

STRUCTURAL AND FUNCTIONAL  
CHARACTERISATION OF MUCUS ADHESION  
PROTEINS OF *LACTOBACILLUS REUTERI*

Sabrina Etzold, M.Sc.

Thesis submitted for the degree of Doctor of Philosophy

To

The University of East Anglia

Institute of Food Research

School of Biology  
(University of East Anglia)

September 2013

© 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

Mucus is the first point of contact between the gut microbiota and the host. Mucus adhesins are thought to be key mediators in the mucus adhesion of commensal *Lactobacillus* species. However, knowledge on the structural or functional basis of adhesin interaction with mucin glycoproteins, the main component of mucus, is limited. This work describes the biochemical and structural properties of two cell-surface proteins from *Lactobacillus reuteri*, the mucus-binding protein (MUB) and the Lar0958 protein, and their mucin binding ability.

MUB from *L. reuteri* ATCC 53608 consists of 14 Mub repeats, six type 1 and eight type 2. Single and tandem Mub repeats were heterologously expressed and purified for structural and functional studies. The three-dimensional structure of the Mub type 1 MubRV was determined by X-ray crystallography and revealed two structural domains, B1 and B2. Furthermore, structural homology between MubRV and fibre-like adhesins of Gram-positive pathogens was identified. Small angle X-ray scattering experiments of single and tandem Mub repeats suggested an elongated structure of MUB in a 'beads on a string' arrangement. Functional studies of recombinant Mub repeats and the full-length native MUB isolated from *Lactobacillus* spent culture media, demonstrated binding to different mucins in vitro. Sugar inhibition experiments and glycan arrays suggested the involvement of sugar recognition in MUB protein binding to mucins.

Lar0958 is a modular protein of six Lar0958 repeats present on the cell-surface of *L. reuteri* DSM 20016<sup>T</sup>. The crystal structure of a single recombinant Lar0958 repeat was solved at 1.5 Å, demonstrating a similar protein fold to Mub repeats. In addition, the Lar0958 repeat shows structural similarity to internalin proteins of the pathogen *Listeria monocytogenes*.

Taken together these results provide new insights into the structural organisation of lactobacilli mucus adhesins and their interaction with mucins, highlighting similarities with Gram-positive adhesins of pathogenic bacteria.

## ACKNOWLEDGEMENTS

I would not have been able to complete the journey of the past four years without the support of numerous people.

First and foremost, I would like to express my deep appreciation to my supervisors: Dr. Nathalie Juge, Dr. Andrew Hemmings and Prof. Rob Field. You taught me well and I thank you for all your guidance, support and advice throughout my Ph.D. Especially to Nathalie for your dedication and the time and effort put into this Ph.D. thesis especially during the writing up process.

I would like to thank my examiners Prof. James Naismith and Dr. Gary Rowley for giving me the opportunity to discuss my work.

I gratefully acknowledge the contribution of my collaborators to this work by providing biological material and technical expertise: Prof. Gunnar Hansson, Prof. Sabine Flitsch, Prof. Lokesh Joshi, Dr. David Bolam, Prof. William Willats and Prof. Michael McGuckin.

Many thanks go to the past and present members of the Juge lab: Donald MacKenzie, Louise Tailford, Emmanuelle Crost, Christine Fuell, Faye Jeffers, Olivia Kober and Elizabeth Thursby, and to my friends of the IFR lunch crowd and 'across the road' for their help and for making the last four years so enjoyable.

To Dagmara, I could not have wished to find a better friend in Norwich, you made it a better place.

To my friends back home, Steffi, Steffi and Tanja, thank you for believing in me, not giving up on me and for being there at the other end of the phone.

Finally, I would like to express my undying gratitude to my parents for their continuing love, deep trust and immeasurable support, it is thanks to you that I arrived where I am now.

*Zum Schluss möchte ich meinen Eltern meine unendliche Dankbarkeit ausdrücken für ihre fortdauernde Liebe, ihr tiefes Vertrauen und ihre grenzenlose Unterstützung, ich habe es euch zu verdanken, dass ich heute hier angekommen.*

# TABLE OF CONTENTS

<b>ABSTRACT</b>	<b>2</b>
<b>ACKNOWLEDGEMENTS</b>	<b>3</b>
<b>TABLE OF CONTENTS</b>	<b>4</b>
<b>LIST OF FIGURES</b>	<b>7</b>
<b>LIST OF TABLES</b>	<b>9</b>
<b>PREFACE</b>	<b>10</b>
<b>ABBREVIATIONS</b>	<b>12</b>
<b>CHAPTER 1 INTRODUCTION</b>	<b>15</b>
<b>1.1 THE GASTROINTESTINAL (GI) TRACT</b>	<b>15</b>
1.1.1 PHYSIOLOGY OF THE MAMMALIAN GI TRACT	15
1.1.2 THE GI MUCUS LAYER	17
<b>1.2 MUCINS IN THE GI TRACT</b>	<b>20</b>
1.2.1 THE MEMBRANE-BOUND MUCIN MUC1/MUC1	23
1.2.2 THE GEL-FORMING MUCIN MUC2/ MUC2	24
1.2.3 MUC2 O-GLYCOSYLATION	26
<b>1.3 MECHANISMS OF HOST-MICROBE INTERACTION IN THE GI TRACT</b>	<b>28</b>
1.3.1 THE GASTROINTESTINAL MICROBIOTA COMPOSITION	28
1.3.2 BACTERIAL INTERACTION AT THE MUCOSAL SURFACE	31
<b>1.4 BACTERIAL ADHESINS</b>	<b>32</b>
1.4.1 MUCUS ADHESINS IN PATHOGENIC BACTERIA	35
1.4.2 MUCUS ADHESINS OF <i>LACTOBACILLUS</i> SPECIES	38
<b>1.5 STRUCTURAL BASIS OF BACTERIAL MUCUS ADHESINS</b>	<b>43</b>
<b>1.6 AIMS AND OBJECTIVES OF THIS RESEARCH PROJECT</b>	<b>48</b>
<b>CHAPTER 2 MATERIAL AND METHODS</b>	<b>49</b>
<b>2.1 MATERIALS</b>	<b>49</b>
2.1.1 STANDARD BUFFERS	49
2.1.2 SUGARS AND MUCINS	49
2.1.3 LECTINS AND ANTIBODIES	50
2.1.4 BACTERIAL STRAINS, MEDIA COMPOSITION AND CULTURE CONDITIONS	51
<b>2.2 MOLECULAR BIOLOGY</b>	<b>53</b>
2.2.1 POLYMERASE CHAIN REACTION (PCR)	53
2.2.2 PLASMID DNA PURIFICATION	53
2.2.3 DNA AGAROSE GEL ELECTROPHORESIS	54
2.2.4 DNA CLONING IN EXPRESSION VECTORS	54
2.2.5 BACTERIAL TRANSFORMATION	55
2.2.6 RECOMBINANT DNA ANALYSIS	55
2.2.7 ESTIMATION OF DNA CONCENTRATION	56
<b>2.3 BIOCHEMISTRY</b>	<b>56</b>

2.3.1	RECOMBINANT PROTEIN PRODUCTION	56
2.3.1.1	Protein expression	56
2.3.1.2	Protein extraction	56
2.3.1.3	Protein purification	57
2.3.2	MUB PURIFICATION	58
2.3.3	ESTIMATION OF PROTEIN CONCENTRATION	58
2.3.4	PROTEIN GEL ELECTROPHORESIS	59
2.3.4.1	Polyacrylamide gel electrophoresis (PAGE)	59
2.3.4.2	Agarose-polyacrylamide composite gel electrophoresis (AgPAGE)	59
2.3.4.3	Isoelectric focusing (IEF)	60
2.3.4.4	Staining of gels	60
2.3.4.5	Western blotting	61
2.3.5	PROTEIN DETECTION VIA ANTIBODIES AND LECTINS	61
2.3.6	PROTEIN BINDING ASSAYS	62
2.3.6.1	Membrane protein binding assays after electrophoresis	62
2.3.6.2	Glycan array	63
2.3.6.3	Mass Spectrometry glycan array	63
2.3.6.3.1	Functionalisation of Au-surfaces	63
2.3.6.3.2	Protein binding on functionalised Au-glycan chips	64
2.3.6.4	Slot-blot assay	64
2.3.6.5	Microtitre plate assays	65
<b>2.4</b>	<b>BIOPHYSICS</b>	<b>66</b>
2.4.1	ISOTHERMAL TITRATION CALORIMETRY (ITC)	66
2.4.2	CIRCULAR DICHROISM (CD)	66
2.4.3	ANALYTICAL ULTRACENTRIFUGATION (AUC)	67
2.4.4	DYNAMIC LIGHT SCATTERING (DLS)	67
<b>2.5</b>	<b>STRUCTURAL BIOLOGY</b>	<b>68</b>
2.5.1	X-RAY CRYSTALLOGRAPHY	68
2.5.1.1	Protein crystallisation	68
2.5.1.2	Crystal soaking and co-crystallisation studies	69
2.5.1.3	X-ray data set analysis	69
2.5.2	SMALL ANGLE X-RAY SCATTERING (SAXS) STUDIES	70
2.5.2.1	SAXS data collection	70
2.5.2.2	SAXS data analysis	71
<b>2.6</b>	<b>BIOINFORMATICS</b>	<b>71</b>
2.6.1	GENERAL DATA ANALYSIS	71
2.6.2	SEQUENCE ANALYSIS	72
2.6.3	PROTEIN STRUCTURE ANALYSIS	72
2.6.4	CRYSTAL DIFFRACTION AND SAXS DATA ANALYSIS	72
 <b>CHAPTER 3 PURIFICATION AND BIOCHEMICAL CHARACTERISATION OF MUB PROTEINS AND NATIVE MUB</b>		 <b>73</b>
<b>3.1</b>	<b>CLONING, HETEROLOGOUS EXPRESSION AND PURIFICATION OF MUB PROTEINS</b>	<b>74</b>
3.1.1	MUB TYPE 1 AND 2 REPEAT PROTEINS	74
3.1.2	HIS <sub>6</sub> -TAGGED N-TERMINAL MUB DOMAIN CONSTRUCTS	86
<b>3.2</b>	<b>STRUCTURAL AND BIOPHYSICAL PROPERTIES OF MUB PROTEINS</b>	<b>89</b>
<b>3.3</b>	<b>PURIFICATION AND CHARACTERISATION OF NATIVE MUB</b>	<b>93</b>
3.3.1	MUB PURIFICATION	93
3.3.2	MUB CHARACTERISATION	95
3.3.2.1	MUB characterisation by native PAGE and AUC	96
3.3.2.2	MUB detection by lectins	98
<b>3.4</b>	<b>DISCUSSION</b>	<b>101</b>

<b>CHAPTER 4 STRUCTURAL CHARACTERISATION OF MUB AND MUB REPEATS</b>	<b>103</b>
<b>4.1 MUBR5 CRYSTAL SOAKING AND CO-CRYSTALLISATION STUDIES</b>	<b>103</b>
<b>4.2 STRUCTURE DETERMINATION OF A MUB TYPE 1 REPEAT PROTEIN</b>	<b>106</b>
4.2.1 CRYSTALLISATION EXPERIMENTS OF MUBRI	107
4.2.2 STRUCTURE DETERMINATION OF MUBRV	108
4.2.3 FUNCTIONAL ANNOTATION ANALYSIS OF MUBRV AND MUBR5	114
<b>4.3 SOLUTION STRUCTURE DETERMINATION AND DOMAIN ORGANISATION OF MUB</b>	<b>115</b>
<b>4.4 DISCUSSION</b>	<b>121</b>
<b>CHAPTER 5 FUNCTIONAL CHARACTERISATION OF MUB ADHESION PROPERTIES TO MUCUS</b>	<b>126</b>
<b>5.1 BINDING OF MUB PROTEINS TO MUCUS AND MUCIN</b>	<b>126</b>
5.1.1 MUB PROTEIN BINDING TO PORCINE AND HUMAN MUCIN IN MEMBRANE-ADHESION ASSAYS	126
5.1.2 ASSESSMENT OF MUB AND MUB REPEAT BINDING TO MUCIN GLYCANS	131
5.1.3 MUB BINDING STUDIES USING MUCIN AND NEOGLYCOCONJUGATE ARRAYS	138
<b>5.2 BINDING OF MUB PROTEINS TO GLYCANS</b>	<b>143</b>
5.2.1 ASSESSMENT OF MUBR5 BINDING TO MANNOSE LIGANDS USING ITC	143
5.2.2 SCREENING OF MONO- AND OLIGOSACCHARIDES FOR MUB PROTEIN BINDING BY GLYCAN ARRAYS	145
<b>5.3 DISCUSSION</b>	<b>150</b>
<b>CHAPTER 6 STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF LAR0958</b>	<b>156</b>
<b>6.1 CLONING, EXPRESSION, PURIFICATION AND BIOPHYSICAL CHARACTERISATION OF LAR0958</b>	<b>156</b>
<b>6.2 STRUCTURE DETERMINATION OF A SINGLE LAR0958 REPEAT</b>	<b>159</b>
<b>6.3 EXPRESSION AND PURIFICATION OF INTERNALIN PROTEINS</b>	<b>165</b>
<b>6.4 ASSESSMENT OF LAR0958 AND INTERNALIN BINDING TO MUCIN</b>	<b>167</b>
<b>6.5 DISCUSSION</b>	<b>171</b>
<b>CHAPTER 7 GENERAL DISCUSSION AND FUTURE PROSPECTIVES</b>	<b>174</b>
<b>APPENDIX</b>	<b>178</b>
<b>I. LDMII MEDIA COMPOSITION</b>	<b>178</b>
<b>II. PRIMERS FOR AMPLIFICATION AND SEQUENCING</b>	<b>179</b>
<b>III. IN-FUSION CLONING SYSTEM</b>	<b>181</b>
<b>IV. ALIGNMENT OF MUB REPEAT SEQUENCES USING CLUSTAWL</b>	<b>182</b>
<b>V. CRYSTAL STRUCTURE SCREEN CONDITIONS</b>	<b>184</b>
<b>VI. AUC SEDIMENTATION VELOCITY PROFILE OF MUB IN PBS</b>	<b>190</b>
<b>VII. MUBR5 CRYSTAL SOAKING AND CO-CRYSTALLISATION DATA SETS</b>	<b>191</b>
<b>VIII. MUBRI DATA PROCESSING PARAMETERS</b>	ERROR! BOOKMARK NOT DEFINED.
<b>IX. SAXS DATA SET STATISTICS</b>	<b>193</b>
<b>X. KRATKY ANALYSIS OF SAXS SCATTERING DATA</b>	<b>194</b>
<b>XI. COMPARISON OF EXPERIMENTAL AND CALCULATED SAXS SCATTERING DATA OF MUBRV AND MUBR5</b>	<b>195</b>
<b>XII. SCREENING OF GLYCOCONJUGATES FOR MUB BINDING</b>	<b>196</b>
<b>XIII. SCREENING OF THE MAMMALIAN GLYCAN ARRAY (CFG) FOR MUB BINDING</b>	<b>197</b>
<b>XIV. MUB BINDING TO PLANT DERIVED GLYCANS</b>	<b>221</b>
<b>LIST OF REFERENCES</b>	<b>222</b>

## LIST OF FIGURES

FIGURE 1.1 SCHEMATIC REPRESENTATION OF THE HUMAN (MAMMALIAN) GUT WITH ITS SUBDIVISIONS (TAKEN AND ADAPTED FROM WWW.MEDTRONIC-GASTRO-URO.COM.AU) .....	15
FIGURE 1.2 SCHEMATIC REPRESENTATION OF THE HUMAN COLONIC MUCOSA .....	16
FIGURE 1.3 REPRESENTATION OF MUCUS THICKNESS AND ORGANISATION IN RATS .....	17
FIGURE 1.4 DEFENCE MECHANISMS AT THE GI MUCOSAL SURFACE .....	19
FIGURE 1.5 SCHEMATIC REPRESENTATION OF THE DOMAIN ARCHITECTURE OF SECRETED AND MEMBRANE-BOUND MUCINS (TAKEN FROM [51]).....	20
FIGURE 1.6 MUCIN TYPE O-GLYCAN CORE STRUCTURES .....	22
FIGURE 1.7 COMMON CORE BUILDING BLOCK OF N-GLYCANS.....	23
FIGURE 1.8 SCHEMATIC REPRESENTATION OF THE MEMBRANE-BOUND MUCIN MUC1 .....	24
FIGURE 1.9 SCHEMATIC REPRESENTATION OF THE HUMAN MUC2 DOMAIN ORGANISATION .....	25
FIGURE 1.10 REPRESENTATION OF THE MUC2 NET FORMATION.....	25
FIGURE 1.11 PHYLOGENETIC TREE OF GI MICROBIOTA .....	29
FIGURE 1.12 PHYLOGENETIC ANALYSIS OF <i>L. REUTERI</i> ISOLATED FROM DIFFERENT VERTEBRATE HOSTS .....	30
FIGURE 1.13 SCHEMATIC REPRESENTATION OF BACTERIAL-SURFACE INTERACTIONS .....	33
FIGURE 1.14 CELL WALL ARCHITECTURE AND SURFACE MOLECULES OF GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA .....	34
FIGURE 1.15 SCHEMATIC REPRESENTATION OF MUB OF <i>L. REUTERI</i> 53608.....	40
FIGURE 1.16 PRESENTATION OF THE MUB TYPE 2 MUBR5 X-RAY CRYSTAL STRUCTURE .....	43
FIGURE 1.17 MUB, MUCBP OR RELATED CRYSTAL STRUCTURES OF BACTERIAL CELL-SURFACE PROTEINS.....	44
FIGURE 1.18 PRESENTATION OF THE CRYSTAL STRUCTURES OF THE <i>E. COLI</i> ADHESINS F17-G AND FIMH .....	45
FIGURE 1.19 EXAMPLES OF INTERNALINS WITH IG-LIKE FOLD DOMAINS .....	46
FIGURE 3.1 SCHEMATIC REPRESENTATION OF CELL-SURFACE ANCHORED MUB AND NEIGHBOUR JOINING TREE FOR MUB REPEAT SEQUENCES .....	73
FIGURE 3.2 SCHEMATIC REPRESENTATION OF MUB WITH RECOMBINANT MUB PROTEINS.....	74
FIGURE 3.3 PURIFICATION OF MUBR5 BY IEC AND SEC .....	76
FIGURE 3.4 PURIFICATION OF MUBRI BY IEC AND SEC .....	78
FIGURE 3.5 PURIFICATION OF MUBRV BY IEC AND SEC .....	79
FIGURE 3.6 PURIFICATION OF MUBR8-V BY IEC AND SEC.....	81
FIGURE 3.7 PURIFICATION OF MUBRV-VI BY IEC AND SEC .....	83
FIGURE 3.8 PURIFICATION OF MUBRI-II-III BY IEC AND SEC .....	85
FIGURE 3.9 PURIFICATION OF NTERM BY IMAC AND SEC .....	87
FIGURE 3.10 PURIFICATION OF NTERMUBRI BY IMAC AND SEC.....	88
FIGURE 3.11 SECONDARY STRUCTURE DETERMINATION OF MUB PROTEINS BY CIRCULAR DICHROISM (CD).....	90
FIGURE 3.12 PI DETERMINATION OF MUB REPEATS BY IEF.....	92
FIGURE 3.13 PURIFICATION OF NATIVE MUB BY SEC.....	94
FIGURE 3.14 SDS-PAGE OF MUB BEFORE TRYPSIN DIGEST AND MS ANALYSIS .....	95
FIGURE 3.15 MUB ANALYSIS BY NATIVE PAGE.....	96
FIGURE 3.16 SEDIMENTATION EQUILIBRIUM DATA FITS OF MUB .....	97
FIGURE 3.17 DETECTION OF MUB (PURIFICATION BATCH 1) VIA LECTINS.....	99
FIGURE 3.18 DETECTION OF MUB (PURIFICATION BATCH 2) BY LECTINS.....	100
FIGURE 4.1 MUBR5 CRYSTAL SOAKED WITH MAN.....	104
FIGURE 4.2 MUBR5 STRUCTURE WITH MANNOSE (MAN) .....	106
FIGURE 4.3 MUBRI CRYSTALS IN HARVESTING LOOP .....	107
FIGURE 4.4 MUBRV CRYSTAL .....	109
FIGURE 4.5 X-RAY CRYSTAL STRUCTURE OF THE MUB TYPE 1 REPEAT MUBRV .....	112
FIGURE 4.6 OVERLAY OF THE B2 DOMAIN OF MUBRV WITH THE N2 DOMAIN OF GBS52 PILIN .....	113
FIGURE 4.7 PREDUS AND SCREEN ANALYSIS OF MUBR5 AND MUBRV .....	115

FIGURE 4.8 SAXS DATA ANALYSIS AND SHAPE RECONSTRUCTION OF SINGLE AND DOUBLE MUB REPEATS .....	118
FIGURE 4.9 SAXS DATA ANALYSIS AND SHAPE RECONSTRUCTION OF MUBRI-II-III AND THE N-TERMINAL DOMAIN OF MUB .....	120
FIGURE 4.10 SCHEMATIC REPRESENTATION OF MUB ON THE BACTERIAL CELL SURFACE .....	124
FIGURE 5.1 BINDING OF MUB PROTEINS TO PPSM AND PSIM IN A SLOT-BLOT ASSAY .....	128
FIGURE 5.2 BINDING OF MUBR5 AND MUBRI-II-III TO COLONIC MUC2 FROM HUMAN BIOPSY SAMPLES .....	130
FIGURE 5.3 BINDING OF MUB TO PORCINE AND HUMAN MUCIN.....	132
FIGURE 5.4 BINDING OF MUB, MUBRI-II-III AND MUBR5 TO PORCINE MUCIN.....	134
FIGURE 5.5 COMPETITION OF MUB BINDING TO PPGM WITH SOLUBLE SUGARS .....	135
FIGURE 5.6 BINDING OF MUB TO MUC1 GLYCOPEPTIDES.....	137
FIGURE 5.7 MUCIN BINDING OF MUB ASSESSED USING A GI MUCIN ARRAY .....	139
FIGURE 5.8 CONCENTRATION-DEPENDENT BINDING OF MUB TO GI MUCINS .....	140
FIGURE 5.9 MUB BINDING TO GI MUCINS IN THE PRESENCE OF SUGAR MOLECULES .....	142
FIGURE 5.10 BINDING OF MUBR5 TO MANNOSIDES INVESTIGATED BY ITC .....	144
FIGURE 5.11 GLYCAN BINDING ABILITY OF MUBR5, MUBRI, NTERMUBRI AND MUB ASSESSED USING A GLYCAN ARRAY .....	146
FIGURE 5.12 SUGAR BINDING OF MUBRI-II-III ASSESSED USING A SUGAR AU-ARRAY .....	148
FIGURE 6.1 SCHEMATIC REPRESENTATION OF LAR0958.....	156
FIGURE 6.2 LAR0958 PROTEIN PURIFICATION VIA IEC .....	157
FIGURE 6.3 IEF AND CD OF LAR0958 .....	158
FIGURE 6.4 REPRESENTATIONS OF THE LAR0958 REPEAT X-RAY CRYSTAL STRUCTURE .....	162
FIGURE 6.5 STRUCTURAL OVERLAY OF LAR0958 WITH MUBR5 AND MUBRV.....	163
FIGURE 6.6 INLC CRYSTAL STRUCTURE AND OVERLAY WITH LAR0958.....	164
FIGURE 6.7 PURIFICATION OF INLJ AND INLJ-LRR-IR BY IMAC .....	166
FIGURE 6.8 PURIFICATION OF INLJ AND INLJ-LRR-IR BY SEC.....	167
FIGURE 6.9 BINDING OF LAR0958 TO PPSM AND PSIM ASSESSED IN A SLOT BLOT ASSAY .....	168
FIGURE 6.10 DETECTION OF LAR0958, INLJ AND INLJ-LRR-IR VIA ANTIBODIES AND OF MUCIN VIA WGA .....	169
FIGURE 6.11 BINDING OF LAR0958, INLJ AND INLJ-LRR-IR TO MUCIN OR MUCUS .....	170
FIGURE 6.12 DOMAIN ORGANISATION OF MEMBERS OF THE INTERNALIN PROTEIN FAMILY .....	172

## LIST OF TABLES

TABLE 1.1 PREDOMINANT OLIGOSACCHARIDE STRUCTURES IN HUMAN MUC2 .....	26
TABLE 1.2 ADHESINS AND POTENTIAL HOST RECEPTORS OF SELECTED PATHOGENS .....	35
TABLE 1.3 COMMENSAL ADHESINS OF <i>LACTOBACILLUS</i> INVOLVED IN HOST SURFACE INTERACTION ....	39
TABLE 2.1 USED PRIMARY AND SECONDARY ANTIBODIES .....	51
TABLE 3.1 PURIFICATION CHARACTERISTICS OF MUB PROTEINS.....	89
TABLE 3.2 PROPORTION OF SECONDARY STRUCTURAL ELEMENTS IN MUB REPEATS BY CD (AND AFTER X-RAY CRYSTAL STRUCTURE DETERMINATION) <sup>A</sup> .....	91
TABLE 3.3 SUGAR SPECIFICITY OF USED LECTINS .....	99
TABLE 4.1 MUBR5 (MAN) DATA COLLECTION AND REFINEMENT PARAMETERS .....	105
TABLE 4.2 DATA COLLECTION AND REFINEMENT STATISTICS FOR MUBRV.....	110
TABLE 5.1 GLYCOCONJUGATES RECOGNISED BY MUB.....	141
TABLE 6.1 BROMIDE SITES WITH FIVE HIGHEST OCCUPANCIES FOUND BY AUTOSOL .....	160
TABLE 6.2 DATA COLLECTION AND REFINEMENT STATISTICS OF LAR0958 DIFFRACTION DATA SETS..	161

## PREFACE

The work presented in this Ph.D. thesis has resulted in the following publications:

### Journal articles

**Etzold S.**, Kober O.I., MacKenzie D.A., Tailford L.E., Gunning P., Hemmings A.M., Juge N. (2013). Structural basis for adaptation of lactobacilli to gastrointestinal mucus. *Environmental Microbiology*, *accepted*

### Abstracts

**Etzold S.**, MacKenzie D.A., Tailford L.E., Field R.A., Hemmings A.M., Juge N. (2011). Structure determination of bacterial mucus-binding proteins and their functional role in adhesion to host glycans. *Glycoconjugate Journal*, *28 (5): 241-42*

**Etzold S.**, MacKenzie D.A., Tailford L.E., Field R.A., Hemmings A.M., Juge N. (2012). Bacterial mucus-binding proteins: Structural and functional insights into their role in the adhesion to host glycans. *Glycobiology*, *22 (11): 1633*

### Oral presentations

**Etzold S.**, MacKenzie D.A., Gunning P., Hemmings A.M., Juge N. (2012). Structural organisation of *Lactobacillus* adhesins determined by X-ray crystallography and small angle X-ray scattering (SAXS). *EEMaX (East of England Macromolecular X-ray crystallography) symposium, Norwich, UK*

**Etzold S.**, MacKenzie D.A., Gunning P., Hemmings A. M., Juge N. (2012). Structure determination of *Lactobacillus* adhesins and their functional role in adhesion to mucin. *IFR Student science showcase, Norwich, UK (awarded best oral presentation)*

**Etzold S.**, MacKenzie D.A., Field R.A., Hemmings A.M., Juge N. (2011). Structure determination of bacterial mucus-binding proteins and their functional role in adhesion to host glycans. *GLYCO XXI, 21st International Symposium on Glycoconjugates, Vienna, Austria*

## Poster presentations

**Etzold S.**, MacKenzie D.A., Tailford L.E., Gunning P., Hemmings A.M., Juge N. (2013). Bacterial mucus-binding proteins: Structural and functional insights into their role in the adhesion to mucin glycans. *12<sup>th</sup> Mucins in Health and Disease conference, Cambridge, UK*

**Etzold S.**, MacKenzie D.A., Hemmings A.M., Juge N. (2013) Bacterial mucus-binding proteins - structural and functional insights into their role in the adhesion to host glycans. *Annual Gut Health and Food Safety ISP meeting, John Innes Conference Centre, Norwich, UK*

**Etzold S.**, MacKenzie D.A., Tailford L.E., Hemmings A.M., Juge N. (2012). Bacterial mucus-binding proteins: Structural and functional insights into their role in the adhesion to host glycans. *Joint meeting of the Society of Glycobiology and American Society for Matrix Biology, San Diego, USA*

**Etzold S.**, MacKenzie D.A., Hemmings A.M., Juge N. (2011). Structure determination of bacterial mucus-binding proteins and their functional role in adhesion to the host. *11<sup>th</sup> Mucins in Health and Disease conference, Cambridge, UK*

**Etzold S.**, MacKenzie D.A., Tailford L.E., Field R. A., Hemmings A.M., Juge N. (2011). Structure determination of bacterial mucus-binding proteins and their functional role in adhesion to host glycans. *1<sup>st</sup> International conference on Glycobiology of Human Milk Oligosaccharides, Copenhagen, Denmark*

MacKenzie D.A., **Etzold S.**, Fais M., Tailford L.E., Fairhurst S., Field R.A., Hemmings A.M., Juge N. (2010). Structure and function of mucus-binding proteins in bacterial adhesion to the host. *BBSRC-INRA-WUR Workshop on Food and Gut Health IFR, Norwich, UK*

MacKenzie D.A., **Etzold S.**, Fais M., Tailford L.E., Fairhurst S., Field R.A., Hemmings A.M., Juge N. (2010). Structure and function of mucus-binding proteins in bacterial adhesion to the host. *Structural glycobiology and human health - Biochemical Society Meeting, Royal Holloway, University of London, London, UK*

## ABBREVIATIONS

<b>A<sub>280</sub></b>	Absorbance measured at 280 nm
<b>Abs</b>	Absorbance
<b>AgPAGE</b>	Agarose polyacrylamide gel electrophoresis
<b>AP</b>	Alkaline phosphatase
<b>APS</b>	Ammonium persulphate
<b>AUC</b>	Analytical ultracentrifugation
<b>ASU</b>	Asymmetric unit
<b>BLI</b>	Bio-layer interferometry
<b>BSA</b>	Bovine serum albumin
<b>CD</b>	Circular dichroism
<b>CD</b>	Crohn's disease
<b>CK</b>	Cysteine knot
<b>DLS</b>	Dynamic light scattering
<b>DTT</b>	Dithiothreitol
<b>ECM</b>	Extracellular matrix
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>EF-Tu</b>	Elongation factor Tu
<b>EPS</b>	Exopolysaccharide
<b>FITC</b>	Fluorescein isothiocyanate
<b>FPLC</b>	Fast protein liquid chromatography
<b>Fuc</b>	Fucose
<b>GAG</b>	Glycosaminoglycans
<b>Gal</b>	Galactose
<b>GalNAc</b>	<i>N</i> -acetylgalactosamine
<b>Glc</b>	Glucose
<b>GlcNAc</b>	<i>N</i> -acetylglucosamine
<b>GI</b>	Gastrointestinal
<b>IBD</b>	Inflammatory bowel disease
<b>IEF</b>	Isoelectric focusing
<b>Ig</b>	Immunoglobulin
<b>IMAC</b>	Immobilised metal ion affinity chromatography
<b>IPTG</b>	Isopropyl- $\beta$ -D-1-thiogalactopyranoside
<b>ITC</b>	Isothermal titration calorimetry
<b>Lac</b>	Lactose
<b>LacNAc</b>	<i>N</i> -acetylglucosamine
<b>LPS</b>	Lipopolysaccharides
<b>LTA</b>	Lipoteichoic acid
<b>Man</b>	Mannose

<b>MME</b>	Monomethylether
<b>MES</b>	2-( <i>N</i> -morpholino)ethanesulfonic acid
<b>MR</b>	Molecular replacement
<b>MS</b>	Mass spectrometry
<b>Msa</b>	Mannose specific adhesin
<b>MSCRAMM</b>	Microbial surface components recognising adhesive matrix molecules
<b>Mub</b>	Mucus binding
<b>MUC</b>	Human mucin
<b>muc</b>	Non-human mucin
<b>MW</b>	Molecular weight
<b>MWCO</b>	Molecular weight cut off
<b>NeuAc</b>	<i>N</i> -acetylneuraminic acid
<b>O/N</b>	Over night
<b>PBS</b>	Phosphate buffered saline
<b>PBST</b>	Phosphate buffered saline with Tween-20
<b>PCR</b>	Polymerase chain reaction
<b>PEG</b>	Polyethylene glycol
<b>PGM</b>	Porcine gastric mucin
<b>pPGM</b>	Purified porcine gastric mucin
<b>pl</b>	Isoelectric point
<b>PSIM</b>	Porcine small intestinal mucus
<b>PTS domain</b>	Proline/ threonine/ serine rich domain
<b>PVDF</b>	Polyvinylidene fluoride
<b>RCA</b>	<i>Ricinus communis</i> agglutinin
<b>RT</b>	Room temperature
<b>SAXS</b>	Small angle X-ray scattering
<b>SEA</b>	sea-urchin sperm protein, enterokinase and agrin
<b>SDP</b>	Sortase-dependant protein
<b>SDS</b>	Sodium dodecylsulphate
<b>SDS-PAGE</b>	Sodiumdodecylsulphate-polyacrylamide gel
<b>SEC</b>	Size exclusion chromatography
<b>SI</b>	Small intestine
<b>SNA</b>	<i>Sambus nigra</i> agglutinin
<b>SOC</b>	Super optimal broth with catabolite repression
<b>SPR</b>	Surface plasmon resonance
<b>TAE</b>	Tris-acetate-EDTA
<b>TCA</b>	Trichloroacetic acid
<b>TE</b>	Tris-EDTA
<b>TEMED</b>	N,N,N,N'-tetramethylethylenediamine
<b>TMB</b>	3,3',5,5'-Tetramethylbenzidine

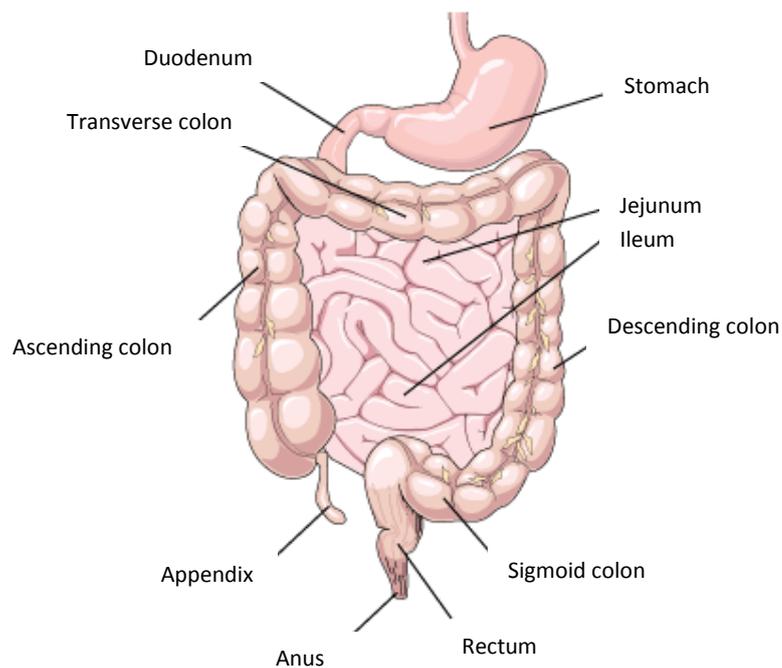
<b>Tris</b>	2-Amino-2-hydroxymenthyl-propane-1,3-diol
<b>UEA</b>	<i>Ulex europaeus</i> agglutinin
<b>UC</b>	Ulcerative colitis
<b>UV</b>	Ultra-violet
<b>VNTR</b>	Variable number of tandem repeats
<b>WGA</b>	Wheat germ agglutinin
<b>WHO</b>	World Health Organisation

# CHAPTER 1 INTRODUCTION

## 1.1 The gastrointestinal (GI) tract

### 1.1.1 Physiology of the mammalian GI tract

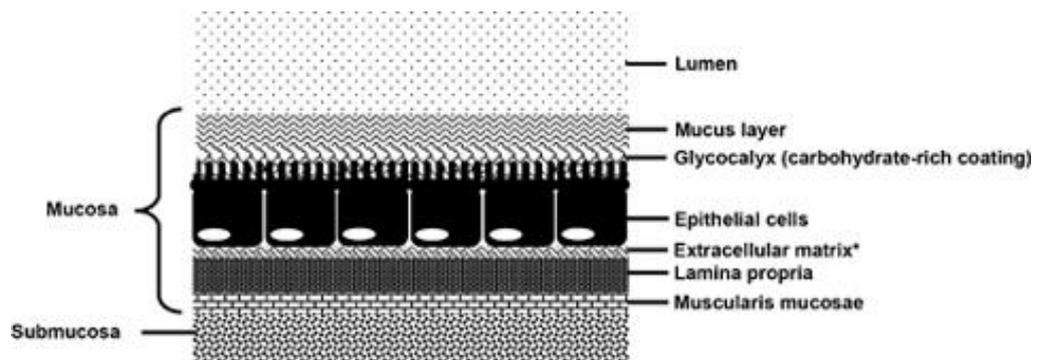
The gastrointestinal (GI) tract is a large organ with a surface area of more than 300 cm<sup>2</sup> and 7 metres long in human adults [1]. Its main function is the supply of nutrients and energy via conversion and adsorption of food components and water. The mammalian GI tract can generally be divided into three major compartments: stomach, small intestine and large intestine. The small intestine is comprised of duodenum, jejunum and ileum, while the large intestine can be further subdivided into appendix, cecum, colon (ascending, transverse, descending and sigmoid) and rectum (Figure 1.1).



**Figure 1.1 Schematic representation of the human (mammalian) gut with its subdivisions**  
(taken and adapted from [www.medtronic-gastro-uro.com.au](http://www.medtronic-gastro-uro.com.au))

The esophagus, stomach and parts of the duodenum are anatomically known as the foregut. The midgut reaches from the duodenum to the transverse colon and the hindgut is comprised of descending colon and rectum [2].

The distinct sections of the GI tract differ with regard to physiological environment and structural composition offering biological niches for various microbial ecosystems. For example, a difference in pH can be observed along the gut, where the pH rises significantly from a highly acidic pH of about 1.0 to 2.5 in the stomach to pH 6.6 in the proximal small intestine (jejunum) reaching a neutral pH of 7.3 to 7.6 in the distal small intestine (ileum) [3-4]. The mean pH in the colon is pH 6.3. Additionally, transient times vary between different parts of the GI tract due to peristaltic movements from 1 h in the stomach, about 10 h in the small intestine and up to 17 h in the colon [4-6]. Furthermore, the microbial biomass differs between distinct regions of the GI tract with the lowest bacterial cell density ( $10^{2-3}$  cells/mL) in the stomach. A significant increase in bacterial cell density can be observed in the jejunum and ileum with  $10^{4-5}$  and  $10^8$  cells/mL, while the bacterial density is highest in the colon with  $10^{10}$  cells/mL [7].



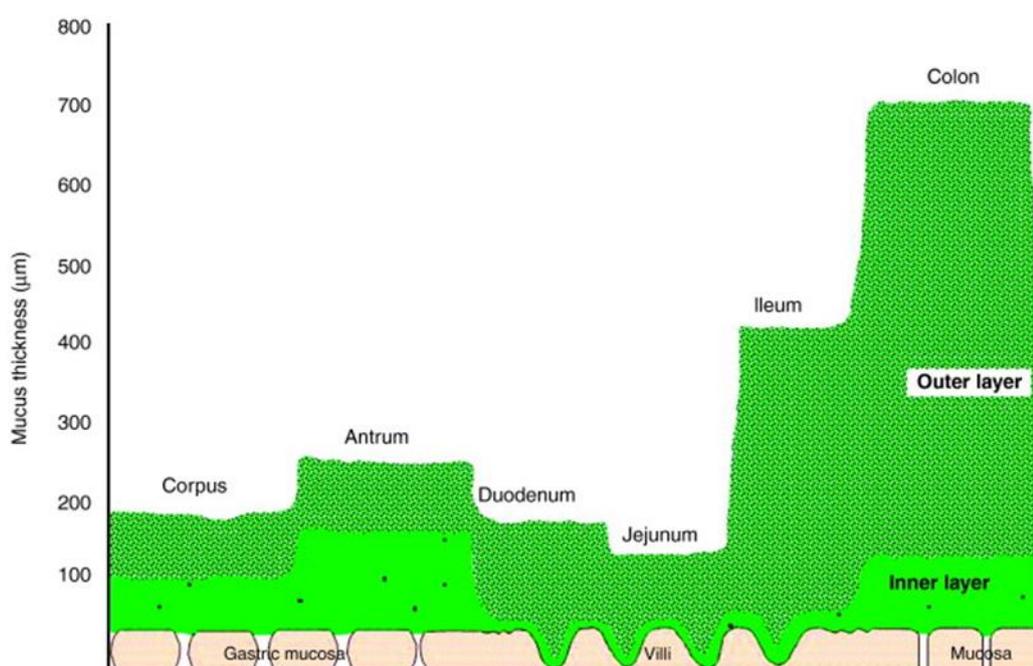
**Figure 1.2 Schematic representation of the human colonic mucosa**

The luminal exposed surface of the gut, the mucosa, consists of muscularis mucosae, lamina propria, extracellular matrix (ECM), epithelium, glycocalyx and the mucus layer (taken and adapted from [8]).

The luminal surface of the GI tract is called the mucosa, which forms a physical barrier between the luminal content and the host, prevents microbial invasion and maintains a healthy state (Figure 1.2). Its most important structural elements are the extracellular matrix (ECM), the epithelial cell layer, the glycocalyx and the mucus layer [8].

### 1.1.2 The GI mucus layer

The mucus layer forms a dynamic, viscous and physical barrier, which separates the underlying GI epithelium from the luminal content, i.e. bacteria and food components, serves as a lubricant to aid passage, and provides a habitat for the diverse microbial community termed the gut microbiota [9]. The intestinal epithelium is comprised of 4 different epithelial cell lineages, enterocytes, enteroendocrine cells, Paneth cells, and goblet cells. They are derived from intestinal epithelial stem cells located at the base of epithelial invaginations named crypts [10]. Goblet cells are the main source of mucus production and release with fast renewal rates of several hours as demonstrated for colonic mucus by *in vivo* labelling [11-12]. In humans up to 10 L of small intestinal and colonic mucus are secreted per day [12]. Continuous secretion is necessary due to microbial enzymatic degradation and mechanical shear forces [13-15]. The mucosal release is a constitutive exocytose process and enhanced mucus secretion is induced by certain secretagogues, i.e. physical disruption or intestinal microbes [16-21].



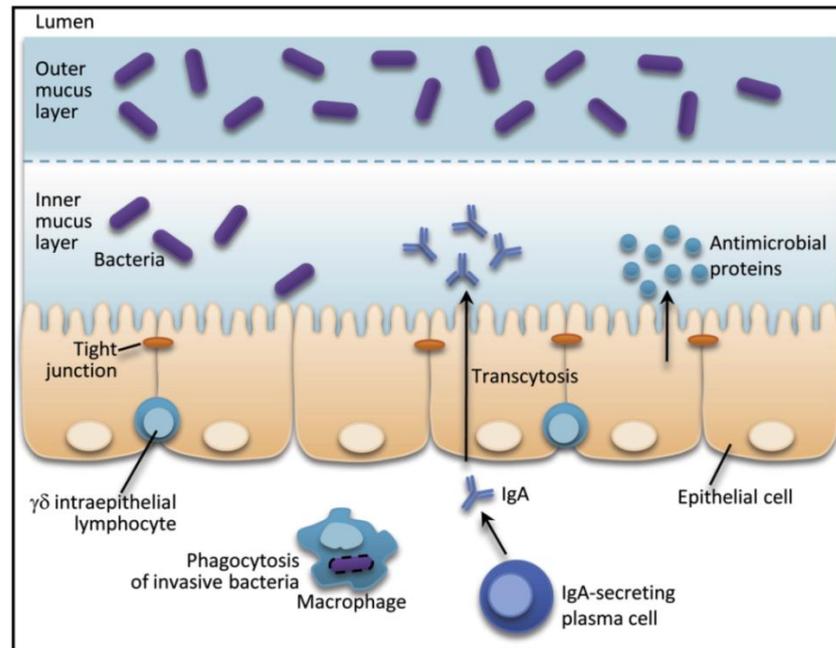
**Figure 1.3 Representation of mucus thickness and organisation in rats**  
(taken and adapted from [22])

The mucus layer covers the epithelial surface from the stomach to the colon varying in thickness, composition and biophysical properties (Figure 1.3) [23-24]. In the stomach

and the colon, mucus is composed of two structurally distinct layers, whereas latest studies identified only a single layer in the small intestine [24]. The outer luminal layer of gastric and colonic mucus is loosely adherent, can easily be removed and is the habitat for the commensal gut flora especially in the colon [25]. In contrast, the underlying colonic mucus layer is firmly attached, of stratified, ordered appearance and mainly devoid of bacteria [25]. However, the inner adherent mucus layer in the stomach and the single small intestinal mucus layer are likely to be penetrable by bacteria [24]. The relative contribution of each layer to the overall thickness of the mucus gel, which is highest in the colon, varies throughout the GI tract (Figure 1.3). Mucus measurements in rat, mice and human showed an increased thickness of the loosely adherent outer mucus layer of 7-, 3- and 2 to 3-fold, respectively, in comparison to the firm mucus blanket in direct contact to the underlying epithelial cells [23, 25-26]. In total the colonic mucus extends about 800  $\mu\text{m}$  in rat, 150  $\mu\text{m}$  in mice and 134  $\mu\text{m}$  in human above the epithelial surface.

The main structural component of the colonic loose and firm mucus, as revealed via a proteomic approach, is the secreted-gel forming mucin 2 (Muc2, mouse or MUC2, human), a highly glycosylated protein of the secreted mucin protein family (see 1.2) [27]. These observations suggest the formation of the loose outer mucus layer from the firm inner presumably via proteolytic cleavage [24, 28]. Due to its intrinsic domain properties, Muc2/ MUC2 is able to form net-like polymers making up a hydrated mucus mesh with certain pore-size. It functions as a physical filter for the GI bacteria keeping them away from the mucosal epithelium (see 1.2.1) (Figure 1.4) [29]. The absence of the Muc2 mucus layer allows bacteria to come into direct contact with the epithelial cell surface causing severe inflammation and eventually colon carcinoma in Muc2<sup>-/-</sup> mice [25, 30].

Besides Muc2, the mucus layer contains a high number of other proteins forming a fairly consistent core proteome in different sections of the murine GI tract from stomach to colon [28]. One example is the Fc-gamma binding-protein, which is strongly associated with colonic Muc2. It was found to bind immunoglobulin (Ig) G but may have additional roles as a Muc2 cross-linker [27, 31].



**Figure 1.4 Defence mechanisms at the GI mucosal surface**

Direct epithelial contact and penetration of gut microbes prevented by the mucus layer, acting as a physical barrier, antimicrobial proteins secreted by Paneth cells and secretory IgA produced by plasma cells (taken from [32]).

Furthermore, the mucus gel contains high levels of secretory IgA, whose production is stimulated by the intestinal microbiota (Figure 1.4) [33-34]. IgA is produced by plasma cells and dimeric IgA is transported across the epithelium via a polymeric-immunoglobulin receptor (pIgR) on columnar epithelial cells [33, 35]. Secreted IgA mediates the neutralisation of viruses and bacterial toxins, and prevents mucus penetration by GI microbes [36-37]. Additionally, large amounts of antimicrobial proteins, such as  $\alpha$ -defensins, lysozyme and RegIII $\gamma$ , are secreted into the mucus layer mainly by Paneth cells as part of the innate immunity (Figure 1.4) [38-43]. They restrict numbers of intestinal microbes in proximity to the GI epithelium by inducing cell lysis [44-46]. While  $\alpha$ -defensins are fairly abundant, and show constitutive as well as inducible expression, RegIII $\gamma$  production is niche-specific and activated upon bacterial stimuli [47-48]. RegIII $\gamma$  seems to preferentially target Gram-positive bacteria by recognition of their cell wall component peptidoglycan (see 1.3.3) [49].

## 1.2 Mucins in the GI tract

Mucins are highly glycosylated proteins constituting the major component of all mucosal surfaces of the respiratory, digestive and urogenital tract, where they build a protective barrier against the external environment. To date 20 members of the mucin family have been identified, which are characterised by a highly-regulated, tissue-specific and overlapping expression [50].

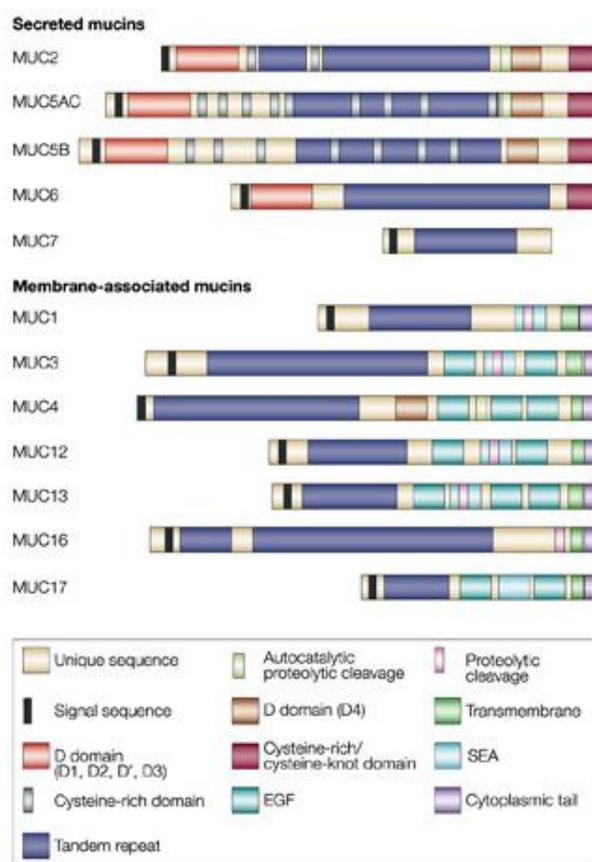


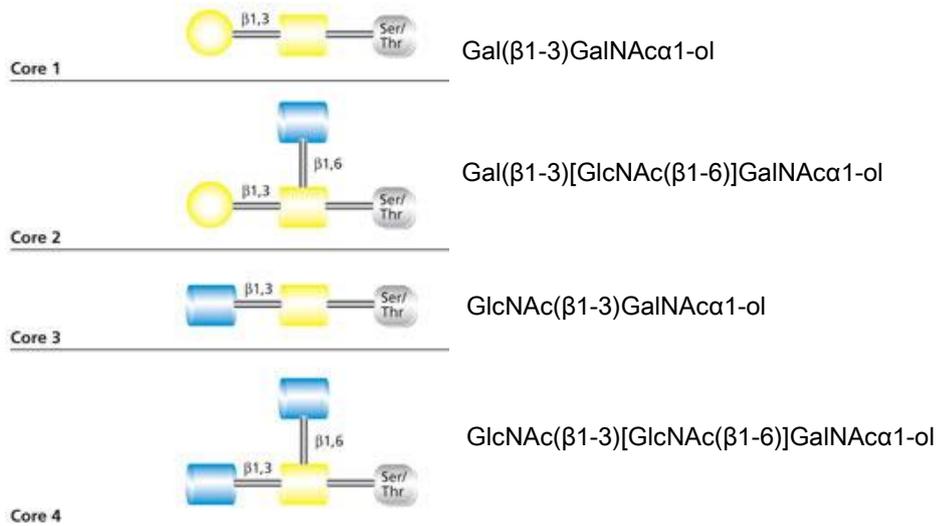
Figure 1.5 Schematic representation of the domain architecture of secreted and membrane-bound mucins (taken from [51])

Apomucins vary in size from about 30 to over 1000 kDa, i.e. the human *MUC2* gene product is more than 5100 amino acids in size giving a 540 kDa apoprotein [52-53]. Due to difference in location and molecular organisation mucins can be divided into three subfamilies: secreted gel forming (MUC2, MUC5AC, MUC5B, MUC6 and MUC19), secreted non-gel forming (MUC7, MUC8 and MUC9) and membrane-bound mucins (MUC1, -3A, -3B, -4, -11, -12, -13, -15, -16, -17, -18, -20, -21) (Figure 1.5). In the

GI tract, 15 different mucins are expressed including the main gel-forming mucins MUC2 (small intestine, colon), MUC6 (stomach) and MUC5AC (stomach), and for example the cell surface mucins: MUC1 (stomach, small intestine), MUC3A/B (small intestine, colon), MUC4 (colon, stomach), MUC12 (colon, small intestine), MUC13 (colon, small intestine) and MUC17 [54-60]. Gel-forming mucins are the main constituent of the hydrated mucus gel, whereas membrane-bound mucins on mucosal epithelial cells constitute the glycocalyx.

Membrane-bound mucins possess a C-terminal cytoplasmic domain for signal transduction and share common extracellular domains such as the epidermal growth factor-like (EGF) domain or Sea urchin sperm protein Enterokinase, and Agrin (SEA) domain [61-64]. The SEA domain, close to the cell surface, undergoes proteolytic cleavage facilitating mucin shedding into the mucosal environment [65-66]. Secreted-gel forming mucins are characterised by cysteine-knot (CK) domains mediating C-terminal dimerisation and the N-terminal D domains allow further oligomerisation [67]. D domains are designated as van Willebrand (VW) domains due to sequence and structural similarities to the VW factor, a disulphide-bridged multimeric glycoprotein [68-69].

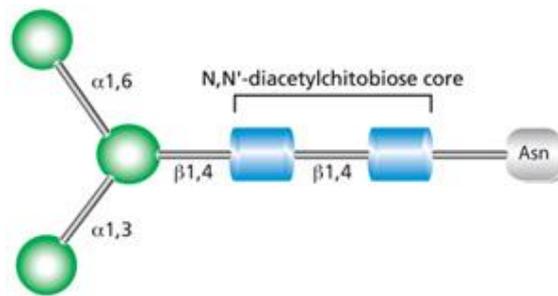
All mucins possess one or more regions with variable numbers of tandem repeat (VNTR) sequences called PTS domains, which are rich in the amino acids proline (P), threonine (T) and serine (S). Thr and Ser residues of the P-T/S motif are potential *O*-glycosylation sites allowing the decoration of the apomucin with a diverse array of *O*-glycans, giving it its filamentous 'bottle-brush' appearance of several hundred nanometers in length [70]. Mucin *O*-glycosylation is initiated in the Golgi-apparatus by an *N*-acetylgalactosamine (GalNAc)-transferase attaching GalNAc to the hydroxyl group of Ser or Thr. The initial GalNAc-Ser/Thr precursor, termed Tn-antigen, is then further modified by glycosyltransferases generating mainly 4 most common *O*-glycan core types (core 1 to 4) along the GI tract (Figure 1.6) [71].



**Figure 1.6 Mucin type *O*-glycan core structures**

Core 1 to 4 *O*-glycan structures with Gal (Galactose) (yellow sphere), GalNAc (*N*-acetylgalactosamine) (yellow square) and GlcNAc (*N*-acetylglucosamine) (blue square) attached to Ser/Thr of the polypeptide chain (taken and adapted from [www.sigma-aldrich.com](http://www.sigma-aldrich.com)).

For further elongation, Gal and GlcNAc molecules are attached to these core structures. The *O*-glycan chains are further modified by differentially linked Fuc, Neu5Ac and GalNAc residues and different terminal structures are formed, such as blood group antigens, resulting in an enormous complexity of mucin associated *O*-glycans [57]. Additionally, mucin glycans often carry sulphate groups adding another level of diversity. The mucin glycosylation pattern depends on genetic background, is tissue-specific and altered in disease, for example inflammatory bowel disease (IBD), ulcerative colitis (UC), or cancer due to the differential expression or presence of certain glycosyltransferase genes [72-79].



**Figure 1.7 Common core building block of *N*-glycans**

Man (green sphere) and GlcNAc (blue square) residues attached to the polypeptide chain via asparagine (Asn) (taken from [www.sigma-aldrich.com](http://www.sigma-aldrich.com)).

Besides the predominant *O*-glycans, mucins additionally carry a small number of *N*-glycans where a GlcNAc residue is attached to asparagine (Asn, N) within the consensus sequences NXS/T, which is located outside of the PTS domains (Figure 1.7) [80-81]. In contrast to *O*-glycans, *N*-glycans contain Man as part of their common core building block (Man<sub>3</sub>GlcNAc<sub>2</sub>). Mucin *N*-glycosylation has a role in the correct dimerisation of mucin molecules [82-84].

### 1.2.1 The membrane-bound mucin MUC1/Muc1

MUC1 is the major cell-surface mucin of the GI tract and constituent of the glycocalyx, the carbohydrate-rich and final protective coating of mucosal epithelial cells [85]. It contains a PTS domain, the site for *O*-glycosylation (as described above), with 20 amino acid tandem-repeats, an extracellular SEA domain, a hydrophobic membrane spanning region and a cytoplasmic domain (Figure 1.8) [86-87]. The latter contains phosphorylated tyrosine (Tyr) residues, which are thought to mediate signal transduction [62, 88-89]. MUC1, as present at the cell-surface, contains two non-covalently linked subunits. Autoproteolysis during mucin biosynthesis and subsequent reassociation of the two subunits is mediated via the SEA domain [90-91]. Although MUC1 is mainly associated with the cell membrane, it is also present in soluble form as a splice variant lacking the cytoplasmic tail or as a proteolytic cleavage product [92-94].



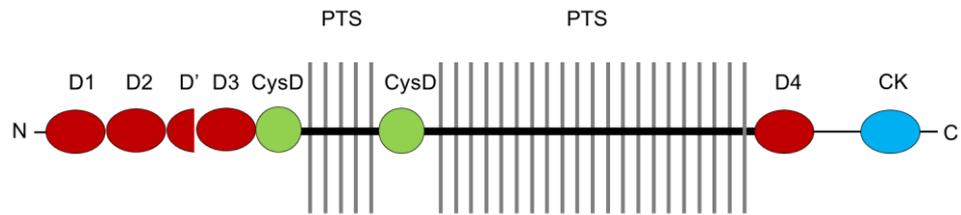
**Figure 1.8 Schematic representation of the membrane-bound mucin MUC1**

Cell membrane (blue) MUC1 presented with its cytoplasmic domain, and extracellular SEA (green) and PTS (black) domains with attached *O*-glycans (grey).

MUC1 *O*-glycans, which account for up to 80% of the protein mass in colonic epithelia cells, share common features with other mucin glycans, including the expression of poly-LacNAc chains, and show tissue-specific glycosylation patterns (see 1.2) [57, 92, 95]. Aberrant and under-glycosylated MUC1 is found in gastric, colonic, breast or ovarian carcinoma, and often characterised by the reduction of *O*-glycans chains to the core type level and increased levels of sialylation [96-99]. For example, certain glycan epitopes, such as the sialyl-Tn (Neu5Ac $\alpha$ 2-6GalNAc $\alpha$ ) and the core 2 TF-antigen (Gal $\beta$ 1-3GalNAc $\alpha$ ), sialyl-Le<sup>x</sup> but not the core 1 Tn-antigen (GalNAc $\alpha$ ), show elevated expression in colon carcinoma [100-101].

### 1.2.2 The gel-forming mucin MUC2/ Muc2

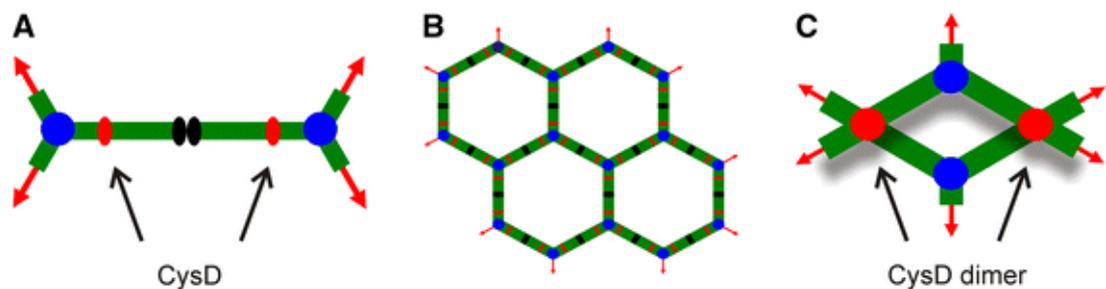
The human MUC2 and its homologue Muc2 expressed in mice, rats and pigs are the major gel-forming mucins in the small intestine and colon [54, 78, 102-103]. MUC2 contains several domains: three complete (D1-3) and one incomplete (D') cysteine (Cys)-rich von Willebrand (VW) D domains in N-terminal, two CysD domains, one adjacent to and one interspersed between the two central PTS domains, two complete VW D-domains (D3-4) in C-terminal, and Cys-knot (CK) domain (Figure 1.9) [104].



**Figure 1.9 Schematic representation of the human MUC2 domain organisation**

MUC2 with 4 complete van Willebrand (VW) D domains (D1-4) (blue), two CysD domains (red), two PTS domains (black) with glycan residues represented as grey lines and a CK-domain (black).

The C-terminal CK domain facilitates dimerisation of MUC2 in the endoplasmic reticulum (ER) and further trimerisation occurs via the N-terminal D domains in the Golgi-apparatus (Figure 1.10 A) [82, 105]. The CysD domain which separates the central PTS domains, has recently been reported to mediate non-covalent crosslinking of MUC2, thereby contributing to the net-like appearance, particularly of the inner mucus layer, and determining the pore size of the mucus mesh (Figure 1.10 B and C) [106]. Ambort and co-workers demonstrated the storage of MUC2 in goblet cell granules as hollow mucin tubules around a concatenated, hexameric ring [107].



**Figure 1.10 Representation of the MUC2 net formation**

(A) MUC2 dimerisation via the C-terminal CK domains (black) and trimerisation via the N-terminal D-domains (blue). (B) Model for a net-like polymeric MUC2 sheet with concatenated, hexameric ring structures. (C) Crosslinking of two MUC2 sheets via non-covalent dimerisation of CysD domains (taken from [29]).

Storage, release and expansion of the mucus-net may be regulated by  $\text{Ca}^{2+}$ - and pH-dependant interactions of structural mucin domains, mainly of the VW D domains, as observed for the VW factor [108-110]. The two central PTS domains of MUC2 contain about 100 tandem repeat sequences of 23 amino acids with a high frequency

of Thr *O*-glycosylation sites. *O*-glycans account for up to 80% of the molecular weight of MUC2 forming a mucin monomer of over 2.5 MDa in mass [52, 111-112].

### 1.2.3 MUC2 *O*-glycosylation

The MUC2 *O*-glycans present in different regions of the normal human gut reveal a tremendous structural complexity and diversity, while only limited variability was observed between individuals [113-114]. More than 100 different, both linear and branched, glycan structures with up to 12 monosaccharide residues, have been identified in human MUC2 [113-114]. An increasing gradient of sialic acid (NeuAc) and sulphate residues was found along the human GI tract from the ileum to the rectum, while the abundance of Fuc decreases [115].

Composition of oligosaccharide alditols and terminal epitopes	
LacNAc	Gal(β1-3)GlcNAc (type 1) Gal(β1-4)GlcNAc (type 2)
Core 3 based	Gal(β1-3/4)GlcNAc(β1-3)[NeuAc(α2-6)]GalNAc-ol
Tn	GalNAc-ol
Sialyl-Tn	NeuAc(α2-6)GalNAc-ol
Terminal NeuAc	NeuAc(α2-3)Galβ
Blood group H	[Fuc(α1-2)]Gal(β1-3)GlcNAcβ
Blood group A	GalNAc(α1-3)[Fuc(α1-2)]Galβ
Sd <sup>a</sup> /Cad	GalNAc(β1-4)[NeuAc(α2-3)]Galβ
Le <sup>a</sup>	Gal(β1-3)[Fuc(α1-4)]GlcNAcβ
Le <sup>b</sup>	[Fuc(α1-2)]Gal(β1-3)[Fuc(α1-4)]GlcNAcβ
Le <sup>x</sup>	Gal(β1-4)[Fuc(α1-3)]GlcNAcβ
Sulphated structures	Gal-3[SO <sub>3</sub> -Gal-GlcNAc-6]GalNAc-ol Gal-3[SO <sub>3</sub> -Gal-(Fuc)GlcNAc-6]GalNAc-ol Fuc-Gal-3[SO <sub>3</sub> -Gal-(Fuc)GlcNAc-6]GalNAc-ol

**Table 1.1 Predominant oligosaccharide structures in human MUC2**  
NeuAc (*N*-acetylneuraminic acid or sialic acid) and Fuc (Fucose)

The core 3 structure is the main building block of nearly all human colonic MUC2 *O*-glycans, whereas in mice core 2 and core 1 structures are more abundant [113-114, 116-117]. It is modified with *N*-acetyllactosamine-units (LacNAc),  $\beta$ 1-3 or  $\beta$ 1-4 linked Gal and GlcNAc forming type 1 or type 2 LacNAc, respectively, and terminated by Neu5Ac, Fuc and SO<sup>3-</sup> (Table 1.1) [113-114, 118-120]. A few structures in the distal colon are built around the core type 2 and are mostly sulphated, while core 4 *O*-glycans found in the ileum possess a high Fuc content ( $\alpha$ 1-4 or  $\alpha$ 1-2 linked) [113-114]. The tetrasaccharide Gal( $\beta$ 1-3/4)GlcNAc( $\beta$ 1-3)[NeuAc( $\alpha$ 2-6)]GalNAc-ol or the sialyl Tn-antigen (NeuAc $\alpha$ 2-6GalNAc-ol), both based on the core 3 structure, were detected in the majority of carbohydrate chains [113]. A high degree of sialylation and sulphation often coexisting was observed on MUC2 oligosaccharides (Table 1.1). Sulphation occurs at position C3 of Gal and C6 of GlcNAc. NeuAc was typically found  $\alpha$ 2-6 linked to GlcNAc and  $\alpha$ 2-3 linked to Gal often forming a  $\alpha$ 2-6 sialylated core type 3 (NeuAc( $\alpha$ 2-6)GlcNAc-ol) or a terminal NeuAc epitope. Several terminal ABO blood group determinants including blood group H and A epitopes as well as Le<sup>a</sup>, Le<sup>b</sup> and Le<sup>x</sup> structures can be identified in the intestine (Table 1.1). Modified blood group Sd<sup>a</sup>/Cad-determinants are also present in colonic mucins [114, 118].

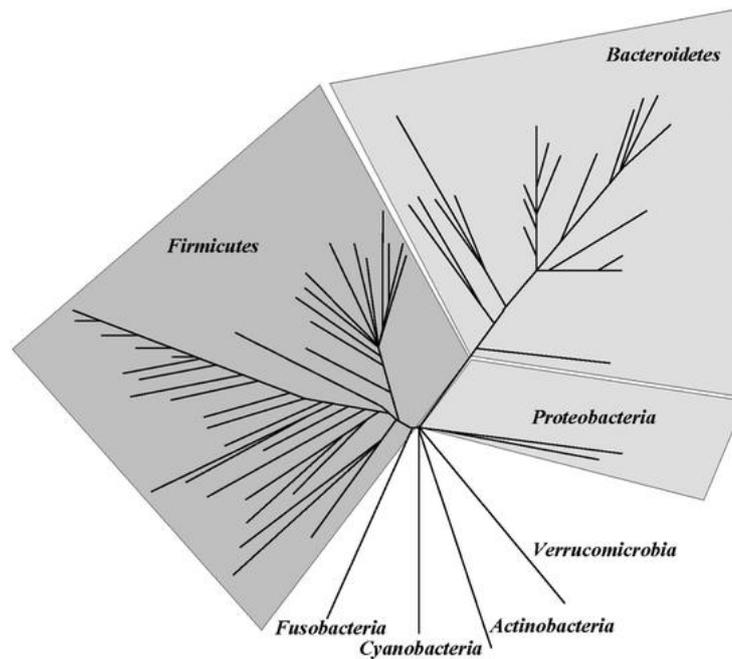
MUC2 *O*-glycosylation is an important intrinsic mucus property, facilitating for example the hydration of the mucus gel after release due to its acidic nature. Additionally, a high content of sialic acid and sulphate groups in the colon of humans is beneficial to prevent degradation via bacterial or host glycosidases [121-123]. Numerous members of the residential gut microflora produce glycosidases providing them with the ability to utilise mucin as a carbon source [124-128]. Moreover, Muc2 *O*-glycans are critical to maintain the mucus barrier properties. Studies in mice showed that the aberration of certain core-derived *O*-glycans results in increased mucus permeability. This allows greater translocation of bacteria into mucosal tissues, which leads to inflammation or colitis [129-130]. In addition, mucin glycans are believed to serve as attachment sites for the gut bacteria, which have adapted to the mucosal environment by expressing complementary adhesins [22, 131]. Hence, a correlation between bacterial colonisation, and mucin expression and glycosylation can be deduced.

### 1.3 Mechanisms of host-microbe interaction in the GI tract

The mucosal surface of the GI tract is in constant contact with an enormous resident microbiota, which offers physiological advantages for the host including degradation of indigestible nutrients, development of the immune system and enhancement of the mucosal barrier facilitating pathogen colonisation resistance [14, 132-135]. Dysbiosis of the gut microbiota has been associated with a growing number of gut-related diseases, for example UC, Crohn's disease (CD) and obesity, indicating the importance of maintaining a homeostatic relationship with the resident gut bacteria [136-140]. Mucus is the first point of contact of GI bacteria with the host and mucus-microbe interactions are critical to maintain a healthy interplay and for the selection of a beneficial microbial community in the GI tract [141].

#### 1.3.1 The gastrointestinal microbiota composition

The colonisation of the gut starts at birth and the composition of the intestinal bacterial community changes over time establishing a fairly stable gut microbiota in adults [142-145] with an estimated number of  $10^{13}$  to  $10^{18}$  microbes, which outnumber our own body cells by a factor of 10 [14]. In the established microflora, an increase in bacterial population density and diversity can be observed from the stomach to the colon [146-148]. Despite the enormous number of colonic bacteria they only represent 8 phyla (division), namely *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, *Proteobacteria*, *Fusobacteria*, *Verruomicrobia*, *Cyanobacteria* and *Spirochaetes*, but vary highly at species (subdivision) level (Figure 1.11) [14, 149]. The majority of the resident commensal microflora belongs to the phyla *Bacteroidetes* followed by *Firmicutes*, as several genomic based approaches mainly focussing on bacterial populations in faeces have shown over the last years [149-150]. From these studies only about 400 species were found to be culturable but a total of several thousand species can be estimated in the colon, of which at least 160 are present in one individual, according to the latest reports using 16S ribosomal RNA sequencing techniques [149, 151-152].



