU by folch-partitioning with Chloroform: Methanol (2:1), H₂O added and vortexed. The upper layer containing the sugar was removed and dried using a speed vacuum (Concentrator Plus; Eppendorf, Germany). For fetuin, the reaction was enclosed in dialysis tubing placed in 100 ml of 100 μ M ammonium formate; with 2,7-anhydro-Neu5Ac able to cross the 10-kDA MWCO while the fetuin and RgNanH / GH33 proteins remained inside the membrane. The resulting product was freeze dried, dissolved in H₂O and separated from salts using a P2 biogel column (BIORAD, USA). The resulting 2,7-anhydro-Neu5Ac was freeze-dried and its purity determined by NMR.

2.5.2.7-anhydro-Neu5Ac detection and quantification

2.5.1 Charged Aerosol Detector (CAD)

HPLC-CAD was performed using an Ultimate 3000 HPLC system (Thermo Scientific). Separation of sialic acid compounds was achieved by Hydrophobic interaction liquid chromatography (HILIC) using a luna 5 μ HILIC-OH column with a gradient of 95% to 10% acetonitrile with 0.1% formic acid over 10 column volumes. Detection of compounds eluted from the HILIC column was achieved using a charged aerosol detector (CAD; Thermo Scientific). Injections of standards were analysed on an analytical size column using 5 μ l injections of 1 mg/ml solutions in H₂O or 10 mM or 100 mM sodium phosphate. Purification of 2,7-anhydro-Neu5Ac was achieved using a semi-preparative size column, injections were between 1 and 5 mg in 100 or 200 μ l H₂O.

2.5.2 Mass Spectrometry

For quantification of sialic acid derivatives, multiple reaction monitoring (MRM) transitions were determined by direct injection of sialic acid standards on a 4000 QTrap (Sciex) mass spectrometer and validated using a Xevo TQ-S Micro (Waters) (Table 3). Standards were prepared at 0.1 mg/ml in H₂O. Separation of sialic acid molecules was achieved either using a bioZen Glycan column (Phenomenex) or an ACQUITY UPLC BEH Amide Column (Waters). For both columns, a linear gradient of 95% to 40% acetonitrile containing 5 mM or 50 mM ammonium acetate was used. A volume of 5 μ l was used for injections of standards or caecal content (see below).

Table 3: Multiple reaction monitoring (MRM) transitions used in this work.

Compound	Parent Ion (Da)	Fragment ion (Da)
2,7-anhydro-Neu5Ac	290.1	87
2,7-anhydro-Neu5Ac	290.1	126
Neu5Ac	308	87
Neu5Ac	308	170
2-OMeNeu	322	87
2-OMeNeu	322	290

2.5.2.1 Caecal content preparation

Caecal content was collected from BI57/6J mice, 50 mg of caecal content was suspended in 1 ml 5% acetic acid and 1 μ l of 31 mM 2-OMeNeu was added to the sample as an internal standard. Samples were transferred to Lysing Matrix E tubes (MP Biomedicals) and homogenised in a FastPrep® Instrument (MP Biomedicals) 3 times for

20 sec at a speed setting of 6.0. To remove salts, hydrophobic compounds and oligosaccharides samples were loaded onto a cation exchange column packed in house (DOWEX 50w x8 H⁺), followed by a graphitised carbon column (EnviCarb 250 mg, Sigma Aldrich), both columns were eluted with a total of 5 ml 5% acetic acid. The remaining sugars were dried using a speedvac (Eppendorf) and resuspended in 100 μ l H₂O, 400 μ l acetonitrile was added to the samples prior to loading on the LC-MS.

2.6 Binding assays

2.6.1 Differential Scanning Fluorimetry (DSF)

DSF experiments were conducted in white 96 well plates (Grenier), 16 μ l protein (12 μ M stock) and 2 μ l of ligand (2,7-anhydro-Neu5Ac or Neu5Ac 1mg/ml) were mixed, added to the plate and incubated on ice for 10 min. Immediately before starting the experiment, 2 μ l SYPRO orange was added to each well. The reactions were run using a Biorad CFX96 Real-time PCR, with a 10 min 10°C hold followed by temperature increases of 0.5°C per 15 sec to 90°C with readings every 15 sec. Readings were performed using FRET with excitation at 470 nm and absorbance at 570 nm, results were analysed using the CFX Manager (BIO-Rad) and the temperature at which 50% of the protein was denatured exported to Excel for analysis.

2.6.2 Fluorescence spectroscopy

All protein fluorescence experiments were carried out using a FluoroMax 3 fluorescence spectrometer (Horiba) with a connecting water bath at 37°C. Due to the presence of 15 tyrosine residues, RgSBP was excited at 297 nm with slit widths of 5 nm. RgSBP was used at a concentration of 0.2 μ M in 50 mM Tris-HCl pH 7.5 for all fluorescence experiments. Initial binding was assed by sequential 10 μ M additions of 2,7-anhydro-Neu5Ac or Neu5Ac to a maximum of 70 μ M. Binding kinetics were determined by cumulative fluorescence changes from titration of the protein with 2,7-anhydro-Neu5Ac were plotted in GraphPad and fitted to a single rectangular hyperbola. The K_d values reported were averaged from three separate ligand titration experiments. Ligand displacement assays were performed using one addition of 10 μ M 2,7-anhydro-Neu5Ac, followed by 6 sequential additions of 10 μ M Neu5Ac or one addition of 10 μ M Neu5Ac, followed by 6 sequential additions of 10 μ M 2,7-anhydro-Neu5Ac.

2.6.3 Isothermal titration calorimetry (ITC)

ITC experiments were performed using the PEAQ-ITC system (Malvern, UK) with a cell volume of 200 μ l. Prior to titration, RgSBP was exhaustively dialysed into 50 mM Tris-HCl pH 7.5 and the ligands (2,7-anhydro-Neu5Ac or Neu5Ac) were dissolved in the dialysis buffer. The cell protein concentration was 100 μ M and the syringe ligand concentration was 2 mM. Controls with titrant (sugar) injected into the buffer only were subtracted from the data. The analysis was performed using the Malvern software, based on a single-binding site model. Experiments were carried out in triplicate.

2.6.4 Saturation Transfer Difference Nuclear Magnetic Resonance (STD-NMR)

For sample preparation, an Amicon centrifuge filter unit with a 10 kDa MW cut-off was used to exchange the protein in 25 mM d_{19} -2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol pH 7.4 (uncorrected for the deuterium isotope effect on the pH glass electrode) D₂O buffer and 50 mM NaCl. 2,7-anhydro-Neu5Ac and Neu5Ac were dissolved in 25 mM d_{19} -2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol pH 7.4, 50 mM NaCl. STD-NMR experiments were then performed by Ridvan Nepravishta (University of East Anglia, UK). Characterization of ligand binding by STD-NMR Spectroscopy (Mayer and Meyer, 1999) was performed on a Bruker Avance 800.23 MHz at 298 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 sec to 4 sec, for binding epitope mapping determination while only 0.5 sec of saturation time for each selected frequency was used to perform the DEEP-STD NMR experiments (Monaco *et al.*, 2017). The water signal was suppressed by using the excitation sculpting technique (Hwang and Shaka, 1995), while the remaining protein resonances were filtered using a T₂ filter of 40 ms. All spectra were performed with a spectral width of 10 KHz and 32768 data points using 256 or 512 scans. For RgSBP, due to the absence of a 3D structure, it was impossible to derive the resonances for saturation of aliphatic and aromatic residues found in the binding site as required by the DEEP-STD NMR technique. Moreover, RgSBP being a high molecular weight protein, the NMR spectra assignment is precluded. Therefore we adopted a search for druggable sites strategy using 4-hydroxy-1-oxy-2,2,6,6-tetramethylpiperidine (TEMPO) as recently described (Nepravishta *et al.*, 2019). Briefly, 1H-1H TOCSY spectra of the protein (500 μ M) were acquired in the presence or in the absence of TEMPO (2.5 mM and 12.5 mM). The spectra were performed with a spectral width of 10 kHz using a time domain of 2056 data points in the direct dimension and 32 scans. The indirect dimension was acquired using the non-uniform sampling (NUS) technique acquiring a NUS amount of 50% of the original 256

increments resulting in 64 hypercomplex points. The spectra were processed with the Topspin 3.1 compressed sensing routine. The final selected resonances were those identified by the TEMPOL PRE effect, and not overlapping with ligand signals. The DEEP-STD NMR data obtained were used to derive the average orientation of the ligand bound to *Rg*SBP by averaging the DEEP-STD factors obtained from each saturated region. The DEEP-STD NMR and binding epitope mapping analysis were performed using previously published procedures (Nepravishta *et al.*, 2019, Monaco *et al.*, 2017, Mayer and James, 2004).

2.7 Enzyme activity assays

2.7.1 Thin Layer Chromatography (TLC)

TLC was used to analyse *Rg*NanA enzymatic reactions with sialic acid derivatives. Briefly, purified *Rg*NanA (0.1 mg/ml) was incubated with 1 mg/ml Neu5Ac or 2,7-anhydro-Neu5Ac in 100 mM ammonium formate at 37°C overnight. Reaction products were concentrated using a concentrator plus speedvac (Eppendorf) and spotted onto a Silica gel 60 F₂₅₄ TLC plate (Merck) alongside ManNAc, Neu5Ac and/or 2,7-anhydro-Neu5Ac as references. The elution was performed in 2:1:1 Butanol: Acetic acid: H₂O, the plate was then dried and sprayed with the Orcinol solution (20 ml conc. Sulfuric acid, 150 ml 3,5-dihydroxytoluene (360 mg in EtOH), and 10 ml water), the plate was heated with a heat-gun until the sugars were revealed.

2.7.2 HPLC analysis of 2-AB labelled products

To quantify the amount of ManNAc produced during the *Rg*NanA catalysed enzymatic reaction, 2-AB labelling was carried out on the products from the above reactions. Briefly, samples were dried using a Concentrator Plus (Eppendorf) and 5 µl of labelling reagent was added and incubated at 65°C for 3 hours. The labelling reagent was prepared by dissolving 50 mg 2-aminobenzamide (2-AB) in a solution containing 300 µl acetic acid and 700 µl DMSO, followed by addition of 60 mg sodium cyanoborohydride. The sample was made up to 100 µl using H₂O, transferred to a HPLC vial and loaded onto a HyperClone 3u ODS (C18) 120A 150x4.6 mm 3 µ column. Mobile phases of 0.25% n-butylamine, 0.5% phosphoric acid, 0.1% tetrahydrofurane; 50% methanol; acetonitrile and H₂O were used at a 0.7 ml/min flow rate.

2.7.3 HPLC analysis of Diamino-4,5-methylenedioxybenzene dihydrochloride (DMB) labelled reaction products

To assay RUMGNA_02695 potential enzymatic activity, the purified recombinant protein was incubated in 100 μ l reactions at 37 °C overnight with 1 mM 2,7-anhydro-Neu5Ac in 50 mM sodium phosphate buffer pH 7.0, in the presence or absence of 500 μ M NADH. The reactions were dried using a Concentrator Plus (Eppendorf) for 1 h. Samples were then resuspended in 50 μ l of water and 50 μ l of DMB-reaction buffer and incubated for 2 hours at 55°C in the dark. Ten vol. of DMB-reaction buffer contained 1.74 mg of DMB (Carbosynth, UK), 324.6 μ l MilliQ water, 88.6 μ l glacial acetic acid, 58.2 μ l of β -mercaptoethanol and 79.3 μ l of sodium hydrosulphite. The samples were then centrifuged for 1 min and filtrated using a 0.45 μ m filter into a glass HPLC vial and directly analysed by HPLC as follows.

DMB-labelled samples were analysed by injecting 10 μ l onto a Luna 5 μ m C-18(2) LC column 250x4.6 mm (Phenomenex) at 1 ml/min. Mobile phases methanol/acetonitrile/water were used for separation of fluorescently labelled sialic acids. The settings of the fluorescence detector were 373 nm excitation, 448 nm emission. Samples were run alongside a 100 ng/ml Neu5Ac standard.

2.7.4 Kinetic enzymatic assays

The enzymatic activity of *RgNanA* was determined using a coupled enzymatic assay. Each reaction was performed at 37 °C in a total volume of 100 μ l and contained 150 μ M NADH, 0.5 U LDH, 20, 10, 5, 4, 2, 1, 0.4, 0.2, 0.1 mM Neu5Ac and 1.5 μ g purified *RgNanA* in 50 mM sodium phosphate buffer (pH 7.0). The decrease in absorbance at 340 nm was measured as NADH is oxidised to NAD⁺ by lactate dehydrogenase (LDH) as pyruvate is produced during the conversion of sialic acid to ManNAc by *RgNanA*. The reactions were monitored using FLUOstar OPTIMA (BMG LABTECH) and the initial rate of reaction determined for each Neu5Ac concentration in triplicate before analysis was performed by fitting the data to a Michaelis-Menten using Graph Pad Prism (V 5.03). To determine the kinetic parameters of RUMGNA_02695 enzymatic reaction, the same coupled reaction was used using an excess of *RgNanA* (15 μ g) and 10 μ g RUMGNA_02695 in each reaction. For the kinetics assays, 1, 0.4, 0.2, 0.1, 0.04, 0.02 and 0.01 mM 2,7-anhydro-Neu5Ac was used and the initial rate of reaction determined for each concentration in triplicate before analysis was performed by fitting the data to a Michaelis-Menten using Graph Pad Prism (V 5.03).

2.7.5 NMR monitoring of the RUMGNA_02695 (*RgNanOx*) reaction

All experiments were performed by Serena Monaco (University of East Anglia, UK) using a Bruker Avance I 500MHz spectrometer with a 5 mm PATXI 1H/D-13C/15N Z-GRD probe at 293 K. To follow the kinetics of the reaction and assess the position of deuteration, two samples containing 2 mM 2,7-anhydro-Neu5Ac, 100 μ M NADH and 60 μ M *RgNanOx* were used, one in PBS deuterated buffer (PBS/D₂O) and one in PBS standard buffer (PBS/H₂O, still containing 10% D₂O for locking purposes). The reaction was followed by 1D NMR acquiring experiments at 15 min intervals, over 24 hours. The reaction reached an equilibrium after about 18 hours, for both samples. The equilibrium ratio was 1:0.16:0.09 for Neu5Ac:2,7-anhydro-Neu5Ac: “xy intermediate”, as defined by integration of the respective acetyl methyl group signals. The standard zg pulse sequence was used for the D₂O sample, while excitation sculpting was used for the H₂O sample (pulse sequence: zgesgp). 1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS, Sigma) was added to each sample as an internal reference and calibrated to 0 ppm.

To characterise the “xy intermediate”, a new sample containing 3 mM 2,7-anhydro-Neu5Ac, 100 μ M NADH and 15 μ M *RgOx* in PBS/D₂O was prepared, to slow down the reaction enough to acquire an HSQC (hsqcetgpsi) in order to capture the intermediate.

2.8 ClosTron mutagenesis

R. gnavus mutants were generated using the ClosTron methodology (Heap *et al.*, 2010), which inserts an erythromycin resistance cassette into the gene of interest. Target sites were identified using the Perutka method (Perutka *et al.*, 2004). The re-targeted introns were synthesised and ligated into the pMTL007C-E2 vector by ATUM (MenloPark, USA). The plasmids were then transformed into *E. coli* CA434 using the heat-shock protocol (as described in 2.3.8.1), and the recombinant clones selected for chloramphenicol resistance. Recombinant *E. coli* cells were grown overnight in 10 ml LB, 1 ml of the overnight culture was pelleted and washed with PBS. The *E. coli* cell pellet was resuspended in 200 μ l of an *R. gnavus* overnight culture and the cell suspension spotted onto a non-selective BHI-YH plate. Following incubation for 8 hours at 37°C the bacteria were washed from the plate using PBS and plated onto BHI-YH supplemented with 250 μ g/ml cycloserine (Sigma) and 15 μ g/ml thiamphenicol (Sigma) and grown for 72 hours to select against *E. coli* and for transfer of the plasmid to *R. gnavus*. Individual colonies were grown in non-selective BHI-YH broth overnight to allow expression of the plasmid and genomic recombination. The culture was then plated onto a BHI-YH medium

containing 250 µg/ml cycloserine and 10 µg/ml erythromycin (Sigma) to select clones with successful genomic recombination. PCR and sequencing were used to confirm recombination in the gene of interest, as described in section 2.3.4.

2.9 Batch cultures

Batch culture experiments were performed using fresh faecal material from participants recruited onto the QIB Colon Model study (described below). The criteria for participation are healthy adults aged 18 or over within 10 miles of the Norwich Research Park. Eligible participants displayed a normal bowel habit, regular defecation between three times a week and three times a day, with an average stool type of 3 – 5 on the Bristol Stool Chart, and no diagnosed chronic gastrointestinal health problems, such as irritable bowel syndrome, inflammatory bowel disease, or coeliac disease. Participants had not taken antibiotics or probiotics within the last four weeks, had not experienced gastrointestinal complaints, such as vomiting or diarrhoea, within the last 72 hours, and had not recently returned from travel. The study was approved by the Quadram Institute Bioscience Human Research Governance committee (IFR01/2015), and London – Westminster Research Ethics Committee (15/LO/2169). The informed consent of all participating subjects was obtained, and the trial is registered at <http://www.clinicaltrials.gov> (NCT02653001).

For sample preparation, ten grams of fresh faecal sample was diluted 1:10 in pre-reduced PBS, and homogenised using a Stomacher 400 (Seward, UK) at 230 rpm for 45 sec. The working volume of each vessel was 50 mL made up of, 45 ml of pre-reduced batch culture media supplemented with appropriate carbon source (see section 2.1.1.7) and 5 mL of processed faecal inocula. The pH was controlled and maintained between 6.8–7.2 using pH control units Fermac 260 (Electrolab, United Kingdom), connected to 1 M NaOH and 1 M HCL solutions. Vessels were kept at 37 °C by a circulating water jacket and anaerobiosis was maintained by continuous bubbling of the system with oxygen-free nitrogen gas. 2 ml samples were taken at 0, 8, and 24 hours from each vessel and the bacteria pelleted by centrifugation at 6,000 g for 5 min, the supernatant was removed, and the pellet stored at -80 °C until DNA extraction (see section 2.3.1.2).

2.10 Mouse studies

The impact of the *nan* cluster on *R. gnavus* fitness was assessed by comparing the ability of *R. gnavus nan* mutant (see section 2.8) or wild-type strain to colonise germ-free C57BL/6J mice. A group of four 7-9 week old germ-free mice were gavaged with 10⁸

CFU of *R. gnavus* ATCC 29149 wild-type or antisense *nan* mutant in 100 µl PBS, individually or in combination. Faecal samples were collected from each mouse at 3, 7 and 14 days post gavage. Following sacrifice by rising concentration of CO₂ gas and cervical dislocation, the caecal content was collected for DNA extraction and RNA extraction in RNA later (see section 2.3.2.2). Sections of colon were also taken into RNA later (see section 2.3.2.3) or fixed in methacarn (60% dry methanol, 30% chloroform and 10% acetic acid) for FISH analysis (see section 2.10.1).

The impact of *R. gnavus nan* mutant on *S. Typhimurium* infection was assessed using six groups of five 7-9 week old BL57/6J mice. Mice were orally gavaged with 20 mg streptomycin (Sigma, USA) at day 0. On day 1 mice were gavaged with either 10⁸ *R. gnavus* ATCC 29149, 10⁸ *R. gnavus nan* mutant or PBS and on day 2 with 10⁸ *S. Typhimurium* SL1344 or PBS as detailed in section 6.2.3. Faecal samples and welfare checks were taken daily. Welfare checks used a 0 (no signs) – 5 (severe signs) scale to assess symptoms of ill health. These included piloerection, abnormal respiration, oculonasal discharge, tremors and/or convulsions, hunched, self-mutilation, intermittent vocalisation and peer interaction. The weight of each mouse was also monitored and any mouse trending towards a 20% weight loss were sacrificed immediately. Following sacrifice by rising concentration of CO₂ gas and cervical dislocation, caecal content was collected for DNA extraction. Mucus was scrapped from the small intestine into 6M guanidium chloride (GuCl) and stored at 4°C until MUC 2 purification, the mucus from all five mice in each group was pooled (see section 2.10.2).

Care and treatment of animals was in accordance with guidelines and approval by the University of East Anglia Disease Modelling Unit, following the 3Rs principles. The assessment of *R. gnavus* fitness in germ-free mice was performed under the project licence NB 70/8929 and the *S. Typhimurium* infection experiment under project licence RK 70/08957.

2.10.1 Fluorescent in situ hybridisation (FISH)

For FISH analysis, the colonic tissue was fixed in methacarn, was processed and embedded in paraffin as previously described (Johansson *et al.*, 2011b). Tissue sections were then prepared at 8-10 µm and subsequent analysis performed by Laura Vaux in the group. Paraffin sections were dewaxed and washed in 95% ethanol. The tissue sections were incubated with 100 µl of Alexa Fluor 555-conjugated Erec482 probe (5' – GCTTCTTAGTCARGTACCG -3') at a concentration of 10 ng/µl, in hybridisation buffer (20 mM Tris-HCl, pH 7.4, 0.9M NaCl, 0.1% SDS) at 50°C overnight. The sections were

then incubated in a 50°C prewarmed wash buffer (20m M Tris-HCl, pH 7.4, 0.9 M NaCl) for 20 min. All subsequent steps were performed at 4°C. The sections were washed with PBS, then blocked with TNB buffer (0.5% w/v blocking reagent in 100 mM Tris-HCl, pH 7.5, 150 mM NaCl) supplemented with 5% goat serum. To detect mucin, the sections were then counterstained with an anti-Muc2 antibody (sc-15334; Santa Cruz biotechnology, USA) at 1:100 dilution in TNB buffer overnight. The sections were washed in PBS, then goat anti-rabbit antibodies (diluted 1:500; Sigma, USA) were used for immunodetection. The sections were counterstained with Sytox blue (S11348, ThermoFisher, USA) diluted 1:1000 in PBS and mounted in Prolong gold anti-fade mounting medium (ThermoFisher, USA). The slides were imaged using a Leica TCS SP2 confocal microscope with a x63 objective. The distance between the leading front of bacteria and the base of the mucus layer was measured with FIJI (Schindelin *et al.*, 2012). A total of 70 images from 8 mice were analysed. The association between genotype and distances was estimated by a linear mixed model, including fixed effects of genotype and area and random effects of mouse and each individual image. There was substantial spatial correlation between adjacent observations and so an AR (1) correlation structure was added. The resulting model had no residual autocorrelation as judged by visual inspection of autocorrelation function. The nmle package version 3.1-137 using R version 3.5.3 was used to estimate the model.

2.10.2 Glycan analysis of MUC2

Mouse mucus from the small intestine was collected as described in section 2.10. To separate MUC2 from other mucins the GuCl containing mouse mucus were centrifuged at 12,000 rpm at 4°C for 30 min to pellet the MUC2. MUC2 samples were prepared for glycan analysis by Dimitris Latousakis as previously described (Leclaire *et al.*, 2018). Following separation from other mucins by centrifugation, the MUC2 samples were subjected to β -elimination and desalting followed by permethylation. Samples were analysed using Matrix-assisted laser desorption and ionization–time of flight mass spectrometry (MALDI-TOF MS) and TOF/TOF-MS data were acquired with the Autoflex analyser mass spectrometer (Bruker, UK) in the positive-ion and reflectron mode.

Chapter 3

2,7-anhydro-Neu5Ac synthesis and detection methods

3.1 Introduction:

The term sialic acid covers a diverse range of nine-carbon acidic monosaccharides commonly found in the terminal location of glycan chains of proteins and lipids (Varki, 2008). Sialic acids have been implicated in many biological processes including cell recognition and interaction, pathogenic infection and tumour progression (Li and Ding, 2019, Cornelissen *et al.*, 2019, Haines-Menges *et al.*, 2015, Schauer, 2009). For instance, some pathogens can utilise sialic acid to evade the host immune system, whereas other pathogens seek out free sialic acid in the gut environment to use as a nutrient source (Langereis *et al.*, 2015, Ng *et al.*, 2013). The most common form of sialic acid in humans is Neu5Ac, indeed the term sialic acid is often used to refer to Neu5Ac. However, it is known that more than 50 forms of sialic acid exist in nature, each of these being a modification of one of the three basic sialic acids, Neu5Ac, Neu5Gc or KDN, distinguished by their C5 substitution. Further modifications such as additional acetylation, lactylation, methylation and sulfation gives rise to the diversity of sialic acids (see section 1.5).

The action of IT-sialidases add another layer of complexity by releasing 2,7-anhydro forms of sialic acid from sialylated substrates, unlike the more common hydrolytic sialidases which release Neu5Ac. The bacterial IT-sialidase enzymes characterised to date, from *S. pneumoniae* and *R. gnavus*, show specificity for α 2-3 linked sialic acids, whereas hydrolytic sialidases are described as active against α 2-3,4,6 or 8 linked residues (Tailford *et al.*, 2015b, Xu *et al.*, 2011).

Earlier work reported the presence of 2,7-anhydro-Neu5Ac in rat urine and human ear wax and it was suggested to be bactericidal and/or serve as a reservoir of sialic acids (Monestier *et al.*, 2017, Suzuki *et al.*, 1985). However, the importance of 2,7-anhydro-Neu5Ac in biological systems, has not been investigated. Unlike other sialic acids it is not available commercially and methods to synthesise the product have not proven effective. This has limited research into 2,7-anhydro-Neu5Ac compared to Neu5Ac which has been extensively studied and for which the bacterial metabolic pathways are known (Brigham *et al.*, 2009, Plumbridge and Vimr, 1999).

With the recent discovery of an IT-sialidase in the gut symbiont *R. gnavus* (RgNanH) (Tailford *et al.*, 2015b) it has been proposed that the enzyme plays a role in *R. gnavus* adaptation in the gut. The *R. gnavus* IT-sialidase targets the terminal sialic acid residues of mucin glycans, a valuable nutrient source in the gut mucosal niche. We proposed that 2,7-anhydro-Neu5Ac is a metabolic substrate for *R. gnavus* and may give this bacterium

a competitive advantage over other microbes (Crost *et al.*, 2016). To confirm this hypothesis and to then investigate the metabolic process by which 2,7-anhydro-Neu5Ac is utilised, mg - g quantities of the substrate are required. A further aim will be to analyse the level of 2,7-anhydro-Neu5Ac in the gut of caecal content of mice.

The chemical synthesis of 2,7-anhydro-Neu5Ac has been previously reported following chemical or enzymatic protocols. Lifely and Cottee showed that intramolecular glycosylation of sialic acid under methanolysis conditions can be used to give rise to the methyl ester of 2,7-anhydro-Neu5Ac (Lifely and Cottee, 1982). Later, the first direct synthesis of 2,7-anhydro-Neu5Ac was described by intramolecular glycosidation of the S-(1-phenyl-1H-tetrazol-5-yl) glycoside derivative of Neu5Ac (Kimio *et al.*, 1991). More recently, the chemical synthesis of 2,7-anhydro-Neu5Ac from the methyl ester in three steps (per-O-trimethylsilylation, intramolecular anomeric protection and desilylation) has been described (Asressu and Wang, 2017). An alternative approach to chemical synthesis is to use an IT-sialidase enzyme to liberate the 2,7-anhydro-Neu5Ac from sialylated substrates. This was demonstrated by Li *et al.* using 2'-(4-methylumbelliferyl)- α -D-N-acetylneuraminic acid (MU-NANA) as the substrate for a leech IT-sialidase (Li *et al.*, 1990). Despite the ability of these methods to produce 2,7-anhydro-Neu5Ac, the cost associated with the respective reagents and/or substrates prohibits synthesis on a large scale, therefore preventing the biological impact of 2,7-anhydro-Neu5Ac to be evaluated. New methods to produce high quantities of 2,7-anhydro-Neu5Ac are described here.

In addition, a further challenge of working with 2,7-anhydro-Neu5Ac over the lack of availability of the product is the difficulty in detecting the compound. It has no natural absorbance or fluorescence properties that can be exploited and furthermore it cannot be labelled by 1,2-Diamino-4,5-methylenedioxybenzene (DMB). Most sialic acids can be DMB-labelled allowing their detection by fluorescence, however, this requires the C2 position to be available (Figure 15a). While in Neu5Ac the C2 position is free and available (Figure 15b), in 2,7-anhydro-Neu5Ac the C2 is occupied by the C2-C7 ether bond (Figure 15c), making it inaccessible to the DMB labelling.

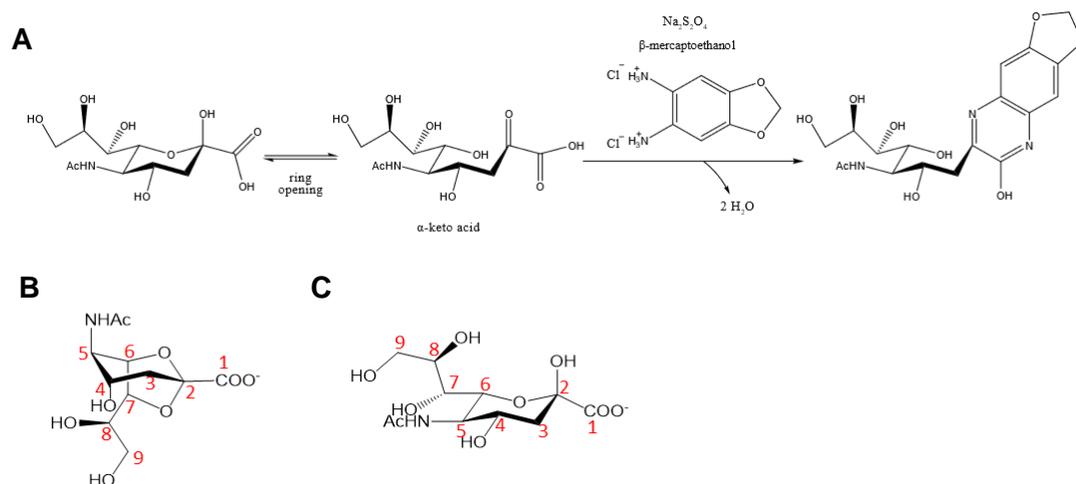


Figure 15: DMB labelling of sialic acid derivatives

A) Reaction mechanism involved in DMB labelling of sialic acids. The reaction results in the addition of DMB at the C1 and C2 positions of Neu5Ac. **B)** Structure of 2,7-anhydro-Neu5Ac with carbon atoms numbered in red, showing that the C2 position is occupied by the C2, C7 anhydro bridge. **C)** Structure of Neu5Ac with carbon atoms numbered in red.

All synthesis work relies on Electrospray ionisation – mass spectrometry (ESI-MS), NMR and Refractive index measures for detection which requires specialised equipment and good standards to ensure accurate peak detections. If peak separation is not optimal these detection methods can also be masked by other compounds. This chapter includes the development of methods that can be used to both monitor and quantify 2,7-anhydro-Neu5Ac in biological systems.

3.2 Results

3.2.1 Enzymatic synthesis of 2,7-anhydro-sialic acid derivatives:

We first used purified recombinant *RgNanH* to produce 2,7-anhydro-Neu5Ac from MU-NANA substrate (Tailford *et al.*, 2015b). The product was then purified using three rounds of precipitation and folch partitioning as previously reported (Crost *et al.*, 2016). However, the yield was very low, the protocol labour-intensive and the price of MU-NANA used as substrate prohibitive.

Next, a membrane-enclosed multienzyme (MEME) method was developed using fetuin as a substrate. Fetuin was chosen as the substrate as it is a commercially available large glycoprotein containing approximately 8% sialic acid by weight. Being a large glycoprotein means that it can be enclosed within a dialysis membrane along with the recombinant *RgNanH* protein. The small monosaccharide product is free to diffuse across the membrane where it is more easily recovered and separated from the substrate (Figure 16a). The product was then purified by size exclusion chromatography on a Biogel P2 column to remove salts. The purity of the obtained product was assessed by ESI-MS and ¹H and ¹³C NMR spectroscopy (Figure 16bc). The spectra obtained from these methods confirmed the nature of the 2,7-anhydro-Neu5Ac product, however, they also revealed that the product contained around 17% Neu5Ac. As Neu5Ac remained after purification on a P2 column, it was clear that this step was not sufficient to separate 2,7-anhydro-Neu5Ac and Neu5Ac, due to their very similar molecular weights.

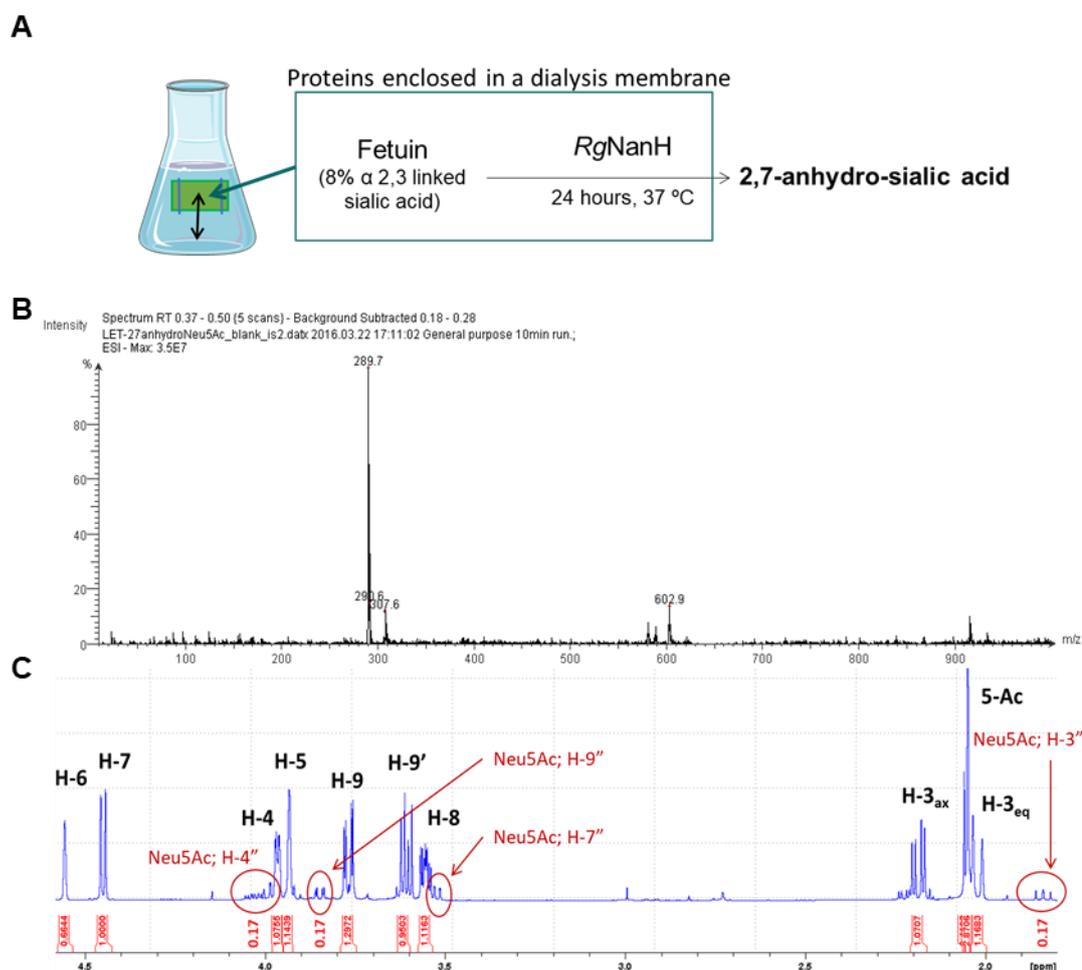


Figure 16: Synthesis of 2,7-anhydro-Neu5Ac

A) Schematic representation of the 2,7-anhydro-Neu5Ac synthesis method. *RgNanH* releases 2,7-anhydro-Neu5Ac from the membrane enclosed fetuin. The 2,7-anhydro-Neu5Ac is able to cross the membrane for simple purification from the substrate. **B)** ESI(-)-MS spectra of the reaction product following purification on a Biogel P2 column, showing 2,7-anhydro-Neu5Ac (289.7) as the major product with ~17% Neu5Ac bi-product (307.6). **C)** NMR spectra (600 MHz) of the reaction product showing 2,7-anhydro-Neu5Ac as the major product, denoted by black annotations, and ~ 17% Neu5Ac by-product denoted by red annotation.

To remove the Neu5Ac by-product, a sialic acid aldolase was added to the membrane enclosed reaction (Figure 17a). We used a commercial *E. coli* sialic acid aldolase (*EcNanA*) which converts Neu5Ac into ManNAc and pyruvate, but does not act on 2,7-anhydro-Neu5Ac. The smaller ManNAc and pyruvate products separate more readily from the 2,7-anhydro-Neu5Ac in size exclusion chromatography. This additional step yielded 2,7-anhydro-Neu5Ac at 96% purity and the Neu5Ac trace was limited to <1%, with the remaining contaminants identified as protein residues, as assessed by ESI-MS and NMR (Figure 17b,c). This methodology now published (Monestier *et al.*, 2017) is an efficient, low cost method accessible to research labs to produce up to 20 mg of 2,7-anhydro-Neu5Ac and other 2,7-anhydro-sialic acid derivatives including Neu5Gc and

KDN. The development and optimisation of the method was led by Marie Monestier from our group, I contributed to the enzyme production and the set up of the biogel-P2 column purification process in our lab.

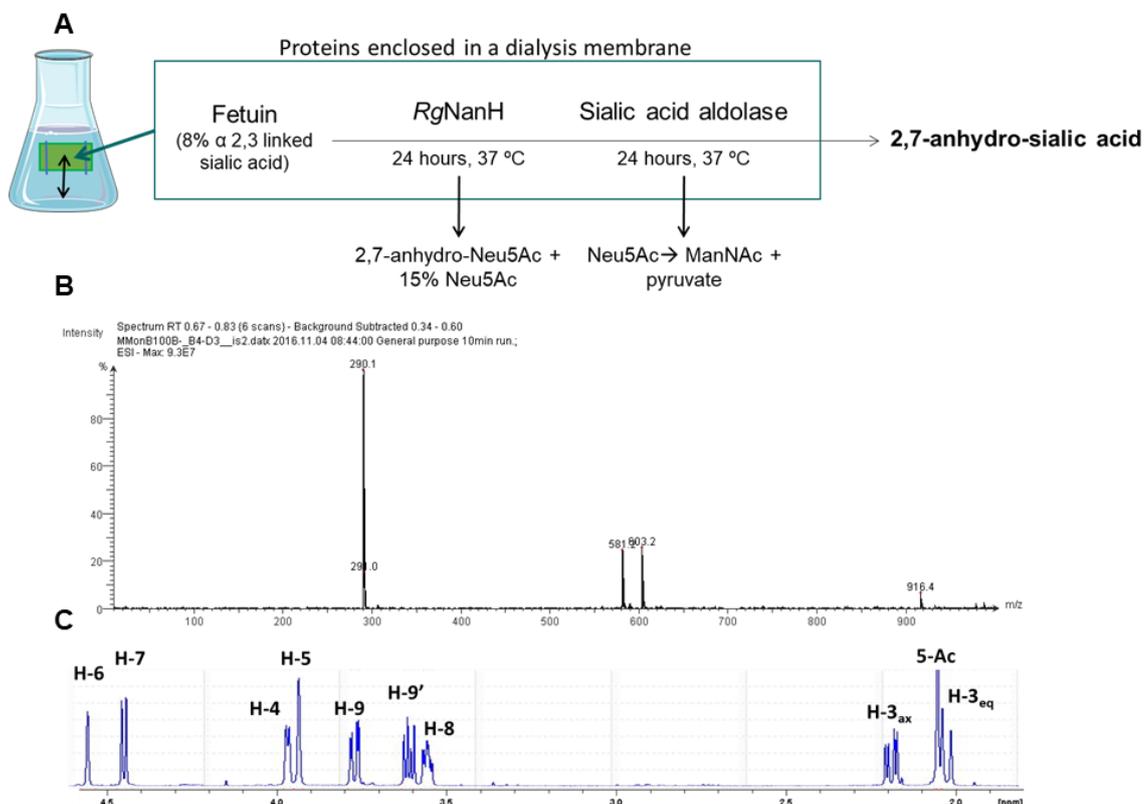


Figure 17: Synthesis of 2,7-anhydro-Neu5Ac with addition of a sialic acid aldolase

A) Schematic representation of the 2,7-anhydro-Neu5Ac synthesis method. RgNanH releases 2,7-anhydro-Neu5Ac as the major product and the Neu5Ac by-product from the membrane enclosed fetuin. Neu5Ac is further degraded to ManNAc and pyruvate by the action of the sialic acid aldolase, which is inactive against 2,7-anhydro-Neu5Ac. **B)** ESI(-)-MS spectra of the reaction product following purification on a Biogel P2 column, showing 2,7-anhydro-Neu5Ac (290.1) with no Neu5Ac visible. **C)** NMR spectra (600 MHz) of the reaction product showing 2,7-anhydro-Neu5Ac and no Neu5Ac.

Following the publication of our synthesis method, Dr. Xi Chen's group (Department of chemistry, University of California) further optimised the method (Xiao *et al.*, 2018). Using a One-Pot Multienzyme (OPME) strategy, Neu5Ac was first synthesised from ManNAc and pyruvate using a sialic aldolase from *Pasteurella multocida* (PmNanA). The Neu5Ac was then activated by a CMP-sialic acid synthase in the presence of CTP (cytidine triphosphate), then transferred to lactose to form 3'SL; the 3'SL was then used as a substrate for the IT-sialidase from *S. pneumoniae* (SpNanB), releasing the 2,7-anhydro-Neu5Ac and also regenerating the lactose acceptor (Figure 18). This method allowed the production of gram quantities of 2,7-anhydro-Neu5Ac which we obtained in

collaboration with Xi Chen's group, however, we found small traces of Neu5Ac. This was first identified by low-level activity of a Neu5Ac specific sialic acid aldolase against the compound (Figure 19a), and later confirmed by ESI-MS to be approximately 2% (Figure 19b).

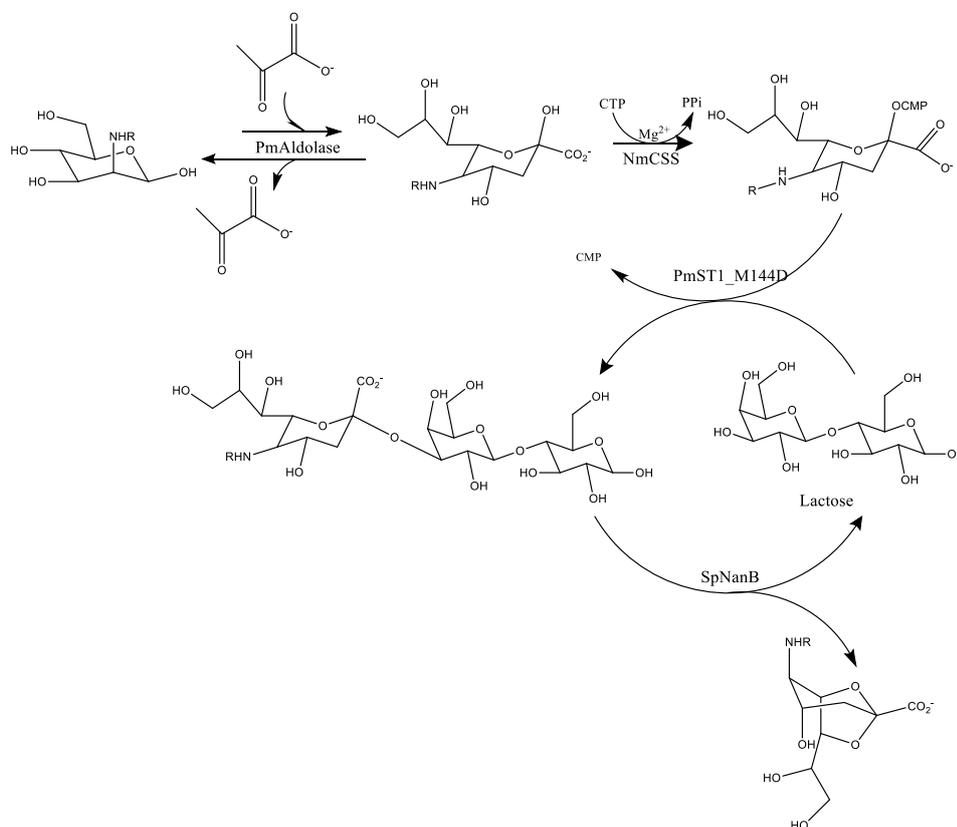


Figure 18: Schematic representation of the One-Pot Multienzyme (OPME) 2,7-anhydro-Neu5Ac synthesis

Briefly, ManNAc and pyruvate are converted to Neu5Ac by the *P. multocida* aldolase, this is then activated by a CMP-sialic acid synthases (CSS) and attached to lactose by a sialyltransferase to produce 3'SL. The activity of the *S. pneumoniae* IT-sialidase is then used to produce 2,7-anhydro-Neu5Ac.

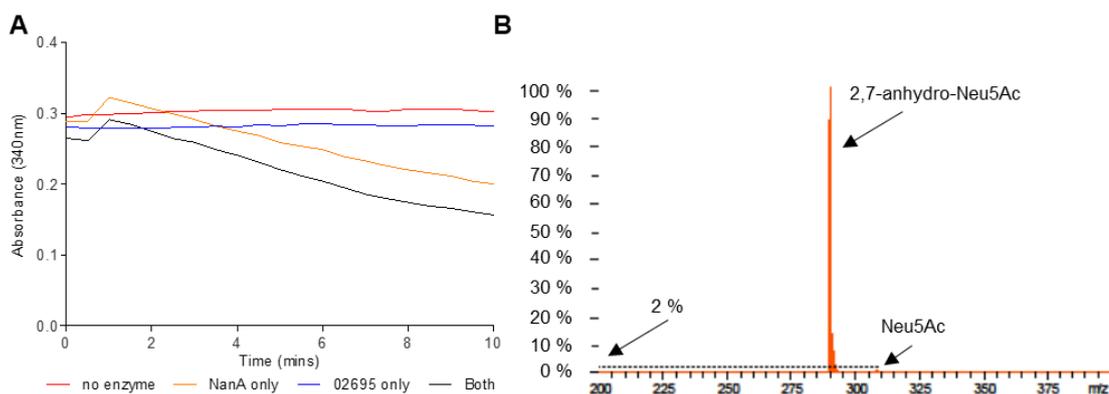


Figure 19: Assessment of Neu5Ac contamination in samples of 2,7-anhydro-Neu5Ac

A) Reactions of 2,7-anhydro-Neu5Ac provided by Xi Chen with *RgNanA* (*R. gnavus* sialic acid aldolase), detection is by coupling the pyruvate produced to a lactate dehydrogenase enzyme which further converts pyruvate to lactate and the associated conversion of NADH (absorbance at 340 nm) to NAD⁺ (no absorbance at 340 nm) is recorded. *RgNanA* has been shown to be inactive against 2,7-anhydro-Neu5Ac and active against Neu5Ac (see section 4.2.7). *RgNanA* showed activity against this 2,7-anhydro-Neu5Ac batch suggesting the presence of Neu5Ac. **B**) ESI(-)-MS spectra of the 2,7-anhydro-Neu5Ac batch showing 2,7-anhydro-Neu5Ac (289.7) as the major peak with ~ 2% Neu5Ac (307.8).

This level of contamination prompted us to develop a new purification method based on the use of Hydrophobic interaction liquid chromatography (HILIC), and detection of 2,7-anhydro-Neu5Ac by a charged aerosol detector (CAD). The method used a gradient from high to low percentage acetonitrile with 0.1% formic acid to maintain the pH. A peak corresponding to the contaminating Neu5Ac could not be seen in the 2,7-anhydro-Neu5Ac sample (Figure 20a) and spiking the sample with Neu5Ac showed that separation was sufficient to allow purification (Figure 20b). The peaks downstream of Neu5Ac are likely to correspond to salts coming from the original synthesis and purification. The peak corresponding to 2,7-anhydro-Neu5Ac was broad with a long tail which at high concentrations begins to overlap with Neu5Ac (Figure 20c). As a consequence, only small amounts (< 1 mg) of product could be loaded at a time resulting in a long purification process. Further efforts are needed to refine the 2,7-anhydro-Neu5Ac purification and achieve a more efficient separation while maintaining a baseline separation from Neu5Ac even at high concentrations.

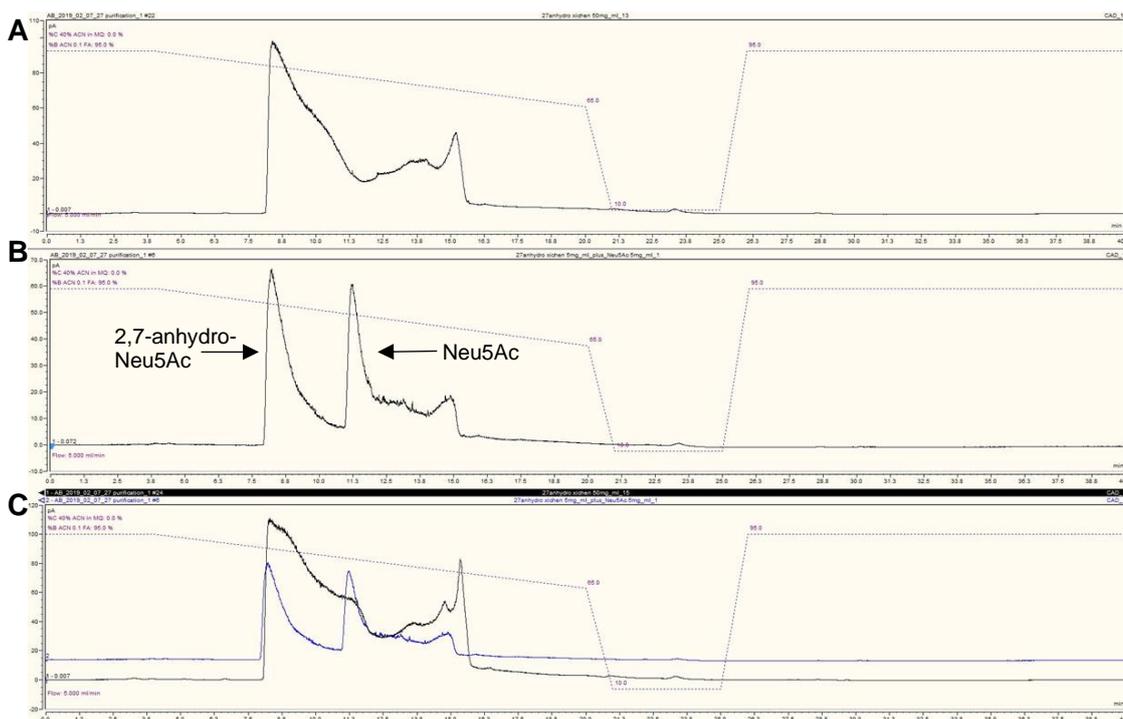


Figure 20: HPLC purification of 2,7-anhydro-Neu5Ac

A Luna 5 μ HILIC-OH column (Phenomenex) was used for the separation of peaks corresponding to 2,7-anhydro-Neu5Ac and Neu5Ac are indicated by the arrows. **A)** 0.5 mg injection of 2,7-anhydro-Neu5Ac (Xi Chen batch). **B)** 0.5 mg injection of 2,7-anhydro-Neu5Ac (Xi Chen batch) with 0.5 mg Neu5Ac. **C)** 5 mg injection of 2,7-anhydro-Neu5Ac (Xi Chen batch) with an overlay of **B**.

3.2.2 Detection and quantification of 2,7-anhydro-Neu5Ac

HILIC technology has proved capable of separating 2,7-anhydro-Neu5Ac from Neu5Ac, however, the CAD detection method provides limitations with regards to the buffers that can be used, as salts are strongly detected by CAD. Although standards of 2,7-anhydro-Neu5Ac, Neu5Ac and ManNAc showed clear separation (Figure 21abc), in the presence of buffers such as sodium phosphate, the Neu5Ac peak was obscured, and much larger peaks for salts were visible. In 100 mM sodium phosphate, the Neu5Ac peak almost doubled in size and the contaminant peak was also much higher than in the standards (Figure 21d). In 10 mM sodium phosphate, the Neu5Ac and contamination peaks overlapped each other making quantification impossible (Figure 21e). Due to CAD being able to detect any molecule that passes through the detector, it is difficult to confidently identify the peaks from mixed samples such as mouse caecal content (Figure 21f).

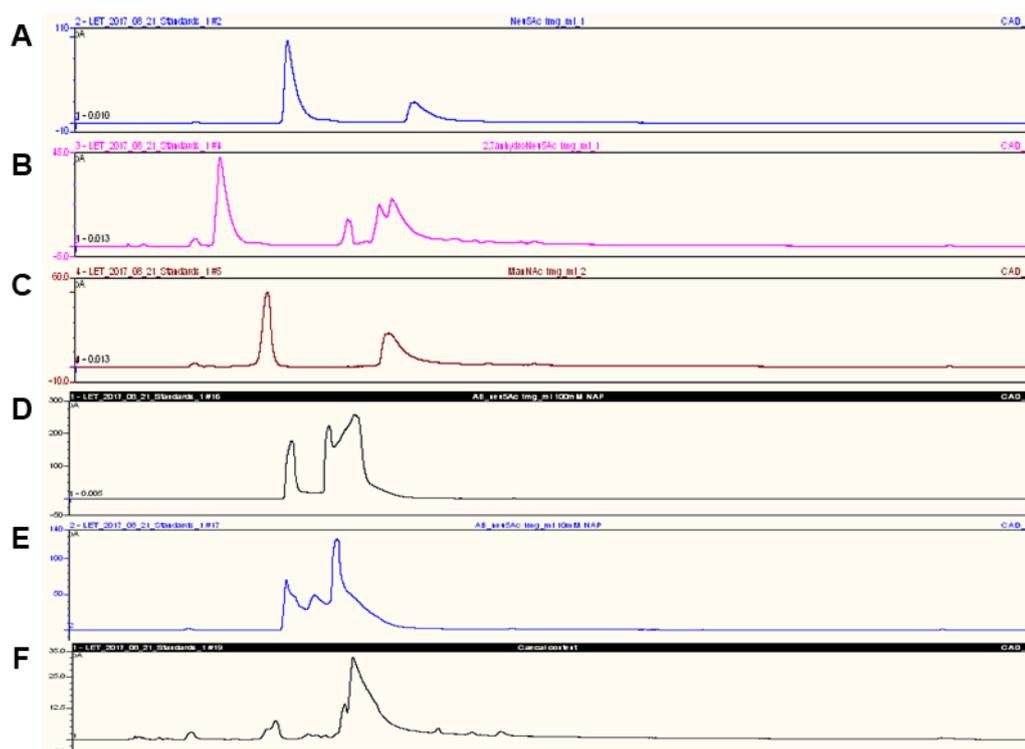


Figure 21: HPLC analysis of sialic acids using a Luna 5 μ HILIC-OH column (Phenomenex) and detection by CAD

A) 1 mg/ml Neu5Ac in H₂O. **B)** 1 mg/ml 2,7-anhydro-Neu5Ac in H₂O. **C)** 1 mg/ml ManNAc in H₂O. **D)** 1 mg/ml Neu5Ac in 100 mM Sodium phosphate. **E)** 1 mg/ml Neu5Ac in 10 mM Sodium phosphate. **F)** Caecal content from a mouse orally gavaged with *R. gnavus* ATCC 29149.

As a result of these limitations, a new detection method was developed based on the use of ESI-MS. The advantage of this method over CAD is that the exact mass of the product eluting from the column can be determined. Furthermore, the product can be ionised, and the mass of individual ions detected. Knowledge of the fragmentation pattern of the different target molecules, which can be determined from standards, provides a higher degree of confidence in the chemical nature of the molecule being detected. In addition, it is possible to generate traces for individual ions, co-eluting with a different molecular mass or different ionisation patterns which would, in principle, allow us to remove these compounds from quantification analysis. Multiple reaction monitoring (MRM) transitions for all the sialic acids under investigation were determined on a 4000 QTrap (Sciex, USA) mass spectrometer following direct injection of each sialic acid standard separately and validated on a Xevo TQ-S Micro (Waters, USA) (Table 3). These were used to detect and quantify the different sialic acids by LC-MS (liquid chromatography – mass spectrometry).

As with the HPLC-CAD method described above, a linear gradient of high to low percentage acetonitrile was applied with a bioZen Glycan column (Phenomenex, USA). The method was trialled in buffers containing either 5 mM or 50 mM ammonium acetate as suggested by the manufacturer. Both methods allowed for separation of 2,7-anhydro-Neu5Ac from Neu5Ac. With 5 mM ammonium acetate, the sialic acids separated very well, however, they eluted at the lowest acetonitrile concentration used, suggesting that an isocratic elution may improve the separation process (Figure 22a). The sialic acid compounds also separated in 50 mM ammonium acetate along the acetonitrile gradient (Figure 22b). However, in both cases, the peak for Neu5Ac was found to be very broad making accurate quantification challenging.

A second column, ACQUITY UPLC BEH amide column, was also tested, again using the same high to low percentage acetonitrile gradient. The use of a UHPLC column is restricted to specific instruments but enabled a very good separation of 2,7-anhydro-Neu5Ac and Neu5Ac (Figure 22c). The resolution of the peak corresponding to Neu5Ac also appeared to be improved compared to the other two columns.

For analysis of crude biological samples by LC-MS, *N*-Acetyl-2-*O*-methyl- α -D-neuraminic acid (2-OMeNeu) was chosen as an internal standard as it is a sialic acid that does not occur naturally and therefore can be used as a reference for quantification from biological samples. To remove salts and other contaminants from caecal contents, a clean-up process was used. This involved desalting on an in-house packed cation exchange column (DOWEX 50w x8 H⁺) and exclusion of hydrophobic compounds and oligosaccharides using graphitised carbon cartridges, to give a clean reaction product. The same caecal content sample was analysed using both ACQUITY UPLC BEH Amide and bioZen Glycan columns and the buffer compositions described above. As with analysis of the standards, separation could be seen between Neu5Ac, 2,7-anhydro-Neu5Ac and 2-OMeNeu. The signal for 2-OMeNeu was very inconsistent between samples, with it not detected in a number of samples (Figure 23). It is possible that this compound interacts strongly with the columns used in the purification process. From crude samples, the resolution of the Neu5Ac peak was poor using the bioZen glycan column with 5 mM ammonium acetate or with the ACQUITY UPLC BEH Amide column (Figure 22d,f). The sharpness and resolution of the peak is important for accurate quantification. Using 50 mM ammonium acetate with the bioZen Glycan column, the Neu5Ac peak was much better defined (Figure 22e).

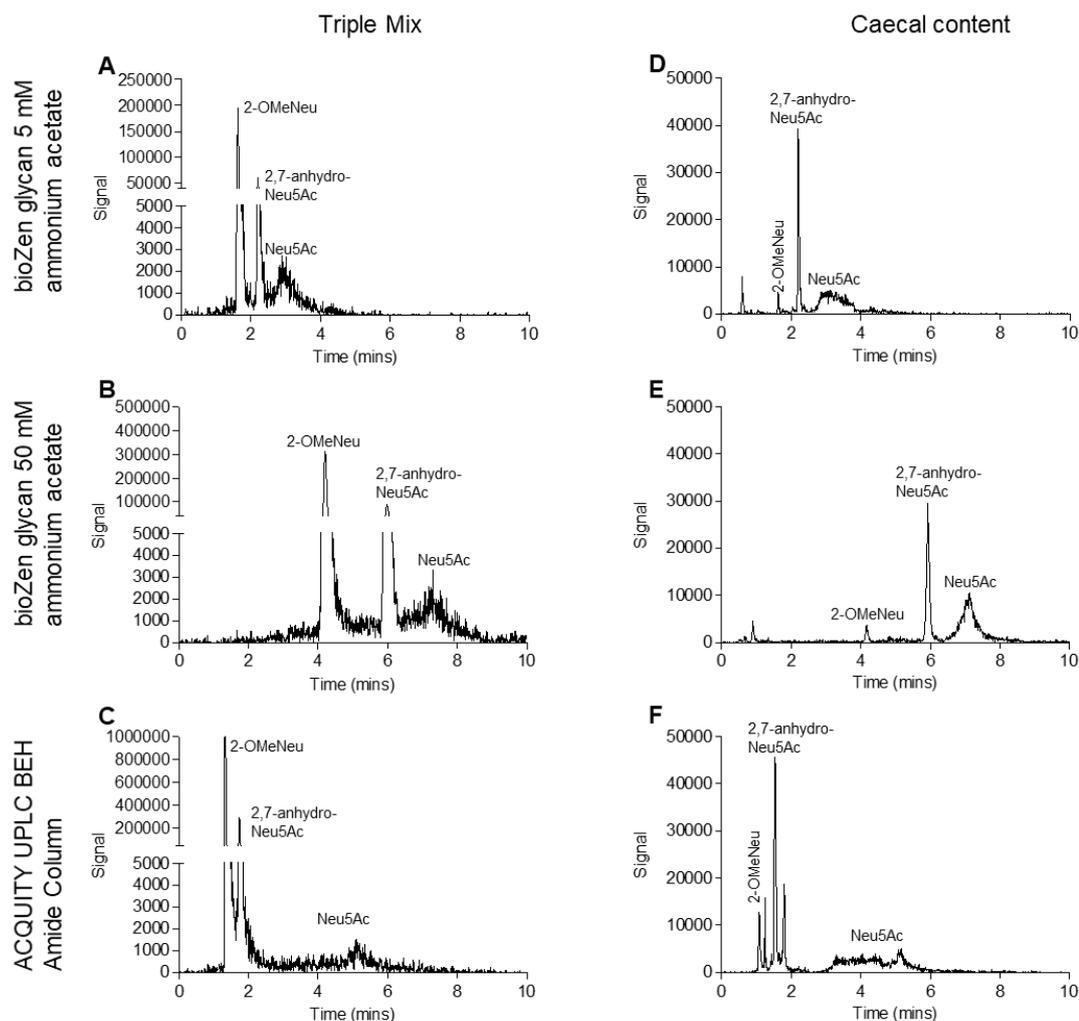


Figure 22: Separation of sialic acids using LC-MS

The separation of the sialic acids 2-OMeNeu, 2,7-anhydro-Neu5Ac and Neu5Ac is shown for a triple mix standard (**A – C**) and mouse caecal content (**D – F**) under different conditions, bioZen glycan column using 5 mM ammonium acetate (**A, D**), bioZen glycan column using 50 mM ammonium acetate (**B, E**), ACQUITY UPLC BEH column using 5 mM ammonium acetate (**C, F**). The detection of sialic acids was by MS and identification of sialic acid compounds was confirmed using MRMs (Table 3). In each case a linear gradient of 95 % to 40 % acetonitrile was used, with a 5 μ l injection volume.

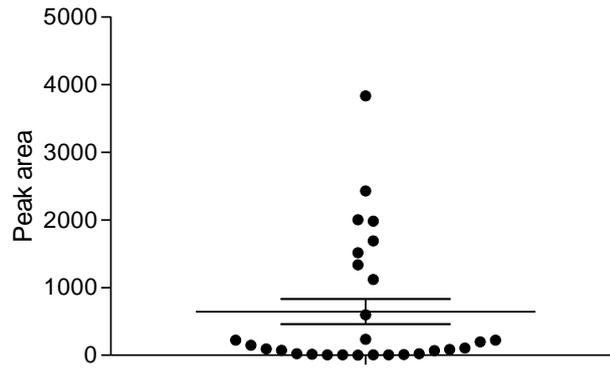


Figure 23: Peak area of internal standard (2-OMeNeu) added to 30 caecal content samples

Each point represents the peak area for internal standard (2-OMeNeu) from a mouse caecal content sample. 1 μ l of 31mM internal standard was added to each caecal content sample and purification performed on cation exchange column and graphitised carbon cartridge before analysis using LC-MS.

3.3 Discussion

This chapter addresses the challenges of working with 2,7-anhydro-Neu5Ac, from its synthesis to detection and quantification in biological samples. The development of a highly efficient MEME method scalable to produce 20 mg of 2,7-anhydro-Neu5Ac is presented. The limiting factor in this method is the final purification on a Biogel P2 column. The price of fetuin as the starting material may also become prohibitive for large scale synthesis. One option to overcome this is to re-cycle fetuin, as shown for lactose in the OPME method (Xiao *et al.*, 2018) as adding sialic acid back to the asialofetuin has been shown to be possible (Monestier *et al.*, 2017). This gives rise to the possibility of producing other forms of 2,7-anhydro-sialic acids such 2,7-anhydro-Neu5Gc or 2,7-anhydro-KDN.

The advantages of the OPME method (Xiao *et al.*, 2018) over the MEME method is largely in the cost effectiveness and scalability of the method, leading to gram quantities of the sugar. Furthermore, the OPME method has greater capacity to produce 2,7-anhydro-sialic acid variants (Xiao *et al.*, 2018). Purification of the compound was a three-phase purification including Biogel P2, a silica column and C18 column. Applying this three-phase purification process may help increase the scalability of the described MEME method. However, this three-phase method does not appear to completely remove all the Neu5Ac as seen in our analysis, performed in house. A further option for improving the purity is to use the HILIC purification method shown here. This method gives a good separation of 2,7-anhydro-Neu5Ac from Neu5Ac which can be crucial for many applications. However, the resolution of the peak corresponding to 2,7-anhydro-Neu5Ac currently limits purification to 1 mg per 40-min run. Significant improvement of this method is therefore required to make it high-throughput and efficient enough for gram scale synthesis. More recently Xi Chen's group have developed a 2-step method for synthesising 2,7-anhydro-Neu5Ac using substrate tagging and a 2-enzyme method giving rise to a multigram scale synthesis (Li *et al.*, 2019) with purification based on the use of a C18 column. An alternative enzymatic approach to produce 2,7-anhydro-Neu5Ac will be discussed in chapter 4.

We also presented the development of methods for the detection and quantification of 2,7-anhydro-Neu5Ac in biological samples. The lack of natural fluorophore or absorbance signature limits the options for detecting these molecules, which is further complicated by the occupation of C2 in the anhydro bridge. This occupation prevents labelling by DMB which is routinely used for other sialic acid molecules (Monestier *et al.*, 2017). Here different methods were trialled for quantification and detection. Firstly, CAD

was used for separation of sialic acids using a HILIC-OH column, this method allowed for detection of sialic acids, however, it showed limited use due to the impact of small contaminants of salts interfering with the peaks of interest. Mass spectrometry methods greatly improved on this by being able to confidently identify the eluted products using molecular weight of parent ions and fragmentation products. This also gives increased confidence in the accuracy of quantification as co-eluting molecules can be removed from the analysis by evaluating the fragmentation patterns.

The chromatography methods tested here allowed the separation and detection of 2,7-anhydro-Neu5Ac and Neu5Ac in biological samples. However, accurate quantification of Neu5Ac using these methods remained challenging due to poor resolution of the peaks. A further challenge with the quantification from biological samples comes from the choice of the internal standard which should be better retained during the purification process used. Although quantification of 2,7-anhydro-Neu5Ac using these methods still requires optimisation, the separation between Neu5Ac and 2,7-anhydro-Neu5Ac gives great potential as a purification step in the synthesis of 2,7-anhydro-Neu5Ac. Progress in these methods and processes will advance the understanding of the biological importance of this molecule.

Chapter 4

Metabolism of 2,7-anhydro-Neu5Ac

by *R. gnavus*

4.1 Introduction

R. gnavus is a gut symbiont belonging to the Lachnospiraceae family of the Clostridia class (cluster XIVa) in the Firmicutes phylum and is considered as a prevalent member of the 'normal' gut microbiota; being present in more than 90% of human faecal samples (Qin *et al.*, 2010). Further, *R. gnavus* has been shown to be disproportionately represented in several diseases including IBD (Olbjorn *et al.*, 2019, Henke *et al.*, 2019, Hall *et al.*, 2017). *R. gnavus* is one of a number of mucus associating gut microbes in the GI tract. These microbes utilise the mucin glycan chains, accounting for the bulk of mucus, as a means of attachment to the gut and/or as a nutrient source (Etienne-Mesmin *et al.*, 2019, Marcobal *et al.*, 2013). Mucin glycans are based on eight core glycan structures with 1 – 4 commonly found in intestinal mucins (Thomsson *et al.*, 2012). These core structures can then be elongated with galactose, GalNAc, or GlcNAc. These glycan chains are terminated by one of numerous epitopes which often include fucosylated or sialylated structures (Tailford *et al.*, 2015a). Mucin glycosylation varies along the GI tract with an increase in sialic acid towards the colon in humans, the reverse trend is observed in mice (Robbe *et al.*, 2003). Due to its terminal location and abundance in the colon, sialic acid is a much-coveted nutrient for gut microbes.

To gain access to this substrate, bacteria first need sialic acid to be released from the mucin glycan chains. Several gut bacteria species encode sialidases to release sialic acid, these sialidases are usually extracellular so the released sialic acid remains free in the environment, where it can be imported by the bacteria or scavenged by other microbes. Interestingly, some bacteria, such as *B. thetaiotaomicron* ATCC 29148, encode sialidases but do not encode genes required for sialic acid utilisation (Brigham *et al.*, 2009). In such cases, it is proposed that the release of the terminal sialic acid residues grants them greater access to the underlying sugars of the glycan chain, which they can then use as a nutrient source. Conversely other bacterial species, including the enteric pathogens *C. difficile* or *S. Typhimurium*, encode the full repertoire of genes required for sialic acid utilisation but lack the sialidase to release it from the glycans, thus relying on sialidase-producing organisms in close proximity to release the sialic acid (Ng *et al.*, 2013).

The genes required for sialic acid utilisation are commonly found clustered together in 'nan' operons. These operons have been described in numerous bacterial species, the majority of which colonise mucosal regions of the body (Juge *et al.*, 2016). The canonical *nanA/E/K* cluster was first described in *E. coli* and encodes the proteins required to transport sialic acid (NanT) and to metabolise sialic acid to GlcNAc-6-P. In this process,

sialic acid is first converted by the *N*-acetylneuraminase lyase (NanA) to ManNAc and pyruvate, the ManNAc is then phosphorylated to ManNAc-6-P by a kinase (NanK), and finally converted to GlcNAc-6-P by an epimerase (NanE) (Plumbridge and Vimr, 1999) (Figure 13). An alternative pathway for sialic acid metabolism also exists and was first described in *Bacteroides fragilis* (Brigham *et al.*, 2009). In this pathway, *nanLET* utilises an MFS transporter (NanT) and relies on the action of an aldolase (NanL), a novel ManNAc-6-P epimerase (also named NanE), and a hexokinase (RokA), to convert Neu5Ac into GlcNAc-6-P (Figure 13). In both cases GlcNAc-6-P is then converted into fructose-6-P, which is a substrate in the glycolytic pathway, by NagA and NagB, the genes encoding these proteins are usually not associated with the *nan* operons but encoded elsewhere in the genome.

Since sialic acid cleavage is an extracellular process, bacteria have evolved multiple strategies to take up this important nutrient. At least five different classes of sialic acid transporter have been described to date (Thomas, 2016). These include the NanT MFS H⁺-coupled symporter used by *E. coli* and *B. fragilis* (Martinez *et al.*, 1995), a secondary transporter of the sodium solute symport (SSS) family utilised by the pathogens *C. difficile* and *S. Typhimurium* (Severi *et al.*, 2010, Almagro-Moreno and Boyd, 2009), and high-affinity sialic acid transporters mediated by substrate-binding proteins, including a tripartite ATP-independent periplasmic (TRAP) transporter, SiaPQM, and a number of different ATP-binding cassette (ABC) transporters (Allen *et al.*, 2005). To date all these transporters have been characterised transporting Neu5Ac, with some additionally being able to transport related sialic acids Neu5Gc and KDN (Hopkins *et al.*, 2013).

The ability of *R. gnavus* to utilise mucin glycans is strain-dependent and associated with the expression of an IT-sialidase. The IT-sialidase encoded for by *R. gnavus* is also part of an extended gene cluster including the canonical NanA/E/K cluster of genes as well as genes encoding for a predicted ABC transporter and genes of unknown function (Croft *et al.*, 2016). ABC transporters have been further subdivided into three classes, Sialic acid transporter (SAT), SAT2 and SAT3. ABC transporters are classified by their three domain structures including two permeases and a solute binding protein (SBP), with an ATPase domain fused with one of the permeases or shared with other systems. The SBP proteins are further classified into clusters according to sequence identity (Scheepers *et al.*, 2016). The first SAT system was characterised from *Haemophilus ducreyi* which uses an SBP of the cluster C family (Post *et al.*, 2005), and where one of the permeases is a fusion with the ATPase domain. The SBPs of the SAT2 and SAT3 systems come from the cluster B family and do not encode an ATPase domain within the cluster, a characteristic of the CUT1 family of carbohydrate ABC transporters (Marion

et al., 2011a). The *R. gnavus* transporter is predicted to encode a SAT2 system (Crost *et al.*, 2016), a SAT2 system is also present in *S. pneumoniae* which additionally encodes a SAT3 system which is required for growth on Neu5Ac, while the function of the SAT2 system is unknown in this bacterium (Marion *et al.*, 2011b).

Despite encoding the necessary genes for sialic acid utilisation, *R. gnavus* is not able to grow on Neu5Ac as a sole carbon source (Crost *et al.*, 2013). The IT-sialidase encoded by *R. gnavus* was found to cleave off terminal α 2-3-linked sialic acid from glycoproteins, releasing 2,7-anhydro-Neu5Ac instead of Neu5Ac (Tailford *et al.*, 2015b). This cluster, which is not found in non-mucin glycan foraging strains of *R. gnavus*, was shown to be upregulated during growth on mucin (Crost *et al.*, 2016). Together these data suggest that the extended *nan* cluster in *R. gnavus* may be dedicated to 2,7-anhydro-Neu5Ac and not Neu5Ac utilisation. This mechanism may be part of selfish strategy employed by *R. gnavus* releasing the much-coveted sialic acid in a form that only it can use.

4.2 Results

4.2.1 Growth of *R. gnavus* on sialylated substrates

To investigate the specificity of *R. gnavus* toward sialylated substrates, a mucin glycan foraging strain, *R. gnavus* ATCC 29149 was grown under strict anaerobic conditions on a range of sialylated and sialic acid carbon sources, as well as glucose as a positive control (Figure 24). This strain was able to grow on glucose, 2,7-anhydro-Neu5Ac and 3'SL to similar densities, but was unable to grow on Neu5Ac, Neu5Gc, lactose or 6'SL. This agrees with the reported specificity of the IT-sialidase for 3'SL but not 6'SL and demonstrates that *R. gnavus* can utilise the product of the IT-sialidase, 2,7-anhydro-Neu5Ac, as a preferential carbon source.

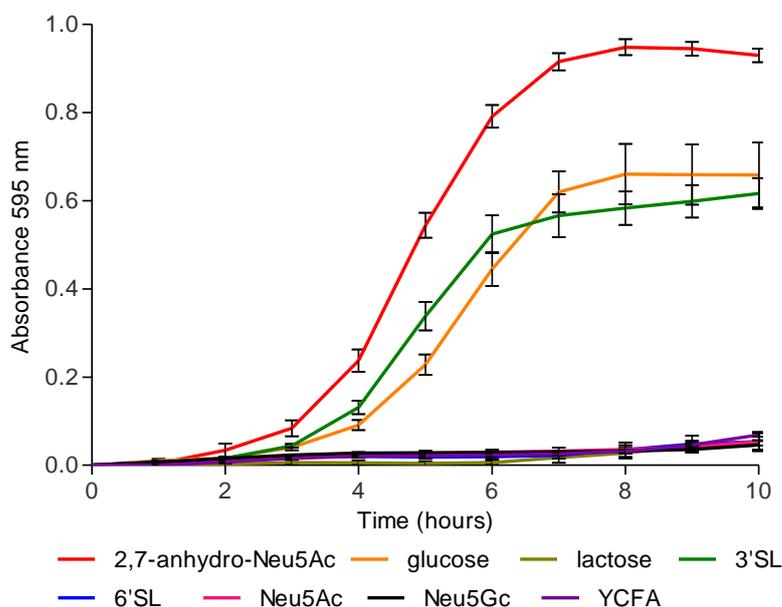


Figure 24: Growth curves of *R. gnavus* on sialic acids

Change in OD₅₉₅ absorbance were measured over time of microtiter cultures of *R. gnavus* ATCC 29149 grown in YCFA using 11.1 mM 2,7-anhydro-Neu5Ac, glucose, lactose, 3'SL, 6'SL, Neu5Ac or Neu5Gc as sole carbon sources.

4.2.2 The *R. gnavus nan* cluster is transcribed as one operon

The *R. gnavus nan* cluster (Figure 25a) is made up of 11 protein encoding genes, the first two proteins RUMGNA_02701 and 02700 have unknown functions, while RUMGNA_02699 is homologous to the AraC type transcriptional regulator. RUMGNA_02698-02696 are predicted to form an ABC transporter of the SAT2 type with two permeases and an SBP and RUMGNA_02695 is a predicted oxidoreductase of

unknown function. The remaining genes are RUMGNA_02694 (IT-sialidase / *RgNanH*) and homologs of the NanE (RUMGNA_02693), NanA (RUMGNA_02692) and NanK (RUMGNA_02691) proteins found in the canonical *nan* clusters.

It was previously reported that six genes of the cluster including homologs of the NanA/E/K canonical *nan* cluster form part of the same operon which is induced during *R. gnavus* growth on 3'SL and mucin (pPGM) (Crost *et al.*, 2013). To investigate whether the five other genes (RUMGNA_02701 – RUMGNA_02697; Figure 25a) form part of the same operon, a transcriptomic analysis was performed on *R. gnavus* ATCC 29149 when grown on 3'SL as a sole carbon source. PCR amplification was performed across adjacent genes including RUMGNA_02694 (which is part of the operon described in Crost *et al.*, 2013) with gDNA, cDNA (from *R. gnavus* grown on 3'SL), and a no reverse transcriptase control (-RT) used as templates. The results show amplification for each primer pair from the cDNA samples, suggesting that these six genes are transcribed as the same transcript forming part of the *R. gnavus* extended *nan* operon (Figure 25b). This indicates that the full 11 genes function together in an operon.

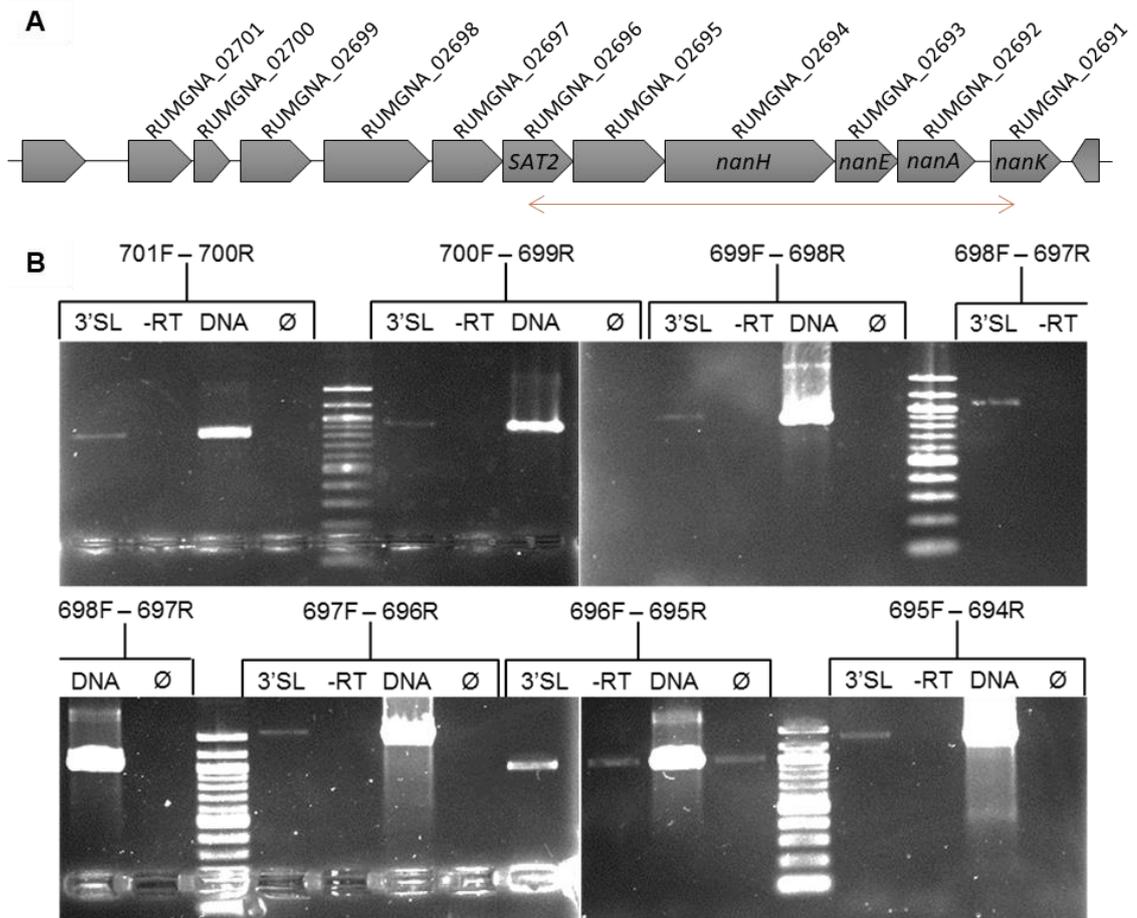


Figure 25: Composition of the *R. gnavus nan* cluster

A) Diagram depicting the layout of the *nan* cluster in *R. gnavus* ATCC 29149. The red arrow indicates genes previously shown to belong to the same operon (Crost *et al.*, 2013) **B)** PCR to determine other genes that are part of the *nan* cluster, each panel contains cDNA from *R. gnavus* ATCC 29149 grown with 3'SL (3'SL), a control where the reverse transcriptase was omitted from cDNA synthesis (-RT), gDNA (DNA) from *R. gnavus* ATCC 29149 and a blank (Ø). Primer pairs used (abbreviated to the last 3 numbers of the gene ID (see **A**) with F-forward and R-reverse) are indicated for each panel, with DNA marker sizes from 100 bp ladders (New England Biolabs), bright bands are 500 bp and 1 kb.

4.2.3 *R. gnavus nan* operon is induced during growth on 2,7-anhydro-Neu5Ac or 3'SL

To confirm that the *R. gnavus nan* operon is implicated in the utilisation of 2,7-anhydro-Neu5Ac, expression analysis of the operon was carried out when *R. gnavus* ATCC 29149 was grown on 2,7-anhydro-Neu5Ac or 3'SL as sole carbon and compared to growth on glucose (Figure 26). All 11 genes within the operon were upregulated when 2,7-anhydro-Neu5Ac or 3'SL was used as the sole carbon source, whilst the expression of the two flanking genes, RUMGNA_02690 and 02702, did not change. This result both defines the limit of the operon and confirms the relevance of the extended *nan* operon in the metabolism of 2,7-anhydro-Neu5Ac.

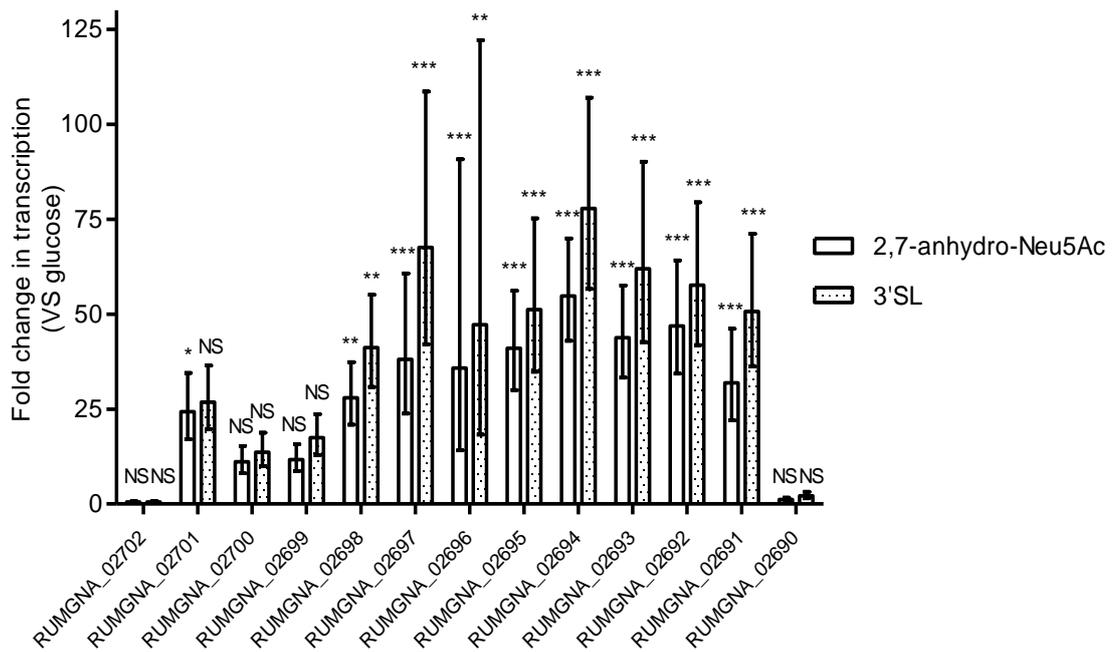


Figure 26: Expression of the *R. gnavus nan* cluster on 2,7-anhydro-Neu5Ac and 3'SL

qPCR analysis showing fold changes in expression of *nan* operon genes when *R. gnavus* ATCC 29149 was grown with 3'SL or 2,7-anhydro-Neu5Ac compared to glucose using $2^{-\Delta\Delta C_t}$ calculation. The bars indicate the average fold change in gene expression with standard deviation indicated with data based on three biological samples each analysed in triplicate. Statistical significance was determined using one-way analysis of variance (ANOVA) with a Dunnett's multiple comparison test. NS, no significant change in expression ($p > 0.05$), * $p < 0.05$, $p < 0.01$, *** $p < 0.001$.

Several hypotheses could be proposed to explain the specificity of *R. gnavus* to utilise 2,7-anhydro-Neu5Ac and not Neu5Ac. One scenario is that the transporter is specific to 2,7-anhydro-Neu5Ac and does not allow Neu5Ac to enter the cells (Figure 27a). This would then require *RgNanA* or another enzyme to be active against 2,7-anhydro-Neu5Ac inside the cell. A second scenario is that both 2,7-anhydro-Neu5Ac and Neu5Ac are transported into the cell. In this case, the first metabolic enzyme involved in sialic acid metabolism, *RgNanA* or an enzyme of unknown function, would be specific for 2,7-anhydro-Neu5Ac and not active against Neu5Ac (Figure 27b).

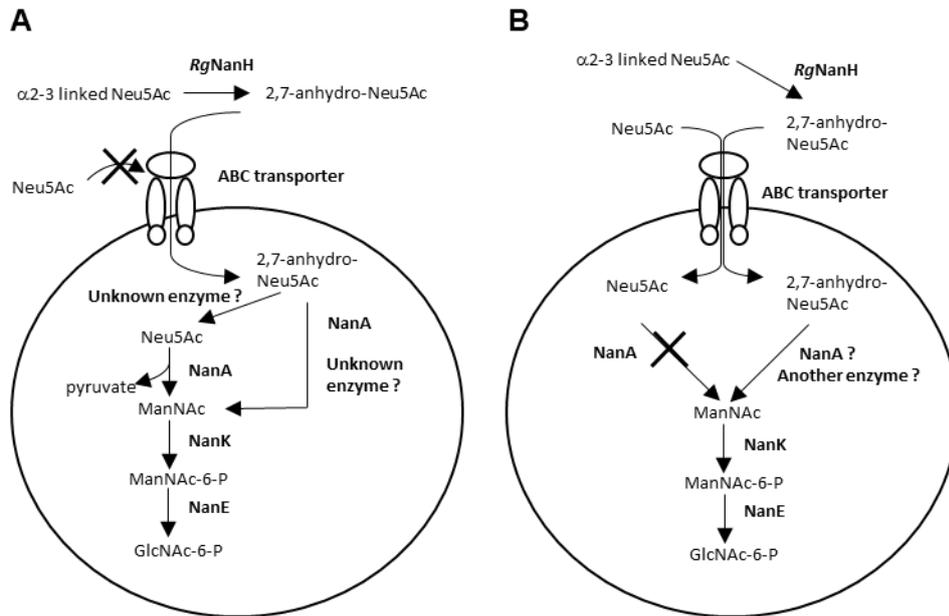


Figure 27: Diagram depicting the possible scenarios for *R. gnavus* metabolism of 2,7-anhydro-Neu5Ac

2,7-anhydro-Neu5Ac may be specifically transported inside the bacteria before being converted into Neu5Ac and metabolised (A) or alternatively both 2,7-anhydro-Neu5Ac and Neu5Ac may be transported inside the bacteria and metabolic enzymes could be specific to 2,7-anhydro-Neu5Ac (B).

4.2.4 Bioinformatic analysis of the *R. gnavus* nan operon

Previous bioinformatic analyses predicted the function of some of the genes of the cluster (Table 4). Here, further bioinformatic analyses were used to identify which genes of the *nan* operon may be involved in 2,7-anhydro-Neu5Ac metabolism. Sequence Similarity Network (SSN) is a tool used for predicting the function of uncharacterised proteins within protein families based on sequence identity and domain architecture with other characterised proteins (Gerlt *et al.*, 2015).

Table 4: Functions or predicted functions of *R. gnavus* nan operon proteins.

Gene identifier	Function	Reference
RUMGNA_02701	Unknown (putative sialic acid esterase)	
RUMGNA_02700	Unknown (putative sugar isomerase)	
RUMGNA_02699	Predicted transcriptional regulator (AraC type)	
RUMGNA_02698	Predicted ABC transporter Soluble binding protein	
RUMGNA_02697	Predicted ABC transporter trans-membrane domain	
RUMGNA_02696	Predicted ABC transporter trans-membrane domain	
RUMGNA_02695	Unknown (putative oxidoreductase)	
RUMGNA_02694	IT-sialidase	(Tailford <i>et al.</i> , 2015b)
RUMGNA_02693	Predicted NanE	
RUMGNA_02692	Predicted NanA	
RUMGNA_02691	Predicted NanK	

SSN analysis was first performed on *RgNanH*, the only characterised protein of the cluster known to be involved in 2,7-anhydro-Neu5Ac metabolism (Crost *et al.*, 2016, Tailford *et al.*, 2015b). *RgNanH* belongs to the GH33 family of glycoside hydrolases encompassing sialidases (Lombard *et al.*, 2014). SSN analyses of the GH33 sequences showed the presence of two distinct groups we predicted represent the hydrolytic sialidases and IT-sialidases. *R. gnavus* sequences were found exclusively in the IT-sialidase group. *S. pneumoniae* was represented in both groups, as expected from previous reports showing that *S. pneumoniae* encode as many as 3 sialidase genes, including one (NanB) with similarity to the *R. gnavus* IT-sialidase (Xu *et al.*, 2011) (Figure 28a).

Transport of 2,7-anhydro-Neu5Ac is likely to be via the predicted ABC transporter (RUMGNA_02696 – 698). The specificity of ABC transporters is conferred by the SBP, which in *R. gnavus* is encoded by RUMGNA_02698. The Uniprot families for the ABC SBPs were too large to perform a meaningful SSN analysis. Therefore, BLAST analyses

were carried out with the EFI-ESN top BLAST feature using the top 2,500 hits against RUMGNA_02698 (Figure 28c). The RUMGNA_02698 protein clustered with a group of uncharacterised proteins predicted to be SBPs, some predicted to be involved in carbohydrate transport. This group contained strains of *S. pneumoniae* as shown for the analysis of the sialidases.

For RUMGNA_02692 (*RgNanA*), the predicted sialic acid aldolase, the SSN analysis was performed for the full family of sialic acid aldolases identified in the uniprot database (N-acetylneuraminase lyase; IPR005264). The analysis showed that *RgNanA* forms part of a cluster that is connected to, but distinct from the other sialic acid aldolases (Figure 28b). As was the case for *RgNanH* and *RgSBP*, the cluster containing the *R. gnavus* sialic acid aldolase protein also contains several representative strains from *S. pneumoniae*. The similar makeup of the clusters containing the *RgNanH*, *RgSBP* and *RgNanA* suggests that these proteins may be involved in the specificity of *R. gnavus* 2,7-anhydro-Neu5Ac metabolism/transport over Neu5Ac.

For the SSN analysis of the three proteins that do not have a predicted function, RUMGNA_02695 (Figure 28f), RUMGNA02700 (Figure 28e) and RUMGNA_02701 (Figure 28d), the EFI-ESN top BLAST feature was once again used. The RUMGNA_02701 proteins grouped with sequences found in species closely related to *R. gnavus* including *Blautia* strains, the proteins are all uncharacterised or predicted GDSL-lipase-like proteins. Some of the proteins are predicted as acylhydrolases PFAM family PF13472. This could point to a role in deacetylation of sialic acids like the NanS proteins (Rangarajan *et al.*, 2011), but this protein does not appear to be present in species such as *S. pneumoniae* found co-occurring with the *RgNanH* and *RgNanA* and *RgSBP* proteins in this analysis.

SSN analysis of RUMGNA_02700 showed that the *R. gnavus* protein belongs to a small subset of eight other uncharacterised proteins from species including *Blautia* strains (Figure 28e). These uncharacterised proteins fall into the PFAM group PF04074 which classifies these proteins as having a domain of unknown function. The YhcH/YjgK/YiaL superfamily these proteins are predicted to be part of include proteins involved in biofilm formation (Kim *et al.*, 2009), or a possible sugar isomerase of sialic acid metabolism (Teplyakov *et al.*, 2005). Again, the cluster containing *R. gnavus* proteins does not include species found co-occurring with *RgNanH*, *RgNanA* or *RgSBP*.

The RUMGNA_02695 SSN analysis grouped the *R. gnavus* protein with proteins from nearly 400 strains including *S. pneumoniae* as was the case with *RgNanA* and *RgNanH*.

Other species found in this cluster include strains of *E. coli*, *Staphylococcus pseudintermedius* and *S. Typhimurium*. The proteins all belong to PFAM family PF02894; which is a sub family of GFO_IDH_MocA containing an NAD binding domain. Characterised proteins in this family have a range of targets including glucose, galactose and fructose (Taberman *et al.*, 2016). The species found in the RUMGNA_02695 protein cluster are similar to those found for *RgNanH*, *RgNanA* and *RgSBP*, again possibly suggesting a shared role in 2,7-anhydro-Neu5Ac metabolism.