**Figure 1.11 Phylogenetic tree of GI microbiota**

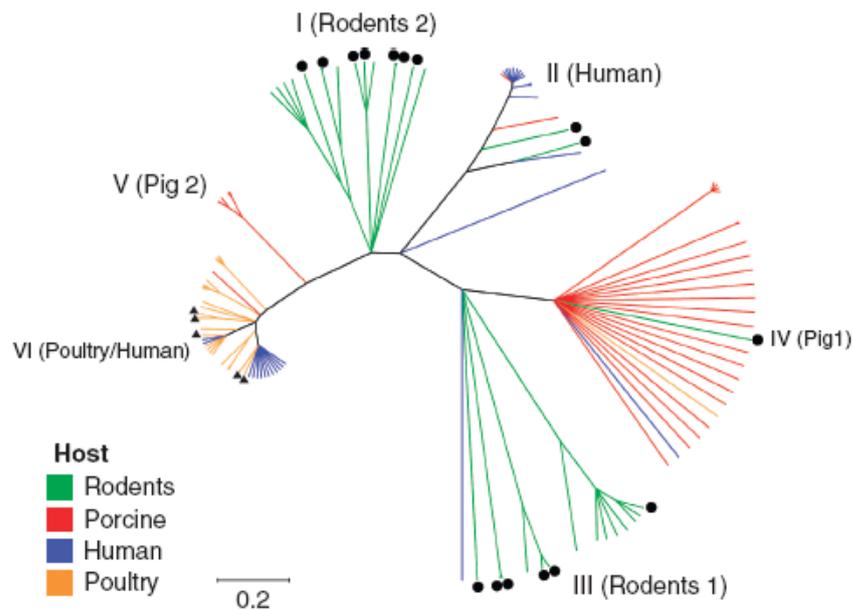
Phylogenetic analysis of the bacterial flora in human mucosal sites and faeces via 16S rRNA analysis (taken from [153]).

Despite variations in the microbial population between individuals, even between twins, there is evidence for a common phylogenetic core [154-155]. Representatives of the *Bacteroidetes* and *Firmicutes*, mainly from *Clostridium* clusters, were found to be predominant inhabitants of mucus [156-157]. Furthermore, mucosa-associated lactobacilli have been identified in human intestinal tissue (stomach, small intestine and colon) and faeces samples including *L. rhamnosus*, *L. plantarum* and *L. reuteri* strains [158-160].

Some beneficial bacteria associated with the human gut are termed probiotics among them 17 identified *Lactobacillus* species as well as *Bifidobacteria* and some beneficial *Streptococcus* strains [161-162]. Probiotics are defined by the FAO/WHO as “live microorganisms that when administered in adequate amounts confer a health benefit on the host”, when fulfilling certain criteria including a demonstrated health benefit, lack of adverse side effects, and prolonged residency and survival in the GI tract [163]. For example, *Bifidobacterium* and *Lactobacillus* strains, which possess mucus-binding ability, are capable of strain dependant inhibition of adhesion via direct competition or upregulation of *MUC* gene expression, as well as displacement of different pathogens [19-20, 164-168]. A therapeutic effect of probiotics, including lactobacilli species, in

intestinal diarrhoea and gastroenteritis as well as prevention of traveller's diarrhoea has been reported in humans [169-170]. Generally, probiotic actions of *Lactobacillus* strains comprise pathogen inhibition, enhancement of the epithelial barrier function and immunomodulation [171].

Several *Lactobacillus* species, including *L. reuteri*, *L. gasseri*, *L. crispatus*, *L. salivarius* and *L. ruminis*, are autochthonous residents of the vertebrate gut, which form stable populations and occupy a biological niche albeit at low bacterial numbers in humans [137, 162, 172-175]. In contrast, other allochthonous species such as *L. plantarum*, *L. rhamnosus*, *L. johnsonii* or *L. adicophilus*, have not been proven to stably inhabit the GI tract, but are often found in food produces [162].



**Figure 1.12 Phylogenetic analysis of *L. reuteri* isolated from different vertebrate hosts**

Genotyping of *L. reuteri* strains originating from rodents, pig, human and poultry (colour coded) and clustering into *L. reuteri* populations (taken from [175-176]).

In particular, the gut symbiont *L. reuteri* shows association with different vertebrate hosts as diverse as humans, pigs, rats, mice and some bird species, as well as close host co-evolution and specialisation (Figure 1.12) [175, 177]. Hence, *L. reuteri* is a good model gut inhabitant to study the mechanisms that mediate lactobacilli persistence in the GI tract. Indeed, oral administration of an autochthonous human *L. reuteri* isolate strain to human subjects showed high levels of bacterial colonisation

and persistence [178]. Probiotic effector molecules of *L. reuteri*, which mediate bacterial adherence and aggregation, are for example collagen-binding proteins (CnbP/MapA), a mucus-binding protein (Mub), a large surface protein (Lsp) and glycoyltransferases (GtfA/Inu) [175].

### 1.3.2 Bacterial interaction at the mucosal surface

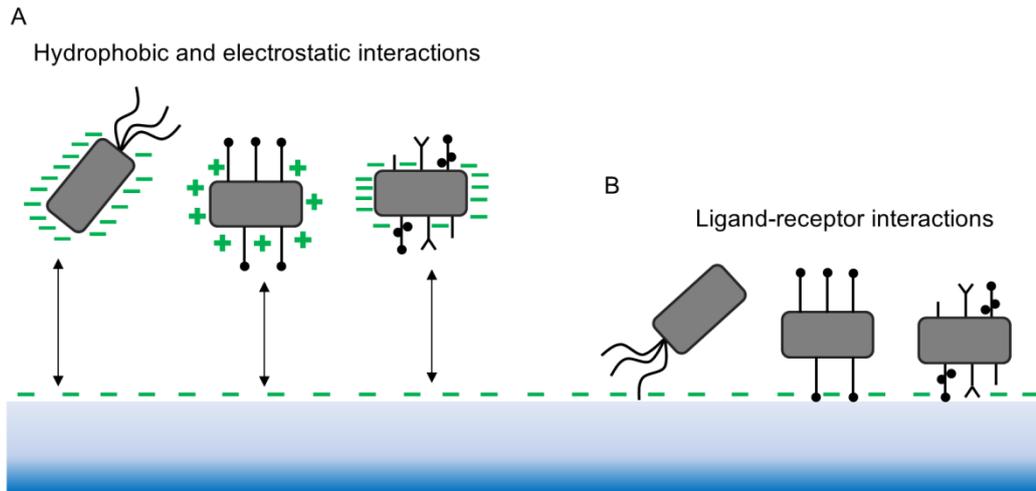
Intestinal bacterial communities show adaptation to various ecological niches impacted by physiological conditions, i.e. pH, mucin glycosylation, nutrient requirements and metabolism [125, 179-182]. Gut bacteria express a repertoire of metabolic genes for hydrolysis, uptake and degradation of dietary carbohydrates depending on location in the gut and host diet [183-187]. Additionally, *Bacteroides*, *Akkermansia* and *Ruminococcus* species produce mucin-degrading enzymes allowing them to utilise mucin-glycans, abundant and constitutively expressed in GI mucus (see 1.2.2), as a nutrient source [126, 128, 186, 188-189]. Efficient nutrient metabolism allows commensal inhabitants of the gut to outcompete pathogenic bacteria and directly inhibit pathogen growth or virulence by production of metabolites such as short chain fatty acids (SCFA) [190-194].

Mucin *O*-glycans are not only a substrate for intestinal bacteria, but are also considered important for the selection of the bacterial gut flora by providing attachment sites for GI bacteria that have adapted to the mucosal environment by expressing the right complement adhesins [22]. Variations in the mucin glycosylation pattern along the GI tract, including core type structures, fucosylation, sialylation and sulphation as well as the expression of certain glycan epitopes, such as the Sd<sup>a</sup>/Cad antigen or ABO blood group antigens (see 1.2.2), coincide with and directly influence variations in bacterial composition among different parts of the gut [148, 195-198]. Alterations in mucin expression and glycosylation have been reported in intestinal inflammation, whether this is a cause or consequence of changes in the microbial flora remains elusive [79, 199-202]. Lactobacilli can induce MUC2 or MUC3 expression [19-20] and *Bacteroides thetaiotaomicron* has been reported to enhance the expression of terminal Fuc residues on mucin glycans, which the bacterium can then harvest and use as a carbon source [186, 203]. The expression of gel-forming and membrane-bound mucins, including Muc2 and Muc1, is also altered by pathogenic bacteria such as *Listeria monocytogenes*, *Pseudomonas aeruginosa* or *Citrobacter rodentium*, which is

used as a surrogate for *Escherichia coli* in *in vivo* infection studies of human gastroenteritis, and is considered as a host defence mechanism against pathogenic infection [204-206]. In addition, membrane-bound mucins such as MUC1 are shed into the mucus layer, where they act as soluble decoy receptors facilitating pathogen exclusion [207]. However, GI pathogens such as *Helicobacter pylori* and *Entamoeba histolytica* have developed strategies allowing them to penetrate the mucus layer, for example via local pH changes or MUC2 proteolysis [208-209]. In contrast, potential probiotic *Lactobacillus* and *Bifidobacterium* species do not disrupt the mucus due to a lack of secretion of MUC2 proteases, stressing their role as commensal inhabitants of mucus [210].

#### **1.4 Bacterial adhesins**

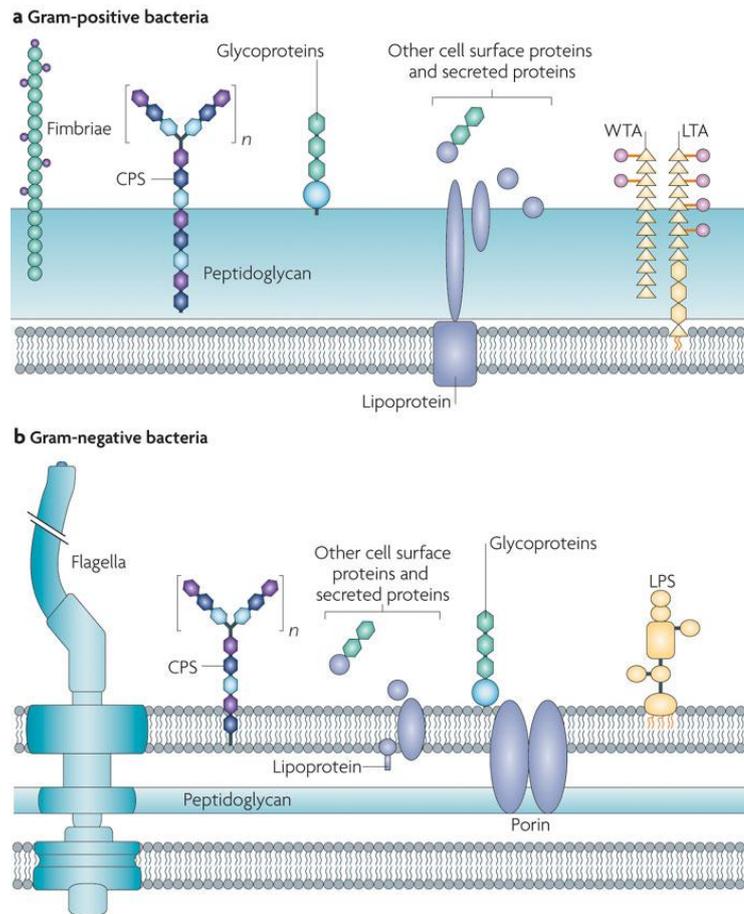
Bacterial adhesion to the host is an important first step to promote colonisation and persistence, penetration of the host cell barrier or induction of signalling pathways [211-216]. The initial adhesion of bacterial cells to solid surfaces is mainly driven by hydrophobic properties of the microbial surface (van der Waals interaction) and additionally influenced by cell-cell or cell-surface electrostatic interactions (Figure 1.13) [217-220]. While the majority of bacteria, regardless of whether they are Gram-positive or Gram-negative, possess a negative surface charge such as the surfaces they adhere to, adhesive positively charged organism have been identified as well [221-223].



**Figure 1.13 Schematic representation of bacterial-surface interactions**

(A) Initial adhesion of bacteria (grey) with mostly negative or positive surface charge to a negatively charged surface via hydrophobic and electrostatic interactions. (B) Bacteria close to the surface engaging in ligand-receptor interactions via bacterial cell-surface molecules.

These initial mechanisms of contact between bacteria and cell surfaces allow a more specific and closer host-bacteria interaction mediated by host receptors and bacterial cell-surface molecules [224]. The latter include adhesion proteins such as pili (also known as fimbriae), flagella or other cell components like lipopolysaccharides (LPS) of Gram-negative, and exopolysaccharides (EPS) or lipoteichoic acid (LTA) of Gram-positive bacteria (Figure 1.14) [141].



**Figure 1.14 Cell wall architecture and surface molecules of Gram-positive and Gram-negative bacteria**

Structural components of the bacterial cell wall of (a) Gram-positive and (b) Gram-negative bacteria including fimbriae (pili) and flagella (found in both bacterial groups), cell-surface proteins (including adhesins), CPS (cell-wall associated polysaccharides), and WTA (wall teichoic acid) or LTA (lipoteichoic acid) of Gram-positive and LPS (lipopolysaccharides) of Gram-negative bacteria (taken from [141]).

Several cell surface molecules of both pathogenic and commensal species have been suggested to be involved in the specific interaction with host receptors, including mucins, mucin-like glycans or ECM proteins. These adhesins include flagella and pili, extended cell-surfaces appendages found in Gram-positive and Gram-negative bacteria, MSCRAMM (microbial surface components recognising adhesive matrix molecules), sugar recognising proteins (i.e. lectins), Serine (Ser) rich proteins or other often modular domain surface proteins [225-227].

### 1.4.1 Mucus adhesins in pathogenic bacteria

Most of the current knowledge on microbe-mucus interaction and the specific adhesins and ligands involved are limited to enteric pathogens (Table 1.2), but information on the biochemical and structural basis of interaction is often rudimentary.

Pathogen	Adhesin	Ligand	Reference
<i>Clostridium difficile</i>	FliC and FliD (Flagella)	Mouse mucus	[228]
<i>Pseudomonas aeruginosa</i>	FliD (flagellum)	Respiratory mucins (Muc1)	[229-230]
<i>Escherichia coli</i> O126:H6	Flagellum	Bovine mucus, mucins and ECM proteins	[231]
<i>E. coli</i> O157:H7	Flagellum	Bovine mucus, mucins and ECM proteins	[231]
<i>E. coli</i> O157:H7	Pili (type 1)	Intestinal mucin (via mannose)	[232-233]
<i>E. coli</i>	FimH (pili)	mannose	[234-236]
<i>E. coli</i>	F17-G (pili)	Intestinal mucin (via GlcNAc)	[237-238]
<i>Streptococcus pneumoniae</i>	Spr1345	PGM, bovine submaxillary mucin, polysaccharides	[239-240]
<i>Helicobacter pylori</i>	BabA	Le <sup>b</sup> (on MUC5AC and MUC1)	[241-244]
	SasA	sialyl-Le <sup>x</sup> (on glycolipids)	
<i>Vibrio cholerae</i>	GbpA	Intestinal mucin (via GlcNAc)	[245]
<i>Entamoeba histolytica</i>	Gal/GalNAc lectin	Human and rat colonic mucins (via GalNAc and Gal)	[246-248]
<i>Listeria monocytogenes</i>	Internalins	Human MUC2	[249]

Table 1.2 Adhesins and potential host receptors of selected pathogens

Flagella are flexible cell appendages important for bacterial motility, chemotaxis and virulence in pathogens [250-251]. The binding of flagella to bovine mucus and mucins for the enteropathogenic and enterohemorrhagic *E. coli* (EPEC and EHEC) strains of *E. coli* O157:H7 and *E. coli* O126:H6 was demonstrated [231]. In addition, the flagellum of *E. coli* O126:H6 also showed binding to ECM proteins. The FliD flagellum protein of *P. aeruginosa* is involved in the bacterial adhesion to human respiratory mucins and rodent Muc1 [229-230]. Furthermore, the FliC and FliD flagella proteins, both recombinant and native, of *Clostridium difficile* bind to mouse mucus but not to porcine stomach mucus [228]. However, whether mucin glycans are involved in the mucin/mucus binding of the described flagella has not been addressed.

Pili are important virulence factors in pathogenic bacteria relevant for host tissue colonisation, and their binding ability to host ligands is especially well described in Gram-negative species [227]. The *E. coli* F17-G pili protein demonstrated binding to GlcNAc as well as GlcNAc $\beta$ 1-3/ $\beta$ 1-6Gal $\beta$ 1- in intestinal mucin *O*-glycans [238]. In fact, the binding of the F17-G adhesin to several sugar molecules including GlcNAc saccharides has been characterised by surface plasmon resonance (SPR) experiments and the GlcNAc binding site identified via X-ray crystallography [237]. Another extensively studied type 1 pili adhesin is FimH present in *E. coli* and *Salmonella enterica*, which binds specifically to mannose demonstrated [234-236]. Several mannose containing glycan receptors of FimH on urinary epithelial cells have been suggested including ECM proteins, but no interaction with mucins has been reported yet [236, 252-253]. Additionally, the FimH has been structurally described and the mannose binding site identified by X-ray crystallography [235, 254]. Another example for a fimbrial adhesin is Std of *S. enterica* which mediates binding to intestinal epithelial cells *in vitro* [255]. Inhibition studies using the intestinal epithelial cell line Caco-2, which expresses mucin glycans, demonstrated reduced binding of Std positive bacterial cells in the presence of the blood group epitope H2 (Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc), while LacNAc (Gal $\beta$ 1-4GlcNAc) did not have an effect [256-257]. In addition, binding of isolated Std pili proteins to mouse intestinal mucus was competitively inhibited in the presence of the Fuc $\alpha$ 1-2 specific agglutinin from *Ulex europaeus* (UEA). Gram-positive pili have mainly been studied for species commonly encountered in the respiratory tract where they bind to epithelial cells, the place of respiratory mucin production [227, 258-259]. However the cell receptors for pili proteins such as Spa of *Corynebacterium*

*diphtheriae* or GBS52 of *Streptococcus agalactiae* have not yet been identified [260-263].

The Spr1345 cell-surface protein from *Strept. pneumoniae* contains a mucin-binding protein (MucBP) domain (Pfam database PF06458), which is often found in lactobacilli proteins (see 1.4.2). Recombinant MucBP, for which the crystal structure has been determined by X-ray crystallography, binds to porcine gastric mucin (PGM) and bovine submaxillary gland mucin as well as sulphated and non-sulphated glycosaminoglycans (GAGs), long unbranched polysaccharides [239-240].

The interaction of *H. pylori* to gastric mucins, which is mediated via the recognition of ABO blood group antigens by two adhesins, BabA (blood group antigen-binding adhesin) and SasA (sialic acid-binding adhesin), has been extensively studied [264]. BabA binds to Le<sup>a</sup> (Gal $\beta$ 1-3[Fuca1-4]GlcNAc) on MUC5AC or MUC1, while SasA recognises sialylated Lewis antigens, sLe<sup>x</sup> (NeuAc $\alpha$ 2-3Gal $\beta$ 1-4[Fuca1-3]GlcNAc) and sLe<sup>a</sup> (NeuAc $\alpha$ 2-3Gal $\beta$ 1-3[Fuca1-4]GlcNAc) [50, 241-242]. These epitopes are found on glycolipids and their expression is upregulated during gastritis [244, 265].

A Gal/GalNAc lectin of the enteric parasite *E. histolytica* has been identified and isolated via Gal affinity chromatography [266]. *E. histolytica* binds to rat and human intestinal mucosa and adherence can be inhibited by GalNAc [248]. Additionally, the binding of the parasite to Chinese hamster ovary cells can be inhibited by rat and human colonic mucins as well as Gal [246].

*Vibrio cholerae* colonises the host intestine via adherence to mucosal surfaces, which is mediated by different cell-surface molecules including the chitin-binding protein GbpA [245, 267-268]. Recombinant GbpA showed binding to mouse mucus and isolated mucin in microtitre plate assays and mucin binding was abolished in the presence of GlcNAc [245].

Members of the internalin (InI) protein family of *Listeria monocytogenes* have been demonstrated to facilitate entry into host cells, including intestinal epithelial cells, via interaction with different receptors [269-270]. Furthermore, a study suggested binding of human intestinal MUC2 to InIJ, InIB and InIC, but no further characterisation of the binding has been performed [249].

#### 1.4.2 Mucus adhesins of *Lactobacillus* species

In contrast to the often well characterised pathogenic adhesins and their specific host receptors, information on commensal adhesion molecules and host ligand structures, especially mucosal surface components, is rudimentary. The majority of studies on commensal adhesins have been carried out in *Lactobacillus* species due to their postulated probiotic effects [161, 168, 271]. Indeed, the ability of lactobacilli to adhere to intestinal mucus is regarded as a probiotic factor and the varying mucus adhesion capability has been demonstrated for a variety of *Lactobacillus* strains [272-274]. Immobilised mucus/ mucin, intestinal epithelial cell culture and whole tissues are frequently used *in vitro* adhesion models to study mucosal adhesion [275]. Recently, more sensitive techniques to study bacterial mucin binding have emerged including surface plasmon resonance (SPR), atomic force microscopy (AFM) and flow cytometry (FCM) [276-278].

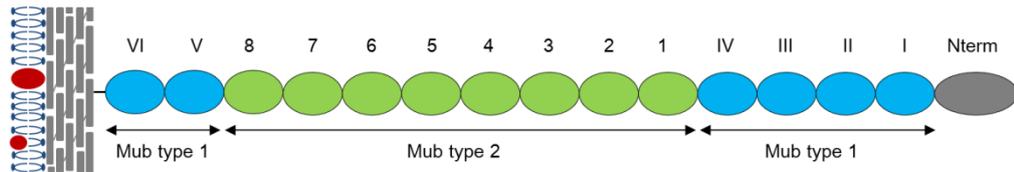
The identification of putative cell-surface adhesion molecules mediating the binding of bacteria to intestinal mucosal surfaces, benefitted greatly from genome sequencing of *Lactobacillus* species (Table 1.3) [181, 279-281].

Commensal	Adhesin	Ligand	Reference
<i>L. reuteri</i> ATCC 53608	MUB	Porcine and chicken mucus, porcine gastric mucin, Igs	[282-283]
<i>L. reuteri</i> 104R	MapA	Small intestinal mucus, porcine gastric mucin	[284-285]
[ <i>L. reuteri</i> NCIB 11951	CnBP	Collagen	[286-287]
<i>L. rhamnosus</i> GG	Spa C and F (Spa B)	Human colonic mucus	[288-289]
	MBF	Human colonic mucus	[290]
<i>Lactococcus. lactis</i> TIL448	YhgE2	Human intestinal cells	[291]
<i>L. acidophilus</i> NCFM	Mub SlpA FbpA	Human intestinal cell	[292]
<i>L. plantarum</i> WCFS1	Msa	Mannose	[293]
<i>L. plantarum</i> LA 318	GAPDH	Human colonic mucin, ABO blood antigens	[294]
<i>L. mucosae</i> ME-340	Lam29	ABO blood group antigens H3 protein (in human colonic mucus)	[295-296]
<i>L. johnsonii</i> La1 NCC 533	EF-Tu GroEL	Human intestinal cells, mucus	[297-298]
<i>L. salivarius</i> UCC118	LspA	Human intestinal cells	[299]
<i>L. fermentum</i> BCS87	32-Mmubp	Porcine mucus and mucin	[300]

**Table 1.3 Commensal adhesins of *Lactobacillus* involved in host surface interaction**

Many *Lactobacillus* adhesins, which have been implicated in mucus or intestinal surface binding, belong to a class of sortase-dependant proteins (SDP) [8, 280]. They are characterised by an N-terminal signal sequence (YSIRK) for transport to the cell wall and a C-terminal LPXTG-motif recognised by an enzyme called sortase, which facilitates covalent cell-wall anchoring via a Thr residue to a peptidoglycan amino group [225, 301].

The canonical sortase-dependent mucus-binding protein (MUB) (353 kDa) of *L. reuteri* ATCC 53608 has been initially functionally characterised by Roos and Jonsson [283]. It is composed of 14 tandemly arranged repeats of two types, Mub type 1 and Mub type 2, present in 6 (R1 to R6) and 8 (R1 to R8) copies, respectively, of about 183 to 206 residues (Figure 1.15).



**Figure 1.15 Schematic representation of MUB of *L. reuteri* 53608**

Mub type 1 (blue) and type 2 (green) labelled with Roman and Arabic numbers, respectively. N-terminal domain (Nterm) coloured grey and the LPXTG-motif presented as black line.

*L. reuteri* ATCC 53608 binds to porcine, chicken and mouse mucus, which is mainly mediated by the presence of MUB on the bacterial cell surface [283, 302]. Fusion proteins of selected Mub type 1 and Mub type 2 repeats with a mannose-binding protein (MBP) showed adhesion to porcine and chicken mucus, as well as porcine gastric mucin [283]. Mub-MBP binding was inhibited by mucin and glycoproteins, fetuin and asialofetuin, suggesting interaction of MUB with a glycan receptor. A synthetic peptide (MUB<sub>70</sub>), comprising the first 70 residues of MubR5, showed binding to Muc2 in colonic mucus tissue sections [303]. In addition, recombinant type 1 and type 2 Mub repeats bind human secretory IgA, IgG and the IgG Fab fragment *in vitro* [282].

A comparative protein database search identified 48 Mub domain-containing proteins in lactic acid bacteria (LAB) of which 30 contained more than 3 or 4 copies of the Mub domain [304]. The majority of those were identified in *Lactobacillus* species, while 4 Mub domain-containing proteins are present in *Lactococcus* and *Pedococcus*. A more recent *in silico* analysis identified 47 mucus-binding proteins in the extracellular proteomes of six *Lactobacillus* genomes [280]. The Mub domain appears to be abundant in LAB, in fact a recent database search by John Walshaw (Institute of Food Research, Norwich, UK) identified a total of 147 Mub domains in *Lactobacillus* species

of the GI tract alone, and 28 Mub domains were present only in *L. reuteri* ATCC 53608 [305].

Mub domains contain a distinct shorter sequence domain termed mucin-binding protein (MucBP) domain as found in the Pfam database (PF06458) (<http://pfam.sanger.ac.uk/>), which, in Mub repeats of *L. reuteri* ATCC 53608, is located at the C-terminal end and folds as a structurally distinct domain (see 1.6) [282].

Another Mub domain protein that was suggested to be involved in bacterial binding to intestinal cells is Mub of *L. acidophilus* NCFM, which contains 15 Mub domains [292, 304]. Additionally, two other adhesion proteins were identified in the same strain, the S-layer protein SlpA and the fibronectin-binding protein FbpA, which contribute to the overall binding to intestinal epithelial cells. The surface (S)-layer is the crystalline proteinaceous and outermost layer of the Gram-positive cell wall and found in some lactobacilli [306].

LPXTG-like pili structures have so far only been identified in *L. johnsonii*, *L. ruminis*, *L. rhamnosus* and recently in *Lactococcus lactis* [288, 291, 307-308]. The pili, present on the cell-surface of *L. lactis* TIL448, is responsible for bacterial adhesion to human intestinal cells, while the adhesive phenotype is nearly absent in a mutant strain with disruptions in the pili genes [291]. *L. rhamnosus* GG encodes two different pilin fibres, SpaCBA and SpaFED, which show a typical pili architecture of Gram-positive bacteria with different pilin subunits [309]. SpaCBA and SpaFED are comprised of major backbone pilin subunits, SpaA and SpaD, minor pilins, (SpaB, SpaC and SpaE, SpaF) [288]. SpaC of the SpaCBA pili has been demonstrated to mediate the bacterial binding to human intestinal mucus [288]. SpaB and SpaF of the SpaFED pili also showed binding to colonic mucus, although the mucus binding ability of SpaB may be due to electrostatic interactions [289]. Another mucus-adhesin, present on the cell surface of *L. rhamnosus* GG, is the mucus-binding factor (MBF), which contains 4 MucBP domains and an anchoring motif [290]. Recombinant MBF has been demonstrated to bind to human colonic mucus and mucus adhesion of bacterial cells was reduced with an anti-MBF antiserum [290].

The sortase-dependant protein MapA of *L. reuteri* 104R (formerly known as *L. fermentum* 104R), was identified by its ability to bind porcine small intestinal mucus and gastric mucin via dot-blot assay after mucin-affinity chromatography [285]. It has

also been demonstrated to be involved in bacterial binding to mucin-expressing human intestinal cells and recombinant MapA has been localised on intestinal cells by immunofluorescence microscopy [284]. LspA of *L. salivarius* UCC118 contains 7 repeats similar to Mub domains of *L. reuteri* ATCC 53608 and has been implicated in bacterial adhesion to colonic epithelial cells [299]. The cell-surface protein Lam29 of *L. mucosae* ME-380 binds to ABO blood group antigens A (GalNAc $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal) and B (Gal $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal) as well as the human histone 3 (H3) protein from colonic mucus, as observed in SPR and microtitre plate assays [295-296]. A mucus-adhesin (32-Mmubp) has been identified in *L. fermentum* BCS87 via its ability to bind porcine mucus and mucin [300].

Additionally, a few normally cytoplasmic proteins show surface association and have been demonstrated to possess mucin-binding ability. The elongation factor Tu (EF-Tu) and the heat shock protein GroEL of *L. johnsonii* NCC533 bind to intestinal epithelial cells and human colonic mucin. In addition, binding of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of *L. plantarum* LA 318 to human colonic mucin and ABO blood group antigens A, and H (Fuc $\alpha$ 1-2Gal) was characterised in SPR experiments [294, 310].

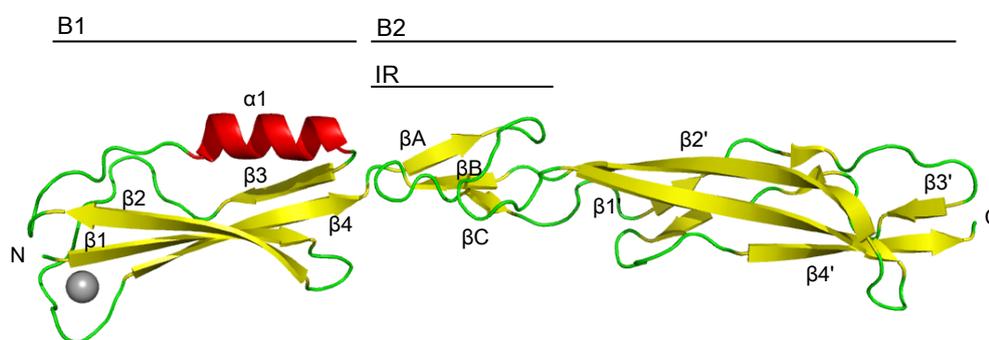
The mannose-dependent binding of a few *L. plantarum* strains to human colonic cells has been observed [311] and a mannose-specific adhesin (Msa) has been identified in *L. plantarum* WCFS1 by mannose-dependent yeast agglutination assays as well as several other *L. plantarum* strains [293, 312]. Msa proteins encompass domains similar to Mub domains found in *L. reuteri* ATCC 53608 and a ConA-like lectin domain of *Staph. aureus* generally responsible for mannose-recognition [293].

The collagen binding protein CnaB of *L. reuteri* NCIB 11951 binds to an alternative cell surface receptor, the ECM component collagen [286-287].

Generally, for many of the identified adhesins described above there is a lack of direct evidence for their binding ability to mucus or mucin. For the majority of these adhesion molecules no further functional characterisation on the nature of the interaction with intestinal cells, mucus or mucins has been performed. Additionally, there is only a limited amount of structural information available on commensal adhesins, which could support their functional characterisation.

## 1.5 Structural basis of bacterial mucus adhesins

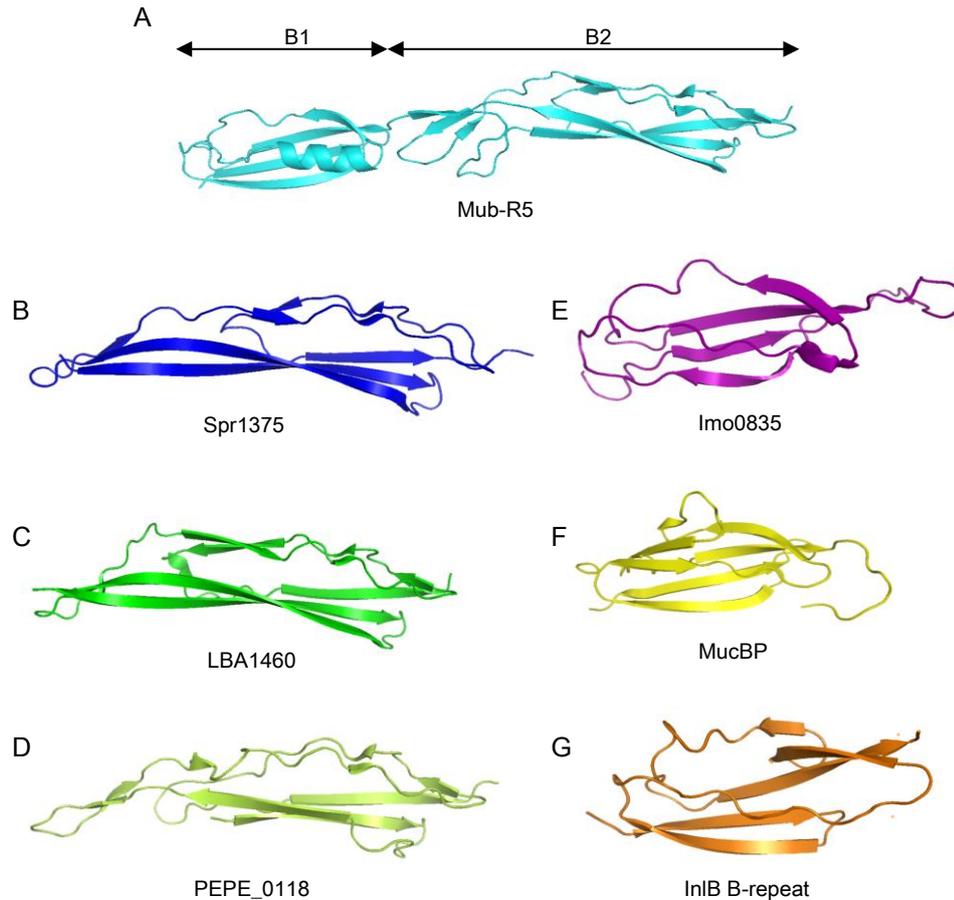
The MUB protein of *L. reuteri* ATCC 53608 is the first commensal mucus adhesin which was structurally characterised. The structure of the Mub type 2 repeat MubR5 (PDB entry 3I57) was determined at 1.8 Å via X-ray crystallography in our lab [282]. MubR5 folds like an elongated distorted cylinder 110 Å long and 25 Å in diameter revealing two distinct domains, an N-terminal B1 (2-75 residues) and a C-terminal B2 domain (76-184 residues) with an inter domain region (IR domain) forming a 3 stranded  $\beta$ -sheet (Figure 1.16).



**Figure 1.16 Presentation of the Mub type 2 MubR5 X-ray crystal structure**

Protein fold of MubR5 with  $\alpha$ -helix (red) and  $\beta$ -sheets (yellow). N- and C-termini and structural elements are labelled.  $\text{Ca}^{2+}$ -ion presented as grey sphere. Structural domains are labelled with B1, B2 and IR (inter domain region).

The B1 domain has an ubiquitin-like  $\beta$ -grasp fold most similar to that found in the Ig-superfamily [313]. This fold contains two pairs of antiparallel  $\beta$ -strands ( $\beta_1$ - $\beta_2$  and  $\beta_3$  and  $\beta_4$ ) in a 4-stranded sheet which are connected by an  $\alpha$ -helix. A  $\text{Ca}^{2+}$ -ion is coordinated by residues of the loop connecting strands  $\beta_3$  and  $\beta_4$  and two water molecules complete the coordination sphere stabilising the loop. Calcium binding is characteristic for calcium-dependant lectins (C-type lectins) recognising carbohydrate ligands, however MubR5 does not structurally resemble a lectin fold [314-315]. High structural similarity of the B1 has been observed to the Ig-binding protein L (PpL) from *Peptostreptococcus magnus* (PDB entry 1HEZ) [316]. The B1 domain is also structurally similar to the B2 domain, which has a modified ubiquitin-like  $\beta$ -grasp fold, where the two antiparallel  $\beta$ -strands in the  $\beta$ -sheet are connected by a  $\beta$ -strand instead of an  $\alpha$ -helix as in the B1 domain.

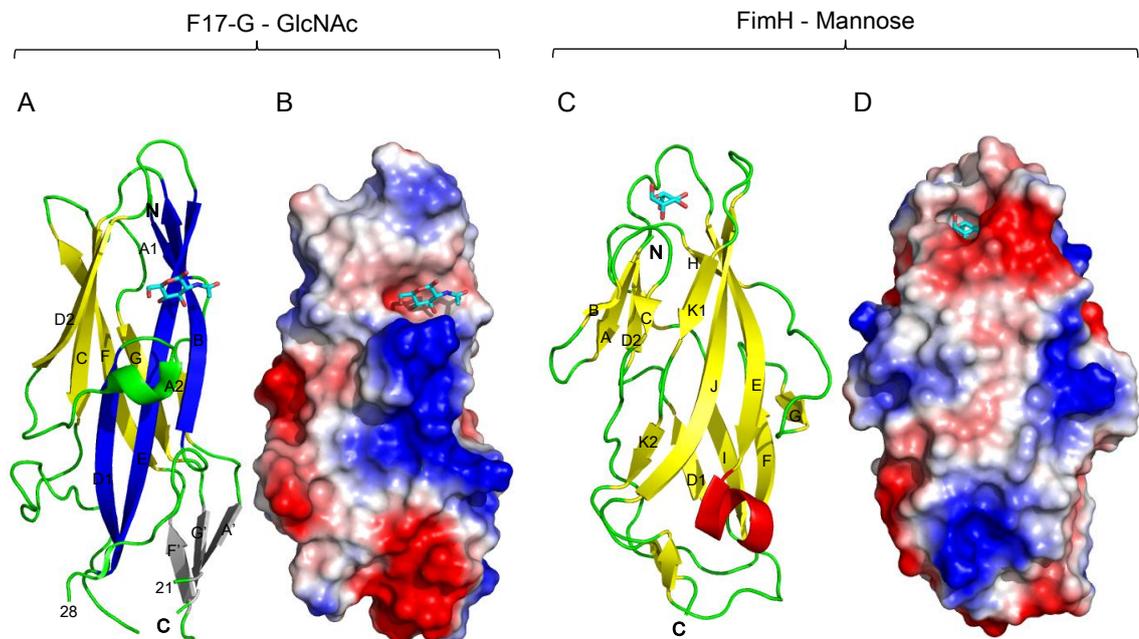


**Figure 1.17 MUB, MucBP or related crystal structures of bacterial cell-surface proteins**

(A) Mub-R5 of *L. reuteri* ATCC 53608 with B1 and MucBP containing B2 domain (PDB entry 3I57). (B) Functionally characterised MucBP domain Spr1375 of *Strept. pneumoniae* (PDB entry 3NZ3). Annotated MucBP domains (C) LBA1460 of *L. acidophilus* (PDB entry 3Q69), PEPE\_0118 of *Pediococcus pentosaceus* (PDB entry 3LYY) (E) and Imo0835 of *L. monocytogenes* (PDB entry 2KVZ). (D) MucBP domain of an adhesion exoprotein from *P. pentosaceus* (PDB entry 2KYW). (G) Structurally related InIB B-repeat of *L. monocytogenes* (PDB entry 2K5B). Structures superimposed on common framework. (taken from [22]).

The B2 domain, exclusive of the IR domain, coincides with the MucBP domain as defined in the Pfam database (PF06458), which, as evident from the MubR5 crystal structure, describes a structurally distinct protein domain (see 1.5). The functionally characterised MucBP domain of the Spr1345 protein (PDB entry 3NZ3) from *Strep. pneumoniae*, which exhibits mucin and polysaccharide binding ability, is structurally similar to the B2 domain of MubR5 (Figure 1.17 B) [240]. Other 3D structures annotated as MubBP have been predicted as cell-surface adhesins including PEPE\_0118 of *Pediococcus pentosaceus* (PDB entry 3LYY) and LBA1460 of *L. acidophilus*. Furthermore, structural similarity for B2 of MubR5 was observed to the

B-repeat of the cell-surface adhesion protein InIB from *L. monocytogenes*, which has an ubiquitin-like  $\beta$ -grasp fold most similar to small ubiquitin-like modifiers (SUMOs) (see 1.4.1) [317].

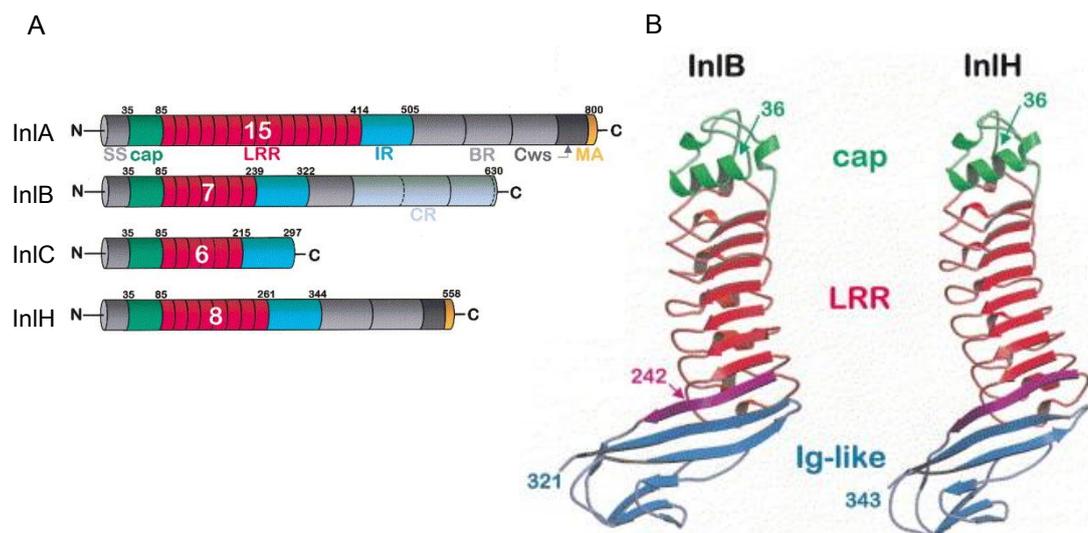


**Figure 1.18 Presentation of the crystal structures of the *E. coli* adhesins F17-G and FimH**

(A) Ribbon diagram of F17-G (PDB entry 1O9W) with the  $\beta$ -sheets of the Ig-like lectin domain shown in yellow (back sheet) and blue (front sheet) and minor  $\beta$ -strands shown in grey. Bound GlcNAc presented in light blue, and  $\beta$ -strands, N- and C-termini labelled. Residues 22 to 27 are missing in structure file (present residues labelled). (B) Electrostatic surface representation of F17-G in complex with GlcNAc with negative and positive charged surface area presented in red and blue, respectively. (C) Ribbon diagram of the N-terminal receptor-binding domain of FimH (PDB entry 1KLF) in complex with mannose with  $\beta$ -strands in yellow.  $\beta$ -strands, N- and C-termini labelled. (D) Surface representation of the FimH receptor-binding domain in complex with mannose with negatively and positively charged surface area presented in red and blue, respectively.

Pathogenic adhesins are generally better functionally and structurally characterised, especially cell-surface adhesion molecules of pathogenic *E. coli* [318-319]. The structure of the F17-G adhesin (PDB entry 1O9W) has been solved in the presence of a GlcNAc molecule allowing the identification of the receptor binding pocket located at the N-terminal end of the protein (Figure 1.18 A and B) [237]. The interaction of amino acid residues of the sugar binding site via their carbonyl groups, side chains or nitrogens with GlcNAc is mediated by 11 hydrogen bonds, four of which involving water

molecules [237]. The F17-G adhesin fold is a variant of the Ig-like fold, which is characterised by a  $\beta$ -sandwich of 7  $\beta$ -strands arranged in two  $\beta$ -sheets of 4 and 3  $\beta$ -strands. [320]. The  $\beta$ -sandwich present in F17-G consists of a 5-stranded sheet (A2, G, F, C and D2) and a 4-stranded sheet (A1, B, E and D1) (Figure 1.18 A). Additional minor  $\beta$ -strands (A', F' and G') extend those of the  $\beta$ -sandwich. The mannose-binding adhesin FimH (PDB entry 1KLF) of *E. coli* contains two domains, an N-terminal receptor-binding domain and a C-terminal pilin domain. While the latter has an Ig-like fold, the receptor-binding domain is an 11-stranded  $\beta$ -barrel (Figure 1.18 C) [235]. The mannose molecule is located in the negatively charged binding pocket at the N-terminal end of the FimH receptor-binding domain (Figure 1.18 D) and ligand interaction is mediated by hydrogen bonding and hydrophobic interactions.



**Figure 1.19 Examples of internalins with Ig-like fold domains**

(A) Domain organisation of *L. monocytogenes* internalins with signal sequence (SS), cap domain (green), characteristic LRR (Leucine rich repeat) (red), interrepeat-region (IR) (blue), B-repeat (BR), C-repeat (CR), cell-wall spanning region (Cws) and membrane anchor (MA) (orange). (B) X-ray crystal structures of InIB and InIH with Ig-like fold IR domains (blue), LRR (red) and cap (green) domains with the N-terminal  $\beta$ -strand of the IR domain coloured magenta (adapted from [321]).

Additionally, several InI proteins of the invasive pathogen *L. monocytogenes* contain protein domain adjacent to their main Leucine rich (LRR) effector domain known as inter repeat (IR) domains (Figure 1.19 A). They describe an Ig-like fold dominated by a 4-stranded  $\beta$ -sheet, where the N-terminal strand is an extension of the C-terminal

strand of the LRR domain (Figure 1.19 B) [321]. While the receptors recognised by the LRR domains of the two major invasive Inls, Inl A and B, are known, no functional information is available for the Ig-like IR domains (see 1.4.1) [212, 322-323].

## 1.6 Aims and objectives of this research project

The overall aim of this research project is to gain structural insight into mucus adhesins of *Lactobacillus reuteri*, to investigate their adhesive properties to intestinal mucus and mucins, and to functionally characterise their mechanism of adhesin binding to structural mucus/ mucin components.

The specific objectives of this project are:

- to purify and characterise mucus adhesins from *L. reuteri* ATCC 53608 and *L. reuteri* DMS 20016<sup>T</sup>
- to obtain structural information on protein domains of mucus adhesins using X-ray crystallography and to investigate their domain organisation by small angle X-ray scattering (SAXS)
- to determine the binding ability of *L. reuteri* adhesins to intestinal mucus, mucins and mucin glycans *in vitro* using binding assays, isothermal titration calorimetry (ITC) and X-ray crystallography

## CHAPTER 2 MATERIAL AND METHODS

### 2.1 Materials

#### 2.1.1 Standard buffers

Deionised ultrapure water was used to a resistance of 18.0 M $\Omega$ /cm (Barnstead Nanopure Diamond, Barnstead Thermolyne Corporation, New Hampshire, USA). All used chemicals were supplied by Sigma-Aldrich (Poole, UK) unless otherwise stated.

Standard buffers used in this study were phosphate buffered saline (PBS) (8.1 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, 2.7 mM KCl, 137 mM NaCl, pH 7.4), PBST (PBS with 0.05% Tween-20), sodium phosphate buffer (10 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.5) and Tris-HCl (10 mM Tris base (Formedium, Hunstanton, UK), pH 7.5).

Pierce Protein-Free (PBS) blocking buffer (Thermo Scientific, Hemel, UK) was used as a blocking agent, if not otherwise stated.

#### 2.1.2 Sugars and mucins

The sugars used in this study for crystallisation (see 2.4.1.2) or binding experiments (see 2.3.6) were Fucose (Fuc), Mannose (Man), Galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc) and *N*-acetylneuraminic acid (Neu5Ac) (all supplied by Sigma-Aldrich, Poole, UK), *N*-acetyllactosamine (LacNAc) (Dextra Laboratories, Reading, UK) and 6'-Sialyllactose (6'SL) (Glycom, Kgs. Lyngby, Denmark). Mannose- $\alpha$ 1-6-mannose (Man $\alpha$ 1-6man) and Mannose- $\alpha$ 1-3(Mannose $\alpha$ 1-6)mannose (Man(Man $\alpha$ 1-6)man) used for isothermal titration calorimetry (ITC) (see 2.4.1) experiments were supplied by (Dextra Laboratories, Reading, UK).

Porcine small intestinal mucus (PSIM) [324] and porcine gastric mucin type III (PGM) (Sigma-Aldrich, Poole, UK) (further purified by ethanol-precipitation (pPGM)), were used in membrane, slot-blot and microtitre plate binding studies (see 2.3.6).

Human MUC2, purified from urine samples of patients with artificial bladder derived from intestinal tissues [325-326], was kindly provided by Michael McGuckin (Mater

Medical Research Institute, South Brisbane, Australia) and used in microtitre plate assays (see 2.3.6.4).

Human MUC2 from biopsy samples for membrane binding assays (see 2.3.6.1) and MUC1-glycopeptides (with a mouse immunoglobulin G (IgG) Fc fusion tag) carrying the specific glycan antigens Tn (GalNAc $\alpha$ ), sialyl-Tn (Sia $\alpha$ 2-6GalNAc $\alpha$ ), TF (Gal $\beta$ 1-3GalNAc $\alpha$ ) or sialyl-TF (Sia $\alpha$ 2-3Gal $\beta$ 1-3GalNAc $\alpha$ ) [327-328], used in membrane binding assays (see 2.3.6.1), were kindly provided by Gunnar Hansson (University of Gothenburg, Sweden).

### **2.1.3 Lectins and antibodies**

All lectins (*Ricinus communis* agglutinin I (RCA), Wheat germ agglutinin (WGA), *Sambucus nigra* agglutinin (SNA) and *Ulex europaeus* agglutinin (UEA)) used in this study were supplied by Vector Labs (Peterborough, UK) as fluorescein (f-) labelled or biotin (b-) conjugated proteins.

Primary antibodies and secondary antibody-conjugates used in this study for protein detection on membrane or in microtitre plate assays are listed below (Table 2.1). Polyclonal rabbit antibodies were raised against recombinant Mub type proteins MubR5 and MubRI, and Lar0958, and were produced by Bio Genes (Berlin, Germany) at a titre of > 200,000 (MubR5 and MubRI) or > 100,000 (Lar0958).

Antibody	Target	raised in	Supplier
Anti-MubR5	Recombinant Mub type 2 repeat RI	rabbit	BioGenes (Berlin, Germany)
Anti-MubRI	Recombinant Mub type 2 repeat RI	rabbit	BioGenes (Berlin, Germany)
Anti-Lar0958	Recombinant Lar_0958 repeat	rabbit	BioGenes (Berlin, Germany)
Anti-His <sub>5</sub>	Five consecutive histidine residues	mouse	Merck KGaA (Darmstadt, Germany)
Anti-MUC2C3	C-terminal human MUC2 peptide PHYVTFDGLYYSYGNC <sup>a</sup>	rabbit	Provided by Gunnar Hansson (University of Gothenburg, Sweden)
Anti-Muc2.3	murine Muc2 peptide CPEDRPIYDEDLKK	rabbit	Provided by Michael McGuckin (Mater Medical Research Institute, South Brisbane, Australia)
Anti-rabbit-HRP <sup>b</sup>	rabbit IgG	donkey	Amersham Bioscience, GE Healthcare (Little Chalfont, UK)
Anti-mouse- HRP <sup>b</sup>	mouse IgG	goat	Invitrogen, Life technologies Ltd (Paisely, UK)
Anti-rabbit-AP <sup>c</sup>	rabbit IgG	goat	Sigma-Aldrich (Poole, UK)
Anti-mouse-AP	mouse IgG	horse	Vector Lab (Peterborough, UK)
Anti-rabbit- AlexaFluor®555	rabbit IgG	donkey	Abcam (Cambridge, UK)

**Table 2.1 Used primary and secondary antibodies**

<sup>a</sup>[25], <sup>b</sup>Horseradish peroxidase (HRP) and <sup>c</sup>Alkaline phosphatase (AP)

#### 2.1.4 Bacterial strains, media composition and culture conditions

Bacterial strains used in this study were *Lactobacillus reuteri* ATCC 53608 for MUB protein isolation, *L. reuteri* DSM 20016<sup>T</sup>, *Escherichia coli* DH5 $\alpha$  (Novagen, Merck KGaA, Darmstadt, Germany) for transformation and plasmid amplification, and *E. coli* Tuner(DE3)pLacI<sup>2</sup> (Novagen, Merck KGaA, Darmstadt, Germany), *E. coli* BL21 (DE3) (Novagen, Merck KGaA, Darmstadt, Germany) and *E. coli* BL21Star<sup>TM</sup> (DE3) (Novagen,

Merck KGaA, Darmstadt, Germany) for recombinant protein expression. The Tuner(DE3)pLacI<sup>2</sup> strain is a lacZY deletion mutant of *E. coli* BL21 and carries a chloramphenicol resistance gene. *E. coli* BL21Star™ (DE3) strains harbouring the pET101-InJ or pET101-InJ-LRR-IR plasmids, were kindly provided by Pascale Cossart (Institute Pasteur, Bacteria-cell interactions Unit, Paris, France).

Bacterial culture media used in this study were modified standard Luria broth (LB), 'de Man, Rogosa and Sharpe' (MRS), and *Lactobacillus* defined media (LDMII) (see Appendix I). For *E. coli* bacterial cell growth, LB media or LB agar was used containing 10 mg/mL Bacto tryptone, 5 mg/mL Bacto yeast extract and 10 mg/ml sodium chloride, or additionally 15 g/L agar, and the pH was adjusted to pH 7.5. *L. reuteri* cultures were grown in MRS broth containing 10 mg/mL peptone, 8 mg/mL 'Lab-Lemco', 4 mg/mL yeast extract, 20 mg/ml glucose, 2 mg/ml di-potassium hydrogen phosphate, 5 mg/mL sodium acetate 3H<sub>2</sub>O, 2 mg/ml tri-ammonium citrate, 0.2 mg/mL magnesium sulphate 7H<sub>2</sub>O, 0.05 mg /mL manganese sulphate 4H<sub>2</sub>O and 1 mL sorbitan mono-oleate. Super Optimal broth with Catabolite repression (SOC) medium (20 mg/mL Bacto tryptone, 5 mg/mL Bacto yeast extract, 10 mM NaCl, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, 20 mM glucose) was used for cell regeneration after bacterial transformation (see 2.2.5).

Long term stocks of *L. reuteri* in MRS with 20% (v/v) glycerol and of *E. coli* in LB medium with 50% (v/v) glycerol were stored at -80°C. *L. reuteri* ATCC 53608 cells were grown to stationary phase from long term stocks at 37°C for 20 h in a static incubator. For native MUB protein isolation, cells were sub-cultured at 0.1% (v/v) into LDMII medium (see Appendix I) at 37°C for 24 h in a static incubator. *E. coli* strains were cultured from long term stocks at 37°C for 16 h in a shaker incubator in LB broth with 1% (w/v) glucose or on LB agar supplemented with antibiotics, carbenicillin at 50 µg/mL, chloramphenicol at 34 µg/mL or kanamycin at 30 µg/mL, as needed.

## 2.2 Molecular biology

### 2.2.1 Polymerase chain reaction (PCR)

Amplification of target gene fragments (*mubRV*, *mubRVI*, *mubR8-V*, *mubRV-VI*, *Nterm*, *NtermmubRI* and *lar0958*) for vector cloning was achieved by polymerase chain reaction (PCR) with whole bacterial cells as a template. *L. reuteri* ATCC 53608 and DSM 20016<sup>T</sup> cells were grown in MRS broth to stationary phase (see 2.1.4), cells pelleted by centrifugation (1342×g, 5 min, 15°C), washed twice in ultrapure water and re-suspended at OD<sub>600</sub> 5.5 in ultrapure water. For whole cell PCR, 10 µL cell suspension and 10 pmol gene specific forward and reverse primers (see Appendix II) were added to the HotStarTaq Master Mix with 125 U HotStarTaq DNA polymerase (Qiagen, Crawley, UK) and the reactions carried out in a Biometra® T gradient thermocycler (Biometra, Goettingen, Germany) with the following reaction profile: initial polymerase activation for 5 min at 95°C; 35 cycles of denaturation for 30 sec at 94°C, annealing for 30 sec at 56°C (*Nterm* and *mubRVI*), 50°C (*NtermmubRI*), 49°C (*mubRV* and *mubRV-VI*) or 51°C (*mubR8-V*), and extension 180 sec (*mubNterm* and *mubNtermRI*), 45 sec (*mubRV* and *mubRVI*) and 75 sec for (*mubR8-V* and *mubRV-VI*) at 72°C; and a final extension for 10 min at 72°C.

Amplified DNA was separated and analysed by agarose gel electrophoresis (see 2.2.3) and purified using a QIAquick® PCR purification kit (Qiagen, Hilden, Germany). Alternatively, DNA fragments were excised from gel and purified using a QIAquick® gel extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

### 2.2.2 Plasmid DNA purification

Plasmid DNA was purified from *E. coli* cells grown in LB broth for 16 h (see 2.1.1) using a QIAprep® Spin Miniprep Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, and purified DNA was eluted with sterile ultrapure water for use in sequencing or restriction analysis, or by a TENS buffer DNA precipitation method (40 mM Tris base, 1 mM ethylenediaminetetraacetic acid (Na<sub>2</sub>EDTA), 0.1 N sodium hydroxide and 0.5% sodium dodecyl sulphate). For this, *E. coli* cells were separated from culture media by centrifugation at 16.200×g for 10 sec and the supernatant decanted. TENS buffer (300 µL) and 3 M sodium acetate (pH 5.2) (150 µL) were then added for cell lysis, and cell debris and chromosomal DNA separated by

centrifugation at 16,200×g and 4°C for 3 min. The supernatant was transferred to a clean tube, plasmid DNA was precipitated with ice-cold absolute ethanol and pelleted by centrifugation at 16.200×g for 5 min (4°C). The DNA pellet was rinsed with 70% (v/v) ethanol (3 min, 16.200×g, 4°C), dried at RT and re-suspended in Tris ethylenediaminetetraacetic acid (EDTA) (TE) buffer (40 mM Tris base, 1 mM Na<sub>2</sub>EDTA, pH 7.7-8.0) with 2.5-5 U RNase ONE™ Ribonuclease (Promega, Madison, USA).

### **2.2.3 DNA agarose gel electrophoresis**

Agarose gel electrophoresis was used for DNA separation after PCR (see 2.2.1), plasmid purification (see 2.2.2) or restriction digest (see 2.2.6) with 1% (w/v) agarose gels. DNA samples were prepared by adding Orange G loading buffer (50 mM EDTA Na<sub>2</sub>, pH 7.5; 70% (w/v) sucrose; 0.1 % (w/v) Orange G dye) at 60% (v/v) before loading onto the gels. Gels were run in Tris acetate EDTA (TAE) buffer (40 mM Tris, 2 mM sodium acetate 1 mM Na<sub>2</sub> EDTA, pH 7.7-8.0) for 30 min at 100 V and stained with ethidiumbromide (EtBr) for up to 30 min, rinsed with water and DNA was detected under UV light. The Gibco 1kb DNA ladder and the Gibco 100 bp ladder (Invitrogen, Life Technologies Ltd., Paisley, UK) were used as DNA standards.

### **2.2.4 DNA cloning in expression vectors**

Cloning of PCR amplified *mubRV*, *mubRVI*, *mubR8-V*, *mubRV-VI* and *lar0958* DNA fragments (see 2.2.1) into pETBlue-1 AccepTor™ (Novagen, Merck KGaA, Darmstadt, Germany) was achieved by direct ligation of linearised vector DNA with single 3'-dU overhangs and PCR products with single 3'-dA overhangs produced by PCR using Taq DNA polymerase (see 2.2.1). For ligation, insert and vector DNA were combined in a 5:1 or 10:1 molar ratio and added to a Clonables™ 2× ligation premix (Novagen, Merck KGaA, Darmstadt, Germany) and the ligation reaction incubated at 16°C for 2 h. The In-Fusion™ PCR cloning kit (Clontech, Mountain View, California, USA) was used for cloning of *nterm* and *ntermmubRI* DNA into the pOPINF vector by ligation-independent fusion of vector and gene insert (see Appendix III) [329]. PCR primers (see Appendix II) for amplification of *nterm* and *ntermmubRI* were designed to incorporate 15 bp sequence-extensions with homology to the pOPINF vector to facilitate DNA insertion via homologous recombination using an In-Fusion™ enzyme

and linearised vector DNA (Clontech, Mountain View, California, USA) according to the manufacturer's instructions. The pOPINF was linearised in a single restriction digest with the restriction endonucleases HindIII (Roche, Mannheim, Germany) and KpnI (Roche, Mannheim, Germany) (see 2.2.6).

The pETBlue-1 AccepTor™ and the pOPINF vectors allow high level protein expression under the control of a *T7lac* promoter upon induction with IPTG, when cloned into the expression strains *E. coli* Tuner (DE3)pLacI<sup>2</sup> and *E. coli* BL21(DE3), respectively. The pOPINF vector facilitates the expression of target proteins with an N-terminal fusion tag of six consecutive histidine (His) residues cleavable by 3C peptidase. Both vectors encode an ampicillin resistance gene facilitating bacterial growth in the presence of carbenicillin (*E. coli* BL21(DE3) pOPINF) or carbenicillin and chloramphenicol (*E. coli* Tuner(DE3)pLacI<sup>2</sup> pETBlue-1 AccepTor™) (see 2.1.4).

### **2.2.5 Bacterial transformation**

For heat shock transformation of *E. coli* DH5 $\alpha$ , *E. coli* Tuner(DE3)pLacI<sup>2</sup> and *E. coli* BL21(DE3) (see 2.1.4), cells were thawed on ice and 45  $\mu$ L of the cell suspension were mixed with vector DNA (see 2.2.4) (2.5  $\mu$ L In-Fusion reaction or 1  $\mu$ L plasmid ligation product), and incubated on ice for 5 min. The heat shock was performed in a 42°C water bath for 30 sec, followed by 2 min incubation on ice. For cell regeneration, 250  $\mu$ L SOC medium (see 2.1.4) were added to the transformation reaction and the cell suspension was incubated at 37°C for 1 h. Cells were then plated on LB-agar (see 2.1.4) with antibiotics and incubated O/N at 37°C.

### **2.2.6 Recombinant DNA analysis**

Clones were tested for the presence of the target genes by whole cell PCR using bacterial colony scrapes as a template (see 2.2.1) and by restriction analysis of plasmid DNA. For restriction digest, plasmid DNA were purified as described earlier (see 2.2.2) and incubated with HindIII (Roche, Mannheim, Germany) and KpnI (Roche, Mannheim, Germany) at 37°C for 6 h in restriction buffer A (pOPINF) or with XbaI (Roche, Mannheim, Germany) and EcoRI (Roche, Mannheim, Germany) at 37°C for 2 h in restriction buffer H (pETBlue-1 AccepTor). The integrity of the cloned sequences

was verified by automated DNA sequencing (Genome Enterprise Ltd. BBSRC Genome Analysis Centre, Norwich, UK; or Eurofins MWG operon, Ebersberg, Germany).

### **2.2.7 Estimation of DNA concentration**

DNA concentrations were estimated using ND-1000 spectrophotometer (Thermo Scientific, Waltham, USA), where an absorbance of 1 at 260 nm equalled 100 ng/ $\mu$ L, or after agarose gel electrophoresis, when compared to DNA standard.

## **2.3 Biochemistry**

### **2.3.1 Recombinant protein production**

#### **2.3.1.1 Protein expression**

For protein expression, cells were sub-cultured at 3% (v/v) from an O/N starter culture (see 2.1.4) into fresh LB media (500 mL in 2 L flasks with chicanes) and grown to an OD<sub>600</sub> of 0.6-0.8 and recombinant proteins expressed in *E. coli* strains Tuner(DE3)pLacI<sup>2</sup>, BL21(DE3) or BL21Star™ (DE3) after induction with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) at 37°C for 3-5 h or at 30°C overnight (MubR8-V). Bacterial cells were harvested by centrifugation at 12,250  $\times$ g and cells stored at -20°C until further use.

#### **2.3.1.2 Protein extraction**

Recombinant proteins were extracted using a freeze-thaw method (10 min freezing and 20 min thawing) (MubR5, -RI, RV, -RV-VI and -R8-V, Lar0958), using BugBuster® HT Protein Extraction Reagent (Novagen, Merck KGaA, Darmstadt, Germany) (Nterm, NtermMubRI) for cell lysis according to the manufacturer's instructions or via ultrasonication for cell disruption (Mub-RI-II-III, InIJ, InIJ-LRR-IR). For this, bacterial cell suspensions were ultrasonicated using a Status 70 MS72 homogeniser (Philip Harris Scientific; Ashby-de-la-Zouch, UK) at 50% power for 3  $\times$  30 sec bursts with 60 sec incubation on ice in between bursts or a Soniprep 150 homogeniser (MSE; Sanyo, London, UK) set at 6  $\mu$ m amplitude for 8  $\times$  15 sec bursts with 30 sec cooling on ice. Protein extracts were clarified at 13,500  $\times$ g for 20 min at 4°C before further purification.

### 2.3.1.3 Protein purification

Non-tagged Mub-repeat proteins and Lar0958 protein extracts were loaded onto an ion exchange chromatography (IEC) column (Mono Q HR 10/10) (GE Healthcare, Little Chalfont, UK) equilibrated with 10 mM sodium phosphate buffer and proteins were eluted with a linear gradient of 10 mM sodium phosphate buffer containing 1 M NaCl using a AKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare, New Jersey, USA).

His-tagged fusion proteins (Nterm, NtermMubRI, InIJ and InIJ-LRR-IR) were purified by immobilised metal affinity chromatography (IMAC) with an AKTA FPLC system using a HisTrapHP column (GE healthcare, Uppsala, Sweden) (NtermMubRI, Nterm) or by gravity flow using a HisBind resin column (Novagen, Madison, USA) (InIJ and InIJ-LRR-IR) according to the manufacturer's instructions. Protein extracts were loaded onto a HisTrapHP column pre-equilibrated with 20 mM sodium phosphate wash buffer (20 mM Na<sub>2</sub>HPO<sub>4</sub>, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.5 M NaCl, 20 mM imidazole; pH 7.5) and bound proteins were eluted step-wise with 20 mM sodium phosphate elution buffer (20 mM Na<sub>2</sub>HPO<sub>4</sub>, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.5 M NaCl, 500 mM imidazole; pH 7.5). HisBind purification was performed according to the manufacturer's instruction with the exception of a sequential column wash with two different wash buffers (20 mM Tris-HCl, 500 mM NaCl; pH. 7.9) containing 20 mM or 40 mM imidazole. Bound proteins were eluted with elution buffer (20 mM Tris-HCl, 500 mM NaCl, 1 M imidazole: pH. 7.9).

When needed, pooled protein fractions were further purified via size exclusion chromatography (SEC) on a Superdex 75 16/60 prep grade or a Superdex 200 16/60 prep grade (GE Healthcare, Little Chalfont, UK) (only MubRI-II-III, InIJ and InIJ-LRR-IR) with PBS or 10 mM Tris-HCl containing 150 mM NaCl (pH 7.5) at a flow rate of 1 ml/min. For calibration of SEC columns proteins from standard high and low molecular weight (LMW and HMW) gel filtration calibration kits (Amersham Pharmacia Biotech, Little Chalfont, UK), namely Ribonuclease A (13.7 kDa), Chymotrypsinogen A (25 kDa), Carbonic anhydrase (29 kDa), Ovalbumin (43 kDa), Albumin (67 kDa), Alcohol dehydrogenase (150 kDa) or Aldolase (158 kDa), were used. For use in X-ray crystallisation studies (see 2.5.1) or small angle X-ray scattering experiments (see 2.5.2), proteins were dialysed twice in 10 mM sodium phosphate buffer or 10 mM

Tris-HCl at 4°C for a minimum of 4 h using a 3,500 molecular weight cut-off (MWCO) Spectra/Por dialysis tube (Spectrum, Breda, The Netherlands). Proteins were then concentrated to at least 10 mg/ml using centrifugal filter units (10,000 or 3,000 MWCO), immediately frozen in liquid nitrogen and stored at -80°C till further use.

### **2.3.2 MUB purification**

The full-length native MUB was purified from spent media of a *L. reuteri* ATCC 53608 culture. Briefly, bacterial cells were grown till stationary phase in LDMII and separated from spent medium at 7,500×g for 15 min at 4°C. Medium solution was further clarified by vacuum-filtration using 0.45 µm and 0.2 µm filters consecutively and then concentrated by tangential flow filtration using Vivaflow 200 cassettes (100,000 MWCO PES) (Vivascience AG, Hannover, Germany). The concentrated extract was dialysed twice in 4 L PBS at 4°C for at least 4 h using a 3,500 MWCO Spectra/Por membrane (Spectrum, Breda, The Netherlands), filtered using 0.45 µm Ultrafree-CI spin columns (Millipore, Merck KGaA, Darmstadt, Germany) and concentrated in 100,000 MWCO spin concentrators (Sartorius, Surrey, UK). The MUB protein solution was purified by size exclusion chromatography (SEC) using a Superose 6 prep grade resin column (GE Healthcare, Little Chalfont, UK) equilibrated with PBS at a flow rate of 0.4 mL/min using an AKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare, New Jersey, USA).

### **2.3.3 Estimation of protein concentration**

Protein concentrations were estimated by measuring the absorbance (Abs) of 1 to 2 µL sample at 280 nm in a NanoDrop 2000 or ND-1000 spectrophotometer (Thermo Scientific, Waltham, USA) according to the Beer-Lambert's law [ $A_{280} = \epsilon \cdot l \cdot c \cdot MW$ ; with  $\epsilon$  = extinction coefficient,  $c$  = molar concentration in mol/L,  $MW$  = molecular weight in g/mol, and  $l$  = optical path length in cm). The protein extinction coefficient and the Abs 0.1%, absorbance value, where the protein concentration equals 1 mg/mL, were calculated from amino acid sequence using the ExPasy protein parameter tool (<http://web.expasy.org/protparam/>).

### **2.3.4 Protein gel electrophoresis**

All electrophoresis gels, running buffers, sample buffers and Western transfer buffer were supplied by Invitrogen, Life Technologies Ltd. (Paisley, UK) unless otherwise stated.

#### **2.3.4.1 Polyacrylamide gel electrophoresis (PAGE)**

For reducing denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of recombinant protein samples, NuPAGE® 4-12% Bis Tris gels (12 well, 1.0 mm) or RunBlue 12% (Expedeon, Harston, UK) (17 wells, 1.0 mm) were used with 2-(N-morpholino)ethanesulfonic acid (MES) (35 min at 200 V) or 3-(N-morpholino)propanesulfonic acid (MOPS) SDS running buffer (55 min at 200 V) for proteins with molecular weights (MW) of  $\leq 50$  kDa or  $> 50$  kDa, respectively. For the MUB protein, NuPAGE® 4-8% Tris-acetate gels (15 well, 1.0 mm) gels (55 min at 150 V) were used with Tris-Acetate SDS running buffer. Samples were prepared with lithium dodecyl sulphate sample buffer and 10 mM dithiothreitol (DTT) as a reducing agent, and incubated for 10 min at 70°C before loading onto the gels. The Broad Range Protein Molecular Weight Marker (10-225 kDa) (Promega, Southampton, UK), Broad Range Prestained Protein Standard (7-175 kDa) (New England Biolabs, Ipswich, UK) and HiMark™ Unstained High Molecular Weight Protein Standard (30-460 kDa) (Invitrogen, Life Technologies Ltd., Paisley, UK) were used as MW standards.