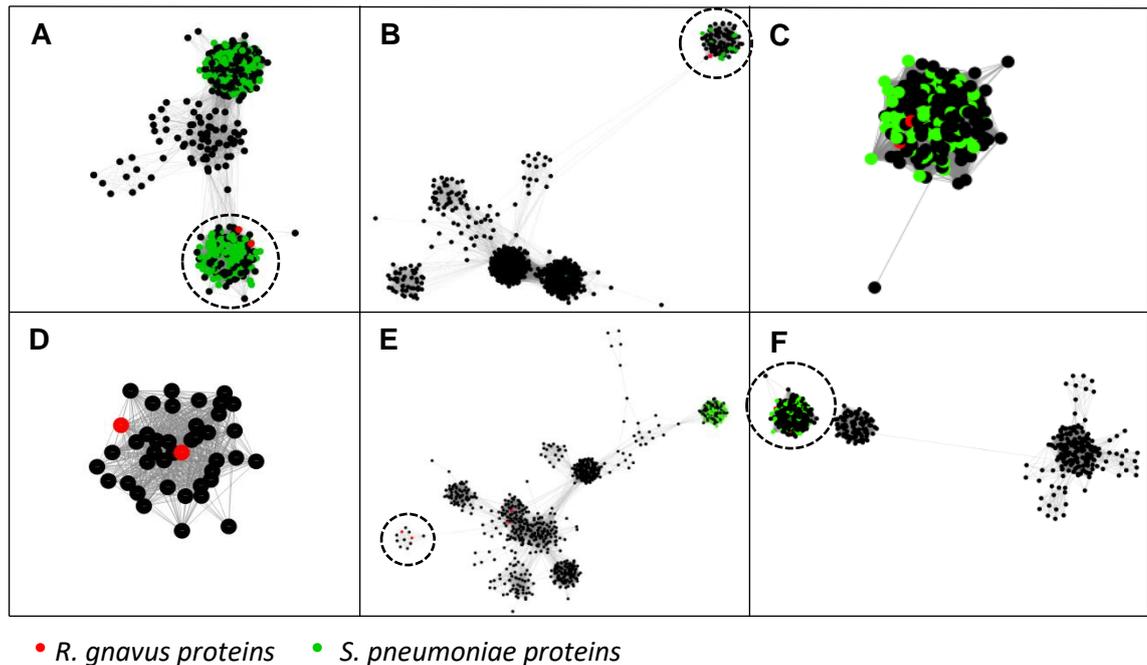


Figure 28: Sequence Similarity Networks (SSN) of predicted proteins in the *R. gnavus nan* cluster. Nodes representing proteins from *R. gnavus* species (red) and *S. pneumoniae* (green) are highlighted. Clusters containing proteins from the *nan* cluster are shown using a dashed circle, and only clusters with edges to the *R. gnavus* cluster are shown. **A)** InterPro family of sialidases. **B)** InterPro family of sialic acid aldolases. **C)** Top 2500 Blast hits of *RgSBP*. **D)** Top 2500 Blast hits of RUMGNA_02701. **E)** Top 2500 Blast hits of RUMGNA_02700. **F)** Top 2500 Blast hits of RUMGNA_02695.

Species that encode proteins within the same SSN clusters as the *R. gnavus* proteins are shown in Table 5. The data show that, of the unknown proteins, RUMGNA_02695 shares the most co-occurring species with *RgNanH*, *RgNanA* and *RgSBP*, suggesting a potential role in 2,7-anhydro-Neu5Ac metabolism. RUMGNA_02701 and 02700 do not share so many co-occurring species so their role may not be specific to this pathway. From bioinformatics analysis we hypothesised that RUMGNA_02695, *RgNanH*, *RgNanA* and *RgSBP* may be involved in the specific metabolism of 2,7-anhydro-Neu5Ac. These predicted proteins were targeted for heterologous expression and biochemical

characterisation. For the other predicted proteins of unknown function (RUMGNA_02700 and RUMGNA_02701), heterologous expression will be carried out to determine function.

Table 5: Identification of species co-occurring with SSN analysis of *R. gnavus* proteins, grey shading denotes that the given species encodes a co-occurring protein.

Species	RgNanH	RgNanA	RgSBP	R_02695	R_02700	R_02701
<i>Acinetobacter baumannii</i>						
<i>Actinobacillus minor</i>						
<i>Actinomyces glycerinitolerans</i>						
<i>Actinomyces johnsonii</i>						
<i>Actinomyces naeslundii</i>						
<i>Actinomyces oris</i>						
<i>Actinomyces viscosus</i>						
<i>Aggregatibacter aphrophilus</i>						
<i>Alloiococcus otitis</i>						
<i>Avibacterium paragallinarum</i>						
<i>Anaerobium acetethylicum</i>						
<i>Bacillus bataviensis</i>						
<i>Bacillus galactosidilyticus</i>						
<i>Bacillus oleivorans</i>						
<i>Bacillus sp. B-jedd</i>						
<i>Bacillus sp. FJAT-18017</i>						
<i>Beutenbergia cavernae</i>						
<i>Blautia hansenii</i>						
<i>Blautia sp. YL58</i>						
<i>Brachyspira hampsonii</i>						
<i>Brevibacillus borstelensis</i>						
<i>Brevibacillus brevis</i>						
<i>Brevibacillus choshinensis</i>						
<i>Brevibacillus formosus</i>						
<i>Chelonobacter oris</i>						
<i>Chlamydia trachomatis</i>						
<i>Clostridium bolteae</i>						
<i>Clostridium chromiireducens</i>						
<i>Clostridium clostridioforme</i>						
<i>Clostridium hathewayi</i>						
<i>Clostridium papyrosolvans</i>						
<i>Clostridium symbiosum</i>						
<i>Clostridium tertium</i>						
<i>Dolosicoccus paucivorans</i>						
<i>Dolosigranulum pigrum</i>						
<i>Dorea longicatena</i>						
<i>Enterococcus canintestini</i>						

<i>Enterococcus cecorum</i>						
<i>Enterococcus faecalis</i>						
<i>Escherichia coli</i>						
<i>Eubacterium coprostanoligenes</i>						
<i>Fictibacillus enclensis</i>						
<i>Fictibacillus solisalsi</i>						
<i>Firmicutes bacterium</i>						
<i>Gallibacterium anatis</i>						
<i>Gallibacterium salpingitidis</i>						
<i>Gemella haemolysans</i>						
<i>Glaesserella parasuis</i>						
<i>Granulicatella balaenopterae</i>						
<i>Haemophilus haemoglobinophilus</i>						
<i>Haemophilus parainfluenzae</i>						
<i>Haemophilus pittmaniae</i>						
<i>Haemophilus sputorum</i>						
<i>Hungatella hathewayi</i>						
<i>Intestinimonas butyriciproducens</i>						
<i>Klebsiella pneumoniae</i>						
<i>Lachnoanaerobaculum saburreum</i>						
<i>Lactobacillus salivarius</i>						
<i>Leptotrichia goodfellowii</i>						
<i>Leptotrichia sp. oral</i>						
<i>Lysinibacillus contaminans</i>						
<i>Lysinibacillus sphaericus</i>						
<i>Lysinibacillus varians</i>						
<i>Macrococcus canis</i>						
<i>Macrococcus caseolyticus</i>						
<i>Macrococcus goetzii</i>						
<i>Marvinbryantia formatexigens</i>						
<i>Microterricola viridarii</i>						
<i>Mobiluncus mulieris</i>						
<i>Muribacter muris</i>						
<i>Paenibacillus antarcticus</i>						
<i>Paenibacillus ihbetae</i>						
<i>Paenibacillus macquariensis</i>						
<i>Paenibacillus odorifer</i>						
<i>Paenibacillus pectinilyticus</i>						
<i>Paenibacillus pini</i>						
<i>Paenisporosarcina quisquiliarum</i>						
<i>Pasteurella multocida</i>						
<i>Pediococcus acidilactici</i>						
<i>Pediococcus pentosaceus</i>						
<i>Psychrobacillus psychrotolerans</i>						

<i>Rodentibacter pneumotropicus</i>						
<i>Ruminococcus gnavus</i>						
<i>Ruminococcus torques</i>						
<i>Ruthenibacterium lactatiformans</i>						
<i>Salmonella enterica</i>						
<i>Spirochaetes bacterium</i>						
<i>Sporosarcina psychrophila</i>						
<i>Staphylococcus cohnii</i>						
<i>Staphylococcus delphini</i>						
<i>Staphylococcus intermedius</i>						
<i>Staphylococcus lentus</i>						
<i>Staphylococcus lutrae</i>						
<i>Staphylococcus pseudintermedius</i>						
<i>Streptobacillus moniliformis</i>						
<i>Streptococcus australis</i>						
<i>Streptococcus canis</i>						
<i>Streptococcus cristatus</i>						
<i>Streptococcus gordonii</i>						
<i>Streptococcus infantis</i>						
<i>Streptococcus merionis</i>						
<i>Streptococcus mitis</i>						
<i>Streptococcus oralis</i>						
<i>Streptococcus parasanguinis</i>						
<i>Streptococcus pneumoniae</i>						
<i>Streptococcus salivarius</i>						
<i>Streptococcus sanguinis</i>						
<i>Streptococcus suis</i>						
<i>Treponema lecithinolyticum</i>						
<i>Trueperella bernardiae</i>						
<i>Trueperella pyogenes</i>						
<i>Tepidanaerobacter syntrophicus</i>						
<i>Tissierella creatinophila</i>						
<i>Vibrio ponticus</i>						
<i>Virgibacillus phasianinus</i>						
<i>Virgibacillus soli</i>						

MultiGeneBlast analysis of the *R. gnavus nan* cluster revealed that this cluster is shared by a limited number of species including 37 different *S. pneumoniae*, *S. suis*, *Blautia hansenii*, *Blautia* sp. YL58 and *Intestinimonas butyriciproducens* (Figure 29). This analysis was limited to those cluster containing a sialidase. All 37 *S. pneumoniae* clusters have roughly the same layout, with the area between the two sub-clusters being the most variable. These genes have been shown to be responsible for Neu5Ac metabolism in *S. pneumoniae* TIGR4 (Marion *et al.*, 2011b). The *Streptococcus* clusters have an RpiR-

type regulator, where as an AraC-type regulator is observed in others, including *R. gnavus*. *Blautia* sp. YL58 is the only other cluster that contains the RUMGNA_RS11885 lipase/esterase homolog, but *S. suis* A7 and *I. butyriciproducens* AF211 both have a different type of esterase (acetyl xylan esterase; Figure 29). This finding suggests that the specialisation of the *R. gnavus nan* cluster, may be unique to *R. gnavus* and confer a competitive advantage over other members of the gut microbiota.

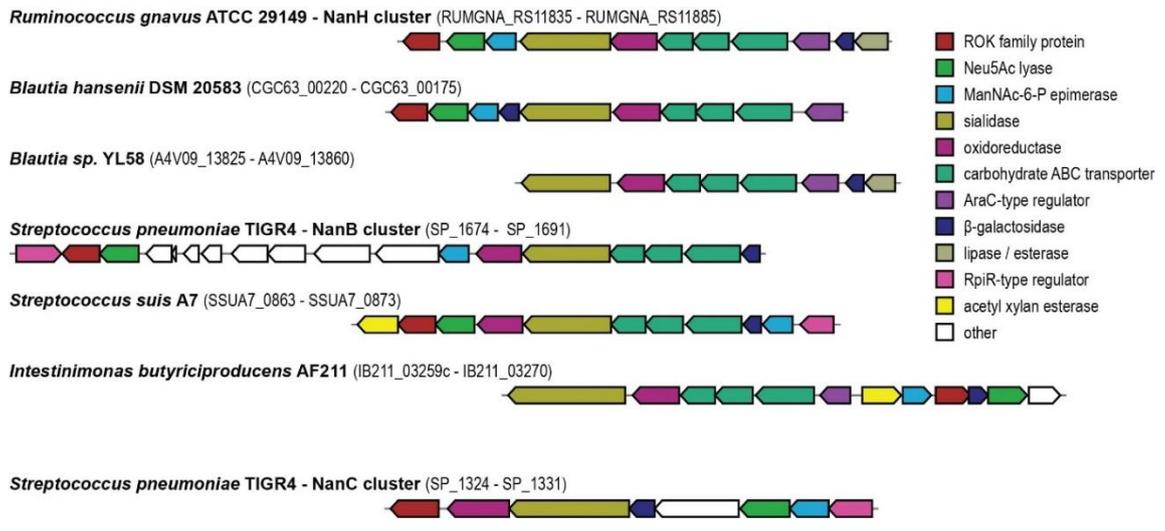


Figure 29: Gene organization in predicted homologs of the *R. gnavus nan* cluster

Schematic representation of the variety of *in silico* identified *nan* cluster homologs. The 37 *S. pneumoniae* cluster organisations are highly similar and represented here by the NanB of NanC clusters from *S. pneumoniae* TIGR4. Cluster locus tag ranges are bracketed, and genes are colour coded by predicted function as described in the inset.

4.2.5 The predicted transporter of *R. gnavus nan* operon is specific to 2,7-anhydro-Neu5Ac

The RUMGNA_02698 gene, encoding the predicted SBP of the transporter (*RgSBP*), was cloned into the pEHISTEV expression vector fused to an N-terminal HIS-tag and the protein recombinantly expressed in *E. coli*. The SBP contains a predicted signal peptide to target the protein for secretion, with the signal peptide predicted to be cleaved between residue 29A and 30A by both Signal 3-L and Signal P prediction tools. The recombinant protein was produced without this domain. Following successful expression, the HIS-tagged recombinant protein was purified by Immobilised Metal Affinity Chromatography (IMAC) and gel filtration. SDS-PAGE analysis showed the presence of a band at the expected molecular weight of 49.7 kDa (Figure 30).

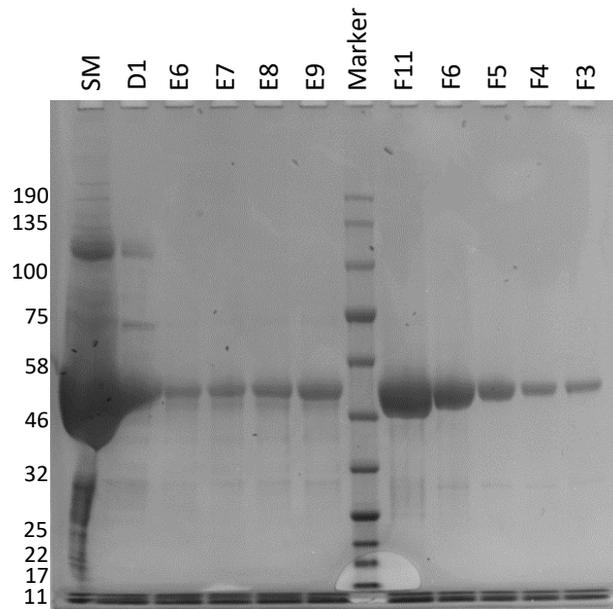


Figure 30: SDS-PAGE analysis of recombinant RUMGNA_02698 (*RgSBP*) purification

Fractions were analysed on a 4-12% BIS-TRIS gel, SM – starting material before gel filtration, marker – N.E.B broadrange pre-stained protein marker (expected size 49.7 kDa) D1 corresponds to small peak identified in gel filtration, E6 – F3 correspond to representative fractions from the major peak following gel filtration.

The ligand specificity of the purified recombinant protein, *RgSBP*, was tested by Differential Scanning Fluorimetry (DSF) against 2,7-anhydro-Neu5Ac and Neu5Ac (Figure 31). In complex with a ligand, proteins are more stable, and the temperature required to melt the protein increases, by measuring the difference in melting temperature between the apoprotein and bound protein the relative strength of binding to different ligands can be calculated (Figure 31a). Using this approach, the data showed that the addition of Neu5Ac led to a minor increase in the protein melting temperature of 1°C. This small increase could also reflect a change in pH or buffer composition on adding Neu5Ac. On the other hand, 2,7-anhydro-Neu5Ac led to a larger difference in melting temperature of 5°C (Figure 31b), suggesting that the protein is able to bind 2,7-anhydro-Neu5Ac.

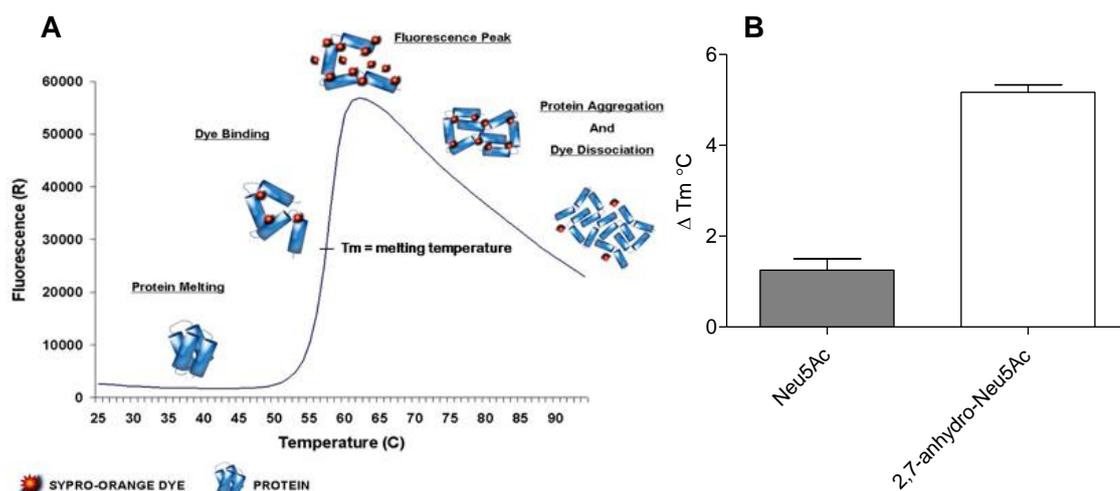


Figure 31: Differential scanning fluorimetry of *RgSBP* with Neu5Ac or 2,7-anhydro-Neu5Ac

A) Diagram showing the principle of DSF, as the temperature increases the protein begins to unfold, this allows SYPRO-orange dye to bind to the exposed hydrophobic regions. The exclusion of water from the dye increases its fluorescence, the melting temperature is determined at half of the maximal fluorescence. In complex with a ligand, proteins are more stable and so the melting temperature increases. **B)** The ΔT_m is the melting temperature of the protein with ligand subtract the melting temperature of the protein with no ligand added.

To confirm the ligand specificity of *RgSBP*, changes in protein fluorescence upon addition of 2,7-anhydro-Neu5Ac or Neu5Ac were evaluated. Due to the presence of 15 tyrosine residues in *RgSBP*, fluorescence changes were measured by exciting the protein at 297 nm. Addition of 10 μM or 20 μM 2,7-anhydro-Neu5Ac resulted in a change in the spectrum intensity with a significant shift at 350 nm. 2,7-anhydro-Neu5Ac caused a 16% quench in the fluorescence of the protein at ligand saturation (Figure 32a). In contrast, addition of Neu5Ac at 10 μM , 20 μM or 70 μM did not change the spectra, indicating no interaction (Figure 32b). Titration of 0.5 μM *RgSBP* with 2,7-anhydro-Neu5Ac was performed in triplicate and change in intensity at 350 nm measured. When the data were fit with a hyperbolic curve, a K_d of 1.349 μM (\pm 0.046) (Figure 32c) was calculated. The lack of binding to Neu5Ac was further demonstrated by measuring changes in fluorescence following sequential additions of 10 μM ligands. Following six sequential additions of 10 μM Neu5Ac, no change in intensity was observed at 350 nm, whereas the subsequent addition of 10 μM 2,7-anhydro-Neu5Ac led to a large decrease in intensity (Figure 32d). Conversely, addition of 10 μM 2,7-anhydro-Neu5Ac resulted in a large decrease in intensity and six subsequent additions of 10 μM Neu5Ac caused no further reduction or reverse in the intensity (Figure 32d), indicating that Neu5Ac is unable to displace 2,7-anhydro-Neu5Ac, and further supporting the specific interaction between *RgSBP* and 2,7-anhydro-Neu5Ac. I performed these fluorescence assays during a 1

week visit to the lab of Gavin H. Thomas at the University of York, who assisted me in experimental set up and data analysis.

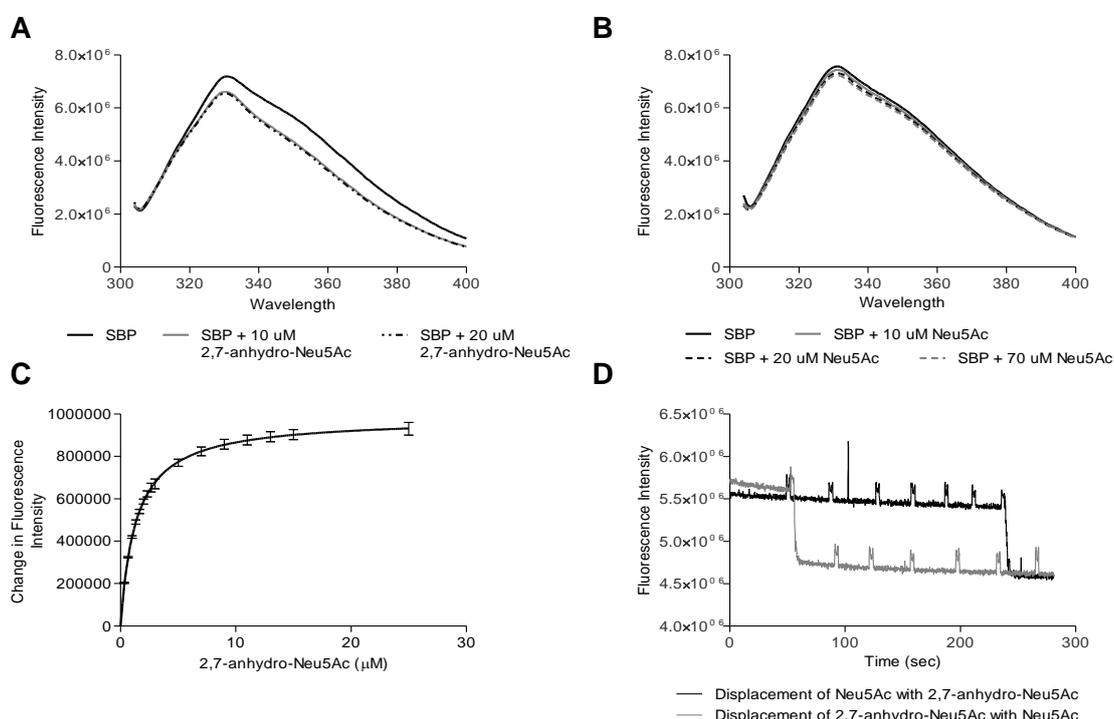


Figure 32: Steady-state fluorescence analysis of ligand binding to *RgSBP*

Fluorescence emission spectrum of 0.5 μ M *RgSBP* excited at 297 nm in the presence or absence of 2,7-anhydro-Neu5Ac **A** or Neu5Ac **B**. **C**) Titration of 0.5 μ M *RgSBP* with 2,7-anhydro-Neu5Ac, average of triplicate readings with standard error. **D**) Displacement of Neu5Ac with 2,7-anhydro-Neu5Ac, six sequential additions of 10 μ M Neu5Ac to 0.5 μ M *RgSBP* followed by one addition of 10 μ M 2,7-anhydro-Neu5Ac, and displacement of 2,7-anhydro-Neu5Ac with Neu5Ac, one addition of 10 μ M 2,7-anhydro-Neu5Ac followed by 6 subsequent additions of 10 μ M Neu5Ac. The signal peaks are artefacts attributed to external light during sample addition.

The binding was further assessed by isothermal titration calorimetry (ITC) using 100 μ M *RgSBP*, and sequential additions of 2 mM ligand. Binding to 2,7-anhydro-Neu5Ac gave a K_d of 2.42 ± 0.27 μ M, conversely no binding was observed with Neu5Ac (Figure 33). This is in agreement with the results obtained by fluorescence spectroscopy. Together these data clearly show that *RgSBP* specifically binds to 2,7-anhydro-Neu5Ac but not to Neu5Ac, in line with the growth of *R. gnavus* ATCC 29149 on 2,7-anhydro-Neu5Ac but not Neu5Ac (Croft et al., 2016).

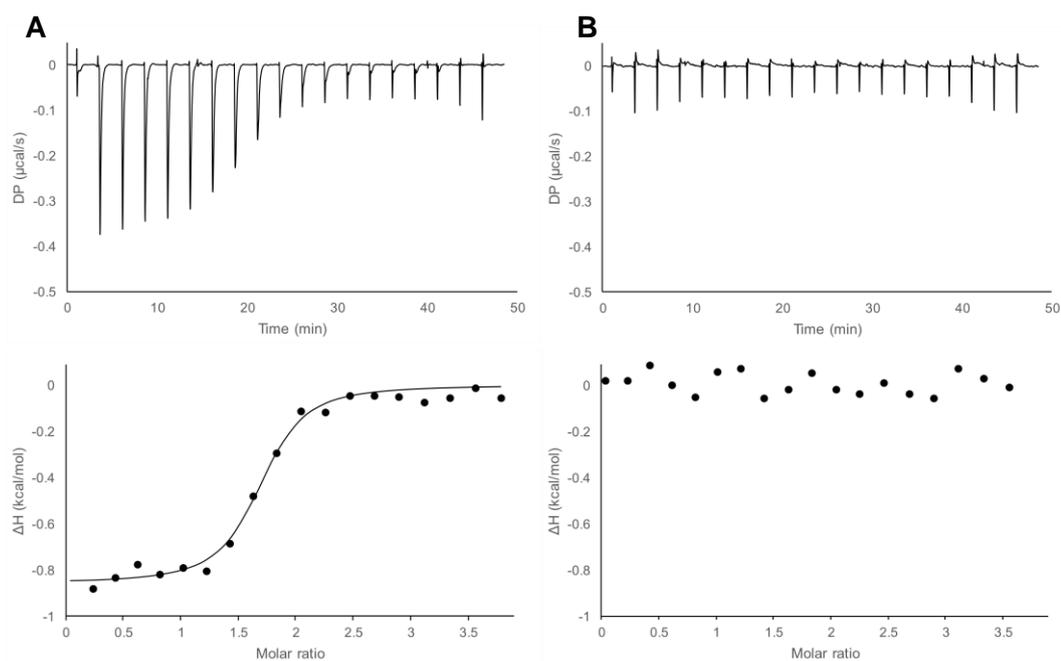


Figure 33: ITC isotherms of *RgSBP* binding to sialic acid derivatives

A) *RgSBP* binding to 2,7-anhydro-Neu5Ac. **B)** *RgSBP* binding to Neu5Ac. The data shown are representative of triplicate readings.

4.2.6 Structural studies of *RgSBP*

To gain structural insights into the ligand specificity of *RgSBP*, several batches of purified recombinant protein were provided for X-ray crystallography studies. Unfortunately, despite numerous trials by David Owen at the Diamond Light Source, the structure has not yet been solved. The binding of *RgSBP* to 2,7-anhydro-Neu5Ac was further analysed by STD (saturation transfer difference) NMR epitope mapping and DEEP (differential epitope) -STD NMR studies by Ridvan Nepravishta working in the group led by Jesus Angulo at the University of East Anglia (Norwich, UK). Epitope mapping (Figure 34a) revealed that protons H3, H4 and H6 showed the highest STD (%) factors, indicating that these protons would interact most closely with the protein. Conversely, protons H7, H8, H9 and those belonging to the CH₃ group showed lower STD (%) and are likely to not associate directly with the protein. Due to the lack of a crystal structure and the large molecular weight of *RgSBP* (49 kDa), it was not possible to determine individual residues involved in the binding of 2,7-anhydro-Neu5Ac by DEEP-STD NMR. It was however possible to determine that protons H4, H6, H7, H8 and H9' were favourably oriented towards aromatic residues and that H3 and protons belonging to the CH₃ group preferentially orientate towards aliphatic residues (Figure 34b).

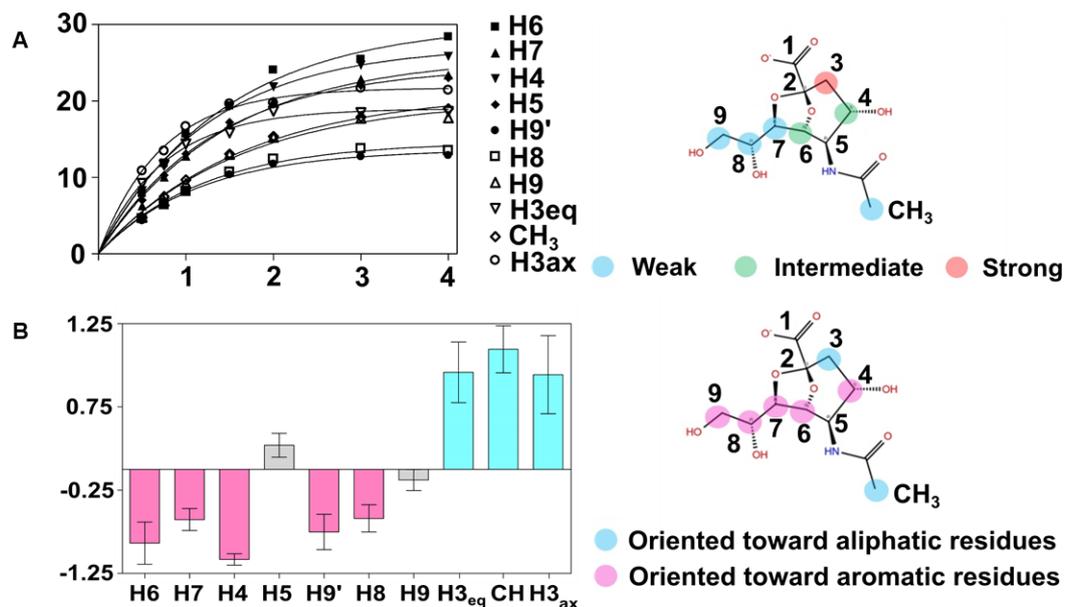


Figure 34: STD NMR analysis of the interaction between *RgSBP* and 2,7-anhydro-Neu5Ac

A) Binding epitope mapping of 2,7-anhydro-Neu5Ac interacting with *RgSBP*. The initial slopes STD_0 (%) were normalised against the highest STD_0 , assigned as 100%. The obtained factors were classified as weak (0-60%), intermediate (60-80%) or strong (80-100%). **B)** Average DEEP-STD NMR factors for 2,7-anhydro-Neu5Ac obtained saturating *RgSBP* in spectral regions 0.6, 0.78, 1.44 ppm for aliphatic and 7.5, 7.23, 7.27 for aromatic residues.

4.2.7 Neu5Ac is the preferential substrate for *RgNanA*

The specificity of the transporter explains why *R. gnavus* ATCC 29149 is able to grow on 2,7-anhydro-Neu5Ac but not Neu5Ac. However, how 2,7-anhydro-Neu5Ac is then metabolised inside the cell remains unknown. SSN analysis suggested that the *R. gnavus* aldolase falls into a distinct sialic acid aldolase clade (Figure 28). To investigate this further, the amino acid sequence of the *R. gnavus* aldolase (*RgNanA*) was compared with aldolase sequences of known bacterial Neu5Ac utilisers, *E. coli*, *S. Typhimurium*, *P. multocida* and *C. difficile*. This bioinformatics analysis showed that whilst the key active site residues Y139 and K167 were conserved in *RgNanA*, there were changes in residues previously identified to be involved in the binding to Neu5Ac in these strains. These amino acid changes are T50S, T169S and S211G (Figure 35) and potentially point to the sialic acid aldolase being active against a modified sialic acid substrate.

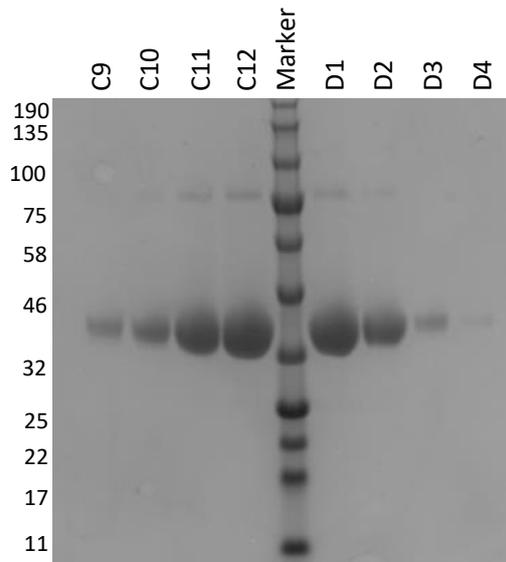


Figure 36: SDS-PAGE analysis of recombinant RUMGNA_02692 (*RgNanA*) purification

Fractions were analysed on a 4-12% BIS-TRIS gel, SM – starting material before gel filtration, marker – N.E.B broadrange pre-stained protein marker (expected size 37.1 kDa). C9 – D4 correspond to representative fractions from the major peak following gel filtration.

The ability of *RgNanA* to convert sialic acid into ManNAc and pyruvate was first assessed by Thin Layer Chromatography (TLC). The data showed that Neu5Ac was converted to ManNAc by *RgNanA* or by the *E. coli* aldolase (*EcNanA*) used as a control, but there was no conversion of 2,7-anhydro-Neu5Ac (Figure 37).

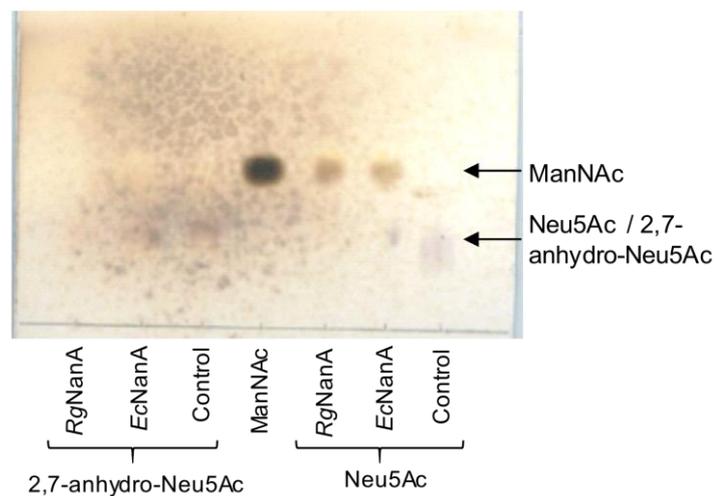


Figure 37: TLC analysis of sialic acid aldolase reaction products

2,7-anhydro-Neu5Ac or Neu5Ac were incubated with *RgNanA*, a commercial *E.coli* sialic acid aldolase (*EcNanA*) or no enzyme (control) at 37 °C in NaP for 24 hours. The TLC was eluted in 2:1:1 Butanol: Acetic acid: H₂O.

To confirm the nature of the reaction products, the products were 2-AB labelled and analysed by HPLC. Incubating *RgNanA* or *EcNanA* with Neu5Ac led to the appearance of a peak corresponding to ManNAc. However, no peak corresponding to ManNAc was observed when the enzymes were incubated with 2,7-anhydro-Neu5Ac (Figure 38).

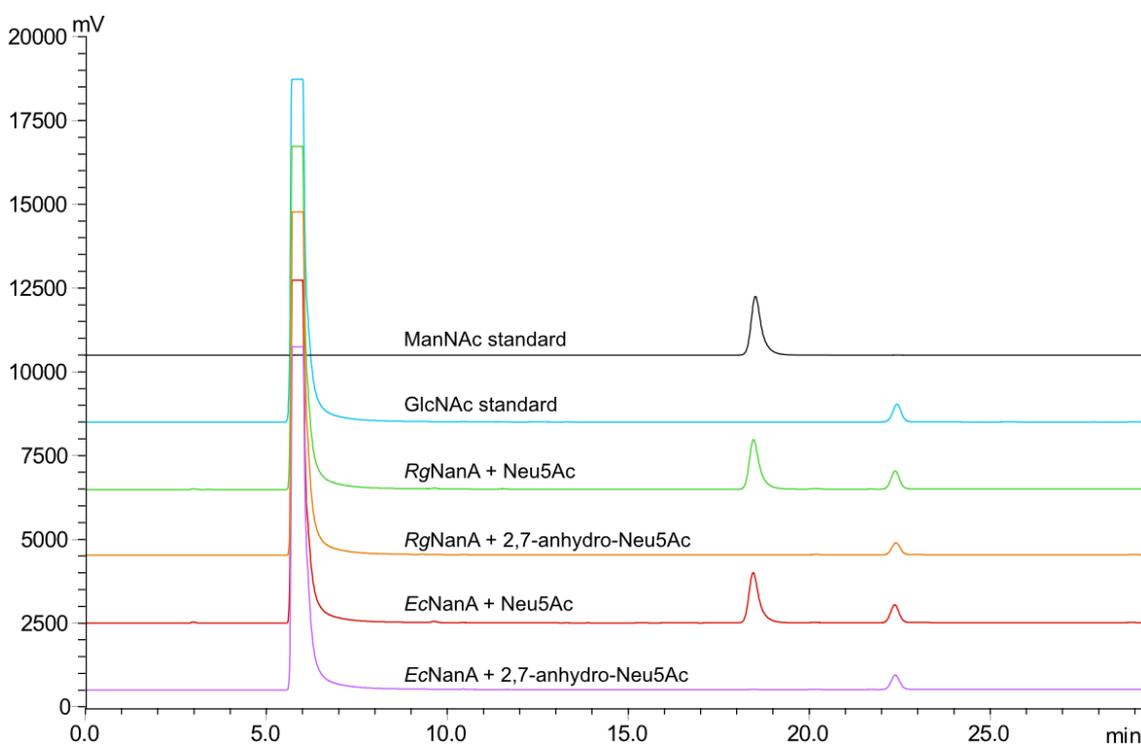


Figure 38: HPLC analysis of 2-AB labelled sialic acid aldolase reactions

RgNanA or a commercial *E. coli* sialic acid aldolase were incubated with 2,7-anhydro-Neu5Ac or Neu5Ac at 37 °C in NaP for 24 hours. GlcNAc was added as an internal standard in each analysis.

The kinetic parameters of *RgNanA* were then determined using a coupled activity assay where the pyruvate released from the conversion of Neu5Ac to ManNAc was monitored by coupling to lactate dehydrogenase and measuring the loss of absorbance at 340 nm as NADH was converted to NAD⁺ (Figure 39a). The initial rate of reaction with increasing concentrations of Neu5Ac was determined from the Michaelis-Menten curve (Figure 39b). The k_{cat} was calculated to be $2.757 \pm 0.033 \text{ s}^{-1}$ and the K_M $1.473 \pm 0.098 \text{ mM}$. These results are in line with kinetic parameters of other bacterial sialic acid aldolases functionally characterised to date (Table 6) (Wang *et al.*, 2018).

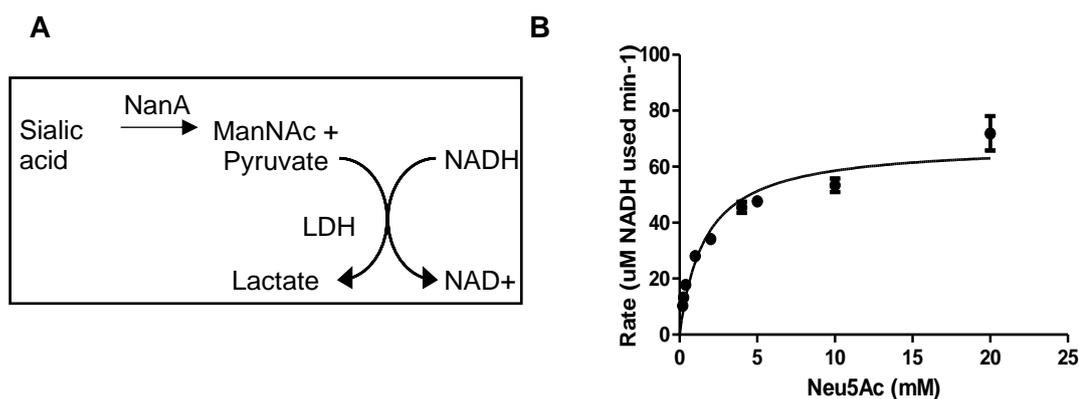


Figure 39: Enzymatic assay of *RgNanA* with Neu5Ac as substrate

A) Schematic representation of coupled sialic acid aldolase enzymatic assay. **B)** Michaelis-Menten plot of the rate of reaction for *RgNanA* with increasing concentration of Neu5Ac substrate in 50 mM NaP pH 7.0. The rate of reaction at each Neu5Ac concentration was determined in triplicate by measuring the drop in absorbance at 340 nm due to conversion of NADH to NAD⁺ in the coupled reaction.

Table 6: Kinetic parameters of characterised NanA proteins using figures taken from Wang *et al.*, 2018.

Species	K_m (mM)	K_{cat} (s ⁻¹)	K_{cat} / K_m (s ⁻¹ mM ⁻¹)
<i>R. gnavus</i>	1.5	2.757	1.838
<i>S. mycoplasma</i>	1.8	6.588	3.660
<i>C. glutamicum</i>	33.5	9.380	0.280
<i>E. coli</i>	3.5	83.125	23.750
<i>P. multocida</i>	4.9	18.375	3.750
<i>C. perfringens</i>	3.2	16.032	5.010
<i>S. carnosus</i>	2.0	4.000	2.000
<i>L. plantarum</i>	1.8	10.080	5.600

In parallel, the purified recombinant protein was provided for X-ray crystallography studies and the crystal structure of *RgNanA* solved at 1.6 Å resolution by David Owen (Diamond Light Source, Didcot, UK). *RgNanA* showed the typical fold of sialic acid aldolases consisting of a (β/α)₈ TIM barrel with an adjacent three-helix bundle. The *RgNanA* active site was structurally homologous to that of known Neu5Ac utilisers including *C. difficile*, *E. coli* and *P. multocida* (Figure 40a). To obtain a crystal structure in complex with Neu5Ac, an active site mutant (K167A) was generated by site directed

mutagenesis and purified as above. Soaking of *RgNanA* K167A with Neu5Ac yielded a crystal structure in complex with the open chain ketone form of Neu5Ac. Neu5Ac forms extensive interactions with the enzyme active site, hydrogen bonding to the side chains of Ser49, Ser50, Ser 169, Asp194, Glu 195 and Tyr257 and to the backbone of Ser50, Gly192, Asp194, Gly 211 (Figure 40b). A close inspection of the *RgNanA* active site showed that residues found to be important for the catalytic activity of the *E. coli* enzyme (Daniels *et al.*, 2014) are conserved in the *R. gnavus* protein with the exception of *E. coli* Thr167 which is Ser169 in *RgNanA*, however, the hydroxyl groups of both superimpose (Figure 40a). Furthermore, the *EcNanA* T167S mutation has been reported to have no impact on the catalytic activity of the enzyme (Daniels *et al.*, 2014). *RgNanA* also showed similar kinetic parameters to previously characterised bacterial sialic acid aldolases (Figure 39, Table 6). Despite differences in highly conserved residues found in the active site as compared to other bacterial sialic acid aldolases, these experiments demonstrate that Neu5Ac is the substrate for *RgNanA* and that the enzyme does not act on 2,7-anhydro-Neu5Ac.

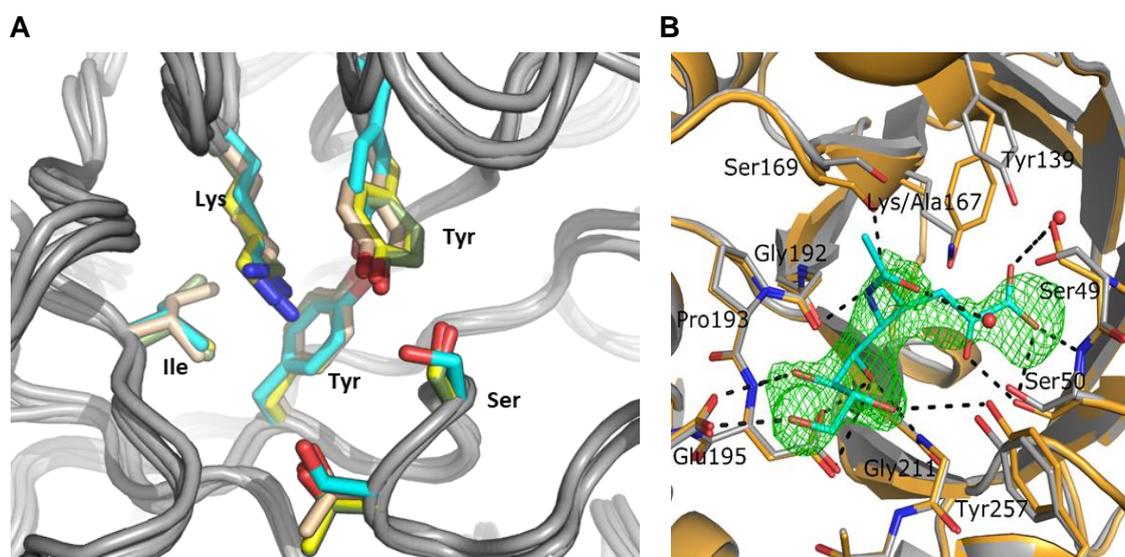


Figure 40: Comparison of *RgNanA* crystal structure with characterised bacterial sialic acid aldolases

A) Crystal structure comparison of the active site of sialic acid aldolases from *R. gnavus* (cyan), *C. difficile* (yellow), *E. coli* (green) and *P. multocida* (brown), the backbone of each protein is shown in grey ribbons and residues important for binding are displayed as coloured sticks. **B)** The *RgNanA* K167A active site is shown in orange with the open chain Neu5Ac bound shown in cyan. The green mesh represents the Neu5Ac $F_o - F_c$ difference map at the 3σ level. Hydrogen bonding interactions are depicted using black dashed lines. In addition, the unbound *RgNanA* wild type active site is shown in grey.

4.2.8 RUMGNA_02695 catalyses the interconversion of 2,7-anhydro-Neu5Ac to Neu5Ac

The findings that *RgSBP* is specific for 2,7-anhydro-Neu5Ac (see section 4.2.5) and that *RgNanA* is specific for Neu5Ac (see section 4.2.7) suggest that, once transported into the cell, 2,7-anhydro-Neu5Ac needs to be converted to Neu5Ac to become a substrate for *RgNanA*. The conversion of 2,7-anhydro-Neu5Ac to Neu5Ac is likely to be carried out by one of the enzymes within the *R. gnavus nan* cluster. Bioinformatic analysis (see section 4.2.4) showed that RUMGNA_02695 co-occurred with a similar group of species to the *RgNanA* and *RgSBP* (Figure 28, Table 5) now shown to be involved in the 2,7-anhydro-Neu5Ac utilisation pathway, whereas RUMGNA_02700 and 02701 did not co-occur with similar species in the same way, suggesting that RUMGNA_02695 is a strong candidate for the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac. To test this hypothesis, the full gene sequence was amplified from *R. gnavus* ATCC 29149 genomic DNA, cloned and heterologously expressed in *E. coli*, before purification of the recombinant enzyme by IMAC, Ion Exchange (IEX) and gel filtration. SDS-PAGE analysis showed the presence of a band at the expected molecular weight of 45.2 kDa (Figure 41).

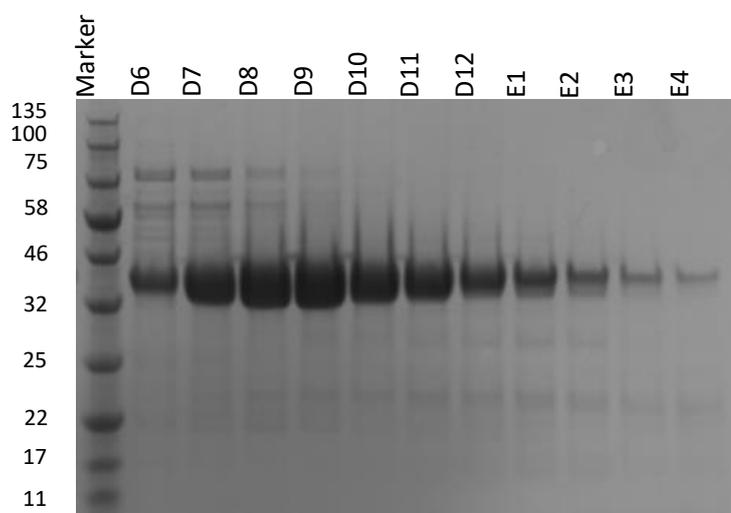


Figure 41: SDS-PAGE analysis of recombinant RUMGNA_02695 purification

Fractions were analysed on a 4-12% BIS-TRIS gel, SM – starting material before gel filtration, marker – N.E.B broadrange pre-stained protein marker (expected size 45.2 kDa). D6 – E4 correspond to representative fractions from the major peak following gel filtration.

The uniprot annotation of RUMGNA_02695 describes an oxidoreductase with a Rossmann fold, suggesting that the enzyme activity may require either NAD or FAD cofactors to accept or donate electrons. Therefore, the activity of the recombinant protein was tested with 2,7-anhydro-Neu5Ac as a potential substrate in the presence of NAD⁺,

NADH or FAD as a potential cofactor. The products of each reaction were DMB labelled to enable detection of Neu5Ac by HPLC as previously reported (Monestier *et al.*, 2017). It is of note that 2,7-anhydro-Neu5Ac cannot be labelled by DMB as carbon 2 is engaged in the C2, C7 anhydro bridge which makes it not accessible to the DMB dye (see section 3.1). Neu5Ac was detected as a reaction product when the enzyme was incubated with 2,7-anhydro-Neu5Ac in the presence of either NAD⁺ or NADH, but not in the presence of FAD or in the absence of a cofactor (Figure 42).

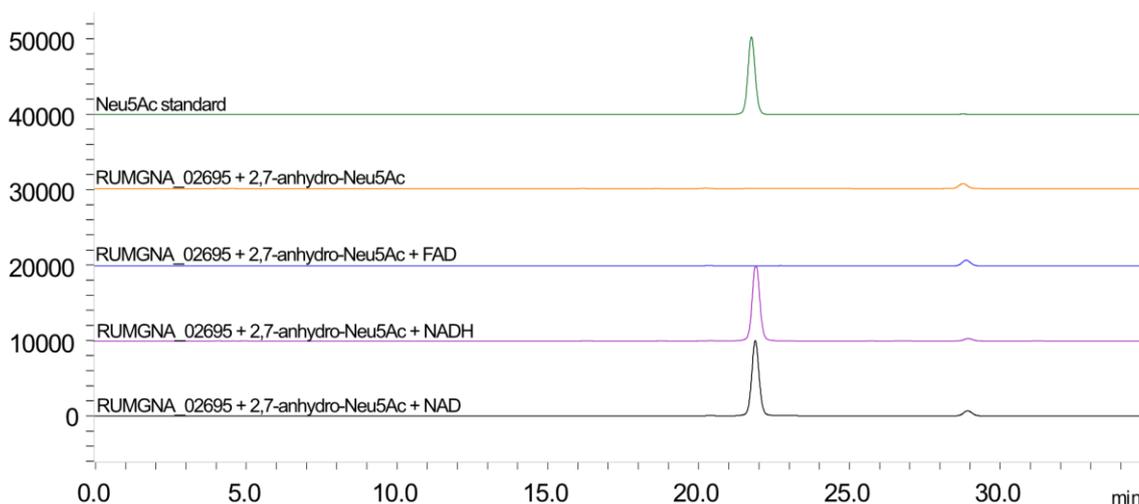


Figure 42: HPLC analysis of recombinant RUMGNA_02695 reactions with 2,7-anhydro-Neu5Ac HPLC analysis of DMB-labelled RUMGNA_02695 reactions with 2,7-anhydro-Neu5Ac using different co-factors. NAD (black), NADH (pink), FAD (blue), no co-factor (brown), and a Neu5Ac standard (green).

The enzymatic reaction was further monitored by direct injection negative mode ESI-MS as this also allows detection of 2,7-anhydro-Neu5Ac (see section 3.2.2). This analysis confirmed the conversion to Neu5Ac observed by HPLC when NAD⁺ or NADH was used as a cofactor. Interestingly, the reactions did not go to completion showing a ratio of 1:2 for 2,7-anhydro-Neu5Ac:Neu5Ac (Figure 43a). This result suggested that the enzymatic reaction may be reversible. To test this hypothesis, the recombinant enzyme was incubated with Neu5Ac in the presence of either NAD⁺ or NADH and the reaction products analysed by ESI-MS. The analysis showed that the enzyme could produce 2,7-anhydro-Neu5Ac from Neu5Ac (Figure 43b), reaching an equilibrium with a ratio of approximately 1:2 as observed previously (Figure 43a), confirming that the reaction is reversible, favouring Neu5Ac as a reaction product.

To further investigate the role of the cofactors in the enzymatic reaction, experiments were set up to monitor the fate of NAD⁺/NADH over the course of the reaction. Since

NADH but not NAD⁺ absorbs at 340 nm, the redox state of NAD⁺/NADH could be monitored in a simple plate reader assay. Whether the enzyme was incubated with Neu5Ac or 2,7-anhydro-Neu5Ac as starting substrate and NAD⁺ or NADH as a cofactor, no change in absorbance at 340 nm was detected, suggesting that the enzyme mechanism may involve oxidation and reduction of NAD⁺/NADH cofactor (Figure 43c).

Since no net change in NADH concentration was observed during the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac by RUMGNA_02695, the kinetic parameters of the enzymatic reaction were determined by coupling with the *RgNanA* assay in the presence of an excess of *RgNanA* and LDH (Figure 43d). To confirm that a sufficient excess of *RgNanA* was used, preliminary reactions were conducted using 10 mM 2,7-anhydro-Neu5Ac and, following establishment of an initial rate, 1 mM Neu5Ac was added. This led to a rapid drop in the absorbance, confirming that the *RgNanA* excess was sufficient and that RUMGNA_02695 was the rate limiting enzyme. For the kinetics studies, an increasing concentration of 2,7-anhydro-Neu5Ac from 0.01 mM to 1 mM was used. Under these conditions the k_{cat} was calculated to be $0.0824 \pm 0.0043 \text{ s}^{-1}$ and the K_{M} $0.074 \pm 0.014 \text{ mM}$ (Figure 43d).

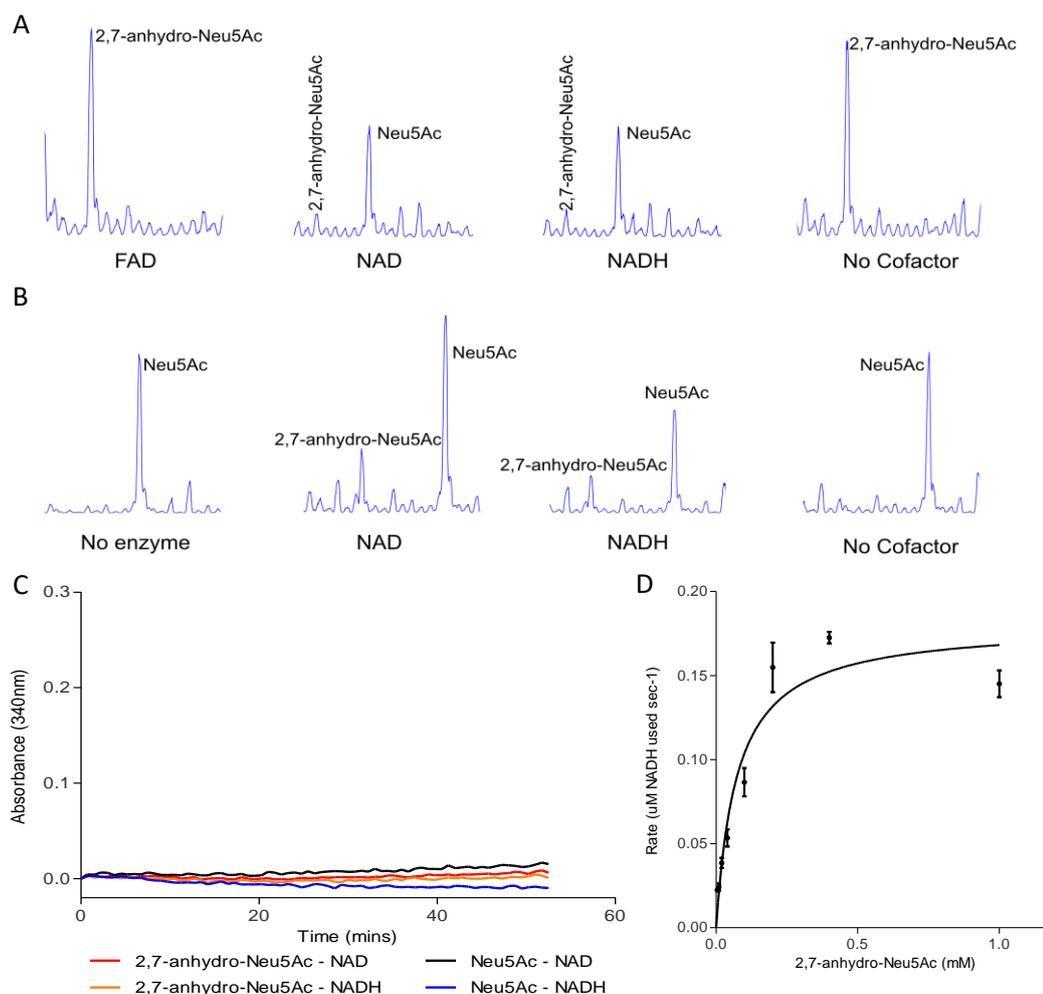


Figure 43: Cofactor requirement and enzyme kinetics of RUMNGNA_02695

Direct injection negative mode Mass Spectrometry analysis of RUMNGNA_02695 reactions with **A**) 2,7-anhydro-Neu5Ac (MW-1 = 290) or **B**) Neu5Ac (MW-1 = 308), with different cofactors defined in the figure. **C**) Change in absorbance at 340 nm over time for reactions of RUMNGNA_02695 with either 2,7-anhydro-Neu5Ac or Neu5Ac as substrate and either NAD⁺ or NADH as cofactor. **D**) Michaelis-Menten plot of the rate of reaction for RUMNGNA_02695 with increasing concentrations of 2,7-anhydro-Neu5Ac. The rate of reaction (μM NADH) at each concentration was determined in triplicate by measuring $A_{340\text{nm}}$ change and using a standard curve.

To further understand the mechanism of action of RUMNGNA_02695, the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac was monitored by ¹H NMR, in collaboration with Serena Monaco in Jesus Angulo's group (University of East Anglia, Norwich, UK). The conversion of 2,7-anhydro-Neu5Ac to Neu5Ac was monitored at 15 min intervals for 24 hours in heavy water (D₂O) or light water (H₂O) at 20 °C. This analysis revealed the presence of a third species (XY), which appeared at its highest concentration at early time points before reducing, suggesting that it is an intermediate product (Figure 44a). Equilibrium in the reaction was reached at 18 hours and integration of the respective

acetyl methyl groups showed a final ratio of 1: 0.16: 0.09 for Neu5Ac: 2,7-anhydro-Neu5Ac: XY. In the D₂O sample, the peaks associated to the axial protons on C3 and C5 of the Neu5Ac product were absent (and the multiplicity of the neighbouring protons was simplified), strongly suggesting that the reaction chemistry occurs around these carbons. Further, heteronuclear single quantum correlation (HSQC) analysis of the reaction mixture containing the intermediate at higher concentration was carried out and compared with the HSQCs of Neu5Ac and 2,7-anhydro-Neu5Ac alone (Figure 44b). For Neu5Ac and 2,7-anhydro-Neu5Ac eight proton-carbon peaks are present, however, by HSQC, only seven proton-carbon peaks not belonging to the substrate or product can be counted, suggesting that one of the XY intermediate carbons could be fully oxidised.

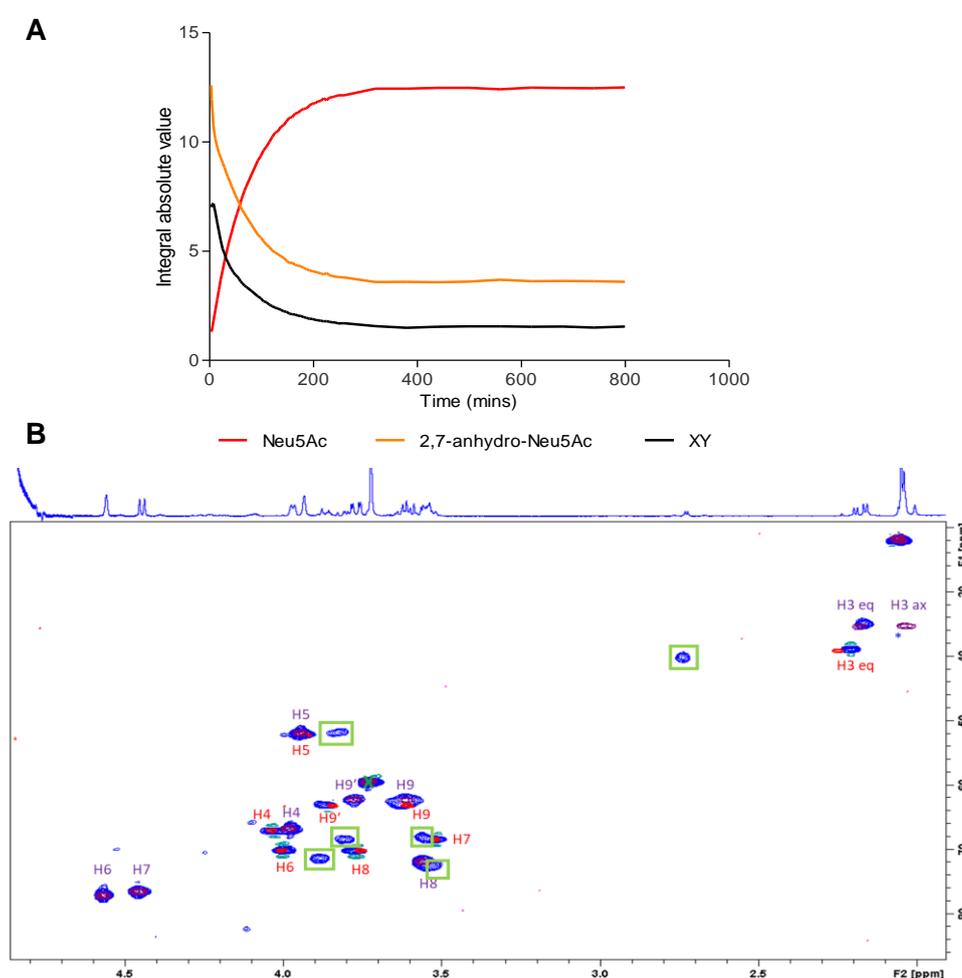


Figure 44: NMR characterisation of the RUMGNA_02695 reaction

A) integration of the acetyl methyl groups for Neu5Ac, 2,7-anhydro-Neu5Ac or XY intermediate from 2D NMR analysis of the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac via the XY intermediate. Spectra were obtained every 15 min. **B)** HSQC analysis of the reaction compared with the HSQCs of Neu5Ac and 2,7-anhydro-Neu5Ac alone. Proton-carbon peaks corresponding to the reactions are shown in blue, those corresponding to Neu5Ac in purple and those corresponding to 2,7-anhydro-Neu5Ac in red. Proton-carbon peaks not corresponding to Neu5Ac or 2,7-anhydro-Neu5Ac are denoted by green boxes.

The recombinant RUMGNA_02695 was provided to Jim Naismith's lab for X-ray crystallography and the crystal structure of the apo protein was solved by Micah Lee, using the structure of an SDR (short chain dehydrogenase/reductase) enzyme from *Agrobacterium radiobacter* as a scaffold (PDB 5UI9_B). The protein shows a typical SDR type fold, characterised by a central β -sheet with helices on either side, with a dimeric conformation and the NAD cofactor can be seen bound to the protein (Figure 46a). In contrast to the majority of SDR proteins, RUMGNA_02695 does not appear to have the complete catalytic tetrad typical of SDR proteins (Persson *et al.*, 2009), lacking the asparagine residue in the active site. A putative active site containing the putative active site lysine is shown in Figure 46b. So far attempts to co-crystallise the protein with the substrate have been unsuccessful but a predicted substrate binding pocket can be identified close to the NAD cofactor.

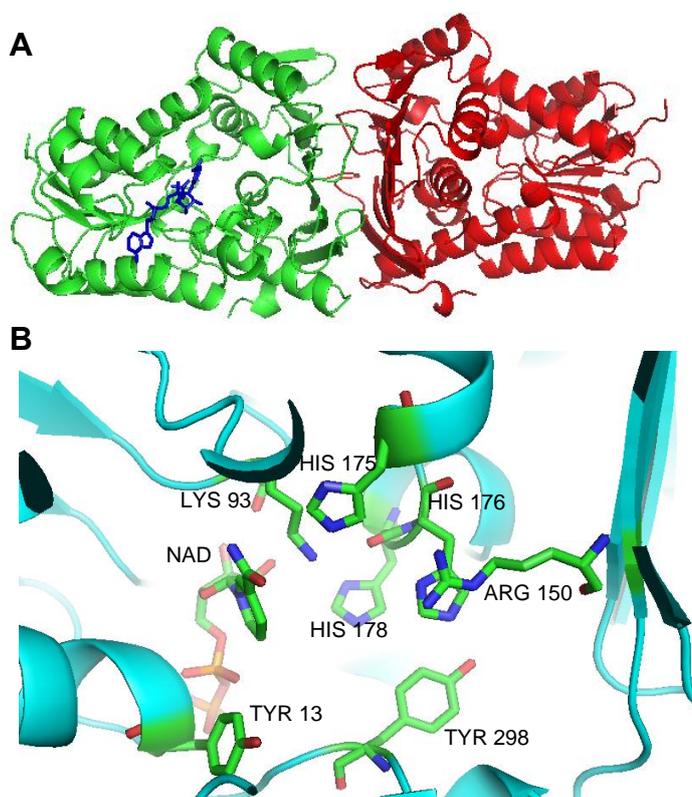


Figure 46: Crystal structure of recombinant RUMGNA_02695

A) Dimeric structure of RUMGNA 02695 shown in green and red, with the NAD cofactor bound (blue). **B)** Structure of putative active site of RUMGNA_02695, protein backbone is shown in cyan cartoon with residues and NAD shown in sticks.

These data indicate that RUMGNA_02695 is a novel oxidoreductase whose activity is critical for the conversion of 2,7-anhydro-Neu5Ac into Neu5Ac, which will then become a substrate for *RgNanA*. We therefore refer to RUMGNA_02695 as *RgNanOx* in the rest of the study.

4.2.9 Remaining 'uncharacterised' proteins from the *R. gnavus nan* cluster

Among the proteins of unknown function in the *R. gnavus nan* cluster, RUMGNA_02699 is a predicted transcriptional regulator, RUMGNA_02697 and 02696 are predicted to be the permeases for the ABC transporter protein, and RUMGNA_02693 and 02691 share a high degree of similarity to characterised NanE and NanK proteins involved in Neu5Ac metabolism, and where therefore not characterised here. RUMGNA_02701 shares some homology with NanS proteins responsible for de-acetylation of sialic acids (Rangarajan *et al.*, 2011, Steenbergen *et al.*, 2009), and one of the proteins co-occurring with RUMGNA_02700 in the SSN analysis is annotated as a potential isomerase of sialic acid metabolism. To assess the function of these two predicted proteins, the full gene sequences were amplified, cloned and heterologously expressed in *E. coli*. Unfortunately, the RUMGNA_02701 recombinant protein produced remained insoluble under the conditions tested, including expression in different media and changing the temperature and volume used for expression (Figure 47). For RUMGNA_02700 little or no expression was observed despite attempts to optimise the expression by altering growth media, temperature or volume (Figure 48).

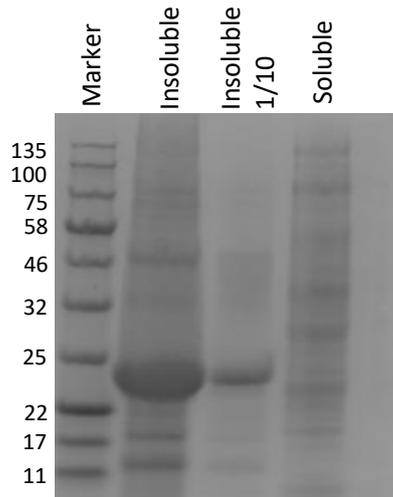


Figure 47: SDS-PAGE analysis of recombinant RUMGNA_02701 expression

The soluble and insoluble fractions following expression were analysed. The insoluble fraction was resuspended in an equal volume of extraction buffer and diluted 1/10. Marker – N.E.B broadrange pre-stained protein marker (expected size 45.2 kDa).

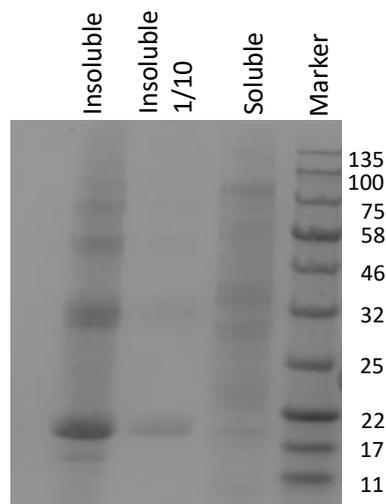


Figure 48: SDS-PAGE analysis of recombinant RUMGNA_02700 expression

The soluble and insoluble fractions following expression were analysed. The insoluble fraction was resuspended in an equal volume of extraction buffer and diluted 1/10. Marker – N.E.B broadrange pre-stained protein marker (expected size 18.4 kDa).

4.3 Discussion

This chapter discusses the identification and characterisation of proteins involved in the metabolic pathway for 2,7-anhydro-Neu5Ac utilisation by *R. gnavus*. The ability to metabolise sialic acid has been shown to be important for *R. gnavus* in mucin glycan utilisation *in vitro* (Crost *et al.*, 2016). The *R. gnavus nan* cluster associated with sialic acid metabolism was previously shown to be upregulated during growth on mucin, and here it was demonstrated that growth on 2,7-anhydro-Neu5Ac induces expression of the extended cluster, suggesting that the *nan* cluster is dedicated to 2,7-anhydro-Neu5Ac metabolism. To metabolise an exogenous substance, a transport system is required for the uptake of the compound into the cells. The *R. gnavus nan* cluster encodes a SAT2 ABC transporter made up of two predicted permeases (RUMGNA_02696 and 02697) and one binding protein (RUMGNA_02698 or *RgSBP*) (Crost *et al.*, 2016). Studying the SBP revealed that this ABC system is specific for 2,7-anhydro-Neu5Ac, whilst showing no affinity to Neu5Ac; explaining why *R. gnavus* is unable to utilise Neu5Ac despite encoding a cluster for sialic acid utilisation. *RgSBP* binds 2,7-anhydro-Neu5Ac with a K_d of $2.42 \pm 0.27 \mu\text{M}$, this is lower than the typical nm range seen for characterised SAT transporters for Neu5Ac (Gangi Setty *et al.*, 2018), perhaps reflecting the exclusive access *R. gnavus* has towards 2,7-anhydro-Neu5Ac. This is the first characterised SAT2 transporter, and the first shown to transport 2,7-anhydro-Neu5Ac. This type of transporter has also been identified in *S. pneumoniae*, *S. gordonii* and *S. sanguinis* suggesting that these species may also be able to transport 2,7-anhydro-Neu5Ac (Almagro-Moreno and Boyd, 2009).

Once inside the cell 2,7-anhydro-Neu5Ac can be metabolised. In Neu5Ac utilisation the first step in metabolism is the conversion of Neu5Ac to ManNAc and pyruvate by a sialic acid aldolase (Brigham *et al.*, 2009, Plumbridge and Vimr, 1999). Bioinformatic analyses of *RgNanA* showed that although classified as a sialic aldolase, it formed part of a unique clade in SSN analysis with differences within the binding domain which we hypothesised may be involved in the recognition of a different sialic acid substrate. However, recombinant *RgNanA* was shown to be active against Neu5Ac and not 2,7-anhydro-Neu5Ac with kinetics similar to other characterised sialic acid aldolases. In addition, its crystal structure revealed a well conserved binding pocket with previously characterised Neu5Ac aldolases. Although these amino acid changes seemed to have little impact on protein function, they could point to a different evolutionary lineage for the *R. gnavus* like aldolases and therefore provide additional means of distinguishing Neu5Ac and 2,7-anhydro-Neu5Ac utilisers based on sequence analysis.

To become a substrate for *RgNanA*, 2,7-anhydro-Neu5Ac needs to be converted to Neu5Ac. Here, a novel enzymatic reaction was identified catalysed by *RgNanOx*. The enzyme, which is predicted to contain a Rossmann fold domain, utilises NAD⁺/NADH as a cofactor to reversibly convert 2,7-anhydro-Neu5Ac into Neu5Ac, with no net change in the oxidation state of NADH. The reversibility of the enzyme gives rise to a potential new method for 2,7-anhydro-Neu5Ac synthesis, using Neu5Ac as a starting material. The desired 2,7-anhydro-Neu5Ac can be separated from Neu5Ac using the methods developed in chapter 3. The method may also be useful for generating 2,7-anhydro-Neu5Ac derivatives, such as 2,7-anhydro-Neu5Gc or 2,7-anhydro-KDN as well as C¹³ labelled versions of 2,7-anhydro-Neu5Ac and optimisation is ongoing in the lab.

The *RgNanOx* apo structure was solved and found to have a typical SDR protein family fold. SDR proteins represent a large family of proteins utilising NAD(P)/H as cofactor. The enzymes in this family have a broad functional diversity with substrates ranging alcohols, sugars, steroids, aromatic compounds, lipids and xenobiotics (Persson *et al.*, 2009). Together with NMR analysis, the enzymatic and structural data allowed us to propose a mechanism of action for *RgNanOx*. Future research will aim to produce inactive mutants of the protein to obtain the crystal structure of *RgNanOx* in complex with the substrate or intermediate. A detailed characterisation of the active site and substrate binding pocket will allow identification of compounds that could be used to validate the proposed enzymatic mechanism. In addition, manipulation of the active site may enable the equilibrium of the protein to shift towards the production of 2,7-anhydro-Neu5Ac which could lead to the development of a high throughput economically viable protocol for the synthesis of 2,7-anhydro-Neu5Ac.

Once converted into ManNAc and pyruvate by *RgNanA*, ManNAc is expected to be further metabolised into GlcNAc-6-phosphate, following phosphorylation and epimerisation of ManNAc by *RgNanK* and *RgNanE* which both show high level similarity to the enzymes described to date. In summary, this work unravelled a novel pathway for *R. gnavus* 2,7-anhydro-Neu5Ac metabolism as depicted in Figure 49.

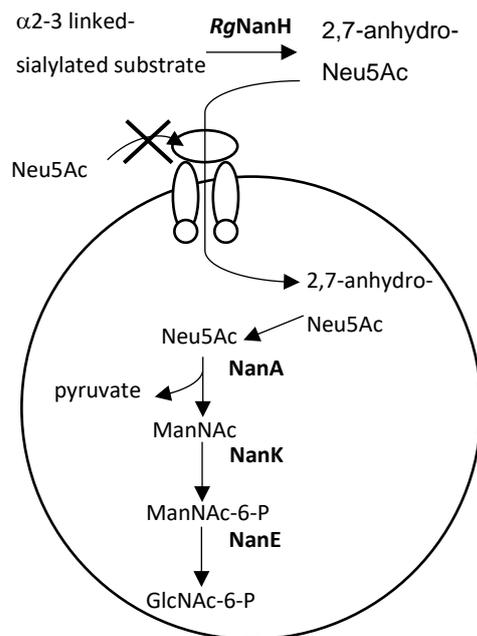


Figure 49: Proposed pathway of 2,7-anhydro-Neu5Ac metabolism by *R. gnavus*

2,7-anhydro-Neu5Ac is specifically transported inside the bacteria by *RgSBP* before being converted into Neu5Ac by an unknown mechanism (orange arrow), where it becomes a substrate for *RgNanA*, which converts into ManNAc which is further metabolised by NanK and NanE.

The RUMGNA_02701 protein shares some similarity with NanS proteins. These proteins are responsible for the deacetylation of sialic acids prior to metabolism (Phansopa *et al.*, 2015, Steenbergen *et al.*, 2009). As the *R. gnavus nan* cluster is dedicated to 2,7-anhydro-Neu5Ac it may be that RUMGNA_02701 is dedicated to deacetylation of 2,7-anhydro-Neu5Ac derivatives. Future work will aim to heterologously express soluble protein and assess its ability to deacetylate sialic acid derivatives. This could be achieved through further optimisation of the expression conditions, by producing modified constructs or refolding of the protein from urea or guanidine.

A protein co-occurring with RUMGNA_02700 in the SSN analysis is predicted to be an isomerase of Neu5Ac metabolism. One possibility is that the predicted protein may facilitate the opening of the Neu5Ac ring structure into the linear form catalysed by sialic acid aldolases to produce ManNAc and pyruvate. The recombinant protein was poorly expressed under the conditions tested here, purifying a sufficient amount of protein will help test for an increased activity when the protein is added to *RgNanA* kinetics assays.

The *R. gnavus nan* cluster encodes an AraC type transcriptional regulator (RUMGNA_02699). The arabinose regulator in *E. coli* represses cluster regulation in the absence of arabinose and induces expression in the presence of arabinose (Schleif,

2010). It is likely that the regulation of the *nan* cluster is mediated by 2,7-anhydro-Neu5Ac or a downstream metabolite. This could be confirmed in *R. gnavus* by creating mutants of specific *nan* cluster genes (using the methods discussed in section 5.2.1.1), such as *RgNanOx*, that would block the metabolic process and assessing the expression of the cluster when 2,7-anhydro-Neu5Ac is added exogenously.

Bioinformatics analysis revealed that this cluster is a unique sialic acid metabolic pathway, specific to few bacterial species, this analysis was limited to those clusters that contained an IT-sialidase. However, it is known that some bacteria are able to scavenge Neu5Ac but lack the sialidase required to release it (Nishiyama *et al.*, 2018, Ng *et al.*, 2013), in future work it would be interesting to investigate whether any of the bacteria co-occurring with *RgSBP* or *RgNanOx* as shown by SSN analysis, are able to utilise 2,7-anhydro-Neu5Ac, independent of whether or not they encode an IT-sialidase. Furthermore, there are species that encode homologs of *RgNanOx*, but do not appear to have a homolog of the transporter protein including strains of *E. coli* and *S. Typhimurium* LT2, which may suggest the presence of other 2,7-anhydro-Neu5Ac transporters or, *RgNanOx* may have another role in these bacteria.

Chapter 5

Impact of the *R. gnavus nan* cluster in gut symbionts

5.1 Introduction

In the previous chapter, the complete pathway of 2,7-anhydro-Neu5Ac metabolism in *R. gnavus* was unravelled. How widely distributed among the gut microbiota this pathway is has yet to be determined, although our SSN analysis and multigene cluster analysis suggested that the distribution of the *R. gnavus* extended *nan* cluster was restricted to a small subset of bacterial species (see section 4.2.4). We therefore hypothesised that 2,7-anhydro-Neu5Ac may be metabolised by a limited number of species in the gut, therefore conferring a competitive advantage *in vivo* over other members of the microbiota and that it may be used to selectively promote *R. gnavus* growth.

In this chapter we will first investigate the effect of 2,7-anhydro-Neu5Ac supplementation on the faecal microbial composition using a human-based *in vitro* batch culture system. Taxonomic analysis will give insight into the bacterial species able to benefit either directly or indirectly from 2,7-anhydro-Neu5Ac.

In order to test this hypothesis that the *nan* cluster confers *R. gnavus* strains a competitive advantage, we developed a method which enabled for the first time to genetically manipulate *R. gnavus*. We generated *R. gnavus nan* mutants which were characterised *in vitro* and used for *in vivo* colonisation assays in germ-free mice.

5.2 Results

5.2.1 2,7-anhydro-Neu5Ac preferentially promotes *R. gnavus* growth in an *in vitro* colon model

To determine the impact of 2,7-anhydro-Neu5Ac on the faecal microbial community, anaerobic *in vitro* batch culture experiments were carried out using 2,7-anhydro-Neu5Ac, 3'SL or Neu5Ac as a sole carbon source in a simulated model colon under anaerobic conditions. The faecal microbiome of seven human donors were first screened in the group by Sandra Tribolo for the presence of *R. gnavus* using qPCR and primers specific to the 16S of *R. gnavus*, specific to *RgSBP* and universal bacterial 16S primers. Despite the report that 90% of the population harbour *R. gnavus* (Qin *et al.*, 2010), only two out of the seven donors tested were found to be positive for *R. gnavus*, CM001 and CM100 (Figure 50a). Of these, the sample from donor CM001 showed a Ct value (the number of cycles taken for the signal to cross the threshold) of 22.47 whereas the sample of CM100 donor displayed a Ct value of 29.81, compared to values of 34-36 for the other samples which is similar to value obtained with no template controls. However, while CM001 showed the presence of the *RgSBP* gene, as determined by qPCR, CM100 did not (Figure 50a), suggesting that the strain(s) of *R. gnavus* present in CM100 cannot utilise 2,7-anhydro-Neu5Ac. CM001 was therefore chosen as the donor for the batch culture experiments.

The faecal slurry from the CM001 donor was used to seed batch models containing 2,7-anhydro-Neu5Ac, 3'SL, Neu5Ac or no additional carbon source (control). In the control and Neu5Ac supplemented vessels, no growth of *R. gnavus* was observed, whereas in the 3'SL and 2,7-anhydro-Neu5Ac supplemented vessels *R. gnavus* growth was detected by qPCR (Figure 50b). *R. gnavus* growth was more apparent in the 2,7-anhydro-Neu5Ac batch, supporting the hypothesis that 2,7-anhydro-Neu5Ac can be metabolised by a limited number of species.

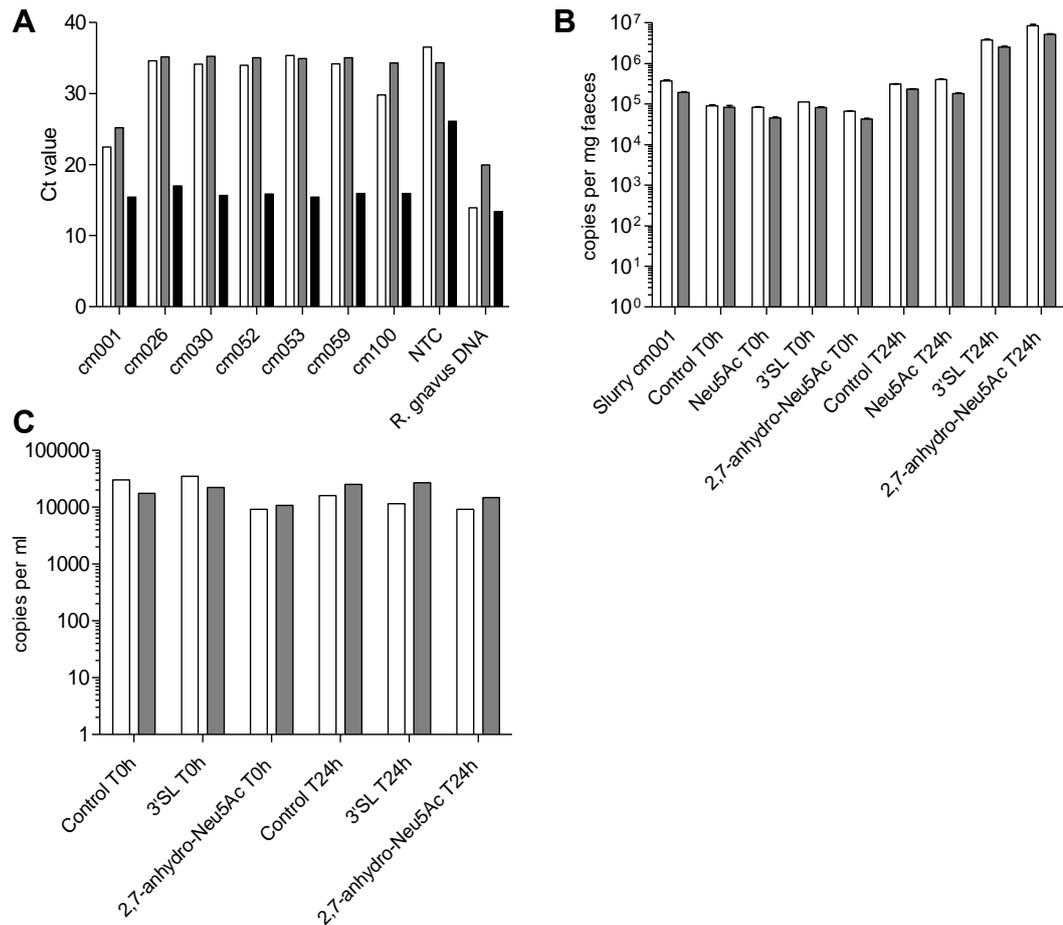


Figure 50: Determination of *R. gnavus* growth in an *in vitro* colon model

qPCR was carried out using primers specific to *R. gnavus* 16S (white bars), *RgSBP* (grey bars) or universal 16S primers (black bars). **A**) Ct values for qPCR of DNA extracted from faecal samples of seven donors, to identify donors with *R. gnavus* in their microbiome. **B**) *R. gnavus* and *RgSBP* copies per mg faeces of samples taken from batch culture fermentations, using Neu5Ac, 3'SL, 2,7-anhydro-Neu5Ac or no carbon source (control) after 0 and 24 hours performed by Sandra Tribolo. **C**) Repeat of the experiment in B without Neu5Ac supplementation with data presented per ml of sample taken.

I repeated the experiment using the same donor (CM001) and conditions. However, the growth of *R. gnavus* in the presence of 2,7-anhydro-Neu5Ac or 3'SL was not observed (Figure 50c). To investigate this further, anaerobic mono-cultures of *R. gnavus* ATCC 29149 were grown in the medium used in the colon model in the presence of glucose, Neu5Ac or 3'SL as a sole carbon source. *R. gnavus* did not grow under any of these conditions, indicating that the batch culture medium alone is not capable of supporting *R. gnavus* growth (Figure 51a). To determine if a key component was lacking from the medium, *R. gnavus* was grown in a 50/50 mix of the batch medium and BDM (a defined media *R. gnavus* was previously shown to grow in; see section 6.2.4.1). This 50/50

mixture did support *R. gnavus* growth, suggesting that additional supplementation to the batch medium was required (Figure 51b). A detailed comparison of the composition of the two media revealed that histidine and vitamin B₁₂ were present in BDM but not in the *in vitro* colon batch medium. Next, *R. gnavus* mono-cultures were grown in the batch medium supplemented with histidine, vitamin B₁₂ or both. The presence of vitamin B₁₂ allowed for the growth of *R. gnavus*, suggesting that *R. gnavus* is auxotrophic for vitamin B₁₂ in a minimal medium (Figure 51c).

Therefore, batch culture experiments were repeated with vitamin B₁₂ supplementation, this time the bloom of *R. gnavus* was observed with 2,7-anhydro-Neu5Ac or 3'SL supplementation compared to Neu5Ac and control vessels as shown by qPCR (Figure 51d). This agrees with the original data and supports the hypothesis that vitamin B₁₂ is a requirement for *R. gnavus* growth.

The difference in the growth of *R. gnavus* in the two batch model experiments presented in Figure 50b and c may be explained by either a change in the donor's diet, leading to a reduced level of vitamin B₁₂ or in a change in the donor's microbiome, leading to a reduction in the level of vitamin B₁₂ producing microbes. This will be the subject of further investigation using the batch model system.

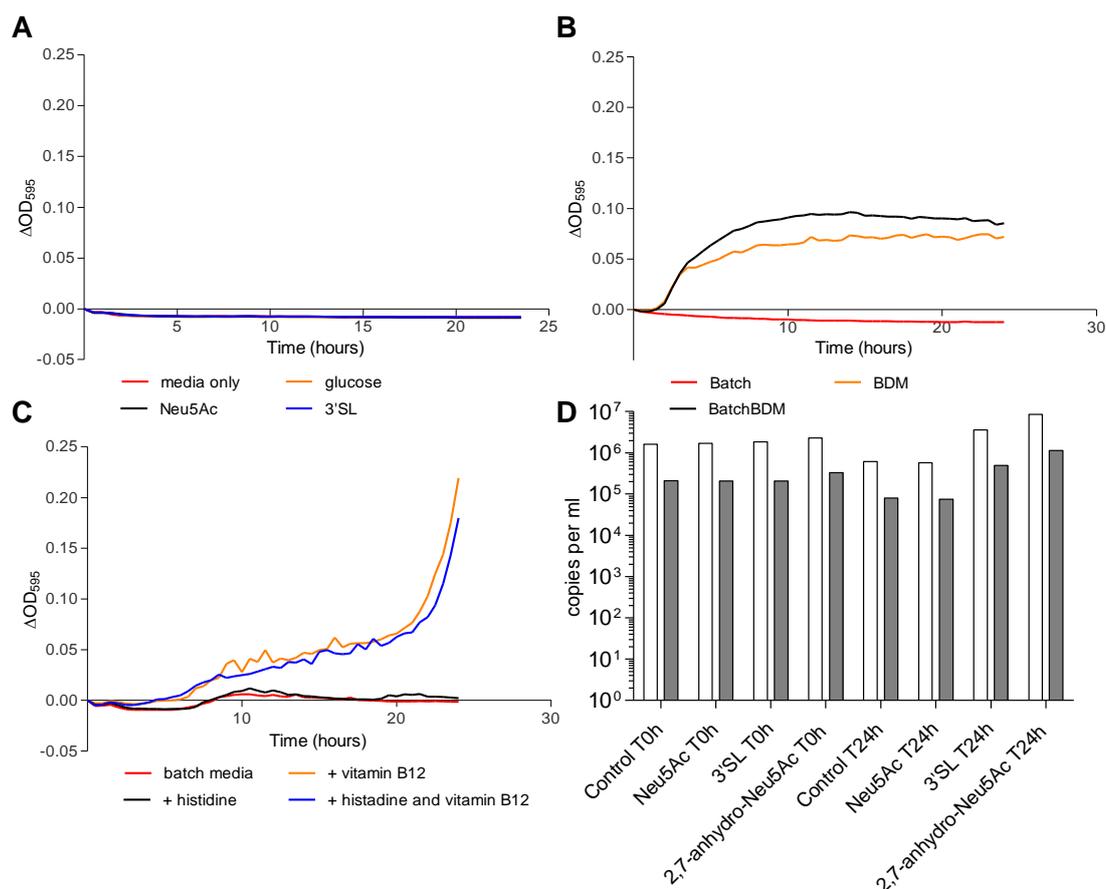


Figure 51: Analysis of *R. gnavus* growth with vitamin B₁₂ and/or histidine supplementation

A) Change in absorbance at OD₅₉₅ over time of microtiter cultures of *R. gnavus* ATCC 29149 grown in batch media using glucose, Neu5Ac or 3'SL as carbon sources. **B)** Change in absorbance at OD₅₉₅ over time of microtiter cultures of *R. gnavus* ATCC 29149 grown in batch medium, BDM or a 50/50 mix of both (BatchBDM). **C)** Change in absorbance at OD₅₉₅ over time of microtiter cultures of *R. gnavus* ATCC 29149 grown in batch media using glucose as carbon sources supplemented with histidine, vitamin B₁₂ or both. **D)** qPCR from colon model experiments using batch media supplemented with vitamin B₁₂, *R. gnavus* (white bars) and *RgSBP* (grey bars) copies per ml of samples taken from batch culture fermentations, using Neu5Ac, 3'SL, 2,7-anhydro-Neu5Ac or no carbon source (control) after 0 and 24 hours are shown.

The impact of 2,7-anhydro-Neu5Ac on the faecal microbial community was determined by 16S sequencing (Figure 52). Samples from the two experiments showing an increase of *R. gnavus* when batch cultures were supplemented with 2,7-anhydro-Neu5Ac or 3'SL were sequenced. The data supports the qPCR data with an increased relative frequency of *R. gnavus* in the 2,7-anhydro-Neu5Ac and 3'SL supplemented vessels after 24 hours. Interestingly, only five other operational taxonomic units (OTUs) showed a similar increase using 2,7-anhydro-Neu5Ac as carbon source, bacteria from the genus *Roseburia* were abundant (~20%) in the donor's microbiome but were not detected after 24 hours of fermentation except for in the 2,7-anhydro-Neu5Ac supplemented vessels.

The other OTUs that showed an increased abundance in the presence of 2,7-anhydro-Neu5Ac and 3'SL were the genera *Sutterella*, *Blautia*, *Phascolarctobacterium*, *Oscillospira* and *Butyricicoccus pullicaecorum*. These data give support to the hypothesis that 2,7-anhydro-Neu5Ac is limited to very few species, as seen by bioinformatics analysis (see section 4.2.4) and may provide these species with a competitive advantage in the mucosal niche.

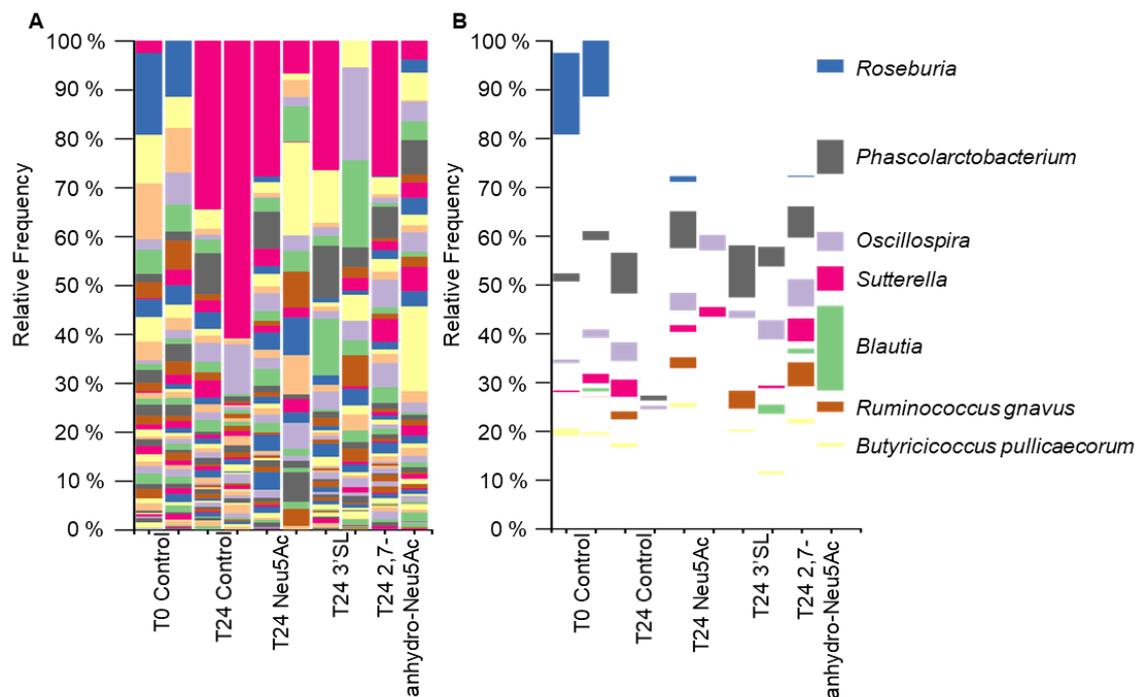


Figure 52: Taxonomic profiles of batch culture experiments supplemented with sialic acids

The results are shown from two independent experiments. **A)** shows the full profile, **B)** shows species identified to have an increased abundance in the 2,7-anhydro-Neu5Ac samples. Sequencing was performed on samples at the beginning of experiment (T0 control) and after 24 hours (T24). Four different experimental conditions were tested, no sialic acid supplementation (Control), Neu5Ac supplementation, 3'SL supplementation and 2,7-anhydro-Neu5Ac supplementation. Taxonomic classifications were made using the QIIME2 program and Greengenes database and are defined in the key presented in Appendix 3.

5.2.1 Importance of the *R. gnavus nan* cluster *in vitro* and *in vivo*

*5.2.1.1 Production of *R. gnavus* mutants using ClosTron technology*

The ClosTron method has been shown to be effective for generating mutants in *Clostridia* species (Heap *et al.*, 2010). The methodology and approach described in Figure 53a was applied to *R. gnavus* ATCC 29149. Briefly, genomic target sites for RgNanH were identified and the retargeted intron system, used to target these regions, was synthesised and ligated into the pMTL007C-E2 vector (Figure 53b). The recombinant plasmid was transformed into *E. coli* CA434 and conjugation performed with *R. gnavus*

ATCC 29149. Selection for successful chromosome recombination was carried out by plating onto plates containing cycloserine, to select against *E. coli* and erythromycin to select for successfully transformed *R. gnavus* (Figure 53a). The recombination event introduces an erythromycin resistance cassette into the gene of interest, both disrupting function of the target gene and providing a selection marker. Two insertion mutants were produced in the sense or antisense direction for *RgNanH*.