For non-denaturing PAGE of MUB, NuPAGE® 4-16% Tris-acetate native gels (12 well, 1.0 mm) were used with Tris-glycine running buffer and samples separated at 150 V for 2 h after adding Tris-glycine sample buffer. The Native Mark™ Protein Standard was used as a MW standard.

#### **2.3.4.2 Agarose-polyacrylamide composite gel electrophoresis (AgPAGE)**

For Agarose-polyacrylamide composite gel electrophoresis (AgPAGE) of high molecular weight mucus and mucin samples, agarose-polyacrylamide gels were prepared by mixing two heated solutions A (1% agarose, 0.375 M Tris-HCl (pH 8.1), 15% acrylamide/bis-acrylamide with (w/w) ratio of 19:1, 10% (v/v) glycerol) and B (0.5 g agarose, 0.375 M Tris-HCl (pH 8.1)). The gels with gradients of 0.5-1% agarose, polyacrylamide 0-6% and 0-10% glycerol were cast at 60°C using a gradient mixer

after adding 3  $\mu$ L APS and TEMED. Gels were left to dry for 3 h at RT and stored at 4° C in a humidified environment for up to a week.

Samples were reduced and alkylated with 2 $\times$  sample loading buffer (0.75 M Tris-HCl (pH 8.1), 60% glycerol, 0.01% bromphenyl blue and 12% SDS) and DTT at a final concentration of 200 mM for 2 h at 37°C, and stored at -20°C till further use. Prior to loading onto gels, samples were boiled at 90°C for 5 min and AgPAGE performed with composite running buffer (192 mM boric acid, 1 mM EDTA, 0.1% SDS; pH 7.6 adjusted with 2 M Tris base) at 4°C and 30 mA per gel for 6.5 h or 12 mA per gel for 16 h. Gels were fixed (see 2.3.4.4) and stained using the Colloidal Blue staining kit according to the manufacturer's instructions or with alcian blue (see 2.3.4.4).

#### **2.3.4.3 Isoelectric focusing (IEF)**

For the determination of the isoelectric point (pI) of proteins, Novex® pH 3-7 IEF gels were run in Novex® Cathode pH 3-7 and Novex® Anode buffer (Invitrogen, Life Technologies Ltd., Paisley, UK) at run conditions of 100 V for 1h, 200 V for 1h and 400 V for 30 min. Samples were prepared with Novex® IEF Sample buffer and the IEF Markers pH 3-10 SERVA liquid mix (Invitrogen, Carlsbad, USA) used as a standard for pI determination of analysed protein samples. Gels were fixed in 12% trichloroacetic acid (TCA) (v/v) for 30 min and stained with Colloidal Blue staining kit (see 2.2.7.2).

#### **2.3.4.4 Staining of gels**

For staining of SDS-PAGE gels, the Colloidal Blue staining kit (Invitrogen, Life Technologies Ltd., Paisley, UK) or the GelCode Blue Stain Reagent (Thermo Scientific, Hemel Hempstead, UK) were used according to the manufacturer's instructions. For Colloidal Blue staining, gels were fixed with 50% (v/v) methanol or ethanol and 10% (v/v) acetic acid for 10 min.

For alcian blue staining of AgPAGE gels, gels were first fixed with 50% (v/v) methanol and 1% (v/v) acetic acid for 1h and then equilibrated in 25% (v/v) ethanol and 10% (v/v) acetic acid (2  $\times$  15 min). Gels were incubated in 0.125% alcian blue solution (25% (v/v) ethanol and 10% (v/v) acetic acid) until sufficiently stained. After Western blotting of AgPAGE onto PVDF membranes, membranes were briefly rinsed with methanol (absolute) and stained with alcian blue solution for 10 min after. Membranes were destained in methanol for 5 $\times$  2 min.

Native gels were stained with Gelcode Blue Stain Reagent and not destained before Western blotting.

A GS-800 calibrated densitometer (Bio-Rad, Hertfordshire, UK) or an Alphamager FluorChemE (Protein Simple, Santa Clara, USA) were used for scanning or imaging gels.

Before Western blotting, gels were destained in 50% (v/v) ethanol and 10% (v/v) acetic acid for several hours. Destained SDS-PAGE gels were then incubated in SDS running buffer containing 1% (w/v) SDS for 1 h, Native PAGE gels in 0.1% (w/v) SDS for 1 min and composite gels in 10 mM Tris-HCl (pH 7.6) containing 1% (w/v) SDS for 1 h.

#### **2.3.4.5 Western blotting**

For Western electroblotting, recombinant purified proteins were transferred onto Pplyvinylidene fluoride (PVDF) Immobilon™-P membrane (Millipore, Watford, UK) of 0.45 µm pore size, native MUB onto nitrocellulose Hybond™-C extra membrane (Amersham Bioscience, GE Healthcare, Little Chalfont, UK) of 0.45 µm pore size, using an XCell™ Blot Module (Invitrogen, Life Technologies Ltd., Paisley, UK) and Western blot transfer buffer (SDS-PAGE gels) or 25 mM Tris-glycine buffer (25 mM Glycine, pH 9.2) (non-denaturing PAGE gels) according to the manufacturer's instructions.

After AgPAGE (see 2.3.4.2), samples were transferred onto PVDF Immobilon P<sup>SQ</sup> membrane (Millipore, Watford, UK) (0.2 µm pore size) with composite blotting buffer (25 mM Tris-HCl, 0.192 M glycine, 0.04% (w/v) SDS and 20% (v/v) methanol) for 5 h at 4°C and 150 V (40 W).

#### **2.3.5 Protein detection via antibodies and lectins**

After transfer of proteins to a membrane by Western blotting (see 2.3.4.6) or slot-blotting (see 2.3.4.3), membranes were incubated in blocking buffer for at least 1 h, washed three times with PBST and once with PBS and proteins were detected via primary and secondary antibodies or fluorescein labelled lectins (see 2.1.3).

For detection via specific antibodies, membranes were incubated with primary anti-His<sub>6</sub> (1:1,000), anti-MubR5 (1:20,000), anti-MubRI (1:20,000), or a mix of anti-MubR5 and anti-MubRI (1:20,000 each), followed by secondary antibodies conjugated with alkaline phosphatase (AP) (1:25,000) or horse reddish peroxidase (HRP) (1:25,000)

(Table 2.1). Nitroblue tetrazolium (100 µg/mL), 5-bromo-4-chloro-3-indolyl phosphate-toluidine (50 µg/mL) in 100 mM Tris-HCl (pH 9.6) with 4 mM MgCl<sub>2</sub> were used as an AP substrate and incubated on membranes until signals showed. Immobilon™ Western Chemiluminescent HRP Substrate (Millipore, Merck KGaA, Darmstadt, Germany) was added to the membranes and incubated for 5 min before imaging with the Alphascreen FluorChemE (Protein Simple, Santa Clara, USA). All incubation steps were performed at RT for at least 1 h followed by washing steps with PBST (3× 5 min) and PBS (1× 5 min).

For protein detection via lectins, membranes were incubated with fluorescein labelled lectins (1:500 in PBS) for 1 h and washed three times with PBST and once with PBS before measuring the fluorescence signal using a Pharos-FX Plus Molecular Imager (BioRad, Hemel Hempstead, UK).

## **2.3.6 Protein binding assays**

### **2.3.6.1 Membrane protein binding assays after electrophoresis**

To investigate the binding of purified MUB (see section 2.3.2) to MUC1-glycopeptides (see 2.1.3), glycoproteins were transferred onto a nitrocellulose membrane via Western blotting (see 2.3.4.5) after SDS-PAGE (see 2.3.4.1). Membranes were incubated in Blocking buffer (see 2.1.1) for at least 1 h and then probed with MUB at 15 µg/mL for 16 h at RT. Bound proteins were detected via primary anti-MubRI and anti-MubR5, and secondary anti-rabbit-HRP as described before (see 2.3.5).

To remove all bound proteins, the membrane was incubated twice for 10 min with stripping buffer (15 g/L Glycine, 1 g/L SDS, 10 mL/L Tween-20; pH 2.2), followed by PBS (2× 10 min) and PBST (2× 5 min), and MUC1-glycopeptides were detected via f-WGA and f-RCA (see 2.3.5).

In order to investigate the binding of MUB and recombinant adhesion proteins (MubRI-II-III and MubR5) to human MUC2, mucins were separated by AgPAGE and transferred onto PVDV membrane (see 2.3.4.2 and 2.3.4.5). After blocking with 5% dried milk powder in PBST for 2 h at RT and washing with PBST (3× 5 min), the membrane was incubated with MUB (10 µg/mL) and MubR5 (50 or 70 µg/mL) for 2 h. Bound proteins were detected via primary anti-MubR5, anti-MubRI (1:5,000 in PBS) for

16 h at RT (see 2.7), followed by incubation with anti-rabbit-HRP (Pierce, Thermo Scientific, Hemel, UK) (1:2,000 in PBS) for 2 h at RT (see 2.3.5).

### **2.3.6.2 Glycan array**

A nitrocellulose glycan array was kindly provided by William Willats (University of Copenhagen, Denmark), where polysaccharides and BSA-conjugated oligosaccharides, mainly plant cell wall components and their building blocks, were printed on nitrocellulose in duplicate at 1 mg/mL, 0.2 mg/mL and 0.04 mg/mL for non BSA-conjugated structures and 2 mg/mL, 0.4 mg/mL and 0.08 mg/mL for BSA-conjugates. After incubation with Blocking buffer for 1 h, the array was probed with MubRI (0.3 mg/mL), MubR5 (0.3 mg/mL), NtermMubRI (0.45 mg/mL) and MUB (60 µg/mL) for 4 h. Specific detection of bound proteins was achieved by primary anti-MubRI (1:5,000), anti-MubR5 (1:5,000) and anti-His<sub>5</sub> (1:1,000), followed by secondary alkaline phosphatase antibody conjugates (see 2.3.5).

To investigate binding to mammalian glycans, MUB proteins and Lar0958 were tested on mammalian printed arrays of the Consortium for Functional Glycomics (CFG).

### **2.3.6.3 Mass Spectrometry glycan array**

The binding of recombinant MUB proteins and native MUB to sugar molecules (Fuc, Gal, Man, GlcNAc, Lac, 6'SL and 3'sialyllactose (3'SL)) was investigated using functionalised Gold(Au)-surfaces and detection by mass spectrometry (MS) in collaboration with Sabine Flitsch (and Mirja Hartmann) (Manchester Institute of Biotechnology, University of Manchester, UK).

#### **2.3.6.3.1 Functionalisation of Au-surfaces**

Self-assembled monolayers (SAM) were formed on Au-chips by mixing linker (0.476 mM) and spacer (0.328 mM) molecules in a 1:4 ratio O/N at RT. Au-plates were washed between incubation steps with pure ethanol and dried in a nitrogen flow. For functionalisation of SAM, an activation mixture of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (35 mg/mL in DMF) and pentafluorophenol (PEP) (32 mg/mL in DMF) was incubated on Au-surfaces for 2 h at RT, followed by incubation with amino-functionised glycosides (Fuc, Gal, Man, GlcNAc and Lac) (50 mM in PBS) O/N at RT. Sialylated Lac sugars, 6'SL and 3'SL, were formed on gold

chips by enzymatic synthesis with sialyl-transferase (with 1 mM CMP-Neu5Ac in 100mM Tris-HCl (pH 8.0)) and trans-sialidase from *Trypanosoma cruzi* (with 4.3 mg/mL fetuin in 100 mM sodium phosphate buffer (pH 7)) at 25°C and 37°C O/N. Successful assembly and functionalisation of Au-surfaces was confirmed by MALDI-ToF mass spectrometry (Ultraflex MALDI TOF/TOF Mass spectrometer, Bruker Daltonics) in linear positive mode using a 2',4',6'-trihydroxyacetophenone monohydrate (THAP) matrix (10 mg/mL in acetone).

#### **2.3.6.3.2 Protein binding on functionalised Au-glycan chips**

After functionalisation (see 2.3.6.3.1), MubRI-II-III (300 µg/mL), MubR8-V (450 µg/mL), MubRV-VI (2.8 mg/mL), MubRV (1.0 mg/mL), MubR5 (1.0 mg/mL), Nterm (230 µg/mL) and MUB (25 µg/mL) were added in duplicate to functionalised glycan arrays and incubated in PBS for 4 h at 37°C. In addition, incubation of protein on Au-surfaces or on non-functionalised SAM was performed to obtain a MS protein profile and as a negative binding control. Au-chips were washed with PBS and left to dry, before matrix solution (sinapinic acid (20 mg/mL) in acetonitrile and 0.1 % trifluoroacetic acid, ratio 70:30) was applied. Mass spectra were collected in a mass-to-charge-ratio range of 6,000 to 160,000 m/z by MALDI-ToF mass spectrometry (Ultraflex MALDI TOF/TOF Mass spectrometer, Bruker Daltonics) in linear positive mode with a matrix containing sinapinic acid (stock concentration 20 mg/mL) in acetonitrile with 0.1% trifluoroacetic acid (TFA) (70:30 ratio). Data were analysed using the open source mass spectrometry tool mMass (version 5.4.1) (<http://www.mmass.org/>).

#### **2.3.6.4 Slot-blot assay**

Purified MUB proteins were slot-blotted onto a nitrocellulose membrane by loading 200 µL protein solutions in serial dilution into the slots of a PR600 24-slot-blot apparatus (Hoefer, Holliston, USA) with a vacuum of 350 mBar using the mini vacuum pump (Biometra, Goettingen, Germany). Slots were washed twice by adding 1 mL PBS and the vacuum increased to about 700 mBar. In order to achieve a better blotting efficiency, the vacuum was applied for another 2 min, the membrane then removed and incubated in blocking buffer for at least 1 h. For analysis of MUB (0-8 µg), Mub repeats (0-64 µg) and Nterm (0-32 µg) binding to mucin, membranes were incubated with PSIM (0.5 mg/mL) and pPGM (1 mg/mL) for 16 h and bound mucin detected by f-RCA (see 2.3.5). The membrane was then incubated with stripping buffer to remove any

bound protein and MUB proteins detected via primary anti-MubRI, anti-MubR5 or anti-His<sub>5</sub> and secondary HRP antibody conjugate (see 2.3.5).

In order to investigate the potential glycosylation of MUB, membranes were incubated with f-RCA, f-WGA, f-UEA and f-SNA as described in section 2.3.5.

As a control for MUB binding studies with MUC1-glycopeptides (see 2.3.6.1), slot-blotted MUB was incubated with an mouse IgG Fc fragment (Pierce, Thermo Scientific, Rockford, USA) (1 µg/mL) at an equal molar concentration as present in MUC1-glycopeptides for 16 h at RT. Bound IgG Fc was detected via anti-mouse-HRP (see 2.3.5).

### **2.3.6.5 Microtitre plate assays**

For binding of MUB proteins to surface immobilised mucin samples, Microlon 600 polystyrene microtitre plates (Greiner Bio-One Ltd., Stonehouse, UK) were incubated with 200 µL PSIM (0.1 mg/mL) and pPGM (1 mg/mL) in PBS at 4°C for 16 h. To remove any free protein, the plate was washed three times with 200 µL PBST and blocked with Blocking Buffer for 5 h. All incubation steps were performed in PBS at RT followed by a PBST wash step as described above. Native MUB and recombinant Mub-repeat proteins were added to the wells in a concentration range of 0-10 µg/mL and 0-200 µg/mL (100 µL per well), respectively, and plates incubated for 16 h. To detect bound MUB proteins, primary anti-MubRI, anti-MubR5 or a combination of both were added to the wells (100 µL at 1:20,000) and incubated for 2 h, followed by incubation with secondary anti-rabbit-AP (100 µL at 1:20,000) for 2 h. Then, SIGMAFAST p-nitrophenol phosphate (pNPP) AP substrate (100 µL of 1 mg/mL pNPP in 0.2 M Tris with 5 mM MgCl<sub>2</sub>, pH 9.6-10.5; Sigma) was added to the wells and the absorbance measured at 405 nm after 1 h incubation in the dark. Assays were performed in triplicate with coated BSA as a control (1 mg/mL) (Sigma-Aldrich, Poole, Dorset, UK).

Inhibition assays were performed as described above, but after incubation with MUB proteins, 100 µL inhibitor solutions containing 6'SL, Neu5Ac and Lac (see 2.1.2) at concentrations of 25, 50, 75 and 200 mM were added to the wells and incubated for 2 h.

A modified microtitre plate binding assay was used to analyse the binding of MUB, Lar0958 and Internalin J proteins to human MUC2, pPGM and PSIM, where purified adhesion proteins (see 2.3.1 and 2.3.2) were coated onto Microlon 600 polystyrene microtitre plates (Greiner Bio-One Ltd., Stonehouse, UK) at 0.4  $\mu\text{g}$  (Lar0958 and internalins) or 0.6  $\mu\text{g}$  (MUB) per well in PBS at 4°C for 16 h. Excess protein was removed by washing three times with 200  $\mu\text{L}$  PBST and wells blocked with 1% (w/v) BSA in PBS for 1 h. All incubation steps were performed in PBS at RT followed by a PBST wash step as described above. PSIM (100  $\mu\text{L}$  of 16  $\mu\text{g}/\text{ml}$ ), pPGM (100  $\mu\text{L}$  of 4  $\text{ng}/\text{mL}$ ) or human MUC2 (100  $\mu\text{L}$  of 6  $\mu\text{g}/\text{mL}$ ) were added to the wells and incubated for 2 h. Bound PSIM components and mucin were either detected via b-WGA (1:500) followed by incubation with ExtrAvidin-peroxidase (1:1,000) (Sigma-Aldrich, Poole, UK) or by anti-Muc2.3 (1:1,000) for PSIM and pPGM, and anti-MUC2C3 (1:1,000) for MUC2, and anti-rabbit-HRP (1:10,000). For detection, 100  $\mu\text{L}$  3,3',5,5'-Tetramethylbenzidine (TMB) peroxidase substrate was added to the wells and the absorbance measured at 640 nm every 10 min for 1 h in a Benchmark Plus™ microplate spectrophotometer (Bio-Rad, Hercules, USA).

## **2.4 Biophysics**

### **2.4.1 Isothermal titration calorimetry (ITC)**

To assess the binding of Mub proteins to sugar ligands, isothermal titration calorimetry (ITC) experiments were conducted in collaboration with David Bolam (Institute for Cell and Molecular Biosciences, Newcastle University, UK). Briefly, Man, Man $\alpha$ 1-6Man and Man $\alpha$ 1-3(Man $\alpha$ 1-6)Man (10 mM) were titrated into MubR5 in 10 mM sodium phosphate buffer at 95  $\mu\text{M}$  or in matching buffer control (25 injections at 25°C).

### **2.4.2 Circular dichroism (CD)**

Circular dichroism (CD) experiments were carried out in a JASCO J-710 spectropolarimeter (Great Dunmow, Cambs, UK) using a 0.1 mm split-cuvette. Purified proteins were transferred into ultrapure water and concentrated to 1  $\text{mg}/\text{mL}$  and 4 or 10 UV CD spectra per sample were accumulated and averaged over a scan range of 180-260 nm at a scan speed of 20  $\text{nm}/\text{min}$  with a band width of 1.0 nm and a response time of 4 sec. Data were manipulated including subtraction of blank spectra (ultrapure

water) using the JASCO Spectra Manager 32 v1.40.00a software (Easton, MD, USA) and the DichroWeb online tool (<http://dichroweb.cryst.bbk.ac.uk>) and CONTIN analysis program [330].

### **2.4.3 Analytical ultracentrifugation (AUC)**

Analytical ultracentrifugation experiments were performed using a Beckman XL-I analytical ultracentrifuge (Beckman Coulter, High Wycombe, UK) equipped with scanning absorbance and interference optics and an An50Ti rotor. Native MUB was freeze-dried directly after purification (see section 2.3.2), stored at  $-80^{\circ}\text{C}$  and re-suspended in PBS, 0.1 M carbonate buffer (pH 9.8, 150 mM NaCl) and 10 mM citrate buffer (pH 4.6, 150 mM NaCl) at a final concentration of 0.1 mg/mL immediately prior sedimentation equilibrium experiments or in PBS for sedimentation velocity experiment. For sedimentation equilibrium experiments, five scans were recorded every four hours at 7,000 rpm and  $20^{\circ}\text{C}$ , and concentration profiles measured using absorbance optics at 280 nm. Sedimentation velocity experiments were recorded at 35,000 rpm using interference optics.

The partial specific volume of MUB was calculated from amino acid sequence using the program SEDNTERP [331] as 0.72 mL/g. Buffer densities were 1.0053 g/mL for PBS and 1.0055 g/mL for carbonate buffer. The program UltraScan [332] was used for AUC data analysis and fit. Sedimentation velocity data were analysed by a radial derivative ( $dC/dr$ ) method to calculation of  $S(20,w)$  value. Sedimentation equilibrium profiles of MUB in carbonate buffer and PBS were fitted to a single component or an independent two component system, respectively.

### **2.4.4 Dynamic light scattering (DLS)**

Dynamic light scattering (DLS) experiments were performed in the temperature controlled DynaPro Protein Solutions DLS device (Wyatt Technology, Santa Barbara, USA) controlled by the DYNAMICS V5 software (Protein Solutions Inc., Charlottesville, USA) to investigate presence of aggregation in purified protein samples. Scattering data of MubR5, -RV, -RV-VI and MUB in PBS and Tris-HCl buffer and MubRI-II-III in PBS, 10 mM sodium phosphate buffer, 10 mM Tris-HCl, 10 mM Tris-HCl with 150 mM NaCl, 0.1 M carbonate buffer (pH 9.8, 150 mM NaCl) and 10 mM citrate buffer (pH 4.6,

150 mM NaCl) were collected at 100% laser intensity with an acquisition time of 10 sec and acquisition number of 10 at 25°C or 4°C. Data were analysed using the DYNAMICS V5 software package.

## **2.5 Structural biology**

### **2.5.1 X-ray crystallography**

Protein crystallisation was performed via hanging or sitting drop vapour diffusion in 24 or 96 well crystallisation trays (Molecular Dimensions, Newmarket, UK). Purified protein solutions in sodium phosphate or Tris-HCl buffer at 10 mg/mL (or 19 mg/mL for MubRI) (see 2.3.1), snap-frozen in liquid nitrogen and stored at -80°C, were thawed on ice immediately before crystallisation experiments, which were set-up using an Oryxnano crystallisation robot (Douglas Instruments, East Garston, Hungerford, UK) with a 1:1 ratio of protein to precipitant solution and a drop size of 0.5 µL. Three different crystallisation screens of 96 conditions were used in this study: Structure Screens 1 and 2, JCSG-plus Screens 1 and 2 (all Molecular Dimensions, Newmarket, UK), and PEG/Ion Screens 1 and 2 (Hampton Research, Aliso Viejo, USA) (see Appendix V). Crystallisation trays were incubated at 16°C and 4°C and crystal growth was monitored using an Olympus SZX9 light microscope (2× or 1.5× objective lens) with a Highlight 2100 Olympus Europe light source (Southend-on-Sea, UK). Crystals were harvested and stored in liquid nitrogen until crystal diffraction data sets were collected at Diamond Light Source (Didcot, UK).

#### **2.5.1.1 Protein crystallisation**

MubRV was crystallised at a protein concentration of 10 mg/mL with a precipitant solution of 0.2 M ammonium acetate and 24% (w/v) PEG 3,350 at 16°C. Crystals were cryo-protected by adding 25% (v/v) DMSO to the reservoir solution. MubRI crystals were grown at a concentration of 18 mg/mL in 0.1 M HEPES (pH 7.5), 10% (v/v) 2-propanol/ 20% (w/v) PEG (polyethylene glycol) 4,000 at 4°C, and crystals soaked in 10 mM ytterbium or mercury chloride solution with 25% (v/v) ethylene glycol as cryoprotectant for multi-anomalous dispersion (MAD) data set collection. Lar0958 was crystallised at 12 mg/mL in 0.2 M ammonium sulphate, 0.1 M sodium acetate (pH 4.5) with 25% (w/v) PEG 4,000 or 30% (v/v) PEG 2,000 MME (monomethylether) at 4°C. Crystals were soaked with 0.5 M potassium bromide for single anomalous dispersion

(SAD) phasing, and soaked and native crystals cryo-protected with 20% (v/v) ethylene glycol.

### **2.5.1.2 Crystal soaking and co-crystallisation studies**

Crystal soaking and co-crystallisation studies of MubR5 with fucose (Fuc), mannose (Man), galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), *N*-acetylneuraminic acid (Neu5Ac) and the disaccharide *N*-acetyllactosamine (LacNAc) were performed at a protein concentration of 8 mg/mL. For crystal soaking studies, single protein crystals were grown in 20% to 30% (w/v) PEG 3350 and 0.2 Gal, GalNAc, GlcNAc, Neu5Ac and LacNAc solutions with 25% to 30% (w/v) PEG 3350 as cryoprotectant and incubated for several minutes at sugar concentrations of 1 mM to 500 mM depending on the crystal stability in sugar solution. For co-crystallisation experiments, crystal structure screens were performed with LacNAc, Neu5Ac and Man at different sugar concentrations of 10 mM, 50 mM, 100 mM and 200 mM and single crystals grown in 25% PEG 4000, 0.2 M magnesium chloride, 0.1 M Tris (pH 8.5).

### **2.5.1.3 X-ray data set analysis**

Integration and reduction of X-ray diffraction data from single crystals of MubR5, MubRI, MubRV and native Lar0958 was achieved using either a combination of MOSFLM [333] and SCALA [334] or by means of the Xia2 automated data reduction system [335]. Crystal diffraction data were analysed using the CCP4i [336] and PHENIX program suits [337].

Model building was performed with the  $\sigma$ -weighted  $2mF_{\text{obs}}-DF_{\text{calc}}$  and  $mF_{\text{obs}}-DF_{\text{calc}}$  Fourier electron density maps [338]. For structural validation, a Ramachandran analysis was performed to assess favoured, allowed or disallowed orientation of amino acids. The amino acid geometry was improved if possible according to the limits of the electron density maps. The solvent content of crystals was estimated according to Matthews, indicating the number of protein molecules in the asymmetric unit (ASU) [339].

Molecular replacement with MubR5 complexes, MubRI and MubRV data sets was performed with MOLREP [340] and PHASER [341] using the MubR5 structure as a search model or a search model based on the MubR5 structure [340-341]. Alignment

models for MubRI or MubRV were generated after amino acid alignment with MubR5 for MubRI and MubRV or with PEPE\_0118 (PDB entry 3LYY) for MubRV using CHAINSAW via pruning of non-conserved residues [342]. For ligand identification in MubR5 data sets of crystal soaking or co-crystallisation, the COOT tool 'unmodelled blobs' [343] was used after initial refinement using REFMAC [344]. Model refinement and alternating manual model building was performed using RESOLVE and COOT [343, 345].

For space group validation of the MubRV data set, the programs LABELIT and ZANUDA were used [346].

Lar0958 data for crystals derivatised with KBr was integrated, reduced and intensities scaled using DENSO and SCALEPACK as part of HKL2000 [347]. Bromide atoms were located via AutoSol and initial SAD phasing performed by PHASER [341, 348]. A first molecular model was built by AUTOBUILD [348]. Manual model building in the molecular graphics program COOT [343] alternated with refinement using PHENIX [348].

## **2.5.2 Small angle X-ray scattering (SAXS) studies**

### **2.5.2.1 SAXS data collection**

The scattering curves of MubR5, MubRV, MubRI, MubR8-V, MubRV-VI, and MubRI-II-III in Tris-HCl were recorded in a concentration range of ~0.6-9 mg/mL as  $10 \times 10$  sec frames at a wavelength of 0.93 Å and a sample-detector distance of 2.4 m covering the momentum transfer range of  $0.04 < s < 0.61 \text{ \AA}^{-1}$  ( $s = 4\pi \sin(\theta)/\lambda$ , where  $2\theta$  is the scattering angle and  $\lambda$  the wavelength) on the ID14-3 beamline, ESRF, Grenoble, France.

Additionally, scattering curves of MubRI-II-III in sodium phosphate buffer and PBS, both supplemented with 2 mM DTT, and Nterm in Tris-HCl were recorded in a concentration range of ~0.5-5 mg/mL for MubRI-II-III and 0.6 to 9 mg/mL for Nterm as  $10 \times 10$  sec frames at a wavelength of 0.99 Å and a sample-detector distance of 2.9 m covering the momentum transfer range of  $0.03 < s < 0.45 \text{ \AA}^{-1}$  on the BM29 beamline, ESRF, Grenoble, France.

### 2.5.2.2 SAXS data analysis

The ATSAS (version 2.4) software was used for SAXS data analysis. Data were normalised subtracting the buffer scattering, scaled for concentration and data points across different concentrations were merged using PRIMUS [349]. The radius of gyration ( $R_g$ ) and scattering at zero angle ( $I(0)$ ) were calculated by Guinier approximation with  $R_{gs} \leq 0.8$  for elongated proteins (Mub-repeats) or with  $R_{gs} \leq 1.0$  for more globular proteins (Nterm), and the distance distribution function ( $P(r)$ ) was generated by GNOM [350] computing the maximum particle diameter ( $D_{max}$ ) and an  $R_g$  value calculated for the whole scattering range. Ten ab initio shapes were reconstructed by GASPOR [351] or by DAMMIF (for MubRI-II-III only) and averaged by the DAMAVER program package [352], generating a  $\chi$  (Chi) value, a measure for the fit of the experimental data to the shape reconstruction, and a normalised spatial discrepancy (NSD), a measure for the agreement between computed shape models. Manual docking of the high-resolution X-ray structures of MubR5 (PDB entry 3I57) and MubRV (PDB entry 4MT5), into low-resolution shape reconstructions was performed using SCULPTOR (version 2.1) [353] and SITUS [354]. The refinement of the docking solution by SITUS calculated a cross correlation coefficient  $R$ , which allows quantitative evaluation of volumetric map and docked structure. The solution scattering of MubR5 and -RV were computed from their atomic structures and fitted to the collected experimental scattering curves using CRY SOL [355]. The molecular weight of solutes was calculated by scaling against reference solutions of BSA providing information on the oligomeric state of the proteins in solution.

## 2.6 Bioinformatics

### 2.6.1 General data analysis

For microtitre plate adhesion assays (see 2.3.6.4), pI determination by isoelectric focusing (see 2.3.4.3) and molecular weight determination by gel filtration or SDS-PAGE, data were analysis in Excel (Microsoft, Washington, USA).

### **2.6.2 Sequence analysis**

The ExPasy protein parameter tool (<http://web.expasy.org/protparam/>) was used to calculate the theoretical molecular weight, isoelectric point, protein extinction coefficient and Abs 0.1%, absorbance value from amino acid sequence of proteins.

The multiple sequence alignment program ClustalW [356] was used for sequence alignment of DNA or protein molecules.

### **2.6.3 Protein structure analysis**

Pairwise structural alignment of proteins was achieved using DaliLite [357] and was employed for protein structure comparison searching the Protein Data Bank (PDB) using the Dali server [358-359].

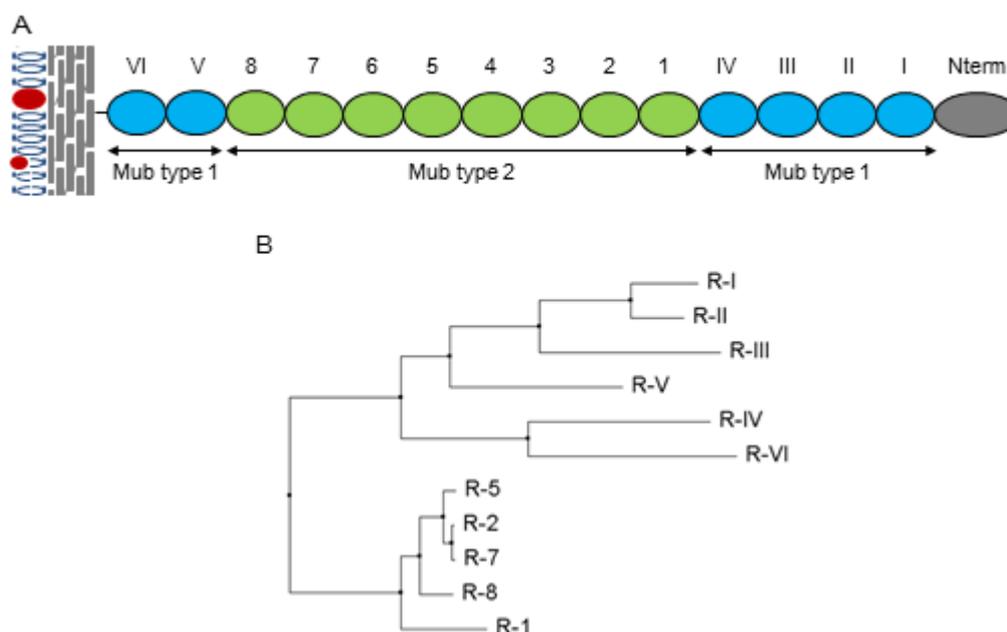
The MarkUs function annotation server [360] was used to investigate structural, biophysical and functional properties of protein structure models.

### **2.6.4 Crystal diffraction and SAXS data analysis**

Crystal diffraction data were analysed using the CCP4i [336] and PHENIX [337] program suites (see 2.5.1.3). For SAXS data analysis, the ATSAS (version 2.4) software was used (see 2.5.2.2). Final figures were made using the PyMOL Molecular Graphics System Version 1.5.0.4 Schrödinger, LLC.

## CHAPTER 3 PURIFICATION AND BIOCHEMICAL CHARACTERISATION OF MUB PROTEINS AND NATIVE MUB

The cell-surface mucus binding (MUB) protein of *L. reuteri* ATCC 53608 (also known as 1063), consists of 14 tandemly arranged Mub type repeats, Mub type 1 and 2, classed based on amino acid sequence. Mub type 1 repeats comprise MubRI, -RII, -RIII, -RIV, -RV and -RVI, and Mub type 2 repeats comprise MubR1, -R2, -R3, -R4, -R5, -R6, -R7 and -R8 (Figure 3.1) In addition, MUB contains an N-terminal domain adjacent to the first Mub repeat RI (see 1.4.2).



**Figure 3.1 Schematic representation of cell-surface anchored MUB and neighbour joining tree for Mub repeat sequences**

(A) Mub type 1 and type 2 repeats are coloured blue and green and numbered in Roman and Arabic, respectively. The N-terminal domain is shown in grey named Nterm. (B) Phylogeny tree calculated by JALVIEW [361] based on percentage identity of aligned Mub repeat sequences (repeats R2, R4, and R6, as well as R3 and R5 are identical).

The Mub type 2 repeats (184 residues) show high sequence identity between 84 and 100% (for sequence alignment and alignment scores see Appendix IV). In contrast, the Mub type 1 repeats (183-206 residues) are more diverse with sequence identities ranging from 29% for RIV and RV to 88% for RI and RII. The lowest similarity (24%)

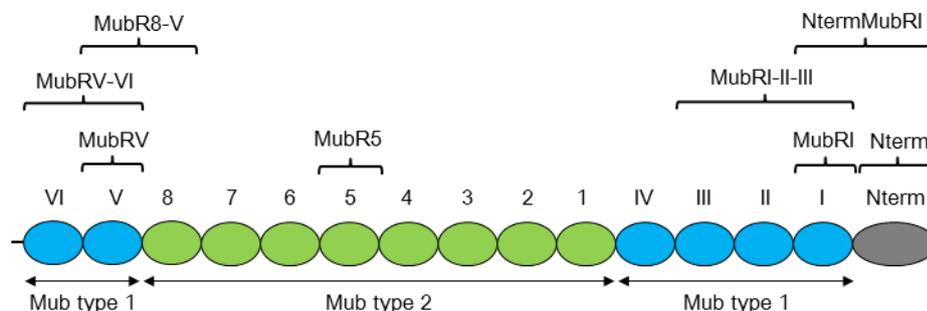
between a type 1 repeat and type 2 repeats was observed for RVI and R5 (R3 and R1 being identical or nearly identical) (for Mub domain borders see Appendix II).

In order to obtain structural information on MUB protein domains and to investigate their functional role in the adhesion to mucin and mucin glycans, recombinant MUB proteins of individual or tandem Mub repeats were cloned, expressed and purified, and native MUB isolated and purified from *L. reuteri* ATCC 53608 culture media.

### 3.1 Cloning, heterologous expression and purification of MUB proteins

#### 3.1.1 Mub type 1 and 2 repeat proteins

For characterisation studies of different Mub repeat proteins, the single type 2 repeat MubR5, the single type 1 repeats MubRI and -RV, the Mub type 1 double repeat MubRV-VI, the mixed type double repeat MubR8-V and the type 1 triple repeat MubRI-II-III, were cloned and the recombinant proteins heterologously expressed in *E. coli* (Figure 3.2).



**Figure 3.2 Schematic representation of MUB with recombinant MUB proteins**

Mub type 1 and type 2 repeats are coloured blue and green, and numbered in Roman and Arabic, respectively. The N-terminal domain is shown in grey named Nterm. Recombinant Mub repeat and Nterm proteins indicated by brackets.

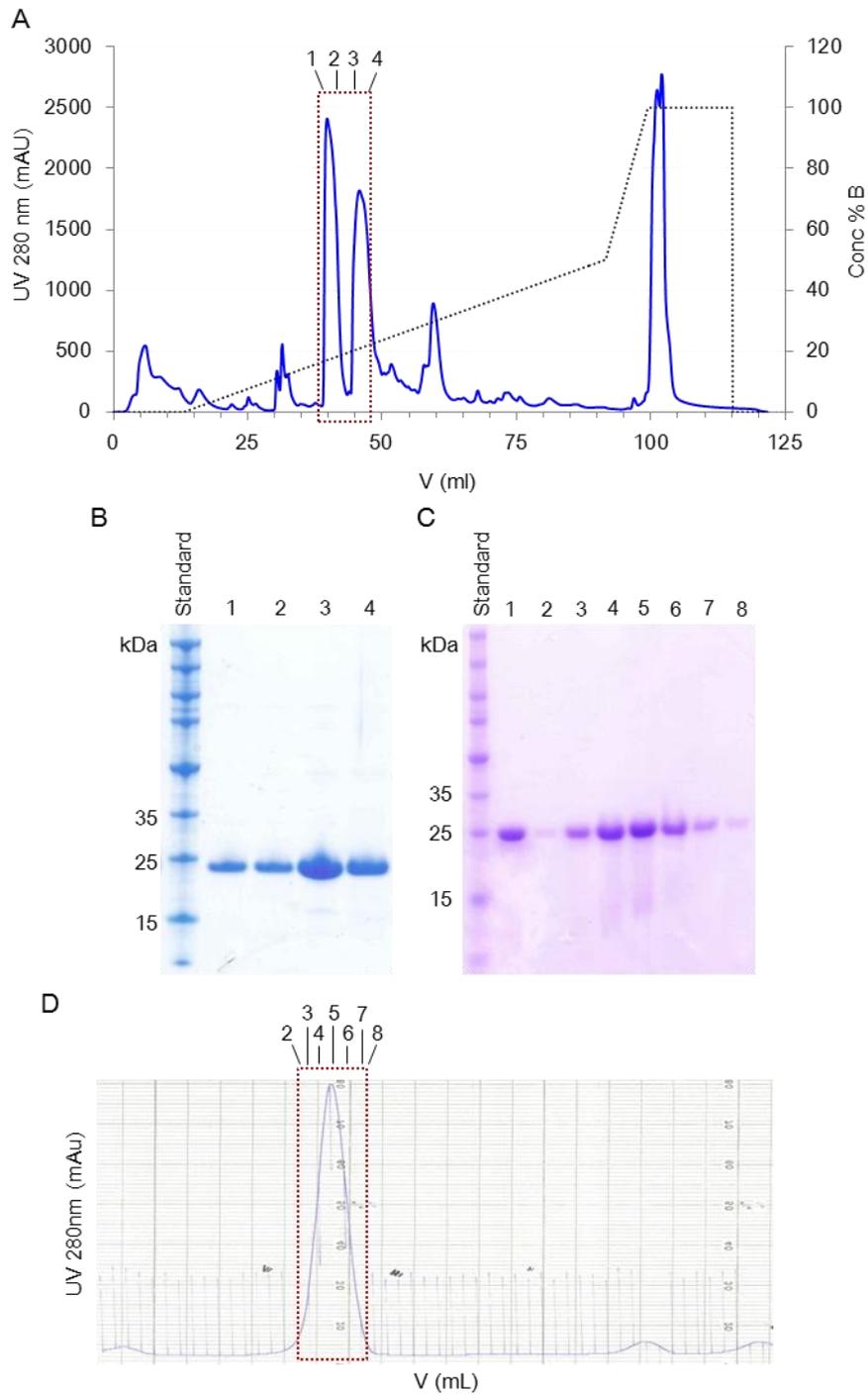
Vector constructs encoding MubR5, MubRI and MubRI-II-III were already available at the beginning of this study and the cloning of the remaining MUB proteins was performed following the same approach. Briefly, oligonucleotide primers for the Mub repeat genes *mubR5*, *mubRI*, *mubRV*, *mubRV-VI*, *mubR8-V* and *mubRI-II-III* were designed to anneal to specific Mub repeat border regions within the mucus-binding

protein gene of *L. reuteri* ATCC 53608 (American Type Culture Collection). The borders of individual Mub repeats were determined by comparative multiple sequence alignments of MUB protein domains from different *Lactobacillus* and *Lactococcus* species [304]. The so defined Mub repeat boundaries are different from those described for the GenBank accession number AF120104 (see Appendix II).

Genes were amplified by PCR (see 2.2.1) from washed bacterial cells, cloned into the pETBlue-1 AccepTor vector (see 2.2.4) and sequence integrity verified by automated DNA sequencing (see 2.2.6). The recombinant vectors carrying the *mub* repeat genes were used to transform *E. coli* Tuner(DE3)pLacI<sup>2</sup> cells for protein expression. MubR5, MubRI, MubRV, MubRV-VI and MubRI-II-III recombinant proteins were produced in a soluble form at 37°C and MubR8-V at 30°C after induction with 1 mM  $\beta$ -D-1-thiogalactopyranoside (IPTG) (see 2.3.1.1), as demonstrated below by SDS-PAGE analysis (Figure 3.3 to 3.8).

Owing to their small size, the single Mub repeat proteins, MubR5, -RI and -RV, were extracted via freeze-thaw method, to avoid cell disruption and reduce the contamination by host proteins in the clarified crude extract (see 2.3.1.2). After an initial purification step via ion exchange chromatography (IEC), all three single Mub proteins showed high sample homogeneity (Figure 3.3 A and B, 3.4 A and B, and 3.5 A and C).

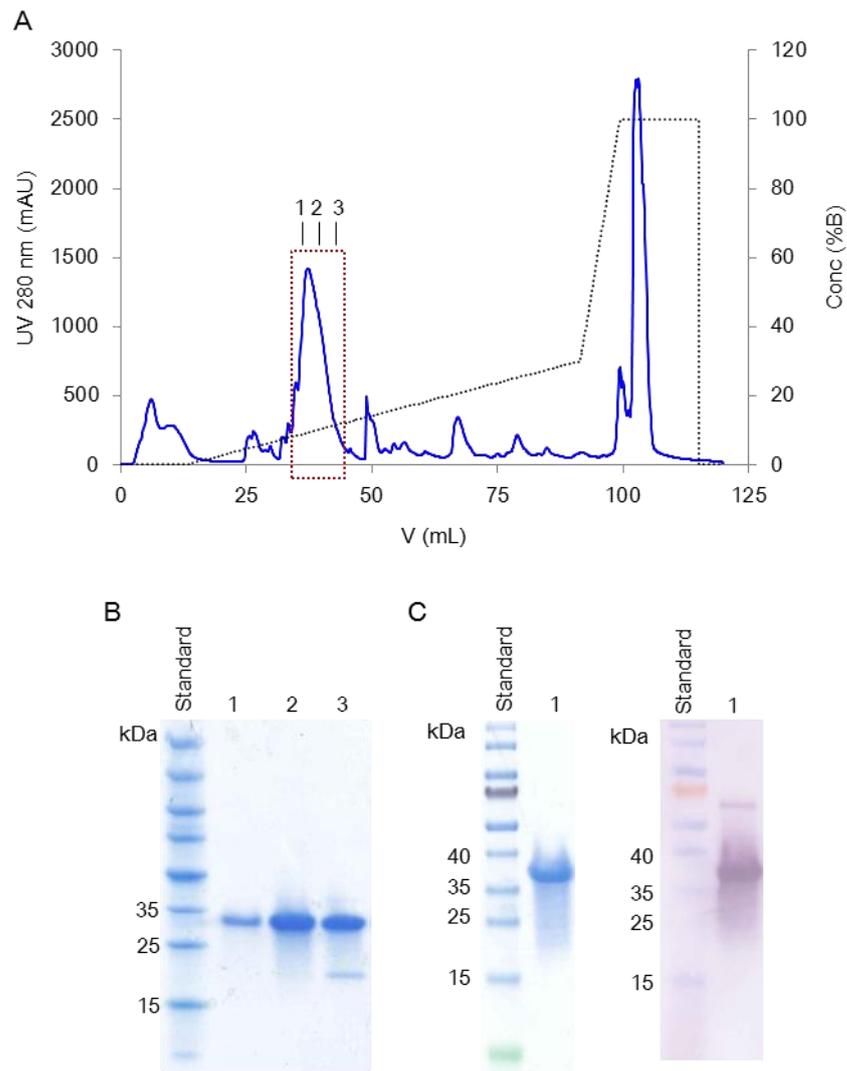
Two major elution peaks were present in the IEC chromatogram of MubR5 (Figure 3.3 A). For both peaks only a single distinct protein band with an apparent molecular weight (MW) of about 22 kDa was observed on the Coomassie stained SDS-PAGE gel (Figure 3.3 B), which is slightly higher than the calculated MW of MubR5 of 20.5 kDa (calculated with ExPasy ProtParam tool, see 2.6.2).



**Figure 3.3 Purification of MubR5 by IEC and SEC**

(A) IEC profile of MubR5 and (B) SDS-PAGE analysis (NuPAGE 4-12% gel) of MubR5 elution fractions (1-4) (as indicated in A). (D) SEC elution profile and (C) SDS-PAGE analysis (RunBlue 12% gel) of MubR5 elution fractions (as indicated in D) (1 sample before injection, 2-8 elution fractions).

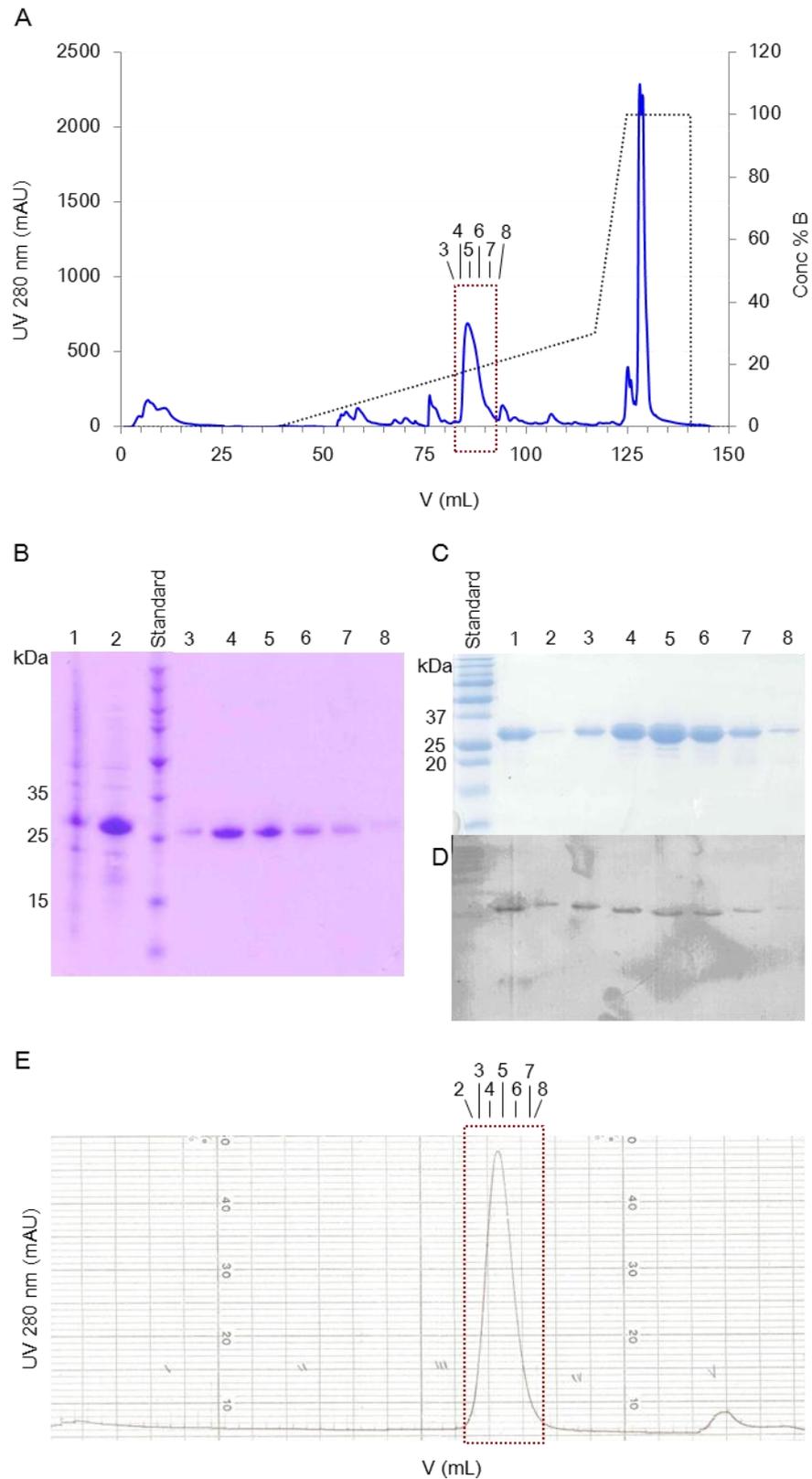
The elution of MubR5 in two distinct peaks is an artefact due to injection of 2×2 mL highly concentrated sample solution onto the IEC column. Nevertheless, both peak samples were further examined for their protein properties by circular dichroism (CD) and isoelectric focusing (IEF) before use in structural studies (see 3.2). In order to investigate potential differences of the two IEC peaks of MubR5, to assess its oligomeric state in solution and to reduce the presence of potential protein aggregation, MubR5 was further purified by size exclusion chromatography (SEC) (see 2.3.1.3). A single elution peak was observed for MubR5 when both IEC peak fractions were tested separately or combined at an elution volume of 67.5 mL corresponding to a MW of about 28.8 kDa (Figure 3.3 D), suggesting no major difference in protein properties between MubR5 of both IEC elution peaks. The SEC column calibration, which was performed with standard globular calibration proteins (see 2.3.1.3), indicated a considerably higher observed MW for MubR5 compared to its theoretical MW (20.5 kDa), which may be explained by the elongated shape of MubR5 as observed in the X-ray crystal structure [282].



**Figure 3.4 Purification of MubRI by IEC and SEC**

(A) IEC profile of MubRI and (B) SDS-PAGE analysis of MubRI peak elution fractions (as indicated in A) (1-3). (C) SDS-PAGE analysis of pooled SEC elution fractions (elution profile not shown) and (D) Western-blot membrane incubated with primary anti-MubRI and secondary anti-rabbit-AP.

For MubRI and MubRV only single elution peaks were observed by IEC (Figure 3.4 and 3.5, A) containing a single protein with apparent MW of about 28 and 27 kDa, respectively, as demonstrated by SDS-PAGE analysis (Figure 3.4 and 3.5, B). The apparent MW observed for the recombinant proteins were higher compared to the calculated MW of 21.3 and 20.3 kDa for MubRI and MubRV, respectively (see 2.6.2).



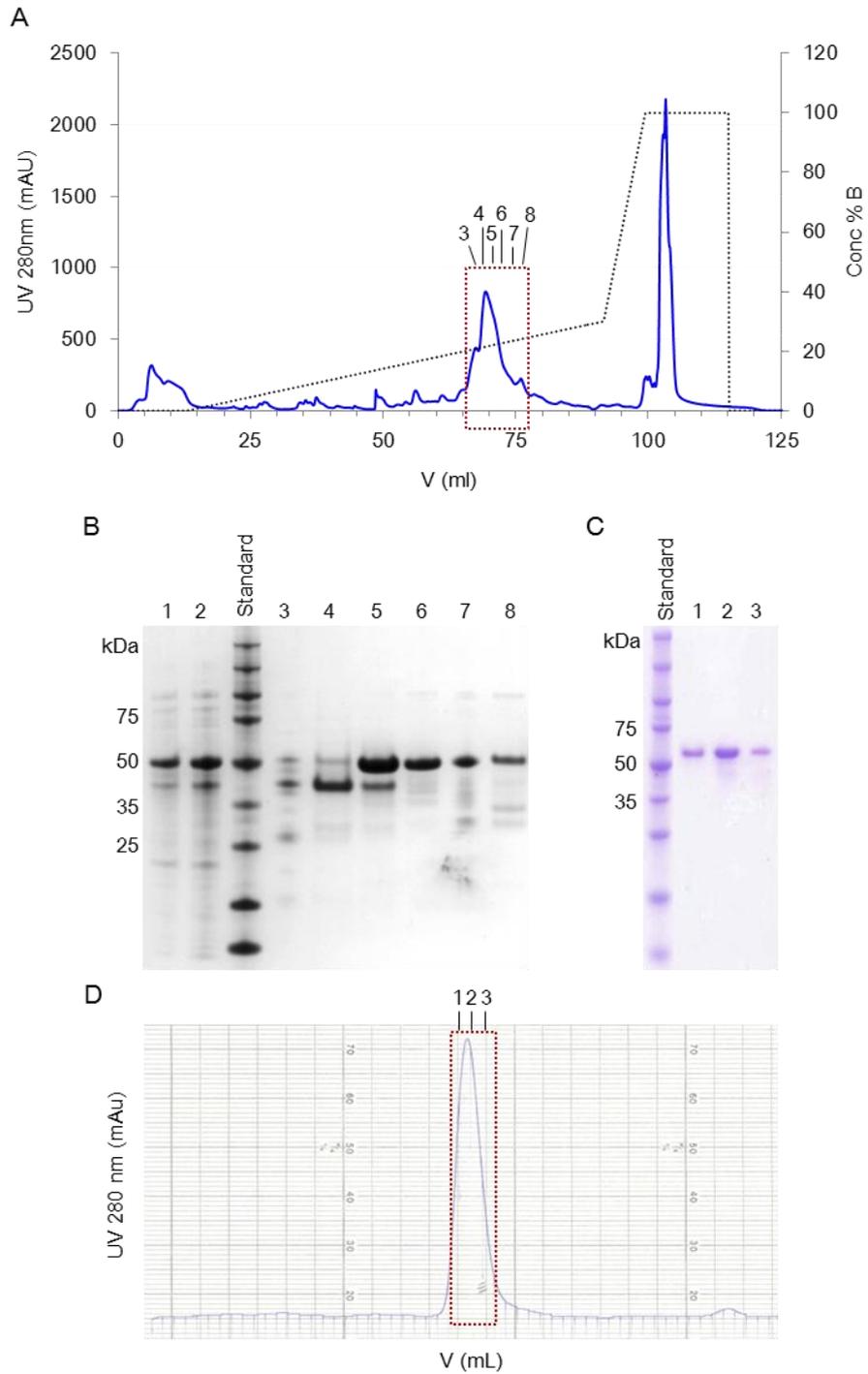
**Figure 3.5 Purification of MubRV by IEC and SEC**

(A) IEC profile of MubRV, (B) SDS-PAGE analysis (RunBlue 12% gel) of MubRV peak elution fractions as indicated in A (1 pellet, 2 crude extract, 3-8 peak fractions). (E) SEC elution profile, (C) SDS-PAGE analysis of MubRV peak elution fractions (as indicated in E)

(1 injection sample, 2-8 peak elution fractions) and (D) Western-blot membrane incubated with primary anti-MubRI and anti-MubR5 followed by anti-rabbit-AP.

Despite high sample homogeneity, MubRI and MubRV were further purified by SEC for structural studies (Figure 3.4 C and D, and 3.5 C to E). Both proteins eluted in a single peak at an elution volume of 64.0 mL for MubRI (data not shown) and 73.0 mL for MubRV (Figure 3.5 E), corresponding to MW of 35.7 kDa and 31.6 kDa, respectively. The MubRI and MubRV protein bands showed reactivity with anti-MubRI and anti-MubR5 after Western-blotting (Figure 3.4 and 3.5, D).

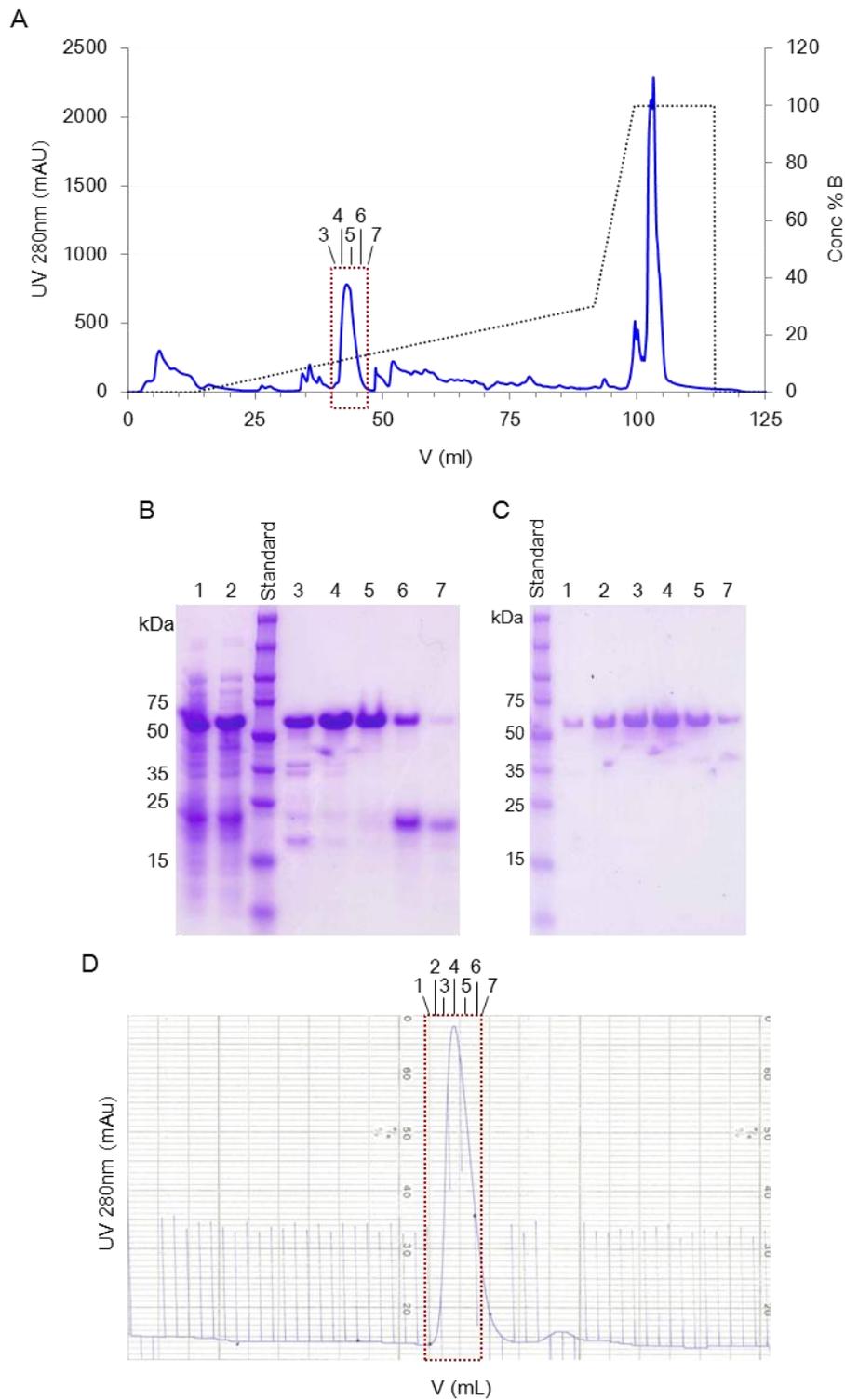
The double Mub-repeat proteins, MubR8-V and MubRV-VI, were extracted via freeze-thaw method (see 2.3.1.2) and purified via IEC (Figure 3.6 and 3.7, A). SDS-PAGE analysis of IEC elution fractions revealed sample heterogeneity (Figure 3.6 and 3.7, B). The predominant protein bands of MubR8-V and MubRV-VI with apparent MW of about 53 and 55 kDa, respectively, showed reactivity with anti-MubRI and anti MubRV after Western-blotting (data not shown). As observed before for the single Mub repeats, the recombinant proteins showed higher apparent MW compared to the calculated MW of MubR8-V and MubRV-VI of 40.6 kDa and 43.1 kDa, respectively (see 2.6.2).



**Figure 3.6 Purification of MubR8-V by IEC and SEC**

(A) IEC profile of MubR8-V and (B) SDS-PAGE analysis (NuPAGE 4-12 % gel) of peak elution fractions (as indicated in A) (1 pellet, 2 crude extract, 3-8 elution fractions). (D) SEC elution profile and (C) SDS-PAGE analysis (RunBlue 12% gel) of peak elution fractions (as indicated in D) (1-3 elution fractions).

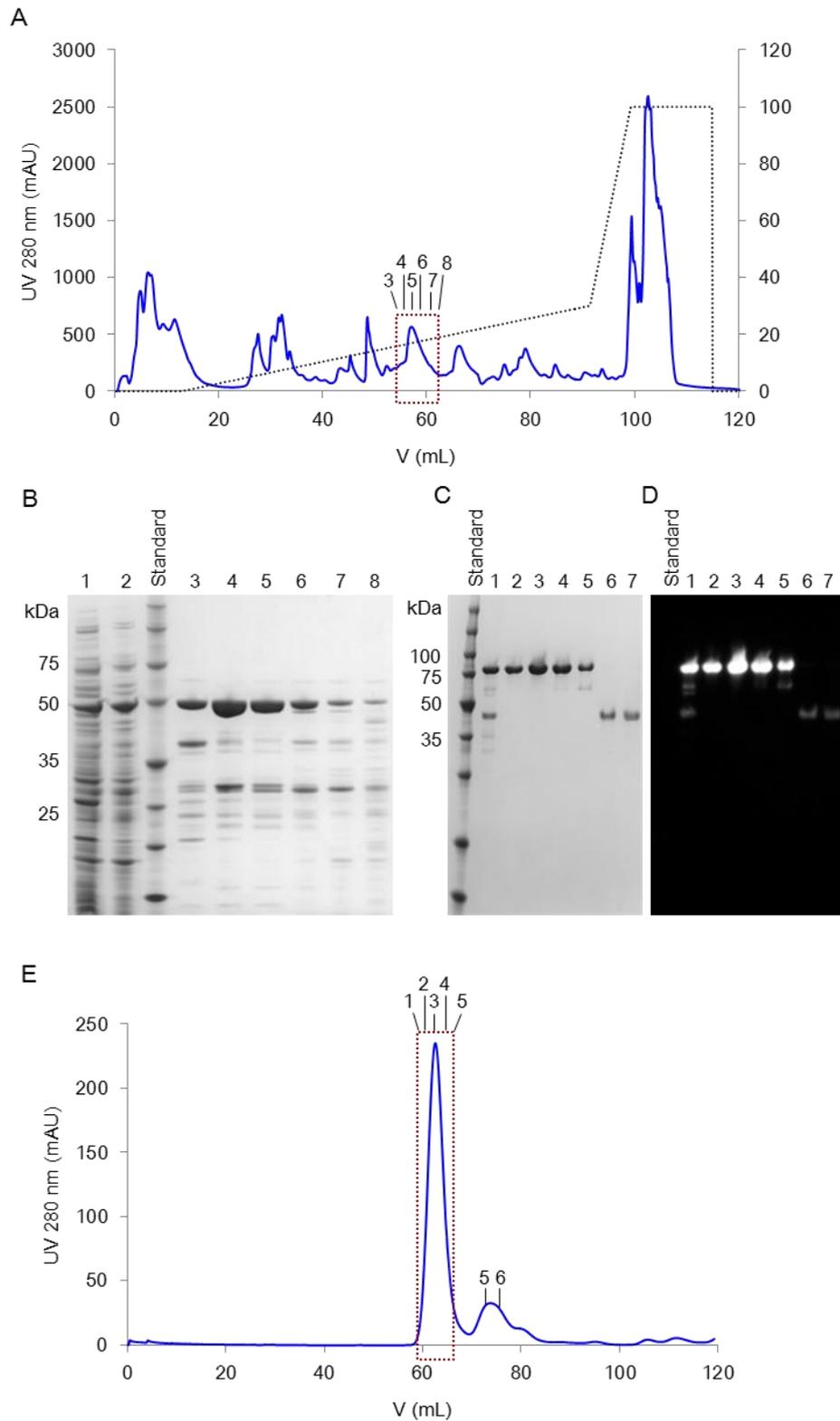
IEC elution fractions containing high target protein content and least contaminating proteins were pooled and the double repeat proteins were further purified by SEC. MubR8-V and MubRV-VI showed main elution peaks at an elution volume of 54.3 mL corresponding to a MW of about 64.9 kDa, which is higher than their calculated MW and in line with observations made for single Mub repeats as described above (Figure 3.6 and 3.7, D). SEC resulted in sufficient sample homogeneity of MubR8-V and MubRV-VI for characterisation and binding studies as demonstrated by SDS-PAGE analysis (Figure 3.6 and 3.7, C).



**Figure 3.7 Purification of MubRV-VI by IEC and SEC**

(A) IEC profile of MubRV-VI and (B) SDS-PAGE analysis of MubRV-VI elution fractions (as indicated in A) (1 pellet, 2 crude extract, 3-7 elution fractions). (D) SEC elution profile and (C) SDS-PAGE analysis of peak elution fractions (as indicated in D) (1-7 elution fractions).

Due to its higher MW, the triple Mub type 1 repeat protein MubRI-II-III was extracted from the bacterial cell pellet by ultrasonication (see 2.3.1.2). The triple repeat protein was first purified from clarified extract by IEC (Figure 3.8 A) and eluted in heterogeneous elution fractions as shown by SDS-PAGE (Figure 3.8 B). A predominant protein band was observed at an apparent MW of about 77 kDa, which is again higher than the theoretical MW of MubRI-II-III of 64 kDa based on amino acid sequence (see 2.6.2). To achieve higher purity, MubRI-II-III elution fractions were pooled and applied onto a SEC column. Two elution peaks showed with elution volumes of 62.6 mL and 73.8 mL corresponding to MW of 162.4 kDa and 71.8 kDa, respectively (Figure 3.8 D). Samples of both peaks showed apparent MW of about 77 kDa and 32 kDa, suggesting that the triple domain elutes in the first predominant peak probably as a dimer. The 77 kDa MubRI-II-III protein showed strong reactivity with the anti-MubRI after Western-blotting, but interestingly weak signals were also observed for the lower MW proteins in the second elution peak (Figure 3.8 C). Both protein species were further analysed by mass spectrometry (MS) (with the help of Fran Mulholland, Institute of Food Research, Norwich, UK) after trypsin digest for peptide mass fingerprint analysis, which confirmed that the high MW protein contained the three Mub repeats RI, RII and RIII, whereas the lower MW protein is a truncated version of the triple domain proteins only comprising the first two repeats RI and RII.



**Figure 3.8 Purification of MubRI-II-III by IEC and SEC**

(A) IEC profile of MubRI-II-III and (B) SDS-PAGE analysis (NuPAGE 4-12 % gel) of peak elution fractions (as indicated in A) (1 pellet, 2 crude extract, 2-8 elution fractions). (E) SEC elution profile, (C) SDS-PAGE analysis (RunBlue 12 % gel) of main peak elution fractions (as indicated in E) (1-5 main peak elution fractions, 6+7 second elution peak fractions) and (D) Western-blot membrane incubated with anti-MubRI and anti-rabbit-HRP.

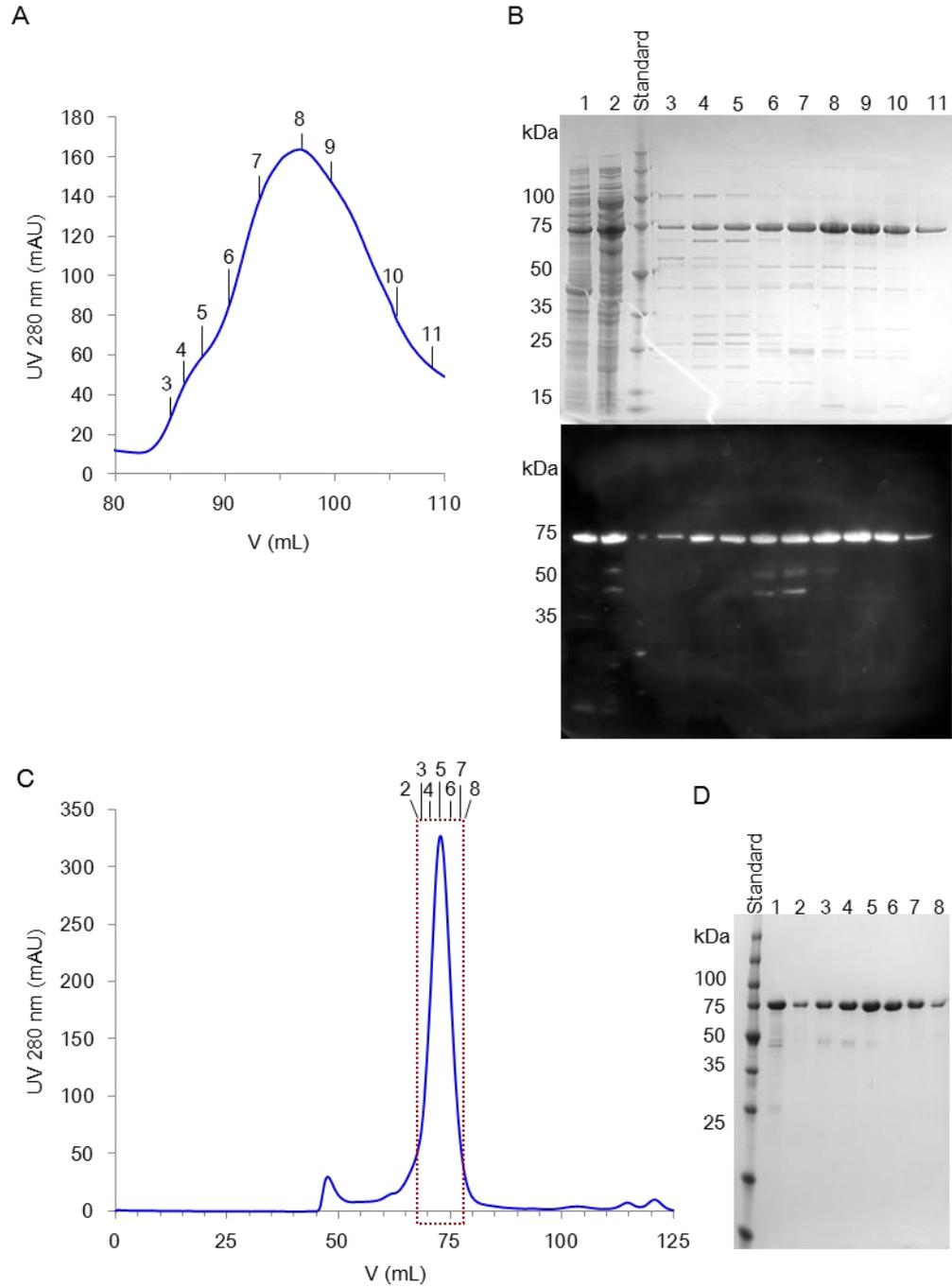
To summarise, all Mub-repeat proteins were successfully purified by IEC and SEC to homogeneity. For all Mub-repeat proteins an elevated apparent MW was observed upon SDS-PAGE and by SEC, possibly due to an elongated, non-globular protein fold. Protein yields varied between the different Mub repeats and were calculated as protein per wet cell weight with 2.4 mg/g for MubR5, 2.2 mg/g for MubRV, 1.8 mg/g for MubRI, 3.5 mg/g for MubR8-V, 1.5 mg/g for MubRV-VI and for 0.28 mg/g MubRI-II-III (see 3.1.2, Table 3.1).