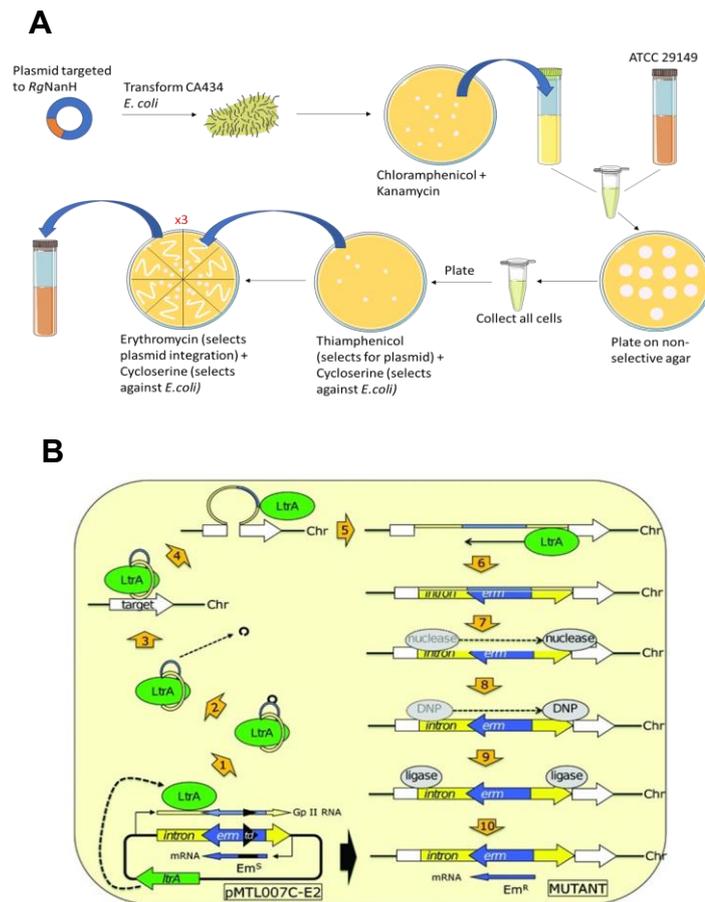


Figure 53: Outline of ClosTron mutagenesis strategy

A) Schematic representation of *R. gnavus* mutant generation using ClosTron. The targeted plasmid was transformed into *E. coli* CA434. The recombinant *E. coli* strain was then conjugated with *R. gnavus* ATCC 29149. *R. gnavus* colonies positive for plasmid transfer were then selected, before chromosome recombination (see **B**) and further selection. **B)** Schematic representation of chromosome recombination using ClosTron. The ClosTron plasmid encompasses an intron (yellow) into which the *ermB* gene (blue) is inserted, which is inactivated by the insertion of a small region of DNA (black). The intron can mediate its own splicing from RNA transcripts, but the process is orientation specific. Transcription of the *ermB* gene results in an mRNA transcript containing the intron insertion, the LtrA protein (green) also encoded by the plasmid binds to the transcript, splicing out the small piece of DNA. The transcript binds (by virtue of the re-targeted sequence) to specific sequences in the target gene within the chromosome (Chr). LtrA “nicks” the DNA target and inserts the RNA. The reverse transcriptase activity of LtrA now synthesises the complementary DNA strand. Host nucleases degrade the insert RNA and DNA polymerase synthesises the opposite DNA strand. Host ligases seal the two gaps, leading to completion of the process, in which a functional *ermB* gene is now present in the target gene (Kuehne and Minton, 2012).

Confirmation that the recombination event took place within the *RgNanH* encoding gene was obtained by PCR using primers spanning the expected insertion site and observing a band corresponding to the amplification product at the expected size (2.6 kb opposed to 450 bp for wild-type) (Figure 54). In addition, DNA sequencing confirmed the correct incorporation of the erythromycin resistance gene in the *R. gnavus* genome.

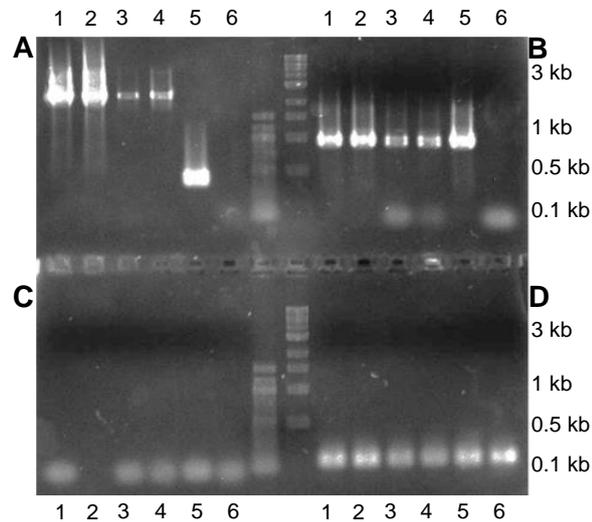


Figure 54: PCR analysis of *R. gnavus* mutants

In each panel 1 and 2 are clones from the sense direction mutations, 3 and 4 are clones from the antisense direction mutagenesis, 5 is gDNA from *R. gnavus* ATCC 29149 and 6 is a no template control. The primers used in each panel were **A**) primers spanning the expected insertion site in *RgNanH* (expected amplicon size is 450 bp for wildtype and 2.6 kb for mutants), **B**) primers from a separate region of the *nan* operon (RUMGNA_02700 – 02699), **C**) universal 16S primers, and **D**) *R. gnavus* specific 16S primers. Panels are separated by both 100 bp 1 kb DNA markers (New England Biolabs) were used, bright bands are 500 bp and 1 kb.

5.2.1.2 The *nan* cluster is essential for *R. gnavus* to utilise sialoglycans or 2,7-anhydro-Neu5Ac *in vitro*

To assess the effect of the *R. gnavus nan* cluster on the ability of *R. gnavus* ATCC 29149 to utilise sialic acid and sialoconjugates, *R. gnavus* ATCC 29149 wild-type, and both sense and antisense mutant strains were grown anaerobically with glucose, Neu5Ac, 3'SL or 2,7-anhydro-Neu5Ac as a sole carbon source. The *R. gnavus* wild-type strain was able to utilise glucose, 3'SL and 2,7-anhydro-Neu5Ac as a sole carbon source, but no growth was detected using the *R. gnavus* mutants on 3'SL or 2,7-anhydro-Neu5Ac whereas mutants grew to levels comparable to the wild-type strain on glucose (Figure 55). The lack of growth on 3'SL is expected as without the IT-sialidase (*RgNanH*), *R. gnavus* would not be able to liberate the 2,7-anhydro-Neu5Ac product from sialylconjugates. The absence of growth on 2,7-anhydro-Neu5Ac for the two mutants is

more surprising as *RgNanH* is not expected to be required for the metabolism of 2,7-anhydro-Neu5Ac (see section 4.3), only its release from sialylated substrates.

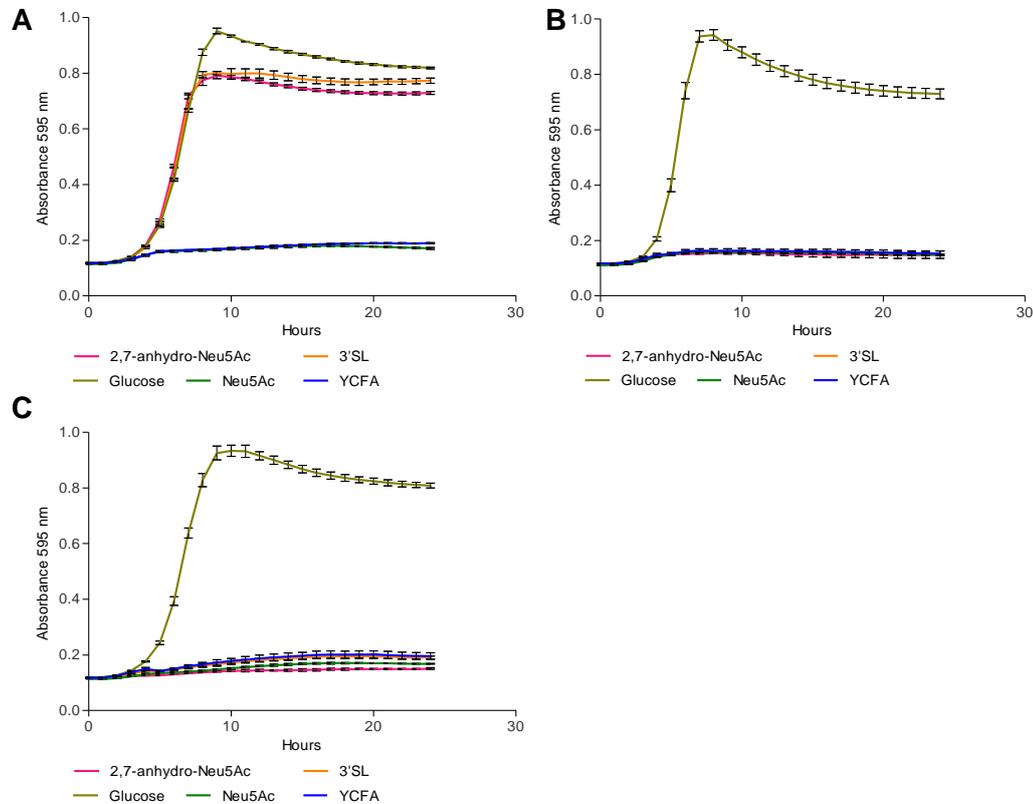


Figure 55: In vitro growth assay of *R. gnavus nan* mutants on sialic acid and sialylated substrates. Growth curves of **A)** *R. gnavus* ATCC 29149 (wild-type), **B)** *R. gnavus* antisense mutant **C)** *R. gnavus* sense mutant using the following sugars as sole carbon sources: media only (YCFA), 2,7-anhydro-Neu5Ac, 3'SL, Neu5Ac, glucose. Centre point is the mean of three replicates with standard error of the mean shown by error bars.

To investigate this further the expression of the full *nan* cluster was tested by qPCR using the same primer sets as for analysis of the cluster expression in wild-type *R. gnavus* on different carbon sources (see section 4.2.3). The expression of the whole operon in both sense and antisense mutant strains was compared to that of the wild-type strain using $2^{-\Delta\Delta Ct}$ calculations with *gyrB* as a housekeeping gene (Figure 56). The expression of the operon in both mutants was found to be negligible from RUMGNA_02698 through to RUMGNA_02690, while some residual expression was detected in the first transcribed genes of the cluster (RUMGNA_02701 through RUMGNA_02699). These data indicate that the *RgNanH* (RUMGNA_02694) insertion affected the expression of other genes in the operon, explaining the mutant phenotype, with regards to 2,7-anhydro-Neu5Ac growth. Taken together, these data demonstrate that the mutation of the *RgNanH* gene disrupts expression of the full operon, making these *nan* cluster mutants.

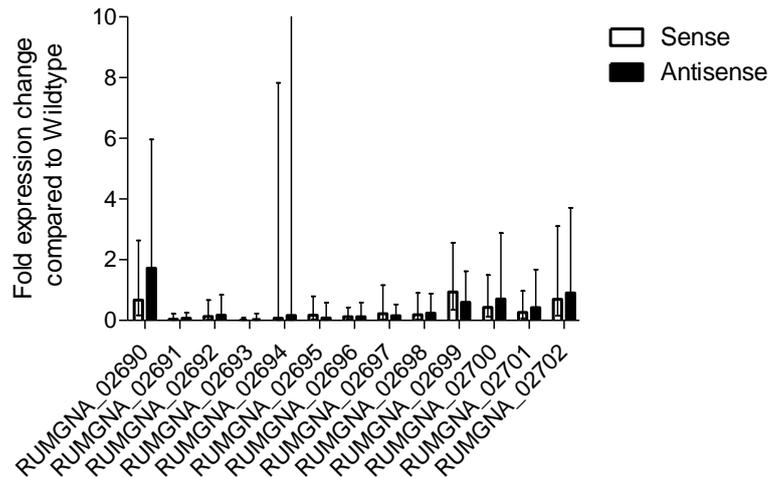


Figure 56: Gene expression analysis of wild-type and mutant *nan* operon

Fold change in expression was determined by qPCR of *R. gnavus nan* mutants compared to *R. gnavus* wild-type using $2^{-\Delta\Delta Ct}$ calculation. Black boxes represent expression changes for the antisense mutant and white boxes represent expression changes for the sense mutant.

5.2.2 In vivo colonisation of germ-free mice by *R. gnavus* wild-type and *nan* mutants

5.2.2.1 The nan cluster is essential for the colonisation and spatial location of R. gnavus in the gut

To assess the impact of the *nan* cluster on the fitness of *R. gnavus in vivo*, germ-free C57BL/6J mice were gavaged with 10^8 CFU *R. gnavus* ATCC 29149 or *R. gnavus* antisense *nan* mutant or a mixture of wild-type and *nan* mutant strains at 10^8 CFU each (Figure 57). During mono-colonisation experiments, both wild-type and mutant strains were detectable in the faecal content at day 3, 7 and 14 post-gavage at mean levels of between 10^6 and 10^7 bacteria per mg of faeces (Figure 57a). Both strains were also detected in the caecal content of mono-colonised mice sacrificed at day 14. In competition experiments, primers based on the insertion in the *RgNanH* gene were used to distinguish between wild-type and *nan* mutant *R. gnavus*. The wild-type strain reached mean colonisation levels comparable to the levels obtained during mono-colonisation, whereas the mutant strain was severely outcompeted, reaching only 10^4 copies per mg at day 3, before decreasing further at day 7 and day 14 to levels below the level of detection, in both the faecal and caecal contents (Figure 57b).

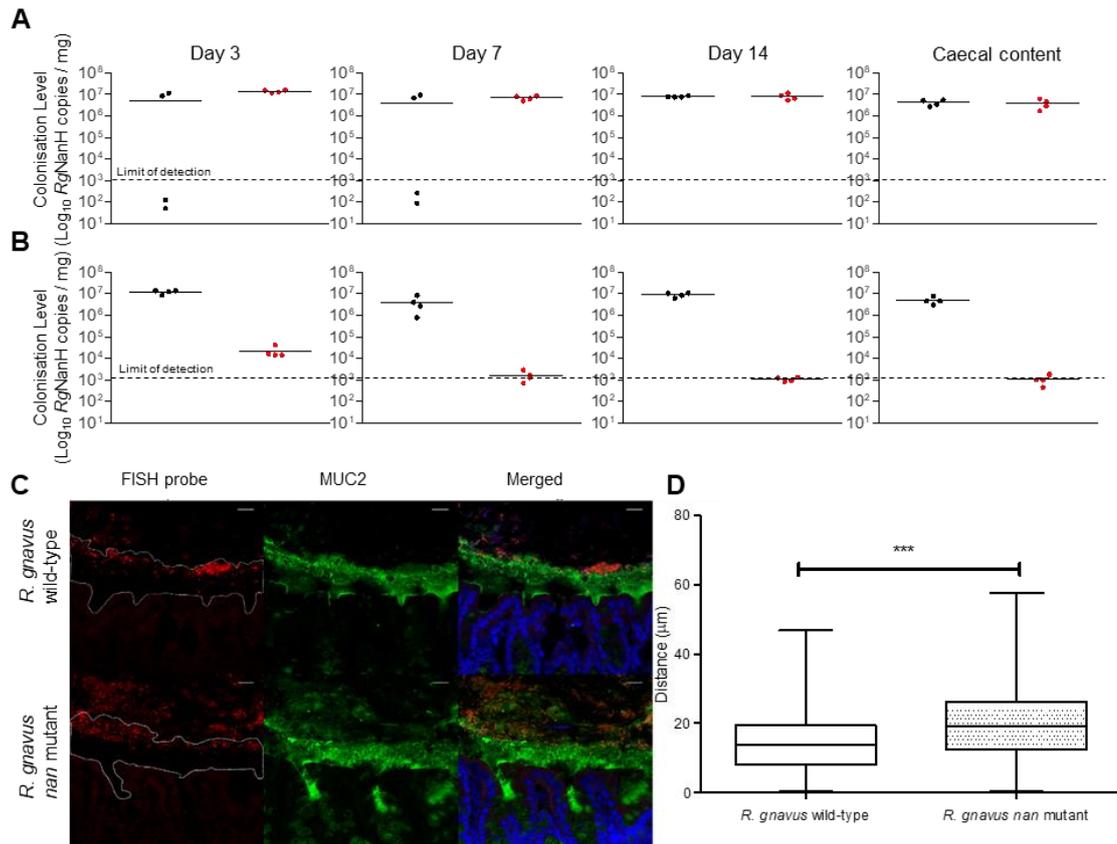


Figure 57: Colonisation of germ-free C57BL/6J mice with *R. gnavus* ATCC 29149 wild-type or *nan* mutant strains

Mice were monocolonised with **A**) *R. gnavus* wild-type (black) or *nan* mutant (red) strains individually or **B**) in competition. Mice were orally gavaged with 10^8 of each strain, faecal samples were analysed at 3,7 and 14 days after inoculation and caecal samples at 14 days after inoculation using qPCR. **C**) Fluorescent in situ hybridisation (FISH) and immunostaining of the colon from *R. gnavus* monocolonised C57BL/6 mice. *R. gnavus* ATCC 29149 and *R. gnavus nan* mutant are shown in red. The mucus layer is shown in green and an outline of the mucus is shown in the first panels. Cell nuclei were counterstained with Sytox blue, shown in blue. Scale bar: 20 µm. **D**) Quantification of the distance between the leading front of bacteria and the base of the mucus layer. A total of 70 images of stained colon from 8 *R. gnavus* monocolonised mice were analysed. The asterisks (***) show the significant significance ($P=0.0135$, by linear mixed model analysis), centre point indicates the mean, box limits, upper and lower quartiles; whiskers, minimum and maximum.

The impact of the *nan* deletion on the ability of *R. gnavus* to penetrate the mucus layer was determined in mono-colonised mice by measuring the distance of the *nan* mutant or wild-type *R. gnavus* strains to the epithelial layer throughout the colon by fluorescent *in situ* hybridization (FISH) staining using confocal microscopy (Figure 57c). The data produced and analysed by Laura Vaux, showed that the *nan* mutant resided 19.70 µm from the epithelial layer, 5.06 µm further away than the wild-type strain, 14.64 µm (Figure 57d).

5.2.2.2 The presence of the *nan* cluster leads to large shifts in colonic metabolites during mono-colonisation

To further investigate whether the difference in colonisation was due to a change in sialic acid metabolism, transcriptomic analyses were carried out on the caecal contents and colonic tissue from mice mono-colonised with *R. gnavus* or antisense *nan* mutant. The mouse transcriptomics data showed that the colonisation with the *nan* mutant did not affect the mouse expression response in comparison to colonisation with wild-type *R. gnavus*. Only one gene was found to be differentially expressed between wild-type and *nan* mutant colonised mice corresponding to a down regulation of an immunoglobulin heavy variable gene in the *nan* mutant colonised mice (Figure 58a). On the bacterial side, a total of 27 genes were found to be differentially expressed between wild-type and *nan* mutant *in vivo* (Figure 58b). All 27 genes were lower expressed in the *nan* mutant, these include 9 of the 11 genes in the *nan* cluster shown to be down regulated *in vitro* (Figure 56). The other genes have limited information on their putative function but appear to fall into three gene clusters associated with phage and capsid proteins (Table 7).

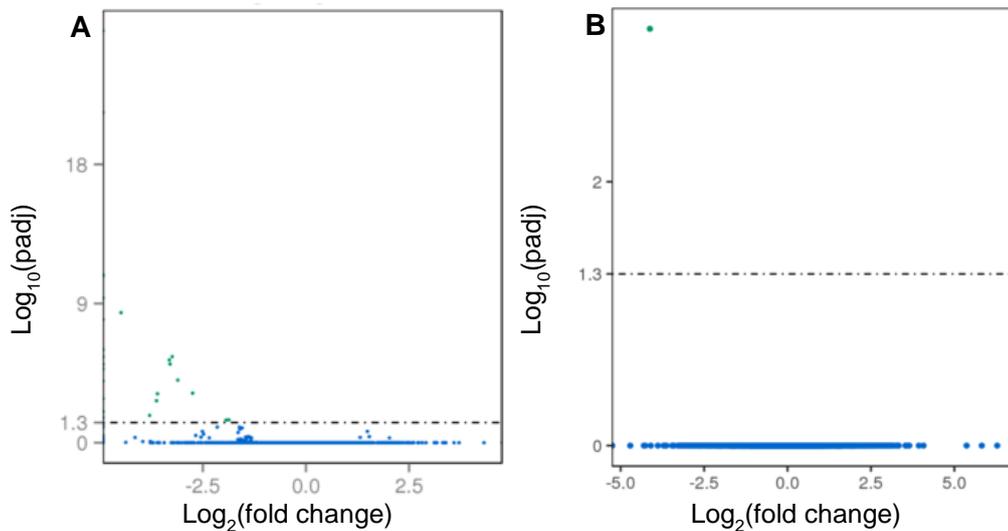


Figure 58: Bacterial and host transcriptomic analysis of mice colonised by *R. gnavus* or *nan* mutant

Volcano plots representing the differentially expressed genes in colonic mouse tissue **A**) or bacteria **B**) when mice were mono-colonised with *R. gnavus* ATCC 29149 or *nan* mutant. Genes are differentially expressed when $\text{Log}_2(\text{fold change})$ is greater than 1 or less than -1 and $\text{Log}_{10}(\text{padj})$ (significance score) is above 1.3. Differentially expressed genes that are down regulated in the *nan* mutant compared to the wild-type strain are shown as green dots, no genes for upregulated in the *nan* mutant for mouse or bacterial transcriptomics. Non-differentially expressed genes are shown as blue dots.

Table 7: Identification of *R. gnavus* genes differentially expressed in nan mutant during colonisation, bold denotes genes in the *R. gnavus nan* cluster.

Gene_id	description
C5Y99_08390	RgNanK
C5Y99_08395	RgNanE
C5Y99_08400	RgNanA
C5Y99_08405	RgNanH
C5Y99_08410	RgNanOx
C5Y99_08415	carbohydrate ABC transporter permease
C5Y99_08420	sugar ABC transporter permease
C5Y99_08425	RgSBP
C5Y99_08440	hypothetical protein
C5Y99_13530	hypothetical protein
C5Y99_16285	hypothetical protein
C5Y99_16290	DUF4236 domain-containing protein
C5Y99_16300	XRE family transcriptional regulator
C5Y99_16305	hypothetical protein
C5Y99_16310	hypothetical protein
C5Y99_16315	3-hydroxyacyl-CoA dehydrogenase
C5Y99_16320	phage antirepressor Ant
C5Y99_16435	PBSX family phage terminase large subunit
C5Y99_16440	phage capsid protein
C5Y99_16445	minor capsid protein
C5Y99_16510	tape measure domain-containing protein
C5Y99_16520	hypothetical protein
C5Y99_16540	XRE family transcriptional regulator
C5Y99_16545	hypothetical protein

C5Y99_16550	hypothetical protein
Novel00028	
sRNA00096	

Interestingly no bacterial genes were found to be upregulated in the mice colonised with the *R. gnavus nan* mutant. *R. gnavus* ATCC 29249 has the ability to grow on other mucin glycans *in vitro* such as fucose and fucosylated substrates (Croft *et al.*, 2013). Therefore, it was hypothesised that *R. gnavus* may induce the expression of these mucin degrading enzymes to compensate for the loss of expression of the *nan* operon. To test that the *nan* mutant was still able to utilise fucose, growth curves were carried out for the *R. gnavus* wild-type and *nan* mutant using fucose and fucosylated substrates as sole carbon sources (Figure 59). The data showed no difference in the ability of wild-type or *nan* mutant to utilise fucose, with both able to grow on fucose, 2' fucosyllactose (2'FL) and 3' fucosyllactose (3'FL), suggesting that the *nan* mutants could also utilise fucose from mucin glycans *in vivo*, in agreement with the level of colonisation observed in mono-colonised mouse experiments (Figure 57).

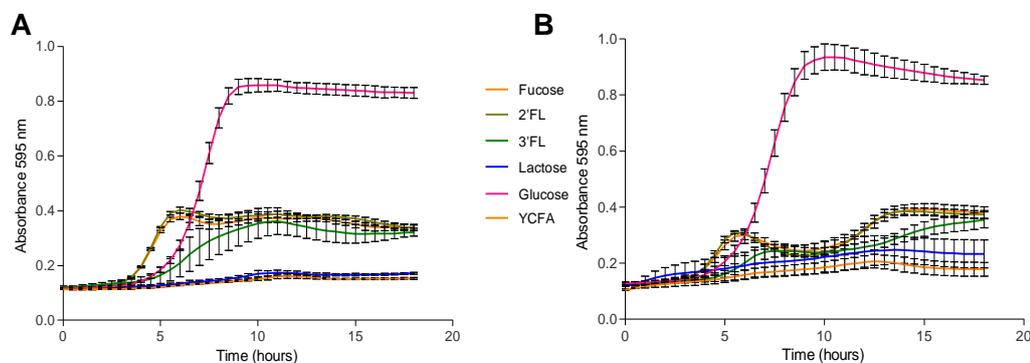


Figure 59: In vitro growth assay of *R. gnavus nan* mutants on fucose and fucosylated substrates. Growth curves of **A)** *R. gnavus* ATCC 29149 (wild-type), **B)** *R. gnavus* antisense mutant using the following sugars as sole carbon sources: media only (YCFA), fucose, 2' fucosyllactose (2'FL), 3' fucosyllactose (3'FL), lactose or glucose. Centre point is the mean of three replicates with standard error of the mean shown by error bars.

In addition to the transcriptomics analyses, caecal content metabolite profiles from mice mono-colonised with *R. gnavus* wild-type or the *nan* mutant were analysed by Metabolon (USA). In contrast to the low number of differentially expressed genes, a total of 119 metabolites were found at significantly different concentrations between the two groups,

106 more abundant and 13 less abundant in the *nan* mutant colonised mice (Figure 60). A further 94 chemicals narrowly missed the significance cut off. These changes in chemical composition were associated with specific pathways, those associated with amino acid metabolism, gamma glutamyl, phosphatidylethanolamine and plasmogen were more abundant in the *nan* mutant colonised mice and those involved in fatty acid metabolism higher in the wild-type *R. gnavus* colonised mice. Also, of note is that sialic acid was found at higher levels in the mice colonised with wild-type as compared to mice colonised with the *nan* mutant, consistent with the release of 2,7-anhydro-Neu5Ac and Neu5Ac by *RgNanH*.

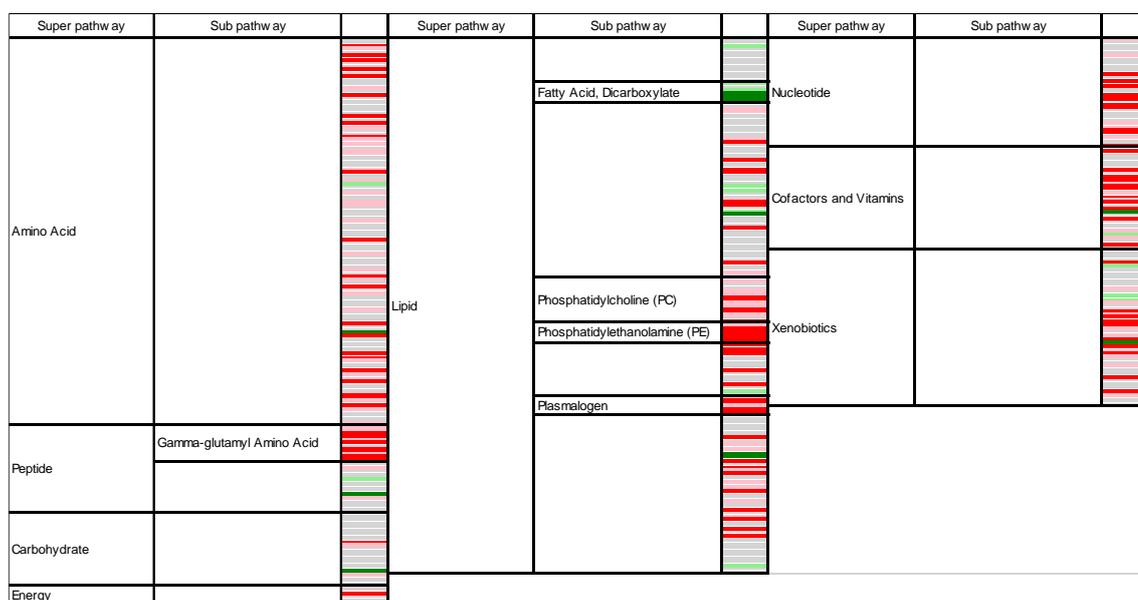


Figure 60: Metabolite changes in germ-free mice colonised with *R. gnavus* wild-type or *nan* mutant

Caecal content from germ-free mice colonised with *R. gnavus* ATCC 29149 or *nan* mutant were analysed for their metabolic profile by Metabolon. Individual metabolites are grouped as part of the sub and super pathways they are part of, and each line represents 1 metabolite. Red indicates those metabolites that are statistically ($p < 0.05$) more abundant in the *nan* mutant group and green more abundant in wild-type colonised mice. Pale red and pale green indicate metabolites trending towards significance ($0.05 < p < 0.10$).

5.3 Discussion

In vitro colon model experiments demonstrated that 2,7-anhydro-Neu5Ac is able to promote the growth of *R. gnavus* as shown using both qPCR assays and 16S sequencing. Furthermore, 16S sequencing identified the *Roseburia*, *Sutterella*, *Blautia*, *Phascolarctobacterium*, *Oscillospira* and *Butyricicoccus* genera as being able to benefit from 2,7-anhydro-Neu5Ac supplementation, though only *Blautia* was identified by the bioinformatics analysis of the *R. gnavus nan* cluster. Since the ability to utilise 2,7-anhydro-Neu5Ac appears species- or strain-specific, as is the case with *R. gnavus*, it is possible that the specific strains do not have a deposited sequenced genome available. *Oscillospira* is a central component of the gut microbiota that has been associated with obesity that is routinely detected using 16S sequencing methods (Gophna *et al.*, 2017, Konikoff and Gophna, 2016). However, it has yet to be cultured and so there is no genomic sequence available, it therefore remains unclear whether members of this genus encode an extended *nan* cluster or not. Members of the genera *Roseburia* and *Butyricicoccus* have been associated with IBD, and *Sutterella* with autism and IBD (Hiippala *et al.*, 2016, Machiels *et al.*, 2014, Wang *et al.*, 2013, Eeckhaut *et al.*, 2013), as has *R. gnavus* (see section 1.3.3). To confirm the 16S sequencing results qPCR analysis can be performed using primers to target each genera of bacteria shown to increase in abundance, as we have done for *R. gnavus*. It would also be beneficial to include more repeat experiments and try 2,7-anhydro-Neu5Ac supplementation with different donors, to identify which other species may benefit from 2,7-anhydro-Neu5Ac supplementation.

The 16S sequencing gives good information about bacteria that can benefit from 2,7-anhydro-Neu5Ac, though it is not possible to determine whether these increases in abundance are a result of primary metabolism of 2,7-anhydro-Neu5Ac or a result of secondary metabolism and cross-feeding. Two approaches could be used to test this, firstly strains of bacteria showing a bloom by 16S sequencing analysis could be grown in mono-culture using 2,7-anhydro-Neu5Ac as a sole carbon source. Alternatively, we could use a stable isotope probing (SIP) approach using ¹³C-labelled 2,7-anhydro-Neu5Ac. This has been performed for Neu5Ac using a ¹³C labelled compound (Young *et al.*, 2015). Recent SIP approaches to assess if a cell or bacterium is metabolically active have used heavy water (D₂O) that can be used to distinguish active from not active cells using Raman microspectroscopy. The resulting fingerprint can detect shifts caused by ²H incorporation by an active cell and as the measurement is non-destructive, cells can be processed and sorted post-measurement (Berry and Loy, 2018). This method may

be preferred when considering 2,7-anhydro-Neu5Ac metabolism due to the difficulties and cost of producing a ^{13}C labelled 2,7-anhydro-Neu5Ac (chapter 3).

As discussed in section 1.3.3, *R. gnavus* has been positively or negatively associated with several disease states, while *R. gnavus* intervention has been shown to have a positive outcome in pre-clinical models of malnutrition and inflammation (Grabinger *et al.*, 2019, Blanton *et al.*, 2016). As *R. gnavus* is a prevalent member of gut microbiota (Qin *et al.*, 2010), a prebiotic that can selectively boost the growth of *R. gnavus* strains has potential medical value in specific conditions. Here, we have shown *in vitro* that 2,7-anhydro-Neu5Ac can selectively promote the growth of *R. gnavus* in a simulated model colon. Future work will determine whether this can also be applied *in vivo*.

The *in vitro* colon model experiments also revealed that vitamin B₁₂ is an essential nutrient for *R. gnavus* which was previously unknown. The inter-individual response to 2,7-anhydro-Neu5Ac in the colon model could be due to changes in the donor's vitamin B₁₂ intake either through supplements or diet, or a change in the abundance of vitamin B₁₂ producing bacteria in the donor's microbiome. This will be investigated further by determining the levels of vitamin B₁₂ from these experiments. Recently a bioinformatics analysis has been conducted to assess the ability of bacteria to synthesise vitamin B₁₂ (Shelton *et al.*, 2019). In this analysis *R. gnavus* was classified as a 'tetrapyrrole precursor salvager', this implies that *R. gnavus* encodes the pathway for vitamin B₁₂ synthesis, but lacks the genes required to synthesise the tetrapyrrole precursors and so must salvage these from the environment to produce vitamin B₁₂. Future experiments will be performed supplementing *R. gnavus* mono-cultures with these precursors, for example porphobilinogen, to validate these findings *in vitro*.

The ClosTron methodology (Heap *et al.*, 2010) has provided a tool to successfully generate *R. gnavus* mutants, the first reported method to genetically manipulate *R. gnavus* strains. However, the use of resistance markers resulted in the down regulation of the complete operon. Future efforts will explore the use of alternative insertions such as stop codons or short nonsense sequence to target individual genes of interest as opposed to full clusters. A variety of different ClosTron vectors and methodology for generating plasmids with a variety of characteristics has been described (Heap *et al.*, 2009). Here the *nan* operon mutant was used to investigate the impact of the *nan* cluster on *in vivo* fitness of *R. gnavus*. In competition with *R. gnavus* wild-type strain, the mutant strain was heavily outcompeted and after 14 days could no longer be detected. In mono-colonised mice, the *R. gnavus nan* mutant also resided further away from the epithelium on the edge of the mucus layer, suggesting that the *nan* operon is important for mucus

penetration. This ability of colonising the mucosal niche may also contribute to the success of *R. gnavus* as one of the most frequent human colonisers (Qin *et al.*, 2010).

The RNAseq analysis showed that the introduction of a *nan* mutation in *R. gnavus* led to the down-regulation of the full *nan* operon but also affected the expression of an additional 18 genes in the mutant strain during *in vivo* colonisation. No genes were upregulated which was unexpected as losing access to an important nutrient source would likely need to be compensated for. It would be interesting to carry out this transcriptomic analysis *in vitro*, to distinguish whether these changes are a consequence of the *nan* mutation or associated with the *in vivo* colonisation.

The transcriptome analysis for colonic tissue of mice colonised with *R. gnavus* wild-type compared to the *R. gnavus nan* mutant revealed only one gene with an altered expression. This was a downregulation of an immunoglobulin heavy variable gene in the mutant colonised mice, and likely reflects a targeted response to a protein not present in the *nan* mutant although the rationale for this remains unclear. In contrast to the low level of transcriptomic changes in the mouse colon or bacteria, 119 chemicals showed an altered concentration in metabolite analysis, most of which were more abundant in the *nan* mutant colonised mice, highlighting the complexity of host-microbe interactions and further supporting the importance of monitoring changes in metabolites, to unravel these interactions and identify therapeutic strategies (Moya and Ferrer, 2016).

In summary, data presented in this chapter showed the potential of 2,7-anhydro-Neu5Ac as a next generation prebiotic for *R. gnavus* which could be used in conditions where increasing *R. gnavus* proportion in the gut is advantageous. The development of a method to allow genetic manipulation opens new possibilities for investigating important biological functions of *R. gnavus* and associations with disease. Here, we did not investigate how bacteria with the *nan* operon may shape the rest of the microbiome. Future work will aim to tackle this question both *in vitro* and *in vivo*. To establish *R. gnavus* in conventional mice an antibiotic treatment is first required to open up the niche, this could introduce some bias in the interpretation of data from an *in vivo* study. Alternatively, the colon model can be used with a donor that does not contain *R. gnavus* which can be seeded with either *R. gnavus* wild-type or the *nan* mutant to determine changes in the microbial community. Further, *in vitro* colon model experiments using ¹³C-labelled 2,7-anhydro-Neu5Ac will provide a more definite answer on the bacteria able to metabolise this compound in the gut microbiota. This information is important to assess the potential therapeutic ability of *R. gnavus* strains or 2,7-anhydro-sialic acid derivatives.

Chapter 6

Impact of the *R. gnavus nan* cluster on enteric pathogens

6.1 Introduction

Sialic acid is an important molecule used by pathogenic organisms in a number of ways. Pathogenic bacteria can use sialic acid expressed on the surface of vertebrate cells or secreted glycans as a target site for adherence, for toxin release or to mask themselves (Severi *et al.*, 2007). Viruses can take up free sialic acid and express it on the cell surface to evade the immune system (Wasik *et al.*, 2016). In bacterial pathogens, sialic acid can be found in cell surface components including LPS, capsular polysaccharides or flagellum (Haines-Menges *et al.*, 2015). In addition, bacteria can utilise sialic acid as a nutrient source. Catabolism of sialic acid has been shown to confer competitive advantages in enteric pathogens such as *S. Typhimurium*, *C. difficile* or *V. cholera*. This can occur via the expression of sialidases releasing the sialic acid followed by its transport and catabolism inside the cell as is the case for *V. cholera* (Almagro-Moreno and Boyd, 2009), but also in gut commensal bacteria such as *B. fragilis*.

Alternatively, some pathogens do not encode a sialidase and scavenge free Neu5Ac released from mucin glycans by members of the gut microbiota as is the case for *S. Typhimurium* and *C. difficile* (Ng *et al.*, 2013). For example, *B. thetaiotaomicron*, a gut commensal bacterium, expresses a sialidase but lacks the genes required to utilise sialic acid, resulting in the release of free Neu5Ac in the environment which becomes available for pathogens to scavenge. Mouse studies revealed that post-antibiotic treatment, there is an increase in the level of free Neu5Ac which is associated with the outgrowth of pathogens. In the previous chapters, *R. gnavus* was shown to cleave sialic acid in the form of 2,7-anhydro-Neu5Ac. Here we will test the ability of pathogens to utilise 2,7-anhydro-Neu5Ac *in vitro* and investigate the hypothesis that the IT-sialidase-expressing *R. gnavus* can reduce the pool of free Neu5Ac available to these pathogens, in turn reducing proliferation of these pathogens *in vivo*.

6.2 Results

6.2.1 Bioinformatics of *S. Typhimurium* and *C. difficile* sialic acid utilisation pathway

To better understand how *S. Typhimurium* and *C. difficile* utilise free sialic acid or sialoconjugates, a bioinformatics analysis was carried out to identify the presence of the *nanA/K/E* cluster and sialidases in these strains. A total of 146 *Salmonella* strains and 71 *C. difficile* strains were analysed by BLAST searching for homologues of the *R. gnavus* ATCC 29149 NanA, NanE and NanK encoded protein sequences against the genome sequences. All strains returned BLAST hits for NanA, NanE and NanK, suggesting that the ability to utilise sialic acid is conserved among *Salmonella* and *C. difficile* species. The IT-sialidase encoding sequence from *R. gnavus* strains is divided into the catalytic GH33 domain, the carbohydrate binding CBM40 domain and the inserted domain (Figure 61). The inserted domain is found within the IT-sialidases of *R. gnavus* but is not present in hydrolytic sialidases; however, its function has yet to be determined biochemically.



Figure 61: Domain structure of *R. gnavus* IT-sialidase

The annotation is based on the IT-sialidase of *R. gnavus* ATCC 29149 green – Signal peptide, orange – CBM40 (carbohydrate binding), blue – catalytic domain (GH33) and red IT domain (domain inserted into catalytic domain).

A bioinformatics search was carried out for the presence of these three domains in *Salmonella* and *C. difficile* genomic sequences. Only two strains of *Salmonella* were found to be positive for the presence of the GH33 domain in their genomes; *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (*S. Typhimurium* LT2) and *Salmonella enterica* subsp. *diarizonae* serovar 60:r:e,n,x,z15 str. 01-0170 whereas no gene encoding a GH33 like domain was identified in *C. difficile* strains. These two sialidases were found to be 97.4 % identical to each other and neither protein produced BLAST hits for the 'IT' domain. Furthermore, whereas the gene encoding the *R. gnavus* IT-sialidase is located within the same cluster as the *nanA*, *nanE* and *nanK* genes, the sialidases in the two positive *Salmonella* strains are found in a different genomic location to the utilisation genes. Together these analyses suggest that all *C. difficile* strains and

the majority of *Salmonella* strains are not able to cleave sialic acid from sialylated substrates (due to the absence of a sialidase) but are likely to utilise sialic acid (due to the presence of the canonical *nan* cluster).

6.2.2 In vitro mono-cultures of enteric pathogens

In order to validate the bioinformatics findings and determine how *Salmonella enterica* and *C. difficile* utilise sialic acid derivatives and sialoconjugates, two strains *S. Typhimurium* LT2 (sialidase positive) and *S. Typhimurium* SL 1344 (no-sialidase) and one *C. difficile* strain (NTCT 12726) were grown using a range of carbon sources. For *S. Typhimurium*, these growth assays were carried out in both aerobic and anaerobic conditions as it is a facultative anaerobe while *C. difficile* was grown under strict anaerobic conditions. As expected from the bioinformatics analysis, *S. Typhimurium* SL1344 was unable to grow on lactose, 3'SL, 6'SL or purified porcine gastric mucin (pPGM), owing to the lack of a sialidase, but could grow to high density on Neu5Ac, Neu5Gc and glucose (Figure 62a). Conversely, *S. Typhimurium* LT2 strain was able to grow on 3'SL as well as glucose, Neu5Ac and Neu5Gc, demonstrating that it can liberate sialic acid. Interestingly it could not grow on 6'SL or pPGM, in agreement with previous work on the reported substrate specificity of the LT2 sialidase, showing strong preference for α 2-3 linked sialic acid over α 2-6; furthermore the sialidase has been shown to have a much reduced activity against complex branched structures such as mucin (Hoyer *et al.*, 1991, Minami *et al.*, 2013). *S. Typhimurium* LT2 was also unable to utilise lactose, indicating that the growth seen on 3'SL is due to the release and metabolism of Neu5Ac and not lactose. In addition, the results demonstrate that while there was no difference in the sugars that could be utilised by *S. Typhimurium* in aerobic vs. anaerobic conditions (Figure 62b), there was a difference in the rate of growth which is to be expected as aerobic respiration generates more adenosine triphosphate (ATP) (Pfeiffer *et al.*, 2001).

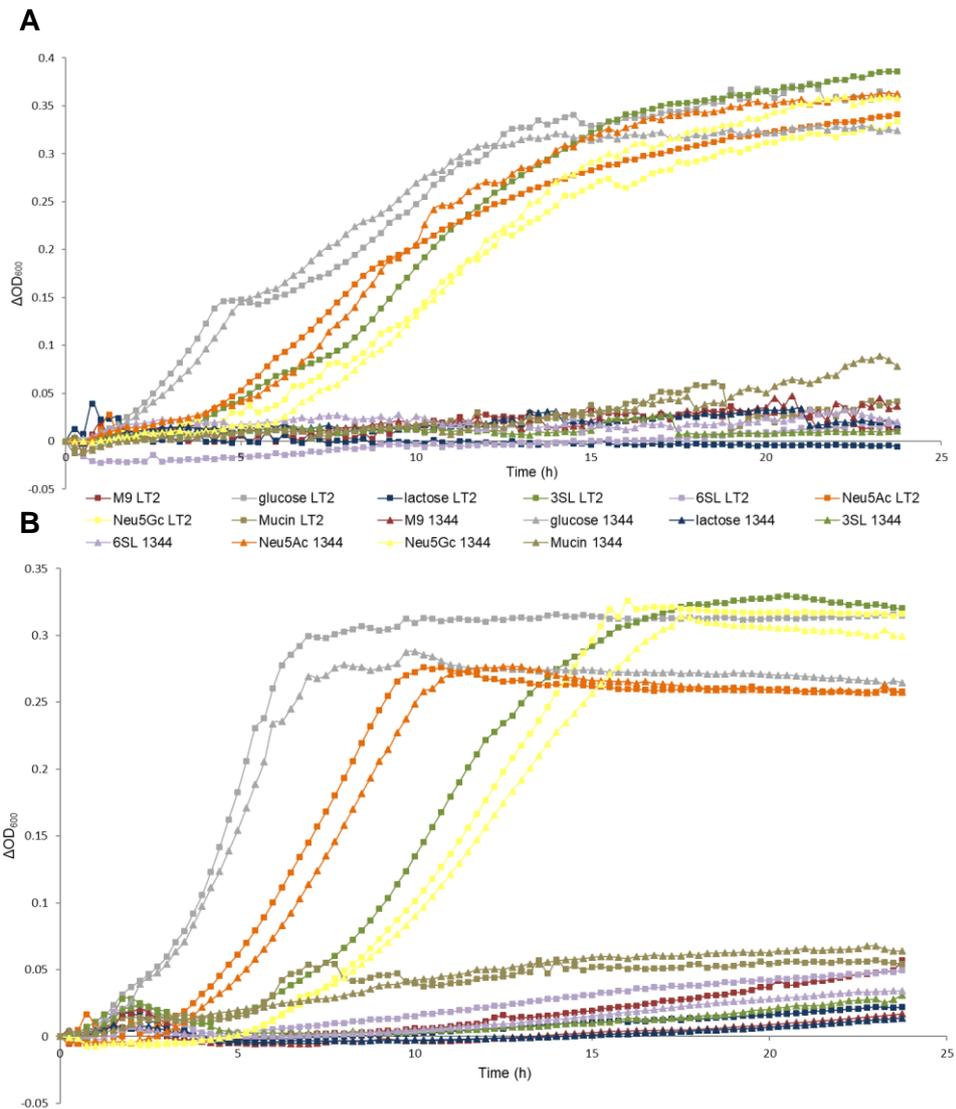


Figure 62: Growth curves of *S. Typhimurium* under aerobic and anaerobic conditions

Change in absorbance at OD600 over time of aerobic **A**) and anaerobic **B**) microtiter cultures of *S. Typhimurium* LT2 and SL 1344 using the following sugars as sole carbon sources: media only (M9), glucose, lactose, 3'SL, 6'SL, Neu5Ac, Neu5Gc, Mucin (pPGM).

We then tested the hypothesis that *S. Typhimurium* and *C. difficile* strains would not be able to grow on 2,7-anhydro-Neu5Ac due to differences in the *nan* operon compared to *R. gnavus* and the lack of an IT-sialidase. Growth curves of *S. Typhimurium* SL1344 and *C. difficile* NCTC 12726 on sialic acid derivatives confirmed that both could utilise Neu5Ac and Neu5Gc but neither could utilise 2,7-anhydro-Neu5Ac (Figure 63). Furthermore, it was confirmed that *C. difficile* could not utilise sialoconjugates due to the absence of a sialidase gene, whilst it could grow on glucose, Neu5Ac and Neu5Gc to high density under the same conditions.

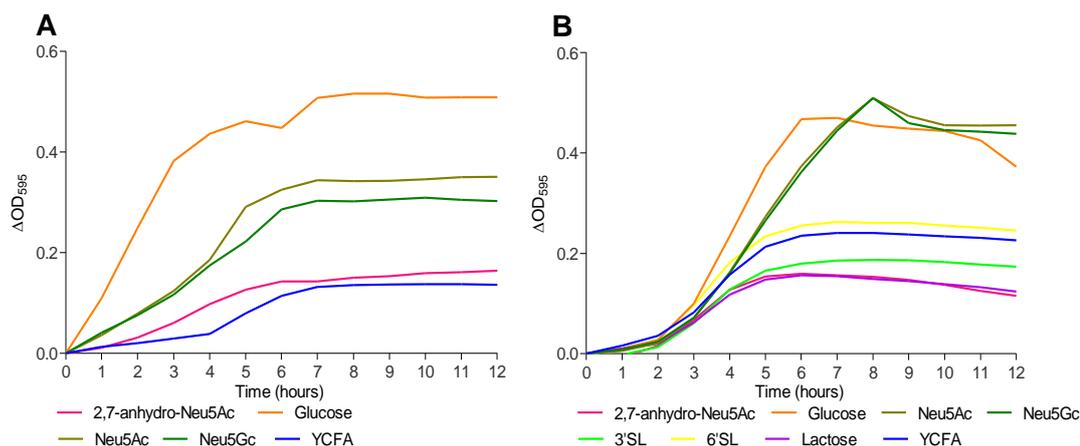


Figure 63: Growth of *S. Typhimurium* and *C. difficile* on sialic acids

Change in absorbance at OD₅₉₅ over time of microtiter cultures of **A)** *S. Typhimurium* SL1344 using 2,7-anhydro-Neu5Ac, glucose, Neu5Ac, Neu5Gc, or YCFA alone. **B)** *C. difficile* NCTC 12726 using 2,7-anhydro-Neu5Ac, glucose, Neu5Ac, Neu5Gc, 3'SL, 6'SL, lactose or YCFA alone.

6.2.3 In vitro co-cultures of *S. Typhimurium* with *R. gnavus* or *B. thetaiotaomicron*

It has been shown that the release of Neu5Ac in the gut by commensal species such as *B. thetaiotaomicron*, and the subsequent scavenging of the Neu5Ac by *S. Typhimurium* and *C. difficile* leads to increased pathogenicity post antibiotic treatment (Ng *et al.*, 2013). *R. gnavus* releases 2,7-anhydro-Neu5Ac from glycans, which could potentially reduce the level of free Neu5Ac available to the pathogens and thus limit the growth of *S. Typhimurium* and *C. difficile*. To assess this *in vitro* the growth of co-cultures of *R. gnavus* ATCC 29149 with *S. Typhimurium* SL1344, and *B. thetaiotaomicron* ATCC 29148 with *S. Typhimurium* SL1344 using 3'SL as the sole carbon source was tested. *B. thetaiotaomicron* has been shown to metabolise 3'SL *in vitro* by using lactose liberated by its hydrolytic sialidase activity, while it is not able to utilise the Neu5Ac liberated (Yu *et al.*, 2013). Based on the mono-culture results showing that *S. Typhimurium* can utilise Neu5Ac but not lactose nor 2,7-anhydro-Neu5Ac, it was hypothesised that *S. Typhimurium* would be able to cross feed on the Neu5Ac released by *B. thetaiotaomicron* sialidase, whilst it should not be able to cross-feed from the activity of the *R. gnavus* IT-sialidase which releases 2,7-anhydro-Neu5Ac and lactose.

6.2.4.1 Choice of minimal media for co-culture assays

In order to perform these co-culture experiments, a minimal medium is required to support the growth of *B. thetaiotaomicron* or *R. gnavus* with *S. Typhimurium* or *C. difficile*. Two minimal media were tested, YCFA which has previously been used for

studies with *R. gnavus*, and BDM which is routinely used for *B. thetaiotaomicron* (see Materials and Methods). Growth curves for each bacterium were conducted under anaerobic conditions using both media with and without glucose supplementation. The data showed that YCFA was capable of supporting the growth of all four bacteria in the presence of glucose, but also showed low levels of growth without glucose supplementation (Figure 64). This growth without glucose is low with *R. gnavus* and *B. thetaiotaomicron* but is substantial for both *S. Typhimurium* and *C. difficile*. This limits the use of YCFA as a minimal medium for these co-culture experiments, as low-level growth on tested carbon sources may be masked. For BDM, no growth of any of the four strains was observed in the absence of glucose. With glucose present, growth was supported for *R. gnavus*, *B. thetaiotaomicron* and *S. Typhimurium*, while *C. difficile* showed no growth. Therefore, BDM can be used for co-cultures of the two commensal strains with *S. Typhimurium* but is not suitable for co-cultures with *C. difficile*.

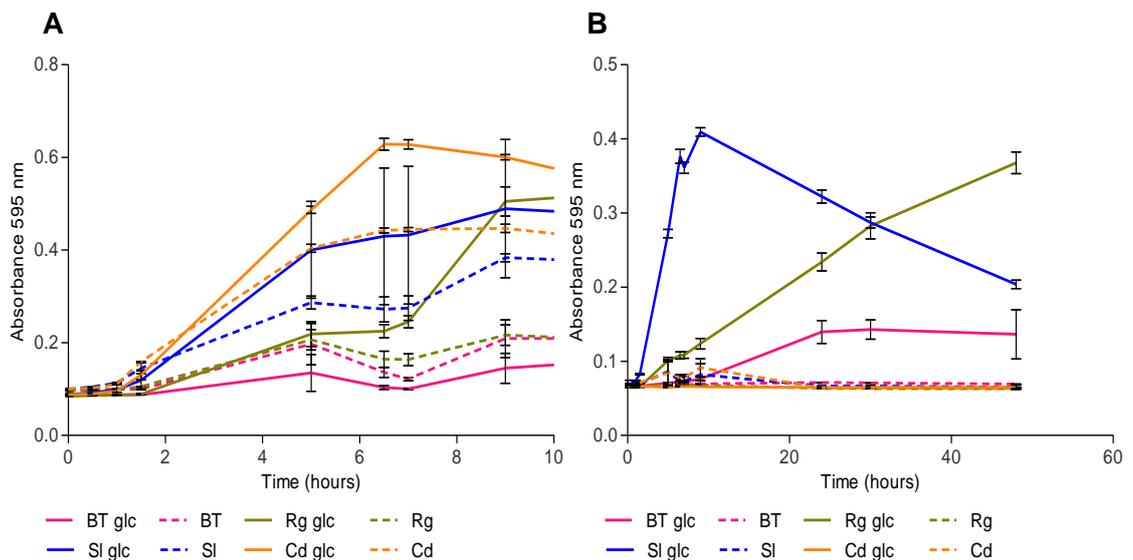


Figure 64: Growth curves of *S. Typhimurium*, *C. difficile*, *R. gnavus* and *B. thetaiotaomicron* on minimal media

Change in absorbance at OD₅₉₅ over time of microtiter cultures of *C. difficile* NCTC 12726 (Cd; orange), *B. thetaiotaomicron* ATCC 29148 (Bt; pink), *S. Typhimurium* SL1344 (St; blue) and *R. gnavus* ATCC 29149 (Rg; green) using either **A**) YCFA or **B**) BDM without carbon source (dashed lines) or with 11.1 mM glucose (solid line).

6.2.4.2 Cross-feeding of *S. Typhimurium* in the presence of bacteria expressing an IT-sialidase or hydrolytic sialidase

To assess the potential of *S. Typhimurium* to cross-feed on sugars liberated by the IT-sialidase of *R. gnavus*, co-culture assays were performed *in vitro* and compared to co-

cultures of *S. Typhimurium* with *B. thetaiotaomicron*. As shown above, *S. Typhimurium* SL1344 cannot utilise 3'SL (Figure 62), therefore any growth would be due to the sialic acid released by either *R. gnavus* or *B. thetaiotaomicron*. Growth of co-cultures in BDM was monitored over time by measuring OD₅₉₅.

The data showed that the *B. thetaiotaomicron* co-culture did not grow within the 24-hour time period used for the experiment, whereas the *R. gnavus* co-culture grew to high density within the same time period (Figure 65). To investigate this further, *R. gnavus* and *B. thetaiotaomicron* mono-cultures were grown on glucose or 3'SL in BDM. These results showed an increased lag-phase of *B. thetaiotaomicron* growth on 3'SL compared to *R. gnavus* with growth only observed after 24 hours (Figure 66). This difference in the lag-phase makes direct comparisons between the *R. gnavus* and *B. thetaiotaomicron* co-cultures in BDM challenging.

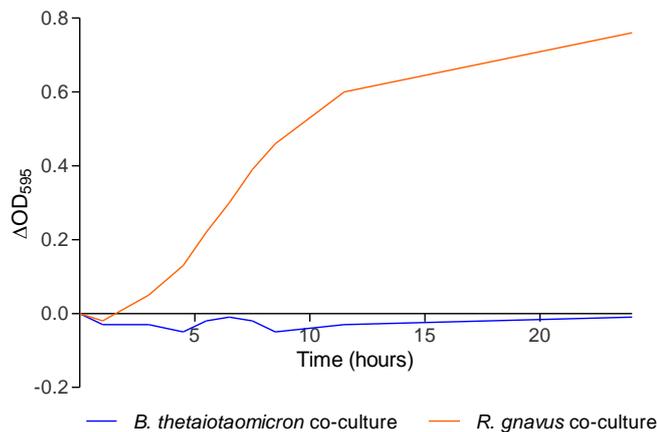


Figure 65: Growth curves of *S. Typhimurium* co-cultures with *R. gnavus* or *B. thetaiotaomicron*. Change in absorbance at OD₅₉₅ over time of co-cultures of *S. Typhimurium* SL1344 with *R. gnavus* ATCC 29149 (orange) or *B. thetaiotaomicron* ATCC 29148 (blue) in BDM with 11.1 mM 3'SL.

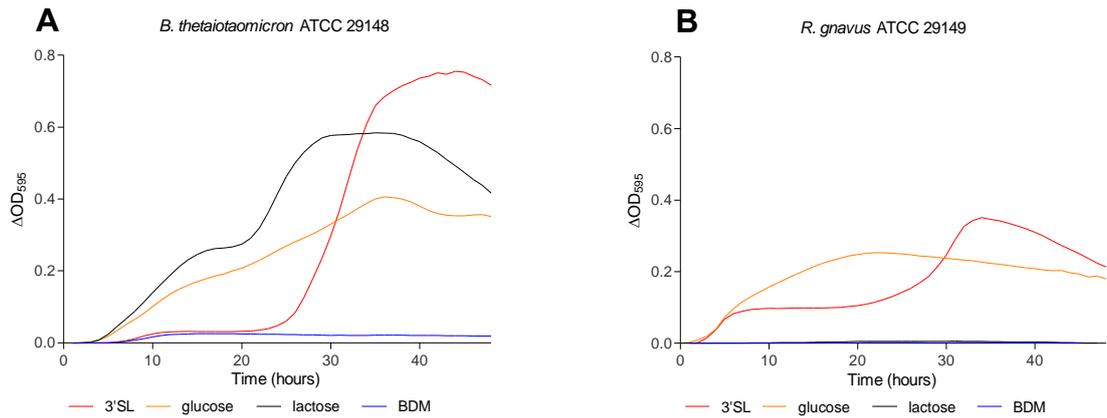


Figure 66: Growth curves of *R. gnavus* or *B. thetaiotaomicron* on 3'SL

Change in absorbance at OD₅₉₅ over time of microtiter cultures of **A)** *B. thetaiotaomicron* ATCC 29148, **B)** *R. gnavus* ATCC 29149 in BDM with 11.1 mM 3'SL (red), glucose (orange), lactose (black) or BDM alone (blue).

To overcome this, co-cultures of *R. gnavus* and *S. Typhimurium* SL1344 grown on 3'SL in BDM were performed alongside a *S. Typhimurium* mono-culture using an equimolar amount of Neu5Ac in BDM to mimic the conditions of a *B. thetaiotaomicron* and *S. Typhimurium* co-culture. These results showed that *S. Typhimurium* growth was reduced 10 fold in co-culture compared to the level of growth observed in the mono-culture of *S. Typhimurium* with Neu5Ac (Figure 67). Although lower, growth of *S. Typhimurium* was still observed in co-cultures with *R. gnavus*. This is likely due to the presence of ~15% Neu5Ac which is a by-product of the IT-sialidase reaction, providing *S. Typhimurium* with a metabolisable carbon source (Monestier *et al.*, 2017).

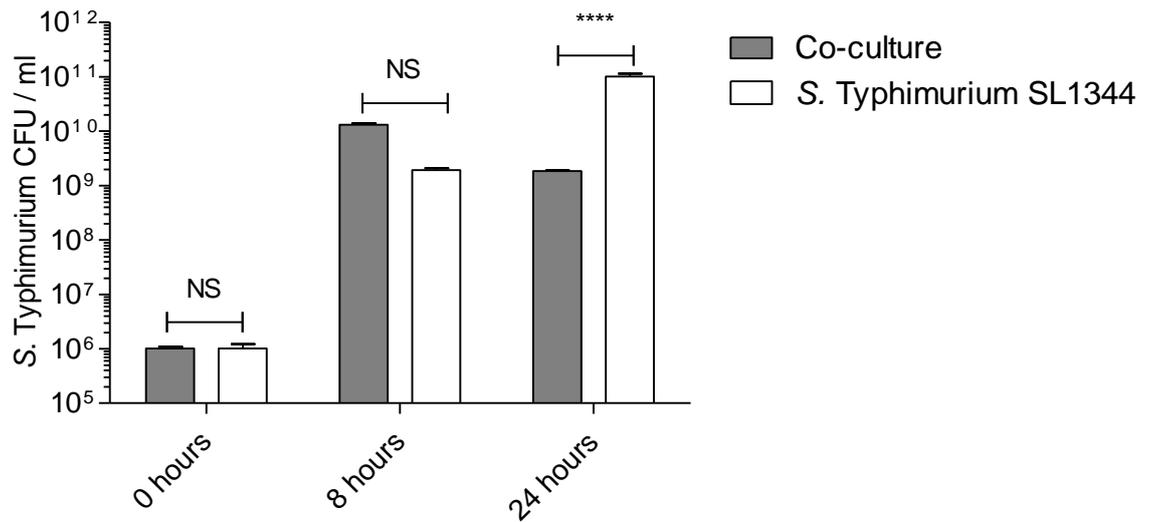


Figure 67: Growth of *S. Typhimurium* in co-culture with *R. gnavus*

S. Typhimurium SL1344 was grown under anaerobic conditions at 37 °C in BDM as a monoculture using Neu5Ac as a sole carbon source (white) or in co-culture with *R. gnavus* ATCC 29149 using 3'SL as a sole carbon source (black). The CFU / ml for *S. Typhimurium* SL1344 at time 0, 8 h and 24 hours was assessed by plating.

Next, co-cultures of *S. Typhimurium* SL1344 with *R. gnavus* or *B. thetaiotaomicron* were carried out using YCFA, taking into account the background growth of *S. Typhimurium* in YCFA unrelated to sialic acid. These results indicated that the presence of *R. gnavus* reduced the growth of *S. Typhimurium* compared to the *B. thetaiotaomicron* co-culture, showing a 10-fold difference at 24 hours, similar to the result obtained in BDM media (Figure 68).

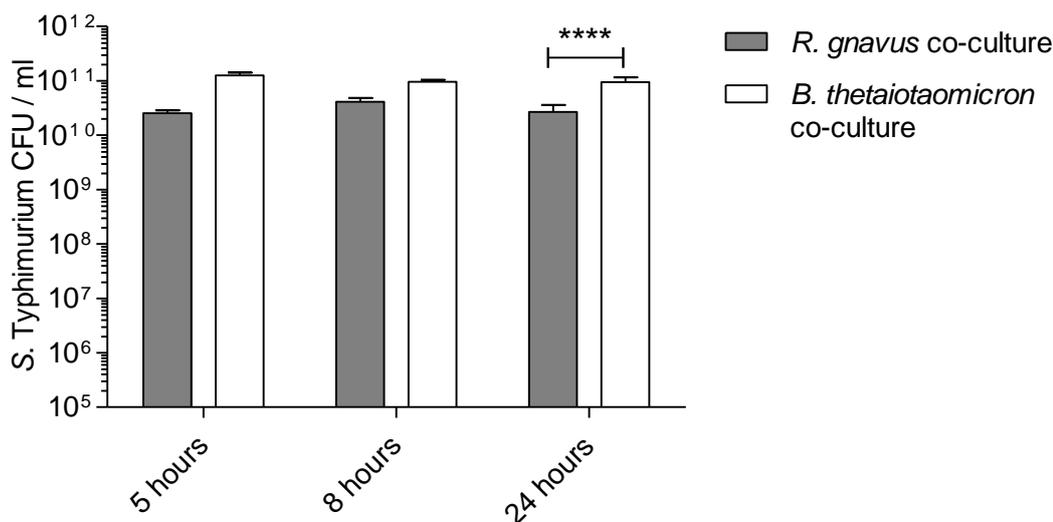


Figure 68: Co-cultures of *S. Typhimurium* in co-culture with *R. gnavus* or *B. thetaiotaomicron*

S. Typhimurium SL1344 was grown in co-culture with *R. gnavus* ATCC 29149 (grey) or *B. thetaiotaomicron* ATCC 29148 (white) using 3'SL as a sole carbon source in YCFA. The CFU / ml for *S. Typhimurium* SL1344 at time 5, 8 and 24 hours was assessed by plating.

Follow up independent growth assays were carried out to test that the *S. Typhimurium* growth in BDM co-cultures with *R. gnavus* was due to the 15% Neu5Ac released by the IT-sialidase and not another nutrient source. Firstly, *R. gnavus* was grown on 3'SL overnight, the *R. gnavus* was pelleted and the resulting supernatant filter sterilised, both pellet and supernatant were used as a potential carbon source for *S. Typhimurium*. Further, *S. Typhimurium* SL1344 growth was tested using 3'SL as a sole carbon source in BDM medium or with addition of either recombinant IT-sialidase from *R. gnavus* or a commercially available hydrolytic sialidase. Finally, Neu5Ac equimolar to the concentration of 3'SL (11.1 mM) used and Neu5Ac at 15% of that concentration (1.7 mM) were used as carbon sources in BDM medium in the presence of *S. Typhimurium* to mimic the expected release of Neu5Ac by the hydrolytic and IT-sialidase, respectively. The results of these growth assays (Figure 69) showed that Neu5Ac and 3'SL with addition of the hydrolytic sialidase allowed for the highest levels of *S. Typhimurium* growth. Furthermore, 3'SL treated with the recombinant IT-sialidase, the *R. gnavus* supernatant or 15% Neu5Ac all showed a similar level of growth but reduced in comparison to growth using Neu5Ac or 3'SL treated with a hydrolytic sialidase. Untreated 3'SL and the *R. gnavus* cell pellet did not support growth of *S. Typhimurium*. Together, these data suggest that the basal growth of *S. Typhimurium* observed in the co-culture experiments with *R. gnavus* is due to the 15% Neu5Ac released by the IT-sialidase reaction.

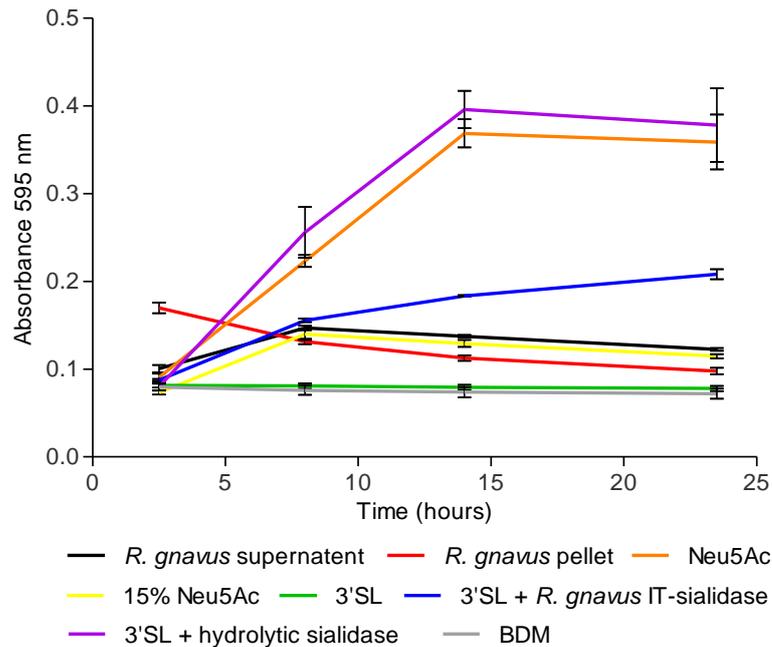


Figure 69: Growth curves of *S. Typhimurium* on Neu5Ac released from exogenous sialidase activity

Absorbance at OD₅₉₅ over time of microtiter cultures of *S. Typhimurium* SL 1344 grown under anaerobic conditions in BDM supplemented with, spent media from *R. gnavus* ATCC 29149 culture grown on 3'SL in BDM overnight (black), *R. gnavus* cell pellet from the same overnight growth (red), 11.1 mM Neu5Ac (orange), 1.7 mM Neu5Ac (15%; yellow), 11.1 mM 3'SL (green), 11.1 mM 3'SL with addition of recombinant *R. gnavus* IT-sialidase (blue) or 11.1 mM 3'SL with addition of a hydrolytic sialidase (purple).

6.2.3 Impact of *R. gnavus* on the colonisation of *S. Typhimurium in vivo*

The *in vitro* data suggest that the presence of *R. gnavus* ATCC 29149 can limit the availability of Neu5Ac released from sialylated substrates compared to bacteria expressing hydrolytic sialidases. This was further investigated in a preliminary *in vivo* study using conventional C57BL/6J mice. Briefly, mice were orally gavaged with streptomycin to disrupt the gut microbiota making mice susceptible to infection and colonisation by *R. gnavus*. After 24 hours, mice were orally gavaged with wild-type *R. gnavus*, *R. gnavus nan* mutant (see section 5.2.1.1) or with PBS as control. After a further 24 hours mice were then challenged with *S. Typhimurium* SL 1344 and monitored for 2 days before culling on day 4 (Figure 70).

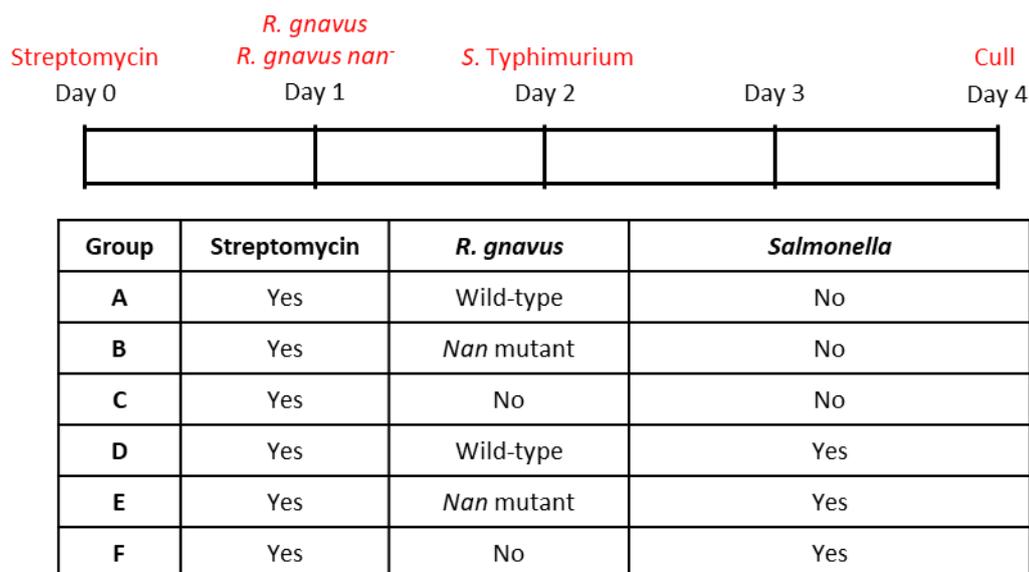


Figure 70: Design of the in vivo study to assess the impact of *R. gnavus* colonisation on *S. Typhimurium* infection

Each group contained 5 mice (2 males and 3 females or 3 males and 2 females), housed in separate cages. Briefly at day 0, all mice were orally gavaged with 20 mg streptomycin, at day 1 mice were gavaged with 10^8 cells *R. gnavus* ATCC 29149, *R. gnavus nan* mutant or PBS, before *S. Typhimurium* SL1344 challenge at day 2 with 10^8 bacteria. Faecal pellets were collected on each day before culling at day 4.

Colonisation of the mice by the *R. gnavus* ATCC 29149 strains was confirmed by qPCR from faecal pellets taken at day 3 or caecal content collected at day 4 (Figure 71). The majority of mice that were gavaged with either the wild-type or *nan* mutant *R. gnavus* strains remained colonised throughout the experiment, however, exceptions were observed. Group A mice were all colonised by wild-type *R. gnavus* at both day 3 and day 4 as expected, however, in group D, mouse 4 was not colonised at day 3 or 4 and while mouse 2 was colonised at day 3, it was no longer colonised at day 4. For the *R. gnavus nan* mutant groups, mouse 2 of group B was not colonised and in group E, all mice were colonised at day 3, but that colonisation was no longer observed in any mice by day 4, this may be due to the *S. Typhimurium* challenge, as we previously reported that the *nan* mutant *R. gnavus* shows impaired colonisation (see section 5.2.1.2).

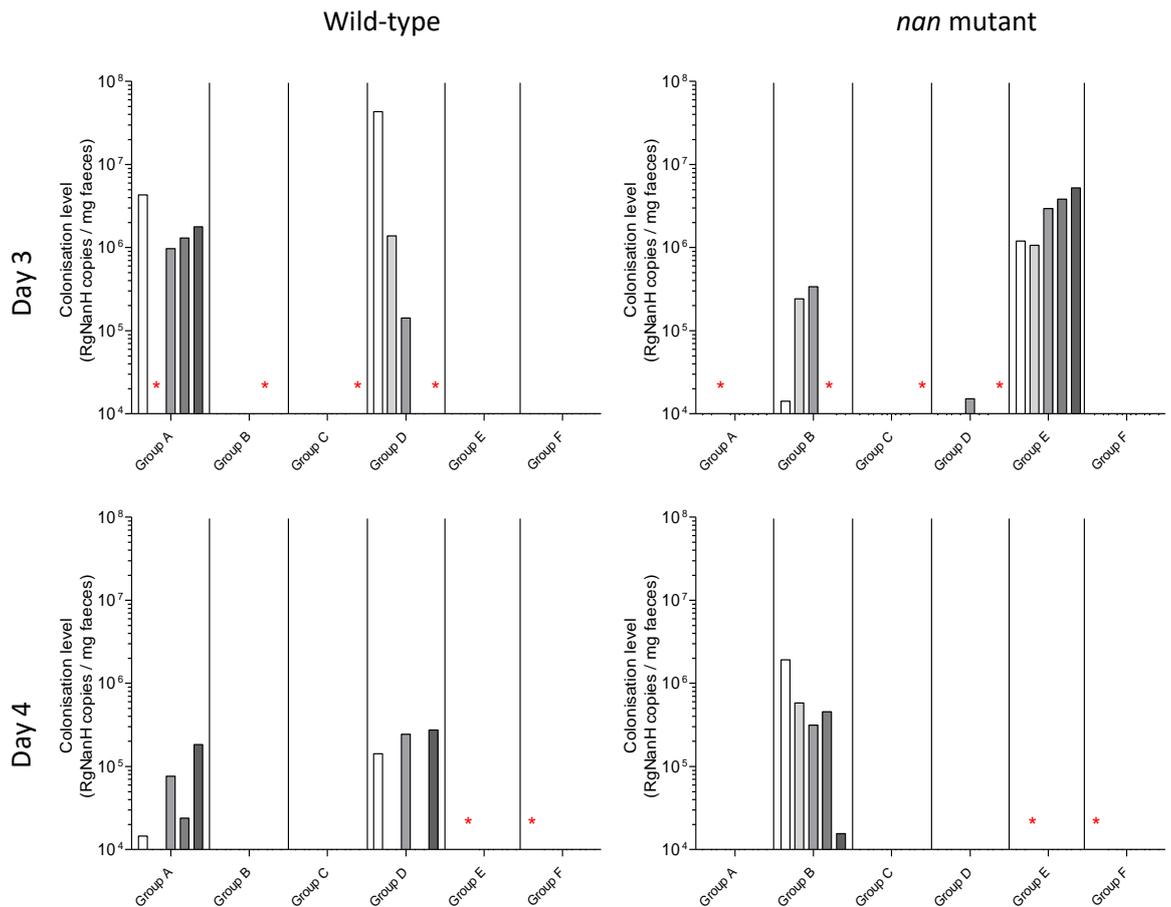


Figure 71: *R. gnavus* colonisation of mice during the *S. Typhimurium* challenge study

In vivo colonisation level by *R. gnavus* ATCC 29149 or *nan* mutant was carried out by qPCR analysis at day 3 and day 4 of the experiment of all 5 mice (mouse 1 white bar – mouse 5 black bar) in each group. Groups A and D were gavaged with wild-type *R. gnavus*, Groups B and E were gavaged with *nan* mutant *R. gnavus* and groups C and F were gavaged with PBS. Red stars indicate where samples were missing or not collected, the limit of detection for the primers used is 10⁴ copies / mg faeces.

Following *S. Typhimurium* challenge, the welfare of mice was monitored each day using a 0 (no signs) – 5 (severe signs) scale to assess symptoms of ill health. These included piloerection, abnormal respiration, oculo-nasal discharge, tremors and/or convulsions, hunched, self-mutilation, intermittent vocalisation and peer interaction. The weight of each mouse was also monitored, and any mouse that was trending towards a loss of 20% body weight before the next check was sacrificed early by schedule 1; this was the case with one mouse in group E and one mouse in group F that were lost at day 3. The total welfare scores were calculated for each mouse to give a ‘wellness’ score, the scores for day 3 were performed blinded by Laura Vaux, the day 4 scores were not blinded. The data showed that the groups of mice colonised with *R. gnavus* wild-type showed less

severe symptoms than the mice colonised with the *nan* mutant or control group at day 3 and day 4 (Figure 72).

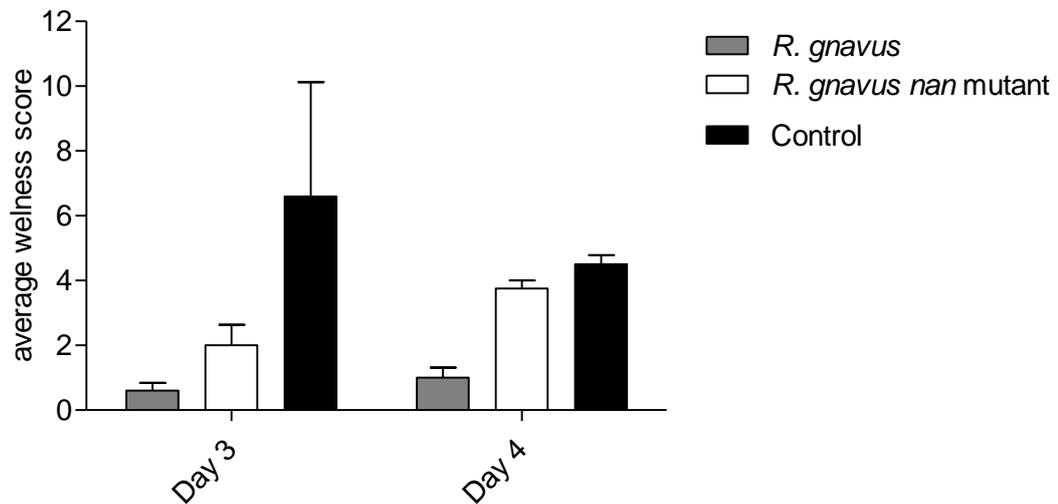


Figure 72: Mouse welfare scores for in vivo *S. Typhimurium* challenge study

Average welfare scores for the three groups challenged with *S. Typhimurium* SL 1344, *R. gnavus* ATC 29149 group – grey bar, *R. gnavus nan* mutant group – white bar and control group black bar. Welfare scores are calculated on a 0 (no signs) – 5 (severe) scale for piloerection, abnormal respiration, oculo-nasal discharge, tremors and/or convulsions, hunched, self-mutilation, intermittent vocalisation and peer interaction. For each mouse the total welfare score was calculated. Scoring at day 3 was blinded and at day 4 was not blinded.

Colonisation of the mice by *S. Typhimurium* SL1344 was also investigated in the mice at day 3 and day 4 by qPCR measuring the copies of the *InvA* gene (Barbau-Piednoir *et al.*, 2013). The *S. Typhimurium* colonisation level in the mice colonised with *R. gnavus* wild-type is skewed by one data point which is 100 times higher than the other mice in the same group. This sample also had a much higher DNA yield (760 ng/mg faeces) compared to the other samples in the same group (24.9 -193.5 ng/mg faeces). The qPCR data obtained for day 3, suggest a reduction in *S. Typhimurium* colonisation in the group colonised with wild-type *R. gnavus* compared to the *nan* mutant and control groups (Figure 73). At day 4, all mice seem to be clearing the *S. Typhimurium* with the *R. gnavus nan* mutant and wild-type group showing a similar colonisation level.

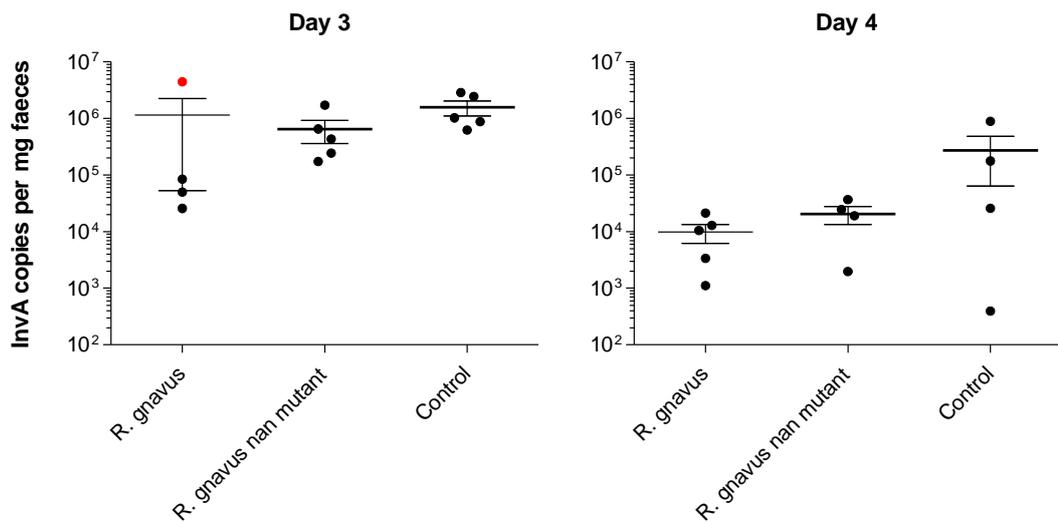


Figure 73: Impact of *R. gnavus* on *S. Typhimurium* infection *in vivo*

qPCR analysis showing the colonisation level *S. Typhimurium* SL1344 at day 3 (left) and day 4 (right) of three groups colonised with *R. gnavus* ATCC 29149 wild-type, *nan* mutant or PBS by quantification of the *InVA* gene.

The microbial composition of the mice was monitored at each day of the experiment (Figure 74). At day 0 the microbiome across all groups was similar and following antibiotic treatment (day 1), the diversity in the microbiome of all mice was depleted. At day 1, the composition of the microbiome was shown to vary across mice in each group and appeared to have a cage-dependent profile (Figure 74). Following addition of *R. gnavus* and an additional 24 hours, the diversity of the microbiome across all groups increased and *R. gnavus* could be detected in groups A, B, D and E in agreement with the qPCR data (Figure 74; day 2). However, there was variability between the gut microbiomes which again appeared to be cage-dependent. For example, although all mice in Group A and Group D were treated in the same way, the microbial profiles of the 2 groups were clearly different, this was also true of groups B and E, and C and F. Following *S. Typhimurium* challenge (Groups D, E and F; day 3 and 4) the microbiome continued to be perturbed, unfortunately the OTU classification could not identify *S. Typhimurium* at the genus or species level, but numerous groups of the family *Enterobacteriaceae* were detected. In the non-*S. Typhimurium* challenged groups (Groups A, B and C) the microbial diversity continued to increase through day 3 and 4 and by day 4, it was almost representative of the day 0 microbiome, with no obvious differences between groups. Therefore, this preliminary *in vivo* mouse study will need

repeating taking into account the variability in microbiome seen throughout the experiment.

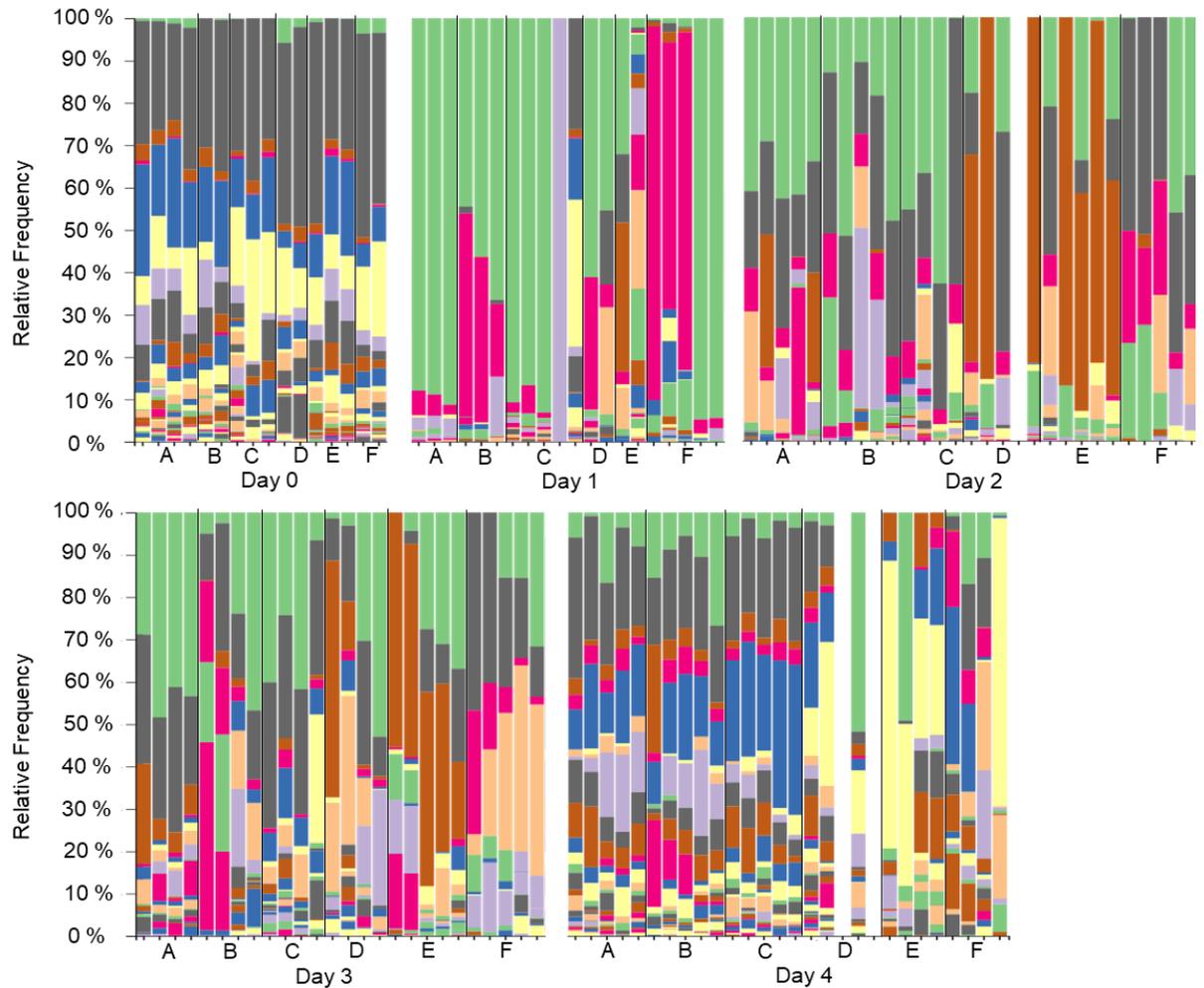


Figure 74: Taxonomic profiles from the *in vivo* study of *S. Typhimurium* colonisation

Faecal samples were taken from mice at day 0, day 1, day 2, day 3, and day 4 of the study, and the taxonomic profile determined by 16S sequencing. Samples are sorted by time point and group number (see Figure 70). Taxonomic classifications were made using the QIIME2 program and Greengenes database and are defined in the key presented in Appendix 4.

We also examined the MUC2 glycosylation pattern in these mice (Appendix 5). The data showed minimal changes in the mucin glycosylation pattern between groups of mice colonised with *R. gnavus* wild-type or *R. gnavus nan* mutant. However, upon *S. Typhimurium* challenge, the mucin glycans were shorter consistent with an inflamed epithelial layer expected during *S. Typhimurium* infection.

6.3 Discussion

Bioinformatics confirmed that while all strains of *Salmonella* and *C. difficile* have the capacity to metabolise Neu5Ac, the majority lack the sialidase required for the release of Neu5Ac from mucin glycans, except for two *Salmonella enterica* strains. In these two strains, the gene encoding the predicted sialidase is not in the same location as the rest of the sialic acid utilisation genes suggesting that they were acquired from different gene transfer events. One of these sialidases, from the *S. Typhimurium* LT2 strain, has been characterised experimentally and is commercially available (Minami *et al.*, 2013). Growth curves of selected strains of both pathogens supported the bioinformatic analyses, with growth observed on Neu5Ac but not on sialoconjugates, except for the LT2 strain which could utilise α 2-3 linked sialic acid oligosaccharides. These *in vitro* growth assays also showed that neither *C. difficile* nor *S. Typhimurium* could utilise 2,7-anhydro-Neu5Ac, supporting the hypothesis that *R. gnavus* ATCC 29149 may have the capability to reduce proliferation of these pathogens *in vivo* by cleaving off sialic acid from mucin glycans and releasing it in a form that they cannot use.

This hypothesis was first tested *in vitro* in co-cultures of *S. Typhimurium* SL1344 and *R. gnavus* ATCC 29149 using 3'SL as a sole carbon source. While *S. Typhimurium* could not utilise 2,7-anhydro-Neu5Ac nor 3'SL, it could still benefit from the IT-sialidase activity of *R. gnavus*. This is likely due to the 15% Neu5Ac released by the *R. gnavus* IT-sialidase (Monestier *et al.*, 2017). This hypothesis was supported by growth assays showing that growth of *S. Typhimurium* using the supernatant of *R. gnavus* ATCC29149 grown on 3'SL, 3'SL treated with the IT-sialidase or Neu5Ac at 15% of the concentration of the 3'SL used, all produced a similar level growth, while *S. Typhimurium* grown on 3'SL treated with a hydrolytic sialidase or equimolar Neu5Ac showed much higher growth levels.

It has been shown *in vivo* that both *S. Typhimurium* and *C. difficile* benefit from the release of free Neu5Ac by the gut symbiont *B. thetaiotaomicron* post antibiotic treatment (Ng *et al.*, 2013). *In vitro* we showed that in co-cultures of *S. Typhimurium* SL1344 with *B. thetaiotaomicron* using 3'SL as a sole carbon source, growth of *S. Typhimurium* was ten times higher than in co-cultures with *R. gnavus*. However, more work is required to support these results such as a transcriptomic analysis to measure expression of genes involved in sialic acid utilisation by *S. Typhimurium* when in co-culture with *R. gnavus*. Taken together, these data provide support to the hypothesis that *R. gnavus* could reduce *S. Typhimurium* proliferation *in vivo*.

This hypothesis was then tested in antibiotic-treated mice colonised with either *R. gnavus* ATCC 29149 wild-type strain or the *R. gnavus nan* mutant followed by *S. Typhimurium* challenge. These studies suggested that *R. gnavus* could reduce *S. Typhimurium* colonisation by approximately 10-fold. Although this was not statistically significant, animal welfare appeared improved in the group treated with *R. gnavus* wild-type. Furthermore, no mice in the *R. gnavus* wild-type colonised group reached the 20% cut off limit for weight loss, whereas one mouse in each of the other groups did pass this limit. A more highly powered study using an attenuated *S. Typhimurium* strain that does not cause systemic infection will need to be carried out to confirm these preliminary findings. Alternatively, a mouse strain genetically resistant to *S. Typhimurium* infection such as CBA/J (Sabag-Daigle *et al.*, 2016, Plant and Glynn, 1976) would allow the monitoring of *S. Typhimurium* colonisation over a longer time frame.

Here the variability in the profile of the gut microbiome following antibiotic treatment makes interpretation of the data more challenging. Therefore, acquiring a more consistent background microbiome at the time of infection would also be beneficial to the study. This could be achieved by giving germ-free mice a defined microbial community containing *R. gnavus* wild-type or the *nan* mutant followed by a period of adaptation before *S. Typhimurium* challenge. This could also be achieved by providing mice a more humanised microbiome which may be more clinically relevant. Furthermore, to better define the gut microbiome of these mice we could consider using metagenome sequencing techniques, this may also give the added benefit being able to detect *S. Typhimurium* levels. Finally, since mice are capable of producing Neu5Gc where humans are not, a more physiologically relevant model would be *CMAH* knock out mice, that no longer produce Neu5Gc (Hedlund *et al.*, 2007).

Importantly for this study, differences in *S. Typhimurium* colonisation and animal welfare were observed between the wild-type *R. gnavus* and the *R. gnavus nan* mutant. These results suggest that the impact on *S. Typhimurium* is not just due to colonisation resistance by *R. gnavus* but is linked to sialic acid metabolism. Following up on work presented in the previous chapter, we propose that 2,7-anhydro-Neu5Ac may be used as a next-generation prebiotic to boost *R. gnavus* in the gut as a strategy to reduce pathogen infection.

Chapter 7

Conclusions and perspectives

Sialic acid is known to be involved in a number of biological processes including immunity, brain development and disease progression, it can also be metabolised as a nutrient source by bacteria (Varki, 2008). As well as its wide ranging and variety of roles, sialic acid comes in many different forms, providing specificity to interactions with different organisms. These interactions are reasonably well studied, though new biological implications of sialic acids continue to be found. One form of sialic acid that has been little researched is 2,7-anhydro-sialic acid, which was first isolated from wet cecum in 1985 (Suzuki *et al.*, 1985). Five years later it was discovered that leech encoded an enzyme (IT-sialidase) that could release 2,7-anhydro-Neu5Ac from sialylated substrates (Li *et al.*, 1990). However, the ability to study this biomolecule has been limited by the effectiveness of methods used to detect and synthesise it. Chemical syntheses with very low production at high cost were for a long time the only described methods for synthesising 2,7-anhydro-Neu5Ac, limiting the scientific community in investigating the potential biological roles of this sialic acid. Owing to the recent discovery by our group of an IT-sialidase in the gut symbiont *R. gnavus* (Tailford *et al.*, 2015b), and the potential of 2,7-anhydro-Neu5Ac in modulating gut microbiota composition in the mucus (this work), renewed efforts have focussed on the synthesis of this compound using enzymatic approaches.