### 3.1.2 His<sub>6</sub>-tagged N-terminal MUB domain constructs

The N-terminal domain gene (*nterm*) of MUB is located between the signal sequence, which facilitates MUB protein transport to the bacterial cell wall, and the first Mub repeat gene *mubRI* (GenBank accession number AF120104) (Figure 3.2). Primers were designed, that shared sequence homology to *nterm* alone or *nterm* and *mubRI* (*ntermmubRI*), as well as to the pOPINF target vector. This allowed homologues recombination dependent cloning of targets genes into the pOPINF vector, which encodes an N-terminal His<sub>6</sub>-tag, using the InFusion cloning system (see 2.2.4) [329]. The pOPINF vectors coding for *nterm* and *ntermmubRI* were successfully cloned into *E. coli* BL21(DE) (see 2.2.4).

Both His<sub>6</sub>-tagged proteins, Nterm and NtermMubRI, were expressed in a soluble form at 37°C after induction with 1 mM IPTG and extracted from the bacterial cell pellet using BugBuster HT (see 2.3.1.1 and 2.3.1.2) (Figure 3.9 and 3.10, B). The clarified extracts were loaded directly onto an immobilised metal affinity chromatography (IMAC) column and the His<sub>6</sub>-tag facilitated specific protein purification and elution with imidazole (see 2.3.1.3).

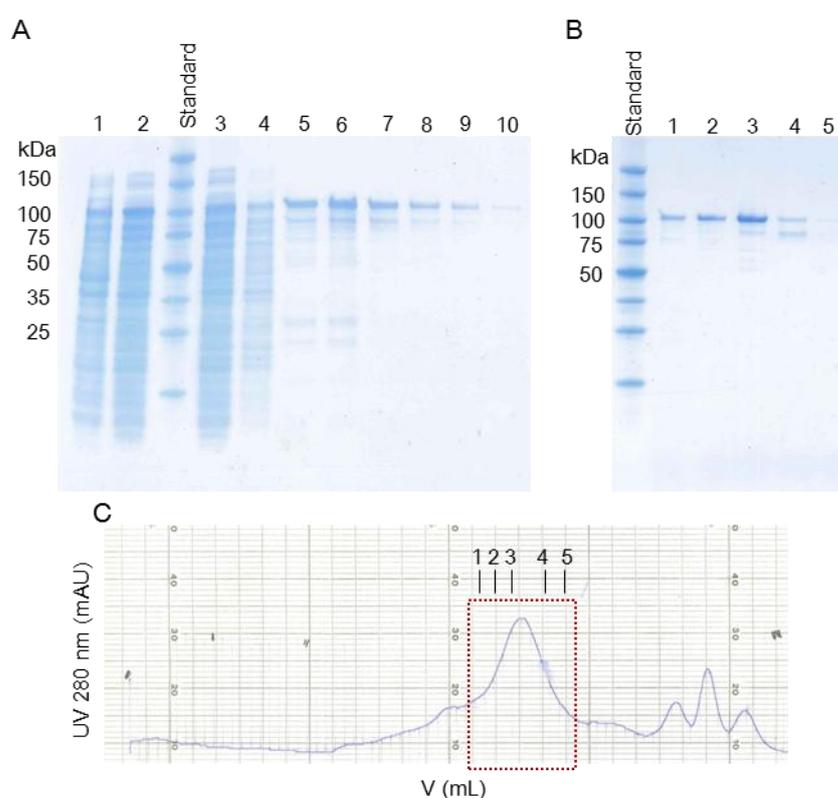
For IMAC elution fractions of Nterm, a predominant band at an apparent MW of about 84 kDa was observed on SDS-PAGE analysis (Figure 3.9 A and B), which was higher than the calculated MW of 56.3 kDa (see 2.6.2). The band showed reactivity with anti-His<sub>5</sub> after Western-blotting (Figure 3.9 C). The IMAC elution fractions contained a high number of host proteins (Figure 3.9 B) and Nterm was therefore further purified by SEC to obtain sample of sufficient homogeneity (Figure 3.9 D and E). The Nterm protein eluted at an elution volume of 73.3 mL corresponding to a MW of about 74.2 kDa.



**Figure 3.9 Purification of Nterm by IMAC and SEC**

(A) IMAC elution profile of Nterm, (B) SDS-PAGE analysis of elution fractions (as indicated in A) (1 pellet, 2 crude extract and lanes 3-11 elution fractions) and (C) Western-blot membrane incubated with primary anti-His<sub>5</sub> and secondary anti-mouse-HRP. (D) SEC elution profile and (E) SDS-PAGE analysis of elution fractions (as indicated in D) (1 sample before SEC, 2-8 peak elution fractions).

NtermMubRI was purified by gravity flow using an IMAC resin column (see 2.3.1.3). SDS-PAGE analysis of protein elution fractions showed a predominant protein band at an apparent MW of about 109 kDa (Figure 3.10 A), for which reactivity with anti-His<sub>6</sub> was observed (data not shown). The calculated MW for NtermMubRI from amino acid sequence including the N-terminal His<sub>6</sub>-tag is 77.4 kDa and is thus much lower than the apparent MW of the recombinant NtermMubRI protein as observed for the recombinant Mub repeat proteins and Nterm (see 3.1.1).



**Figure 3.10 Purification of NtermMubRI by IMAC and SEC**

(A) SDS-PAGE analysis of NtermMubRI elution fractions after IMAC (1 pellet, 2 crude extract, 3 flow through, 4 wash and 5-10 elution fractions). (C) SEC elution profile and (B) SDS-PAGE analysis of main peak elution fractions (as indicated in C) (1-5 elution fractions).

To purify NtermMubRI further, IMAC fractions were pooled, dialysed into gel filtration buffer and applied to a SEC column (Figure 3.10 C). One main elution peak was obtained corresponding to a MW of about 96 kDa (84 mL elution volume) (Figure 3.10 C).

Mub repeat	Theoretical MW [kDa]	Apparent MW <sup>a</sup> [kDa]	Observed MW <sup>b</sup> [kDa]	Purification yield [mg/g]
MubR5	20.5	22	28.8	2.4
MubRI	21.3	28	35.7	1.8
MubRV	20.3	27	31.6	2.2
MubR8-V	40.6	53	64.9	3.5
MubRV-VI	43.1	55	64.9	1.5
MubRI-II-III	64	77	162.4	0.28
Nterm	56.3	84	74.5	1.1
NtermRI	77.4	109	95.9	0.5

**Table 3.1 Purification characteristics of MUB proteins**

<sup>a</sup>apparent MW upon SDS PAGE

<sup>b</sup>observed MW by SEC

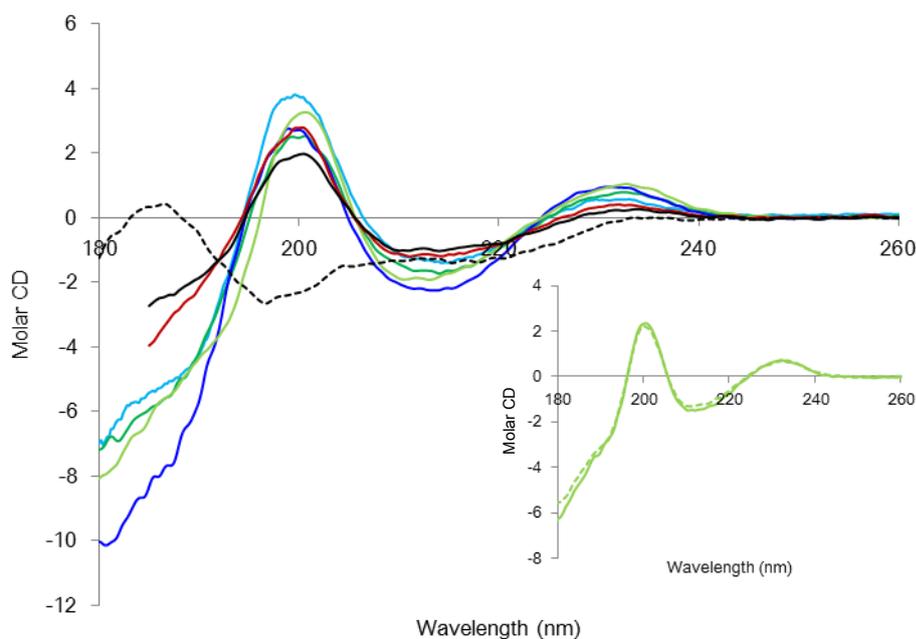
The protein yield of NtermMubRI was 0.5 mg protein per g wet cell weight compared to 1.1 mg/g for Nterm (Table 3.1). Proteolytic degradation of NtermMubRI after storage in Tris-HCl (pH 7.5) and PBS (pH 7.4) for 28 days at 4°C was observed by SDS-PAGE (data not shown). MS analysis of Nterm and NtermMubRI after trypsin digest confirmed the nature of the recombinant proteins and the degradation of NtermMubRI (performed in collaboration with Fran Mulholland, Institute of Food Research, Norwich, UK) (data not shown). Due to low protein yield and proteolytic degradation, NtermMubRI was not used in structural or follow-up functional studies.

### 3.2 Structural and biophysical properties of MUB proteins

In order to verify proper protein folding and to obtain some initial structural information on the secondary structural elements of recombinant purified MUB proteins, circular dichroism (CD) experiments were performed. In addition, isoelectric point (pI) determination was conducted by isoelectric focusing (IEF) to determine the net charge of MUB proteins in experimental buffer solutions.

Structural elements such as  $\beta$ -sheets and  $\alpha$ -helices give defined signatures in the CD spectrum and help to detect unfolding of protein. Wavelength scans of 180 to 260 nm were performed with MubR5 (peak 1 and 2, see 3.1.1), -RV, -RI, -R8-V, -RV-VI, -RI-II-III and Nterm (MubRI and -RI-II-III spectra kindly provided by Donald MacKenzie,

IFR, Norwich, UK) in ultrapure water (see 2.4.2). After subtraction of water blank, recorded spectra were plotted as the molar CD against the wavelength (Figure 3.11).



**Figure 3.11 Secondary structure determination of Mub proteins by circular dichroism (CD)**

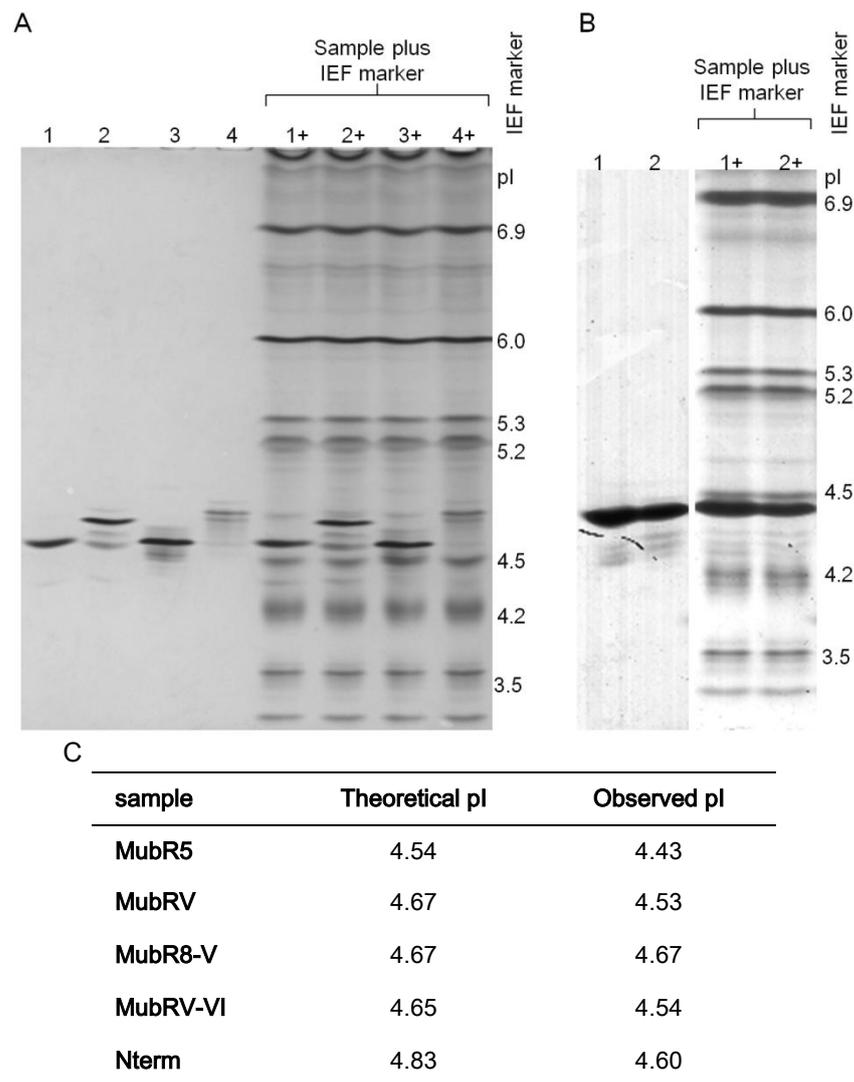
Far UV CD spectra of purified Mub-repeat proteins (RV light blue, R5 light green, RI red, RV-VI dark blue, R8-V dark green, RI-II-III black) and Nterm (black dotted) were collected over a scan range of 180-260 nm. Insert: UV spectra of MubR5 IEF elution peak 1 (light green) and peak 2 (light green, dotted).

All tested Mub repeats showed similar far UV CD spectra suggesting a highly similar overall protein solution structure (Figure 3.11). In contrast, the UV spectra of the Nterm protein differed significantly, indicating a divergent protein shape in solution. The spectra for the two IEC elution peaks of MubR5 (see 3.1.1) were nearly identical, hence a uniform solution structure can be assumed further suggesting that there is no difference between MubR5 of both peaks (Figure 3.11 inset). All sample spectra were further analysed via DichroWeb using the CONTIN analysis program with the reference data set 3 (see 2.4.2). A similar proportion of secondary structural elements was observed for all analysed Mub repeats consisting of 38.7 to 46.4%  $\beta$ -sheets and 21.1 to 28.5%  $\beta$ -turns (Table 3.2). The percentage of helix was very low with only up to 1.5% for MubRI-II-III. In contrast, for Nterm a helix percentage of 6.2% and a total percentage of 56.2%  $\beta$ -sheets and -turns were calculated.

sample	$\alpha$ -helix [%]	$\beta$ -sheets [%]	$\beta$ -turns [%]	$\beta$ -sheets and $\beta$ -turns [%]	unordered
Nterm	6.2	33.1	23.1	56.2	37.6
MubR5	0.3 (6.0) <sup>a</sup>	38.7 (45.9) <sup>a</sup>	21.1	64.8	34.9
MubRI	1.0	44.2	22.2	66.4	32.6
MubRV	1.2 (5.4) <sup>a</sup>	46.4 (46.7) <sup>a</sup>	24.4	70.8	28.1
MubR8-V	0.0	42.5	25.0	67.5	32.5
MubRV-VI	0.0	39.4	28.5	67.9	32.1
MubRI-II-III	1.5	42.0	22.4	64.4	34.1

**Table 3.2 Proportion of secondary structural elements in Mub repeats by CD (and after X-ray crystal structure determination)<sup>a</sup>**

For further characterisation of Mub repeats, IEF experiments were undertaken by gel electrophoresis in a pH range of pH 3 to 7 and a voltage gradient (see 2.3.4.3). The pI value of purified MubR5 (peak 1 and 2), -RV, -R8-V and -RV-VI was determined compared to an IEF standard (Figure 3.12). Theoretical pI values were calculated from amino acid sequence using the ExPASy online tool (Figure 3.12 A and B) (see 2.6.2). Both IEC elution peaks of MubR5 showed an identical pI value of 4.43, which was similar to its theoretical pI of 4.54, further supporting that there is no difference in MubR5 properties of the two elution peaks as also demonstrated by SEC and CD analysis (see 3.3.1). For MubRV, -R8-V, -RV-VI and Nterm, pI values were determined as 4.53, 4.67, 4.54 and 4.60, and were all in good agreement with theoretical values (Figure 3.12 C).



**Figure 3.12 pI determination of Mub repeats by IEF**

IEF PAGE of (A) MubRV, MubR8-V, MubRV-VI, Nterm (1 and 1+ MubRV; 2 and 2+ MubR8-V; 3 and 3+ MubRV-VI; 4 and 4+ Nterm) and (B) MubR5 (two elution peaks) (1 and 1+ MubR5 peak1; 2 and 2+ MubR5 peak 2). (C) Observed pI values determined in comparison to IEF marker by linear regression and theoretical pI values calculated from amino acid sequence using the ExPasy online tool (see 2.6.2).

All tested MUB proteins showed pI values between 4.5 and 4.7 and thus have a net negative charge in the standard buffers used in this study (see 2.1.1).

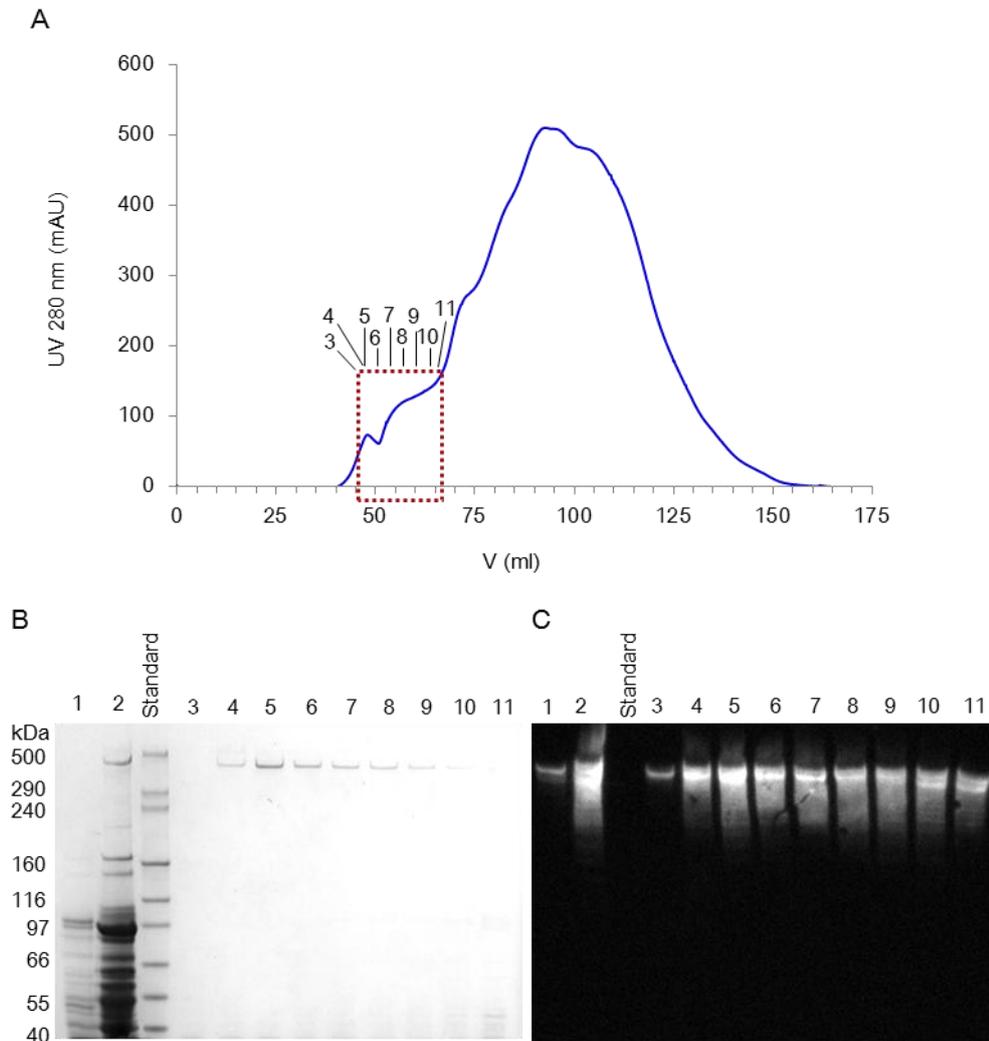
### **3.3 Purification and characterisation of native MUB**

In order to obtain biochemical information on the native MUB protein and to study interaction of the full-length MUB comprising all 14 Mub repeats and the N-terminal domain with potential ligands, we isolated MUB from a *L. reuteri* ATCC 53608 culture.

#### **3.3.1 MUB purification**

The full length native MUB protein is generally covalently attached to the bacterial cell-wall component peptidoglycan via its LPXTG-motif (with X being Q, see GenBank entry AF120104) (see 1.4.2), but under *in vitro* conditions it is also found in lower amounts in the bacterial culture media. The mechanism by which it is released from the cell surface remains elusive.

We developed a method to purify the native mature MUB (353 kDa) from the spent media of a *L. reuteri* ATCC 53608 culture via a multi-step process (see 2.3.2). Briefly, after growth of bacteria in LDMII media at 37°C (see 2.1.4), the culture media was separated from bacterial cells and extensively filtered. The media extract was concentrated and then applied to a SEC for separation of MUB from secreted bacterial proteins (see 2.3.2).



**Figure 3.13 Purification of native MUB by SEC**

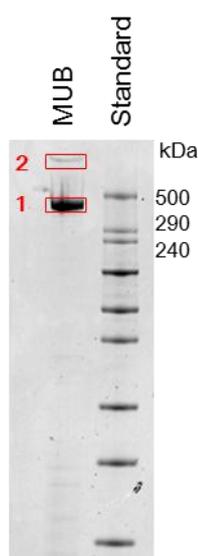
(A) SEC purification profile of native MUB protein and (B) SDS-PAGE analysis of MUB elution fractions (as indicated in A) (1 bacterial pellet, 2 concentrated media extract, 3-11 MUB elution fractions). (C) Western-blot membrane incubated with primary anti-MubRI and anti-MubR5, and secondary anti-rabbit-HRP (1 bacterial pellet, 2 concentrated media extract, 3-11 MUB elution fractions).

The SEC chromatogram of MUB showed a single, broad elution peak (Figure 3.13 A). SDS-PAGE analysis of the peak elution fractions revealed the presence of a single high MW protein band with an apparent MW of about 396 kDa in the early elution fractions (Figure 3.13 B). The theoretical MW of the mature MUB exclusive of the cell wall anchoring motif was determined to be 353 kDa (see 2.6.2). The protein band reacted with anti-MubR5 and anti-MubRI as demonstrated after Western blotting (Figure 3.13 C). The elution of MUB within the void volume of the column suggests the

formation of oligomers or aggregates, which was observed earlier by Roos and Johnsson [283].

### 3.3.2 MUB characterisation

After successful purification of native MUB (see 3.3.1), peptides of pooled elution fractions were initially analysed by MS (in collaboration with Fran Mulholland, Insititue of Food Research, Norwich, UK). Briefly, MUB was separated by SDS-PAGE and the predominant MUB band at about 396 kDa and the fainter protein band above the 500 kDa protein standard band were treated with trypsin before MS analysis (Figure 3.14).



**Figure 3.14 SDS-PAGE of MUB before trypsin digest and MS analysis**

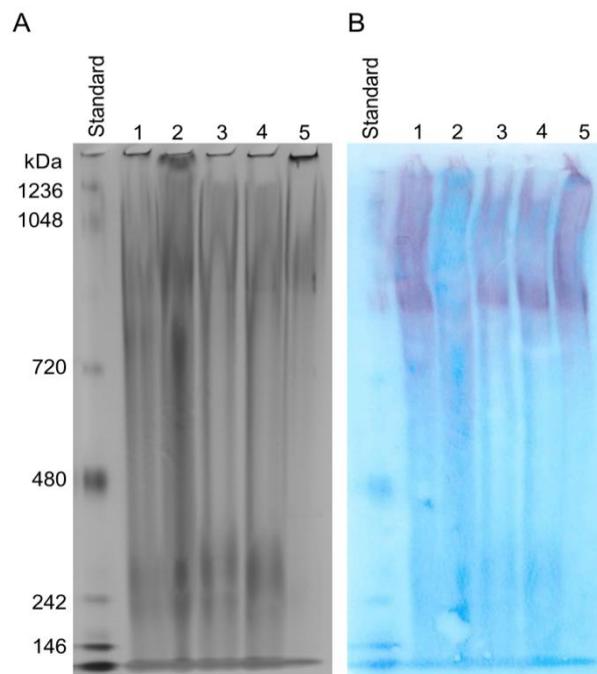
Separation of purified MUB by SDS-PAGE for trypsin digest and MS analysis of stained protein bands 1 and 2.

For the 396 kDa band of MUB, that showed reactivity with anti-MubRI and anti-MubR5 (Figure 3.14) (see 3.3.1), nearly exclusively MUB protein peptides, 243 in total of which 131 were non-duplicates, were identified with a high abundance (score 3176) compared to potential contaminants (score 24). The peptides covered 94.5% of the full length MUB amino acid sequence including the N-terminal methionine (Met) and the MW of MUB was estimated with 358 kDa. No protein peptides were identified for the

last 168 residues at the C-terminus of the protein including the LPQTG-anchor motif, which may explain its presence in the culture media (see 3.3.1). In addition, MUB protein peptides (13 non-duplicates) were present in the second higher MW band in low abundance (score 78) amongst peptides from three additional proteins including a putative glycosyltransferase.

### 3.3.2.1 MUB characterisation by native PAGE and AUC

Purified MUB showed a higher apparent MW compared to its calculated size and SEC suggested the formation of MUB oligomers or aggregates (see 3.3.1). In order to address this matter, different MUB purification samples in PBS were analysed by non-denaturing native PAGE using Tris-glycine buffer (see 2.3.4.1).



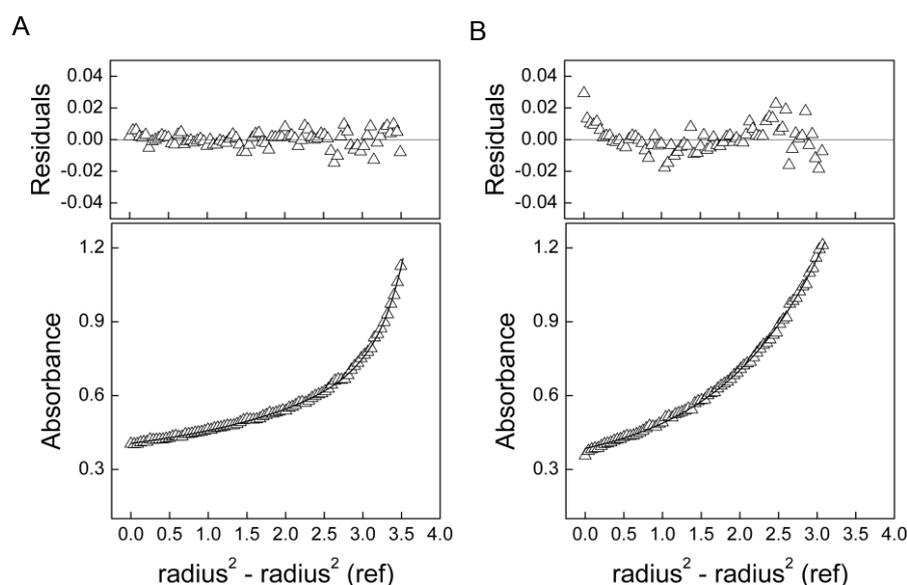
**Figure 3.15 MUB analysis by native PAGE**

(A) Native PAGE analysis of MUB samples of different purification runs (1-5) (signals in black) and (B) Western-blot membrane incubated with primary anti-MubR1 and anti-MubR5, and secondary anti-rabbit-AP (signals in purple, blue background signal derived from Colloidal blue stained native PAGE gel).

No distinct protein bands could be detected on the stained protein gel (see 2.3.4.4) for all tested MUB samples (Figure 3.15 A). In contrast, strongly stained smears were observed between the 1048 and 720 kDa marker band and above 242 kDa.

Anti-MubR5 and anti-MubRI were used for specific MUB detection after Western-blotting and showed only reactivity for protein components above 720 kDa up to 1236 kDa (Figure 3.15 B). Hence, MUB seems to be present in different oligomeric MW species under the experimental conditions, forming at least a trimer and possibly higher oligomers.

In order to further investigate the association and shape of MUB in solution, analytical ultracentrifugation (AUC) experiments were performed in collaboration with Tom Clarke (University of East Anglia, Norwich, UK). For sedimentation equilibrium (SE) experiments, purified and freeze-dried MUB was resuspended in PBS, carbonate or citrate buffer (see 2.4.3). Data scans were recorded for all three conditions, but only data for MUB in PBS and carbonate buffer were of sufficient quality for analysis by UltraScan [332] (see 2.4.3).



**Figure 3.16 Sedimentation equilibrium data fits of MUB**

AUC sedimentation equilibrium data of MUB (triangle) in PBS (A) and carbonate buffer (B) and data fits (black line) to 2 component and 1 component system, respectively.

The data of MUB in PBS were fitted to a two component system with protein species of 160 kDa and 1050 kDa (Figure 3.16 A). In contrast, data for MUB in carbonate buffer was fitted to a one component system with a MW of 250 kDa (Figure 3.16 B). The AUC observations suggested MUB oligomerisation, which seemed to be condition-

dependent and only occurring in standard PBS at pH 7.4 (see 2.1.1). These findings are in line with the native PAGE results described above. The size of the lower MW components observed in PBS and carbonate buffer was significantly lower than the theoretical MW of the 353 kDa MUB, which may indicate protein degradation in these buffer systems or after freeze-drying. MS analysis of MUB in PBS however suggested a MW coverage of about 358 kDa (94.9%) (see 3.3.2). Sedimentation velocity experiments were conducted with MUB in PBS. The radial displacement profiles of different scan intervals, where the interference fringes are plotted against the distance from the rotor centre, showed elongated tails at higher rotor distance till a plateau was reached (see Appendix VI). This AUC profile is characteristic for higher hydrodynamic friction caused by an elongated molecule shape. In addition, a smaller sedimentation coefficient (S) was observed compared to a globular protein of the same MW. For MUB, the sedimentation coefficient, as calculated by a radial derivative (dC/dr) method was 5.7S (see Appendix VI). For comparison, Fibrinogen, a 340 kDa plasma protein, showing a coiled extended fold of 46 nm after X-ray crystallography [362], presents a sedimentation value of 7.9S [363].

### 3.3.2.2 MUB detection by lectins

Protein glycosylation is a common post-translational modification known in eukarya, and has recently been shown to occur in bacteria and archaea [364-365]. Bacterial glycosylation has been described for pathogenic bacteria but glycosylated proteins have recently been identified in *Lactobacillus* species [366-368]. However, the glycosylation pattern and the functional characterisation of lactobacilli glycoproteins remain elusive.

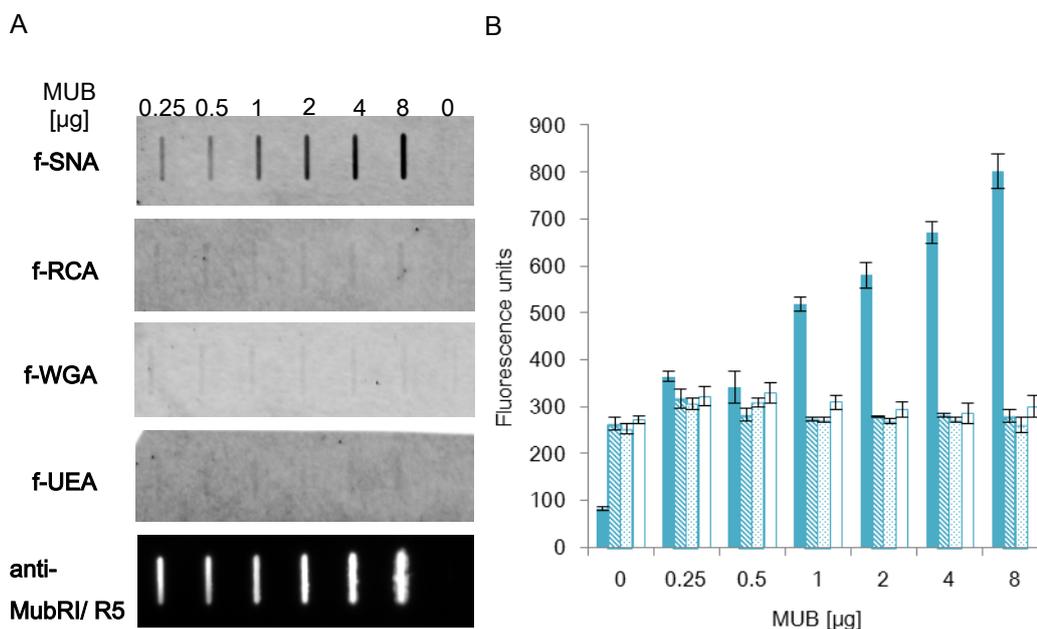
Glycosylated proteins are characterised by an aberrant migration pattern upon SDS-PAGE with a higher observed MW compared to the theoretical value based on amino acid sequence. In order to investigate the potential glycosylation pattern of native MUB isolated from *L. reuteri*, lectin detection experiments were performed. Briefly, increasing amounts of MUB from two independent purifications were slot-blotted onto nitrocellulose and probed with fluorescein isothiocyanate (FITC) labelled lectins f-RCA, f-WGA, f-SNA and f-UEA possessing different sugar specificities (Table 3.3).

Lectin	Abbreviation	Sugar specificity <sup>a</sup>
<i>Ricinus communis</i> agglutinin	RCA	Gal
Wheat germ agglutinin	WGA	GlcNAc, sialylated
<i>Sambucus nigra</i> agglutinin	SNA	$\alpha$ 2-6 sialic acid
<i>Ulex europaeus</i> agglutinin	UEA	$\alpha$ -Fuc

**Table 3.3 Sugar specificity of used lectins**

<sup>a</sup>Galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), and Fucose (Fuc)

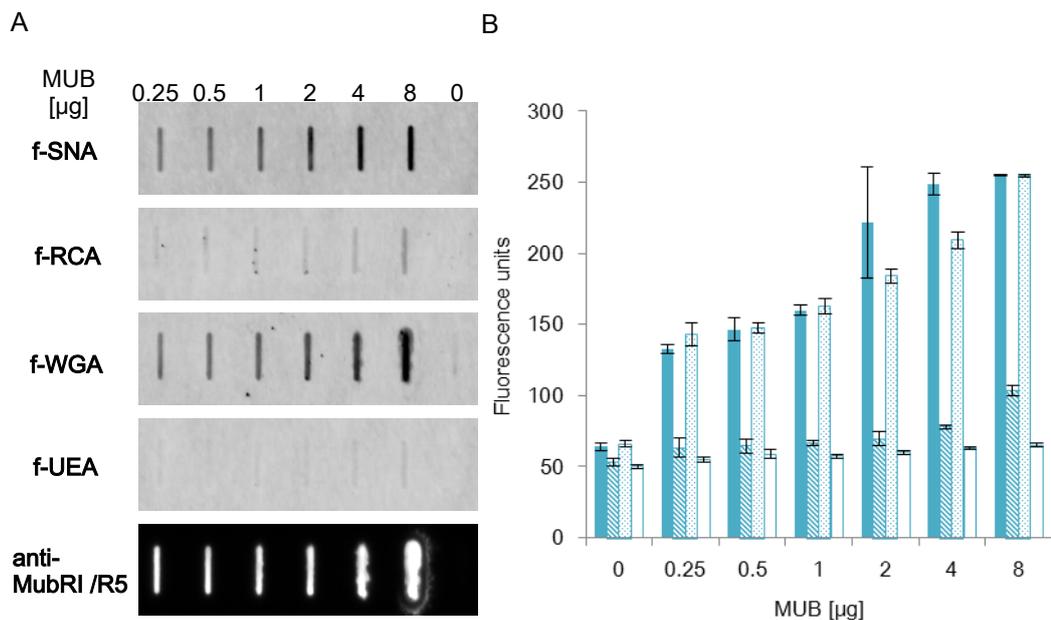
In a first experiment, positive signals were only detected for MUB probed with f-SNA compared to a PBS control (Figure 3.17). Fluorescence signal intensities were semi-quantitatively determined from slot blot membranes and demonstrated a dose-dependent interaction of MUB with the lectin. SNA binds to  $\alpha$ 2-6 sialic acid attached to terminal Gal, and to a weaker extent to  $\alpha$ 2-3 sialic acid.



**Figure 3.17 Detection of MUB (purification batch 1) via lectins**

(A) Slot-blot of purified native MUB probed with FITC labelled f-SNA, f-RCA, f-WGA and f-UEA (signals in black), or incubated with via anti-MubRI and anti-MubR5 followed by anti-rabbit-HRP performed as a control (signals in white). (B) Presentation of quantified fluorescence signals as an average of 5 counts per concentration with standard deviations (blue: f-SNA, blue stripes: f-RCA, blue dots: f-WGA and white: f-UEA).

MUB purified from a different *L. reuteri* ATCC 53608 culture was repeatedly tested for its interaction with f-SNA, f-RCA, f-WGA and f-UEA, as described above. MUB showed equally strong reactivity with f-SNA and f-WGA, and weaker interaction with f-RCA (Figure 3.18) compared to a PBS only negative control. The receptor sugar for WGA is GlcNAc, although interaction with sialylated glycoproteins has also been reported. RCA binds to terminal Gal residues.



**Figure 3.18 Detection of MUB (purification batch 2) by lectins**

(A) Slot-blot of purified native MUB probed with FITC labelled f-SNA, f-RCA, f-WGA and f-UEA (signals in black), or incubated with anti-MubRI and anti-MubR5 followed by anti-rabbit-HRP performed as a control (signals in white). (B) Presentation of quantified fluorescence signals as an average of 5 counts per concentration with standard deviations (blue: f-SNA, blue stripes: f-RCA, blue dots: f-WGA and white: f-UEA).

In summary, native MUB showed reactivity with f-SNA, f-WGA and to a lesser extent with f-RCA, suggesting the presence of sialic acid, GlcNAc or Gal residues in the purified MUB sample. While the reactivity with f-SNA is reproducible for MUB of different purifications, the reactivity of f-WGA and f-RCA seemed to be batch-dependent.

### 3.4 Discussion

Structural information on MUB is restricted to the X-ray crystal structure for the type 2 MubR5 repeat, providing the first structure of a mucin-binding protein (MucBP) domain [282]. We demonstrated the contribution of the cell surface protein MUB of *L. reuteri* ATCC 53068 in the interaction of bacterial cells to the host mucus layer in the GI tract [302]. However, the underlying biochemical binding mechanism and the specificity to ligands remain elusive.

In order to obtain further structural information on different MUB proteins, investigate the domain organisation of MUB and characterise MUB protein-ligand interaction, we successfully cloned, expressed and purified the single type 1 Mub repeats, MubRI and MubRV, the type 2 repeat MubR5, the double mixed or type 1 repeat proteins, MubR8-V and MubRV-VI, the triple Mub type 1 repeat MubRI-II-III and the N-terminal domain proteins, Nterm and NtermMubRI (see 3.1). All recombinant Mub repeat proteins showed a higher apparent MW upon SEC compared to their theoretical MW. These observations suggest an elongated protein shape and thus a similar protein fold of all Mub repeats. CD analysis of MUB proteins demonstrated similar overall secondary structure composition of Mub repeats displaying a high  $\beta$ -sheet content and a very low  $\alpha$ -helix content (see 3.2). MubR5 forms an elongated rod mainly composed of  $\beta$ -sheets with a single short helix when crystallised (see 1.5). In contrast, the secondary structure analysis of the Nterm protein after CD analysis suggested a divergent fold and secondary structural element composition with a higher  $\alpha$ -helix content (see 3.2).

Additionally, we developed a purification protocol to isolate the native full-length MUB from culture media (see 3.3.1). The surface-associated MUB protein carries an LPXTG-motif, which is recognised by an extracellular sortase, a surface peptidase that cleaves between threonine and glycine residues and covalently attaches MUB to cell-wall peptidoglycan [225]. Besides its surface location, MUB is also released into the medium under *in vitro* culture conditions by a yet unknown mechanism. Surface-associated proteins possessing a signal peptide but no surface-retention domains are found in the external culture media [369-370]. The MS analysis of the MUB protein may suggest that the C-terminal protein region containing the anchoring motif is absent in

MUB found in the culture media, but this needs to be further investigated by C-terminal sequencing (see 3.3.2).

Biochemical and biophysical characterisation of native MUB by SEC, native PAGE and AUC sedimentation equilibrium experiments (see 3.3.1 and 3.3.2) suggested the formation of oligomers or soluble aggregates in standard PBS buffer. The S-value of MUB was determined by sedimentation velocity AUC, a hydrodynamic technique providing some structural analysis of protein shape, as the S-value increases with protein mass but decreases with increasing protein asymmetry [363]. MUB showed a small S-value compared to its MW, further suggesting an elongated shape in solution.

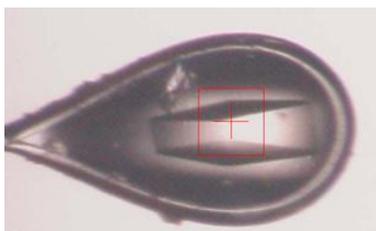
Lectin staining of MUB revealed reactivity with SNA, WGA and to a lower extent RCA, suggesting the presence of sialylated glycans, especially  $\alpha$ 2-6-linked NeuAc, terminal GlcNAc, and Gal or GalNAc structures (see 3.3.2). Putative glycosyltransferases were identified in the genome of *L. reuteri* ATCC 53608 after completion of the whole genome sequencing and gene annotation (unpublished data), which may potentially be involved in the glycosylation of host proteins. Only a few *Lactobacillus* glycoproteins have been described to date including S-layer proteins, a bacteriocin and the major autolysin Acm2 of *L. plantarum*, and the major secreted protein Msp/p75 of *L. rhamnosus* GG [367-368, 371-372]. Glycosylation of these proteins was identified and initially characterised by Periodic-acid Schiff (PAS) stain, lectin detection and MS glycopeptide analysis. Lectin recognition assays using succinylated WGA (GlcNAc), ConA (Man, Concavalin A), UEA (Fuc) MAA ( $\alpha$ 2-3 sialic acid, *Maackia amurensis agglutinin*) and SNA ( $\alpha$ 2-6 sialic acid) indicated the presence of GlcNAc on Acm2 and Man residues on Msp/p75. MS analysis provided evidence for the *O*-glycosylation of Acm2 and Msp/p75 presumably via serine residues. *O*-glycosylation was also observed for serine-rich repeat (SRR) proteins, which are present in Gram-positive streptococci, staphylococci and lactobacilli. Streptococcal SRR glycoproteins contain *O*-linked GlcNAc and GalNAc residues [373]. The predominant glycan in flagella of the gastric pathogens *Campylobacter jejuni* and *Helicobacter pylori* is pseudaminic acid, which is structurally related to sialic acid. Here we suggest, that the MUB protein of *L. reuteri* ATCC 53608 may be glycosylated, but future work is needed to proof this hypothesis including MS glycopeptide analysis and metabolic glycoprotein labeling.

## CHAPTER 4 STRUCTURAL CHARACTERISATION OF MUB AND MUB REPEATS

The previously reported crystal structure determination of the Mub type 2 repeat MubR5 of the cell surface adhesin MUB revealed two structural domains, B1 (N-terminal) and B2 (C-terminal) (see 1.5) [282]. B1 showed structural similarity to an Ig-binding protein and MubR5 was shown to bind to Ig molecules *in vitro* [282], whereas B2 demonstrated structural homology with mucin-binding proteins (MucBP) as annotated in the Pfam database (PF06458) [374]. Here, the ability of MubR5 to bind to mucin glycans was investigated by crystal soaking and co-crystallisation studies with different sugars with the aim to identify potential ligands, reveal the topology of binding sites and understand the biochemical basis of protein-sugar interaction. As structural information on MUB is so far limited to a Mub type 2 repeat, we additionally aimed to determine the X-ray crystal structure of a Mub type 1 repeat. Furthermore, in order to obtain additional information on the domain organisation of MUB in solution, small angle X-ray scattering (SAXS) experiments of single, tandem and triple Mub repeats as well as the N-terminal domain were performed.

### 4.1 MubR5 crystal soaking and co-crystallisation studies

For crystal soaking experiments of MubR5 (183 residues) with potential sugar ligands, recombinant MubR5 was purified to homogeneity, dialysed to remove salt and concentrated to 8 mg/mL (see 3.1.1). Crystals were grown in 20% to 30% (w/v) PEG (polyethyleneglycol) 3,350 and 0.2 M magnesium formate as precipitant, which showed improved crystal growth when compared to the alternative growth conditions published earlier (Figure 4.1) [282]. Crystals were soaked in fucose (Fuc), mannose (Man), galactose (Gal), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), *N*-acetylneuraminic acid (Neu5Ac) and *N*-acetylglucosamine (LacNAc) solution containing 25% to 30% (w/v) PEG 3350 as a cryoprotectant for several minutes at concentrations ranging from 1 mM to 500 mM (see 2.5.1.2). The sugar concentration was dependent on crystal stability in the sugar solution used and was particularly low (1mM and 5mM) for Neu5Ac and LacNAc. Single wavelength diffraction data sets were collected for crystals soaked with 200 mM Fuc, 200 mM Man, 200 mM GalNAc, 500 mM GlcNAc and 5 mM LacNAc to resolutions of 1.7 to 1.8 Å (see Appendix VII).



**Figure 4.1 MubR5 crystal soaked with Man**

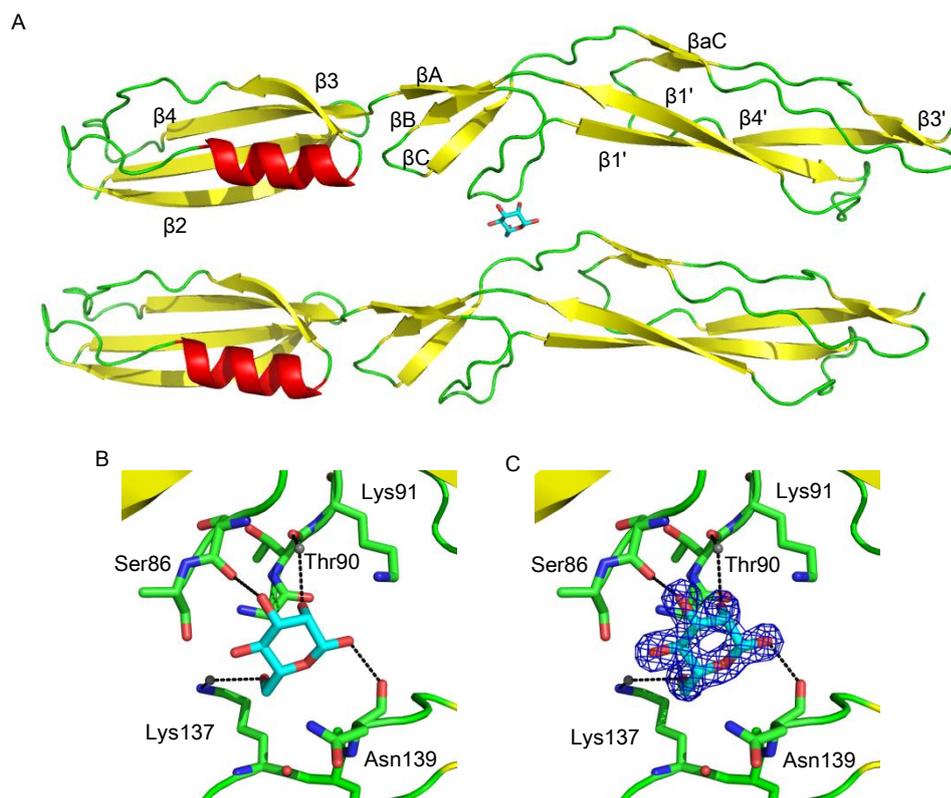
MubR5 crystal grown in 20% to 30% (w/v) PEG 3350 and 0.2 M magnesium formate and soaked with 200 mM Man in harvesting loop flash-cooled to 100 K in a stream of gaseous nitrogen before exposure to X-ray beams.

For co-crystallisation experiments, crystallisation screens were performed with LacNAc, Neu5Ac and Man at sugar concentrations of 10 mM, 50 mM, 100 mM and 200 mM (only Man) (see 2.5.1.2). Crystal growth was observed for all tested ligands and concentrations with 25% (w/v) PEG 4,000, 0.2 M magnesium chloride and 0.1 M Tris (pH 8.5). Single wavelength diffraction data sets of sufficient diffraction quality were collected for crystals grown in the presence of 10 mM Man, 50 and 100 mM LacNAc, and 100 mM Neu5Ac (see Appendix VII). The X-ray diffraction images of all collected SAD data sets were indexed and integrated with MOSFLM in the primitive orthorhombic space group  $P2_12_12_1$  and data subsequently reduced using SCALA [333-334]. The solvent content of the crystals was estimated to be 50% (v/v) with two MubR5 molecules in the asymmetric unit (ASU) [339]. The MubR5 structure (PDB entry 3I57) was used as a model for molecular replacement (MR) using MOLREP and PHASER [340-341]. Alternating rounds of model refinement (without water molecules) and manual model building were performed using REFMAC and COOT [343-344]. The presence of sugar molecules for sufficiently refined protein models of soaked and co-crystallised MubR5 crystals was investigated in COOT using the 'unmodelled blobs' tool in reference to the  $\sigma$ -weighted  $2mF_{obs}-DF_{calc}$  electron density map. The only sugar molecule observed in an electron density map was Man after MubR5 crystal soaking. After a single Man molecule was added to the model, the final MubR5 structure was refined with an  $R_{cryst}$  of 18.9% and an  $R_{free}$  of 21.9% to a resolution of 1.7 Å using PHENIX (Table 4.1) [348]. The average temperature factor of Man is 15.1 and the residues in the near surrounding have temperature factors between 9.4 and 10.7 (Figure 4.2 C).

<b>MubR5 (Man)</b>	
<b>Data collection</b>	
Beamline	Diamond i02
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell parameters: a, b, c (Å), (°)	45.4, 45.8, 197.3; α=β=γ=90
Wavelength (Å)	0.9795
Resolution (Å)	45.4-1.7 (5.3-1.7)
R <sub>sym</sub> (%)	6.3 (21.5)
I/σI	17.1 (5.2)
Unique reflections	47572 (6786)
Completeness (%)	99.7 (99.4)
Multiplicity	3.8 (3.8)
Overall B-factor (Å <sup>2</sup> )	12.3
<b>Refinement statistics</b>	
Molecules per AU	2
Total atoms	3815
Water molecules	921
R <sub>cryst</sub> (%)	18.9 (24.0)
R <sub>free</sub> (%)	21.9 (26.8)
<b>Ramachandran analysis</b>	
Most favoured	99.5
Outliers	0.0
<b>r.m.s.d.</b>	
Bonds (Å)	0.007
Angles (°)	1.023
Planes (Å <sup>2</sup> )	0.005
Mean atomic B-factor (Å <sup>2</sup> )	12.1

**Table 4.1 MubR5 (Man) data collection and refinement parameters**

Man is localised in the inter domain region (IR domain) connecting the B1 and B2 domain of MubR5 sandwiched between two adjacent protein molecules in the crystal lattice (Figure 4.2 A). The amino acid residues in contact with the Man molecule are located in the loop connecting βA and β1' of one MubR5 molecule and in the loop between βC and βaC of a second R5 repeat. The main chain carbonyl group of Ser86 and Thr90 of the first MubR5 forms hydrogen bonds with Man as well as the side chain of Lys137 and the main chain carbonyl group of Asn139 of the second MubR5 molecule (Figure 4.2 B and C). Two water molecules complete the coordination sphere of the Man ligand.



**Figure 4.2 MubR5 structure with mannose (Man)**

(A) Representation of two adjacent MubR5 molecules in contact with a single Man ligand (light blue, sticks) with  $\alpha$ -helices in red and  $\beta$ -sheets in yellow. (B) Man binding region with surrounding residues (sticks) and water molecules (grey spheres) involved in ligand interaction (C) and Man overlaid with  $2mF_{obs} - DF_{calc}$  electron density map contoured at  $1.0 \sigma$  level (dark blue).

The fact that Man was identified as the only ligand in the soaked and co-crystallised MubR5 data sets was surprising, as it is typically found in complex N-glycans, which are less abundant in mucin glycoproteins (see 1.2). However, the location of Man between two adjacent MubR5 molecules seems to indicate that Man does not interact with a distinct carbohydrate binding site, but rather suggests a role of MUB in promoting interaction between MUB molecules on the same bacterial cell or from different cells.

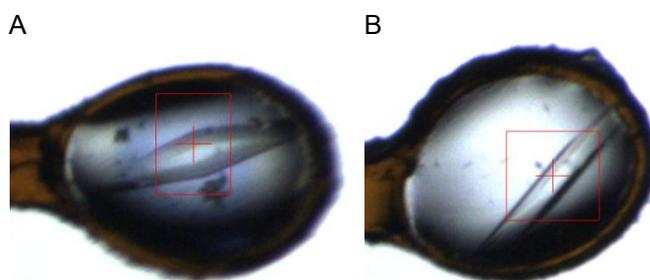
#### 4.2 Structure determination of a Mub type 1 repeat protein

While the Mub type 2 repeats (R1 to R8) of MUB show high sequence identity (84% to 100%), the Mub type 1 repeats (R1 to RVI) are more diverse with sequence similarities of 29% to 88% [282]. In order to investigate, whether Mub type 1 repeats possess

a similar protein fold compared to the type 2 MubR5 repeat, crystallisation experiments were performed using the recombinant type 1 repeats MubRI and MubRV, which share 32.4% and 42.4% sequence identity with MubR5, respectively.

#### 4.2.1 Crystallisation experiments of MubRI

For crystallisation growth condition screens, the recombinant MubRI protein (194 residues) was purified by IEC and SEC, dialysed into sodium phosphate or alternatively in Tris-HCl buffer and concentrated to about 18 mg/mL (see 3.1.1). Crystal structure screens were set up at 4° and 16°C, but did not result in the growth of single crystals of good quality (see 2.5.1.1). Crystal optimisation for various growth conditions including the variation of pH, reservoir solution, salt concentration and precipitation solution was performed. Finally, single, diamond shaped crystals grew in the following precipitation solution: 0.1 M HEPES (pH 7.5), 10% (v/v) 2-propanol/ 20% (w/v) PEG 4000 at 4°C after crystal seeding (Figure 4.3).



**Figure 4.3 MubRI crystals in harvesting loop**

(A+B) Orthogonal views on a diamond shaped MubRI crystal of 210  $\mu\text{m}$   $\times$  50  $\mu\text{m}$  in size grown in 0.1 M HEPES (pH 7.5), 10% (v/v) 2-propanol/ 20% (w/v) PEG 4000 at 4°C in a crystal harvest loop flash-cooled to 100 K in a stream of gaseous nitrogen before exposure to X-ray beams.

Native crystals as well as crystals soaked with ytterbium and mercury chloride were tested for their X-ray diffraction ability. Ytterbium and mercury chloride were used as anomalous scatters to accomplish phasing of crystal diffraction data. Alternatively, MR with various MubR5 models derived from the coordinate file (PDB entry 3I57) was performed. Four SAD data sets were collected for MubRI, two for native crystals and two for halide and heavy atom soaked crystals at resolutions between 2.0 and 2.6 Å (see Appendix VIII).

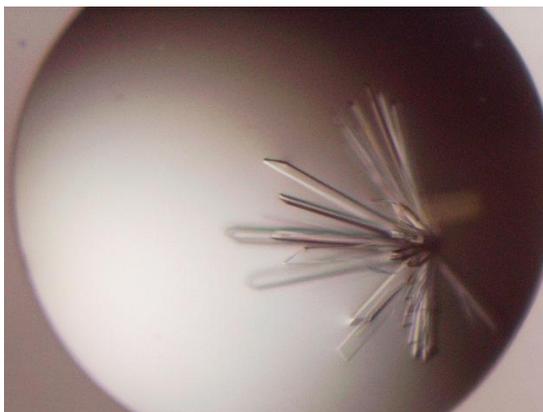
The solvent content analysis of crystal forms giving all 4 collected data sets suggested the presence of two protein molecules per asymmetric unit cell (ASU) with an estimated solvent content of 50% (v/v). The inspection of the anomalous probability plot produced by SCALA suggested the absence of an anomalous signal for the soaked crystals after image data integration and data processing via MOSFLM and SCALA [333-334] in space group P222. Attempts to obtain initial phase information by MR with different MubR5 models using MOLREP and PHASER failed. Data quality analysis by XTRIAGE suggested weak translational pseudosymmetry for all MubRI data files (see Appendix VIII). Pseudosymmetry can occur when more than one molecule is present in the ASU and the non-crystallographic translational symmetry operator relating these molecules is close to a true crystallographic symmetry operator. This can lead to an incorrect unit cell and false space group assignment after autoindexing, making a structure solution difficult or even impossible to obtain [375].

Pseudosymmetry is a crystal property and may be avoided if an alternative crystal form is available. Unfortunately, only a single crystal form was observed for MubRI after crystal growth optimisation (as described above). Hence, attempts to solve the structure of MubRI were abandoned and the recombinant MubRV protein was alternatively used in further crystallisation experiments.

#### **4.2.2 Structure determination of MubRV**

In parallel to MubRI crystallisation experiments (see 4.2.1), crystal screens were set up for the recombinant type 1 MubRV (184 residues) purified via IEC and SEC, after dialysis into sodium phosphate buffer and concentration to 12 mg/mL (see 3.1.1).

Crystals grew as long rods from a single nucleation point for a variety of crystallisation conditions (Figure 4.4), but were readily separated allowing crystal harvest for diffraction analysis. Diffraction data sets were collected for a number of growth conditions. The best X-ray data set was collected to a resolution of 2.6 Å for a crystal grown in 0.2 M ammonium acetate and 24% (w/v) PEG 3,350 at 16°C and used for further data processing.



**Figure 4.4 MubRV crystal**

Representative image of MubRV crystals grown from a single nucleation point.

The initial indexing of the MubRV data set via XDS [376] suggested two potential space groups,  $C222_1$  and  $P222_1$ . For correct space group assignment, the diffraction data were analysed using ZANUDA and LABELIT [346], revealing a higher probability for the space group  $C222_1$ . A similar outcome was achieved by POINTLESS [334]. The data were therefore indexed in a C-centred orthorhombic lattice with cell parameters of  $a=31.4$ ,  $b=271.7$  and  $c=142.9$  (Table 4.2). Data integration and reduction was performed by the Xia2 automated data reduction system [335]. The data did not show any signs of pseudosymmetry or twinning when analysed by XTRIAGE in PHENIX [337]. The solvent content of the crystal was estimated with 50% (v/v) for three molecules in the ASU, however we found only two molecules with a solvent content 67% (v/v).

<b>MubRV (PDB entry 4MT5)</b>	
<b>Data collection</b>	
Beamline	Diamond i04
Space group	C 222 <sub>1</sub>
Cell parameters: a, b, c (Å), (°)	31.4, 271.7, 142.9; α=β=γ=90
Wavelength (Å)	0.9205
Resolution (Å)	49.2-2.6 (11.6-2.6)
R <sub>sym</sub> (%)	8 (36)
I/σI	16.6 (5.5)
Unique reflections	19614 (1413)
Completeness (%)	99.9 (99.9)
Multiplicity	6.0 (6.2)
Overall B-factor (Å <sup>2</sup> )	48.0
<b>Refinement statistics</b>	
Molecules per AU	2
Total atoms	3010
Water molecules	162
R <sub>cryst</sub> (%)	21.1 (31.9)
R <sub>free</sub> (%)	26.9 (35.3)
<b>Ramachandran analysis</b>	
Most favoured	95.9
Allowed	3.9
Outliers	0.3
<b>r.m.s.d.</b>	
Bonds (Å)	0.008
Angles (°)	1.083
Planes (Å <sup>2</sup> )	0.005
Mean atomic B-factor (Å <sup>2</sup> )	36.1

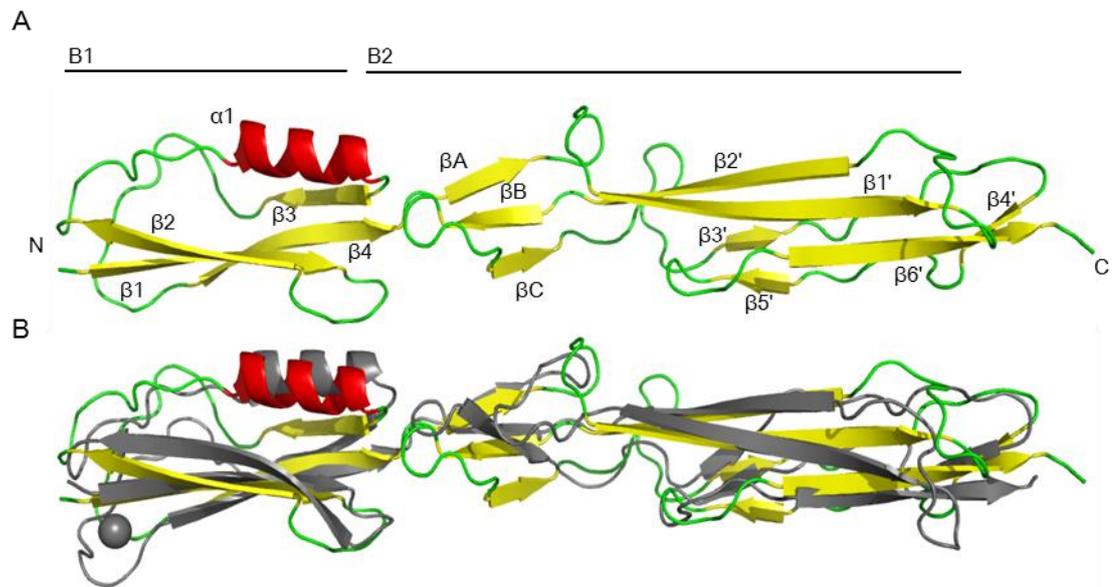
**Table 4.2 Data collection and refinement statistics for MubRV**

In an attempt to obtain initial phases by MR, a structural alignment model for MubRV was generated by CHAINSAW [342] after sequence alignment of MubR5 and MubRV using ClustalW (see 2.6.2). However, MR using PHASER [341] did not produce a convincing solution, as good coverage of Ca-atoms with electron density was only observed for the N-terminal B1 domain. The C-terminal domain of MubRV showed 14% sequence identity to the MucBP domain of the adhesion protein PEPE\_0118 (PDB entry 3LYY) of *Pedococcus pentosaceus* compared to 13% to the B2 domain of MubR5. Hence, the MucBP of PEPE\_0118 and the C-terminal MubR5 B1 domain (residues

1-75) were used as models in MR by PHASER resulting in a convincing solution. The inter domain region was built both automated using AUTOBUILD and manually in COOT with reference to the  $\sigma$ -weighted 2mFobs-DFcalc and mFobs-DFcalc Fourier electron density maps, and refined in REFMAC [343-344, 348].

The final crystal structure of MubRV was refined at a resolution of 2.6 Å in space group C222<sub>1</sub> with an R<sub>cryst</sub> of 21.1% and an R<sub>free</sub> of 26.9%. The two MubRV molecules present in the ASU comprise 184 residues including the N-terminal methionine (Met). They are highly similar showing a Z-score of 20.1 (r.m.s.d. 0.9 Å) over the C $\alpha$ -atoms when aligned by DaliLite (see 2.6.2).

MubRV folds to form an elongated structure 110 Å in length and 24 Å in diameter. It shows the same domain organisation as MubR5, comprising an N-terminal B1 domain and a C-terminal B2 domain with an inter domain region (IR domain) (Figure 4.5 A). The Mub type 1 and Mub type 2 structures share high structural similarity with a Z-score of 15.1 over 176 aligned residues (r.m.s.d. 4.1 Å). Like MubR5, the MubRV B1 domain has an ubiquitin-like  $\beta$ -grasp fold containing two pairs of antiparallel  $\beta$ -strands in a 4-stranded sheet connected by an  $\alpha$ -helix, which is similar to that found in members of the immunoglobulin (Ig)-binding superfamily [377]. However, in contrast to MubR5, there was no evidence for a Ca<sup>2+</sup>-ion coordinated by residues of the loop connecting strands  $\beta$ 3 and  $\beta$ 4 in the electron density map of MubRV (Figure 4.5 B). The MubRV repeat lacks the three residues (Asp60, Asp62 and Asn65) present in the loop region of MubR5 which mediate binding of the Ca<sup>2+</sup>-ion. The B2 domain of MubRV is classed as a mucin-binding domain (MucBP) as annotated in the Pfam database (PF06458) [374], with a modified ubiquitin-like  $\beta$ -grasp fold, in which the outer strands of the 4-stranded  $\beta$ -sheet are connected by a  $\beta$ -strand ( $\beta$ 3') instead of an  $\alpha$ -helix as in the B1 domain. This connecting strand together with an additional  $\beta$ -strand ( $\beta$ 5'), that is located between  $\beta$ 4' and  $\beta$ 6', forms a third antiparallel  $\beta$ -sheet and is referred as the IR domain. The ubiquitin-like  $\beta$ -grasp fold is involved in diverse functions with variations in the core protein fold [378]. Interestingly, the  $\beta$ -grasp fold also shares similarities with  $\beta$ -grasp domains and  $\beta$ -sheet clefts, commonly involved in the binding of sialic acid containing carbohydrate ligands [318].



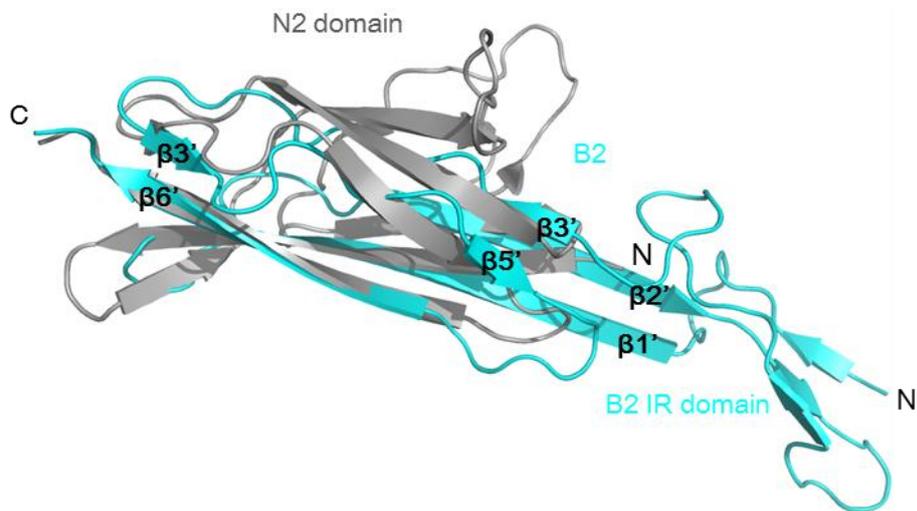
**Figure 4.5 X-ray crystal structure of the Mub type 1 repeat MubRV**

(A) Protein fold of MubRV with  $\alpha$ -helix and  $\beta$ -sheets coloured red and yellow, respectively. The N- and C-termini of the protein and the main structural elements are labelled. (B) Superposition of type 1 MubRV (yellow, red) and type 2 MubR5 (grey, PDB entry 3517) by DaliLite with the  $\text{Ca}^{2+}$ -ion presented as a grey sphere.

When performing a comparison of MubRV with structures in the protein databank (PDB) using DALI [358], the B1 domain of MubRV shows structural similarity to the immunoglobulin (Ig)-binding protein L (PpL) of *Peptostreptococcus magnus* (Z-score 5.4, 15% sequence identity, r.m.s.d. 2.9 Å), as reported earlier for MubR5 (see 1.5) [316]. However, a closer structural homologue in the PDB, characterised by an Ig-like fold, is the B-repeat of the *Listeria* invasion protein internalin B (InIB) (PDB entry 2KVZ) (Z-score 5.9, r.m.s.d 2.6 Å, 13% sequence identity) with unknown receptor specificity. In addition, MubRV shows high structural homology to the cell-surface adhesin Spr1345 of *Streptococcus pneumoniae*, a structure solved by MR using the MubR5 B2 coordinates [240]. The structural alignment of the B2 domain of MubRV with the MucBP domain of Spr1345 (PDB entry 3NZ3) shows an r.m.s.d. of 1.4 Å and a Z-score of 11.6 at 37% sequence identity. MucBP of *Strep. pneumoniae* has been demonstrated to bind to mucins and polysaccharides [239].

More striking, however, is the structural similarity of the B2 domain of MubRV to a number of extended modular adhesins from Gram-positive pathogens. These include pilin proteins such as GBS52 of *Strept. agalactiae* (PDB entry 3PHS) (Z-score 5.4), BcpA of *Bacillus cereus* (PDB entry 3KPT) (Z-score 4.7), RrgB of *Strep. pneumoniae*

(PDB entry 3RPK) (Z-score 4.6), Spy0128 of *Strep. pyrogenes* (PDB entry 3B2M) (Z-score 4.2) and SpaA of *Corynebacterium diphtheriae* (PDB entry 3HR6) (Z-score 4.0) as well as the MSCRAMM (microbial surface components recognizing adhesive matrix molecule) CnaB of *Staphylococcus aureus* (PDB entry 1D2P) (Z-score 2.5) [261, 379-383]. These proteins all contain a similar domain organisation of two to 4 Ig-like domains with maximum dimensions between 85 Å and 134 Å as observed by X-crystallography. The Ig-like fold is a  $\beta$ -sandwich fold of at least 7 predominantly anti-parallel  $\beta$ -strands grouped into two  $\beta$ -sheets. Several distinct topology variants of the Ig-like fold can be distinguished [320, 384-385].