Here we developed a method for the synthesis of 2,7-anhydro-Neu5Ac using the recombinant IT-sialidase from *R. gnavus* ATCC 29149 allowing for mg production of pure 2,7-anhydro-Neu5Ac. Since the IT-sialidase releases around 15% Neu5Ac as a by-product, a commercial sialic aldolase was added to convert the Neu5Ac into ManNAc and pyruvate which could be separated from the 2,7-anhydro-Neu5Ac product more readily in a membrane-enclosed multi-enzyme reaction. Shortly after the publication of our method (Monestier *et al.*, 2017), the group of Xi Chen (Department of chemistry, University of California) developed a method for the gram scale synthesis of 2,7-anhydro-Neu5Ac again based on the action of an IT-sialidase, this time from *S. pneumoniae* (Xiao *et al.*, 2018), and later a more concise protocol using Neu5Ac as starting material (Li *et al.*, 2019). The purification protocols described use a three-phase purification including Biogel P2, a silica column and C18 column. However, these methods did not allow separation of Neu5Ac, which is likely a by-product of the IT-sialidase reaction. We therefore developed a HPLC based protocol to separate 2,7-anhydro-Neu5Ac and Neu5Ac (see section 3.2.1). Following further optimisation, this protocol could provide a robust method of purifying 2,7-anhydro-Neu5Ac to a high level of purity. Alternatively, we could explore the use of *RgNanOx*, an enzyme involved in the *R. gnavus nan* metabolic pathway (identified in this work) to produce 2,7-anhydro-Neu5Ac (see section 4.2.8). As

we showed that this is a reversible reaction, *RgNanOx* could be used for the conversion of Neu5Ac to 2,7-anhydro-Neu5Ac, the excess Neu5Ac would need to be removed using the HPLC method described in this work. Additionally, optimisation of this method would require the shift of the reaction equilibrium towards the production of 2,7-anhydro-Neu5Ac. At 37 °C in sodium phosphate buffer, MS analysis showed a ratio of 2:1 Neu5Ac:2,7-anhydro-Neu5Ac, whereas at 20 °C in PBS (used for NMR experiments) the equilibrium ratio was 10:1. This suggests that it is possible to alter the position of the equilibrium under different experimental conditions. Owing to the resolution of the crystal structure of *RgNanOx* (section 4.2.8) and the identification of key residues in the enzyme active site, it may also be possible to generate mutants of the enzyme to favour the production of 2,7-anhydro-Neu5Ac. Finally, the ability of *RgNanOx* to act on different forms of sialic acid could be tested with the aim of producing 2,7-anhydro-sialic acid derivatives, such as 2,7-anhydro-Neu5Gc and 2,7-anhydro-KDN from Neu5Gc and KDN, respectively.

As well as the synthesis of 2,7-anhydro-Neu5Ac, detection and accurate quantification of this sialic acid derivative from biological samples is challenging. Neu5Ac and its derivatives are generally detected by labelling the compounds with DMB. However, this is dependent on the C2 position of the sialic acid ring being available which is not the case for 2,7-anhydro-Neu5Ac. Here, we have developed an MS approach to allow for the detection of 2,7-anhydro-Neu5Ac using MRMs which accurately identifies this compound. Future work is warranted to improve the method of sample preparation for quantification of 2,7-anhydro-Neu5Ac from complex samples such as caecal contents.

One important question linked to the recent discovery of IT-sialidase expression by *R. gnavus* is whether 2,7-anhydro-Neu5Ac can be metabolised as a nutrient source like Neu5Ac is by certain members of the gut microbiota (Nishiyama *et al.*, 2018, Lewis and Lewis, 2012). We showed that all *R. gnavus* strains expressing the IT-sialidase could grow on 2,7-anhydro-Neu5Ac while they were not able to grow on Neu5Ac. The pathway of Neu5Ac metabolism in bacteria is well understood. Here we elucidated the pathway of 2,7-anhydro-Neu5Ac metabolism in *R. gnavus* (see section 4.2). The IT-sialidase expressed by *R. gnavus* ATCC 29149 has been shown to be part of a larger gene cluster as compared to Neu5Ac clusters (Crost *et al.*, 2016) which we showed is upregulated during growth of *R. gnavus* on 2,7-anhydro-Neu5Ac or 3'SL. To identify the proteins involved in this metabolic pathway, the genes encoded by the *R. gnavus nan* cluster were selected based on bioinformatic analysis and recombinantly expressed in *E. coli*.

These recombinant proteins were then assessed for function using a range of biochemical and biophysical approaches.

The first protein investigated was the SBP of the predicted ABC transporter (*RgSBP*). SBPs are extracellular domains of sugar transport conferring substrate specificity. Using a combination of DSF, fluorescence spectroscopy and ITC analysis, we showed that *RgSBP* was specific for 2,7-anhydro-Neu5Ac and did not bind Neu5Ac. As well as confirming the transport route of 2,7-anhydro-Neu5Ac, this finding also explains why *R. gnavus* is unable to utilise Neu5Ac as a sole carbon source. *RgSBP* is part of a predicted SAT2 transporter and is homologous to the uncharacterised *S. pneumoniae*, *S. gordonii* and *S. sanguinis* SAT2 (Scheepers *et al.*, 2016, Almagro-Moreno and Boyd, 2009). Our SSN analysis of *RgSBP* also identifies these three species as well as 44 other bacterial species as encoding an *RgSBP* like protein (see section 4.2.4, Table 5). It is known that *S. pneumoniae* encodes up to three sialidases including two hydrolytic sialidases and an IT-sialidase (Xu *et al.*, 2011). This bacterium has been shown to utilise Neu5Ac and the presence of an *RgSBP* like protein in *S. pneumoniae* would suggest that it can also utilise 2,7-anhydro-Neu5Ac. The ability to release sialic acid in both the Neu5Ac and 2,7-anhydro-Neu5Ac forms, would potentially provide *S. pneumoniae* a significant advantage in accessing this valuable nutrient source. *RgSBP* is the first characterised domain of a SAT2 transporter and the first shown to bind 2,7-anhydro-Neu5Ac. To date the crystal structure for *RgSBP* has not been resolved.

In the canonical pathway of Neu5Ac metabolism, NanA converts Neu5Ac to ManNAc and pyruvate (Brigham *et al.*, 2009). Bioinformatic analysis of *R. gnavus* NanA (*RgNanA*) revealed that the enzyme differs in sequence compared to Neu5Ac aldolases characterised to date, particularly in key substrate binding residues. It was therefore hypothesised that the substrate specificity may be different. Using a combination of absorbance-based kinetics assays and HPLC, we found that *RgNanA* was specific for Neu5Ac and not 2,7-anhydro-Neu5Ac, which was in agreement with the solved crystal structure of *RgNanA* showing conserved binding and active site structure to previously characterised Neu5Ac aldolases (see section 4.2.7).

With *RgNanA* being active against Neu5Ac and 2,7-anhydro-Neu5Ac being transported by *RgSBP*, it was likely that an intracellular conversion of 2,7-anhydro-Neu5Ac to Neu5Ac was required. Bioinformatics analysis of genes in the *R. gnavus nan* cluster suggested that RUMGNA_02695 (*RgNanOx*) was a likely candidate. Using HPLC and MS methods, we confirmed that the recombinant *RgNanOx* was capable of reversibly converting 2,7-anhydro-Neu5Ac to Neu5Ac (see section 4.2.8). The crystal structure of

RgNanOx was determined using X-ray crystallography showing a dimeric formation with a typical SDR protein fold. This fold is characterised by a central β -sheet with helices on either side and, in most cases, a catalytic tetrad of asparagine, lysine, serine and tyrosine (Kavanagh *et al.*, 2008). *RgNanOx* showed conserved features apart from the asparagine which is lacking in the active site. SDR proteins are abundant in all domains of life and have been associated with a range of substrates including alcohols, sugars, steroids, aromatic compounds, lipids and xenobiotics (Persson *et al.*, 2009, Kavanagh *et al.*, 2008). So far, we have not been able to obtain a crystal structure in complex with a product or intermediate. Future efforts will focus on mutant generation to further understand the reaction mechanism of this novel enzyme, the first shown to be active against 2,7-anhydro-Neu5Ac.

Neu5Ac metabolism is widely spread among members of the gut microbiome, however, to date 2,7-anhydro-Neu5Ac metabolism has only been shown for *R. gnavus* (Croft *et al.*, 2016). Understanding the function of proteins in the extended *nan* cluster of *R. gnavus* allowed us to refine the bioinformatics analysis to identify members of the microbiota that encode homologues of 2,7-anhydro-Neu5Ac metabolism proteins (see section 4.2.3). SSN analysis of the *RgNanOx* identified 69 bacterial species that encode a similar protein, with 40 of these also associated with an *RgSBP* like transporter. For the 29 species not encoding a transporter protein, it is possible that other unknown 2,7-anhydro-Neu5Ac transport proteins exist. Alternatively, if no 2,7-anhydro-Neu5Ac transport mechanism is present, then a novel role for *RgNanOx* may also exist. For example, *E. coli* BW25113 encodes an *RgNanOx* homologue (YjhC), and our preliminary data showed that it has the same function as *RgNanOx*, however, we also showed that this *E. coli* strain cannot use 2,7-anhydro-Neu5Ac as a sole carbon source (data not shown), suggesting an alternative role for *RgNanOx* like enzyme, potentially as a way of storing sialic acid as 2,7-anhydro-Neu5Ac. In addition to the bioinformatics analysis of individual genes involved in 2,7-anhydro-Neu5Ac metabolism, a MultiGeneBlast analysis was carried out. This analysis showed that only species closely related to *R. gnavus* and *S. pneumoniae* strains share a similar *nan* cluster. Taken together these results suggest that the presence of the IT-sialidase and 2,7-anhydro-Neu5Ac metabolism pathway may provide *R. gnavus* with a competitive advantage in the gut by allowing the bacteria to release sialic acid in a form, 2,7-anhydro-Neu5Ac, only it can use.

To investigate this in the context of a complex microbiota, we used a batch culture system seeded with human faecal microbiota. We showed by qPCR that 2,7-anhydro-Neu5Ac and 3'SL were able to promote the growth of *R. gnavus*, and this was also supported by

16S analysis of the microbial community. Interestingly, and in support of the hypothesis, only two other species showed a similar trend in the presence of 2,7-anhydro-Neu5Ac. These were identified as being from the genera *Roseburia*, *Sutterella*, *Blautia*, *Phascolarctobacterium*, *Oscillospira* and *Butyrivibrio*. These results will be confirmed using qPCR analysis of these species as shown for *R. gnavus* in this work. Interestingly bioinformatics analyses only predicted *Blautia* to encode *RgNanOx*, *RgNanH* or *RgSBP* homologs. It is not clear from this sequencing analysis whether these bacteria are primary utilisers of 2,7-anhydro-Neu5Ac or whether they benefit from secondary metabolites. This could be tested using growth assays of monocultures with 2,7-anhydro-Neu5Ac or using SIP methods based on ¹³C-labelled 2,7-anhydro-Neu5Ac, as discussed in section 5.3. The findings presented here could potentially give rise to the use of 2,7-anhydro-Neu5Ac to boost levels of *R. gnavus* as a next generation prebiotic strategy. *R. gnavus* is associated with several diseases including IBD, malnutrition, obesity and brain disorders (see section 1.3.3). As *R. gnavus* is present in the microbiome of 90% of the human population (Qin *et al.*, 2010), its modulation with a prebiotic strategy could lead to novel treatments in conditions associated with an under representation of *R. gnavus*.

To further explore the impact of 2,7-anhydro-Neu5Ac on the host, one approach will be to use mouse models. Although *R. gnavus* was seen to be part of the natural mouse microbiota in earlier studies, *R. gnavus* is not naturally present in C57BL/6J mice, housed at the Disease Modelling Unit (DMU), UEA. We found that to allow colonisation of *R. gnavus* in these mice, an antibiotic treatment was necessary, alternatively we could use 'humanised' mice, taking germ-free mice and gavaging them with a human faecal slurry. This later approach is more representative of the human situation and we could monitor the presence or absence of *R. gnavus* in the faeces of the donor, and add specific strains of *R. gnavus*, for example *R. gnavus* ATCC 29149 wild-type or *nan* mutant. The ability to mutate *R. gnavus* ATCC 29149 as described below will provide novel insights into the mechanisms underpinning the potential impact of 2,7-anhydro-Neu5Ac on the host and microbiome.

Here we used the ClosTron technology (Heap *et al.*, 2010) to generate an *R. gnavus nan* cluster mutant and showed that the *nan* cluster is critical for the fitness of *R. gnavus* in the gut of gnotobiotic mice. The *nan* mutant was less able to penetrate the mucus layer than the wild-type and in co-colonisation was heavily out competed by *R. gnavus* wild-type. In addition to these studies, the large fitness cost of lacking a functional *nan* operon was shown in *S. Typhimurium* infection studies by qPCR (see section 6.2.3). Following *S. Typhimurium* challenge, the *R. gnavus* wild-type strain persisted in the mice, whereas

the *R. gnavus nan* mutant could not be detected in the faecal or caecal content 2 days following *S. Typhimurium* challenge. These data suggest that the *nan* cluster is important for successful *R. gnavus* colonisation of the mucus layer and its persistence in the gut.

Finally, we proposed that *R. gnavus* or IT-sialidase activity may reduce the amount of bound and free Neu5Ac in the gut, by releasing bound sialic acid in the form of 2,7-anhydro-Neu5Ac. It has been shown that an increase of free Neu5Ac post antibiotic treatment leads to an increase in *S. Typhimurium* and *C. difficile* colonisation in mice (Ng *et al.*, 2013). These pathogens do not encode their own sialidases and so rely on commensal organisms in close proximity to release Neu5Ac that they then scavenge. It was therefore proposed that decreasing the level of free Neu5Ac in the gut may be an anti-infective approach, not relying on the use of antibiotics and thus alleviating antimicrobial resistance. Here, we first showed *in vitro* that the enteric pathogens *S. Typhimurium* and *C. difficile* are unable to utilise 2,7-anhydro-Neu5Ac or sialoconjugates as a sole carbon source, whereas they can use Neu5Ac. We next showed that although *S. Typhimurium* was able to benefit from the activity of the IT-sialidase, due to the release of ~15% sialic acid as Neu5Ac, the observed growth was reduced as compared to that from the activity of a hydrolytic sialidase (see section 6.2.4.2). To extend these *in vitro* studies it would be interesting to assess the ability of *S. Typhimurium* to grow on mucins or other sialylated substrates in the presence of a complex microbial community, supplemented with *R. gnavus* wild-type or *nan* mutant in combination with 2,7-anhydro-Neu5Ac as a prebiotic to boost *R. gnavus* growth. These experiments would more closely resemble the gut environment and the competition for sialic acids occurring between gut commensal microbes and pathogens *in vivo*.

In vivo, we observed that pre-colonisation with *R. gnavus* ATCC 29149 afforded some protection against *S. Typhimurium* infection both in terms of animal welfare and colonisation level. This was not observed for mice pre-colonised with the *nan* mutant suggesting that this is linked to sialic metabolism. However, these preliminary data will need supporting through further *in vivo* studies. Future experiments will require long-term colonisation mouse models, using either mice more resistant to *S. Typhimurium* infection such as CBA/J (Sabag-Daigle *et al.*, 2016, Plant and Glynn, 1976) or attenuated strains of *S. Typhimurium*, in order to compare the impact of pre-colonisation with *B. thetaiotaomicron*, a bacterium encoding a hydrolytic sialidase, and *R. gnavus* on *S. Typhimurium* infection.

In conclusion, we showed that 2,7-anhydro-Neu5Ac is metabolised by the *R. gnavus nan* cluster, and that the ability to metabolise 2,7-anhydro-Neu5Ac confers *R. gnavus* with a

competitive advantage over other members of the gut microbiota. A method for mutating *R. gnavus* was also presented and was used to demonstrate the presence of this *nan* cluster plays a vital role in *R. gnavus* fitness in the gut. Due to its ability to reduce Neu5Ac level *in situ*, *R. gnavus* may be considered as a next-generation probiotic to limit the impact of pathogen infection therefore reducing the need for antibiotics. However, this approach may be challenging due to difficulties in upscaling the anaerobic production of this bacterium and finding a delivery method that does not expose *R. gnavus* to high levels of oxygen, which may kill the bacterium. In addition, regulatory procedures surrounding next-generation probiotics are continuing to be developed and may provide further restrictions (O'Toole *et al.*, 2017). Alternatively, given the specificity of 2,7-anhydro-Neu5Ac towards *R. gnavus* without compromising the rest of the gut microbiota, this sialic acid derivative may be considered as a prebiotic in disease conditions which would benefit from an increase in *R. gnavus* strains. One such example is in undernutrition where it has been shown that adding *R. gnavus* to a microbiome from malnourished children in mice could ameliorate growth and metabolic abnormalities compared to mice colonised with the malnourished microbiome alone (Blanton *et al.*, 2016). Another example would be in the context of colitis as it was recently shown that boosting levels of *R. gnavus* led to an alleviation of intestinal inflammation in interleukin 10 deficient mice susceptible to colitis (Grabinger *et al.*, 2019).

In summary, we showed that 2,7-anhydro-Neu5Ac is an important contributor of symbiosis in the gut and a molecular target for microbiome-based strategies to improve human health.

References

- ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., *et al.* (2010) Phenix: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallographica Section D*, 66, 213-221.
- ALLEN, J. M., BERG MILLER, M. E., PENCE, B. D., WHITLOCK, K., NEHRA, V., *et al.* (2015) Voluntary and Forced Exercise Differentially Alters the Gut Microbiome in C57bl/6j Mice. *J Appl Physiol (1985)*, 118, 1059-66.
- ALLEN, S., ZALESKI, A., JOHNSTON, J. W., GIBSON, B. W. & APICELLA, M. A. (2005) Novel Sialic Acid Transporter of Haemophilus Influenzae. *Infection and Immunity*, 73, 5291-300.
- ALMAGRO-MORENO, S. & BOYD, E. F. (2009) Insights into the Evolution of Sialic Acid Catabolism among Bacteria. *BMC Evolutionary Biology*, 9, 118.
- AMBORT, D., JOHANSSON, M. E. V., GUSTAFSSON, J. K., NILSSON, H. E., ERMUND, A., *et al.* (2012) Calcium and Ph-Dependent Packing and Release of the Gel-Forming Muc2 Mucin. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 5645-5650.
- AN, G., WEI, B., XIA, B., MCDANIEL, J. M., JU, T., *et al.* (2007) Increased Susceptibility to Colitis and Colorectal Tumors in Mice Lacking Core 3-Derived O-Glycans. *Journal of Experimental Medicine*, 204, 1417-29.
- ANDERS, S. & HUBER, W. (2010) Differential Expression Analysis for Sequence Count Data. *Genome Biology*, 11, R106.
- ARABYAN, N., PARK, D., FOUTOUHI, S., WEIS, A. M., HUANG, B. H. C., *et al.* (2016) Salmonella Degrades the Host Glycocalyx Leading to Altered Infection and Glycan Remodeling. *Scientific Reports*, 6.
- ARABYAN, N., WEIS, A. M., HUANG, B. C. & WEIMER, B. C. (2017) Implication of Sialidases in Salmonella Infection: Genome Release of Sialidase Knockout Strains from Salmonella Enterica Serovar Typhimurium Lt2. *Genome Announc*, 5.
- ARIKE, L. & HANSSON, G. C. (2016) The Densely O-Glycosylated Muc2 Mucin Protects the Intestine and Provides Food for the Commensal Bacteria. *Journal of Molecular Biology*.
- ARIKE, L., HOLMEN-LARSSON, J. & HANSSON, G. C. (2017) Intestinal Muc2 Mucin O-Glycosylation Is Affected by Microbiota and Regulated by Differential Expression of Glycosyltransferases. *Glycobiology*, 27, 318-328.

- ASKER, N., AXELSSON, M. A. B., OLOFSSON, S. O. & HANSSON, G. C. (1998) Human Muc5ac Mucin Dimerizes in the Rough Endoplasmic Reticulum, Similarly to the Muc2 Mucin. *Biochemical Journal*, 335, 381-387.
- ASNICAR, F., MANARA, S., ZOLFO, M., TRUONG, D. T., SCHOLZ, M., *et al.* (2017) Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems*, 2, e00164-16.
- ASRESSU, K. H. & WANG, C.-C. (2017) Concise Synthesis of 2,7-Anhydrosialic Acid Derivatives and Its Application. *Carbohydrate Research*, 453-454, 44-53.
- ATUMA, C., STRUGALA, V., ALLEN, A. & HOLM, L. (2001) The Adherent Gastrointestinal Mucus Gel Layer: Thickness and Physical State in Vivo. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 280, G922-G929.
- AWAD, W. A., HESS, C. & HESS, M. (2017) Enteric Pathogens and Their Toxin-Induced Disruption of the Intestinal Barrier through Alteration of Tight Junctions in Chickens. *Toxins (Basel)*, 9.
- BACKHED, F., ROSWALL, J., PENG, Y., FENG, Q., JIA, H., *et al.* (2015) Dynamics and Stabilization of the Human Gut Microbiome During the First Year of Life. *Cell Host & Microbe*, 17, 690-703.
- BARBAU-PIEDNOIR, E., BERTRAND, S., MAHILLON, J., ROOSENS, N. H. & BOTTELDOORN, N. (2013) Sybr®Green Qpcr Salmonella Detection System Allowing Discrimination at the Genus, Species and Subspecies Levels. *Applied Microbiology and Biotechnology*, 97, 9811-24.
- BARKER, N. (2014) Adult Intestinal Stem Cells: Critical Drivers of Epithelial Homeostasis and Regeneration. *Nature Reviews Molecular Cell Biology*, 15, 19-33.
- BARKER, N., VAN DE WETERING, M. & CLEVERS, H. (2008) The Intestinal Stem Cell. *Genes & Development*, 22, 1856-1864.
- BENNETT, E. P., MANDEL, U., CLAUSEN, H., GERKEN, T. A., FRITZ, T. A., *et al.* (2012) Control of Mucin-Type O-Glycosylation: A Classification of the Polypeptide Galnac-Transferase Gene Family. *Glycobiology*, 22, 736-756.
- BERGSTROM, J. H., BIRCHENOUGH, G. M., KATONA, G., SCHROEDER, B. O., SCHUTTE, A., *et al.* (2016) Gram-Positive Bacteria Are Held at a Distance in the Colon Mucus by the Lectin-Like Protein Zg16. *Proc Natl Acad Sci U S A*, 113, 13833-13838.
- BERGSTROM, K., FU, J., JOHANSSON, M. E., LIU, X., GAO, N., *et al.* (2017) Core 1- and 3-Derived O-Glycans Collectively Maintain the Colonic Mucus Barrier and Protect against Spontaneous Colitis in Mice. *Mucosal Immunol*, 10, 91-103.

- BERGSTROM, K. S. & XIA, L. (2013) Mucin-Type O-Glycans and Their Roles in Intestinal Homeostasis. *Glycobiology*, 23, 1026-37.
- BERRY, D. & LOY, A. (2018) Stable-Isotope Probing of Human and Animal Microbiome Function. *Trends in Microbiology*, 26, 999-1007.
- BIDOSSI, A., MULAS, L., DECOROSI, F., COLOMBA, L., RICCI, S., *et al.* (2012) A Functional Genomics Approach to Establish the Complement of Carbohydrate Transporters in *Streptococcus Pneumoniae*. *PLoS One*, 7.
- BIRCHENOUGH, G. M., NYSTRÖM, E. E., JOHANSSON, M. E. & HANSSON, G. C. (2016) A Sentinel Goblet Cell Guards the Colonic Crypt by Triggering Nlrp6-Dependent Muc2 Secretion. *Science*, 352, 1535-42.
- BLANTON, L. V., CHARBONNEAU, M. R., SALIH, T., BARRATT, M. J., VENKATESH, S., *et al.* (2016) Gut Bacteria That Prevent Growth Impairments Transmitted by Microbiota from Malnourished Children. *Science*, 351.
- BOLYEN, E., RIDEOUT, J. R., DILLON, M. R., BOKULICH, N. A., ABNET, C. C., *et al.* (2019) Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using Qiime 2. *Nature Biotechnology*, 37, 852-857.
- BREBAN, M., TAP, J., LEBOIME, A., SAID-NAHAL, R., LANGELLA, P., *et al.* (2017) Faecal Microbiota Study Reveals Specific Dysbiosis in Spondyloarthritis. *Annals of the Rheumatic Diseases*, 76, 1614-1622.
- BRIGHAM, C., CAUGHLAN, R., GALLEGOS, R., DALLAS, M. B., GODOY, V. G., *et al.* (2009) Sialic Acid (N-Acetyl Neuraminic Acid) Utilization by *Bacteroides Fragilis* Requires a Novel N-Acetyl Mannosamine Epimerase. *Journal of Bacteriology*, 191, 3629-3638.
- BROCKHAUSEN, I., SCHACHTER, H. & STANLEY, P. (2009) O-Galnac Glycans. *Essentials of Glycobiology*, 2nd eds, 115-127.
- BUNKÓCZI, G. & READ, R. J. (2011) Improvement of Molecular-Replacement Models with Sculptor. *Acta Crystallogr D Biol Crystallogr*, 67, 303-12.
- CANI, P. D. (2018) Human Gut Microbiome: Hopes, Threats and Promises. *Gut*, 67, 1716-1725.
- CANNY, G. O. & MCCORMICK, B. A. (2008) Bacteria in the Intestine, Helpful Residents or Enemies from Within? *Infection and Immunity*, 76, 3360-3373.
- CHAMBERS, E. S., PRESTON, T., FROST, G. & MORRISON, D. J. (2018) Role of Gut Microbiota-Generated Short-Chain Fatty Acids in Metabolic and Cardiovascular Health. *Curr Nutr Rep*, 7, 198-206.
- CHENG, H. & LEBLOND, C. P. (1974) Origin, Differentiation and Renewal of 4 Main Epithelial-Cell Types in Mouse Small Intestine .5. Unitarian Theory of Origin of 4 Epithelial-Cell Types. *American Journal of Anatomy*, 141, 537-561.

- CHU, D. M. & AAGAARD, K. M. (2016) Microbiome: Eating for Trillions. *Nature*, 532, 316-7.
- COHEN, M. & VARKI, A. (2010) The Sialome-Far More Than the Sum of Its Parts. *Omics-a Journal of Integrative Biology*, 14, 455-464.
- COLLADO, M. C., RAUTAVA, S., AAKKO, J., ISOLAURI, E. & SALMINEN, S. (2016) Human Gut Colonisation May Be Initiated in Utero by Distinct Microbial Communities in the Placenta and Amniotic Fluid. *Scientific Reports*, 6, 23129.
- CORFIELD, A. P. (2015) Mucins: A Biologically Relevant Glycan Barrier in Mucosal Protection. *Biochimica Et Biophysica Acta-General Subjects*, 1850, 236-252.
- CORFIELD, A. P. (2018) The Interaction of the Gut Microbiota with the Mucus Barrier in Health and Disease in Human. *Microorganisms*, 6, 78.
- CORFIELD, A. P., MYERSCOUGH, N., WARREN, B. F., DURDEY, P., PARASKEVA, C., *et al.* (1999) Reduction of Sialic Acid O-Acetylation in Human Colonic Mucins in the Adenoma-Carcinoma Sequence. *Glycoconj J*, 16, 307-17.
- CORNELISSEN, L. A. M., BLANAS, A., VAN DER HORST, J. C., KRUIJSSEN, L., ZAAL, A., *et al.* (2019) Disruption of Sialic Acid Metabolism Drives Tumor Growth by Augmenting Cd8+ T Cell Apoptosis. *International Journal of Cancer*, 144, 2290-2302.
- CORNICK, S., TAWIAH, A. & CHADEE, K. (2015) Roles and Regulation of the Mucus Barrier in the Gut. *Tissue barriers*, 3, e982426-e982426.
- CORR, S. C., GAHAN, C. & HILL, C. (2008) M-Cells: Origin, Morphology and Role in Mucosal Immunity and Microbial Pathogenesis. *FEMS Immunology and Medical Microbiology*, 52, 2-12.
- COSTELLO, E. K., LAUBER, C. L., HAMADY, M., FIERER, N., GORDON, J. I., *et al.* (2009) Bacterial Community Variation in Human Body Habitats across Space and Time. *Science*, 326, 1694-1697.
- COWTAN, K. (2006) The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains. *Acta Crystallographica Section D*, 62, 1002-1011.
- CROSNIER, C., STAMATAKI, D. & LEWIS, J. (2006) Organizing Cell Renewal in the Intestine: Stem Cells, Signals and Combinatorial Control. *Nature Reviews Genetics*, 7, 349-359.
- CROST, E. H., TAILFORD, L. E., LE GALL, G., FONS, M., HENRISSAT, B., *et al.* (2013) Utilisation of Mucin Glycans by the Human Gut Symbiont Ruminococcus Gnavus Is Strain-Dependent. *PLoS One*, 8, e76341.
- CROST, E. H., TAILFORD, L. E., MONESTIER, M., SWARBRECK, D., HENRISSAT, B., *et al.* (2016) The Mucin-Degradation Strategy of Ruminococcus Gnavus: The Importance of Intramolecular Trans-Sialidases. *Gut Microbes*, 7, 302-312.

- CULLING, C. F., REID, P. E. & DUNN, W. L. (1979) A Histochemical Comparison of the O-Acylated Sialic Acids of the Epithelial Mucins in Ulcerative Colitis, Crohn's Disease, and Normal Controls. *Journal of Clinical Pathology*, 32, 1272-7.
- DAHAN, S., ROTH-WALTER, F., ARNABOLDI, P., AGARWAL, S. & MAYER, L. (2007) Epithelia: Lymphocyte Interactions in the Gut. *Immunological Reviews*, 215, 243-253.
- DAÏËN, C. I., PINGET, G. V., TAN, J. K. & MACIA, L. (2017) Detrimental Impact of Microbiota-Accessible Carbohydrate-Deprived Diet on Gut and Immune Homeostasis: An Overview. *Front Immunol*, 8, 548.
- DANIELS, A. D., CAMPEOTTO, I., VAN DER KAMP, M. W., BOLT, A. H., TRINH, C. H., *et al.* (2014) Reaction Mechanism of N-Acetylneuraminic Acid Lyase Revealed by a Combination of Crystallography, Qm/Mm Simulation, and Mutagenesis. *ACS Chem Biol*, 9, 1025-32.
- DAVID, L. A., MAURICE, C. F., CARMODY, R. N., GOOTENBERG, D. B., BUTTON, J. E., *et al.* (2014) Diet Rapidly and Reproducibly Alters the Human Gut Microbiome. *Nature*, 505, 559-63.
- DE GOFFAU, M. C., LAGER, S., SOVIO, U., GACCIOLI, F., COOK, E., *et al.* (2019) Human Placenta Has No Microbiome but Can Contain Potential Pathogens. *Nature*.
- DEN BESTEN, G., VAN EUNEN, K., GROEN, A. K., VENEMA, K., REIJNGOUD, D. J., *et al.* (2013) The Role of Short-Chain Fatty Acids in the Interplay between Diet, Gut Microbiota, and Host Energy Metabolism. *Journal of Lipid Research*, 54, 2325-40.
- DEPOMMIER, C., EVERARD, A., DRUART, C., PLOVIER, H., VAN HUL, M., *et al.* (2019) Supplementation with Akkermansia Muciniphila in Overweight and Obese Human Volunteers: A Proof-of-Concept Exploratory Study. *Nature Medicine*, 25, 1096-1103.
- DESAI, M. S., SEEKATZ, A. M., KOROPATKIN, N. M., KAMADA, N., HICKEY, C. A., *et al.* (2016) A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, 167, 1339-+.
- DESANTIS, T. Z., HUGENHOLTZ, P., LARSEN, N., ROJAS, M., BRODIE, E. L., *et al.* (2006) Greengenes, a Chimera-Checked 16s Rrna Gene Database and Workbench Compatible with Arb. *Applied and Environmental Microbiology*, 72, 5069-5072.
- DETHLEFSEN, L., HUSE, S., SOGIN, M. L. & RELMAN, D. A. (2008) The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16s Rrna Sequencing. *PLoS Biology*, 6, e280.

- DETHLEFSEN, L. & RELMAN, D. A. (2011) Incomplete Recovery and Individualized Responses of the Human Distal Gut Microbiota to Repeated Antibiotic Perturbation. *Proc Natl Acad Sci U S A*, 108 Suppl 1, 4554-61.
- DONALDSON, G. P., LEE, S. M. & MAZMANIAN, S. K. (2016) Gut Biogeography of the Bacterial Microbiota. *Nature Reviews Microbiology*, 14, 20-32.
- DUDEK-WICHER, R. K., JUNKKA, A. & BARTOSZEWICZ, M. (2018) The Influence of Antibiotics and Dietary Components on Gut Microbiota. *Prz Gastroenterol*, 13, 85-92.
- EARLE, K. A., BILLINGS, G., SIGAL, M., LICHTMAN, J. S., HANSSON, G. C., *et al.* (2015) Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host & Microbe*, 18, 478-488.
- ECKHAUT, V., MACHIELS, K., PERRIER, C., ROMERO, C., MAES, S., *et al.* (2013) Butyricococcus Pullicaecorum in Inflammatory Bowel Disease. *Gut*, 62, 1745-52.
- EGAN, M., MOTHERWAY, M. O., VENTURA, M. & VAN SINDEREN, D. (2014) Metabolism of Sialic Acid by Bifidobacterium Breve Ucc2003. *Applied and Environmental Microbiology*, 80, 4414-4426.
- EL KAOUTARI, A., ARMOUGOM, F., GORDON, J. I., RAOULT, D. & HENRISSAT, B. (2013) The Abundance and Variety of Carbohydrate-Active Enzymes in the Human Gut Microbiota. *Nature Reviews Microbiology*, 11, 497-504.
- EMSLEY, P. (2017) Tools for Ligand Validation in Coot. *Acta Crystallogr D Struct Biol*, 73, 203-210.
- ERMUND, A., SCHUETTE, A., JOHANSSON, M. E. V., GUSTAFSSON, J. K. & HANSSON, G. C. (2013) Studies of Mucus in Mouse Stomach, Small Intestine, and Colon. I. Gastrointestinal Mucus Layers Have Different Properties Depending on Location as Well as over the Peyer's Patches. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 305, G341-G347.
- ETIENNE-MESMIN, L., CHASSAING, B., DESVAUX, M., DE PAEPE, K., GRESSE, R., *et al.* (2019) Experimental Models to Study Intestinal Microbes-Mucus Interactions in Health and Disease. *FEMS Microbiology Reviews*.
- ETZOLD, S. & JUGE, N. (2014) Structural Insights into Bacterial Recognition of Intestinal Mucins. *Current Opinion in Structural Biology*, 28, 23-31.
- ETZOLD, S., MACKENZIE, D. A., JEFFERS, F., WALSHAW, J., ROOS, S., *et al.* (2014) Structural and Molecular Insights into Novel Surface- Exposed Mucus Adhesins from Lactobacillus Reuteri Human Strains. *Molecular Microbiology*, 92, 543-556.
- EVERARD, A., BELZER, C., GEURTS, L., OUWERKERK, J. P., DRUART, C., *et al.* (2013) Cross-Talk between *Akkermansia muciniphila* and Intestinal

- Epithelium Controls Diet-Induced Obesity. *Proceedings of the National Academy of Sciences*, 110, 9066-9071.
- FABREGA, A. & VILA, J. (2013) Salmonella Enterica Serovar Typhimurium Skills to Succeed in the Host: Virulence and Regulation. *Clinical Microbiology Reviews*, 26, 308-341.
- FAITH, J. J., GURUGE, J. L., CHARBONNEAU, M., SUBRAMANIAN, S., SEEDORF, H., *et al.* (2013) The Long-Term Stability of the Human Gut Microbiota. *Science*, 341, 44-+.
- FASANO, A. & SHEA-DONOHUE, T. (2005) Mechanisms of Disease: The Role of Intestinal Barrier Function in the Pathogenesis of Gastrointestinal Autoimmune Diseases. *Nature Clinical Practice Gastroenterology & Hepatology*, 2, 416-422.
- FLINT, H. J., DUNCAN, S. H. & LOUIS, P. (2017) The Impact of Nutrition on Intestinal Bacterial Communities. *Current Opinion in Microbiology*, 38, 59-65.
- FRESE, S. A., MACKENZIE, D. A., PETERSON, D. A., SCHMALTZ, R., FANGMAN, T., *et al.* (2013) Molecular Characterization of Host-Specific Biofilm Formation in a Vertebrate Gut Symbiont. *PLoS Genetics*, 9.
- FU, C., ZHAO, H., WANG, Y., CAI, H., XIAO, Y., *et al.* (2016) Tumor-Associated Antigens: Tn Antigen, Stn Antigen, and T Antigen. *Hla*, 88, 275-286.
- FU, J., WEI, B., WEN, T., JOHANSSON, M. E., LIU, X., *et al.* (2011) Loss of Intestinal Core 1-Derived O-Glycans Causes Spontaneous Colitis in Mice. *Journal of Clinical Investigation*, 121, 1657-66.
- GAMBLIN, S. J. & SKEHEL, J. J. (2010) Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. *Journal of Biological Chemistry*, 285, 28403-9.
- GANGI SETTY, T., MOWERS, J. C., HOBBS, A. G., MAIYA, S. P., SYED, S., *et al.* (2018) Molecular Characterization of the Interaction of Sialic Acid with the Periplasmic Binding Protein from Haemophilus Ducreyi. *Journal of Biological Chemistry*, 293, 20073-20084.
- GERLT, J. A., BOUVIER, J. T., DAVIDSON, D. B., IMKER, H. J., SADKHIN, B., *et al.* (2015) Enzyme Function Initiative-Enzyme Similarity Tool (Efi-Est): A Web Tool for Generating Protein Sequence Similarity Networks. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1854, 1019-37.
- GIBSON, G. R., HUTKINS, R., SANDERS, M. E., PRESCOTT, S. L., REIMER, R. A., *et al.* (2017) Expert Consensus Document: The International Scientific Association for Probiotics and Prebiotics (Isapp) Consensus Statement on the Definition and Scope of Prebiotics. *Nat Rev Gastroenterol Hepatol*, 14, 491-502.

- GOPHNA, U., KONIKOFF, T. & NIELSEN, H. B. (2017) Oscillospira and Related Bacteria - from Metagenomic Species to Metabolic Features. *Environmental Microbiology*, 19, 835-841.
- GRABINGER, T., GLAUS GARZON, J. F., HAUSMANN, M., GEIRNAERT, A., LACROIX, C., *et al.* (2019) Alleviation of Intestinal Inflammation by Oral Supplementation with 2-Fucosyllactose in Mice. *Front Microbiol*, 10, 1385.
- GUNNING, A. P., KIRBY, A. R., FUELL, C., PIN, C., TAILFORD, L. E., *et al.* (2013) Mining the "Glycocode"-Exploring the Spatial Distribution of Glycans in Gastrointestinal Mucin Using Force Spectroscopy. *FASEB Journal*, 27, 2342-2354.
- GUSTAFSSON, J. K., ERMUND, A., AMBORT, D., JOHANSSON, M. E. V., NILSSON, H. E., *et al.* (2012) Bicarbonate and Functional Cftr Channel Are Required for Proper Mucin Secretion and Link Cystic Fibrosis with Its Mucus Phenotype. *The Journal of Experimental Medicine*, 209, 1263-1272.
- HAIKO, J. & WESTERLUND-WIKSTROM, B. (2013) The Role of the Bacterial Flagellum in Adhesion and Virulence. *Biology*, 2, 1242-67.
- HAINES-MENGES, B. L., WHITAKER, W. B., LUBIN, J. B. & BOYD, E. F. (2015) Host Sialic Acids: A Delicacy for the Pathogen with Discerning Taste. *Microbiol Spectr*, 3.
- HALL, A. B., YASSOUR, M., SAUK, J., GARNER, A., JIANG, X., *et al.* (2017) A Novel Ruminococcus Gnavus Clade Enriched in Inflammatory Bowel Disease Patients. *Genome Medicine*, 9, 103.
- HAMER, H. M., JONKERS, D., VENEMA, K., VANHOUTVIN, S., TROOST, F. J., *et al.* (2008) Review Article: The Role of Butyrate on Colonic Function. *Alimentary Pharmacology & Therapeutics*, 27, 104-119.
- HEAP, J. T., KUEHNE, S. A., EHSAAN, M., CARTMAN, S. T., COOKSLEY, C. M., *et al.* (2010) The Clostron: Mutagenesis in Clostridium Refined and Streamlined. *Journal of Microbiological Methods*, 80, 49-55.
- HEAP, J. T., PENNINGTON, O. J., CARTMAN, S. T. & MINTON, N. P. (2009) A Modular System for Clostridium Shuttle Plasmids. *Journal of Microbiological Methods*, 78, 79-85.
- HECHT, A. L., CASTERLINE, B. W., EARLEY, Z. M., GOO, Y. A., GOODLETT, D. R., *et al.* (2016) Strain Competition Restricts Colonization of an Enteric Pathogen and Prevents Colitis. *Embo Reports*, 17, 1281-1291.
- HEDLUND, M., TANGVORANUNTAKUL, P., TAKEMATSU, H., LONG, J. M., HOUSLEY, G. D., *et al.* (2007) N-Glycolylneuraminic Acid Deficiency in Mice:

- Implications for Human Biology and Evolution. *Molecular and Cellular Biology*, 27, 4340-6.
- HENKE, M. T., KENNY, D. J., CASSILLY, C. D., VLAMAKIS, H., XAVIER, R. J., *et al.* (2019) Ruminococcus Gnavus, a Member of the Human Gut Microbiome Associated with Crohn's Disease, Produces an Inflammatory Polysaccharide. *Proceedings of the National Academy of Sciences*, 201904099.
- HEWS, C. L., TRAN, S. L., WEGMANN, U., BRETT, B., WALSHAM, A. D. S., *et al.* (2017) The Stce Metalloprotease of Enterohaemorrhagic Escherichia Coli Reduces the Inner Mucus Layer and Promotes Adherence to Human Colonic Epithelium Ex Vivo. *Cellular Microbiology*, 19.
- HIIPPALA, K., KAINULAINEN, V., KALLIOMÄKI, M., ARKKILA, P. & SATOKARI, R. (2016) Mucosal Prevalence and Interactions with the Epithelium Indicate Commensalism of Sutterella Spp. *Front Microbiol*, 7, 1706.
- HILDEBRAND, F., MOITINHO-SILVA, L., BLASCHE, S., JAHN, M. T., GOSSMANN, T. I., *et al.* (2019) Antibiotics-Induced Monodominance of a Novel Gut Bacterial Order. *Gut*, gvtjnl-2018-317715.
- HILL, C., GUARNER, F., REID, G., GIBSON, G. R., MERENSTEIN, D. J., *et al.* (2014) Expert Consensus Document. The International Scientific Association for Probiotics and Prebiotics Consensus Statement on the Scope and Appropriate Use of the Term Probiotic. *Nat Rev Gastroenterol Hepatol*, 11, 506-14.
- HIROTA, K., AHLFORS, H., DUARTE, J. H. & STOCKINGER, B. (2012) Regulation and Function of Innate and Adaptive Interleukin-17-Producing Cells. *EMBO Rep*, 13, 113-20.
- HOOPER, L. V., LITTMAN, D. R. & MACPHERSON, A. J. (2012) Interactions between the Microbiota and the Immune System. *Science*, 336, 1268-1273.
- HOOPER, L. V., MIDTVEDT, T. & GORDON, J. I. (2002) How Host-Microbial Interactions Shape the Nutrient Environment of the Mammalian Intestine. *Annual Review of Nutrition*, 22, 283-307.
- HOPKINS, A. P., HAWKHEAD, J. A. & THOMAS, G. H. (2013) Transport and Catabolism of the Sialic Acids N-Glycolylneuraminic Acid and 3-Keto-3-Deoxy-D-Glycero-D-Galactonononic Acid by Escherichia Coli K-12. *FEMS Microbiology Letters*, 347, 14-22.
- HOYER, L. L., ROGGENTIN, P., SCHAUER, R. & VIMR, E. R. (1991) Purification and Properties of Cloned Salmonella-Typhimurium Lt2 Sialidase with Virus-Typical Kinetic Preference for Sialyl Alpha-2-3 Linkages. *Journal of Biochemistry*, 110, 462-467.

- HRYCKOWIAN, A. J., VAN TREUREN, W., SMITS, S. A., DAVIS, N. M., GARDNER, J. O., *et al.* (2018) Microbiota-Accessible Carbohydrates Suppress Clostridium Difficile Infection in a Murine Model. *Nature Microbiology*, 3, 662-669.
- HUNT, K. M., FOSTER, J. A., FORNEY, L. J., SCHÜTTE, U. M. E., BECK, D. L., *et al.* (2011) Characterization of the Diversity and Temporal Stability of Bacterial Communities in Human Milk. *PLoS One*, 6, e21313.
- HUYNH, N., AYE, A., LI, Y. H., YU, H., CAO, H. Z., *et al.* (2013) Structural Basis for Substrate Specificity and Mechanism of N-Acetyl-D-Neuraminic Acid Lyase from Pasteurella Multocida. *Biochemistry*, 52, 8570-8579.
- HWANG, T. L. & SHAKA, A. J. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A*, 112, 275-279.
- IJSSENNAGGER, N., VAN DER MEER, R. & VAN MIL, S. W. C. (2016) Sulfide as a Mucus Barrier-Breaker in Inflammatory Bowel Disease? *Trends in Molecular Medicine*, 22, 190-199.
- JAKOBSSON, H. E., ABRAHAMSSON, T. R., JENMALM, M. C., HARRIS, K., QUINCE, C., *et al.* (2014) Decreased Gut Microbiota Diversity, Delayed Bacteroidetes Colonisation and Reduced Th1 Responses in Infants Delivered by Caesarean Section. *Gut*, 63, 559-66.
- JAKOBSSON, H. E., RODRIGUEZ-PINEIRO, A. M., SCHUTTE, A., ERMUND, A., BOYSEN, P., *et al.* (2015) The Composition of the Gut Microbiota Shapes the Colon Mucus Barrier. *Embo Reports*, 16, 164-177.
- JIMENEZ, E., MARIN, M. L., MARTIN, R., ODRIUZOLA, J. M., OLIVARES, M., *et al.* (2008) Is Meconium from Healthy Newborns Actually Sterile? *Research in Microbiology*, 159, 187-93.
- JOHANSSON, M. E. & HANSSON, G. C. (2016) Immunological Aspects of Intestinal Mucus and Mucins. *Nature Reviews: Immunology*, 16, 639-49.
- JOHANSSON, M. E. V., AMBORT, D., PELASEYED, T., SCHUTTE, A., GUSTAFSSON, J. K., *et al.* (2011a) Composition and Functional Role of the Mucus Layers in the Intestine. *Cellular and Molecular Life Sciences*, 68, 3635-3641.
- JOHANSSON, M. E. V., LARSSON, J. M. H. & HANSSON, G. C. (2011b) The Two Mucus Layers of Colon Are Organized by the Muc2 Mucin, Whereas the Outer Layer Is a Legislator of Host-Microbial Interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 4659-4665.
- JOHANSSON, M. E. V., PHILLIPSON, M., PETERSSON, J., VELCICH, A., HOLM, L., *et al.* (2008) The Inner of the Two Muc2 Mucin-Dependent Mucus Layers in Colon

- Is Devoid of Bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 15064-15069.
- JOSENHANS, C. & SUERBAUM, S. (2002) The Role of Motility as a Virulence Factor in Bacteria. *International Journal of Medical Microbiology*, 291, 605-614.
- JUGE, N. (2012) Microbial Adhesins to Gastrointestinal Mucus. *Trends in Microbiology*, 20, 30-39.
- JUGE, N., TAILFORD, L. & OWEN, C. D. (2016) Sialidases from Gut Bacteria: A Mini-Review. *Biochemical Society Transactions*, 44, 166-75.
- KAMPHUIS, J. B. J., MERCIER-BONIN, M., EUTAMÈNE, H. & THEODOROU, V. (2017) Mucus Organisation Is Shaped by Colonic Content; a New View. *Scientific Reports*, 7, 8527.
- KASHYAP, P. C., MARCOBAL, A., URSELL, L. K., SMITS, S. A., SONNENBURG, E. D., *et al.* (2013) Genetically Dictated Change in Host Mucus Carbohydrate Landscape Exerts a Diet-Dependent Effect on the Gut Microbiota. *Proc Natl Acad Sci U S A*, 110, 17059-64.
- KAVANAGH, K. L., JÖRNVALL, H., PERSSON, B. & OPPERMANN, U. (2008) Medium- and Short-Chain Dehydrogenase/Reductase Gene and Protein Families. *Cellular and Molecular Life Sciences*, 65, 3895.
- KEEGAN, R. M. & WINN, M. D. (2008) Mrbump: An Automated Pipeline for Molecular Replacement. *Acta Crystallogr D Biol Crystallogr*, 64, 119-24.
- KEENEY, K. M., YURIST-DOUTSCH, S., ARRIETA, M. C. & FINLAY, B. B. (2014) Effects of Antibiotics on Human Microbiota and Subsequent Disease. In: GOTTESMAN, S. (ed.) *Annual Review of Microbiology*, Vol 68.
- KIM, D., PERTEA, G., TRAPNELL, C., PIMENTEL, H., KELLEY, R., *et al.* (2013) Tophat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biology*, 14, R36.
- KIM, Y., WANG, X., MA, Q., ZHANG, X.-S. & WOOD, T. K. (2009) Toxin-Antitoxin Systems in Escherichia Coli Influence Biofilm Formation through Yjgk (Taba) and Fimbriae. *Journal of Bacteriology*, 191, 1258-1267.
- KIMIO, F., NOBUYOSHI, T. & HARUO, O. (1991) Synthesis of 2, 7-Anhydro-N-Acetylneuraminic Acid. *Chemical & Pharmaceutical Bulletin*, 39, 817-819.
- KLEEREBEZEM, M. & VAUGHAN, E. E. (2009) Probiotic and Gut Lactobacilli and Bifidobacteria: Molecular Approaches to Study Diversity and Activity. *Annual Review of Microbiology*, 63, 269-90.
- KOENIG, J. E., SPOR, A., SCALFONE, N., FRICKER, A. D., STOMBAUGH, J., *et al.* (2011) Succession of Microbial Consortia in the Developing Infant Gut Microbiome. *Proc Natl Acad Sci U S A*, 108 Suppl 1, 4578-85.