**Figure 4.6 Overlay of the B2 domain of MubRV with the N2 domain of GBS52 pilin**  
 Superposition of the B2 domain (inclusive the IR domain) of MubRV (blue) and the N2 domain of the GBS52 pilin of *Strep. agalactiae* (grey) (PDB entry 3PHS) using DaliLite. B-strand in the B2 domain (red) and N- and C-termini are labelled.

The pilin GBS52 is the closest structural homologue of MubRV and folds in two Ig-like domains, N1 and N2. The latter shows a structural alignment Z-score with the B2 domain of Mub-RV of 5.4 and an r.m.s.d. of 2.4 Å over 64 aligned Ca-atoms (Figure 4.6) [261]. Both N domains show a typical Ig-fold of 7  $\beta$ -strands, but display an alternative arrangement in the order of their  $\beta$ -strands related to a CnaB topology. Interestingly, Ig-like domains of GBS52, BcpA, RrgB, Spy0128 and SpaA all describe a CnaB topology, which was first identified in the collagen binding protein Cna B-region of *Staph. aureus*, to which MubRV shows some level of structural similarity

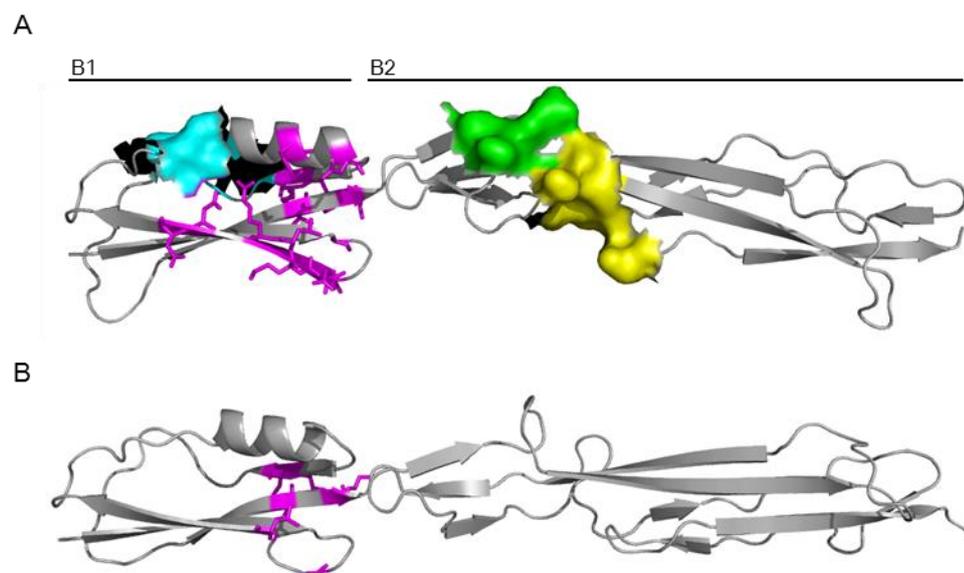
(Z-score 2.5, r.m.s.d. 4.5 Å , 45 aligned Ca atoms). The CnaB topology, albeit reminiscent of the Ig-like fold type C or IgG fold (with ABED/CFG topology), displays an inverse relation of  $\beta$ -strands in a four-(DAGF) and a three-(CBE)  $\beta$ -sandwich arrangement, and is hence designated as an IgG-rev (Ig-like) fold [320, 379]. The  $\beta$ -sandwich in the B2 domain of MubRV (exclusive of the IR domain), comprising the 4-stranded  $\beta$ -sheet ( $\beta 2'$ ,  $\beta 1'$ ,  $\beta 6'$  and  $\beta 4'$ ) and 2  $\beta$ -strands ( $\beta 3'$  and  $\beta 5'$ ), shows a different BAFD/EC topology (Figure 4.6).

In summary, the MubRV type 2 and MubR5 type 1 proteins show a conserved protein fold, suggesting a similar physiological function on the bacterial cell-surface. They possess structural homology to the Ig-binding protein L and the MucBP domain Spr1345 of *Strep. pneumoniae* indicating their involvement in Ig-binding and potential interaction with mucins and glycans. Additionally, MubRV shows structural similarity to cell surface adhesins, pili and MSCRAMM, present in Gram-positive GI pathogens, perhaps suggesting a common niche at the mucosal surface.

#### **4.2.3 Functional annotation analysis of MubRV and MubR5**

In order to further investigate the structure-function relationship of Mub repeats, the crystal structures of MubRV and MubR5 were analysed using the MarkUS function annotation server [386]. The MarkUS server identifies structural neighbours that share a minimum of three secondary structure elements for annotation of common molecular function.

The PredUS analysis tool predicted potential protein-protein interaction sites for MubR5 and MubRV that are exclusively located in the N-terminal B1 domain of both molecules including residues of the  $\alpha$ -helix and mainly of the  $\beta$ -strand  $\beta 2$  and  $\beta 4$  of MubR5, and  $\beta 3$  and  $\beta 4$  of MubRV (Figure 4.7 A and B) [387]. The predicted site is partially identical to the possible interaction site of MubR5 with Igs based on structural alignment and binding modelling with the Ig-binding protein L [282].



**Figure 4.7 PredUS and SCREEN analysis of MubR5 and MubRV**

(A) MubR5 and (B) MubRV structures with protein-protein interaction sites predicted by PredUS (magenta) and solvent accessible cavities predicted by SCREEN shown as surfaces (1 blue, 2 yellow, 3 green) (in A).

Interestingly, three solvent accessible cavities of 7 (cavity 1) and 6 (cavity 1 and 2) residues (Figure 4.7) that may function as non-protein ligand interaction sites were identified for MubR5 but not for MubRV by SCREEN [388]. Cavity (1) is located in the B1 domain of MubR5 and comprises residues preceding the  $\alpha$ -helix and in the loop connecting strands  $\beta$ 2 and  $\beta$ 3 (Figure 4.7 A). It shows a diameter of 7.2 Å, a maximum depth of 3.1 Å and a surface area of 27.8 Å<sup>2</sup>. The other two adjacent cavities (2 and 3) show diameters of 6.1 and 5.1 Å, maximum depth of 3.6 and 3.3 Å, and areas of 24.0 Å<sup>2</sup> and 20.5 Å<sup>2</sup>, respectively. They are located in the IR domain and the B2 domain, where interaction of MubR5 with Man was detected by X-ray crystallography (see 4.2.1). However, they do not comprise the residues that were involved in the interaction with Man. These findings may indicate a functional difference between both Mub type repeats in ligand recognition and the potential interaction of Mub repeats with different types of ligands.

### 4.3 Solution structure determination and domain organisation of MUB

The determination of crystal structures of both Mub type repeats provided valuable information on protein architecture at high resolution but is limited to the low energy state of these proteins in a rigid crystal lattice. In order to obtain information on the

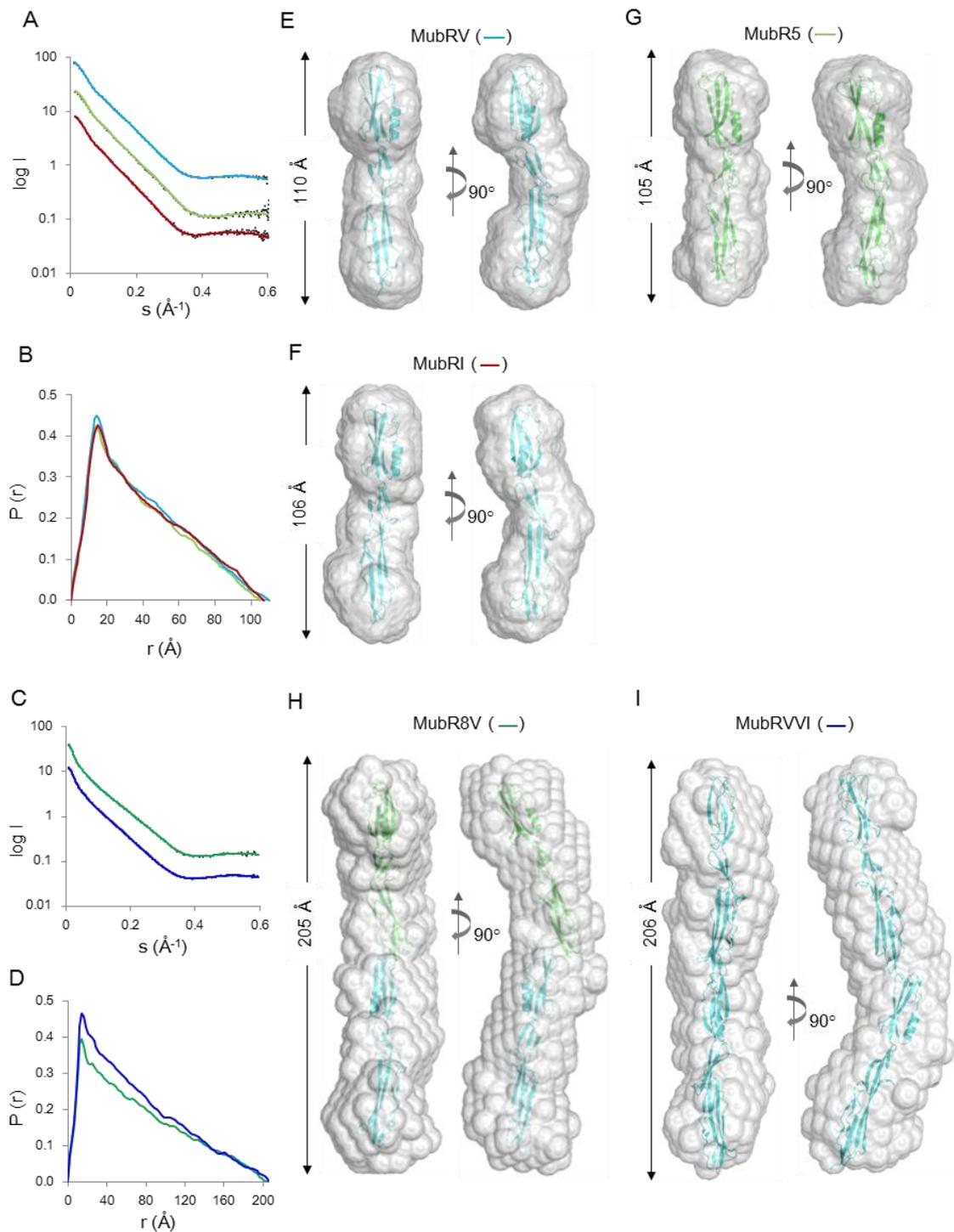
properties of individual Mub repeats freely moving in solution, and of their intra- and inter-domain organisation and flexibility, low resolution SAXS experiments were conducted. In addition, since no structural information is currently available on multiple Mub-repeat proteins, repeat assembly and on the N-terminal region of the native MUB, tandem repeats and the Nterm protein were also analysed by SAXS.

For SAXS analysis, the single Mub repeats, MubR5, -RV and -RI, the type 1 and mixed type double repeats, MubRV-VI and Mub8-V, the triple repeat MubRI-II-III and the Nterm protein, were purified to homogeneity by IEC or IMAC followed by SEC (see 3.1.1 and 3.1.2). The single and double repeats as well as the Nterm domain were dialysed into Tris-HCl and MubRI-II-III was exchanged into 4 different buffers, Tris-HCl, Tris-NaCl (150 mM NaCl), sodium phosphate and PBS. All proteins were concentrated to 10 mg/mL (18 mg/mL for MubRI). The single and double repeat proteins were snap-frozen, and the Nterm protein and MubRI-II-III were stored at 4°C.

Analysis of purified Mub proteins by dynamic light scattering (DLS) (see 2.4.4) revealed the presence of a single protein species for all tested proteins and buffers accounting for 99.9 to 100% of the total mass of the sample and a polydispersity of 13.1 to 23.4%. The presence of other protein species, amounting to only 0.1% of total mass, was low and may indicate a very small amount of protein aggregation. The DLS data demonstrated high homogeneity, low polydispersity and high stability of all tested MUB proteins in the sample buffers.

Scattering curves for MubR5, MubRV, MubRI, MubR8-V, MubRV-VI and Nterm in Tris-HCl were recorded in a concentration range of 0.6 to 9.0 mg/mL. Three data sets were collected for MubRI-II-III in Tris-HCl, sodium phosphate buffer and PBS, the latter two supplemented with 2 mM DTT to reduce the effects of low levels of radiation damage, in a protein concentration range of 0.6 to 0.9 mg/mL (in Tris-HCl) or 0.5 to 5 mg/mL (in sodium phosphate or PBS) (see 2.5.2.1). The scattering data for the triple Mub domain in PBS (2 mM DTT) showed the highest quality of all three collected data sets and was used for further data processing. The scattering profiles of all Mub proteins and the Nterm protein were analysed using PRIMUS and a final scattering curve merged from three individual curves covering concentrations of 1.3 to 4.3 mg/mL (see Appendix IX) (see 2.5.2.2).

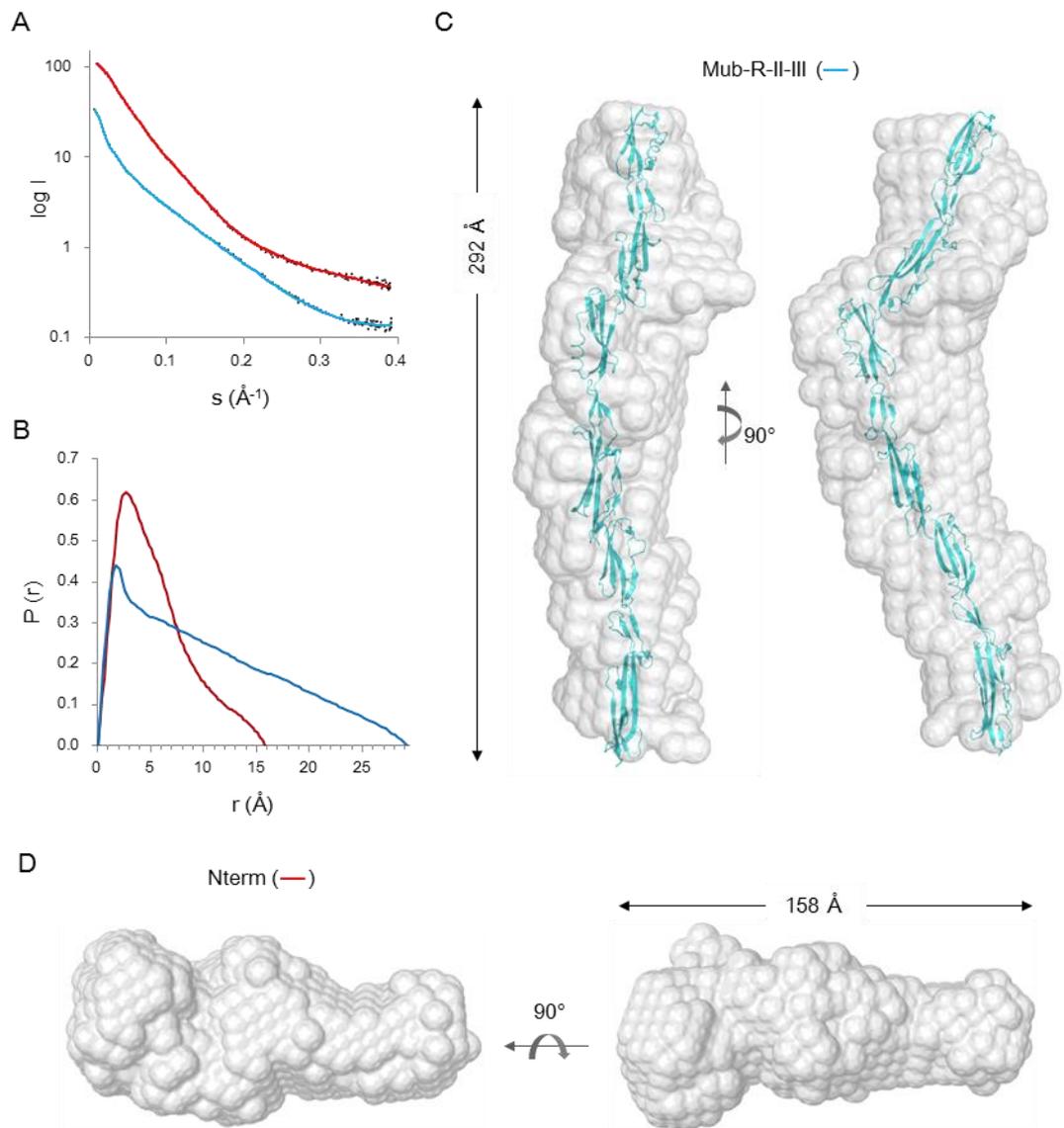
The Kratky analysis of the merged scattering curves of the Mub repeats, where  $s^2I(s)$  is plotted against  $s$ , revealed a shape characteristic for extended or partially flexible molecules with an increasing tail in the larger  $s$ -range (see Appendix X). In contrast, for the Nterm protein, a bell-shape curve with a plateau in the larger  $s$ -range was observed in the Kratky plot, indicating a different protein shape in solution compared to the Mub repeats (see Appendix X). Using the online tool PONDR for structural disorder prediction in the sequence of Mub-repeat proteins, less ordered stretches of residues were identified for all repeats (data not shown) [389].



**Figure 4.8 SAXS data analysis and shape reconstruction of single and double Mub repeats**  
 Experimental scattering curves for (A) the Mub type 1 repeats Mub-RV (light blue) and Mub-RI (red), the Mub type 2 repeat Mub-R5 (light green) and (C) the tandem Mub repeats MubR8-V (dark green) and MubRV-VI (dark blue) shown as the logarithm of the scattering intensity  $I$  (black dots) as a function of the reverse momentum transfer  $s$  and presented offset for better visualisation. Overlaying the scattering profiles, are fits of the reconstructed averaged models for Mub proteins. (B+D) Pair distribution functions  $P(r)$  generated from the experimental scattering. Low resolution shape reconstructions of MubRV (E), MubRI (F), Mub-R5 (G). MubR8-V (H) and MubRV-VI (I) with manually docked high resolution structures of MubRV (blue), MubR5 (green).

Solution envelopes of single and tandem Mub repeats were reconstructed from  $P(r)$  functions using GASPOR (Figure 4.8 E-I), whereas DAMMIF was used for MubRI-II-III shape reconstructions (Figure 4.9 C), revealing an extended boomerang-like solution structure of Mub proteins (see 2.5.2.2). The experimental scattering profiles of MubR5 and MubRV were in good agreement with the computed solution scatterings curves from their crystal structures using CRY SOL, showing  $X$  values of 1.06 and 1.02 (see Appendix XI) [355]. The MubRV and -R5 crystal structures were docked manually into the reconstructed shapes and docking solutions were refined using SITUS (see Appendix IX) [354]. A good fit of low resolution solution envelopes and high resolution structures was observed for the single and tandem repeat proteins (Figure 4.8 E-I), while the docking of three MubRV structures into the MubRI-II-III shape reconstruction proved more challenging (Figure 4.9 C), likely due to the use of rigid structural models. Alternatively, it may indicate that the actual, maximal dimensions of MubRI-II-III in solution are higher than those determined from the collected SAXS data maybe due to sample properties or quality.

The maximal particle diameters for MubRV, -RI and -RV were found to be 110 Å, 106 Å and 105 Å (Figure 4.8 E-G), while the  $D_{\max}$  values for MubR8-V and MubRV-VI were calculated to be 205 Å and 206 Å (Figure 4.8 H-I). Finally, the triple repeat MubRI-II-III showed a maximal particle diameter of 292 Å indicating an elongated conformation of single and multiple Mub-repeat proteins in solution (Figure 4.9 C). In contrast, the  $D_{\max}$  for the Nterm protein was found to be 159 Å and hence significantly lower than that of MubRI-II-III, a protein of similar molecular weight, suggesting a different solution structure of the Nterm domain to the Mub repeats (Figure 4.9 D). The solution envelope of Nterm, reconstructed from its  $P(r)$  function using GASPOR, demonstrated a more globular protein shape with an elongated tail and provided the first structural information, albeit at low resolution, for the N-terminal domain of MUB (Figure 4.9 B and D). To date, no high resolution structural information is available on the Nterm protein of MUB or any homologous proteins.



**Figure 4.9 SAXS data analysis and shape reconstruction of MubRI-II-III and the N-terminal domain of MUB**

(A) Experimental scattering curves for the triple MubRI-II-III (light blue) and the N-terminal domain (red) are shown as the logarithm of the scattering intensity  $I$  (black dots) as a function of the reverse momentum transfer  $s$  and presented offset for better visualisation. Overlaying the scattering profiles, are fits of the reconstructed averaged models for MubRI-II-III and Nterm. (B) Pair distribution functions  $P(r)$  were generated from the experimental scattering. Low resolution shape reconstructions of (C) MubRI-II-III with manually docked high resolution structures of MubRV (blue) and (D) Nterm.

In summary, the results of these SAXS studies indicate the arrangement of Mub repeats as ‘beads on a string’ within the full-length MUB surface protein, with an alternatively-shaped Nterm domain at the protein tip.

## 4.4 Discussion

Adhesion to host tissues is a necessary first step of bacterial colonisation and is thought to be mediated by cell surface adhesion proteins. Considerable progress has been made in the last decade in the structural characterisation of pathogenic Gram-positive adhesins such as MSCRAMM or pili, revealing many interesting and unique features that may explain their adhesion mechanism to host cell receptors [384, 390-391]. In sharp contrast, there is currently a lack of structural information on Gram-positive adhesins, which are suggested to be involved in commensal or probiotic adhesion to mucus.

The cell surface adhesin MUB of *L. reuteri* ATCC 53608 is one of the few adhesins for which the contribution to the overall bacterial adhesion to mucus has been demonstrated [283, 302]. As commonly observed in Gram-positive adhesion molecules, MUB is a LPXTG-anchored cell wall protein with a modular domain organisation. It contains 14 tandemly arranged Mub repeat domains of two types, Mub 1 and Mub 2 (see 1.4.2). The crystal structure of the Mub type 2 repeat MubR5, resembling an elongated structure with two distinct domains, B1 and B2, previously reported in our lab, provided first structural insight into commensal Mub domains [282].

Besides MUB, Mub domain containing proteins have been identified in different lactobacilli species by *in silico* analysis, and are implicated in mucin binding (see 1.4.2) [304]. Additionally, the SpaCBA pili present in *L. rhamnosus* GG exerts mucus-binding ability, more specifically its SpaC subunit, potentially resulting in prolonged residency of *L. rhamnosus* GG in the GI tract [288]. However, their functional characterisation is fragmentary, no structural information is available, and the mechanism of adhesion and the specific ligands recognised remain to be determined. The current hypothesis is that glycan structures found in large numbers on mucin proteins are the preferential binding sites for mucus binding protein but this has yet to be proven.

The potential sugar recognition by MUB was investigated by X-ray crystallography through co-crystallisation and crystal soaking studies with the single Mub repeat MubR5 and various mono- and disaccharides (see 4.1). The only sugar molecule identified bound by MubR5 was Man, mediating the cross-linking of two adjacent MubR5 repeats in the inter domain region of the protein. This finding may suggest a role of carbohydrate-ligand binding by the cell surface MUB protein in bacterial cell

interaction and aggregation. Indeed, MUB-positive *L. reuteri* strains *L. reuteri* ATCC 55739 and 53608 demonstrated auto-aggregation properties, which have been associated with the presence of MUB on the bacterial cell surface [302]. In contrast, the mutant strain 1063N of *L. reuteri* ATCC 53608, expressing a truncated MUB protein, showed significantly less auto-aggregation. The aggregation capability is, besides specific mucosal receptor recognition, an important host colonisation factor of commensal lactic acid bacteria. For example, the aggregating strain *L. crispatus* M247 shows prolonged persistence in the GI tract of mice compared to an aggregation-deficient strain [392]. *L. acidophilus* M29 has been demonstrated to autoaggregate and adhere to intestinal epithelial cells, demonstrating a correlation between aggregation and adhesion [393]. The interaction between sugars and adhesins has recently been suggested to play a role in *Lactobacillus* aggregation [394]. The role of Man as a potential sugar ligand and mediator of bacterial auto-aggregation and the characterisation of its interaction with Mub repeats, however, needs to be further investigated. Isothermal titration calorimetry (IT) experiments of MubR5 with Man sugars were performed to determine their binding affinity (see 5.2.1).

In order to gain additional functional insight into the structure and function of Mub repeats, the crystal structure of the Mub type 1 repeat MubRV was determined at 2.6 Å. MubRV shows high structural similarity to MubR5 with the same overall organisation of two domains, B1 and B2, and an elongated shape 110 Å in length (see 4.2.2). Additionally, MubRV shows structural similarity to a number of pilins and MSCRAMM. Those structures include the CnaB structures of *Strep. aureus*, the N2 domain of GBS52 from *Strep. agalactiae*, that contains the pulmonary binding site, the C-terminal D4-domain of RrgB of *Strep. pneumoniae*, the major Spy0128 pilin of *Strep. aureus* and the SpaA shaft pilin of *C. diphtheriae* [261, 382-383, 395]. Hence, MubRV shows structural similarity to members of the 4 major groups of invasive Gram-positive pathogens: *C. diphtheriae*, group A *Streptococcus* (that is *Strep. pyogenes*), group B *Streptococcus* (that is *Strep. agalactiae*) and *Strep. pneumoniae*.

Pili, long protein filaments, which are composed of major and minor pilin subunits, and MSCRAMM are Gram-positive cell-surface adhesins of pathogenic bacteria. They share a similar modular protein organisation and common structural motifs including an N-terminal Sec-dependent secretion signal for protein transport to the cell wall,

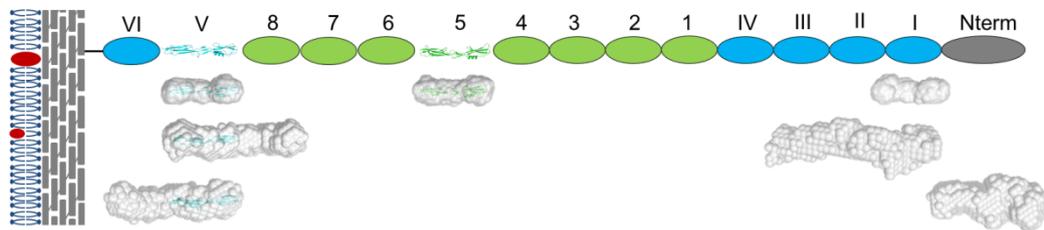
a variable number of modular domains, and a C-terminal cell wall anchor including a canonical LPXTG-motif for sortase-dependent covalent attachment to the cell wall component peptidoglycan. Individual, modular protein repeats in MSCRAMM and minor or ancillary pilins are functional domains that play an important role in host adhesion and biofilm formation and are thus key elements of bacterial pathogenicity [263, 309, 396]. While the interaction of MSCRAMM with extracellular matrix components, such as fibronectin, fibrinogen and collagen is well described, less is known about the specific target receptors and the binding mode of pilins [226, 385, 397-400].

To date, crystal structures are available for a number of MSCRAMM domains or individual pilins revealing a common structural module based on the IgG-constant (IgG-C) domain, designated as an Ig-like fold. The IgG-fold is a  $\beta$ -sandwich characterized by a four-stranded  $\beta$ -sheet (ABED) and a three-stranded  $\beta$ -sheet (CFG), which is different from other Ig-superfamily folds, i.e. V, H and I [320]. Two variants of the Ig-like fold, based on IgG-C have been identified, IgG-rev and DEv-IgG. While the latter shows the same topology as the IgG-C but possess additionally strands, the IgG-rev is characterized by a reverse strand arrangement in two alternative topologies CBE DAGF and CBEF DAG. The IgG-rev fold was first observed in the MSCRAMM Cna B-region of *Staph. aureus*, while the DEv-IgG fold was found in the Cna A-region in the same molecule [379, 400-401]. MubRV shows structural homology to pili and MSCRAMMs, whose folds are particularly similar to the CnaB Ig-like fold or IgG-rev fold.

However, despite the high similarity of the B2 domain of MubRV to these Ig-like fold domains, the  $\beta$ -sandwich in the B2 domain only contains a 4-stranded  $\beta$ -sheet (BAFD) and two  $\beta$ -strands (EC) of different topology. MubRV possesses three additional  $\beta$ -strands forming a  $\beta$ -sheet in the inter domain region between the B1 domain and the C-terminal part of the B2 domain (see 4.2.2).

AUC experiments of full-length native MUB suggested its appearance as an elongated protein in solution (see 3.3.2). SAXS experiment were conducted to investigate the properties of single Mub repeats in solution and to obtain information on shape and organisation of multiple Mub repeat proteins (see 4.3). Single Mub repeats describe boomerang-like solution envelopes 105 to 110 Å (~10 nm) in length, which showed a good fit with the elongated X-ray structures of the Mub type 1 MubRV and the type 2

MubR5 structures 110 Å in length. In addition, the low resolution structure reconstruction of double and triple Mub repeats of about 200 and 300 Å (~20 and 30 nm) in length demonstrated an elongated protein shape of multiple Mub repeats. Additionally, the first structural information on the N-terminal domain of MUB was obtained, which demonstrated a less extended, more globular 3D reconstruction compared to Mub repeats. These findings thus suggest a potential ‘beads on a string’ arrangement of 14 Mub repeats together with a less extended N-terminal domain (see Figure 4.10).



**Figure 4.10 Schematic representation of MUB on the bacterial cell surface**

Mub type 1 and type 2 repeats are coloured blue and green, respectively, as are MubRV and MubR5 crystal structures. SAXS shape reconstructions are shown next to corresponding Mub repeats and MubRV crystal structures is fitted into envelopes of MubRV, -R8-V and -RV-VI.

Filamentous pili structures are fairly abundant and better studied in pathogenic Gram positive bacteria rather than commensals, and can reach a length of 70-200 nm or 0.3-3 µm depending on pili type [402]. For example, the serine-rich pili of *Strep. parasanguinis* has been visualised by electron microscopy (EM) with an estimated length of several 100 nm [391]. Commensal pili genes have to date only been reported in *L. johnsonii* NCC533, *L. lactis* TIL448, two *L. ruminis* strains and the *L. rhamnosus* strains GG and LC705 (see 1.4.2) [288, 291, 307-308]. The SpaCBA pili of *L. rhamnosus* GG was the first to be visualised on commensal bacterial cells as a long extended fiber [288].

A similar protein shape as shown for Mub repeats was observed for Gram-positive pathogenic pilin subunits. A pilin subunit of a serine rich fimbriae of *Strep. pneumoniae* demonstrated a kidney-shaped SAXS solution structure of about 12 nm in length [391]. Another example is the Spy0128 backbone pili of *Strep. pyogenes*, that is, such as

a single Mub repeat, composed of two domains and shows both a crystal and a SAXS solution structure of about 10 nm in length [381, 403]. In addition, the native pili of *Strep. pneumoniae* describes a rod-like structure as observed in its EM density map fitting twice the crystal structure of a RrgB pili subunit [404]. These studies provide the first low and high resolution structure information on a commensal mucus-binding adhesin, indicating high structural similarity with pathogenic fimbrial proteins, such as pilins and MSCRAMM, involved in adhesion and infection.

## **CHAPTER 5    FUNCTIONAL CHARACTERISATION OF MUB ADHESION PROPERTIES TO MUCUS**

The cell-surface protein MUB of the pig isolate strain *L. reuteri* ATCC 53608 is a modular adhesion molecule comprised of 14 Mub repeats, of type 1 and type 2, and an N-terminal domain (Nterm). The role of MUB as an effector molecule in the adhesion of bacterial cells to mucus has been demonstrated in our lab [302]. In addition, studies by Roos and Johnsson showed the interaction of *L. reuteri* ATCC 53608 cells as well as recombinant Mub-repeat proteins, fused to a maltose binding protein, to porcine and hen mucus, and to porcine mucin [283]. However, the distinct mucus or mucin components recognised by MUB proteins and the biochemical mechanism of interaction remain unknown. The current working hypothesis is that cell-surface adhesins mediate bacterial binding to mucosal surfaces by interaction with mucin glycans. In order to investigate and dissect the binding ability of MUB proteins to different types of mucin, and investigate the potential recognition of mucin glycans, the adhesion properties of Mub repeats, MubR5, -RV, -RI, -R8-V, -RV-VI, -RI-II-III and the Nterm domain proteins, Nterm and NtermMubRI, as well as the native full-length MUB were tested in a variety of different mucin and glycan binding studies.

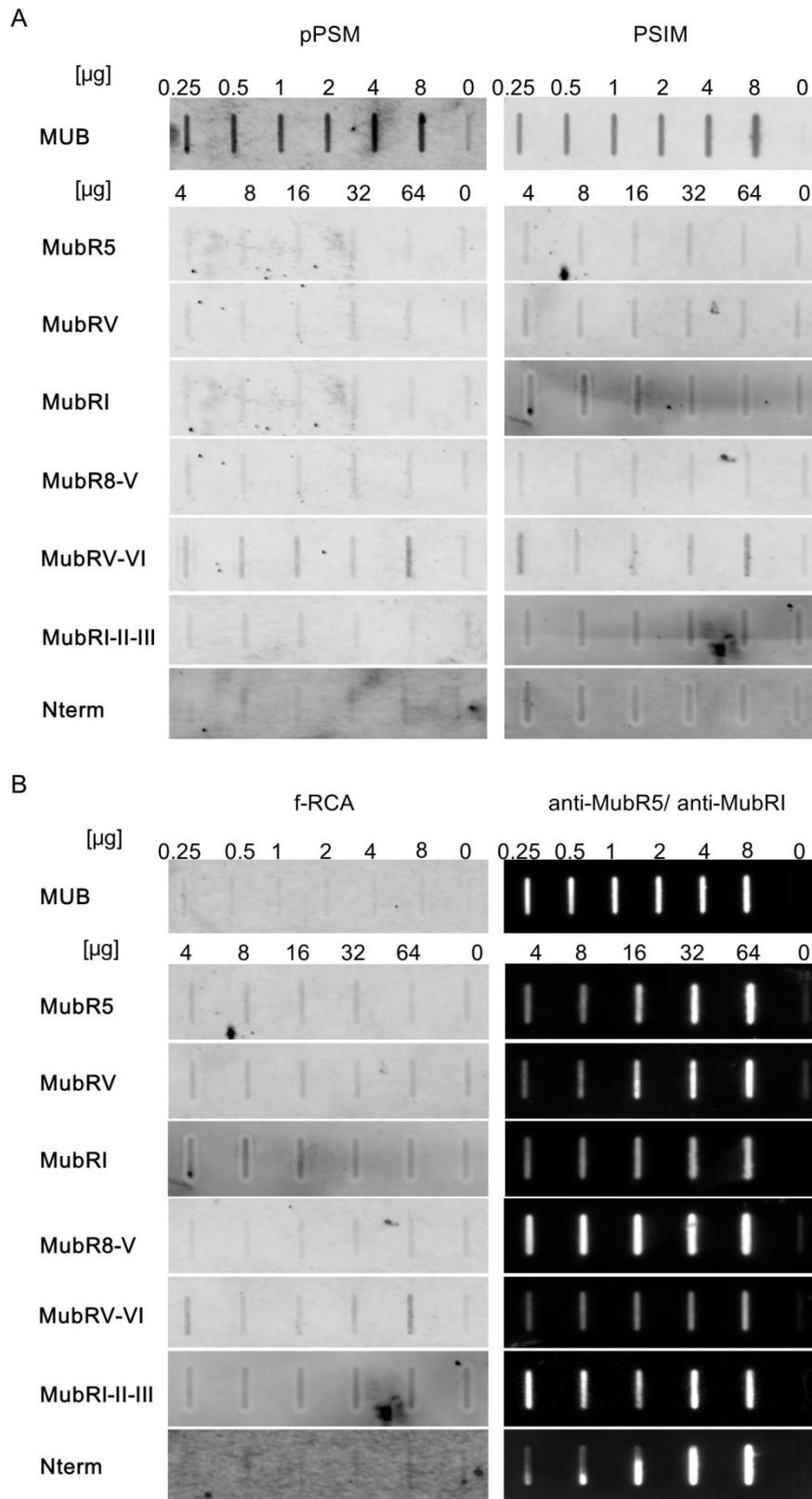
### **5.1    Binding of MUB proteins to mucus and mucin**

#### **5.1.1    MUB protein binding to porcine and human mucin in membrane-adhesion assays**

The binding of the recombinant Mub repeats MubR5, -RV, -RI, -R8-V, -RV-VI and -RI-II-III, the Nterm protein, as well as native MUB, to isolated pig small intestinal mucus (PSIM) and purified commercial porcine gastric mucin (pPGM) was first investigated in a slot-blot binding assay (see 2.1.2 and 2.3.6.4). Briefly, recombinant Mub-repeat proteins and the Nterm protein were purified by IEC and IMAC followed by SEC (see 3.1), and native MUB was purified from a *L. reuteri* ATCC 53608 cell culture via SEC (see 3.3.1). The recombinant MUB proteins (0-64 µg) and native MUB (0-8 µg) were immobilised onto nitrocellulose membranes and probed with PSIM and pPGM in excess. Bound mucin was detected via FITC labelled RCA (f-RCA), which specifically interacts with Gal residues (see 2.3.5), a frequently encountered sugar molecule in

mucin *O*-glycans (see 1.2). The reactivity of PSIM and pPGM with f-RCA was separately verified in a test assay (data not shown).

Native MUB, comprising all 14 Mub repeats and the N-terminal domain, binds to PSIM and pPGM when compared to the PBS negative control (Figure 5.1 A). In contrast, no binding was observed for any of the single, tandem and triple Mub-repeat proteins or the Nterm protein under the current assay conditions, maybe indicating the requirement of several Mub repeats for mucin binding. The presence of immobilised adhesion proteins was verified by detection via specific anti-MubRI and anti-MubR5 for MUB and Mub repeats and anti-His<sub>5</sub> for Nterm (Figure 5.1 B). Recombinant MUB proteins as well as the native MUB did not show reactivity with f-RCA, suggesting specificity of the observed interaction of MUB with mucin samples (Figure 5.1 B).

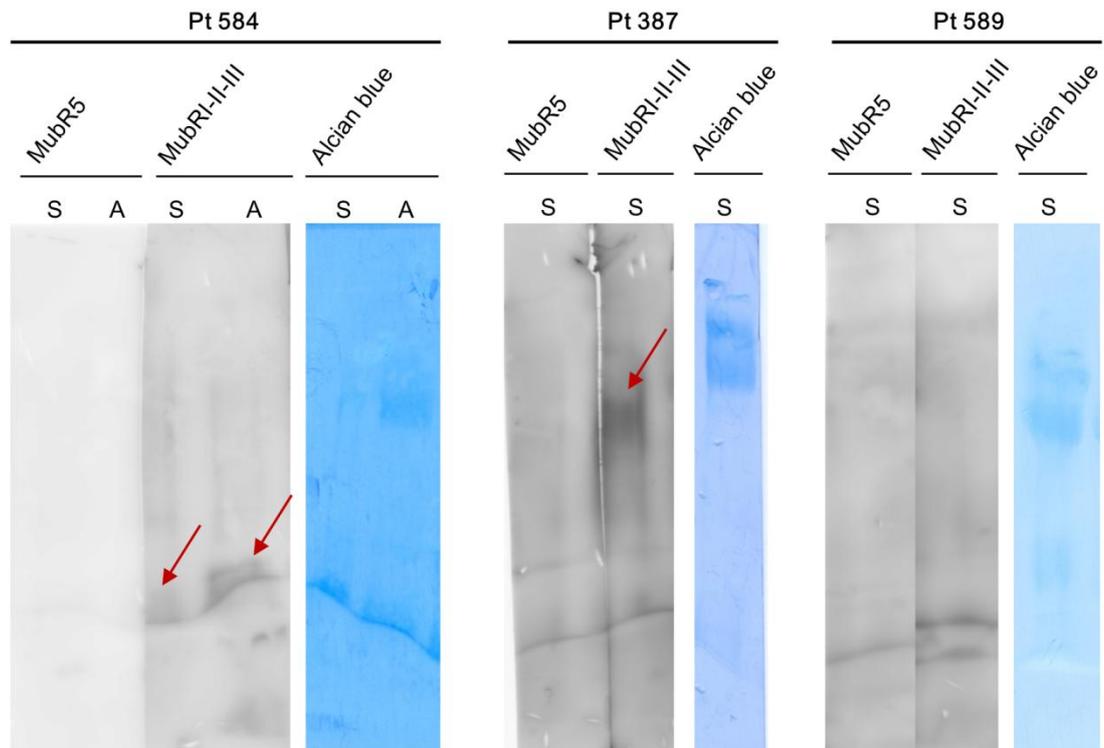


**Figure 5.1 Binding of MUB proteins to pPSM and PSIM in a slot-blot assay**

(A) Slot-blot of recombinant MUB proteins MubR5, -RV, -RI, -R8-V, -RV-VI, -RI-II-III and Nterm, and the native full-length MUB probed with pPSM and PSIM. Bound mucin was detected by f-RCA. (B) Control slot-blot of MUB proteins detected with f-RCA or primary anti-MubR5, anti-MubRI or anti-His<sub>5</sub> followed by secondary anti-rabbit or anti-mouse-HRP.

In addition, the binding of the selected recombinant Mub repeats, MubRI-II-III and MubR5, and native MUB to MUC2 isolated from human biopsy samples of the sigmoid and ascending colon was investigated after agarose polyacrylamide gel electrophoresis (AgPAGE) in collaboration with Gunnar Hansson (and Jessica Holmen-Larsson) at the University of Gothenburg (Sweden). Reduced crude MUC2 samples from three different patients were separated by AgPAGE, blotted onto a PVDF membrane and stained by alcian blue, a cationic dye reacting with acidic mucosal polysaccharides and glycosaminoglycans (see 2.3.4.2 and 2.3.4.4) [405]. Membranes were incubated with MubR5 (50-70  $\mu\text{g/mL}$ ), MubRI-II-III (50-70  $\mu\text{g/mL}$ ) and MUB (10  $\mu\text{g/mL}$ ) and bound proteins detected via primary anti-MubRI and anti-MubR5, and secondary anti-rabbit-HRP (see 2.3.5).

MubRI-II-III showed binding to MUC2 sample components of low and high MW, which stained positive with alcian blue, from the ascending and sigmoid colon of two (Pt 584 and Pt 387) out of three tested patient samples (Figure 5.2), which may indicate a differential glycosylation of MUC2 in patient Pt 589. It is known that glycosylation profiles differ between individuals, however the relative variability is considered to be low (see 1.2.2) [113]. Observed differences in electrophoretic mobility and MW of MUC2 material may be due to differences in sample composition, degree of glycosylation and possibly degradation. The single repeat protein MubR5 did not show binding to MUC2 isolated from any of the patients. Interestingly, Coic and co-workers reported the colocalisation of the synthetic MUB<sub>70</sub> peptide, which comprises the first 70 residues of MubR5, with MUC2 on human colonic tissue samples and suggested its potential use as a human colonic mucus marker [303]. These observations indicate differences between assay types and stress the importance of the careful investigation of MUB protein binding to mucins, e.g. using different assay set ups and sufficient binding controls. Binding experiments with MUB were not conclusive due to ambiguous signals after antibody detection (data not shown).



**Figure 5.2 Binding of MubR5 and MubRI-II-III to colonic MUC2 from human biopsy samples**  
 Western-blot of MUC2 isolated from human biopsy samples of the ascending (A) and/or sigmoid (S) colon of three different patients (Pt 584, 387, 589) after AgPAGE probed with MubR5 and MubRI-II-III. Bound protein detected via primary anti-MubR5 and anti-MubRI, and secondary anti-rabbit-HRP and Western-blot stained with alcian blue.

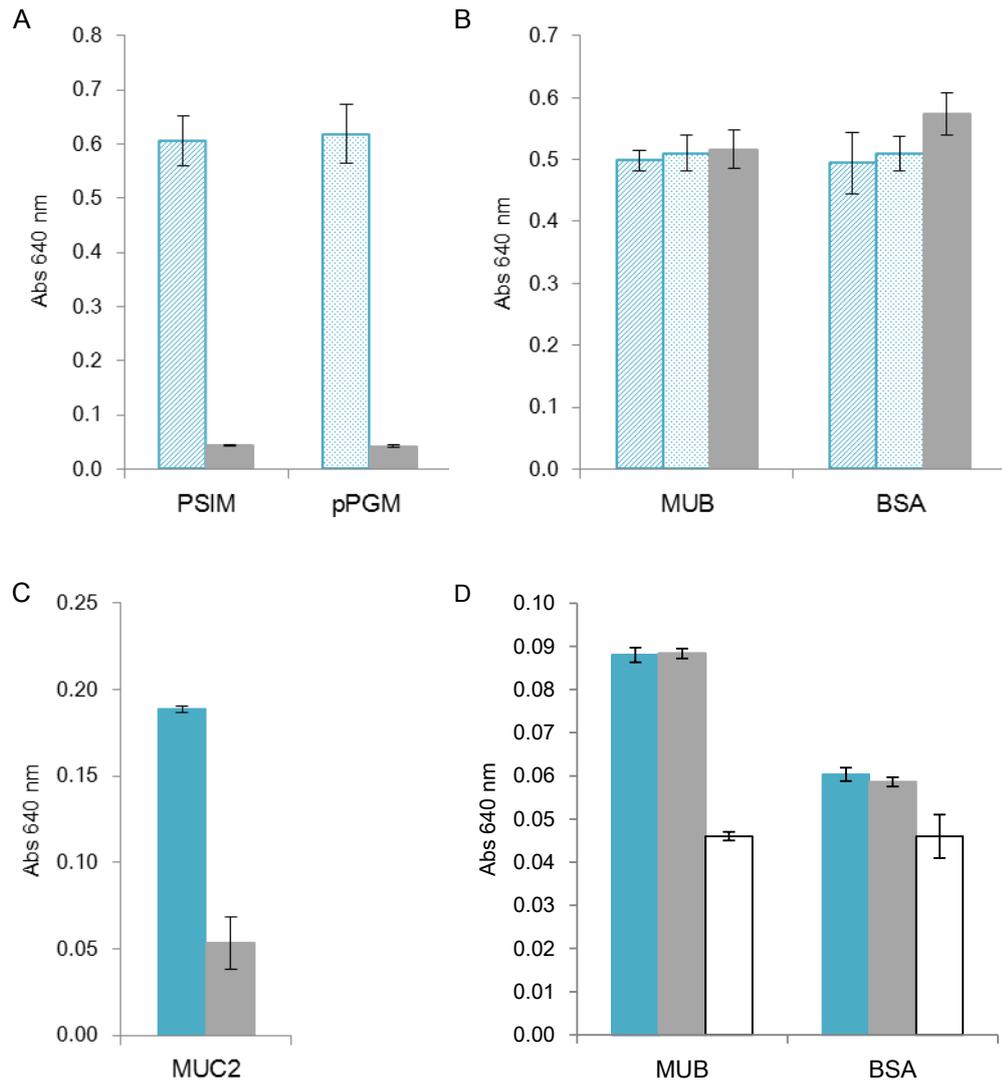
The observation that the triple repeat MubRI-II-III but not the single repeat MubR5 showed binding to colonic MUC2 material may again suggest the presence of several Mub repeats to be necessary for mucin binding. In addition, the binding of MubRI-II-III to MUC2 after AgPAGE but not to PSIM or pPGM after slot-blotting may indicate a potential role of mucin type and presentation for ligand recognition (as described above).

### 5.1.2 Assessment of MUB and MUB repeat binding to mucin glycans

The binding of native MUB and the recombinant Mub repeats, MubR5 and MubRI-II-III, to human and porcine mucin was further investigated using two different microtitre plate assay set-ups modified from those described in literature for adhesin-mucin interaction or previously used in our lab for bacteria-mucus adhesion [249, 283, 302].

Firstly, purified native MUB (0.6 µg) and BSA (0.6 µg) as a control were coated on microtitre plates and then incubated with PSIM (1.6 µg), pPGM (0.4 ng) and purified human MUC2 (0.6 µg) kindly provided by Michael McGuckin (Mater Medical Research Institute, South Brisbane, Australia) (see 2.1.2). Bound mucin was detected via primary anti-Muc2.3 for PSIM and pPGM or anti-MUC2C3 for human MUC2, and secondary anti-rabbit-HRP (see 2.3.5). The binding epitope of anti-MUC2C3 is a 16 amino acid epitope at the C-terminus of human MUC2 (see 2.1.3). Anti-Muc2.3 is specific for a 14 amino acid peptide of murine Muc2, but also reacted with PSIM and pPGM (see 2.1.3). The reactivity and sufficient signal intensity of PSIM, pPGM and human MUC2 detection with the used anti-Muc2.3 and anti-MUC2C3 antibodies was confirmed for mucins immobilised onto microtitre plates (Figure 5.3 A and C). The presence of coated MUB at saturated concentration was confirmed by incubation with primary anti-MubRI and anti-MubR5 (data not shown).

MUB did not show increased binding to PSIM and pPGM when compared to the BSA control (Figure 5.3 B), whereas higher signals were detected against MUC2 (Figure 5.3 D). However, comparable signals were observed in control experiments, where primary and secondary antibodies were incubated on MUB and BSA in the absence of mucin ligands (Figure 5.3 B and D). The secondary anti-rabbit-HRP alone showed reduced signal intensities (Figure 5.3 D), suggesting that the unspecific interaction is mainly due to the reaction of the primary antibodies with the coated proteins.



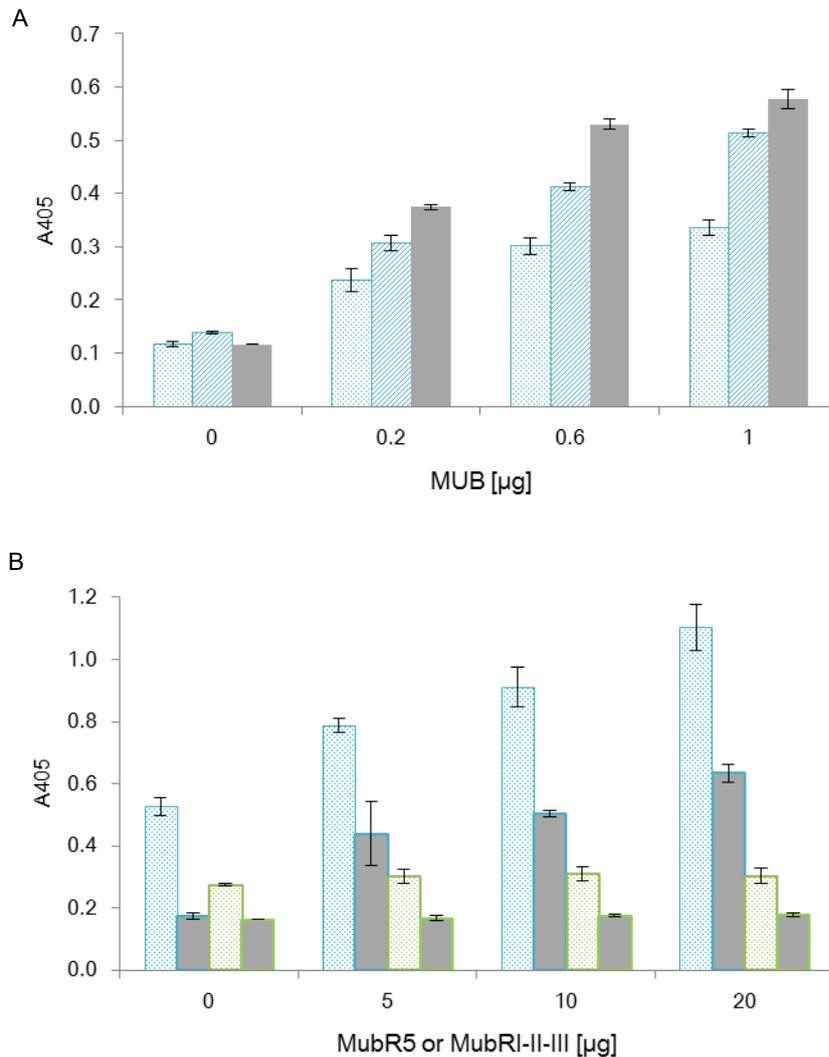
**Figure 5.3 Binding of MUB to porcine and human mucin**

Control assays with (A) immobilised PSIM (1.6  $\mu$ g) (blue stripes) and pPGM (0.4 ng) (blue dots) detected via primary anti-Muc2.3 or (C) human MUC2 (0.6  $\mu$ g) (blue) detected via primary anti-MUC2C3 followed by secondary anti-rabbit-HRP. (A+C) Anti-rabbit-HRP added to mucin coated wells as a secondary antibody control in the absence of primary antibodies (grey). (B) PSIM (blue stripes) and pPGM (blue dots) incubated on MUB and BSA with bound mucin detected by anti-Muc2.3 and anti-rabbit-HRP. (D) Human MUC2 incubated on MUB and BSA with bound mucin detected by anti-MUC2C3 and anti-rabbit-HRP. (B+D) MUB and BSA incubated with primary anti-Muc2.3 or anti-MUC2C3 followed by secondary anti-rabbit-HRP (grey), or anti-rabbit-HRP alone (white) as antibody controls. Error bars represented standard deviation from mean calculated from triplicate.

Secondly, PSIM, pPGM and BSA as a control were coated onto microtitre plates in excess and probed with increasing amounts of native MUB (0-1  $\mu\text{g}$ ) or recombinant MubR5 (0-20  $\mu\text{g}$ ) and MubRI-II-III (0-20  $\mu\text{g}$ ) (see 2.3.6.5). Bound MUB proteins were detected via protein specific primary anti-MubR5 and anti-MubRI, and secondary anti-rabbit-AP (see 2.1.3). The presence of PSIM and pPGM was verified by incubation with biotinylated WGA (b-WGA) and ExtrAvidin-peroxidase. The presence of coated mucins was confirmed by detection with b-WGA and ExtrAvidin-peroxidase (data not shown).

MUB showed concentration-dependent binding to PSIM and pPGM, as observed in slot-blot assays (see 5.1.1) (Figure 5.4 A). However, signals for MUB probed on BSA as a control showed slightly higher signal intensities than on coated mucins. MubR5 and MubRI-II-III also showed interaction with PSIM and pPGM in a concentration dependent-manner, which was increased compared to the BSA control (Figure 5.4 B). Binding of the triple domain MubRI-II-III was higher at all tested concentrations compared to the single repeat MubR5, again indicating a correlation between the number of Mub repeats and binding capability (see 5.1.1).

The high signals for all tested MUB proteins when probed on BSA may indicate unspecific interaction of the adhesins and the need to investigate alternative controls. Studies assessing the binding of lectins to coated mucin glycoproteins performed in a microtitre plate assay, suggested that BSA used as a blocking agent may interfere with binding when studying carbohydrate-protein interactions [406]. A protein free blocking solution was thus alternatively applied, which was also used for the above described binding experiments of MUB proteins to mucins (see 2.1.1).

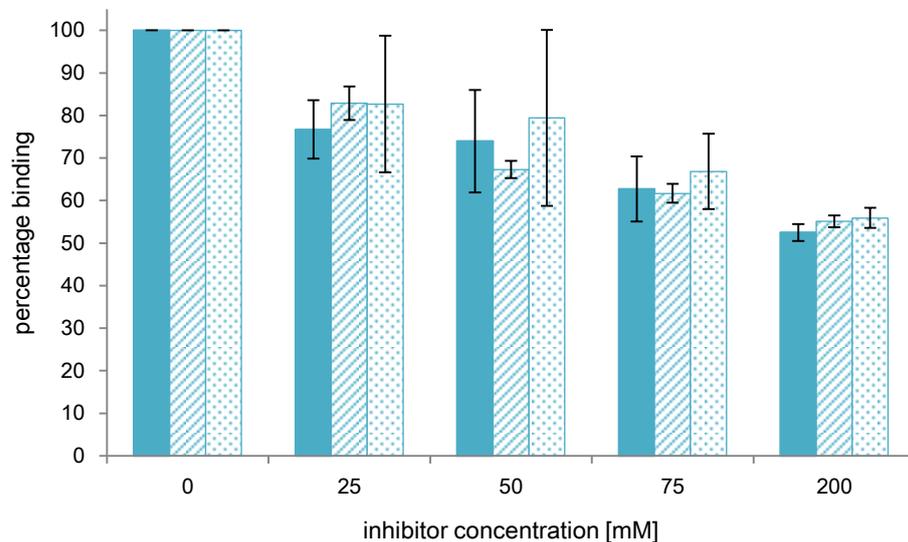


**Figure 5.4 Binding of MUB, MubRI-II-III and MubR5 to porcine mucin**

(A) MUB (0-1  $\mu\text{g}$ ) incubated on pPSM (blue dots) and PSIM (blue stripes), and (B) MubRI-II-III (blue dots) and MubR5 (green dots) incubated on pPSM (B) as well as BSA as a control (grey, grey-blue line, grey-green line). Bound proteins detected via primary anti-MubR5 and anti-MubRI, and secondary anti-rabbit-HRP. Error bars represent standard deviation of the mean calculated from triplicate.

In order to further investigate the nature of the interaction of full-length MUB with pPGM, soluble sugars, 6'sialyllactose (6'SL, Neu5Ac $\alpha$ 2-6Gal $\beta$ 1-4Glc), Neu5Ac and lactose (Lac, Gal $\beta$ 1-4Glc), were used as inhibitors. Sialylation is commonly found in GI mucin *O*-glycans and has been identified in PGM *O*- and *N*-glycans together with high amounts of Gal and GlcNAc, the building blocks of Lac (see 1.1.2) [407]. Inhibition studies were performed in line to the above described microtitre plate assay, but after incubation of MUB (0.6  $\mu\text{g}$ ) on pPGM, binding was competed with sugars (0-200 mM), and bound MUB was detected via protein specific anti-MubRI and anti-MubR5 followed

by anti-rabbit-HRP (see 2.3.6.5). Data are presented as a percentage of binding relative to MUB binding in the absence of soluble sugar inhibitor.



**Figure 5.5 Competition of MUB binding to pPGM with soluble sugars**

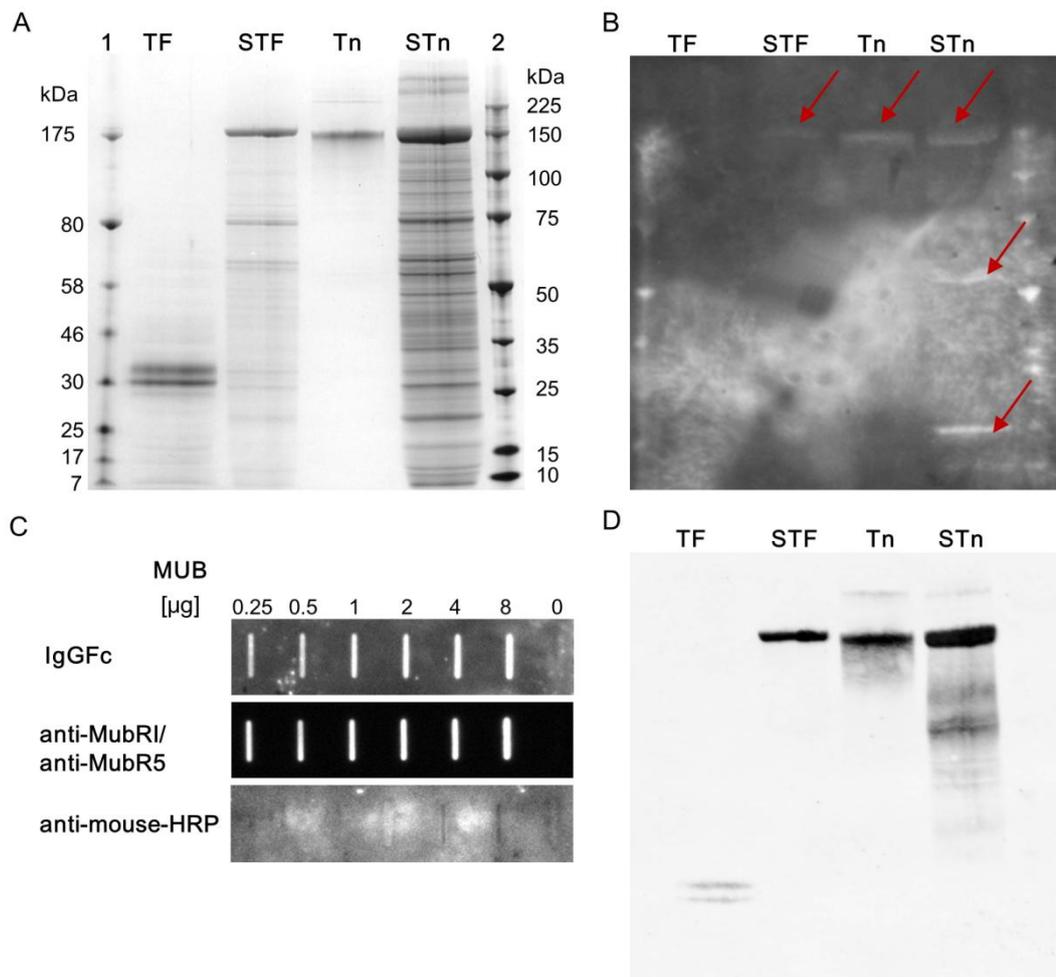
Competition of MUB (0.6  $\mu$ g) binding to pPGM using increasing concentrations of 6'sialyllactose (6'SL) (blue), Neu5Ac (blue dots) and lactose (Lac) (blue stripes). Bound MUB was detected via primary anti-MubR5 and anti-MubRI, and secondary anti-rabbit-HRP. Absorbance measured at 405 nm and absorbance values presented as percentage binding relative to signal intensity at zero inhibitor concentration. Error bars represent standard deviation from the mean of a triplicate.

MUB binding to pPGM was reduced by increasing concentrations of 6'SL, Neu5Ac and Lac up to 200 mM to about 55-60% compared to MUB adhesion in the absence of sugars (Figure 5.5). These data suggest the involvement of different sugar molecules in the binding of MUB to mucin, which offers numerous diverse glycan bindings sites (see 1.2).

The potential role of sialic acid in the binding of mucin by MUB was further investigated using MUC1-glycopeptides carrying the specific non-sialylated and sialylated glycan epitopes: TF (Gal $\beta$ 1-3GalNAc $\alpha$ ), STF (Sia $\alpha$ 2-3Gal $\beta$ 1-3GalNAc $\alpha$ ), Tn (GalNAc $\alpha$ ) and STn (Sia $\alpha$ 2-6GalNAc $\alpha$ ), which were kindly provided by Gunnar Hansson (University of

Gothenburg, Sweden) (see 2.1.2). Briefly, MUC1-glycopeptides were separated via SDS-PAGE and Western-blotted onto a nitrocellulose membrane, before incubation with native MUB (10 µg/mL) (see 2.3.6.1). Bound MUB was detected via anti-MubR5 and anti-MubRI followed by anti-mouse-HRP, and the presence of MUC1-glycopeptides was verified using f-RCA and f-WGA. These lectins commonly recognise Gal and Neu5Ac residues (see 3.3.2, Table 3.3) and show additional reactivity with GalNAc [408]. Since the MUC1-glycopeptides are produced as fusion proteins with a mouse IgG Fc fragment (IgGFc), the interaction of MUB to mouse IgGFc was examined in a control slot-blot assay as described before, and IgGFc detected via anti-mouse-HRP (see 5.1.1).

SDS-PAGE analysis of the MUC1-glycopeptides showed the presence of predominant protein bands for the peptides carrying the STF, Tn and STn antigens at about 150 to 175 kDa, which is in line with earlier observations and in agreement with their theoretical MW (Figure 5.6 A) [327-328]. For MUC1-STF and -STn additional bands at lower MW were observed, which may be caused by protein contamination or degradation. Two strong bands were observed at about 30 kDa for the MUC1-TF peptide, suggesting its degradation in multiple fragments. Our collaborators at the University of Gothenburg demonstrated the reactivity of all visible bands for the 4 tested MUC1-glycoproteins by anti-MUC1 detection after Western-blotting confirming their integrity (data not shown). Additionally, all MUC1-glycopeptides samples showed reactivity with f-RCA and f-WGA (Figure 5.6 D).



**Figure 5.6 Binding of MUB to MUC1 glycopeptides**

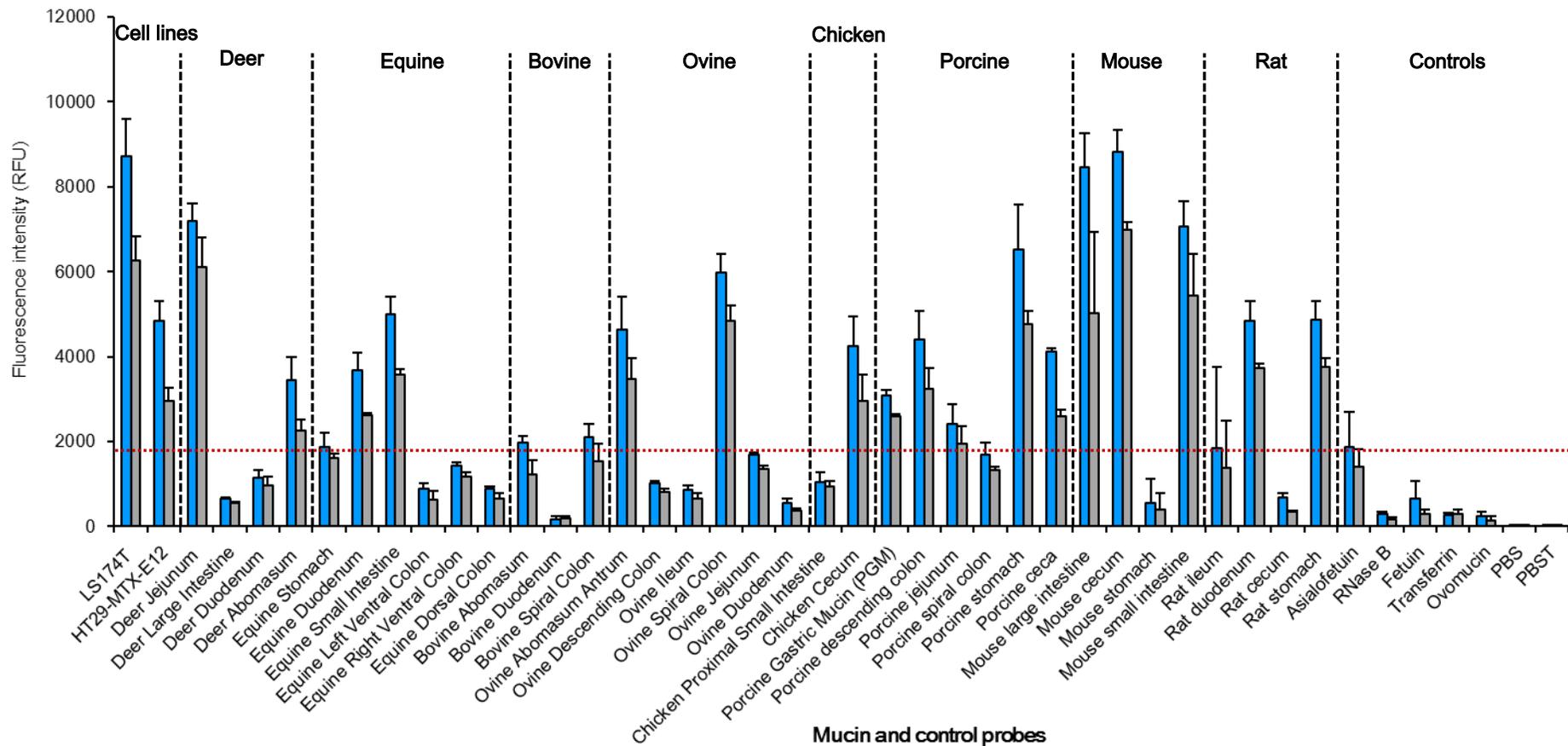
(A) Coomassie-stained SDS-PAGE gel of MUC1-IgGFc-glycopeptides with protein standards (1 and 2). (B) Western-blotted MUC1-IgGFc-glycopeptides probed with MUB, and bound MUB detected via primary anti-R5 and anti-RI, and anti-rabbit-HRP. (C) Control slot-blot assay of MUB probed with mouse IgGFc detected via anti-mouse-HRP. Slot-blotted MUB detected via anti-MubRI and anti-MubRI followed by anti-mouse-HRP, or incubated with anti-mouse-HRP alone as a control. (D) Slot-blot of MUC1-IgGFc-glycopeptides probed with f-RCA and f-WGA.

When probed against the MUC1-TF, -STF, -Tn, and -STn peptides after Western-blotting, MUB showed binding to MUC1-Tn and the high MW species of MUC1-STF and -STn, as well as lower MW fragments of MUC1-STn, while no signals could be observed for MUC1-TF (Figure 5.6 B). However, MUB also showed interaction with mouse IgGFc as demonstrated in a control slot-blot assay, which may contribute to the observed binding of MUB to the MUC1-IgGFc-glycopeptides (Figure 5.6 C). The binding of recombinant Mub-repeat proteins to Igs, including human IgG and IgGFab but not human IgGFc, has been demonstrated in our lab before [282].

### 5.1.3 MUB binding studies using mucin and neoglycoconjugate arrays

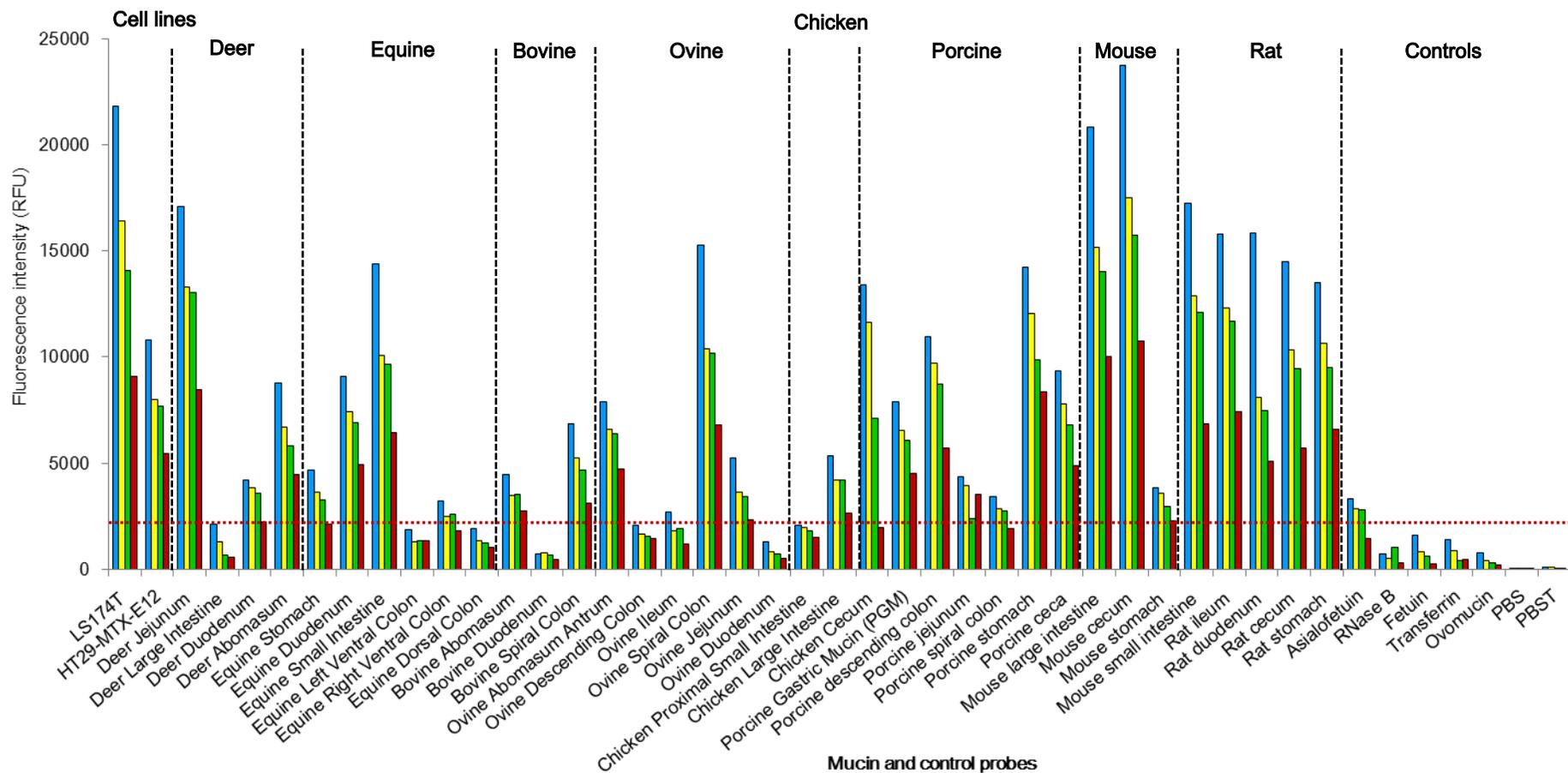
In order to screen a number of GI mucins for binding by native MUB, mucin microarray experiments were performed in collaboration with Lokesh Joshi (and Michelle Kilcoyne) at the University of Ireland (Galway, Ireland). The arrays contained mucins isolated from the GI tracts of different animal species (deer, equine, bovine, ovine, chicken, porcine, mouse and rat) and from two mucus-producing intestinal epithelial cell lines (LS174T and HT29-MTX-E12) as well as several binding controls all printed in replicas of 6. Covalent conjugation of the mucin probes to the microarray polymer surface was achieved via accessible amino groups in the terminal regions of the protein backbone. Mucin array printing and binding experiments were performed as described by Kilcoyne and co-workers [409]. Briefly, MUB was probed on the array in PBST (0.01% Tween-20) at 23°C for one hour, followed by detection via primary anti-MubRI and secondary anti-rabbit-AlexaFluor555 with wash steps in between using PBST (0.005% Tween-20).

Under these conditions, MUB (17 µg/mL) was shown to bind a number of mucins from different species including deer, equine, ovine, chicken, porcine, rat and predominantly mouse, with no obvious preference for mucins isolated from various parts of the GI tract (Figure 5.7). Low binding of MUB to commercial PGM was observed in line with the microtitre plate and slot-blot assays described earlier (see 5.1.1 and 5.1.2). Additionally, MUB bound to mucin originated from mucus-producing intestinal cell lines, especially LS174T over HT29-MTX-E12 (Figure 5.7). No binding signals were observed for the control proteins, asialofetuin, fetuin, RNase B, transferrin and the egg white ovomucin, or in the absence of printed probes. However signals were also observed for primary and secondary antibodies in the absence of MUB, which suggests some level of unspecific interaction of the antibodies used for adhesin detection with the printed probes. However, MUB (5-20 µg/mL) showed concentration-dependent binding to different mucins in control arrays, where constant primary and secondary antibody concentrations were used indicating binding specificity of MUB to mucin probes (Figure 5.8). Nevertheless, further optimisation of the mucin array experiments are desirable to reduce unspecific antibody interaction, for example by using fluorescently labelled MUB protein.



**Figure 5.7 Mucin binding of MUB assessed using a GI mucin array**

MUB (17  $\mu\text{g/mL}$ ) probed on a GI mucin array containing mucin probes from intestinal epithelial cell lines and different sections of deer, equine, bovine, ovine, chicken, porcine, mouse and rat gut as well as binding controls. Bound MUB detected by primary anti-MubRI (1:2,000) and secondary anti-rabbit-AlexaFluor555 (1:1,000) (blue). Control array probed with anti-MubRI (1:2,000) and anti-rabbit-AlexaFluor555 (1:1,000) in the absence of MUB (grey). Data presented as average of three independent experiments with 6 probe replicas per array with error shown as standard deviation of the mean. Signals above threshold of about 1,700 RFU represent binding to probes (red line).



**Figure 5.8 Concentration-dependent binding of MUB to GI mucins**

MUB probed at different concentrations 18 µg/mL (blue), 15 µg/mL (yellow), 10 µg/mL (green) and 5 µg/mL (red)) on a GI mucin array containing mucins from intestinal epithelial cell lines and different animal species as well as binding controls. Bound MUB detected via primary anti-MubR1 (1:1,000) and secondary anti-rabbit-Alexa Fluor555 (1:1,000). Signals above threshold of about 2,000 RFU represent binding to probes (red line).

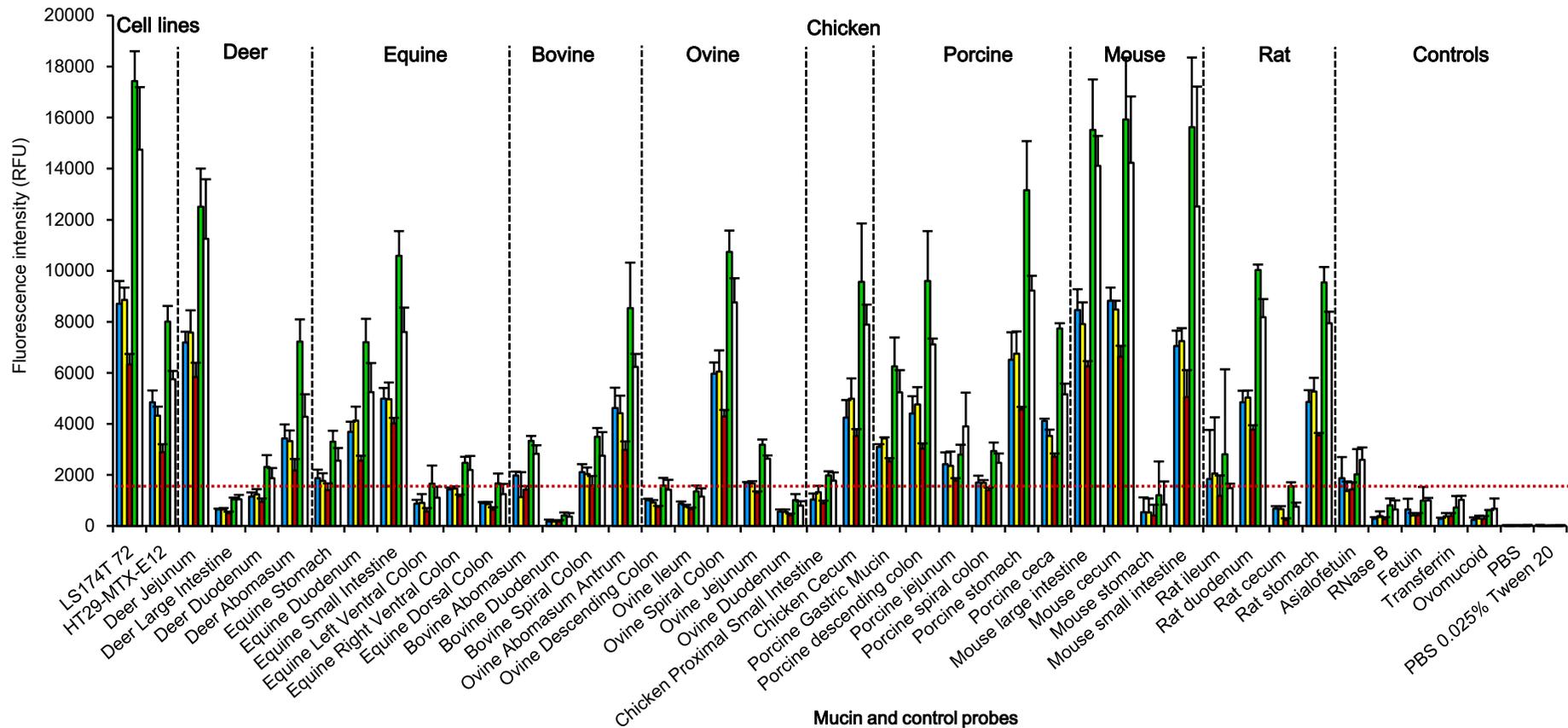
In parallel, a glycoconjugate array experiment was performed using the same array platform as for the GI mucin arrays, in order to identify glycan epitopes that potentially mediate MUB binding to mucins and thus can be tested as sugar inhibitors in the mucin array [410]. The glycoconjugate array contained different glycan structures including blood group epitopes conjugated with bovine or human serum albumin (BSA or HSA), which were immobilised via protein amino groups as for mucin arrays (described above). Bound MUB was detected via primary anti-MubRI and secondary anti-rabbit-AlexaFluor555.

MUB (10 µg/mL, 15 µg/mL, 18 µg/mL) showed concentration-dependent binding to 6 glycan structures with highest signals for GlcNAc and fucosylated LacNAc (Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc-), while no binding was observed for LacNAc alone (Table 5.1, see Appendix XII). All bound glycoconjugates contained Fuc, Gal or GlcNAc residues and nearly all of them were composed of LacNAc type 1 (Gal $\beta$ 1-3GlcNAc) or type 2 (Gal $\beta$ 1-4GlcNAc) and Lac (Gal $\beta$ 1-4Glc) disaccharides, which can be found in mucin glycan structures (see 1.2).

Name	Structure
GlcNAc-	GlcNAc
H2-	Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc
LNFP II-	Fuc $\alpha$ 1-3Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc-
LNFP III-	Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc-
3SuLe <sup>x</sup>	(SO <sub>4</sub> )Gal $\alpha$ 1-3[Fuc $\alpha$ 1-3]GlcNAc-
GlobT	Gal $\alpha$ 1-4Gal $\beta$ 1-4Glc-

**Table 5.1 Glycoconjugates recognised by MUB**

Interestingly, no binding of MUB to sialylated structures linked  $\alpha$ 2-6 or  $\alpha$ 2-3 to Gal was observed, which is in contrast to the results from microtitre plate competition studies showing that Neu5Ac and 6'SL were able to reduce MUB binding to pPGM (see 5.1.2).



**Figure 5.9 MUB binding to GI mucins in the presence of sugar molecules**

MUB (17  $\mu\text{g/mL}$ ) probed on a mucin array with GI mucins from intestinal epithelial cell lines and different animal species as well as binding controls in the absence (blue) or presence of soluble sugars (100 mM), Fuc (yellow), Gal (red), GlcNAc (green) and Lac (white). Bound MUB detected primary anti-MubRI (1:2,000) and secondary anti-rabbit-Alexa Flour555 (1:1,000). Data presented as average of three independent experiments with 6 probe replicas per array with error presented as standard deviation of the mean. Signals above threshold of about 1,700 RFU represent binding to probes (red line).

After identification of these potential sugar ligands, inhibition studies were performed with MUB incubated in the presence of Fuc, Gal, GlcNAc and Lac (100 mM) using the mucin arrays as described above. MUB co-incubated with Fuc or Gal showed the same level of binding compared to MUB in the absence of soluble sugar molecules, while increased binding was observed for GlcNAc and Lac (Figure 5.9). These data indicate that Fuc and Gal may not be involved in the binding of MUB to mucins used in this array and that MUB however interacts with GlcNAc and Lac residues promoting binding to mucin. This enhanced binding may be due to cross-linking of MUB molecules by these sugar molecules causing MUB oligomerisation or even aggregation. Indeed, the ability of MUB to oligomerise in solution has been demonstrated by AUC, albeit in the absence of soluble sugars, and native PAGE experiments, and co-crystallisation studies showed the cross-linking of two MubR5 molecules by Man (see 3.3.2 and 4.1).

In addition to the mucin and neoglycoconjugate microarray experiments (as above), the full-length MUB protein was screened against the mammalian printed glycan array (version 5.0) by the Consortium for Functional Glycomics (CFG) [411]. However, no binding of MUB to mammalian glycans was observed (see Appendix XIII).

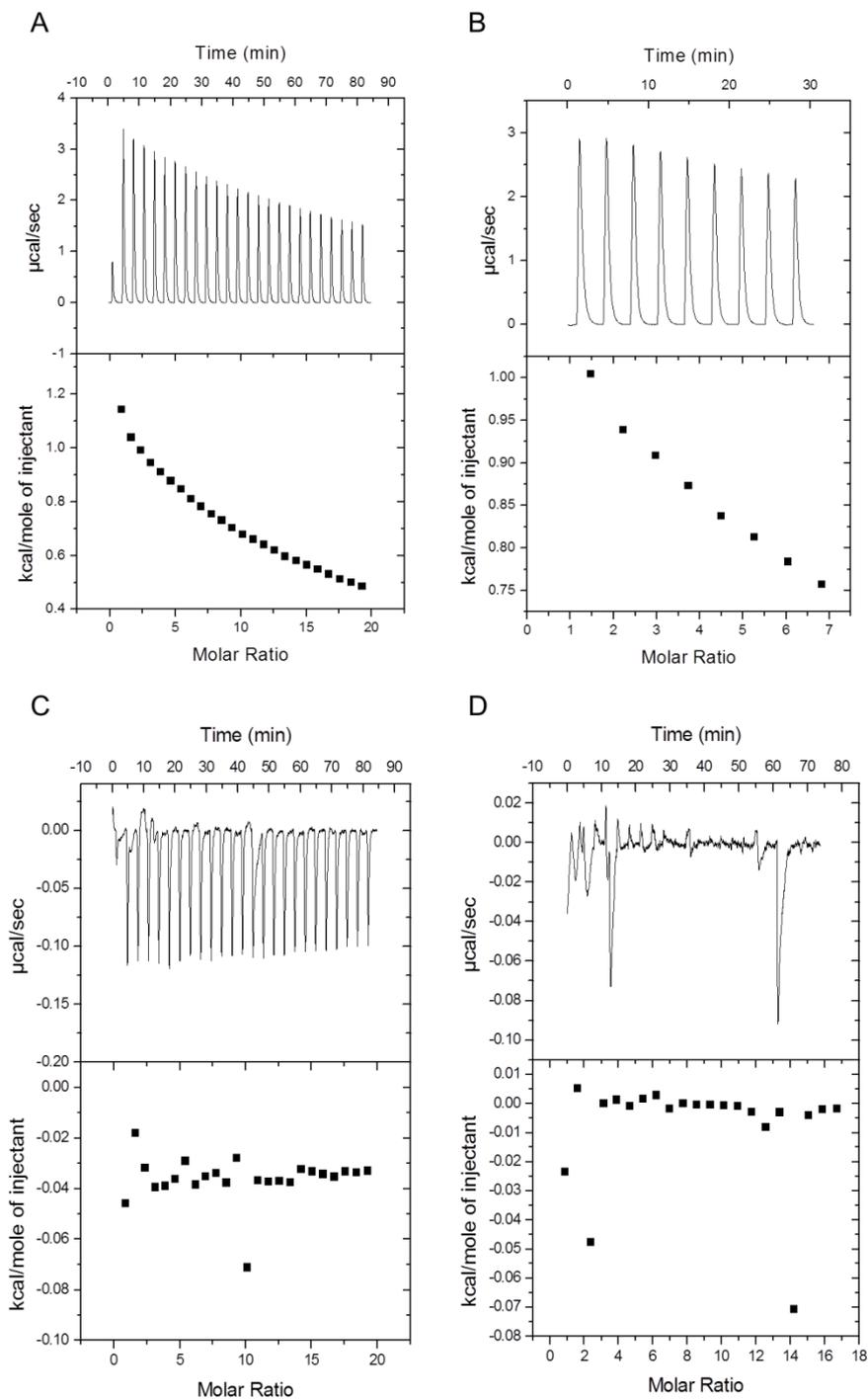
## **5.2 Binding of MUB proteins to glycans**

After demonstrating the binding of MUB proteins to different mucins, the involvement of glycans in this interaction and identification of potential sugar ligands in an initial screening, further studies were performed to investigate the sugar specificity of MUB proteins and characterise their sugar binding ability.

### **5.2.1 Assessment of MubR5 binding to mannose ligands using ITC**

Co-crystallisation studies of MubR5 with soluble mono- and disaccharides commonly encountered in mucin glycans revealed two adjacent MubR5 molecules in contact with a single Man molecule (see 4.1). In order to further investigate the binding ability of MubR5 to Man ligands, isothermal titration calorimetry (ITC) experiments were performed in collaboration with David Bolam (Newcastle University, UK) (see 2.4.1). The tri-, di- and monosaccharides, Mannose $\alpha$ 1-3(Mannose $\alpha$ 1-6)mannose (Man $\alpha$ 1-3(Man $\alpha$ 1-6)Man), Mannose $\alpha$ 1-6mannose (Man $\alpha$ 1-6man) and Man were

titrated into MubR5 (95  $\mu\text{M}$ ) as well as into a matching sample buffer control (see 2.4.1).



**Figure 5.10 Binding of MubR5 to mannosides investigated by ITC**

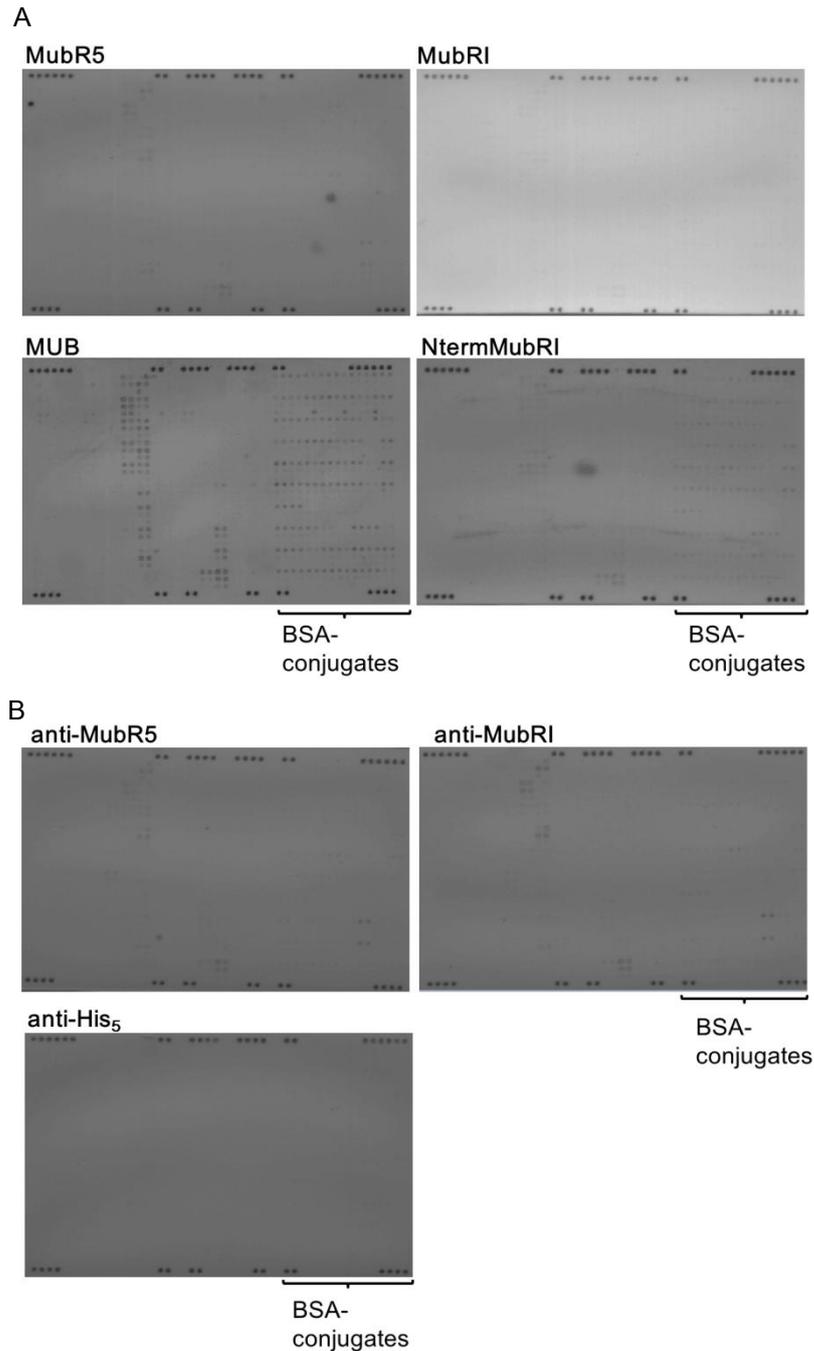
Titration of mannose ligands (10 mM) (A) Mannose- $\alpha$ 1-3(Mannose $\alpha$ 1-6)mannose (Man $\alpha$ 1-3(Man $\alpha$ 1-6)Man) (C) Mannose- $\alpha$ 1-6-mannose (Man $\alpha$ 1-6man) and (D) Man, or (B) PBS buffer control against MubR5 (95  $\mu\text{M}$ ). Upper panel showing injection heats (raw data) and lower panel integrated data fitted to a single site binding model.

MubR5 did not bind to Man $\alpha$ 1-6man or Man at 10 mM, as no change in heat was detected with increasing molar ratio of ligand to receptor (Figure 5.10 C and D). A decrease in heat was observed for (Man $\alpha$ 1-3(Man $\alpha$ 1-6)Man), however similar heat changes showed for the control titration of the trisaccharide into the sample buffer (Figure 5.10 A and B). Hence, MubR5 did not seem to bind to Man saccharides under the tested condition despite previous observations in X-crystallography co-crystallisation studies (see 4.1).

### **5.2.2 Screening of mono- and oligosaccharides for MUB protein binding by glycan arrays**

In order to dissect the binding ability and specificity of MUB and Mub repeats, additional glycan binding studies were performed using a nitrocellulose glycan array. The array contained immobilised plant derived glycans conjugated or non-conjugated to BSA, and BSA as a control, and was kindly provided by William Willats (University of Copenhagen, Denmark) (see 2.3.6.2) [412]. Purified MubR5 (0.3 mg/mL), MubRI (0.3 mg/mL), NtermMubRI (0.45 mg/mL) and native MUB proteins (60  $\mu$ g/mL) were incubated on the nitrocellulose arrays and bound proteins detected via primary anti-MubR5, anti-MubRI and anti-His<sub>5</sub> followed by secondary anti-mouse-AP or anti-rabbit-AP (see 3.1.1, 3.1.2, 3.3.1 and 2.3.5).

Only a few very faint binding signals were detected for the single Mub repeats, MubR5 and MubRI, which were however similar to those observed in the control experiments using primary and secondary antibodies in the absence of adhesins (Figure 5.11 A and B). An increase in positive spot number was observed for NtermMubRI, while no signal was detected for the primary and secondary antibodies in a control experiment (Figure 5.11 A and B). Among those, strongest signals were observed for Man glycans, Glc structures with  $\beta$ 1-4 and  $\beta$ 1-3 linkages as well as for arabinose and xylose glycans. No signal was detected for the BSA control suggesting that adhesin binding to spotted BSA-glycans is mediated by the carbohydrate structures conjugated to BSA and not to the carrier protein itself. In addition, a high number of positive spots corresponding to carrageenan molecules were observed. These carbohydrates are linear, helical polysaccharides isolated from seaweed and composed of repeating  $\alpha$ 1-3 and  $\beta$ 1-4Gal units with varying levels of sulphation. This may indicate interaction of these glycans due to their net negative charge.



**Figure 5.11 Glycan binding ability of MubR5, MubRI, NtermMubRI and MUB assessed using a glycan array**

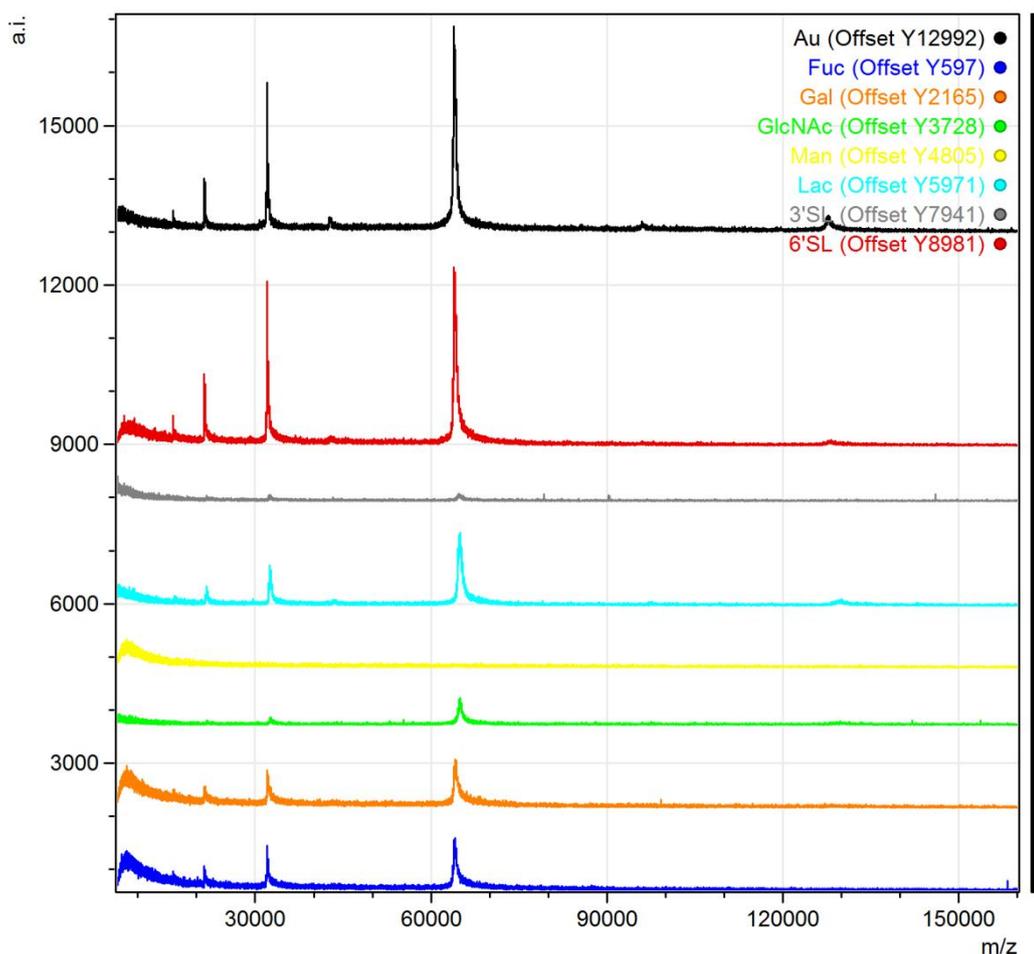
(A) Nitrocellulose glycan array spotted with plant derived glycans partially BSA-conjugated incubated with MubR5 (0.3 mg/mL), MubRI (0.3 mg/mL), MUB (60  $\mu$ g/mL) and NtermMubRI (0.45 mg/ml). Bound protein detected via primary anti-MubR5, anti-MubRI or anti-His<sub>5</sub>, and secondary anti-rabbit-AP or anti-mouse-AP. (B) Control membrane glycan array probed with primary and secondary antibodies in the absence of adhesins.

For the full-length MUB additional strong signals were detected corresponding to carrageenan,  $\beta$ -glucan, pectin, gum and mannan containing Gal, Man, Glc and

galacturonic acid (GalA) residues (for further information see Appendix XIV). Furthermore, signals were detected for the monosaccharides Man, Glc and Gal as well as for the Lac disaccharide.

Taken together, these preliminary data suggest a possible interaction of MUB proteins with glycan structures. However, the potential sugar binding ability of MUB proteins needs to be further investigated in alternative glycan assays. It is particularly important to address, whether the binding to carbohydrate structures depends on specific sugar recognition or is the sum of unspecific, weak ionic interactions, for example by competition studies with soluble sugar ligands or direct protein detection assays. As reported earlier, a correlation between the number of functional domains and the interaction capacity of MUB proteins was observed (see 5.1.1 and 5.1.2).

Moreover, since we showed that antibodies may in some ways interfere with the specificity of the interaction, a gold (Au)-array with mucin-like sugar molecules was used allowing an alternative detection method of bound adhesins based on MS, in collaboration with Sabine Flitsch (and Mirja Hartmann) (University of Manchester, UK). Briefly, the Au-arrays were functionalised with the monosaccharides Fuc, Gal, GlcNAc and Man, the disaccharide Lac, which was further modified via enzymatic synthesis to form 6'SL and 3'SL structures (see 2.3.6.3.1). Functionalised glycan arrays or Au-chips with self-assembled monolayers of linker and spacer molecules before functionalisation with sugar molecules were probed with MUB, MubRI-II-III, MubRV-VI, MubR8-V, MubR5, MubRV and Nterm, to assess binding and unspecific interaction of MUB proteins with linker molecules (see 2.3.6.3.2), respectively.



**Figure 5.12 Sugar binding of MubRI-II-III assessed using a sugar Au-array**

MubRI-II-III probed on a sugar Au-array with Fuc (dark blue), Gal (orange), GlcNAc (green), Man (yellow), Lac (light blue), 3'SL (grey) and 6'SL (red) and on non-functionalised Au-surface (black). MS spectra shown as average intensity (a.i.) plotted against mass-to-charge-ratio in a mass range of 6,500 to 160,000 m/z and presented offset for better visualisation.

MS spectra were obtained for all tested adhesins except for the native MUB after incubation of proteins on untreated Au-surfaces, allowing the detection of any bound protein molecules via MS on sugar-functionalised surfaces. (The presence of characteristic protein peaks but not signal intensity defines the binding event). Only MubRI-II-III showed reproducible binding to the sugar Au-array to Fuc, Gal, GlcNAc, Lac, and both sialylated Lac structures, but not to Man (Figure 5.12). No unspecific interaction for MubRI-II-III probed on non-functionalised linker monolayers was observed (data not shown). These data suggest the recognition of certain sugar molecules by MubRI-II-III and is in line with previous observations, that several Mub repeats may be necessary for receptor recognition. Furthermore, it is in

agreement with the results for MUB binding to glycans on a glycoconjugate array, where MUB bound to Fuc, Gal, GlcNAc and Lac containing glycan structures (see 5.1.3). In addition, previous findings were confirmed, which demonstrated that MubR5 does not recognise Man ligands in ITC experiments (see 5.2.1).

These different screening approaches also highlight the difficulty in assessing glycan-protein interaction and the necessity to carry out biomolecular interaction assays, such as ITC, SPR or bio-layer interferometry (BLI), to confirm or infirm the specificity of the interaction.

### 5.3 Discussion

The mucus layer is the first point of contact for intestinal bacteria with the host and serves as a habitat for the gut microbiota [25]. Bacterial adhesion to mucus is thought to be mediated via the interaction of mucin glycoproteins, the main structural components of mucus, with bacterial cell-surface adhesins [22, 179, 275]. A number of different mucus adhesins have been identified in lactobacilli and implicated in bacterial adhesion to mucus (see 1.4.2.). A correlation between the presence of mucus binding (Mub) or mucin-binding protein (MucBP) domains (Pfam database PF06458) in cell-surface proteins from *Lactobacillus* strains and their mucus binding ability has been reported (see 1.4.2) [8, 304]. The presence of the mucus-binding protein MUB of *L. reuteri* ATCC 53608, comprised of 14 Mub repeats of type 1 and type 2, is important for bacterial adhesion to mucus, as previously demonstrated in our lab (see 1.4.2) [302]. Reduced mucus adhesion was observed for the mutant strain 1063N of *L. reuteri* ATCC 53608, expressing a truncated MUB protein [302]. Besides the reported binding of recombinant Mub-repeat proteins in fusion with a maltose-binding protein (MBP) to chicken and porcine mucin in a microtitre plate assay, no further characterisation of the interaction of individual Mub repeats or of the native full-length MUB has been reported [283]. The current working hypothesis is that bacterial adhesins, such as MUB, mediate mucus interaction via the recognition of mucin glycans.

In order to test this hypothesis, we investigated the binding ability of recombinant single and tandem Mub-repeat proteins, N-terminal domain proteins and the native full-length MUB to different mucins and mucin glycans *in vitro*. The full-length native MUB, demonstrated binding to PSIM and pPGM after slot-blotting onto nitrocellulose membrane, while no binding was observed for Mub-repeat proteins or the Nterm region (see 5.1.1). However, MUB as well as MubR5 and MubRI-II-III showed concentration-dependent binding to PSIM and pPGM when incubated on immobilised mucin. In contrast, no MUB binding to PSIM, pPGM or human MUC2 was observed when mucins were probed on immobilised MUB protein in a microtitre plate assay (see 5.1.2), indicating the potential importance of correct ligand presentation. Additionally, MubRI-II-III but not MubR5 bound to colonic MUC2 from human biopsy samples after AgPAGE and Western-blotting (see 5.1.1). Furthermore, when screening a number of mucins for binding by native MUB using a GI mucin array,

binding was detected to mucins from various animal species, especially mouse, rat, pig, sheep and deer. MUB also bound to mucins isolated from the intestinal epithelial cell lines LS174T and HT29, which express the gel-forming mucins MUC2, MUC6, MUC5AC and MUC5B (see 5.1.3) [413-414]. Taken together these results demonstrate the binding of native MUB and recombinant Mub repeats but not the Nterm protein to different mucins and indicate a correlation between the number of Mub repeats and protein binding capability.

The binding ability of commensal adhesins to mucus and mucins has been demonstrated for a few lactobacilli. For example, the MapA protein of *L. reuteri* 104R showed binding to porcine small intestinal mucin and PGM in dot-blot assay similar to that one used for MUB-protein binding studies (see 5.1.1) [285]. Additionally, interaction with mucins isolated from colonic HT29-MTX epithelial cells was observed for the recombinant GroEL and EF-Tu proteins from *L. johnsonii* NCC 533 in a microtitre plate assay [297-298]. Another example are the fiber-like pilin structures, SpaCBA and SpaFED, identified in *L. rhamnosus* GG, which are similarly found in other *Lactobacillus* strains (see 1.4.2). The recombinant, labelled pilin proteins SpaB, C and F bound to human colonic mucus coated onto microtitre plates and binding was competed adding unlabeled protein. In addition, binding of bacterial cells expressing SpaC, F and B to human mucus was inhibited by SpaC and SpaF, but not by SpaB antiserum [288-289]. The positively charged Spa B (pI 8) seems to adhere to mucus, which has a net negative charge, via non-specific electrostatic interactions under the used assay conditions [289]. In contrast, native MUB and Mub-repeat proteins have low pI values of about 4.5 to 5.0 pI, as calculated from their amino acid sequence (see 2.6.2) and demonstrated via IEF (see 3.2). They thus have a neutral or negative net charge in the GI tract or under the used assay condition, indicating the specificity of the interaction with mucins. However, adhesion of MUB proteins to highly sialylated and sulphated mucins (see 1.2.2) due to electric charge cannot be completely excluded, and thus further investigation of the nature of MUB-mucin adhesion is needed.

From the observations made in this study, it seems that the number of Mub repeats is important for their ability to bind mucin and multivalent interactions are generally common in biology and considered to be necessary to achieve high avidity binding

[415]. For some *L. plantarum* strains, expressing the mannose-specific adhesin Msa, a partial correlation between Man-dependent yeast agglutination and varying numbers of MucBP domains in Msa proteins can be observed (see 1.4.2) [312]. Studies with *Staph. aureus* showed that at least three out of 11 fibronectin-binding repeats (FnBRs) of the cell-surface fibronectin binding protein (FnBP), which possess different binding affinities for fibronectin, are necessary for target cell adhesion and invasion [416-417]. In addition, the MSCRAMM LigB from pathogenic *Leptospira* species contains Ig-like domains, as found in other MSCRAMM (see 4.4), and shows an increase in ligand binding affinity with increasing number of Ig-like domains [418].

In addition, mucin binding control experiments revealed some level of MUB protein binding to BSA and IgGFc fragments (see 5.1.2). This suggests that MUB may interact with mucin proteins via the recognition of the protein backbone or with other protein ligands found in mucus (see 1.1.2). Indeed, the binding ability of Mub repeats to IgA and other Igs has been demonstrated in our lab [282]. Igs, especially IgA, are found in large amounts in mucus [25, 27]. Proteomic analysis of mucus from stomach, small intestine and colon of mice also revealed a high abundance of albumin, however it is unclear whether this is due to contamination during sample extraction [28]. The ability of a mucus-adhesin to bind mucin and another protein in human colonic mucus was reported for 32-MmubP from *L. fermentum* (see 1.4.2) [419]. Further experiments, are warranted to investigate the potential binding of MUB proteins to the mucin protein backbone using for example glycosylated and de-glycosylated mucins after enzymatic or chemical treatment, or by performing sugar inhibition studies [420-422].

The role of mucin glycans in the interaction of MUB proteins was further assessed in sugar competition studies. MUB showed reduced binding to pPGM in the presence of Neu5Ac, Lac and 6'SL in microtitre plate assays (see 5.1.2). In contrast, increased MUB binding was observed for Lac and GlcNAc, while Gal and Fuc did not have an effect on MUB binding to GI mucins in a mucin array (see 5.1.3). This may indicate the cross-linkage of MUB molecules by sugar residues. The tendency of MUB to form oligomers has been demonstrated by native PAGE and AUC experiments albeit without further addition of sugars (see 3.3.2). Furthermore, *L. reuteri* ATCC 53608 but not the mutant strain 1063N, expressing a truncated MUB protein on the cell surface, shows auto-aggregation ability (see 4.4) [302]. Co-crystallisation studies of MubR5

demonstrated the cross-linking of two MubR5 molecules by Man (see 4.1). However, ITC experiments with MubR5 and Man ligands did not confirm the binding of Man by MubR5 (see 5.2.1). Man is part of the core *N*-glycan structure and thus less abundant in mucin glycans, which are mainly *O*-linked (see 1.2). However, MucBP domain containing Msa proteins from *L. plantarum* strains seem to be involved in Man-dependent binding to epithelial cells (see 1.4) [293, 312]. In addition, the binding of Man by the FimH adhesin present in uropathogenic *E. coli* (UPEC), which colonise the urinary bladder mucosa and whose reservoir is thought to be the GI tract, is well established (see 1.4.1) [234, 423]. Even though, interaction of FimH with mucins has not yet been reported, the bladder is interestingly lined by a 'bladder mucus layer' composed of glycosaminoglycans (GAGs) also known as mucopolysaccharides. GAGs are long polysaccharide chains which can be *O*- or *N*-linked to proteins and the GAG keratan sulphate is linked to the *N*-glycan core structure [424-427]. It is thus possible that commensal bacteria adhere to mucins via recognition of less abundant Man sugars by cell-surface adhesins.

Further screening for specific glycan ligands with native MUB was performed using different glycan array platforms. This analysis revealed binding to glycan structures in a neoglycoconjugate array, which were composed of LacNAc and Lac units including terminal blood group epitopes, H2 (Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc) or sulphated Le<sup>x</sup> (see 5.1.3). Additionally, membrane glycan arrays revealed binding of MUB to Man, Gal, Glc and Lac-containing sugar molecules as well as to a number of highly sulphated polysaccharides (see 5.2.2), again indicating a potential role for unspecific electrostatic interactions in MUB binding to glycans as discussed earlier. In an Au-glycan array the recombinant triple domain MubRI-II-III showed binding to Fuc, Gal and GlcNAc monosaccharides, as well as the disaccharide Lac and sialylated Lac structures with  $\alpha$ 2-3 and  $\alpha$ 2-6 linkages. As reported earlier, no binding was observed for single or double Mub-repeat proteins, and the Nterm protein, indicating that MUB sugar binding is mediated by the presence of several Mub repeats (see 5.2.2). Additionally, no difference can be observed between the mucin or sugar binding ability of Mub type 1 or type 2 repeats (see 5.1.1 and 5.2.2). Very few studies have investigated the nature of mucin ligands recognised by mucus adhesins. The adhesin GAPDH from *L. plantarum* LA 318 binds to human colonic mucin, and recent characterisation studies using SPR showed its specificity for ABO blood antigens

H (Fuc $\alpha$ 1-2Gal-), A (GalNAc $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal) and B (Gal $\alpha$ 1-3[Fuc $\alpha$ 1-2]Gal), whereas comparably weaker binding to Neu5Ac was observed [294]. Similarly, the BabA and SabA protein of *H. pylori*, binds to gastric mucins via recognition of ABO blood group antigens, Le<sup>a</sup> or sialyl-Le<sup>a</sup> and sialyl-Le<sup>x</sup>, respectively, which contain Gal, GlcNAc, Fuc and Neu5Ac residues (see 1.4.1) [207, 242]. SPR experiments with the pathogenic *E. coli* F17 G pili protein demonstrated binding to GlcNAc but not to Gal, GalNAc or Glc sugar molecules [237]. The enteric parasite *E. histolytica*, which expresses a cell-surface Gal/GalNAc lectin, binds to human and rat colonic mucin and binding can be inhibited by Gal and GalNAc [246-247]. These data suggest that pathogenic and commensal adhesins may share the same molecular targets at the mucosal surface, although more work is required to further biochemically assess the exact sugar affinity and specificity of MUB proteins.

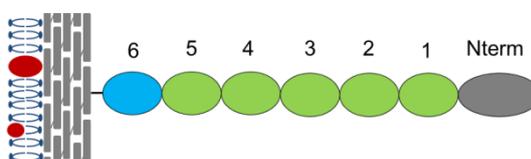
Ligand recognition of fimbrial adhesins is often mediated via their N-terminal protein domain, while additional domains act as a scaffold to present the ligand binding domain and enhance ligand accessibility. For example the PapG, F17-G and FimH adhesins present in *E. coli* pili are located at the N-terminal pili tip, where they mediate binding to Gal, GlcNAc and Man ligands, respectively (see 1.5) [235, 237, 428-431]. Another example are siglecs, which are composed of variable numbers of C-domains and an N-terminal V-domain (V-set Ig-like domain) which mediates sialic acid binding [432]. X-ray crystal structures are available for the V-domain of Siglec-1 (sialoadhesin) and Siglec-7 in complex with Neu5Ac $\alpha$ 2-3Gal and Neu5Ac, respectively [433-434]. However, while the first, N-terminal domain is sufficient for sialic acid recognition in Siglec-1, Siglec-2 (CD22) requires the first two N-terminal domains for ligand binding [435-437]. In the SpaCBA pili present on the cell-surface of *L. rhamnosus* GG, the SpaC adhesin, mediating binding to human intestinal mucus, has been located at the pilus tip as well as along the pilus fiber [288, 438]. In contrast, the N-terminal domain of MUB did not demonstrate any binding to mucin or mucin glycans under the used assay conditions *in vitro*, but binding was observed for MubRI-II-III comprising the first three N-terminal Mub repeats of MUB. These observations may indicate an alternative molecular mechanism of ligand interaction of MUB compared to other elongated adhesion molecules.

This study additionally highlighted potential pitfalls of *in vitro* screening approaches when studying protein-glycan interaction, including for example adequate ligand presentation, appropriate detection methods, blocking agents or suitable binding controls [439]. Significant advances have been made to improve and develop glycan array platforms that account for the complex glycan landscape found in natural systems. These include the natural ligand presentation using glycan material isolated from natural sources and direct immobilisation methods without the use of chemical linkers or use of carrier proteins such as albumin [409, 412, 440-443]. Alternative strategies to obtain glycan material for linker based surface-conjugation are chemical, chemoenzymatic or enzymatic syntheses, which avoid microheterogeneity present in naturally sourced glycans and aim for a natural mimetic surface presentation [444-447]. Although, fluorescent-labelling is one of the most frequently used detection methods for protein-carbohydrate interactions, label-free technologies such as SPR or MS can be combined with glycan-arrays. A recent comparison between two sialoglycan microarrays, differing in the number of diverse sialic acid structures, using fluorescently labelled antibodies for detection, revealed their potential to provide complementary information, however also highlighted possible factors leading to differential glycan recognition, such as glycan immobilisation [448]. These observations stress the need of using additional techniques such as SPR or ITC to biochemically evaluate the interaction of potential sugar ligands identified by glycan array screening.

In summary, we demonstrated the binding of native MUB and recombinant Mub repeats to mucins and different mucin glycans as indicated by sugar inhibition assays and sugar screens (see 5.1 and 5.2). In particular, we showed binding to blood group antigens, Lac, sialylated Lac as well as to Gal, GlcNAc and Fuc monosaccharides or glycan structures containing those sugars using different glycan arrays (see 5.2.2). However, the interaction of potential sugar ligands with MUB proteins requires further characterisation. Inhibition studies with soluble sugar molecules using Au-arrays are ongoing. In order to validate carbohydrate-adhesin interactions, SPR, ITC and BLI experiments need to be performed, which have the additional benefit of label-free detection. Additionally, co-crystallisation experiments of multiple Mub-repeat proteins with sialylated Lac or blood group antigens will provide functional basis for the glycan recognition by MUB proteins.

## CHAPTER 6 STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF LAR0958

Lar0958 is a recently identified cell-surface adhesion protein present in the human isolate strains *L. reuteri* MM4-1a, DSM 20016<sup>T</sup> and JCM 112<sup>T</sup> [302, 324]. It is a modular protein with an N-terminal signal peptide and a C-terminal LPXTG-motif facilitating covalent attachment to the bacterial cell wall.



**Figure 6.1 Schematic representation of Lar0958**

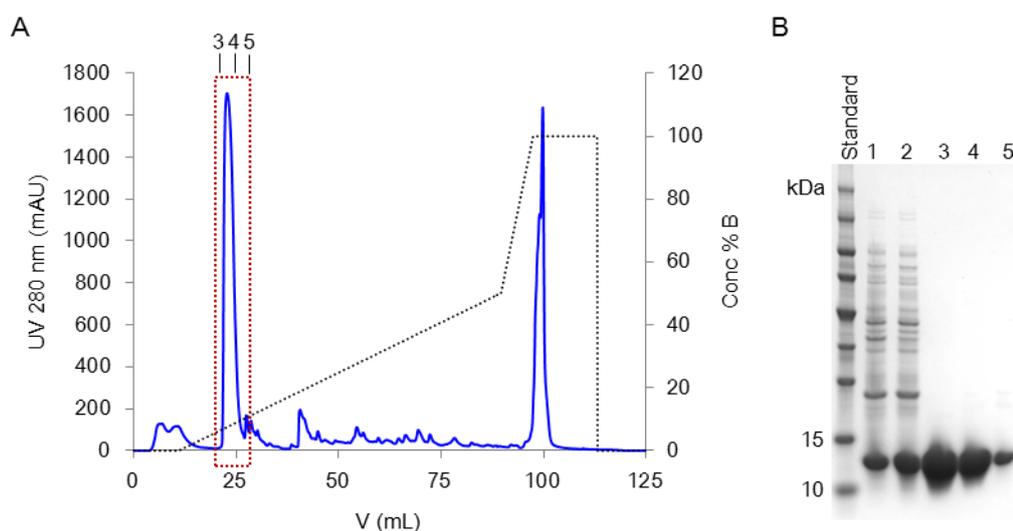
Domain organisation of the Lar0958 cell-surface protein of DSM 20016<sup>T</sup> with five virtually identical repeats (1-5, green), one repeat of lower homology (6, blue) and the N-terminal domain (Nterm, grey). The LPXTG-motif for covalent attachment to cell wall peptidoglycan is presented as a black line.

The full length Lar0958 contains 4 identical repeats of 96 residues, a virtually identical fifth repeat of 97 residues, and a final 94 amino acid repeat of lower homology (46% identity) (Figure 6.1). *L. reuteri* DSM 20016<sup>T</sup> and MM4-1a bind to mouse colonic mucus (MCM), whereas the MM4-1a Lar0958 knock out (KO) strain MM4KO shows reduced binding ability to mucin [302, 324]. The specific ligands involved and the biochemical mechanism of adhesion remains, however, elusive. In order to structurally and functionally characterise Lar0958, we cloned and heterologously expressed a recombinant single Lar0958 repeat in *E. coli*.

### 6.1 Cloning, expression, purification and biophysical characterisation of Lar0958

For the cloning of a single *lar0958* gene, primers were designed to anneal to the border regions of the identical repeats of the full length *lar0958* gene of *L. reuteri* DSM20016<sup>T</sup> (BioProject accession numbers PRJNA15766 and PRJNA58471). The PCR-amplified *lar0958* repeat (see 2.2.1) was cloned into the pETBlue-1 AccepTor

vector (see 2.2.4), the sequence integrity verified by automated DNA sequencing (see 2.2.6) and the pETBlue-1 AccepTor\_*lar0958* vector construct transformed into *E. coli* Tuner(DE3)pLacI<sup>2</sup> cells (see 2.2.5). The single Lar0958 repeat protein was expressed in the periplasm of the expression host in a soluble form at 37°C after induction with 1 mM IPTG (see 2.3.1.1), but some recombinant protein was also present in the insoluble bacterial pellet fraction (Figure 6.2 B). Lar0958 was extracted from bacterial cells via a freeze-thaw method, thus avoiding cell disruption and contamination with host cell proteins in the cell extract (see 2.3.1.2). The clarified crude extract was applied to an IEC column and Lar0958 eluted in a distinct peak (Figure 6.2 A).



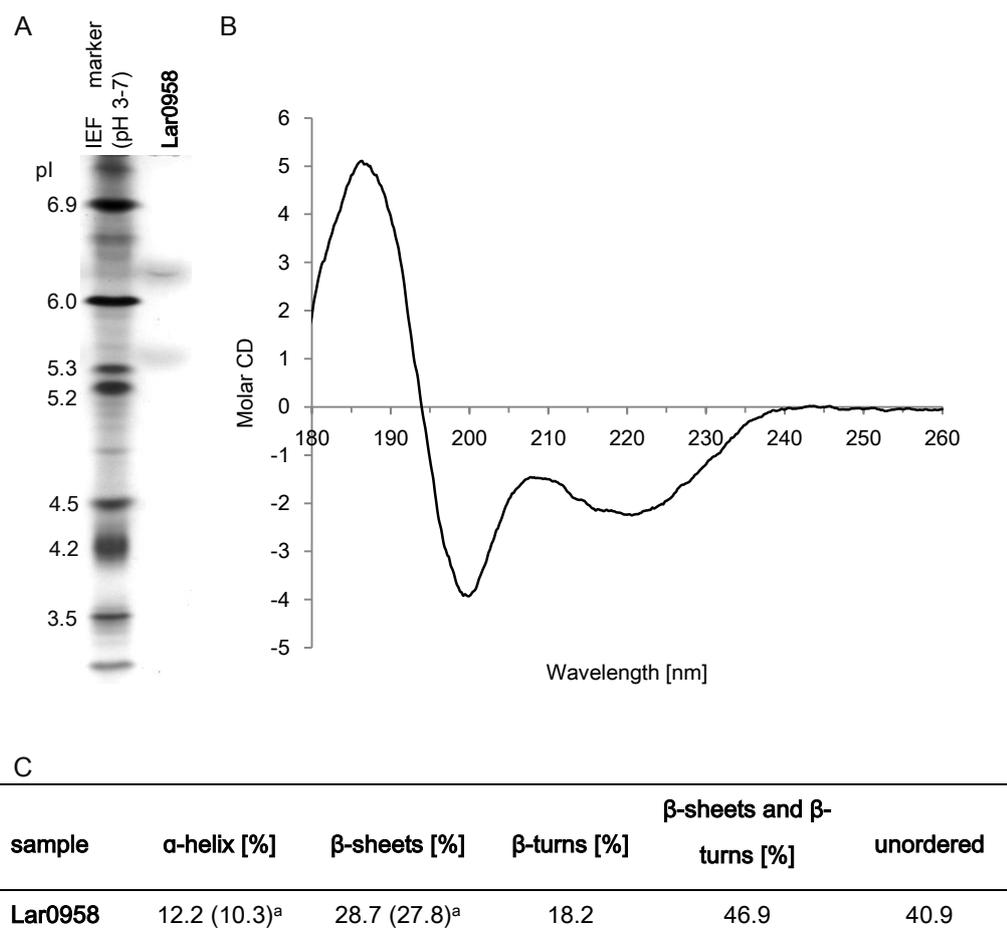
**Figure 6.2 Lar0958 protein purification via IEC**

(A) IEC profile of Lar0958 (red) and (B) SDS-PAGE gel analysis of IEC elution fractions (as indicated in A) (1 pellet, 2 crude extract, 3-5 elution fractions).

The IEC elution fractions were analysed by SDS-PAGE and only a single protein band with an apparent MW of 11.6 kDa was observed, which is in good agreement with the 10.4 kDa theoretical MW of Lar0958 (see 2.6.2). The absence of contaminating bands upon SDS-PAGE analysis revealed the sample homogeneity to be sufficient for further functional and structural studies (Figure 6.2 B).

In order to determine the pI value of Lar0958, to examine proper protein folding and to obtain initial information on secondary structural elements of the recombinant Lar0958

protein, IEF and CD experiments were conducted (see 2.4.1 and 2.4.2). IEF was performed via gel electrophoresis in a pH range of 3 to 7 and a voltage gradient (see 2.3.4.3). The pI value of Lar0958 was determined as 6.23 compared to the IEF standard (Figure 6.3 A) in reasonable agreement with the theoretical pI value of 6.09 (see 2.6.2).



**Figure 6.3 IEF and CD of Lar0958**

(A) IEF of Lar0958 by PAGE for pI determination in comparison to a pI standard. (B) Far UV CD spectra of purified Lar0958 over a scan range of 180-260 nm and (C) secondary structure analysis using the DichroWeb online tool (and based on X-ray crystal structure)<sup>a</sup>.

A far UV CD spectra was recorded for the Lar0958 protein in ultrapure water over a scan range of 180-260 nm and further analysed by the DichroWeb online using the CONTIN analysis program (Figure 6.3 B) (see 2.4.2). Secondary structural elements give distinct spectra characteristics and secondary structure determination of Lar0958

revealed an  $\alpha$ -helix content of 12.2% and a combined  $\beta$ -sheet and  $\beta$ -turn composition of 46.9% (Figure 6.3 C). These data suggest a difference in protein fold between Lar0958 and Mub repeats, as CD analysis of recombinant Mub proteins showed an  $\alpha$ -helix content of 1% (see 3.2).

## 6.2 Structure determination of a single Lar0958 repeat

X-ray crystallography of a purified single Lar0958 repeat protein was used to obtain high-resolution structure information for the Lar0958 cell-surface protein repeat and to gain structural insight into recognition of mucins (see 6.1).

For crystallisation experiments, recombinant purified Lar0958 was dialysed against 10 mM sodium phosphate buffer and concentrated to 12 mg/mL. Initial crystallisation condition screens were performed and single prismatic crystals of high quality were observed for the precipitation solution containing 0.2 M ammonium sulphate, 0.1 M sodium acetate (pH 4.5) and either 25% (w/v) PEG 4,000 or 30% (v/v) PEG 2,000 MME (monomethylester) (see 2.5.1). Higher quality diffraction was observed for crystals grown in 25% (w/v) PEG 4,000 compared to 30% (v/v) PEG 2,000 MME after adding 20% ethylene glycol to the crystallisation mother liquor, and this condition was thus used for further crystallisation experiments.

Due to the lack of structural information on Lar0958 proteins or protein homologues to provide a search model for MR, phasing was attempted by derivatisation of crystals with different anomalous scatterers, including KBr. Following a short soak of a crystal in cryoprotectant solution (as described) containing 0.5 M KBr, an X-ray fluorescence scan at the bromide K-edge was performed, where the inflection and peak energies were calculated to be 13,472.5 eV and 13,475.5 eV, respectively. Subsequently, a single wavelength anomalous dispersion (SAD) data set was collected at 13,472 eV (0.9203 Å) showing measureable diffraction to a resolution of 2.0 Å. After integration and reduction of the SAD diffraction data (see 2.5.1.3), 30 bromide ion sites were located and initial phases estimated using AutoSol [449]. The highest occupancy of a bromide ion located on the surface of a Lar0958 molecule was found to be 0.48 (Table 6.1). A first molecular model of Lar0958 was built by AUTOBUILD and a final model refined with an  $R_{\text{cryst}}$  of 17.6% and an  $R_{\text{free}}$  of 21.4% with RESOLVE [345, 348] (Table 6.2) (see 2.5.1.3).

Bromide atom	Occupancy	B-factor (Å <sup>2</sup> )
1	0.48	17.8
2	0.36	17.2
4	0.35	12.7
3	0.33	20.7
9	0.32	16.3

**Table 6.1 Bromide sites with five highest occupancies found by AutoSol**

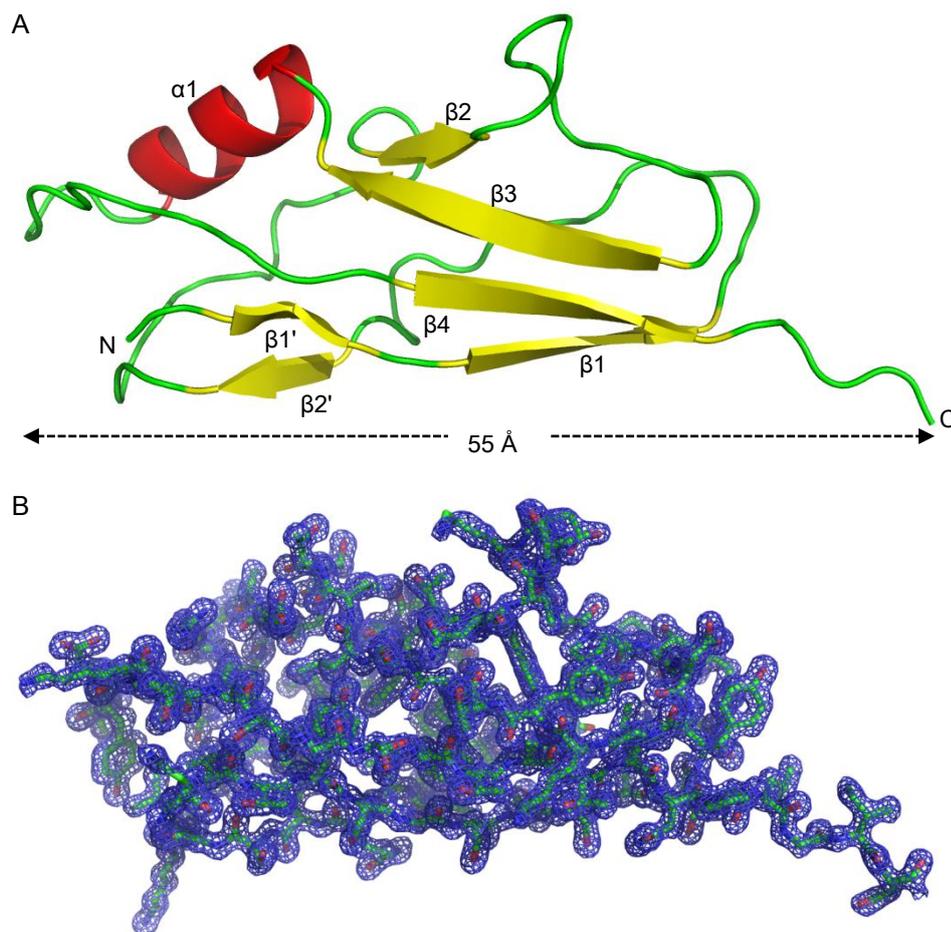
A native Lar0958 crystal diffraction data set was then collected at an X-ray wavelength of 0.9795 Å (12658 eV) to a resolution of 1.5 Å. The corresponding X-ray data collection and reduction parameters are listed in Table 6.2. The solvent content of the crystal was estimated to be 50% for three Lar0958 molecules in the ASU (see 2.5.1.3). The diffraction data were indexed in a primitive monoclinic lattice and processed by MOSFLM and SCALA in the  $P 2_1$  space group (Table 6.2) [333]. The Lar0958 model determined from a KBr derivatised crystal (as described above) was used for MR. After several rounds of manual model building in COOT [343] and refinement, the final crystal structure (PDB entry 4NG0) was refined at a resolution of 1.5 Å with  $R_{\text{cryst}}$  of 15.8% and  $R_{\text{free}}$  of 18.7%.

Three Lar0958 molecules were present in the ASU, as indicated by solvent content analysis of the crystal (see 2.5.1.3), each consisting of 96 residues comprising Lys2 to Asp97, thus only lacking the N-terminal Met. The three Lar0958 molecules showed only minor structural difference when compared by DaliLite (see 2.6.3) [450] with Z-scores of 18.9-19.1 and r.m.s.d. values of 0.3-0.5 Å.

	Lar0958 (KBr SAD data set)	Native Lar0958 repeat (PDB entry 4NG0)
<b>Data collection</b>		
Beamline		Diamond i02
Space group	P 1 2 <sub>1</sub> 1	P 1 2 <sub>1</sub> 1
Cell parameters: a, b, c (Å), (°)	52.8, 36.8, 75.3; $\alpha = \gamma = 90, \beta = 100.4$	52.8, 37.0, 75.6; $\alpha = \gamma = 90, \beta = 100.6$
Wavelength (Å)	0.920	0.9795
Resolution (Å)	37-2.0 (2.1-2.0)	46.8-1.5 (1.6-1.5)
R <sub>sym</sub> (%)	3.6 (51.0)	2.8 (14.2)
I/ $\sigma$ I	30.7 (1.7)	30.1 (9.2)
Unique reflections	30893 (2235)	45978 (3327)
Completeness (%)	99.6 (98.4)	99.4 (97.8)
Multiplicity	6.6 (2.5)	4.0 (3.9)
Anomalous completeness (%)	66.3 (40.0)	
Anomalous multiplicity	3.7 (1.9)	
Overall B-factor (Å <sup>2</sup> )	17.2	13.6
FOM (SAD phasing)	0.364	
<b>Refinement statistics</b>		
Molecules per AU		3
Total atoms		2912
Water molecules		690
R <sub>cryst</sub> (%)		15.8 (18.5)
R <sub>free</sub> (%)		18.7 (23.9)
<b>Ramachandran analysis</b>		
Most favoured		98.9
Allowed (%)		1.0
Outliers (%)		0
<b>r.m.s.d.</b>		
Bonds (Å)		0.006
Angles (°)		0.988
Planes (Å <sup>2</sup> )		0.004
Mean atomic B-factor (Å <sup>2</sup> )		11.7

Table 6.2 Data collection and refinement statistics of Lar0958 diffraction data sets

The Lar0958 structure displayed an ubiquitin-like  $\beta$ -grasp fold (Figure 6.4 A and B). This fold contains a pair of parallel  $\beta$ -strands ( $\beta$ 1,  $\beta$ 4) and a pair of antiparallel  $\beta$ -strands ( $\beta$ 2,  $\beta$ 3), forming a four stranded mixed  $\beta$ -sheet of the order 1432 connected by an  $\alpha$ -helix.

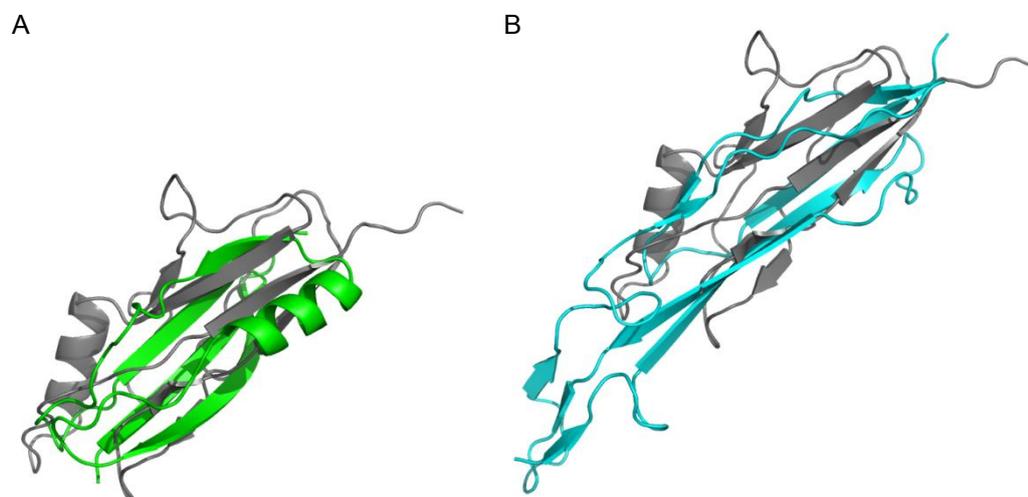


**Figure 6.4 Representations of the Lar0958 repeat X-ray crystal structure**

(A) Crystal structure of Lar0958 at 1.5 Å resolution with helices in red and  $\beta$ -sheets in yellow. Secondary structural elements and, N- and C-termini labeled. (B) Representation of the Lar0958 structure as sticks overlaid with  $2mF_{\text{obs}} - DF_{\text{calc}}$  electron density map contoured at 1.0  $\sigma$  level.

Additionally, Lar0958 possesses a third pair of anti-parallel  $\beta$ -strands ( $\beta 1'$ ,  $\beta 2'$ ). An ubiquitin-like  $\beta$ -grasp fold is also present in the N-terminal B1 domain of MubR5 and MubRV (see 4.2.2). Pair wise structural alignment using DaliLite revealed low similarity of Lar0958 to the B1 domain of MubR5 (Z-score of 1.9, 49 aligned residues, r.m.s.d. 5.9 Å) (Figure 6.5 A). Interestingly, Lar0958 showed a higher structural similarity to the B2 domain of MubRV with a Z-score of 2.0 (61 aligned residues, r.m.s.d. 8.2 Å) than to its B1 domain (Z-score 1.8, 49 aligned residues, r.m.s.d. 6.7 Å) (Figure 6.5 B). The B2 domain of MubR5 and MubRV also forms a modified  $\beta$ -grasp fold, where the connecting helix of the 4 stranded  $\beta$ -sheet is replaced by

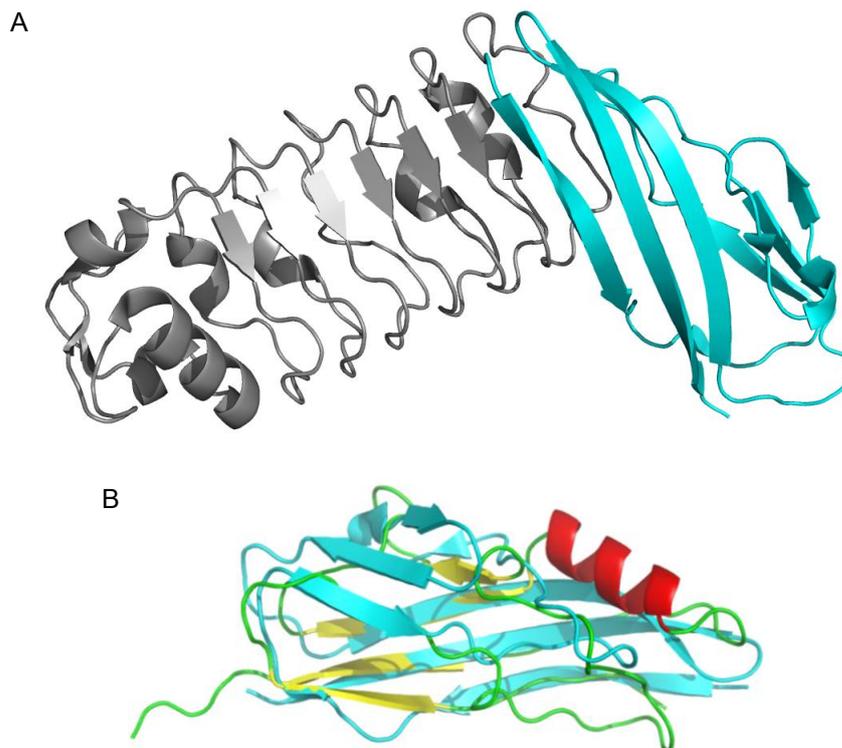
a two-stranded  $\beta$ -sheet. This domain is annotated as a mucus binding protein (MucBP) domain in the Pfam database (PF06458) [451].



**Figure 6.5 Structural overlay of Lar0958 with MubR5 and MubRV**

Structural alignment of Lar0958 (grey) with (A) the B1 domain of MubR5 (green) and (B) the B2 domain of MubRV (blue) by DaliLite.

When comparing Lar0958 with structures in the protein database (PDB) using DALI [358], structural similarity was observed for InIC and other members of the internalin protein family, i.e. InIA (PDB entry 1O6T) and InIB (PDB entry 2WQW). Superposition of Lar0958 and InIC (PDB entry 1XEU) gave a Z-score of 3.8 and an r.m.s.d. of 2.8 Å for 69 aligned residues (Figure 6.6 A) despite a sequence identity of only 7%. Internalins are bacterial cell-surface proteins of *L. monocytogenes*, an intracellular food-borne pathogen that crosses the intestinal barrier and causes severe infections [270, 452]. The structure of internalin proteins comprises an N-terminal cap followed by a concave shaped leucine rich repeat (LRR) domain and C-terminal Ig-like domains including the inter-repeat (IR) and additionally the B-repeat domain [321]. Lar0958 shows highest structural homology to the C-terminal IR-domain of InIC, which lacks the helix bridging the two pairs of  $\beta$ -sheets in an ubiquitin  $\beta$ -grasp fold (Figure 6.6 B). The B1 domain of MubR5 has been identified as structural homologue of the B-repeat of InIB having an r.m.s.d. of 2.4 Å for 61 aligned residues (Z-score of 5.2) (see 4.2.2) [453]. In contrast, Lar0958 does not show structural similarity to the InIB B-repeat as revealed by its low Z-score of 1.3 (r.m.s.d. 6.2 Å over 44 aligned residues).



**Figure 6.6 InIC crystal structure and overlay with Lar0958**

(A) Representation of InIC (PDB entry 1XEU) showing the internalin Leucine rich repeat (LRR) domain (grey) and the IR domain (blue). (B) Overlay of Lar0958 (red, yellow, green) and the IR domain (blue). The alignment has a Z-score of 3.1 over 68 aligned residues and an r.m.s.d. value of 2.8 Å.

The structural similarity of Lar0958 and Mub repeats from *Lactobacillus* to the internalin Ig-like domains, IR and B-repeat, of *Listeria*, may indicate a similar role of these cell-surface protein repeats in adhesion to structural components in the intestine. However, there is a current lack of knowledge on the nature of the ligands and the mechanism of interaction.