- KONIKOFF, T. & GOPHNA, U. (2016) Oscillospira: A Central, Enigmatic Component of the Human Gut Microbiota. *Trends in Microbiology*, 24, 523-524.
- KOROPATKIN, N. M., CAMERON, E. A. & MARTENS, E. C. (2012) How Glycan Metabolism Shapes the Human Gut Microbiota. *Nature Reviews Microbiology*, 10, 323-335.
- KOSTIC, A. D., HOWITT, M. R. & GARRETT, W. S. (2013) Exploring Host-Microbiota Interactions in Animal Models and Humans. *Genes & Development*, 27, 701-718.
- KRISSINEL, E., USKI, V., LEBEDEV, A., WINN, M. & BALLARD, C. (2018) Distributed Computing for Macromolecular Crystallography. *Acta Crystallogr D Struct Biol*, 74, 143-151.
- KRISTENSEN, N. B., BRYRUP, T., ALLIN, K. H., NIELSEN, T., HANSEN, T. H., *et al.* (2016) Alterations in Fecal Microbiota Composition by Probiotic Supplementation in Healthy Adults: A Systematic Review of Randomized Controlled Trials. *Genome Medicine*, 8, 52.
- KUEHNE, S. A. & MINTON, N. P. (2012) Clostron-Mediated Engineering of Clostridium. *Bioengineered*, 3, 247-54.
- LANGDON, A., CROOK, N. & DANTAS, G. (2016) The Effects of Antibiotics on the Microbiome Throughout Development and Alternative Approaches for Therapeutic Modulation. *Genome Med*, 8, 39.
- LANGER, G., COHEN, S. X., LAMZIN, V. S. & PERRAKIS, A. (2008) Automated Macromolecular Model Building for X-Ray Crystallography Using Arp/Warp Version 7. *Nat Protoc*, 3, 1171-9.
- LANGEREIS, M. A., BAKKERS, M. J., DENG, L., PADLER-KARAVANI, V., VERVOORT, S. J., *et al.* (2015) Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep*, 11, 1966-78.
- LARSSON, J. M., KARLSSON, H., CRESPO, J. G., JOHANSSON, M. E., EKLUND, L., *et al.* (2011) Altered O-Glycosylation Profile of Muc2 Mucin Occurs in Active Ulcerative Colitis and Is Associated with Increased Inflammation. *Inflamm Bowel Dis*, 17, 2299-307.
- LARSSON, J. M. H., THOMSSON, K. A., RODRIGUEZ-PINEIRO, A. M., KARLSSON, H. & HANSSON, G. C. (2013) Studies of Mucus in Mouse Stomach, Small Intestine, and Colon. Iii. Gastrointestinal Muc5ac and Muc2 Mucin O-Glycan Patterns Reveal a Regiospecific Distribution. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 305, G357-G363.
- LAVELLE, A., LENNON, G., O'SULLIVAN, O., DOCHERTY, N., BALFE, A., *et al.* (2015) Spatial Variation of the Colonic Microbiota in Patients with Ulcerative Colitis and Control Volunteers. *Gut*, 64, 1553-1561.

- LAWSON, P. A. & FINEGOLD, S. M. (2015) Reclassification of *Ruminococcus Obeum* as *Blautia Obeum* Comb. Nov. *International Journal of Systematic and Evolutionary Microbiology*, 65, 789-93.
- LEBLANC, J. G., MILANI, C., DE GIORI, G. S., SESMA, F., VAN SINDEREN, D., *et al.* (2013) Bacteria as Vitamin Suppliers to Their Host: A Gut Microbiota Perspective. *Current Opinion in Biotechnology*, 24, 160-8.
- LECLAIRE, C., LECOINTE, K., GUNNING, P. A., TRIBOLO, S., KAVANAUGH, D. W., *et al.* (2018) Molecular Basis for Intestinal Mucin Recognition by Galectin-3 and C-Type Lectins. *The FASEB Journal*, 32, 3301-3320.
- LEEDHAM, S. J., BRITTAN, M., MCDONALD, S. A. C. & WRIGHT, N. A. (2005) Intestinal Stem Cells. *Journal of Cellular and Molecular Medicine*, 9, 11-24.
- LEWIS, A. L. & LEWIS, W. G. (2012) Host Sialoglycans and Bacterial Sialidases: A Mucosal Perspective. *Cellular Microbiology*, 14, 1174-82.
- LI, F. & DING, J. (2019) Sialylation Is Involved in Cell Fate Decision During Development, Reprogramming and Cancer Progression. *Protein & Cell*, 10, 550-565.
- LI, H., LIMENITAKIS, J. P., FUHRER, T., GEUKING, M. B., LAWSON, M. A., *et al.* (2015) The Outer Mucus Layer Hosts a Distinct Intestinal Microbial Niche. *Nat Commun*, 6, 8292.
- LI, J., JIA, H., CAI, X., ZHONG, H., FENG, Q., *et al.* (2014) An Integrated Catalog of Reference Genes in the Human Gut Microbiome. *Nature Biotechnology*, 32, 834.
- LI, W., GHOSH, T., BAI, Y., SANTRA, A., XIAO, A., *et al.* (2019) A Substrate Tagging and Two-Step Enzymatic Reaction Strategy for Large-Scale Synthesis of 2,7-Anhydro-Sialic Acid. *Carbohydrate Research*, 479, 41-47.
- LI, Y. T., NAKAGAWA, H., ROSS, S. A., HANSSON, G. C. & LI, S. C. (1990) A Novel Sialidase Which Releases 2,7-Anhydro-Alpha-N-Acetylneuraminic Acid from Sialoglycoconjugates. *Journal of Biological Chemistry*, 265, 21629-21633.
- LIAO, D.-H., ZHAO, J.-B. & GREGERSEN, H. (2009) Gastrointestinal Tract Modelling in Health and Disease. *World Journal of Gastroenterology*, 15, 169-176.
- LIFELY, M. R. & COTTEE, F. H. (1982) Formation and Identification of Two Novel Anhydro Compounds Obtained by Methanolysis of N-Acetylneuraminic Acid and Carboxyl-Reduced, Meningococcal polysaccharide. *Carbohydrate Research*, 107, 187-197.
- LINDEN, S. K., SUTTON, P., KARLSSON, N. G., KOROLIK, V. & MCGUCKIN, M. A. (2008) Mucins in the Mucosal Barrier to Infection. *Mucosal Immunology*, 1, 183-197.

- LIU, S., DA CUNHA, ANDRE P., REZENDE, RAFAEL M., CIALIC, R., WEI, Z., *et al.* (2016) The Host Shapes the Gut Microbiota Via Fecal MicroRNA. *Cell Host & Microbe*, 19, 32-43.
- LOMBARD, V., GOLACONDA RAMULU, H., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. (2014) The Carbohydrate-Active Enzymes Database (Cazy) in 2013. *Nucleic Acids Research*, 42, D490-5.
- LOZUPONE, C. A., STOMBAUGH, J. I., GORDON, J. I., JANSSON, J. K. & KNIGHT, R. (2012) Diversity, Stability and Resilience of the Human Gut Microbiota. *Nature*, 489, 220-30.
- MACHIELS, K., JOOSSENS, M., SABINO, J., DE PRETER, V., ARIJS, I., *et al.* (2014) A Decrease of the Butyrate-Producing Species *Roseburia Hominis* and *Faecalibacterium Prausnitzii* Defines Dysbiosis in Patients with Ulcerative Colitis. *Gut*, 63, 1275-1283.
- MACKENZIE, D. A., JEFFERS, F., PARKER, M. L., VIBERT-VALLET, A., BONGAERTS, R. J., *et al.* (2010) Strain-Specific Diversity of Mucus-Binding Proteins in the Adhesion and Aggregation Properties of *Lactobacillus Reuteri*. *Microbiology-Sgm*, 156, 3368-3378.
- MALMBERG, E. K., NOAKSSON, K. A., PHILLIPSON, M., JOHANSSON, M. E. V., HINOJOSA-KURTZBERG, M., *et al.* (2006) Increased Levels of Mucins in the Cystic Fibrosis Mouse Small Intestine, and Modulator Effects of the Muc1 Mucin Expression. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 291, G203-G210.
- MARCOBAL, A., SOUTHWICK, A. M., EARLE, K. A. & SONNENBURG, J. L. (2013) A Refined Palate: Bacterial Consumption of Host Glycans in the Gut. *Glycobiology*, 23, 1038-+.
- MARION, C., ATEN, A. E., WOODIGA, S. A. & KING, S. J. (2011a) Identification of an Atpase, Msmk, Which Energizes Multiple Carbohydrate Abc Transporters in *Streptococcus Pneumoniae*. *Infection and Immunity*, 79, 4193-200.
- MARION, C., BURNAUGH, A. M., WOODIGA, S. A. & KING, S. J. (2011b) Sialic Acid Transport Contributes to Pneumococcal Colonization. *Infection and Immunity*, 79, 1262-1269.
- MARTENS, E. C., NEUMANN, M. & DESAI, M. S. (2018) Interactions of Commensal and Pathogenic Microorganisms with the Intestinal Mucosal Barrier. *Nature Reviews Microbiology*, 16, 457-470.
- MARTINEZ, J., STEENBERGEN, S. & VIMR, E. (1995) Derived Structure of the Putative Sialic Acid Transporter from *Escherichia Coli* Predicts a Novel Sugar Permease Domain. *Journal of Bacteriology*, 177, 6005-10.

- MATSUSHITA, M., KILPATRICK, D. C., LEE, B. L. & WAKAMIYA, N. (2012) Soluble Host-Defense Lectins. *Journal of Biomedicine and Biotechnology*, 2012, 2.
- MAYER, M. & JAMES, T. L. (2004) Nmr-Based Characterization of Phenothiazines as a Rna Binding Scaffold. *Journal of the American Chemical Society*, 126, 4453-60.
- MAYER, M. & MEYER, B. (1999) Characterization of Ligand Binding by Saturation Transfer Difference Nmr Spectroscopy. *Angewandte Chemie International Edition*, 38, 1784-1788.
- MCCOY, A. J., OEFFNER, R. D., MILLAN, C., SAMMITO, M., USON, I., *et al.* (2018) Gyre and Gimble: A Maximum-Likelihood Replacement for Patterson Correlation Refinement. *Acta Crystallogr D Struct Biol*, 74, 279-289.
- MCDONALD, D., PRICE, M. N., GOODRICH, J., NAWROCKI, E. P., DESANTIS, T. Z., *et al.* (2012) An Improved Greengenes Taxonomy with Explicit Ranks for Ecological and Evolutionary Analyses of Bacteria and Archaea. *Isme j*, 6, 610-8.
- MCFALL-NGAI, M. (2006) Love the One You're With: Vertebrate Guts Shape Their Microbiota. *Cell*, 127, 247-249.
- MCGUCKIN, M. A., LINDEN, S. K., SUTTON, P. & FLORIN, T. H. (2011) Mucin Dynamics and Enteric Pathogens. *Nature Reviews Microbiology*, 9, 265-278.
- MEDEMA, M. H., TAKANO, E. & BREITLING, R. (2013) Detecting Sequence Homology at the Gene Cluster Level with Multigeneblast. *Molecular Biology and Evolution*, 30, 1218-23.
- MINAMI, A., ISHIBASHI, S., IKEDA, K., ISHITSUBO, E., HORI, T., *et al.* (2013) Catalytic Preference of Salmonella Typhimurium Lt2 Sialidase for N-Acetylneuraminic Acid Residues over N-Glycolylneuraminic Acid Residues. *Febs Open Bio*, 3, 231-236.
- MONACO, S., TAILFORD, L. E., JUGE, N. & ANGULO, J. (2017) Differential Epitope Mapping by Std Nmr Spectroscopy to Reveal the Nature of Protein-Ligand Contacts. *Angew Chem Int Ed Engl*, 56, 15289-15293.
- MONDAL, M., NAG, D., KOLEY, H., SAHA, D. R. & CHATTERJEE, N. S. (2014) The Vibrio Cholerae Extracellular Chitinase Chia2 Is Important for Survival and Pathogenesis in the Host Intestine. *PLoS One*, 9, e103119.
- MONESTIER, M., LATOUSAKIS, D., BELL, A., TRIBOLO, S., TAILFORD, L. E., *et al.* (2017) Membrane-Enclosed Multienzyme (Meme) Synthesis of 2,7-Anhydro-Sialic Acid Derivatives. *Carbohydrate Research*, 451, 110-117.
- MONIAUX, N., ESCANDE, F., PORCHET, N., AUBERT, J. P. & BATRA, S. K. (2001) Structural Organization and Classification of the Human Mucin Genes. *Frontiers in Bioscience*, 6, D1192-D1206.
- MOYA, A. & FERRER, M. (2016) Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends in Microbiology*, 24, 402-413.

- NDEH, D. & GILBERT, H. J. (2018) Biochemistry of Complex Glycan Depolymerisation by the Human Gut Microbiota. *FEMS Microbiology Reviews*, 42, 146-164.
- NEISH, A. S. (2009) Microbes in Gastrointestinal Health and Disease. *Gastroenterology*, 136, 65-80.
- NEPRAVISHTA, R., WALPOLE, S., TAILFORD, L., JUGE, N. & ANGULO, J. (2019) Deriving Ligand Orientation in Weak Protein-Ligand Complexes by Deep-Std Nmr Spectroscopy in the Absence of Protein Chemical-Shift Assignment. *ChemBioChem*, 20, 340-344.
- NG, K. M., FERREYRA, J. A., HIGGINBOTTOM, S. K., LYNCH, J. B., KASHYAP, P. C., *et al.* (2013) Microbiota-Liberated Host Sugars Facilitate Post-Antibiotic Expansion of Enteric Pathogens. *Nature*, 502, 96-9.
- NISHIYAMA, K., NAGAI, A., URIBAYASHI, K., YAMAMOTO, Y., MUKAI, T., *et al.* (2018) Two Extracellular Sialidases from Bifidobacterium Bifidum Promote the Degradation of Sialyl-Oligosaccharides and Support the Growth of Bifidobacterium Breve. *Anaerobe*.
- O'TOOLE, P. W., MARCHESI, J. R. & HILL, C. (2017) Next-Generation Probiotics: The Spectrum from Probiotics to Live Biotherapeutics. *Nature Microbiology*, 2, 17057.
- OLBJORN, C., CVANCAROVA SMASTUEN, M., THIIIS-EVENSEN, E., NAKSTAD, B., VATN, M. H., *et al.* (2019) Fecal Microbiota Profiles in Treatment-Naive Pediatric Inflammatory Bowel Disease - Associations with Disease Phenotype, Treatment, and Outcome. *Clin Exp Gastroenterol*, 12, 37-49.
- OWEN, C. D., LUKACIK, P., POTTER, J. A., SLEATOR, O., TAYLOR, G. L., *et al.* (2015) Streptococcus Pneumoniae Nanc: Structural Insights into the Specificity and Mechanism of a Sialidase That Produces a Sialidase Inhibitor. *Journal of Biological Chemistry*, 290, 27736-48.
- PACHECO, A. R., CURTIS, M. M., RITCHIE, J. M., MUNERA, D., WALDOR, M. K., *et al.* (2012) Fucose Sensing Regulates Bacterial Intestinal Colonization. *Nature*, 492, 113-7.
- PAINTER, J. & MERRITT, E. A. (2006) Tlsmid Web Server for the Generation of Multi-Group Tls Models. *Journal of Applied Crystallography*, 39, 109-111.
- PANDA, S., EL KHADER, I., CASELLAS, F., LOPEZ VIVANCOS, J., GARCIA CORS, M., *et al.* (2014) Short-Term Effect of Antibiotics on Human Gut Microbiota. *PLoS One*, 9, e95476.
- PATTERSON, A. M. & WATSON, A. J. M. (2017) Deciphering the Complex Signaling Systems That Regulate Intestinal Epithelial Cell Death Processes and Shedding. *Front Immunol*, 8.

- PELASEYED, T., BERGSTROM, J. H., GUSTAFSSON, J. K., ERMUND, A., BIRCHENOUGH, G. M. H., *et al.* (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.
- PEREZ-COBAS, A. E., ARTACHO, A., KNECHT, H., FERRUS, M. L., FRIEDRICHS, A., *et al.* (2013) Differential Effects of Antibiotic Therapy on the Structure and Function of Human Gut Microbiota. *PLoS One*, 8, e80201.
- PEREZ-VILAR, J. & MABOLO, R. (2007) Gel-Forming Mucins. Notions from in Vitro Studies. *Histology and Histopathology*, 22, 455-464.
- PERSSON, B., KALLBERG, Y., BRAY, J. E., BRUFORD, E., DELLAPORTA, S. L., *et al.* (2009) The Sdr (Short-Chain Dehydrogenase/Reductase and Related Enzymes) Nomenclature Initiative. *Chem Biol Interact*, 178, 94-8.
- PERUTKA, J., WANG, W., GOERLITZ, D. & LAMBOWITZ, A. M. (2004) Use of Computer-Designed Group II Introns to Disrupt Escherichia Coli DEXH/D-Box Protein and DNA Helicase Genes. *Journal of Molecular Biology*, 336, 421-39.
- PETERSON, L. W. & ARTIS, D. (2014) Intestinal Epithelial Cells: Regulators of Barrier Function and Immune Homeostasis. *Nature Reviews Immunology*, 14, 141-153.
- PETERSSON, J., SCHREIBER, O., HANSSON, G. C., GENDLER, S. J., VELCICH, A., *et al.* (2011) Importance and Regulation of the Colonic Mucus Barrier in a Mouse Model of Colitis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 300, G327-G333.
- PFEIFFER, T., SCHUSTER, S. & BONHOEFFER, S. (2001) Cooperation and Competition in the Evolution of ATP-Producing Pathways. *Science*, 292, 504-507.
- PHANSOPA, C., KOZAK, RADOSLAW P., LIEW, LI P., FREY, ANDREW M., FARMILO, T., *et al.* (2015) Characterization of a Sialate-O-Acetyltransferase (Nans) from the Oral Pathogen Tannerella forsythia That Enhances Sialic Acid Release by NanH, Its Cognate Sialidase. *Biochemical Journal*, 472, 157-167.
- PICKARD, J. M., ZENG, M. Y., CARUSO, R. & NUNEZ, G. (2017) Gut Microbiota: Role in Pathogen Colonization, Immune Responses, and Inflammatory Disease. *Immunological Reviews*, 279, 70-89.
- PLANT, J. & GLYNN, A. A. (1976) Genetics of Resistance to Infection with Salmonella Typhimurium in Mice. *Journal of Infectious Diseases*, 133, 72-8.
- PLUMBRIDGE, J. & VIMR, E. (1999) Convergent Pathways for Utilization of the Amino Sugars N-Acetylglucosamine, N-Acetylmannosamine, and N-Acetylneuraminic Acid by Escherichia Coli. *Journal of Bacteriology*, 181, 47-54.

- POST, D. M., MUNGUR, R., GIBSON, B. W. & MUNSON, R. S., JR. (2005) Identification of a Novel Sialic Acid Transporter in *Haemophilus Ducreyi*. *Infection and Immunity*, 73, 6727-35.
- POWELL, D. W., PINCHUK, I. V., SAADA, J. I., CHEN, X. & MIFFLIN, R. C. (2011) Mesenchymal Cells of the Intestinal Lamina Propria. *In: JULIUS, D. & CLAPHAM, D. E. (eds.) Annual Review of Physiology, Vol 73*.
- POWER, S. E., O'TOOLE, P. W., STANTON, C., ROSS, R. P. & FITZGERALD, G. F. (2014) Intestinal Microbiota, Diet and Health. *British Journal of Nutrition*, 111, 387-402.
- PRAHARAJ, A. B., DEHURY, B., MAHAPATRA, N., KAR, S. K. & BEHERA, S. K. (2018) Molecular Dynamics Insights into the Structure, Function, and Substrate Binding Mechanism of Mucin Desulfating Sulfatase of Gut Microbe *Bacteroides Fragilis*. *Journal of Cellular Biochemistry*, 119, 3618-3631.
- QIN, J., LI, R., RAES, J., ARUMUGAM, M., BURGDORF, K. S., *et al.* (2010) A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing. *Nature*, 464, 59-65.
- RANGARAJAN, E. S., RUANE, K. M., PROTEAU, A., SCHRAG, J. D., VALLADARES, R., *et al.* (2011) Structural and Enzymatic Characterization of Nans (Yjhs), a 9-O-Acetyl N-Acetylneuraminic Acid Esterase from *Escherichia Coli* O157:H7. *Protein Science*, 20, 1208-19.
- RAVCHEEV, D. A. & 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. *Front Genet*, 8, 111.
- RAWLS, J. F., MAHOWALD, M. A., LEY, R. E. & GORDON, J. I. (2006) Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-Free Recipients Reveal Host Habitat Selection. *Cell*, 127, 423-433.
- REID, P. E., CULLING, C. F., DUNN, W. L., RAMEY, C. W. & CLAY, M. G. (1984) Chemical and Histochemical Studies of Normal and Diseased Human Gastrointestinal Tract. I. A Comparison between Histologically Normal Colon, Colonic Tumours, Ulcerative Colitis and Diverticular Disease of the Colon. *Histochem J*, 16, 235-51.
- REY, F. E., GONZALEZ, M. D., CHENG, J., WU, M., AHERN, P. P., *et al.* (2013) Metabolic Niche of a Prominent Sulfate-Reducing Human Gut Bacterium. *Proc Natl Acad Sci U S A*, 110, 13582-7.
- RIBET, D. & COSSART, P. (2015) How Bacterial Pathogens Colonize Their Hosts and Invade Deeper Tissues. *Microbes and Infection*, 17, 173-183.

- ROBBE, C., CAPON, C., CODDEVILLE, B. & MICHALSKI, J. C. (2004) Structural Diversity and Specific Distribution of O-Glycans in Normal Human Mucins Along the Intestinal Tract. *Biochemical Journal*, 384, 307-316.
- ROBBE, C., CAPON, C., MAES, E., ROUSSET, M., ZWEIBAUM, A., *et al.* (2003) Evidence of Regio-Specific Glycosylation in Human Intestinal Mucins - Presence of an Acidic Gradient Along the Intestinal Tract. *Journal of Biological Chemistry*, 278, 46337-46348.
- ROBINSON, L. S., LEWIS, W. G. & LEWIS, A. L. (2017) The Sialate O-Acetyltransferase from Gut Bacteroidetes Species Enables Sialidase-Mediated Cross-Species Foraging of 9-O-Acetylated Sialoglycans. *Journal of Biological Chemistry*, 292, 11861-11872.
- ROOKS, M. G. & GARRETT, W. S. (2016) Gut Microbiota, Metabolites and Host Immunity. *Nature Reviews Immunology*, 16, 341.
- SABAG-DAIGLE, A., BLUNK, H. M., GONZALEZ, J. F., STEIDLEY, B. L., BOYAKA, P. N., *et al.* (2016) Use of Attenuated but Metabolically Competent Salmonella as a Probiotic to Prevent or Treat Salmonella Infection. *Infection and Immunity*, 84, 2131-2140.
- SAGHEDDU, V., PATRONE, V., MIRAGOLI, F., PUGLISI, E. & MORELLI, L. (2016) Infant Early Gut Colonization by Lachnospiraceae: High Frequency of Ruminococcus Gnavus. *Front Pediatr*, 4, 57.
- SALYERS, A. A., VERCELLOTTI, J. R., WEST, S. E. H. & WILKINS, T. D. (1977) Fermentation of Mucin and Plant Polysaccharides by Strains of Bacteroides from Human Colon. *Applied and Environmental Microbiology*, 33, 319-322.
- SANCHEZ-WEATHERBY, J., SANDY, J., MIKOLAJEK, H., LOBLEY, C. M. C., MAZZORANA, M., *et al.* (2019) Vmxi: A Fully Automated, Fully Remote, High-Flux in Situ Macromolecular Crystallography Beamline. *J Synchrotron Radiat*, 26, 291-301.
- SANCHO, E., BATLLE, E. & CLEVERS, H. (2003) Live and Let Die in the Intestinal Epithelium. *Current Opinion in Cell Biology*, 15, 763-770.
- SCHAUER, R. (2009) Sialic Acids as Regulators of Molecular and Cellular Interactions. *Current Opinion in Structural Biology*, 19, 507-14.
- SCHEEPERS, G. H., LYCKLAMA, A. N. J. A. & POOLMAN, B. (2016) An Updated Structural Classification of Substrate-Binding Proteins. *FEBS Letters*, 590, 4393-4401.
- SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., *et al.* (2012) Fiji: An Open-Source Platform for Biological-Image Analysis. *Nature Methods*, 9, 676.

- SCHLEIF, R. (2010) Arac Protein, Regulation of the L-Arabinose Operon in Escherichia Coli, and the Light Switch Mechanism of Arac Action. *FEMS Microbiology Reviews*, 34, 779-796.
- SCHMIDT, T. S. B., RAES, J. & BORK, P. (2018) The Human Gut Microbiome: From Association to Modulation. *Cell*, 172, 1198-1215.
- SEVERI, E., HOOD, D. W. & THOMAS, G. H. (2007) Sialic Acid Utilization by Bacterial Pathogens. *Microbiology*, 153, 2817-22.
- SEVERI, E., HOSIE, A. H., HAWKHEAD, J. A. & THOMAS, G. H. (2010) Characterization of a Novel Sialic Acid Transporter of the Sodium Solute Symporter (Sss) Family and in Vivo Comparison with Known Bacterial Sialic Acid Transporters. *FEMS Microbiology Letters*, 304, 47-54.
- SHANAHAN, F., VAN SINDEREN, D., O'TOOLE, P. W. & STANTON, C. (2017) Feeding the Microbiota: Transducer of Nutrient Signals for the Host. *Gut*, 66, 1709-1717.
- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., *et al.* (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13, 2498-504.
- SHELTON, A. N., SETH, E. C., MOK, K. C., HAN, A. W., JACKSON, S. N., *et al.* (2019) Uneven Distribution of Cobamide Biosynthesis and Dependence in Bacteria Predicted by Comparative Genomics. *Isme j*, 13, 789-804.
- SHERWIN, E., DINAN, T. G. & CRYAN, J. F. (2018) Recent Developments in Understanding the Role of the Gut Microbiota in Brain Health and Disease. *Annals of the New York Academy of Sciences*, 1420, 5-25.
- SHETH, R. U., LI, M., JIANG, W., SIMS, P. A., LEONG, K. W., *et al.* (2019) Spatial Metagenomic Characterization of Microbial Biogeography in the Gut. *Nature Biotechnology*, 37, 877-883.
- SHIN, J., NOH, J.-R., CHANG, D.-H., KIM, Y.-H., KIM, M. H., *et al.* (2019) Elucidation of Akkermansia Muciniphila Probiotic Traits Driven by Mucin Depletion. *Front Microbiol*, 10.
- SMART, O. S., WOMACK, T. O., FLENSBURG, C., KELLER, P., PACIOREK, W., *et al.* (2012) Exploiting Structure Similarity in Refinement: Automated Ncs and Target-Structure Restraints in Buster. *Acta Crystallogr D Biol Crystallogr*, 68, 368-80.
- SMITHSON, J. E., CAMPBELL, A., ANDREWS, J. M., MILTON, J. D., PIGOTT, R., *et al.* (1997) Altered Expression of Mucins Throughout the Colon in Ulcerative Colitis. *Gut*, 40, 234-40.
- SNOECK, V., GODDEERIS, B. & COX, E. (2005) The Role of Enterocytes in the Intestinal Barrier Function and Antigen Uptake. *Microbes and Infection*, 7, 997-1004.

- SOMMER, F., ADAM, N., JOHANSSON, M. E. V., XIA, L., HANSSON, G. C., *et al.* (2014) Altered Mucus Glycosylation in Core 1 O-Glycan-Deficient Mice Affects Microbiota Composition and Intestinal Architecture. *PLoS One*, 9, e85254.
- SONNENBURG, J. L., XU, J., LEIP, D. D., CHEN, C. H., WESTOVER, B. P., *et al.* (2005) Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont. *Science*, 307, 1955-1959.
- STEENBERGEN, S. M., JIRIK, J. L. & VIMR, E. R. (2009) Yjhs (Nans) Is Required for Escherichia Coli to Grow on 9-O-Acetylated N-Acetylneuraminic Acid. *Journal of Bacteriology*, 191, 7134-7139.
- SUZUKI, M., SUZUKI, A., YAMAKAWA, T. & MATSUNAGA, E. (1985) Characterization of 2,7-Anhydro-N-Acetylneuraminic Acid in Human Wet Cerumen. *J Biochem*, 97, 509-15.
- TABERMAN, H., PARKKINEN, T. & ROUVINEN, J. (2016) Structural and Functional Features of the Nad(P) Dependent Gfo/Idh/Moca Protein Family Oxidoreductases. *Protein Science*, 25, 778-86.
- TAILFORD, L. E., CROST, E. H., KAVANAUGH, D. & JUGE, N. (2015a) Mucin Glycan Foraging in the Human Gut Microbiome. *Frontiers in genetics*, 6, 81-81.
- TAILFORD, L. E., OWEN, C. D., WALSHAW, J., CROST, E. H., HARDY-GODDARD, J., *et al.* (2015b) Discovery of Intramolecular Trans-Sialidases in Human Gut Microbiota Suggests Novel Mechanisms of Mucosal Adaptation. *Nat Commun*, 6, 7624.
- TANNOCK, G. W., LAWLEY, B., MUNRO, K., GOWRI PATHMANATHAN, S., ZHOU, S. J., *et al.* (2013) Comparison of the Compositions of the Stool Microbiotas of Infants Fed Goat Milk Formula, Cow Milk-Based Formula, or Breast Milk. *Applied and Environmental Microbiology*, 79, 3040-3048.
- TAYLOR, S. L., MCGUCKIN, M. A., WESSELINGH, S. & ROGERS, G. B. (2018) Infection's Sweet Tooth: How Glycans Mediate Infection and Disease Susceptibility. *Trends in Microbiology*, 26, 92-101.
- TEPLYAKOV, A., OBMOLOVA, G., TOEDT, J., GALPERIN, M. Y. & GILLILAND, G. L. (2005) Crystal Structure of the Bacterial Yhch Protein Indicates a Role in Sialic Acid Catabolism. *Journal of Bacteriology*, 187, 5520-5527.
- THE HUMAN MICROBIOME PROJECT, C., HUTTENHOWER, C., GEVERS, D., KNIGHT, R., ABUBUCKER, S., *et al.* (2012) Structure, Function and Diversity of the Healthy Human Microbiome. *Nature*, 486, 207.
- THEODORATOU, E., CAMPBELL, H., VENTHAM, N. T., KOLARICH, D., PUČIĆ-BAKOVIĆ, M., *et al.* (2014) The Role of Glycosylation in Ibd. *Nature Reviews Gastroenterology & Hepatology*, 11, 588.

- THOMAS, G. H. (2016) Sialic Acid Acquisition in Bacteria - One Substrate, Many Transporters. *Biochemical Society Transactions*, 44, 760-765.
- THOMSSON, K. A., HOLMEN-LARSSON, J. M., ANGSTROM, J., JOHANSSON, M. E. V., XIA, L. J., *et al.* (2012) Detailed O-Glycomics of the Muc2 Mucin from Colon of Wild-Type, Core 1-and Core 3-Transferase-Deficient Mice Highlights Differences Compared with Human Muc2. *Glycobiology*, 22, 1128-1139.
- THURSBY, E. & JUGE, N. (2017) Introduction to the Human Gut Microbiota. *Biochemical Journal*, 474, 1823-1836.
- TICKLE, I. J., FLENSBURG, C., KELLER, P., PACIOREK, W., SHARFF, A., & VONRHEIN, C., BRICOGNE, G. (2018) Staraniso. *Cambridge, United kingdom: Global Phasing Ltd.*
- TOGO, A. H., DIOP, A., BITTAR, F., MARANINCHI, M., VALERO, R., *et al.* (2018) Description of *Mediterraneibacter Massiliensis*, Gen. Nov., Sp. Nov., a New Genus Isolated from the Gut Microbiota of an Obese Patient and Reclassification of *Ruminococcus Faecis*, *Ruminococcus Lactaris*, *Ruminococcus Torques*, *Ruminococcus Gnavus* and *Clostridium Glycyrrhizinilyticum* as *Mediterraneibacter Faecis* Comb. Nov., *Mediterraneibacter Lactaris* Comb. Nov., *Mediterraneibacter Torques* Comb. Nov., *Mediterraneibacter Gnavus* Comb. Nov. And *Mediterraneibacter Glycyrrhizinilyticus* Comb. Nov. *Antonie Van Leeuwenhoek*, 111, 2107-2128.
- TOH, M. C. & ALLEN-VERCOE, E. (2015) The Human Gut Microbiota with Reference to Autism Spectrum Disorder: Considering the Whole as More Than a Sum of Its Parts. *Microbial Ecology in Health and Disease*, 26, 26309.
- TRAN, T. H., BOUDRY, C., EVERAERT, N., THEWIS, A., PORTETELLE, D., *et al.* (2016) Adding Mucins to an in Vitro Batch Fermentation Model of the Large Intestine Induces Changes in Microbial Population Isolated from Porcine Feces Depending on the Substrate. *FEMS Microbiology Ecology*, 92.
- TROPINI, C., EARLE, K. A., HUANG, K. C. & SONNENBURG, J. L. (2017) The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host Microbe*, 21, 433-442.
- TSIRTSIKOS, P., FEGEROS, K., BALASKAS, C., KOMINAKIS, A. & MOUNTZOURIS, K. C. (2012) Dietary Probiotic Inclusion Level Modulates Intestinal Mucin Composition and Mucosal Morphology in Broilers. *Poultry Science*, 91, 1860-1868.
- TURNER, J. R. (2009) Intestinal Mucosal Barrier Function in Health and Disease. *Nature Reviews Immunology*, 9, 799-809.

- UBEDA, C. & PAMER, E. G. (2012) Antibiotics, Microbiota, and Immune Defense. *Trends in Immunology*, 33, 459-66.
- VAGIN, A. & TEPLYAKOV, A. (2010) Molecular Replacement with Molrep. *Acta Crystallogr D Biol Crystallogr*, 66, 22-5.
- VAN BEUSEKOM, B., JOOSTEN, K., HEKKELMAN, M. L., JOOSTEN, R. P. & PERRAKIS, A. (2018) Homology-Based Loop Modelling Yields More Complete Crystallographic Protein Structures. *bioRxiv*, 329219.
- VAN NOOD, E., VRIEZE, A., NIEUWDORP, M., FUENTES, S., ZOETENDAL, E. G., *et al.* (2013) Duodenal Infusion of Donor Feces for Recurrent Clostridium Difficile. *New England Journal of Medicine*, 368, 407-415.
- VARKI, A. (2008) Sialic Acids in Human Health and Disease. *Trends in Molecular Medicine*, 14, 351-360.
- VIMR, E. R. (2013) Unified Theory of Bacterial Sialometabolism: How and Why Bacteria Metabolize Host Sialic Acids. *ISRN Microbiol*, 2013, 816713.
- VONRHEIN, C., FLENSBURG, C., KELLER, P., SHARFF, A., SMART, O., *et al.* (2011) Data Processing and Analysis with the Autoproc Toolbox. *Acta Crystallogr D Biol Crystallogr*, 67, 293-302.
- WALKER, A. W., INCE, J., DUNCAN, S. H., WEBSTER, L. M., HOLTROP, G., *et al.* (2011) Dominant and Diet-Responsive Groups of Bacteria within the Human Colonic Microbiota. *Isme j*, 5, 220-30.
- WANG, L., CHRISTOPHERSEN, C. T., SORICH, M. J., GERBER, J. P., ANGLE, M. T., *et al.* (2013) Increased Abundance of Sutterella Spp. And Ruminococcus Torques in Feces of Children with Autism Spectrum Disorder. *Molecular Autism*, 4, 42.
- WANG, S., LI, Y., HAN, Z., CHEN, X., CHEN, Q., *et al.* (2018) Molecular Characterization of a Novel N-Acetylneuraminase Lyase from a Deep-Sea Symbiotic Mycoplasma. *Marine Drugs*, 16.
- WASIK, B. R., BARNARD, K. N. & PARRISH, C. R. (2016) Effects of Sialic Acid Modifications on Virus Binding and Infection. *Trends in Microbiology*, 24, 991-1001.
- WERNER, J. J., KOREN, O., HUGENHOLTZ, P., DESANTIS, T. Z., WALTERS, W. A., *et al.* (2012) Impact of Training Sets on Classification of High-Throughput Bacterial 16s Rrna Gene Surveys. *Isme j*, 6, 94-103.
- WILLIAMS, C. J., HEADD, J. J., MORIARTY, N. W., PRISANT, M. G., VIDEAU, L. L., *et al.* (2018) Molprobit: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Science*, 27, 293-315.

- WINN, M. D., MURSHUDOV, G. N. & PAPIZ, M. Z. (2003) Macromolecular TLS Refinement in Refmac at Moderate Resolutions. *Methods in Enzymology*. Academic Press.
- WOSTMANN, B. S., LARKIN, C., MORIARTY, A. & BRUCKNERKARDOSS, E. (1983) Dietary-Intake, Energy-Metabolism, and Excretory Losses of Adult Male Germfree Wistar Rats. *Laboratory Animal Science*, 33, 46-50.
- XIAO, A., SLACK, T. J., LI, Y., SHI, D., YU, H., *et al.* (2018) Streptococcus Pneumoniae Sialidase Spnanb-Catalyzed One-Pot Multienzyme (Opme) Synthesis of 2,7-Anhydro-Sialic Acids as Selective Sialidase Inhibitors. *The Journal of Organic Chemistry*.
- XU, G., KIEFEL, M. J., WILSON, J. C., ANDREW, P. W., OGGIONI, M. R., *et al.* (2011) Three Streptococcus Pneumoniae Sialidases: Three Different Products. *Journal of the American Chemical Society*, 133, 1718-21.
- YAMAOKA, Y. (2008) Increasing Evidence of the Role of Helicobacter Pylori Saba in the Pathogenesis of Gastrointestinal Disease. *J Infect Dev Ctries*, 2, 174-81.
- YATSUNENKO, T., REY, F. E., MANARY, M. J., TREHAN, I., DOMINGUEZ-BELLO, M. G., *et al.* (2012) Human Gut Microbiome Viewed across Age and Geography. *Nature*, 486, 222-+.
- YOUNG, W., EGERT, M., BASSETT, S. A. & BIBILONI, R. (2015) Detection of Sialic Acid-Utilising Bacteria in a Caecal Community Batch Culture Using Rna-Based Stable Isotope Probing. *Nutrients*, 7, 2109-2124.
- YU, Z.-T., CHEN, C. & NEWBURG, D. S. (2013) Utilization of Major Fucosylated and Sialylated Human Milk Oligosaccharides by Isolated Human Gut Microbes. *Glycobiology*, 23, 1281-1292.
- ZHAO, S., LIU, W., WANG, J., SHI, J., SUN, Y., *et al.* (2017) Akkermansia Muciniphila Improves Metabolic Profiles by Reducing Inflammation in Chow Diet-Fed Mice. *J Mol Endocrinol*, 58, 1-14.
- ZHOU, W., GAY, N. & OH, J. (2018) Reprdb and Pandb: Minimalist Databases with Maximal Microbial Representation. *Microbiome*, 6, 15.
- ZIMMERMANN, M., ZIMMERMANN-KOGADEEVA, M., WEGMANN, R. & GOODMAN, A. L. (2019) Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes. *Nature*.
- ZIVKOVIC, A. M., GERMAN, J. B., LEBRILLA, C. B. & MILLS, D. A. (2011) Human Milk Glycobiome and Its Impact on the Infant Gastrointestinal Microbiota. *Proc Natl Acad Sci U S A*, 108 Suppl 1, 4653-8.

ZOETENDAL, E. G., RAES, J., VAN DEN BOGERT, B., ARUMUGAM, M., BOOIJINK, C. C., *et al.* (2012) The Human Small Intestinal Microbiota Is Driven by Rapid Uptake and Conversion of Simple Carbohydrates. *Isme j*, 6, 1415-26.

Appendix 3: Legend for taxonomic classifications made in Figure 52.

	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;__;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;__;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Faecalibacterium;s__prausnitzii
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Phascolarctobacterium;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;__;_;
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Lachnospira;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira;s__
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Roseburia;s__faecis
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Collinsella;s__aerofaciens
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Ruminococcus;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__producta
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;__;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Ruminococcus;s__gnavus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea;s__longicatena
	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Gemmiger;s__formicilis
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;s__uniformis
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus;s__catus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Butyrivibrio;s__pulliaecorum
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea;s__formicigenans
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Adlercreutzia;s__
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__distasonis
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__adolescentis
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__[Barnesiellaceae];__;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Dorea;s__
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus;s__
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__longum
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Blautia;s__obeum
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__;g__;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Anaerostipes;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus;_
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__;s__
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium;s__celatum
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Clostridium;s__citroniae
	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Eggerthella;s__lenta
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Veillonellaceae;g__Veillonella;_
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae;g__Bifidobacterium;s__bifidum
	k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales;f__Fusobacteriaceae;g__Fusobacterium;s__
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae;g__;s__

k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_[Odoribacteraceae];g_Butyricimonas;s_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Clostridium;s_symbiosum
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_[Mogibacteriaceae];g_;s_
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus;s_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Christensenellaceae;g_;s_
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Clostridium;s_spiroforme
 k_Bacteria;p_Proteobacteria;c_Deltaproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Bilophila;s_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Veillonella;s_dispar
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Veillonella;s_parvula
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pasteurellales;f_Pasteurellaceae;g_Haemophilus;s_painfluenzae
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_[Eubacterium];s_dolichum
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_SMB53;s_
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Porphyrionadaceae;g_Parabacteroides;s_
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Streptococcaceae;g_Streptococcus;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;s_
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_fragilis
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;g_Peptostreptococcus;s_anaerobius
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Morganella;s_morganii
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Clostridium;s_aldense
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Morganella;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;_
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Clostridium;s_amosum
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptococcaceae;g_rc4-4;s_
 k_Bacteria;p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_Collinsella;_
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus;s_casseliflavus
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Turicibacteriales;f_Turicibacteraceae;g_Turicibacter;s_
 k_Bacteria;p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_;s_
 k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;_;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;_;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Dorea;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_[Tissierellaceae];g_Peptoniphilus;s_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Clostridium;s_hathewayi
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Leuconostocaceae;g_Weissella;_
 k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;_;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;g_;s_
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Shigella;_
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_ovatus
 k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;s_breve
 k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Neisseriales;f_Neisseriaceae;g_Eikenella;s_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;_;_
 k_Bacteria;p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_Collinsella;s_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;_
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomonadales;f_Pseudomonadaceae;g_Pseudomonas;s_fragi
 k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;_
 k_Bacteria;p_Firmicutes;_;_;_;_
 k_Bacteria;p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae;g_Eggerthella;s_
 k_Bacteria;_;_;_;_
 k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;_;_;_
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Gemellales;f_Gemellaceae;_;_
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;_;_
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;_

Appendix 4: Legend for taxonomic classifications made in Figure 74.

	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__S24-7
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Ruminococcus;s__gnavus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Enterococcaceae;g__Enterococcus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales
	k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales;f__Coriobacteriaceae;g__Adlercreutzia
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus;s__reuteri
	k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae
	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Allobaculum
	k__Bacteria;p__Deferribacteres;c__Deferribacteres;o__Deferribacterales;f__Deferribacteraceae;g__Mucispirillum;s__schaedleri
	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae;g__Clostridium;s__cocleatum
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Alcaligenaceae;g__Sutterella
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae;g__Coprococcus
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus;s__sciuri
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Lactobacillaceae;g__Lactobacillus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Anaerotruncus
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Rikenellaceae
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Jeotgalicoccus;s__psychrophilus
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae;g__Parabacteroides;s__distasonis
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium
	k__Bacteria;p__Firmicutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae
	k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__Akkermansia;s__muciniphila
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Mogibacteriaceae
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Oscillospira
	k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Corynebacteriaceae;g__Corynebacterium;s__stacti
	k__Bacteria
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
	k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Actinomycetales;f__Yaniellaceae;g__Yaniella
	k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
	k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Dehalobacteriaceae;g__Dehalobacterium
	k__Bacteria;p__Proteobacteria;c__Betaproteobacteria;o__Burkholderiales;f__Oxalobacteraceae;g__Ralstonia

k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Butyricoccus;s_pullicaeorum
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Aerococcaceae
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Candidatus Arthromitus
 k_Bacteria;p_Proteobacteria;c_Deltaproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_C21_c20
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Planococcaceae;g_Sporosarcina
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Aerococcaceae;g_Facklamia
 k_Bacteria;p_TM7;c_TM7-3;o_CW040;f_F16
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia
 k_Bacteria;p_Tenericutes;c_Mollicutes;o_Anaeroplasmatales;f_Anaeroplasmataceae;g_Anaeroplasma
 k_Bacteria;p_Firmicutes
 k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_mitochondria
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae
 k_Bacteria;p_Actinobacteria;c_Coriobacteriia;o_Coriobacteriales;f_Coriobacteriaceae
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Bacillaceae
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Aerococcaceae
 k_Bacteria;p_Firmicutes;c_Bacilli
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Clostridium
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Christensenellaceae
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_ovatus
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Faecalibacterium;s_prausnitzii
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Coprobacillus
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Bacillaceae;g_Anoxybacillus
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Clostridium
 k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Gemmiger;s_formicilis
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Bacillaceae
 k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Corynebacteriaceae;g_Corynebacterium
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Gemellales
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomonadales;f_Moraxellaceae;g_Acinetobacter
 k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae
 k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales
 k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Proteus;
 k_Bacteria;p_Tenericutes;c_Mollicutes;o_RF39
 k_Bacteria;p_Firmicutes;c_Clostridia
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;g_Alistipes
 k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Porphyrimonadaceae;g_Parabacteroides
 k_Bacteria;p_Proteobacteria;c_Deltaproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio
 k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Planococcaceae;g_Staphylococcus;s_saprophyticus
 k_Bacteria;p_Bacteroidetes;c_Sphingobacteriia;o_Sphingobacteriales
 k_Bacteria;p_Bacteroidetes

Appendix 5: Glycosylation pattern of MUC2 from mice in the *S. Typhimurium* challenge study
 The abundance of each glycan is shown for each group (Figure 70) as depicted by the legend.