The MarkUs function annotation server was used to identify structural similarities and surface properties of functional relevance between Lar0958 and structures in the PDB [386]. Four solvent accessible cavities were identified by SCREEN for Lar0958 with surface areas between 66.1 and 25.2 Å<sup>2</sup> and diameters of 8.2 to 5.9 Å, which may function as potential ligand interaction sites [388]. The program VASP used for cavity comparison of Lar0958 with those of known structures did not provide further insight into potential ligands or the mode of interaction [454]. Additionally, Lar0958 contains

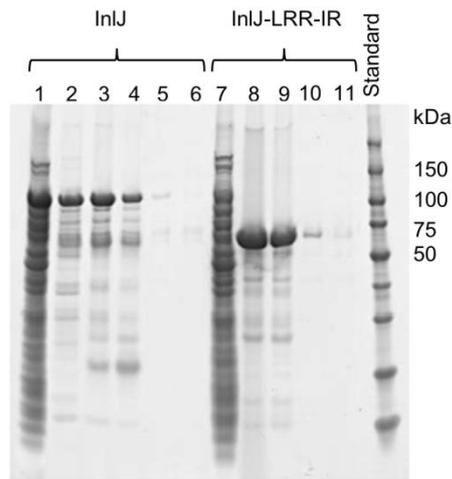
a potential interface for protein-protein interaction (as predicted by PredUs) formed by residues of the sheet  $\beta 3$  and the loop connecting  $\beta 2$  and  $\beta 3$  (Figure 6.4) [387].

### 6.3 Expression and purification of internalin proteins

The X-ray crystal structure of Lar0958 revealed structural similarity to the IR domain of internalin cell-surface proteins of *L. monocytogenes*, especially InIA and InIB (see 6.2). InIA and InIB are the best functionally characterised internalins to date and have been demonstrated to bind to the human host cell receptors E-cadherin and the Met receptor tyrosine kinase, respectively, via the LRR domain to facilitate bacterial cell entry [455-457]. Interestingly, the internalin proteins, InIB, InIC, InIJ and InIJ-LRR-IR, only comprising the LRR domain and the IR domain, bind with different affinities to human intestinal mucin MUC2 *in vitro* by a yet unknown mechanism [249]. In order to investigate the binding of Lar0958 to human MUC2 as compared to internalins, InIJ and InIJ-LRR-IR were heterologously expressed in *E. coli* and purified.

The *E. coli* BL21Star™ (DE3) strains harbouring the pET101-InIJ or pET101-InIJ-LRR-IR vectors (courtesy of P. Cossart, Institute Pasteur, Bacteria-cell interactions Unit, Paris, France), encode the *inIJ* gene, also designated as *Imo2821*, or the *inIJLRRIR* gene fragment of *L. monocytogenes* EGD-e (GenBank accession number CAD01034) [458]. The N-terminal His-tagged proteins InIJ and InIJ-LRR-IR were expressed in *E. coli* at 37°C after induction with 1 mM IPTG (see 2.3.1.1).

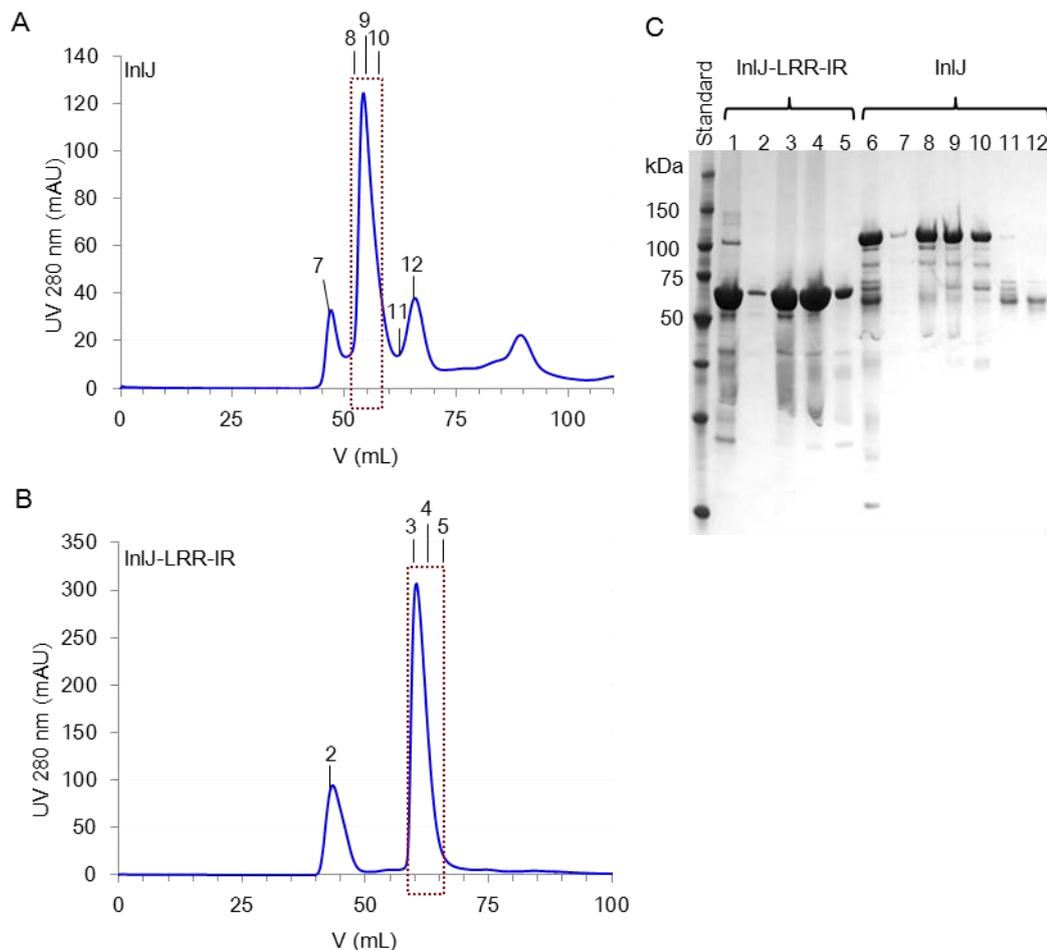
The recombinant internalin proteins were extracted from bacterial cell pellets by ultrasonication and the clarified protein extract applied to an IMAC resin column (see 2.3.1.2 and 2.3.1.2). IMAC wash and elution fractions were analysed by SDS-PAGE and showed a heterogeneous sample composition with predominant protein bands at an apparent MW of about 106 kDa and 65 kDa for InIJ and InIJ-LRR-IR, respectively (Figure 6.7). These apparent MW are higher than the theoretical MW of both proteins calculated as 87.5 kDa and 56 kDa by based on their amino acid sequence (see 2.6.2). Both protein bands reacted with anti-His<sub>5</sub> after Western-blotting (data not shown).



**Figure 6.7 Purification of InIJ and InIJ-LRR-IR by IMAC**

(A) SDS-PAGE analysis of IMAC purification fractions of InIJ (1-6) and InIJ-LRR-IR (7-11) (flow through 1+7, wash 2, eluates 3-6 (InIJ) and 8-11 (InIJ-LRR-IR)).

In order to achieve higher purity, IMAC elution fractions of InIJ and InIJ-LRR-IR were pooled, dialysed into PBS and further purified by SEC (see 2.3.1.3). Major elution peaks showed at elution volumes of 56.1 mL and 60.3 mL corresponding to an apparent MW of about 261 and 192 kDa for InIJ and InIJ-LRR-IR, respectively (Figure 6.8 A and B). The observed MW of recombinant InIJ and InIJ-LRR-IR is about 3-fold or 4-fold higher than the theoretical MW, respectively, perhaps indicating oligomerisation in PBS buffer. SDS-PAGE analysis of SEC elution fractions showed sufficient sample homogeneity for use in binding studies (Figure 6.8 C).



**Figure 6.8 Purification of InIJ and InIJ-LRR-IR by SEC**

SEC elution profiles of (A) InIJ and (B) InIJ-LRR-IR. (C) SDS-PAGE analysis of SEC elution fractions (as indicated in A and B, main target protein elution peak indicated by red box) (InIJ-LRR-IR: 1 IMAC eluate, 2-5 elution fractions; InIJ: 6 IMAC eluate, 7-12 elution fractions).

#### 6.4 Assessment of Lar0958 and internalin binding to mucin

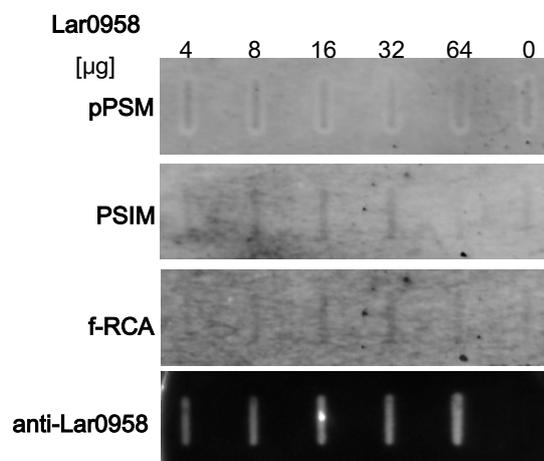
In order to investigate the binding of Lar0958 to mucus and mucin *in vitro* compared to InIJ and InIJ-LRR-IR, slot-blot and microtitre plate assays were performed.

Human MUC2 of the same source as used for the binding studies mentioned above (kindly provided by Michael McGuckin, Mater Medical Research Institute, South Brisbane, Australia), pPGM and PSIM (see 2.1.2) were used as a source of mucin in the performed binding experiments (see 6.3).

For slot-blot assays, purified Lar0958 (0-64  $\mu$ g) was immobilised onto a nitrocellulose membrane and then probed with pPSM and PSIM in excess (see 2.3.6.4). Bound mucin was detected via f-RCA, which recognises Gal residues, commonly found in

mucin *O*-glycans (see 2.3.5) [114, 459]. The reactivity of pPSM and PSIM with f-RCA was confirmed in a membrane test assay (data not shown).

Lar0958 did not bind to pPSM and PSIM when compared to the negative PBS control or to the f-RCA control in the absence of mucin, suggesting that the adhesin does not recognise porcine mucins under the applied assay conditions (Figure 6.9). The presence of Lar0958 was confirmed by primary anti-Lar0958 and secondary anti-rabbit-AP.



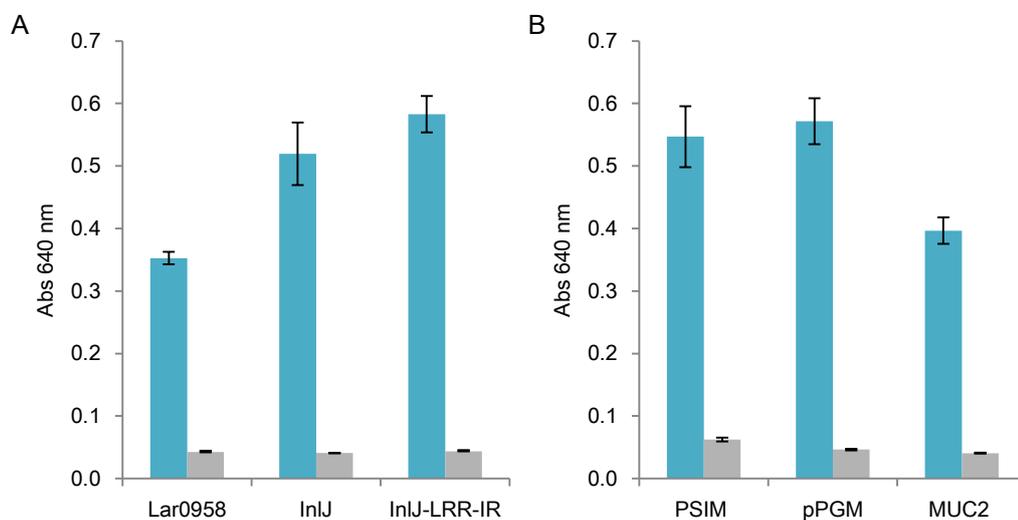
**Figure 6.9 Binding of Lar0958 to pPSM and PSIM assessed in a slot blot assay**

Slot-blot assay of Lar0958 probed with pPSM and PSIM and bound mucin detected by f-RCA. Slot-blot membrane probed with f-RCA in absence of mucin and with by anti-Lar0958 and secondary anti-rabbit-HRP as controls.

In order to investigate the binding of Lar0958, InIJ and InIJ-LRR-IR to human MUC2, PSIM and pPSM, a microtitre plate assay was performed following that described by Linden and co-workers, with modifications, used to demonstrate binding of internalins to MUC2 [460].

Briefly, adhesins (0.4 µg) were coated on microtitre plates and the presence of Lar0958 or His-tagged InIJ and InIJ-LRR-IR was verified using primary anti-Lar0958 or anti-His<sub>5</sub> and secondary anti-rabbit-HRP or anti-mouse-HRP, respectively (Figure 6.10 A). Biotinylated WGA (b-WGA) and ExtrAvidin-peroxidase proved to be most suitable for the detection of MUC2, pPGM and PSIM. An anti-MUC2C3 (kindly provided by Gunnar Hansson, University of Gothenburg, Sweden), anti-Muc2.3

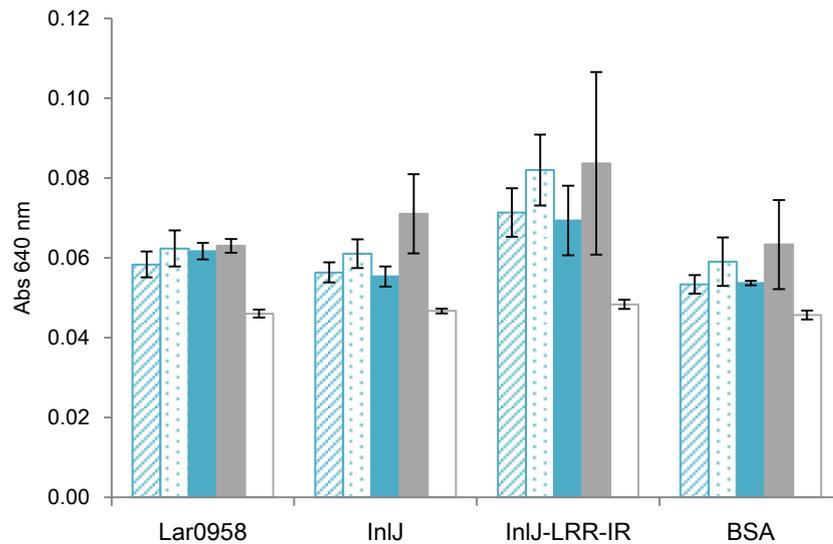
(kindly provided by Michael McGuckin, Mater Medical Research Institute, South Brisbane, Australia), as well as two commercially available anti-Muc2 antibodies, used with secondary HRP-conjugated antibodies, showed no reactivity with human MUC2 or demonstrated high background binding to target adhesins (see 2.1.3) (data not shown). To determine the optimum mucin concentration, microtitre plates were coated with varying amounts of PSIM (1.6-8  $\mu\text{g}$ ), pPGM (0.4 ng-1  $\mu\text{g}$ ) and human MUC2 (0.3-0.6  $\mu\text{g}$ ) (data not shown). Sufficient signal intensities were observed for 1.6  $\mu\text{g}$  PSIM, 0.4 ng pPGM and 0.6  $\mu\text{g}$  human MUC2 (Figure 6.10 B).



**Figure 6.10 Detection of Lar0958, InIJ and InIJ-LRR-IR via antibodies and of mucin via WGA**

(A) Immobilised Lar0958, InIJ and InIJ-LRR-IR incubated with via specific primary antibodies and secondary HRP antibody conjugates (blue). (B) Immobilised PSIM (1.6  $\mu\text{g}$ ), pPGM (0.4 ng) and human MUC2 (0.6  $\mu\text{g}$ ) (blue) probed with b-WGA and ExtrAvidin-peroxidase. (A+B) Adhesins or mucins incubated with secondary antibodies or ExtrAvidin-peroxidase as control (grey). Error bars represent standard deviation from mean.

In summary, in order to assess the binding of Lar0958, InIJ and InIJ-LRR-IR to mucin, adhesins were coated onto micro titre plates along with BSA as a control, probed with PSIM, pPGM and MUC2, and bound mucin detected via b-WGA and ExtrAvidin-peroxidase.



**Figure 6.11 Binding of Lar0958, InIJ and InIJ-LRR-IR to mucin or mucus**

Immobilised PSIM (blue stripes), pPGM (blue dots) and human MUC2 (blue) incubated on Lar0958, InIJ and InIJ-LRR-IR and BSA with bound mucin detected by b-WGA and ExtrAvidin-peroxidase. Adhesins incubated with B-WGA and ExtrAvidin-peroxidase (grey), or ExtrAvidin-peroxidase alone (white) as controls. Error bars represent standard deviation from mean.

Lar0958, InIJ and InIJ-LRR-IR did not show increased binding to PSIM, pPGM or MUC2 compared to the BSA control (Figure 6.11). As described before, BSA may interfere with binding when studying carbohydrate-protein interactions (see 5.1.1) [406]. In addition, signal intensities for mucins incubated on the immobilised adhesins and BSA were similar for those of b-WGA and ExtrAvidin-peroxidase alone in the absence of mucin. The interaction of b-WGA seems to vary slightly between Lar0958, InIJ, InIJ-LRR-IR and BSA, whereas the background signal for ExtrAvidin-peroxidase was comparably low for all tested proteins.

## 6.5 Discussion

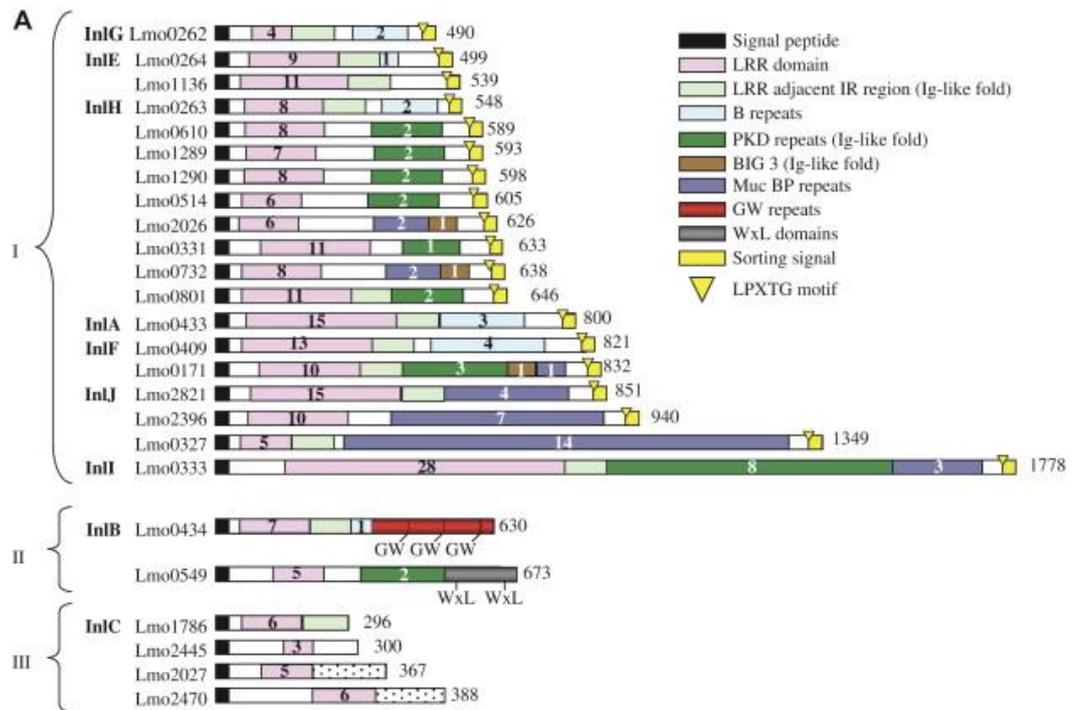
The Lar0958 protein of the human isolate strains *L. reuteri* MM4-1a, DSM 20016<sup>T</sup> and JCM 112<sup>T</sup> is a newly-identified cell-surface adhesion molecule, that shares a similar domain organisation with other Gram-positive cell surface adhesins (see 1.4.2). It is a modular protein covalently attached to the bacterial cell wall via an LPXTG-anchor, but it is additionally found in the bacterial culture supernatant [324]. The Lar0958 protein of *L. reuteri* MM4-1a and DMS 20016<sup>T</sup> seems to be involved in the adhesion of bacterial cells to mucus [302, 324], but no structural or functional characterisation of the Lar0958 protein has been conducted so far.

In order to obtain structural information on the novel Lar0958 cell-surface adhesin and to identify potential ligands, a single Lar0958 protein repeat was cloned, heterologously expressed and purified (see 6.1). The purified Lar0958 repeat protein forms a fully-folded, independent protein domain with a higher  $\alpha$ -helix content than the Mub-repeat proteins as demonstrated by CD (see 3.2 and 6.1). Like recombinant Mub repeats, recombinant Lar0958 has a negative overall charge at neutral pH in standard PBS used for functional studies.

The X-ray crystal structure of the Lar0958 protein was determined at 1.5 Å resolution revealing a ubiquitin-like  $\beta$ -grasp fold, showing some level of similarity to the B1 and B2 domains of MubR5 and MubRV, respectively (see 6.2). Lar0958 showed generally low structural similarity to structures found in the PDB, apart from internalin proteins, especially InIC.

Internalin proteins are cell surface proteins of the Gram-positive food borne pathogen *L. monocytogenes* [461]. To date, 25 members of the internalin protein family have been identified in *Listeria*, which are either covalently anchored to the cell wall (InIA, InIJ), non-covalently attached (InIB) or secreted (InIC) [462-463]. They show modular domain organisation consisting of an N-terminal cap structure, an LRR-domain, and an Ig-like IR-domain. In addition, some internalins possess additional domains containing a  $\beta$ -grasp fold, such as B-repeats (InIA or InIB) or MucBP repeats (InIJ), which show structural similarity to MucBP domains as defined in the Pfam database (PF06458) [321, 453] (Figure 6.12). The  $\beta$ -grasp fold is a widespread fold involved in a variety of physiological functions [378].

Lar0958 is structurally similar to the IR domains of InIC, InIA and InIB, but not to the B-repeat of InIB. The level of structural homology between Mub repeats, Lar0958 repeats and Internalin protein domains may indicate a related function of these protein domains.



**Figure 6.12 Domain organisation of members of the internalin protein family**

Internalin protein classification in three families of LPXTG-anchored (I), surface associated (via GW- or WxL module) (II) and secreted (III) internalins. Homologous domains are presented in the same colour and the number of repeats indicated (taken from [462]).

Functional characterisation of internalin protein domains has focussed so far on the characteristic LRR domains, mainly of InIA and InIB, which interact with the host cell receptors E-cadherin and Met, respectively [455, 457]. The IR domain is believed to be important for stability of the LRR domain, as the first N-terminal  $\beta$ -strand of the IR domain extends into the LRR domain [321]. The recombinant InIB-LRR protein lacking the IR domain shows significantly reduced stability compared to the InIB-LRR-IR protein [464]. Additionally, there is some evidence in the literature that the IR domain as well as the B-repeat fulfil an additional functional role in infection. The Ig-like IR domain and the LRR domain of InIA are necessary for host infection [465]. The InIB B-repeat seems to recognise another, unidentified host cell receptor different

to the one bound by its LRR domain [453, 466]. The GW modules present only in InIB interact with GAGs on the host cell surface, which is important for bacterial invasion [467]. Interestingly, Linden and co-workers demonstrated binding of several internalins including InIB and InIJ to human MUC2 *in vitro* [460]. However, the functional characterisation of the binding of internalins to mucins remains elusive.

Based on the functional similarities of the recombinant Lar0958 repeat to internalin protein domains, the binding ability of Lar0958 to pPGM, PSIM and human MUC2 was investigated in comparison to InIJ and InIJ-LRR-IR. No increased binding to mucin was observed for neither the recombinant single Lar0958 repeat nor the InIJ and InIJ-LRR-IR protein compared to BSA (see 6.4), which may indicate that a single repeat is not sufficient for ligand interaction, as previously observed in MUB protein binding studies (see 5.1 and 5.2).

To address the question, whether the presence of multiple Lar0958 repeats may be necessary for ligand binding, we purified the full-length native Lar0958 from the culture media of *L. reuteri* MM4-1a [324]. However, the low recovery of the native protein prevented its use in binding studies. Alternatively, for future functional characterisation of Lar0958, multi-protein repeats will need to be cloned and expressed in recombinant form for binding studies including GI mucin and sugar Au-arrays (see 5.1.3 and 5.2.2) [409].

## CHAPTER 7 GENERAL DISCUSSION AND FUTURE PROSPECTIVES

Despite recent advances in the identification of commensal mucus-adhesins enabled by genome sequencing of *Lactobacillus* species (see 1.4.2) [280], there is a current lack of knowledge on the structural and functional characterisation of these adhesins thought to mediate adherence of commensal bacteria to mucus and mucin (see 1.4.2). Lactobacilli, which are autochthonous members of the microbiota in different vertebrate hosts and commonly encountered in food produces, play a critical role in maintaining the healthy state in the gut [162, 175].

This study provides first insights into the low and high resolution structural organisation of a commensal mucus-adhesin, the modular-repeat containing MUB protein of *L. reuteri* ATCC 53608, and additionally reveals structural similarities to Gram-positive pathogenic adhesins. The Mub type 1 repeat MubRV folds to form an elongated structure of two distinct domains, an N-terminal ubiquitin-like  $\beta$ -grasp fold domain (B1) and a C-terminal MucBP domain (B2), as determined by X-ray crystallography (see 4.2.2). The MubRV structure is highly similar to the rod-like structure of the Mub type 2 repeat MubR5, showing only limited flexibility around the inter domain region, suggesting a similar protein fold between Mub repeats as relatively rigid building blocks of the cell-surface MUB protein (see 4.2.2). However, the MubRV structure lacks the  $\text{Ca}^{2+}$ -ion, which stabilises the loop conformation preceding strand  $\beta_4$  in the N-terminal domain of MubR5 (see 4.4.2). Whether calcium is required for the folding of some Mub repeats or is also important for ligand binding, as for example observed in C-type lectins, needs to be further investigated [468-469]. Extensive structural similarities of MubRV to a number of pilins and MSCRAMM including GBS52 of *Strep. agalactiae*, BcpA of *B. cereus*, RrgB of *Strep. pneumoniae*, Spy0128 of *Strep. pyogenes* and SpaA of *C. diphtheriae* has been demonstrated (see 4.2.2). They share a common Ig-like fold designated as an IgG-rev and frequently encountered in pili and MSCRAMM. The IgG-rev fold is reminiscent of the Ig-superfamily IgG fold (or Ig-like fold type C) with an ABED/CFG topology, however it shows a reverse strand arrangement in two alternative topologies (CBE/DAGF and CBEFT/DAG) (see 4.4). Despite the structural similarity of the B2 domain of MubRV to Ig-like pilin domains, it shows a different BAFD/EC topology (see 4.2.2). Pilins are components of long,

filamentous pili structures present on the bacterial cell surface, which are abundant in pathogens but rather rare in *Lactobacillus* species (see 4.4). SAXS experiments with single and tandem Mub repeats demonstrated their presence as extended structures with limited flexibility in solution, as observed in MubRV and MubR5 crystal structures (described above). These findings suggest the arrangement of Mub repeats as 'beads on a string' to form a relatively rigid fibre-like MUB protruding from the bacterial cell surface and capped by a differentially shaped Nterm region at the tip (see 4.3). An elongated appearance of native MUB was also indicated by sedimentation velocity AUC experiments (see 3.3.2). The organisation of Mub repeats along the length of the MUB fibre potentially enhances their ability to interact with multiple ligands for higher avidity binding. A similar mechanism of mucin binding may be employed by the *L. rhamnosus* GG pili, since the binding receptor SpaC, which contains two copies of a CnaB-type domain, is located along the pilus shaft [470]. Whether the Nterm region is involved in ligand interaction, remains to be demonstrated, however no mucin binding was observed in preliminary binding experiments (see 5.1.1). This is in contrast with the mode of interaction employed by bacterial flagella and pili, as well as most siglecs, where binding is mediated by the N-terminal tip domain (see 5.3 and 1.4.1). In addition to the relatively comprehensive structural information presented here on MUB, we provided the first structural information on the newly identified modular mucus-adhesin Lar0958 present on the cell-surface of *L. reuteri* ATCC 53608. In line with observations made for MUB, the X-ray crystal structure of a single Lar0958 repeat shows structural similarity to pathogenic cell-surface proteins from *Listeria* involved in host cell attachment and invasion (see 6.2 and 1.5). These findings further highlight structural similarities between *L. reuteri* adhesins, containing tandemly-arranged repeats, and filamentous, pathogenic adhesion molecules with similar modular organisation, which may be indicative of a similar mode of ligand binding and specificity.

We investigated the binding ability of recombinant Mub repeats and native MUB to mucin and mucin glycans *in vitro*, and demonstrated that mucin binding seems to be mediated via sugar recognition (see 5.1.2 and 5.1.3). Additionally, we showed binding of MUB proteins to several glycan structures and blood-group epitopes (see 5.2). More work is required to biochemically evaluate and characterise MUB protein interactions with potential sugar ligands including Au-array inhibition studies, SPR, ITC and

co-crystallisation experiments. Other commensal adhesins that recognise glycans or ABO blood group antigens, are Msa of *L. acidophilus* NCFM, which contains MucBP domains, GAPDH of *L. plantarum* LA 318 and Lam29 of *L. mucosae* ME-380 (see 1.4.2). Additionally, sugar binding has been reported for a number of pathogenic bacterial adhesins, for example BabA and Sab of *H. pylori*, the MucBP domain of Spr1345 from *Strep. pneumoniae* as well as *E. coli* pilins and GspB of *Strep. gordonii*, the last two containing Ig-like protein domains found in Gram-positive adhesins (as described above) (see 1.4.1) [384]. The MucBP domain of Spr1345, which has additionally been demonstrated to bind different mucins, shows structural similarity to the B2 domains of MubR5 and MubRV (see 1.4.1), while their B1 domains are structurally similar to the Ig-binding protein L (see 4.1.2) [282]. These findings indicate a dual receptor specificity of MUB to glycan and protein ligands (such as IgA) present in the mucus layer (see 1.1.2 and 1.2), as suggested by mucin adhesion assays (see 5.1.2 and 5.3) and demonstrated earlier in our lab [302]. Mucin binding studies further suggested a multivalency binding of Mub repeats, as observed in FnBPs from Gram-positive bacteria (see 5.3) [416, 471]. Multivalent interactions may potentiate MUB-dependent binding of bacteria to mucin glycans or proteins in the GI mucus and increase the avidity of the binding enabled by the elongated fiber-like structural organisation of MUB, similarly reported for the *L. rhamnosus* GG pili (as described above). Furthermore, the cross-linking of two MubR5 molecules by Man, which we showed using X-crystallography, suggests a role of MUB in bacterial aggregation as demonstrated in our lab (see 4.1) [302], and may yet be another mechanism by which sugar-adhesin interactions facilitate lactobacilli mucus adhesion [394]. These findings provide insights into mechanisms by which gut bacteria may achieve pathogen exclusion, i.e. via competition for mucosal attachment sites, and facilitate persistence in the gut (see 1.3.1). This knowledge may be used to design strategies for the targeted selection of suitable commensal lactobacilli to be used as probiotics.

Here we report the potential glycosylation of native MUB isolated from a *L. reuteri* ATCC 53608 culture (see 3.3.2) as shown via lectin staining (see 3.3.3.2). A few glycoproteins have been identified in commensal bacteria including *Lactobacillus* or *Bacteroides* species (see 3.4) [472], demonstrating that protein glycosylation is not restricted to pathogenic organisms such as *Pseudomonas*, *Clostridium*, *Campylobacter* and *Helicobacter* species, where glycosylation is often found in flagella

or pili structures [365-366]. For example, the extracellular autolysin protein Acm2 from *L. plantarum*, or the major secreted protein Msp1/p75 from *L. rhamnosus*, which has been suggested to be involved in intestinal epithelial cell signalling, have been demonstrated to contain GlcNAc or Man residues, respectively. Lectin staining of MUB (with WGA) indicated the presence of GlcNAc molecules (see 3.3.3.2) [367, 473]. Further work is required, to determine the glycan profile of MUB by MS glycopeptide analysis and metabolic glycoprotein labeling. Probiotic cell-surface molecules may contribute to the maturation of the host immune system. For example, *L. rhamnosus* pili have been shown to possess immunomodulating properties in intestinal epithelial cells [474]. Glycosylation of lactobacilli cell-surface proteins such as MUB may yet be another mechanism contributing to the overall health benefit of commensals on the host immune system to establish homeostasis in the gut [138, 475]. However, further work is required to investigate potential immunomodulating effects of MUB and the role of glycans in this process including the *in vitro* investigation of the expression levels of pro- and anti-inflammatory markers after treatment of intestinal epithelial cells with glycosylated and de-glycosylated MUB. This finding indicates a dual function of MUB in mediating bacterial adhesion to mucins supporting its role as an important probiotic factor for *Lactobacillus* host colonisation (see 1.3.1).

Further investigations, as proposed here, as well as *in vitro* experiments including the application of MUB-expressing and –non-expressing *L. reuteri* cells to intestinal epithelial cells with intact or aberrant mucin glycosylation, will help us to identify relevant probiotic features important to maintain or promote a beneficial relationship with our gut microbiota.

## APPENDIX

### I. LDMII media composition

Ingredients	Amount per L
K <sub>2</sub> HPO <sub>4</sub> (anhydrous)	1.5 g
KH <sub>2</sub> PO <sub>4</sub> (anhydrous)	1.5 g
Sodium acetate	15 g
Sodium citrate	0.22 g
Tryptophan	50 mg
Asparagine	50 mg
Cysteine	50 mg
Glycine	50 mg
Serine	50 mg
Alanine	50 mg
Phenylalanine	50 mg
Histidine	50 mg
Isoleucine	50 mg
Methionine	50 mg
Proline	50 mg
Threonine	50 mg
Valine	50 mg
Tyrosine	50 mg
Leucine	50 mg
Glutamine	50 mg
Aspartic acid	50 mg
Glutamic acid	50 mg
Thiamine-HCl	0.2 mg
<i>para</i> -Aminobenzoic acid	0.04 mg
Calcium pantothenic acid	0.4 mg
Niacin	1.0 mg
Pyridoxine-HCl	0.5 mg
Biotin	0.05 mg
Folic acid	0.1 mg
Riboflavin	0.4 mg
Adenine sulphate	10 mg
Uracil	20 mg
Guanine-HCl	10 mg
Cytidine (acid)	50 mg
Thymidine	1.6 µg
Tween-80	1.0 ml
MgSO <sub>4</sub> H <sub>2</sub> O	0.163 g
MnSO <sub>4</sub> 7H <sub>2</sub> O	23.4 mg
FeSO <sub>4</sub> 7H <sub>2</sub> O	13 mg
Sucrose	30 g
Optional amino acids (arginine and lysine)	50 mg (each)

## II. Primers for amplification and sequencing

Protein	Primer	Sequence	Nucleotide alignment position in target gene <sup>a/b</sup>	Amino acid residue limits (size in amino acids)
MubRV	RV_for	5'- <b>ATG</b> CAAACAGCCTACGTCAAG-3'	8515-9063 (& 7963-7980)	2799-2981 (183)
	RV_rev	5'- <b>TTAGGG</b> ATCACCAACATAAACGA-3'	9044-9063	
MubR8-V	RV_for	5'- <b>ATG</b> CAAACAGCCTACGTCAAG-3'	7963-7980 (& 8515-9063)	2615-2798 (367)
	RV_rev	5'- <b>TTAGGG</b> ATCACCAACATAAACGA-3'	9044-9063	
MubRV-VI	RV_for	5'- <b>ATG</b> CAAACAGCCTACGTCAAG-3'	8515-9063 (& 7963-7980)	2799-3187 (389)
	RVI_rev	5'- <b>TTA</b> ATCAAGCTTCTTGTAGGT-3'	9664-9681	
MubR5	MucB1-R4F	5'- <b>ATG</b> CAAAAGGTTACGTTCAA-3'	6307-6324 & 6859-6876	2063-2246 (184)
	MucB2-R4R	5'- <b>TTAGGC</b> ATCAGCCGTGTAGA-3'	6842-6858 & 7394-7410	
MubRI	MucB1-RIF	5'- <b>ATG</b> CAAGAAGCTGCCATCAG-3'	1789-1805	557-749 (193)
	MucB2-RIR	5'- <b>TA</b> CGTGT CAGCGGTATAGT-3'	2351-2367	
MubRI-II-III	MucB1-RIF	5'- <b>ATG</b> CAAGAAGCTGCCATCAG-3'	1789-1805	557-1134 (578)
	MucB2-RIIIR	5'- <b>TTAGGC</b> ATTCCCAACATAAAC-3'	3505-3522	
Nterm	NtermMUB-pOPINF-for	5'-AAGTTCTGTTTCAGGGCCCCGGCAACTACTGAATCG-3'	241-287	50-556 (507)
	NtermMUB-pOPINF-rev	5'-ATGGTCTAGAAAGCTTTACGTGTCAGCTACATAAAC-3'	1768-1788	without His <sub>6</sub> -tag
NtermRI	NtermMUB-pOPINF-for	5'-AAGTTCTGTTTCAGGGCCCCGGCAACTACTGAATCG-3'	241-287	50-749 (700) without His <sub>6</sub> -tag

---

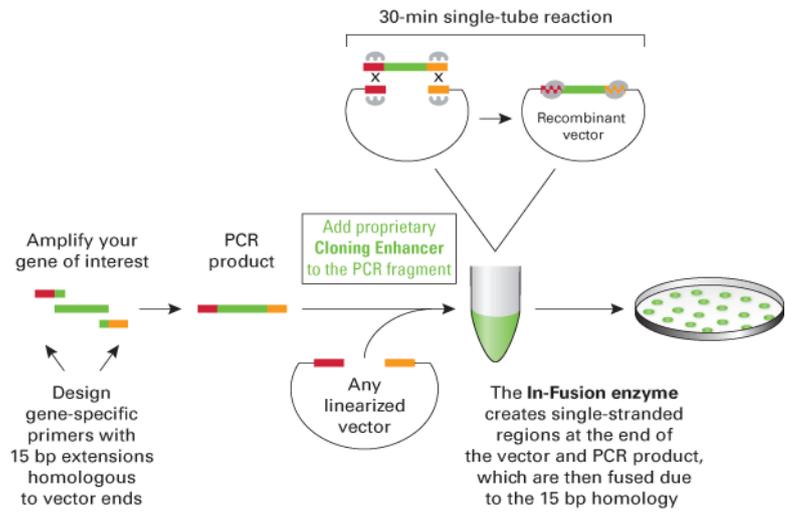
	MubRI-pOPINF- rev	5'-ATGGTCTAGAAAGCTTTACGTGTCAGCGGTATAGT-3'	2351-2347	
	pOPINFrev (sequencing)	5'-TCAGATGCTCAAGGGGCTT-3'		
	midNterm-1 (sequencing)	5'-GCTGCTGATGTAGAAACAGT-3'	592-612	
	midNterm-2 (sequencing)	5'-CAAAAAGGTCTGGATGGTCA-3'	1189-1208	
Lar0958	Lar0958for	5'- <b>ATG</b> AAAGTTACCTATAGTGGTAG-3'	1729-1748, 2017-2036, 2305- 2324 & 2593-2612	577-672, 675-768, 769- 864 or 865-960 (95)
	Lar0958rev	5'- <b>TTA</b> ATCAATCCCTAGTGGATT-3'	1998-2016, 2286-2244, 2574- 2592 & 2862-2880	

---

<sup>a</sup> Nucleotide alignment position in target gene *mub* based on DNA sequence from GenBank accession number AF120104

<sup>b</sup> Nucleotide alignment position in target gene *lar\_0958* based on DNA sequence from GenBank accession number BAG25474

### III. In-Fusion cloning system



#### IV. Alignment of Mub repeat sequences using ClustaWI

```

R1      QHAVINYIDGESDEILHT--DKVNGHSDEKINYST--ADMIKQLEAKGYE
R2      QKVHVQYIDGETDQMLRQ--DDL DGYTDETIPIYST--AEGIKKFEGDGYE
R3      QKVHVQYIDGETDQMLRQ--DDL DGYTDETIPIYST--AEGIKKFEGDGYE
R4      QKVHVQYIDGETDQMLRQ--DDL DGYTDETIPIYST--AEGIKKFEGDGYE
R5      QKVHVQYIDGETDQMLRQ--DDL DGYTDETIPIYST--AEGIKKFEGDGYE
R6      QKVHVQYIDGETDQMLRQ--DDL DGYTDETIPIYST--AEGIKKFEGDGYE
R7      QKVHVQYIDGETDQMLRQ--DDL DGYTDETIPIYST--AEGIKKFEGDGYE
R8      QTAYVKYVDDTTGETLRQ--DDL HGYTDETIPIYST--AEGIKKFEGDGYE
R.I     QEAAISFYDETDHKPLNDQTIQLTGKTGEKISHTEA-NQTLAKLKGQGYV
R.II    QEAAINFYDETGHKLLDNQTIHLTGKTGEKVDRQA-DQTLADLVKQGYV
R.III   QEAAINFYDETGHKLLDNQTIHLTGKTGEKVDRQA-DQTLAELEKQGYV
R.IV    QEAAQIFYDETTGKEISGTREIATGKDEETISFTKDPNEVVKELEKQGYV
R.V     QTAYVKYVDDTTGETLRQD--DLHGYTDETIPIYSTA-E-GIKKYEKQGYV
R.VI    QEAAQIFYDETTGKEISNTREIVNGKTDETIGFTKDPNEVVKELEKQGYV
* . : * : : * :.*: : : . :.**

R1      LFKDN-----FPAGEKFDNDDTNDQFYTVIFKHHRENVDPNHSSA-----
R2      LFKDN-----FPAGEKFDNDDKNDQTYTVIFKHHRENVDPNHSSA-----
R3      LFKDN-----FPAGEKFDNDDTNDQFYTVIFKHHRENVDPNHSSA-----
R5      LFKDN-----FPAGEKFDNDDTNDQFYTVIFKHHRENVDPNHSSA-----
R6      LFKDN-----FPAGEKFDNDDKNDQTYTVIFKHHRENVDPNHSSA-----
R7      LFKDN-----FPAGEKFDNDDKTDQTYTVIFKHHRENVDPNHSSA-----
R8      LFKDN-----FPAGEKFDNDDKTDQTYTVIFKHHRENVDPNHSSA-----
R.I     VDQNT-----FADDATYDNDTQAPQEFTIYLKHDTHHTDATSSKA-----
R.II    LDKENTA-KAFPADAVYDNDQTPQEFTIYLKHGTHHTDATSSKA-----
R.IIII  LDENNTK-LGFPSNAAYDDDDVKPQEFTIYLKHGMTHHTDATDKNA-----
R.IV    FDKDNAKNNVFAAGTAYDNKSEVHQYFKYYLKHGHATVTPDQ-----
R.V     LVSDG-----FKPGTKFVGV---TPTYEVHFKHGMTHHTDATDKNA-----
R.VI    FDKDNANNVFAAGTTYDKNSEVHQYFKYYFTHATTIVTPDNPKTPADVL
. . : * . : . : : : :.* . .

R1      -----DGTKGKTLTETVHYKYANGTKA----AEDQTAQVTF
R2      -----DGTKGKTLTETVHYKYADGKA----AEDQTAQVTF
R3      -----DGTKGKTLTETVHYKYANGTKA----AEDQTAQVTF
R4      -----DGTKGKTLTETVHYKYADGKA----AEDQTAQVTF
R5      -----DGTKGKTLTETVHYKYANGTKA----AEDQTAQVTF
R6      -----DGTKGKTLTETVHYKYADGKA----AEDQTAQVTF
R7      -----DGTKGKTLTETVHYKYADGKA----AEDQTAQVTF
R8      -----DGTKGKTLTETVHYKYADGKA----AEDQTAQVTF
R.I     -----D----QKTVSETHYVYKDVNANKPVADDANTTVTFK
R.II    -----D----QKTVSETHYVYKDVNANKPVADDANTTVTFK
R.IIII  -----E----QKIVTETHYVYENNTA----KTDYTSVDFK
R.IV    -----DPQKGQKTVTQTIKYEYADGTAT---GLADNVQTLTFK
R.V     -----E----QKTVTETHYVDENNQTV---QPDSTTAVTFK
R.VI    PDNPGKNYPSGVAKDDL NKTVTRTINITTPDGKTQ---T---ITQKAEFT
* : :.*: : : . :

R1      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYAPSEAVVKR
R2      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYTPSEAVVKR
R3      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYAPSEAVVKR
R4      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYTPSEAVVKR
R5      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYAPSEAVVKR
R6      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYTPSEAVVKR
R7      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYTPSEAVVKR
R8      RNGVLDDVTG-IVAWGKWNE-----ASQSYKALTSPTIAGYTPSEAVVKR
R.I     RGYTTDKVTGKIVSYDPWTVDGKQADSKTFDAVKSPVIAGYADQAEVAA
R.II    RGYTTDKVTGKIVSYDPWTVDGKQADSKTFDAVKSPVIAGYADQAEVAA
R.IIII  RGYTTDNVTHKIIISYDPMWV-----SSKKFGFVKSPAIEGYTPNHSQIDE
R.IV    RTGDKDLVTH-EVTPDWST----VAGQQTSVVTPALKGYTADTNEIPA
R.V     RGYTTDNVTGKVVSYDPWTVDGNQADSKTFAAVPSPAVEGYTPNHSQINE
R.VI    RSATVDEVTG-EVTYGPWSK-----N-----
* * ** : : *

R1      SSNSDAEQGPTLTVIYTADA
R2      SSNSDAEQGPTLTVIYTADA
R3      SSNSDAEQGPTLTVIYTADA
R4      SSNSDAEQGPTLTVIYTADA
R5      SSNSDAEQGPTLTVIYTADA
R6      SSNSDAEQGPTLTVIYTADA
R7      SSNSDAEQGPTLTVIYTADA
R8      SSNSDAEQGPTLTVIYTADA
R.I     QTVTPDSQNINKTVYYTADT
R.II    QTVTPDSQNINKTVYYTADT
R.IIII  ITVTPDSKDVVKTVVYVGNV
R.IV    ITYHAGDSVTVYVVKYNADV
R.V     FTVTPDSKDIVKTVVYVGNV
R.VI    -----VV

```

	Alignment score													
Mub repeat sequence	R1	R2	R3	R4	R5	R6	R7	R8	RI	RII	RIII	RIV	RV	RVI
R1	100	85.3	87.5	85.3	87.5	85.3	84.8	83.7	32.1	38.1	31.5	30.4	35.0	23.9
R2	85.3	100	97.8	100	97.8	100	99.5	93.5	33.7	31.5	25.0	32.6	42.6	26.6
R3	87.5	97.8	100	97.8	100	97.8	97.3	91.3	32.6	30.4	25.0	31.5	42.1	23.9
R4	85.3	100	97.8	100	97.8	100	99.5	93.5	33.7	31.5	25.0	32.6	42.6	26.6
R5	87.5	97.8	100	97.8	100	97.8	97.3	91.3	32.6	30.4	25.0	31.5	42.1	23.9
R6	85.3	100	97.8	100	97.8	100	99.5	93.5	33.7	31.5	25.0	32.6	42.6	26.6
R7	84.8	99.5	97.3	99.5	97.3	99.5	100	94.0	33.7	32.1	25.0	32.6	42.6	26.6
R8	83.7	93.5	91.3	93.5	91.3	93.5	94.0	100	36.4	32.1	28.3	35.9	48.6	29.9
RI	32.1	33.7	32.6	33.7	32.6	33.7	33.7	36.4	100	87.6	59.6	39.6	49.2	32.6
RII	38.1	31.5	30.4	31.5	30.4	31.5	32.1	32.1	87.6	100	67.6	40.1	42.6	32.0
RIII	31.5	25.0	25.0	25.0	25.0	25.0	25.0	28.3	59.6	67.6	100	38.8	54.1	31.4
RIV	30.4	32.6	31.5	32.6	31.5	32.6	32.6	35.9	39.6	40.1	38.8	100	29.0	55.7
RV	35.0	42.6	42.1	42.6	42.1	42.6	42.6	48.6	49.2	42.6	54.1	29.0	100	31.2
RVI	23.9	26.6	23.9	26.6	23.9	26.6	26.6	29.9	32.6	32.0	31.4	55.7	31.2	100

## V. Crystal structure screen conditions

### Structure Screen 1

(Molecular Dimensions, Newmarket, UK; April 2010)

ID	Salt	Buffer	pH	Precipitant
1	0.02 M calcium chloride	0.1 M sodium acetate	4.6	30 % v/v MPD
2	0.2 M ammonium acetate	0.1 M sodium acetate	4.6	30 % w/v PEG 4K
3	0.2 M ammonium sulfate	0.1 M sodium acetate	4.6	25 % w/v PEG 4K
4	None	0.1 M sodium acetate	4.6	2.0 M sodium formate
5	None	0.1 M sodium acetate	4.6	2.0 M ammonium sulfate
6	None	0.1 M sodium acetate	4.6	8 % w/v PEG 4K
7	0.2 M ammonium acetate	0.1 M tri-sodium citrate	5.6	30 % w/v PEG 4K
8	0.2 M ammonium acetate	0.1 M tri-sodium citrate	5.6	30 % v/v MPD
9	None	0.1 M tri-sodium citrate	5.6	20 % v/v 2-propanol, 20%w/v PEG 4K
10	None	0.1 M tri-sodium citrate	5.6	1.0 M ammonium dihydrogen phosphate
11	0.2 M calcium chloride	0.1 M sodium acetate	4.6	20 % v/v 2-propanol
12	None	0.1 M sodium cacodylate	6.5	1.4 M sodium acetate
13	0.2 M tri-sodium citrate	0.1 M sodium cacodylate	6.5	30 % v/v 2-propanol
14	0.2 M ammonium sulfate	0.1 M sodium cacodylate	6.5	30 % w/v PEG 8K
15	0.2 M magnesium acetate	0.1 M sodium cacodylate	6.5	20 % w/v PEG 8K
16	0.2 M magnesium acetate	0.1 M sodium cacodylate	6.5	30 % v/v MPD
17	None	0.1 M imidazole	6.5	1.0 M sodium acetate
18	0.2 M sodium acetate	0.1 M sodium cacodylate	6.5	30 % w/v PEG 8K
19	0.2 M zinc acetate	0.1 M sodium cacodylate	6.5	18 % w/v PEG 8K
20	0.2 M calcium acetate	0.1 M sodium cacodylate	6.5	18 % w/v PEG 8K
21	0.2 M tri-sodium citrate	0.1 M Na HEPES	7.5	30 % v/v MPD
22	0.2 M magnesium chloride	0.1 M Na HEPES	7.5	30 % v/v 2-propanol
23	0.2 M calcium chloride	0.1 M Na HEPES	7.5	28 % v/v PEG 400
24	0.2 M magnesium chloride	0.1 M Na HEPES	7.5	30 % v/v PEG 400
25	0.2 M tri-sodium citrate	0.1 M Na HEPES	7.5	20 % v/v 2-propanol
26	None	0.1 M Na HEPES	7.5	0.8 M K/Na tartrate
27	None	0.1 M Na HEPES	7.5	1.5 M lithium sulfate
28	None	0.1 M Na HEPES	7.5	0.8 M sodium dihydrogen phosphate/ 0.8 M K dihydrogen phosphate
29	None	0.1 M Na HEPES	7.5	1.4 M tri-sodium citrate
30	None	0.1 M Na HEPES	7.5	2 % v/v PEG 400 , 2.0 M ammonium sulfate
31	None	0.1 M Na HEPES	7.5	10 % v/v 2-propanol, 20% w/v PEG 4K
32	None	0.1 M Tris	8.5	2.0 M ammonium sulfate
33	0.2 M magnesium chloride	0.1 M Tris	8.5	30 % w/v PEG 4K
34	0.2 M tri-sodium citrate	0.1 M Tris	8.5	30 % v/v PEG 400
35	0.2 M lithium sulfate	0.1 M Tris	8.5	30 % w/v PEG 4K
36	0.2 M ammonium acetate	0.1 M Tris	8.5	30 % v/v 2-Propanol
37	0.2 M sodium acetate	0.1 M Tris	8.5	30 % w/v PEG 4K
38	None	0.1 M Tris	8.5	8 % w/v PEG 8K
39	None	0.1 M Tris	8.5	2.0 M ammonium dihydrogen phosphate
40	None	None	-	0.4 M K/Na Tartrate
41	None	None	-	0.4 M ammonium dihydrogen phosphate
42	0.2 M ammonium sulfate	None	-	30 % w/v PEG 8K
43	0.2 M ammonium sulfate	None	-	30 % w/v PEG 4K
44	None	None	-	2.0 M ammonium sulfate
45	None	None	-	4.0 M sodium formate
46	0.05 M potassium dihydrogen phosphate	None	-	20 % w/v PEG 8K
47	None	None	-	30 % w/v PEG 1.5K
48	None	None	-	0.2 M magnesium formate

**Structure Screen 2**

(Molecular Dimensions, Newmarket, UK; April 2010)

<b>ID</b>	<b>Salt</b>	<b>Buffer</b>	<b>pH</b>	<b>Precipitant</b>
1	0.1 M sodium chloride	0.1 M Bicine	9.0	30 % v/v PEG 550 MME
2	None	0.1 M Bicine	9.0	2.0 M magnesium chloride
3	None	0.1 M Bicine	9.0	2 % v/v 1,4-Dioxane/10 % w/v PEG 20,000
4	0.2 M magnesium chloride	0.1 M Tris	8.5	3.4 M 1,6-hexanediol
5	None	0.1 M Tris	8.5	25 % v/v tert-butanol
6	0.01 M nickel chloride	0.1 M Tris	8.5	1.0 M lithium sulfate
7	1.5 M ammonium sulfate	0.1 M Tris	8.5	12 % v/v glycerol
8	0.2 M ammonium dihydrogen phosphate	0.1 M Tris	8.5	50 % v/v MPD
9	None	0.1 M Tris	8.5	20 % v/v ethanol
10	0.01 M nickel chloride	0.1 M Tris	8.5	20 % w/v PEG 2000 MME
11	0.5 M ammonium sulfate	0.1 M Na HEPES	7.5	30 % v/v MPD
12	None	0.1 M Na HEPES	7.5	10 % w/v PEG 6000, 5% v/v MPD
13	None	0.1 M Na HEPES	7.5	20 % v/v Jeffamine M-600
14	0.1 M sodium chloride	0.1 M Na HEPES	7.5	1.6 M ammonium sulfate
15	None	0.1 M Na HEPES	7.5	2.0 M ammonium formate
16	0.05 M cadmium sulfate	0.1 M Na HEPES	7.5	1.0 M sodium acetate
17	None	0.1 M Na HEPES	7.5	70 % v/v MPD
18	None	0.1 M Na HEPES	7.5	4.3 M sodium chloride
19	None	0.1 M Na HEPES	7.5	10 % w/v PEG 8000, 8 % v/v ethylene glycol
20	None	0.1 M MES	6.5	1.6 M magnesium sulfate
21	0.1 M potassium phosphate + 0.1 M sodium phosphate	0.1 M MES	6.5	2.0 M sodium chloride
22	None	0.1 M MES	6.5	12 % w/v PEG 20,000
23	1.6 M ammonium sulfate	0.1 M MES	6.5	10 % v/v Dioxane
24	0.05 M caesium chloride	0.1 M MES	6.5	30 % v/v Jeffamine M-600
25	0.01 M cobalt chloride	0.1 M MES	6.5	1.8 M ammonium sulfate
26	0.2 M ammonium sulfate	0.1 M MES	6.5	30 % w/v PEG 5000 MME
27	0.01 M zinc sulfate	0.1 M MES	6.5	25 % v/v PEG 550 MME
28	None	0.1 M Na HEPES	7.5	20 % w/v PEG 10,000
29	0.2 M potassium sodium tartrate	0.1 M Na citrate	5.6	2.0 M ammonium sulfate
30	0.5 M ammonium sulfate	0.1 M Na citrate	5.6	1.0 M lithium sulfate
31	0.5 M sodium chloride	0.1 M Na citrate	5.6	4 % v/v polyethyleneimine
32	None	0.1 M Na citrate	5.6	35 % v/v tert-butanol
33	0.01 M ferric chloride	0.1 M Na citrate	5.6	10 % v/v Jeffamine M-600
34	0.01 M manganese chloride	0.1 M Na citrate	5.6	2.5 M 1,6-hexanediol
35	None	0.1 M Na acetate	4.6	2.0 M sodium chloride
36	0.2 M sodium chloride	0.1 M Na acetate	4.6	30 % v/v MPD
37	0.01 M cobalt chloride	0.1 M Na acetate	4.6	1.0 M 1,6-hexanediol
38	0.1 M cadmium chloride	0.1 M Na acetate	4.6	30 % v/v PEG 400
39	0.2 M ammonium sulfate	0.1 M Na acetate	4.6	30 % w/v PEG 2000 MME
40	2.0 M sodium chloride	None	None	10 % w/v PEG 6000
41	0.01 M CTAB	None	None	0.5 M sodium chloride, 0.1 M magnesium chloride
42	None	None	None	25 % v/v ethylene glycol
43	None	None	None	35 % v/v dioxane
44	2.0 M ammonium sulfate	None	None	5 % v/v 2-propanol
45	None	None	None	1.0 M imidazole pH 7.0
46	None	None	None	10 % w/v PEG 1000, 10 % w/v PEG 8000
47	1.5 M sodium chloride	None	None	10 % v/v ethanol
48	None	None	None	1.6 M sodium citrate pH 6.5

**JCSG-plus Screen 1**

(Molecular Dimensions, Newmarket, UK; April 2010)

<b>ID</b>	<b>Salt</b>	<b>Buffer</b>	<b>pH</b>	<b>Precipitant</b>
1	0.2 M lithium sulfate	0.1 M sodium acetate	4.5	50 % v/v PEG 400
2	None	0.1 M sodium citrate	5.5	20 % w/v PEG 3000
3	0.2 M di-ammonium hydrogen citrate	None	-	20 % w/v PEG 3350
4	0.02 M calcium chloride	0.1 M sodium acetate	4.6	30 % v/v MPD
5	0.2 M magnesium formate	None	-	20 % w/v PEG 3350
6	0.2 M lithium sulfate	0.1 M phosphate/citrate	4.2	20 % w/v PEG 1000
7	None	0.1 M CHES	9.5	20 % w/v PEG 8000
8	0.2 M ammonium formate	None	-	20 % w/v PEG 3350
9	0.2 M ammonium chloride	None	-	20 % w/v PEG 3350
10	0.2 M potassium formate	None	-	20 % w/v PEG 3350
11	0.2 M ammonium dihydrogen phosphate	0.1 M Tris	8.5	50 % v/v MPD
12	0.2 M potassium nitrate	None	-	20 % w/v PEG 3350
13	None	0.1 M citrate	4.0	0.8 M ammonium sulfate
14	0.2 M sodium thiocyanate	None	-	20 % w/v PEG 3350
15	None	0.1 M Bicine	9.0	20 % w/v PEG 6000
16	None	0.1 M HEPES	7.5	10 % w/v PEG 8000/ 8 % v/v Ethylene glycol
17	None	0.1 M sodium cacodylate	6.5	40 % v/v MPD/ 5 % w/v PEG 8000
18	None	0.1 M phosphate/citrate	4.2	40 % v/v Ethanol/ 5 % w/v PEG 1000
19	None	0.1 M sodium acetate	4.6	8 % w/v PEG 4000
20	0.2 M magnesium chloride	0.1 M Tris	7.0	10 % w/v PEG 8000
21	None	0.1 M citrate	5.0	20 % w/v PEG 6000
22	0.2 M magnesium chloride	0.1 M sodium cacodylate	6.5	50 % v/v PEG 200
23	None	None	6.5	1.6 M tri-sodium citrate
24	0.2 M tri-potassium citrate	None	-	20 % w/v PEG 3350
25	0.2 M sodium chloride	0.1 M phosphate/citrate	4.2	20 % w/v PEG 8000
26	1.0 M lithium chloride	0.1 M Na citrate	4.0	20 % w/v PEG 6000
27	0.2 M ammonium nitrate	None	-	20 % w/v PEG 3350
28	None	0.1 M Na HEPES	7.0	10 % w/v PEG 6000
29	None	0.1 M Na HEPES	7.5	0.8 M sodium dihydrogen phosphate 0.8 M potassium dihydrogen phosphate
30	None	0.1 M phosphate/citrate	4.2	40 % v/v PEG 300
31	0.2 M zinc acetate	0.1 M sodium acetate	4.5	10 % w/v PEG 3000
32	None	0.1 M Tris	8.5	20 % v/v Ethanol
33	None	0.1 M Na/K phosphate	6.2	25 % v/v 1,2-propanediol 10 % v/v Glycerol
34	None	0.1 M Bicine	9.0	10 % w/v PEG 20,000/ 2% v/v Dioxane
35	None	0.1 M sodium acetate	4.6	2.0 M ammonium sulfate
36	None	None	-	10 % w/v PEG 1000/ 10 % w/v PEG 8000
37	None	None	-	24 % w/v PEG 1500/ 20 % v/v Glycerol
38	0.2 M magnesium chloride	0.1 M Na HEPES	7.5	30 % v/v PEG 400
39	0.2 M sodium chloride	0.1 M Na/K phosphate	6.2	50 % v/v PEG 200
40	0.2 M lithium sulfate	0.1 M sodium acetate	4.5	30 % w/v PEG 8000
41	None	0.1 M HEPES	7.5	70 % v/v MPD
42	0.2 M magnesium chloride	0.1 M Tris	8.5	20 % w/v PEG 8000
43	0.2 M lithium sulfate	0.1 M Tris	8.5	40 % v/v PEG 400
44	None	0.1 M Tris	8.0	40 % v/v MPD
45	0.17 M ammonium sulfate	None	-	25.5 % w/v PEG 4000/ 15 % v/v Glycerol
46	0.2 M calcium acetate	0.1 M sodium cacodylate	6.5	40 % v/v PEG 300
47	0.14 M calcium chloride	0.07 M sodium acetate	4.6	14 % v/v 2-propanol/

48	0.04 M potassium dihydrogen phosphate	None	-	30 % v/v Glycerol 16 % w/v PEG 8000/ 20 % v/v
----	---------------------------------------	------	---	--

**JCSG-plus Screen 2** (Molecular Dimensions, Newmarket, UK; April 2010)

ID	Salt	Buffer	pH	Precipitant
1	None	0.1 M sodium cacodylate	6.5	1.0 M tri-sodium citrate
2	0.2 M sodium chloride	0.1 M sodium cacodylate	6.5	2.0 M ammonium sulfate
3	0.2 M sodium chloride	0.1 M HEPES	7.5	10 % v/v 2-propanol
4	0.2 M lithium sulfate	0.1 M Tris	8.5	1.26 M ammonium sulfate
5	None	0.1 M CAPS	10.5	40 % v/v MPD
6	0.2 M zinc acetate	0.1 M imidazole	8.0	20 % w/v PEG 3000
7	0.2 M zinc acetate	0.1 M sodium cacodylate	6.5	10 % v/v 2-propanol
8	None	0.1 M sodium acetate	4.5	1.0 M di-ammonium hydrogen phosphate
9	None	0.1 M MES	6.5	1.6 M magnesium sulfate
10	None	0.1 M Bicine	9.0	10 % w/v PEG 6000
11	0.16 M calcium acetate	0.08 M sodium cacodylate	6.5	14.4 % w/v PEG 8000/ 20 % v/v glycerol
12	None	0.1 M imidazole	8.0	10 % w/v PEG 8000
13	0.05 M caesium chloride	0.1 M MES	6.5	30 % v/v Jeffamine M-600
14	None	0.1 M Na Citrate	5.0	3.2 M ammonium sulfate
15	None	0.1 M Tris	8.0	20 % v/v MPD
16	None	0.1 M HEPES	7.5	20 % v/v Jeffamine M-600
17	0.2 M magnesium chloride	0.1 M Tris	8.5	50 % v/v ethylene glycol
18	None	0.1 M Bicine	9.0	10 % v/v MPD
19	None	None	7.0	0.8 M succinic acid
20	None	None	7.0	2.1 M DL-malic acid
21	None	None	7.0	2.4 M sodium malonate
22	1.1 M sodium malonate	0.1 M HEPES	7.0	0.5 % v/v Jeffamine ED-2001
23	1.0 M succinic acid	0.1 M HEPES	7.0	1 % w/v PEG 2000 MME
24	None	0.1 M HEPES	7.0	30 % v/v Jeffamine M-600
25	None	0.1 M HEPES	7.0	30 % v/v Jeffamine ED-2001
26	0.02 M magnesium chloride	0.1 M HEPES	7.5	22 % w/v polyacrylic acid 5100 sodium salt
27	0.01 M cobalt chloride	0.1 M Tris	8.5	20 % w/v polyvinylpyrrolidone K15
28	0.2 M tri-methylamine N-oxide	0.1 M Tris	8.5	20 % w/v PEG 2000 MME
29	0.005 M cobalt chloride 0.005 M cadmium chloride 0.005 M magnesium chloride 0.005 M nickel chloride	0.1 M HEPES	7.5	12 % w/v PEG 3350
30	0.2 M sodium malonate	None	7.0	20 % w/v PEG 3350
31	0.1 M succinic acid	None	7.0	15 % w/v PEG 3350
32	0.15 M DL - malic acid	None	7.0	20 % w/v PEG 3350
33	0.1 M potassium thiocyanate	None	-	30 % w/v PEG 2000 MME
34	0.15 M potassium bromide	None	-	30 % w/v PEG 2000 MME
35	None	0.1 M Bis Tris	5.5	2.0 M ammonium sulfate
36	None	0.1 M Bis Tris	5.5	3.0 M sodium chloride
37	None	0.1 M Bis Tris	5.5	0.3 M magnesium formate
38	1.0 M ammonium sulfate	0.1 M Bis Tris	5.5	1 % w/v PEG 3350
39	None	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
40	0.2 M calcium chloride	0.1 M Bis Tris	5.5	45 % v/v MPD
41	0.2 M ammonium acetate	0.1 M Bis Tris	5.5	45 % v/v MPD
42	0.1 M ammonium acetate	0.1 M Bis Tris	5.5	17 % w/v PEG 10000
43	0.2 M ammonium sulfate	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
44	0.2 M sodium chloride	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
45	0.2 M lithium sulfate	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
46	0.2 M ammonium acetate	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
47	0.2 M magnesium chloride	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
48	0.2 M ammonium acetate	0.1 M HEPES	7.5	45 % v/v MPD

**PEG/Ion Screen 1**

(Hampton Research, Aliso Viejo, USA; May 2011)

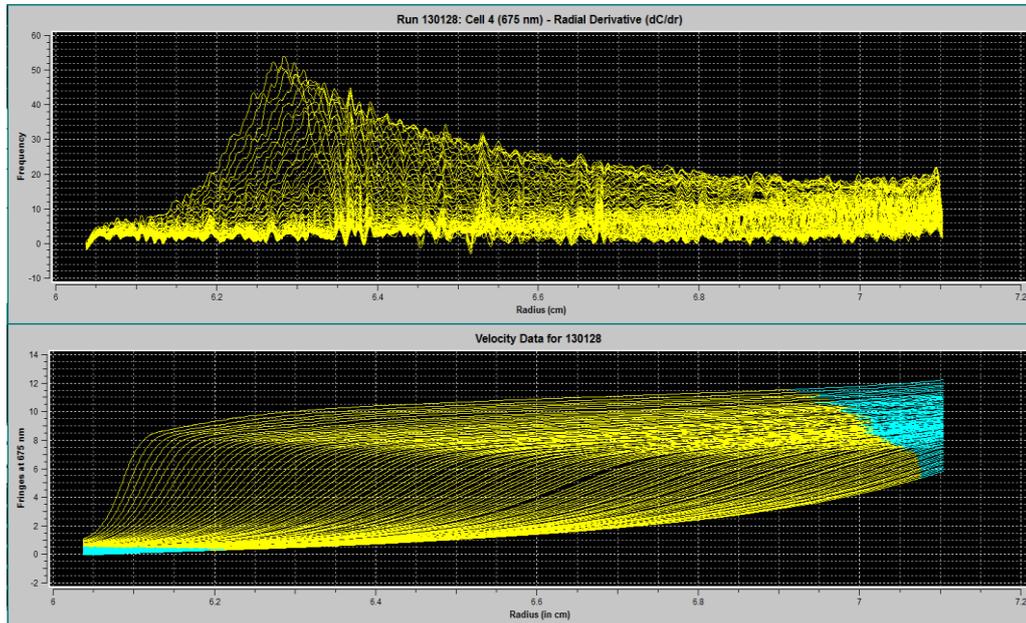
<b>ID</b>	<b>Salt</b>	<b>Buffer</b>	<b>pH</b>	<b>Precipitant</b>
1		0.2 M imidazole malate	5.5	15% v/v PEG 600
2		0.2 M imidazole malate	5.5	24 %v/v PEG 600
3		0.2 M imidazole malate	5.5	33 % v/v PEG 600
4		0.2 M imidazole malate	5.5	42 % v/v PEG 600
5		0.2 M imidazole malate	7	10 % w/v PEG 4000
6		0.2 M imidazole malate	7	15 % w/v PEG 4000
7		0.2 M imidazole malate	7	20 % w/v PEG 4000
8		0.2 M imidazole malate	7	25 % PEG w/v 4000
9		0.2 M imidazole malate	8.5	7.5 % w/v PEG 10,000
10		0.2 M imidazole malate	8.5	12.5 % w/v PEG 10,000
11		0.2 M imidazole malate	8.5	17.5 % w/v PEG 10,000
12		0.2 M imidazole malate	8.5	22.5 % w/v PEG 10,000
13		0.15 M sodium citrate	5.5	0.75 M ammonium sulfate
14		0.15 M sodium citrate	5.5	1.0 M ammonium sulfate
15		0.15 M sodium citrate	5.5	1.5 M ammonium sulfate
16		0.15 M sodium citrate	5.5	2.0 M ammonium sulfate
17		0.8 M NaH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub>	7	-
18		1.32 M NaH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub>	7	-
19		1.6 M NaH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub>	7	-
20		2.0 M NaH <sub>2</sub> PO <sub>4</sub> /K <sub>2</sub> HPO <sub>4</sub>	7	-
21		0.01 M sodium borate	8.5	0.75 M sodium citrate
22		0.01 M sodium borate	8.5	1.0 M sodium citrate
23		0.01 M sodium borate	8.5	1.2 M sodium citrate
24		0.01 M sodium borate	8.5	1.5 M sodium citrate
25		0.1 M Na HEPES	8.2	30 % v/v PEG 550 MME
26		0.1 M Na HEPES	8.2	40 % v/v PEG 550 MME
27		0.1 M Na HEPES	8.2	50 % v/v PEG 550 MME
28		0.1 M Na HEPES	8.2	60 % v/v PEG 550 MME
29		0.1 M Na HEPES	7.5	18 % v/v PEG 600
30		0.1 M Na HEPES	7.5	27 % v/v PEG 600
31		0.1 M Na HEPES	7.5	36 % v/v PEG 600
32		0.1 M Na HEPES	7.5	45 % v/v PEG 600
33		0.1 M sodium cacodylate	6.5	18 % w/v PEG 2000 MME
34		0.1 M sodium cacodylate	6.5	27 % w/v PEG 2000 MME
35		0.1 M sodium cacodylate	6.5	36 % w/v PEG 2000 MME
36		0.1 M sodium cacodylate	6.5	45 % w/v PEG 2000 MME
37		0.2 M imidazole malate	6	8 % w/v PEG 4000
38		0.2 M imidazole malate	6	15 % w/v PEG 4000
39		0.2 M imidazole malate	6	20 % w/v PEG 4000
40		0.2 M imidazole malate	6	30 % w/v PEG 4000
41		0.1 M sodium acetate	5.5	12% PEG w/v 5000 MME
42		0.1 M sodium acetate	5.5	18% PEG w/v 5000 MME
43		0.1 M sodium acetate	5.5	24% PEG w/v 5000 MME
44		0.1 M sodium acetate	5.5	36% PEG w/v 5000 MME
45		0.1 M ammonium acetate	4.5	9 % w/v PEG 10,000
46		0.1 M ammonium acetate	4.5	15 % w/v PEG 10,000
47		0.1 M ammonium acetate	4.5	22.5 % w/v PEG 10,000
48		0.1 M ammonium acetate	4.5	27 % w/v PEG 10,000

**PEG/Ion Screen 2**

(Hampton Research, Aliso Viejo, USA; May 2011)

<b>ID</b>	<b>Salt</b>	<b>Buffer</b>	<b>Precipitant</b>
1	0.2 M Sodium fluoride		20% (w/v) Polyethylene glycol 3,350
2	0.2 M Potassium fluoride		20% (w/v) Polyethylene glycol 3,350
3	0.2 M Ammonium fluoride		20% (w/v) Polyethylene glycol 3,350
4	0.2 M Lithium chloride		20% (w/v) Polyethylene glycol 3,350
5	0.2 M Magnesium chloride hexahydrate		20% (w/v) Polyethylene glycol 3,350
6	0.2 M Sodium chloride		20% (w/v) Polyethylene glycol 3,350
7	0.2 M Calcium chloride dihydrate		20% (w/v) Polyethylene glycol 3,350
8	0.2 M Potassium chloride		20% (w/v) Polyethylene glycol 3,350
9	0.2 M Ammonium chloride		20% (w/v) Polyethylene glycol 3,350
10	0.2 M Sodium iodide		20% (w/v) Polyethylene glycol 3,350
11	0.2 M Potassium iodide		20% (w/v) Polyethylene glycol 3,350
12	0.2 M Ammonium iodide		20% (w/v) Polyethylene glycol 3,350
13	0.2 M Sodium thiocyanate		20% (w/v) Polyethylene glycol 3,350
14	0.2 M Potassium thiocyanate		20% (w/v) Polyethylene glycol 3,350
15	0.2 M Lithium nitrate		20% (w/v) Polyethylene glycol 3,350
16	0.2 M Magnesium nitrate hexahydrate		20% (w/v) Polyethylene glycol 3,350
17	0.2 M Sodium nitrate		20% (w/v) Polyethylene glycol 3,350
18	0.2 M Potassium nitrate		20% (w/v) Polyethylene glycol 3,350
19	0.2 M Ammonium nitrate		20% (w/v) Polyethylene glycol 3,350
20	0.2 M Magnesium formate dihydrate		20% (w/v) Polyethylene glycol 3,350
21	0.2 M Sodium formate		20% (w/v) Polyethylene glycol 3,350
22	0.2 M Potassium formate		20% (w/v) Polyethylene glycol 3,350
23	0.2 M Ammonium formate		20% (w/v) Polyethylene glycol 3,350
24	0.2 M Lithium acetate dihydrate		20% (w/v) Polyethylene glycol 3,350
25	0.2 M Magnesium acetate tetrahydrate		20% (w/v) Polyethylene glycol 3,350
26	0.2 M Zinc acetate dihydrate		20% (w/v) Polyethylene glycol 3,350
27	0.2 M Sodium acetate trihydrate		20% (w/v) Polyethylene glycol 3,350
28	0.2 M Calcium acetate hydrate		20% (w/v) Polyethylene glycol 3,350
29	0.2 M Potassium acetate		20% (w/v) Polyethylene glycol 3,350
30	0.2 M Ammonium acetate		20% (w/v) Polyethylene glycol 3,350
31	0.2 M Lithium sulfate monohydrate		20% (w/v) Polyethylene glycol 3,350
32	0.2 M Magnesium sulfate heptahydrate		20% (w/v) Polyethylene glycol 3,350
33	0.2 M Sodium sulfate decahydrate		20% (w/v) Polyethylene glycol 3,350
34	0.2 M Potassium sulfate		20% (w/v) Polyethylene glycol 3,350
35	0.2 M Ammonium sulfate		20% (w/v) Polyethylene glycol 3,350
36	0.2 M Sodium tartrate dibasic dihydrate		20% (w/v) Polyethylene glycol 3,350
37	0.2 M Potassium sodium tartrate tetrahydrate		20% (w/v) Polyethylene glycol 3,350
38	0.2 M Ammonium tartrate dibasic		20% (w/v) Polyethylene glycol 3,350
39	0.2 M Sodium phosphate monobasic monohydrate		20% (w/v) Polyethylene glycol 3,350
40	0.2 M Sodium phosphate dibasic dihydrate		20% (w/v) Polyethylene glycol 3,350
41	0.2 M Potassium phosphate monobasic		20% (w/v) Polyethylene glycol 3,350
42	0.2 M Potassium phosphate dibasic		20% (w/v) Polyethylene glycol 3,350
43	0.2 M Ammonium phosphate monobasic		20% (w/v) Polyethylene glycol 3,350
44	0.2 M Ammonium phosphate dibasic		20% (w/v) Polyethylene glycol 3,350
45	0.2 M Lithium citrate tribasic tetrahydrate		20% (w/v) Polyethylene glycol 3,350
46	0.2 M Sodium citrate tribasic dihydrate		20% (w/v) Polyethylene glycol 3,350
47	0.2 M Potassium citrate tribasic monohydrate		20% (w/v) Polyethylene glycol 3,350
48	0.2 M Ammonium citrate dibasic		20% (w/v) Polyethylene glycol 3,350

## VI. AUC sedimentation velocity profile of MUB in PBS



## VII. MubR5 crystal soaking and co-crystallisation data sets

sugar ligand	MubR5 data collection								
	Fuc	Man		GalNAc	GlcNAc	LacNAc			Neu5Ac
soak/ co-crystallisation	S	S	C	S	S	S	C	C	C
Ligand concentration (mM)	200	200	10	200	500	5	50	100	200
Beamline	Diamond i02	Diamond i02	Diamond i04	Diamond i02	Diamond i02	Diamond i02	Diamond i04	Diamond i03	Diamond i03
Wavelength (Å)	0.9795	0.9795	0.9763	0.9795	0.9795	0.9795	0.9763	0.9709	0.9709
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>								
Cell parameters: a, b, c (Å)	45.3, 45.6, 197.3	45.3, 45.5, 197.9	45.2, 45.7, 197.2	46.8, 47.0, 194.6	44.8, 45.4, 197.8	45.3, 45.6, 197.8	45.0, 45.6, 197.8	45.0, 45.1, 196.6	46.0, 45.0, 196.4
Resolution (Å)	45-1.8	45-1.7	37-1.5	45-1.7	45-1.8	45-1.7	49-1.8	39-1.5	41-1.7
R <sub>sym</sub> (%)	8.8 (32.6)	6.3 (21.5)	3.6 (11.4)	5.3 (18.4)	6.8 (31.8)	9.1 (41.4)	6.5 (26.1)	6.9 (37.4)	7.2 (22.9)
I/σ	9.0 (3.3)	17.1 (5.2)	17.8 (8.2)	21.9 (8.4)	11.8 (3.5)	10.2 (2.9)	15.8 (5.1)	13.5 (4.3)	10.6 (4.2)
Unique reflections	37602	47572	67337	45219	40084	46769	39964	56162	53950
Completeness (%)	97.4 (99.0)	99.7 (99.4)	97.9 (98.5)	95.7 (98.9)	98.8 (97.8)	99.9 (99.9)	99.1 (98.0)	90.1 (91.0)	99.5 (99.5)
Multiplicity	3.4 (3.6)	3.8 (3.8)	2.7 (2.6)	6.1 (6.4)	3.9 (4.0)	3.9 (3.9)	3.8 (3.8)	5.7 (6.1)	3.8 (3.8)

### VIII. MubRI data processing parameters

	MubRI data collection			
	Native	Native	Ytterbium	Mercury chloride
Ligand concentration (mM)	-	-	10	200
Beamline	Diamond i03	Diamond i04	Diamond i04	Diamond i04
Wavelength (Å)	0.97090	0.98000	1.38670	1.00780
Space group	P222	P222	P222	P222
Cell parameters: a, b, c (Å)	44.1, 49.5, 216.5	44.2, 49.3, 217.1	44.3, 49.6, 217.6	44.1, 49.3, 217.3
Resolution (Å)	49.5-1.9 (2.0-1.9)	49.2-2.0 (2.1-2.0)	54.4-2.4 (2.5-2.4)	49.0-2.6 (2.7-2.6)
R <sub>sym</sub> (%)	21.6 (198.8)	16.7 (124.6)	12.7 (50.9)	9.3 (41.6)
I/σ	6.2 (1.1)	3.6 (3.5)	6.1 (6.6)	5.6 (5.5)
Unique reflections	39405 (4948)	34167 (4398)	20040 (3090)	16201 (2621)
Completeness (%)	99.0 (76.2)	99.6 (71.8)	99.7 (79.9)	99.4 (86.8)
Multiplicity	7.3 (7.3)	3.6 (3.5)	6.1 (6.6)	5.6 (5.5)
p-value <sup>a</sup>	2.384e-02	2.2520e-03	3.435e-03	3.688e-03

<sup>a</sup>p-values determined by XTRIAGE for reflection data quality analysis; p-value < 0.05 indicates weak translational pseudosymmetry; p-value < 1e-03 indicates strong pseudosymmetry

## IX. SAXS data set statistics

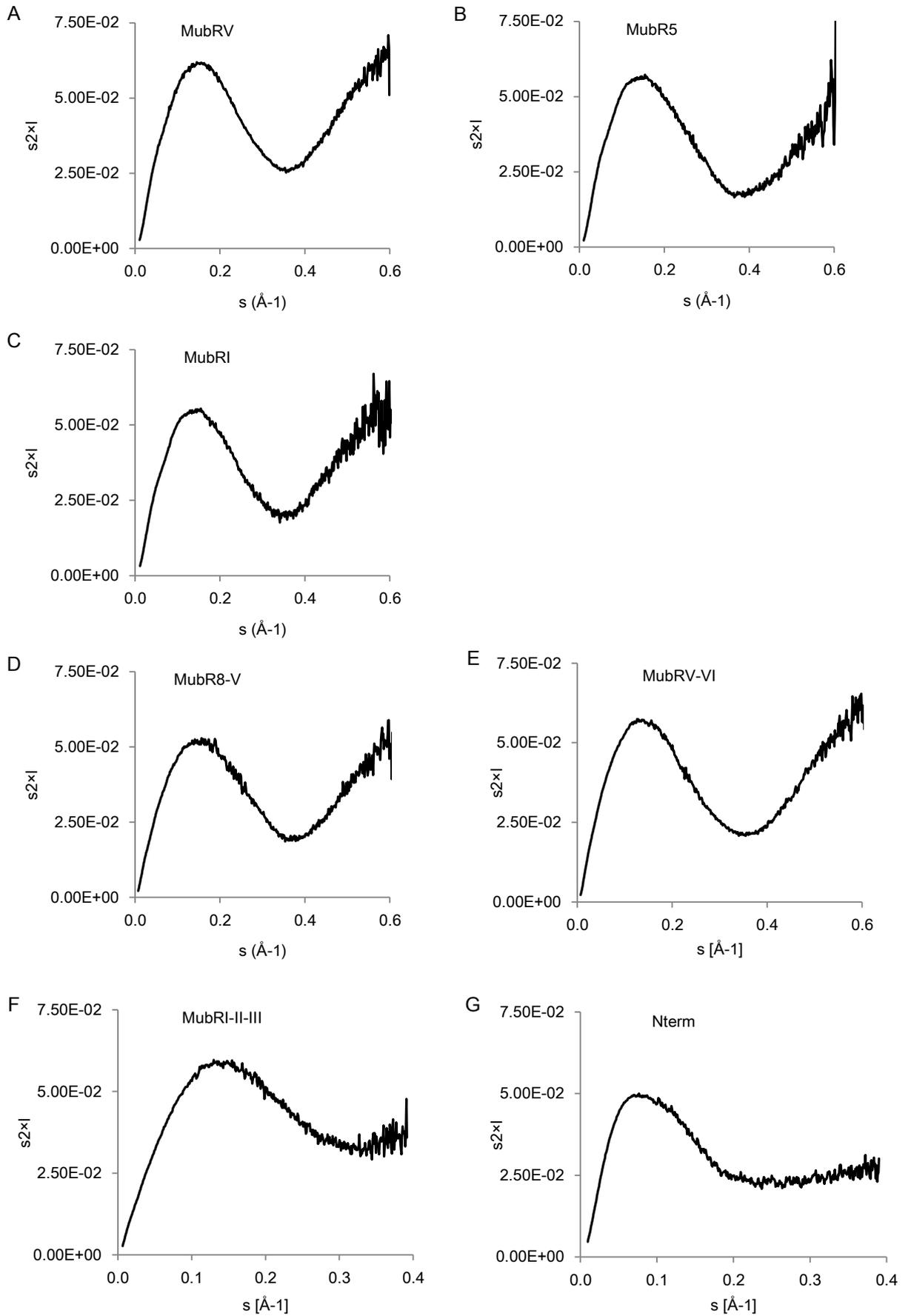
sample	Molecular weight calculated from sequence	Molecular weight determined from SAXS data <sup>a</sup>	R <sub>g</sub> (Guinier analysis)	R <sub>g</sub> (GNOM analysis)	D <sub>max</sub>	GASPOR		SITUS/ Sculptor	CRY SOL
	[kDa]	[kDa]	[nm]	[nm]	[nm]	X <sup>b</sup>	NSD <sup>c</sup>	R	X
MubR5	20.5	24.4	3.00 ± 0.05	3.15 ± 0.01	10.5	1.8	0.91 ± 0.04	0.83	1.06
MubRV	20.3	25.9	3.09 ± 0.05	3.23 ± 0.01	11.0	2.18	0.87 ± 0.01	0.82	1.02
MubRI	21.3	25.2	3.10 ± 0.04	3.29 ± 0.00	10.6	1.92	0.92 ± 0.02	0.84	-
MubR8-V	40.6	48.3	5.87 ± 0.10	6.03 ± 0.02	20.5	3.15	1.55 ± 0.08	0.60	-
MubRV-VI	38.7	48.7	5.79 ± 0.15	5.37 ± 0.02	20.6	3.34	1.48 ± 0.12	0.54	-
Nterm	56.3	53.3	4.53 ± 0.05	4.47 ± 0.02	15.9	1.40	1.54 ± 0.05	-	-
						DAMMIF			
MubRI-II-III	64.1	71.1	8.66 ± 0.44	8.61 ± 0.50	29.2	1.55	1.02 ± 0.07	0.45	-

<sup>a</sup>particle molecular weight calculated using the formula  $I(0) \div I(0)_{BSA} \times MW_{BSA}$ ; where  $MW_{BSA}$  was 66.5 kDa and  $I(0)_{BSA}$  was 65.9 or 66.8 (Nterm or MubRI-II-III)

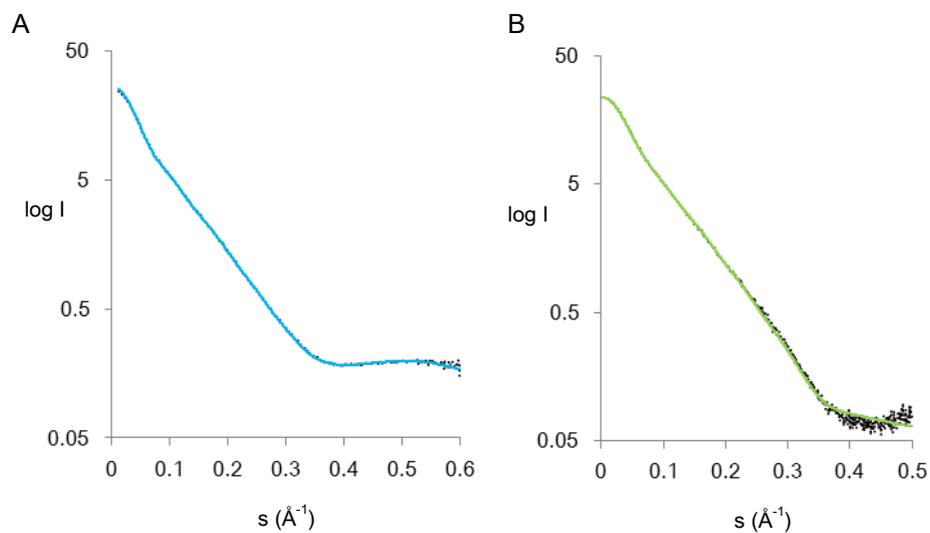
<sup>b</sup>average discrepancy of computed scattering curves from 10 ab-initio shape reconstructions compared to experimental scattering curves

<sup>c</sup>NSD value between 10 individual shape reconstructions

## X. Kratky analysis of SAXS scattering data

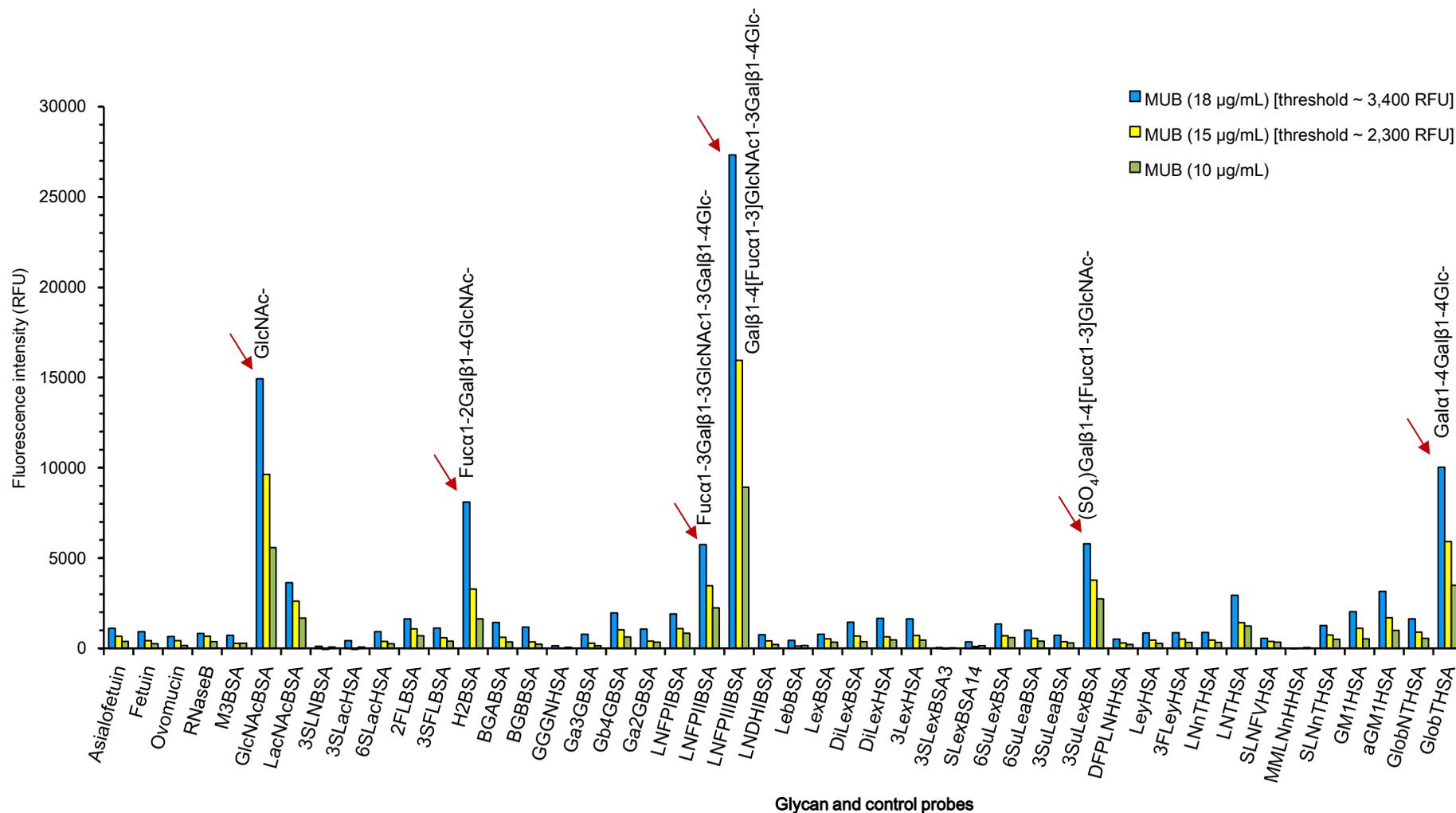


## XI. Comparison of experimental and calculated SAXS scattering data of MubRV and MubR5



Comparison of experimental scattering data of (A) MubRV and (B) MubR5 (black dotted line) and calculated SAXS scattering profiles (blue for Mub-RV, green for Mub-R5) from X-ray crystal structures of Mub-RV and Mub-R5 using CRY SOL.

## XII. Screening of glycoconjugates for MUB binding



### XIII. Screening of the mammalian glycan array (CFG) for MUB binding

Chart Number	MUB detected by anti-MubR5 CFG#2311 Slide#:13810 09/12/2011 at 488nm	Average RFU	StDev	% CV
1	Gala-Sp8	17	23	136
2	Glca-Sp8	12	14	118
3	Mana-Sp8	14	6	44
4	GalNAca-Sp8	12	6	54
5	GalNAca-Sp15	8	8	93
6	Fuca-Sp8	11	7	59
7	Fuca-Sp9	5	9	176
8	Rhaa-Sp8	21	12	58
9	Neu5Aca-Sp8	17	17	95
10	Neu5Aca-Sp11	10	4	39
11	Neu5Acb-Sp8	9	4	48
12	Galb-Sp8	9	14	150
13	Glcb-Sp8	8	1	13
14	Manb-Sp8	7	4	52
15	GalNAcb-Sp8	15	12	79
16	GlcNAcb-Sp0	12	9	72
17	GlcNAcb-Sp8	3	7	219
18	GlcN(Gc)b-Sp8	11	10	96
19	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAca-Sp8	11	6	57
20	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)GalNAc-Sp14	16	10	67
21	GlcNAcb1-6(GlcNAcb1-4)(GlcNAcb1-3)GlcNAc-Sp8	5	5	104
22	6S(3S)Galb1-4(6S)GlcNAcb-Sp0	13	4	30
23	6S(3S)Galb1-4GlcNAcb-Sp0	14	7	52
24	(3S)Galb1-4(Fuca1-3)(6S)Glc-Sp0	6	6	105
25	(3S)Galb1-4Glcb-Sp8	10	12	113

26	(3S)Galb1-4(6S)GlcB-Sp0	11	5	44
27	(3S)Galb1-4(6S)GlcB-Sp8	10	10	99
28	(3S)Galb1-3(Fuca1-4)GlcNAc-Sp8	8	5	61
29	(3S)Galb1-3GalNAc-Sp8	3	5	139
30	(3S)Galb1-3GlcNAc-Sp0	10	7	68
31	(3S)Galb1-3GlcNAc-Sp8	4	1	19
32	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp0	7	7	98
33	(3S)Galb1-4(Fuca1-3)GlcNAc-Sp8	14	8	59
34	(3S)Galb1-4(6S)GlcNAc-Sp0	10	15	153
35	(3S)Galb1-4(6S)GlcNAc-Sp8	7	2	30
36	(3S)Galb1-4GlcNAc-Sp0	10	5	48
37	(3S)Galb1-4GlcNAc-Sp8	6	4	68
38	(3S)Galb-Sp8	4	3	91
39	(6S)(4S)Galb1-4GlcNAc-Sp0	9	16	186
40	(4S)Galb1-4GlcNAc-Sp8	10	7	70
41	(6P)Mana-Sp8	5	5	100
42	(6S)Galb1-4GlcB-Sp0	9	7	78
43	(6S)Galb1-4GlcB-Sp8	0	4	-848
44	(6S)Galb1-4GlcNAc-Sp8	9	10	119
45	(6S)Galb1-4(6S)GlcB-Sp8	10	4	39
46	Neu5Aca2-3(6S)Galb1-4GlcNAc-Sp8	4	7	181
47	(6S)GlcNAc-Sp8	8	6	75
48	Neu5,9Ac <sub>2</sub> a-Sp8	6	3	44
49	Neu5,9Ac <sub>2</sub> a2-6Galb1-4GlcNAc-Sp8	3	2	75
50	Mana1-6(Mana1-3)Manb1-4GlcNAc1-4GlcNAc-Sp12	5	5	120
51	Mana1-6(Mana1-3)Manb1-4GlcNAc1-4GlcNAc-Sp13	6	4	71
52	GlcNAc1-2Mana1-6(GlcNAc1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc-Sp12	15	10	69
53	GlcNAc1-2Mana1-6(GlcNAc1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc-Sp13	2	2	84
54	Galb1-4GlcNAc1-2Mana1-6(Galb1-4GlcNAc1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc-Sp12	18	14	79
55	Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAc1-2Mana1-3)Manb1-4GlcNAc1-4GlcNAc-Sp12	8	5	59

56	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	6	5	81
57	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Man-a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	12	10	87
58	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	16	3	19
59	Fuca1-2Galb1-3GalNAcb1-3Gala-Sp9	1	3	296
60	Fuca1-2Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp9	7	3	49
61	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb-Sp8	14	6	45
62	Fuca1-2Galb1-3GalNAca-Sp8	1	3	341
63	Fuca1-2Galb1-3GalNAca-Sp14	5	4	80
64	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0	6	5	81
65	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp9	10	6	65
66	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp8	1	2	130
67	Fuca1-2Galb1-3GlcNAcb1-3Galb1-4Glc-Sp10	11	7	64
68	Fuca1-2Galb1-3GlcNAcb-Sp0	5	5	109
69	Fuca1-2Galb1-3GlcNAcb-Sp8	9	4	51
70	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	2	100
71	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	2	49
72	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp0	8	4	44
73	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb-Sp8	2	2	131
74	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	4	4	101
75	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	11	11	105
76	Fuca1-2Galb1-4GlcNAcb-Sp0	14	8	58
77	Fuca1-2Galb1-4GlcNAcb-Sp8	11	6	57
78	Fuca1-2Galb1-4Glc-Sp0	-1	3	-280
79	Fuca1-2Galb-Sp8	2	1	60
80	Fuca1-3GlcNAcb-Sp8	11	8	77
81	Fuca1-4GlcNAcb-Sp8	8	3	31
82	Fucb1-3GlcNAcb-Sp8	8	4	49
83	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	5	4	76

84	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	9	9	98
85	(3S)Galb1-4(Fuca1-3)Glc-Sp0	4	5	105
86	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	7	6	90
87	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	4	5	104
88	GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	2	4	185
89	GlcNAcb1-3Galb1-3GalNAca-Sp8	9	6	68
90	GalNAca1-3(Fuca1-2)Galb-Sp8	5	3	56
91	GalNAca1-3(Fuca1-2)Galb-Sp18	7	7	91
92	GalNAca1-3GalNAcb-Sp8	5	2	36
93	GalNAca1-3Galb-Sp8	1	2	199
94	GalNAca1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	0	1	316
95	GalNAcb1-3GalNAca-Sp8	6	4	56
96	GalNAcb1-3(Fuca1-2)Galb-Sp8	4	2	58
97	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	4	4	95
98	GalNAcb1-4(Fuca1-3)GlcNAcb-Sp0	12	4	35
99	GalNAcb1-4GlcNAcb-Sp0	7	4	63
100	GalNAcb1-4GlcNAcb-Sp8	5	5	84
101	Gala1-2Galb-Sp8	2	4	228
102	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	8	6	77
103	Gala1-3(Fuca1-2)Galb1-3GlcNAcb-Sp8	7	9	131
104	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	3	105
105	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp8	6	8	135
106	Gala1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	6	7	102
107	Gala1-3(Fuca1-2)Galb1-4Glc-Sp0	3	4	144
108	Gala1-3(Fuca1-2)Galb-Sp8	7	10	133
109	Gala1-3(Fuca1-2)Galb-Sp18	12	7	60
110	Gala1-4(Gala1-3)Galb1-4GlcNAcb-Sp8	10	3	31
111	Gala1-3GalNAca-Sp8	18	3	14
112	Gala1-3GalNAca-Sp16	1	6	544
113	Gala1-3GalNAcb-Sp8	0	6	2084
114	Gala1-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	6	1	23

115	Gala1-3Galb1-3GlcNAcb-Sp0	5	2	49
116	Gala1-3Galb1-4GlcNAcb-Sp8	12	4	30
117	Gala1-3Galb1-4Glc-Sp0	11	3	23
118	Gala1-3Galb1-4Glc-Sp10	2	3	164
119	Gala1-3Galb-Sp8	6	5	84
120	Gala1-4(Fuca1-2)Galb1-4GlcNAcb-Sp8	1	4	316
121	Gala1-4Galb1-4GlcNAcb-Sp0	7	4	53
122	Gala1-4Galb1-4GlcNAcb-Sp8	17	15	89
123	Gala1-4Galb1-4Glc-Sp0	9	10	107
124	Gala1-4GlcNAcb-Sp8	9	4	45
125	Gala1-6Glc-Sp8	13	13	101
126	Galb1-2Galb-Sp8	9	5	55
127	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	4	6	150
128	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	2	5	220
129	Galb1-3(Fuca1-4)GlcNAc-Sp0	22	10	46
130	Galb1-3(Fuca1-4)GlcNAc-Sp8	4	4	110
131	Fuca1-4(Galb1-3)GlcNAcb-Sp8	5	3	50
132	Galb1-4GlcNAcb1-6GalNAca-Sp8	2	3	157
133	Galb1-4GlcNAcb1-6GalNAc-Sp14	14	7	47
134	GlcNAcb1-6(Galb1-3)GalNAca-Sp8	5	6	110
135	GlcNAcb1-6(Galb1-3)GalNAca-Sp14	3	6	202
136	Neu5Aca2-6(Galb1-3)GalNAca-Sp8	8	8	105
137	Neu5Aca2-6(Galb1-3)GalNAca-Sp14	10	6	65
138	Neu5Acb2-6(Galb1-3)GalNAca-Sp8	5	2	35
139	Neu5Aca2-6(Galb1-3)GlcNAcb1-4Galb1-4Glc-Sp10	10	5	53
140	Galb1-3GalNAca-Sp8	4	2	58
141	Galb1-3GalNAca-Sp14	13	10	75
142	Galb1-3GalNAca-Sp16	4	3	78
143	Galb1-3GalNAcb-Sp8	4	2	47
144	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0	10	8	80
145	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0	4	4	110

146	Galb1-3GalNAcb1-4Galb1-4GlcB-Sp8	5	4	66
147	Galb1-3Galb-Sp8	8	8	93
148	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	5	1	26
149	Galb1-3GlcNAcb1-3Galb1-4GlcB-Sp10	4	2	60
150	Galb1-3GlcNAcb-Sp0	8	8	98
151	Galb1-3GlcNAcb-Sp8	7	5	72
152	Galb1-4(Fuca1-3)GlcNAcb-Sp0	10	7	66
153	Galb1-4(Fuca1-3)GlcNAcb-Sp8	11	3	25
154	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	4	1	16
155	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	12	6	49
156	Galb1-4(6S)GlcB-Sp0	8	8	96
157	Galb1-4(6S)GlcB-Sp8	12	12	99
158	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	10	6	65
159	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	7	4	59
160	Galb1-4GlcNAcb1-3GalNAca-Sp8	2	3	146
161	Galb1-4GlcNAcb1-3GalNAc-Sp14	4	3	60
162	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	4	135
163	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	12	9	70
164	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	1	3	362
165	Galb1-4GlcNAcb1-3Galb1-4GlcB-Sp0	3	3	133
166	Galb1-4GlcNAcb1-3Galb1-4GlcB-Sp8	6	3	57
167	Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp8	8	1	13
168	Galb1-4GlcNAcb1-6(Galb1-3)GalNAc-Sp14	4	4	95
169	Galb1-4GlcNAcb-Sp0	7	5	67
170	Galb1-4GlcNAcb-Sp8	6	2	34
171	Galb1-4GlcNAcb-Sp23	9	5	54
172	Galb1-4GlcB-Sp0	0	1	-193
173	Galb1-4GlcB-Sp8	1	2	114
174	GlcNAca1-3Galb1-4GlcNAcb-Sp8	2	2	133
175	GlcNAca1-6Galb1-4GlcNAcb-Sp8	9	3	31
176	GlcNAcb1-2Galb1-3GalNAca-Sp8	4	2	54

177	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp8	2	3	108
178	GlcNAcb1-6(GlcNAcb1-3)GalNAca-Sp14	5	6	130
179	GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAcb-Sp8	7	5	74
180	GlcNAcb1-3GalNAca-Sp8	8	6	76
181	GlcNAcb1-3GalNAca-Sp14	2	2	83
182	GlcNAcb1-3Galb-Sp8	7	5	77
183	GlcNAcb1-3Galb1-4GlcNAcb-Sp0	12	4	36
184	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	9	7	85
185	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	2	3	123
186	GlcNAcb1-3Galb1-4Glc-Sp0	5	6	125
187	GlcNAcb1-4-MDPLys	13	13	98
188	GlcNAcb1-6(GlcNAcb1-4)GalNAca-Sp8	7	3	44
189	GlcNAcb1-4Galb1-4GlcNAcb-Sp8	12	8	64
190	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	1	2	236
191	GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-4GlcNAcb1-Sp8	8	5	68
192	GlcNAcb1-4GlcNAcb1-4GlcNAcb-Sp8	3	3	93
193	GlcNAcb1-6GalNAca-Sp8	7	3	42
194	GlcNAcb1-6GalNAca-Sp14	8	7	84
195	GlcNAcb1-6Galb1-4GlcNAcb-Sp8	16	8	48
196	Glca1-4Glc-Sp8	1	2	170
197	Glca1-4Glca-Sp8	12	7	58
198	Glca1-6Glca1-6Glc-Sp8	9	2	22
199	Glc-Sp8	10	10	103
200	Glc-Sp8	5	2	44
201	G-ol-Sp8	11	7	60
202	GlcAa-Sp8	5	4	85
203	GlcAb-Sp8	10	7	69
204	GlcAb1-3Galb-Sp8	7	4	63
205	GlcAb1-6Galb-Sp8	4	2	55
206	KDNa2-3Galb1-3GlcNAcb-Sp0	5	7	148
207	KDNa2-3Galb1-4GlcNAcb-Sp0	6	8	133

208	Mana1-2Mana1-2Mana1-3Mana-Sp9	5	3	56
209	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana-Sp9	7	5	70
210	Mana1-2Mana1-3Mana-Sp9	10	9	91
211	Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	3	2	87
212	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	5	4	76
213	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	5	3	59
214	Mana1-6(Mana1-3)Mana-Sp9	0	2	973
215	Mana1-2Mana1-2Mana1-6(Mana1-3)Mana-Sp9	12	9	73
216	Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	18	12	65
217	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	9	113
218	Manb1-4GlcNAcb-Sp0	4	3	68
219	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	11	14	122
220	(3S)Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	12	8	64
221	Fuca1-2(6S)Galb1-4GlcNAcb-Sp0	7	5	65
222	Fuca1-2Galb1-4(6S)GlcNAcb-Sp8	4	3	63
223	Fuca1-2(6S)Galb1-4(6S)GlcNAcb-Sp0	13	15	115
224	Neu5Aca2-3Galb1-3GalNAca-Sp8	7	4	54
225	Neu5Aca2-3Galb1-3GalNAca-Sp14	10	5	51
226	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	5	4	78
227	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	5	4	78
228	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAcb-Sp0	10	8	79
229	GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	9	6	60
230	Neu5Aca2-8Neu5Aca2-8Neu5Aca-Sp8	5	3	75
231	Neu5Aca2-3(6S)Galb1-4(Fuca1-3)GlcNAcb-Sp8	12	5	42
232	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	2	2	93
233	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp8	11	9	84
234	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	5	5	101
235	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb-Sp0	6	5	73
236	Neu5Aca2-6(Neu5Aca2-3)GalNAca-Sp8	8	3	34
237	Neu5Aca2-3GalNAca-Sp8	7	3	39

238	Neu5Aca2-3GalNAcb1-4GlcNAcb-Sp0	12	8	65
239	Neu5Aca2-3Galb1-3(6S)GlcNAc-Sp8	5	3	73
240	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp8	7	9	127
241	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	6	3	54
242	Neu5Aca2-3Galb1-4(Neu5Aca2-3Galb1-3)GlcNAcb-Sp8	12	7	60
243	Neu5Aca2-3Galb1-3(6S)GalNAca-Sp8	7	9	129
244	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp8	5	3	63
245	Neu5Aca2-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	8	3	41
246	Neu5Aca2-3Galb-Sp8	14	12	91
247	Neu5Aca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0	10	3	35
248	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	6	4	68
249	Fuca1-2(6S)Galb1-4Glc-Sp0	21	6	27
250	Neu5Aca2-3Galb1-3GlcNAcb-Sp0	9	3	40
251	Neu5Aca2-3Galb1-3GlcNAcb-Sp8	11	3	24
252	Neu5Aca2-3Galb1-4(6S)GlcNAcb-Sp8	4	6	146
253	Neu5Aca2-3Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	-1	2	-356
254	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	11	7	61
255	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	6	199
256	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp8	5	4	81
257	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb-Sp8	11	6	58
258	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	4	4	100
259	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	2	3	154
260	Neu5Aca2-3Galb1-4GlcNAcb-Sp0	9	7	81
261	Neu5Aca2-3Galb1-4GlcNAcb-Sp8	5	7	134
262	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	4	5	133
263	Fuca1-2Galb1-4(6S)Glc-Sp0	6	1	11
264	Neu5Aca2-3Galb1-4Glc-Sp0	5	7	133
265	Neu5Aca2-3Galb1-4Glc-Sp8	4	3	78
266	Neu5Aca2-6GalNAca-Sp8	11	10	85
267	Neu5Aca2-6GalNAcb1-4GlcNAcb-Sp0	13	3	22
268	Neu5Aca2-6Galb1-4(6S)GlcNAcb-Sp8	7	7	90

269	Neu5Aca2-6Galb1-4GlcNAcb-Sp0	1	3	181
270	Neu5Aca2-6Galb1-4GlcNAcb-Sp8	3	5	164
271	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	2	75
272	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	9	8	91
273	Neu5Aca2-6Galb1-4Glc-Sp0	7	2	31
274	Neu5Aca2-6Galb1-4Glc-Sp8	7	5	67
275	Neu5Aca2-6Galb-Sp8	3	5	149
276	Neu5Aca2-8Neu5Aca-Sp8	7	5	77
277	Neu5Aca2-8Neu5Aca2-3Galb1-4Glc-Sp0	6	1	13
278	Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	16	10	65
279	Neu5Acb2-6GalNAca-Sp8	9	5	48
280	Neu5Acb2-6Galb1-4GlcNAcb-Sp8	6	6	93
281	Neu5Gca2-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	4	6	133
282	Neu5Gca2-3Galb1-3GlcNAcb-Sp0	6	3	52
283	Neu5Gca2-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	5	8	146
284	Neu5Gca2-3Galb1-4GlcNAcb-Sp0	9	18	189
285	Neu5Gca2-3Galb1-4Glc-Sp0	1	7	1317
286	Neu5Gca2-6GalNAca-Sp0	2	3	112
287	Neu5Gca2-6Galb1-4GlcNAcb-Sp0	5	9	161
288	Neu5Gca-Sp8	10	7	67
289	Neu5Aca2-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	8	3	41
290	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	1	2	339
291	Galb1-4(Fuca1-3)(6S)GlcNAcb-Sp0	6	3	56
292	Galb1-4(Fuca1-3)(6S)Glc-Sp0	10	4	43
293	Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	8	8	98
294	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	6	5	82
295	Neu5Aca2-3Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	8	7	84
296	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	9	2	20
297	4S(3S)Galb1-4GlcNAcb-Sp0	13	8	62
298	(6S)Galb1-4(6S)GlcNAcb-Sp0	3	1	42
299	(6P)Glc-Sp10	4	2	41

300	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3)GalNAca-Sp14	5	4	73
301	Galb1-3Galb1-4GlcNAcb-Sp8	11	6	50
302	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	3	2	86
303	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	13	14	110
304	GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4GlcNAc-Sp0	6	6	99
305	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	2	4	238
306	Galb1-4GlcNAcb1-6Galb1-4GlcNAcb-Sp0	11	3	25
307	GalNAcb1-3Galb-Sp8	11	2	17
308	GlcAb1-3GlcNAcb-Sp8	10	2	17
309	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	10	12	119
310	GlcNAcb1-3Man-Sp10	8	7	94
311	GlcNAcb1-4GlcNAcb-Sp10	5	6	111
312	GlcNAcb1-4GlcNAcb-Sp12	8	7	85
313	MurNAcb1-4GlcNAcb-Sp10	5	4	89
314	Mana1-6Manb-Sp10	14	10	74
315	Mana1-6(Mana1-3)Mana1-6(Mana1-3)Manb-Sp10	13	8	61
316	Mana1-2Mana1-6(Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	4	4	105
317	Mana1-2Mana1-6(Mana1-2Mana1-3)Mana1-6(Mana1-2Mana1-2Mana1-3)Mana-Sp9	4	6	167
318	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	7	8	112
319	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	13	6	49
320	Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	5	5	98
321	GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	7	3	50
322	Neu5Aca2-8Neu5Acb-Sp17	8	4	51
323	Neu5Aca2-8Neu5Aca2-8Neu5Acb-Sp8	4	2	42
324	Neu5Gcb2-6Galb1-4GlcNAc-Sp8	4	3	59
325	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	9	5	60
326	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	23	9	38

327	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	12	9	77
328	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	11	6	53
329	Neu5,9Ac2a2-3Galb1-4GlcNAcb-Sp0	1	5	402
330	Neu5,9Ac2a2-3Galb1-3GlcNAcb-Sp0	4	7	161
331	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	-1	5	-517
332	Neu5Aca2-3Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-3(Fuca1-4)GlcNAcb-Sp0	16	10	61
333	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	4	5	124
334	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	2	6	398
335	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	6	4	65
336	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	5	3	59
337	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	6	6	102
338	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAc-Sp14	5	5	93
339	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	8	6	74
340	GlcNAca1-4Galb1-4GlcNAcb-Sp0	6	3	52
341	GlcNAca1-4Galb1-3GlcNAcb-Sp0	16	10	63
342	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	9	11	126
343	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	3	1	41
344	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	4	6	142
345	GlcNAca1-4Galb1-3GalNAc-Sp14	13	6	44
346	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	4	4	113
347	Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	5	7	126
348	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	7	4	56
349	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	11	3	23
350	Galb1-4GlcNAcb1-2Mana1-3Manb1-4GlcNAcb1-4GlcNAc-Sp12	10	7	67
351	Galb1-4GlcNAcb1-2Mana1-6Manb1-4GlcNAcb1-4GlcNAc-Sp12	12	8	67
352	Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	2	1	44
353	GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	11	7	60
354	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	11	12	110
355	Galb1-3GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	5	4	69

356	(6S)GlcNAcb1-3Galb1-4GlcNAcb-Sp0	3	1	23
357	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	7	6	83
358	KDNa2-6Galb1-4GlcNAc-Sp0	3	3	115
359	KDNa2-3Galb1-4Glc-Sp0	5	3	59
360	KDNa2-3Galb1-3GalNAca-Sp14	2	4	190
361	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	5	6	138
362	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	4	3	67
363	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	7	2	23
364	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	7	8	112
365	Galb1-4GlcNAcb1-2Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	9	5	60
366	Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-6(Fuca1-4(Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	5	1	29
367	Neu5Aca2-6GlcNAcb1-4GlcNAc-Sp21	4	2	57
368	Neu5Aca2-6GlcNAcb1-4GlcNAcb1-4GlcNAc-Sp21	7	4	64
369	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	1	3	204
370	Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	8	7	85
371	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	9	7	79
372	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	15	8	53
373	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	9	8	89
374	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	7	3	38
375	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	3	3	93

376	Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	2	3	179
377	Neu5Aca2-3Galb1-4GlcNAcb1-3GalNAc-Sp14	4	5	107
378	Neu5Aca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	2	4	204
379	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	13	4	27
380	GalNAcb1-4GlcNAcb1-2Mana1-6(GalNAcb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	12	6	51
381	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	9	3	32
382	Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAc-Sp0	7	8	109
383	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp0	2	5	218
384	Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	4	6	166
385	Galb1-4GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	5	13	275
386	Galb1-4(Fuca1-3)GlcNAcb1-6(Fuca1-4(Fuca1-2Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	-1	2	-145
387	Galb1-3GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	2	4	189
388	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	19	6	32
389	GlcNAcb1-2Mana1-6(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	4	7	193
390	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4Glc-Sp0	0	4	-843
391	Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	3	3	106
392	Galb1-3GlcNAcb1-3GalNAca-Sp14	4	4	112
393	GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcNAcb1-3GalNAca-Sp14	10	5	47
394	GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	4	2	47
395	Gala1-3Galb1-3GlcNAcb1-2Mana1-6(Gala1-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	13	11	84
396	Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Gala1-3Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	5	2	41
397	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp19	6	3	48
398	GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	10	8	82
399	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp12	6	6	109
400	Neu5Aca2-3Galb1-3GlcNAcb1-3GalNAca-Sp14	9	12	133
401	Fuca1-2Galb1-4GlcNAcb1-3GalNAca-Sp14	6	5	78
402	Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	10	2	18

403	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	1	2	354
404	Gala1-4Galb1-3GlcNAcb1-2Mana1-6(Gala1-4Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	2	3	178
405	Gala1-4Galb1-4GlcNAcb1-2Mana1-6(Gala1-4Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	15	8	53
406	Gala1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	6	5	90
407	Galb1-3GlcNAcb1-6Galb1-4GlcNAcb-Sp0	5	6	112
408	Galb1-3GlcNAca1-6Galb1-4GlcNAcb-Sp0	5	1	26
409	GalNAcb1-3Gala1-6Galb1-4Glc-Sp8	1	5	626
410	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)Glc-Sp21	9	10	116
411	Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	9	4	50
412	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0	10	10	99
413	Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0	2	3	110
414	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	7	4	61
415	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	6	4	65
416	GalNAca1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0	5	3	73
417	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3GalNAca-Sp14	5	4	91
418	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	2	4	174
419	GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb1-3GalNAc-Sp14	8	3	44
420	Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	6	2	35
421	Fuca1-2Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	8	3	45
422	GlcNAcb1-2(GlcNAcb1-6)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	4	9	203
423	Fuca1-2Galb1-3GlcNAcb1-3GalNAc-Sp14	-1	3	-284
424	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	9	8	80
425	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	4	2	54
426	Gala1-3Galb1-3GlcNAcb1-3GalNAc-Sp14	5	6	126
427	Fuca1-2Galb1-3GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	2	3	109
428	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	10	6	60

429	Galb1-3GlcNAcb1-6(Galb1-3GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	15	6	41
430	Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	3	3	106
431	Fuca1-3GlcNAcb1-6(Galb1-4GlcNAcb1-3)Galb1-4Glc-Sp21	2	4	198
432	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	1	1	96
433	GlcNAcb1-2Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	9	4	47
434	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	2	4	202
435	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	2	3	113
436	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	-1	2	-122
437	Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	4	5	110
438	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	7	2	30
439	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	8	6	78
440	Galb1-4Galb-Sp10	6	7	105
441	Galb1-6Galb-Sp10	12	4	39
442	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb-Sp8	2	3	167
443	GalNAcb1-6GalNAcb-Sp8	6	6	103
444	(6S)Galb1-3GlcNAcb-Sp0	10	10	92
445	(6S)Galb1-3(6S)GlcNAc-Sp0	6	3	57
446	Fuca1-2Galb1-4 GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-2(Fuca1-2Galb1-4GlcNAcb1-4)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	6	3	39
447	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-4(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	13	15	116
448	Galb1-4(Fuca1-3)GlcNAcb1-6GalNAc-Sp14	12	6	44
449	Galb1-4GlcNAcb1-2Mana-Sp0	10	9	91
450	Fuca1-2Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4GlcNAcb1-3)GalNAc-Sp14	7	3	40

451	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6(Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	8	6	85
452	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3)GalNAc-Sp14	3	2	85
453	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp0	4	4	123
454	GalNAcb1-4Galb1-4Glc-Sp0	6	5	78
455	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	11	8	74
456	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	6	6	110
457	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)Galb1-4Glc-Sp21	13	2	13
458	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp22	8	7	88
459	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	13	3	24
460	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	4	4	122
461	Neu5Aca2-3Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	12	6	52
462	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	4	5	151
463	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-4)(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	5	3	68
464	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	11	4	32
465	Neu5Aca2-6Galb1-4GlcNAcb1-4Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	5	6	120
466	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	1	6	518
467	Neu5Aca2-6Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-4)(Neu5Aca2-6Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	5	3	62
468	Gala1-3(Fuca1-2)Galb1-3GalNAca-Sp8	10	11	108
469	Gala1-3(Fuca1-2)Galb1-3GalNAcb-Sp8	2	4	215
470	Glca1-6Glca1-6Glca1-6Glc-Sp10	7	3	52

471	Glca1-4Glca1-4Glca1-4Glc- Sp10	14	7	47
472	Neu5Aca2-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	9	8	83
473	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	4	4	96
474	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6(Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	8	13	158
475	Neu5Aca2-3Galb1-3GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2)Mana1-6(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp19	21	7	36
476	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	7	7	99
477	Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp21	10	14	141
478	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3)Galb1-4Glc- Sp21	7	7	99
479	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana-Sp0	6	4	64
480	Neu5Aca2-3Galb1-4GlcNAcb1-6GalNAca-Sp14	9	6	69
481	Neu5Aca2-6Galb1-4GlcNAcb1-6GalNAca-Sp14	7	2	27
482	Neu5Aca2-6Galb1-4 GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3)GalNAca-Sp14	5	4	71
483	Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	13	15	115
484	Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	13	8	67
485	Mana1-6(Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp19	3	5	139
486	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	9	7	77
487	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-6(GlcNAcb1-4)(Neu5Aca2-3Galb1-3GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAc-Sp21	17	6	38
488	Neu5Aca2-6Galb1-4GlcNAcb1-6(Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	6	4	64
489	Galb1-3GlcNAcb1-6GalNAca-Sp14	14	12	81
490	Gala1-3Galb1-3GlcNAcb1-6GalNAca-Sp14	4	2	42
491	Galb1-3(Fuca1-4)GlcNAcb1-6GalNAca-Sp14	10	10	98
492	Neu5Aca2-3Galb1-3GlcNAcb1-6GalNAca-Sp14	2	2	79
493	(3S)Galb1-3(Fuca1-4)GlcNAcb-Sp0	1	5	977
494	Galb1-4(Fuca1-3)GlcNAcb1-6(Neu5Aca2-6(Neu5Aca2-3Galb1-3)GlcNAcb1-3)Galb1-4Glc-Sp21	6	6	104

495	Fuca1-2Galb1-4GlcNAcb1-6GalNAca-Sp14	12	1	10
496	Gala1-3Galb1-4GlcNAcb1-6GalNAca-Sp14	11	9	80
497	Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	10	2	21
498	Fuca1-2(6S)Galb1-3GlcNAcb-Sp0	5	3	59
499	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAca-Sp14	8	4	55
500	Fuca1-2Galb1-4GlcNAcb1-2Mana-Sp0	1	3	321
501	Fuca1-2Galb1-3(6S)GlcNAcb-Sp0	5	2	46
502	Fuca1-2(6S)Galb1-3(6S)GlcNAcb-Sp0	11	10	84
503	Neu5Aca2-6GalNAcb1-4(6S)GlcNAcb-Sp8	3	4	164
504	GalNAcb1-4(Fuca1-3)(6S)GlcNAcb-Sp8	14	12	85
505	(3S)GalNAcb1-4(Fuca1-3)GlcNAcb-Sp8	9	2	18
506	Fuca1-2Galb1-3GlcNAcb1-6(Fuca1-2Galb1-3GlcNAcb1-3)GalNAca-Sp14	10	7	67
507	GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAca-Sp14	10	3	31
508	GlcNAcb1-6(GlcNAcb1-2)Mana1-6(GlcNAcb1-4)(GlcNAcb1-4(GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	3	4	113
509	Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-4)Galb1-4GlcNAcb1-4(Galb1-4GlcNAcb1-2)Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	13	16	123
510	Galb1-3GlcNAca1-3Galb1-4GlcNAcb-Sp8	7	6	83
511	Galb1-3(6S)GlcNAcb-Sp8	11	8	69
512	(6S)(4S)GalNAcb1-4GlcNAc-Sp8	4	12	294
513	(6S)GalNAcb1-4GlcNAc-Sp8	4	3	59
514	(3S)GalNAcb1-4(3S)GlcNAc-Sp8	5	1	27
515	GalNAcb1-4(6S)GlcNAc-Sp8	8	5	63
516	(3S)GalNAcb1-4GlcNAc-Sp8	1	4	291
517	(4S)GalNAcb-Sp10	8	13	161
518	Galb1-4(6P)GlcNAcb-Sp0	16	8	47
519	(6P)Galb1-4GlcNAcb-Sp0	21	12	58
520	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-6GalNAc-Sp14	9	9	110
521	Neu5Aca2-6Galb1-4GlcNAcb1-2Man-Sp0	8	9	118
522	Gala1-3Galb1-4GlcNAcb1-2Mana-Sp0	6	2	33
523	Gala1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana-Sp0	7	4	54

524	GalNAca1-3(Fuca1-2)Galb1-4 GlcNAcb1-2Mana-Sp0	10	16	161
525	Galb1-3GlcNAcb1-2Mana-Sp0	8	4	53
526	Gala1-3(Fuca1-2)Galb1-3GlcNAcb1-6GalNAc-Sp14	7	3	45
527	Neu5Aca2-3Galb1-3GlcNAcb1-2Mana-Sp0	1	3	250
528	Gala1-3Galb1-3GlcNAcb1-2Mana-Sp0	13	7	59
529	GalNAcb1-4GlcNAcb1-2Mana-Sp0	8	3	37
530	Neu5Aca2-3Galb1-3GlcNAcb1-4Galb1-4Glc-Sp0	5	8	161
531	GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	20	21	103
532	Galb1-4GlcNAcb1-2 Mana1-6(GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	7	3	46
533	Galb1-4GlcNAcb1-2 Mana1-6(Galb1-4GlcNAcb1-4)(Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAc-Sp21	14	6	41
534	Fuca1-4(Galb1-3)GlcNAcb1-2 Mana-Sp0	4	4	86
535	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana-Sp0	7	7	109
536	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3)Galb1-4GlcNAc-Sp0	6	8	131
537	GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	6	3	50
538	Gala1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp21	14	13	88
539	Galb1-3GalNAcb1-3Gal-Sp21	4	3	71
540	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	2	2	76
541	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	4	5	142
542	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	6	71
543	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	6	4	68
544	Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	8	3	34
545	Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Fuca1-2Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	5	2	47
546	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	18	6	30

547	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	5	5	109
548	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	13	8	61
549	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	2	3	146
550	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	15	6	42
551	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	6	5	80
552	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	4	2	42
553	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb-Sp0	7	3	42
554	Neu5Aca2-8Neu5Gca2-3Galb1-4GlcNAcb-Sp0	4	3	89
555	Neu5Gca2-8Neu5Aca2-3Galb1-4GlcNAcb-Sp0	7	7	94
556	Neu5Gca2-8Neu5Gca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	10	4	37
557	Neu5Gca2-8Neu5Gca2-6Galb1-4GlcNAcb-Sp0	1	4	273
558	Neu5Aca2-8Neu5Aca2-3Galb1-4GlcNAcb-Sp0	1	2	258
559	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Man a1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	8	4	51
560	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Mana1-4GlcNAcb1-4GlcNAcb-Sp24	6	6	93
561	Gala1-3Galb1-4GlcNAcb1-2Mana1-6(Gala1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp24	6	12	191
562	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-3)GalNAca-Sp14	6	5	78
563	GalNAcb1-3GlcNAcb-Sp0	9	6	73
564	GalNAcb1-4GlcNAcb1-3GalNAcb1-4GlcNAcb-Sp0	9	5	53
565	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	7	3	46

566	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	5	2	35
567	GlcNAcb1-3Galb1-3GalNAc-Sp14	11	7	64
568	Galb1-3GlcNAcb1-6(Galb1-3)GalNAc-Sp14	3	7	260
569	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp25	9	7	77
570	(3S)GlcAb1-3Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	3	7	221
571	(3S)GlcAb1-3Galb1-4GlcNAcb1-2Mana-Sp0	4	3	89
572	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	17	10	58
573	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2)Mana1-6(Galb1-3GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	16	5	30
574	Neu5Aca2-8Neu5Aca2-3Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp21	3	4	117
575	GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	12	4	31
576	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	3	3	117
577	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	3	3	118
578	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	8	3	42
579	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	24	11	46
580	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4(Fuca1-6)GlcNAcb-Sp24	12	2	20



596	GlcNAcb1-3Galb1-4GlcNAcb1-6(GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	6	3	55
597	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	7	5	71
598	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	18	11	64
599	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3GalNAca-Sp14	5	4	70
600	Galb1-4GlcNAcb1-3Galb1-3GalNAca-Sp14	6	6	94
601	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	7	3	39
602	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	7	2	29
603	Neu5Aca2-6Galb1-4GlcNAcb1-6(Galb1-3)GalNAca-Sp14	5	2	34
604	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	12	10	89
605	GlcNAcb1-6(Neu5Aca2-3Galb1-3)GalNAca-Sp14	4	3	71
606	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3)GalNAca-Sp14	5	5	113
607	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	17	15	86
608	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	8	6	80
609	Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-6(Neu5Aca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-2Mana1-3)Manb1-4GlcNAcb1-4GlcNAcb-Sp12	9	2	24
610	GlcNAcb1-3Fuca-Sp21	15	7	46
611	Galb1-3GalNAcb1-4(Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3)Galb1-4Glc-Sp21	5	6	120

#### XIV. MUB binding to plant derived glycans

non-BSA-conjugated structures		
Name	Glycan structure	Description
Sugar beet pectin	( $\alpha$ 1-4)-galacturonic acid	Possible modifications galactose, rhamnose, arabinose, fucose
Lime pectin	( $\alpha$ 1-4)-galacturonic acid	as above
Arabinan	D-arabinose (C5)	Component of hemicullose or pectin
Gum arabic	( $\beta$ 1-3)( $\beta$ 1-6)-galactose (plus Gal, Rha, Ara, GlcA sidechains ( $\beta$ 1-6) linked)	Dietary and functional fibre for (thickening agent)
Gum guar	Galactomannan, ( $\beta$ 1-4)-mannose with 1,6 linked galactose (every second mannose)	guar beans, dietary and functional fibre
Carrageenan	( $\alpha$ 1-3),( $\beta$ 1-4)-galactose, sulfated	polysaccharide from red seaweeds
Tomato mannan	( $\beta$ 1-4)-mannose	plant polysaccharide
$\beta$ -glucan	(1-3)- $\beta$ -D-Glucan	yeast
Pachyman	(1-3)- $\beta$ -D-Glucan	
BSA-conjugated structures		
Arabinose	( $\alpha$ 1-5)-L-arabinobiose(C5), feruloyted (feruloyl, phenolic acid)	hemicullose or pectin
Chitin	( $\alpha$ 1-5)-L-arabinotriose, feruloyted	
Glucose	( $\beta$ 1-4)-chitotobiose to -chitohexaose	
Laminarin	D-glucose	storage glucan of algae
Cellulose	( $\beta$ 1-3)-glucotriose to -hexaose ( )	
Maltose	( $\beta$ 1-4)-glucotriose to glucohexaose	
Mlg (Glc)	( $\alpha$ 1-4)-glucobiose to -decaose	
Xylose	( $\alpha$ 1-6), ( $\alpha$ 1-4)-glucotetraose	
XG	( $\beta$ 1-3),( $\beta$ 1-4)-glucotriose to – glucopentaose (A,B,C)	hemicullose precursor
Isoprimeverose	( $\beta$ 1-4)-xylotriase to -xylohexaose (C5), XG-heptamer	XG blood group antigen, glycoprotein
Mannose	xylosyl- $\beta$ 1-6-glucose	Hydrolysis product of xyloglucan
Galactomannan	D-mannose to ( $\beta$ 1-4)-mannohexaose	
Gal2M5	galactosyl-( $\beta$ 1-4)-mannopentaose	
GalM2	galactosyl-( $\beta$ 1-4)-mannobiose	
Galactose	$\beta$ -D-galactose	
Lactose	( $\beta$ 1-4)-galactotobiose	
Galactouronic acid (GalA)	galactosyl-( $\beta$ 1-4)-galactobiose, feruloyted	
	galactosyl-( $\beta$ 1-4)-glucose	
	( $\alpha$ 1-4)-hexaglactouronate (and derivates) + -octagalacturonate	Main component of pectin

## LIST OF REFERENCES

- [1] Bengmark S (1998) Ecological control of the gastrointestinal tract. The role of probiotic flora. *Gut* 42(1):2-7.
- [2] Chivers DJ & Langer P (1994) Gut form and functions: variations and terminology. *The digestive system in mammals: food, form and function*. (Press Syndecate of the University of Cambridge, Cambridge, UK), pp 3-8.
- [3] Evans DF, Pye G, Bramley R, Clark AG, Dyson TJ, Hardcastle JD (1988) Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* 29(8):1035-1041.
- [4] Fallingborg J, Christensen LA, Ingeman-Nielsen M, Jacobsen BA, Abildgaard K, Rasmussen HH (1989) pH-profile and regional transit times of the normal gut measured by a radiotelemetry device. *Aliment Pharmacol Ther* 3(6):605-613.
- [5] Fallingborg J, Christensen LA, Ingeman-Nielsen M, Jacobsen BA, Abildgaard K, Rasmussen HH, Rasmussen SN (1990) Measurement of gastrointestinal pH and regional transit times in normal children. *J Pediatr Gastroenterol Nutr* 11(2):211-214.
- [6] Fallingborg J, Christensen LA, Ingeman-Nielsen M, Jacobsen BA, Abildgaard K, Rasmussen HH, Rasmussen SN (1990) Gastrointestinal pH and transit times in healthy subjects with ileostomy. *Aliment Pharmacol Ther* 4(3):247-253.
- [7] Walter J & Ley R (2011) The human gut microbiome: ecology and recent evolutionary changes. *Annu Rev Microbiol* 65:411-429.
- [8] Velez MP, De Keersmaecker SC, Vanderleyden J (2007) Adherence factors of *Lactobacillus* in the human gastrointestinal tract. *FEMS Microbiol Lett* 276(2):140-148.
- [9] Forstner JF & Forstner GG (1994) Gastrointestinal mucus. *Physiology of the Gastrointestinal Tract*, ed Johnson LR (Raven Press, New York), pp 1255-1283.
- [10] Potten CS, Booth C, Pritchard DM (1997) The intestinal epithelial stem cell: the mucosal governor. *Int J Exp Pathol* 78(4):219-243.
- [11] Johansson ME (2012) Fast renewal of the distal colonic mucus layers by the surface goblet cells as measured by in vivo labeling of mucin glycoproteins. *PLoS One* 7(7):e41009.
- [12] Powell DW (1987) Ion and water transport in the intestine. *Physiology of Membrane Disorders*, eds Andreoli TE & Schultz SG (Plenum, New York), Vol 2, pp 559-596.
- [13] Allen A (1989) Gastrointestinal mucus. *Handbook of Physiology. The Gastrointestinal System. Salivary, Gastric, Pancreatic, and Hepatobiliary Secretion.*, (Physiol. Soc., Bethesda, MD, USA), Vol 3, pp 359-382.
- [14] Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) Host-bacterial mutualism in the human intestine. *Science* 307(5717):1915-1920.

- [15] Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307(5717):1955-1959.
- [16] Coconnier MH, Dlissi E, Robard M, Laboisse CL, Gaillard JL, Servin AL (1998) *Listeria monocytogenes* stimulates mucus exocytosis in cultured human polarized mucosecreting intestinal cells through action of listeriolysin O. *Infect Immun* 66(8):3673-3681.
- [17] Halm DR & Halm ST (2000) Secretagogue response of goblet cells and columnar cells in human colonic crypts. *Am J Physiol Cell Physiol* 278(1):C212-233.
- [18] Lencer WI, Reinhart FD, Neutra MR (1990) Interaction of cholera toxin with cloned human goblet cells in monolayer culture. *Am J Physiol* 258(1 Pt 1):G96-102.
- [19] Mack DR, Ahrne S, Hyde L, Wei S, Hollingsworth MA (2003) Extracellular MUC3 mucin secretion follows adherence of *Lactobacillus* strains to intestinal epithelial cells in vitro. *Gut* 52(6):827-833.
- [20] Mack DR, Michail S, Wei S, McDougall L, Hollingsworth MA (1999) Probiotics inhibit enteropathogenic *E. coli* adherence in vitro by inducing intestinal mucin gene expression. *Am J Physiol* 276(4 Pt 1):G941-950.
- [21] Miyake K, Tanaka T, McNeil PL (2006) Disruption-induced mucus secretion: repair and protection. *PLoS Biol* 4(9):e276.
- [22] Juge N (2012) Microbial adhesins to gastrointestinal mucus. *Trends Microbiol* 20(1):30-39.
- [23] Atuma C, Strugala V, Allen A, Holm L (2001) The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *Am J Physiol Gastrointest Liver Physiol* 280(5):G922-929.
- [24] Ermund A, Schuette A, Johansson ME, Gustafsson JK, Hansson GC (2013) Gastrointestinal mucus layers have different properties depending on location - 1. Studies of mucus in mouse stomach, small intestine, Peyer's patches and colon. *Am J Physiol Gastrointest Liver Physiol*.
- [25] Johansson ME, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci U S A* 105(39):15064-15069.
- [26] Pullan RD, Thomas GA, Rhodes M, Newcombe RG, Williams GT, Allen A, Rhodes J (1994) Thickness of adherent mucus gel on colonic mucosa in humans and its relevance to colitis. *Gut* 35(3):353-359.
- [27] Johansson ME, Thomsson KA, Hansson GC (2009) Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the Fcgbp protein. *J Proteome Res* 8(7):3549-3557.
- [28] Rodriguez-Pineiro AM, Bergstrom JH, Ermund A, Gustafsson JK, Schuette A, Johansson ME, Hansson GC (2013) Gastrointestinal mucus proteome reveals Muc2

- and Muc5ac accompanied by a set of core proteins - 2. Studies of mucus in mouse stomach, small intestine, and colon. *Am J Physiol Gastrointest Liver Physiol*.
- [29] Johansson ME, Larsson JM, Hansson GC (2011) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* 108 Suppl 1:4659-4665.
- [30] Velcich A, Yang W, Heyer J, Fragale A, Nicholas C, Viani S, Kucherlapati R, Lipkin M, Yang K, Augenlicht L (2002) Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* 295(5560):1726-1729.
- [31] Harada N, Iijima S, Kobayashi K, Yoshida T, Brown WR, Hibi T, Oshima A, Morikawa M (1997) Human IgGfC binding protein (FcγBP) in colonic epithelial cells exhibits mucin-like structure. *J Biol Chem* 272(24):15232-15241.
- [32] Duerkop BA, Vaishnav S, Hooper LV (2009) Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* 31(3):368-376.
- [33] Brandtzaeg P (2007) Induction of secretory immunity and memory at mucosal surfaces. *Vaccine* 25(30):5467-5484.
- [34] Shroff KE, Meslin K, Cebra JJ (1995) Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut. *Infect Immun* 63(10):3904-3913.
- [35] Shimada S, Kawaguchi-Miyashita M, Kushiro A, Sato T, Nanno M, Sako T, Matsuoka Y, Sudo K, Tagawa Y, Iwakura Y, Ohwaki M (1999) Generation of polymeric immunoglobulin receptor-deficient mouse with marked reduction of secretory IgA. *J Immunol* 163(10):5367-5373.
- [36] Johansen FE, Pekna M, Norderhaug IN, Haneberg B, Hietala MA, Krajci P, Betsholtz C, Brandtzaeg P (1999) Absence of epithelial immunoglobulin A transport, with increased mucosal leakiness, in polymeric immunoglobulin receptor/secretory component-deficient mice. *J Exp Med* 190(7):915-922.
- [37] Sait LC, Galic M, Price JD, Simpfendorfer KR, Diavatopoulos DA, Uren TK, Janssen PH, Wijburg OL, Strugnell RA (2007) Secretory antibodies reduce systemic antibody responses against the gastrointestinal commensal flora. *Int Immunol* 19(3):257-265.
- [38] Cash HL, Whitham CV, Hooper LV (2006) Refolding, purification, and characterization of human and murine RegIII proteins expressed in *Escherichia coli*. *Protein Expr Purif* 48(1):151-159.
- [39] Cunliffe RN (2003) Alpha-defensins in the gastrointestinal tract. *Mol Immunol* 40(7):463-467.
- [40] Erlandsen SL, Parsons JA, Taylor TD (1974) Ultrastructural immunocytochemical localization of lysozyme in the Paneth cells of man. *J Histochem Cytochem* 22(6):401-413.
- [41] Ouellette AJ, Miller SI, Henschen AH, Selsted ME (1992) Purification and primary structure of murine cryptdin-1, a Paneth cell defensin. *FEBS Lett* 304(2-3):146-148.

- [42] Pellegrini A, Thomas U, von Fellenberg R, Wild P (1992) Bactericidal activities of lysozyme and aprotinin against gram-negative and gram-positive bacteria related to their basic character. *J Appl Bacteriol* 72(3):180-187.
- [43] Selsted ME & Ouellette AJ (1995) Defensins in granules of phagocytic and non-phagocytic cells. *Trends in cell biology* 5(3):114-119.
- [44] Hallock KJ, Lee DK, Ramamoorthy A (2003) MSI-78, an analogue of the magainin antimicrobial peptides, disrupts lipid bilayer structure via positive curvature strain. *Biophys J* 84(5):3052-3060.
- [45] Kagan BL, Ganz T, Lehrer RI (1994) Defensins: a family of antimicrobial and cytotoxic peptides. *Toxicology* 87(1-3):131-149.
- [46] Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, Ley R, Wakeland EK, Hooper LV (2011) The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* 334(6053):255-258.
- [47] Kaiser V & Diamond G (2000) Expression of mammalian defensin genes. *J Leukocyte Biol* 68(6):779-784.
- [48] Vaishnava S, Behrendt CL, Hooper LV (2008) Innate immune responses to commensal bacteria in the gut epithelium. *J Pediatr Gastroenterol Nutr* 46 Suppl 1:E10-11.
- [49] Cash HL, Whitham CV, Behrendt CL, Hooper LV (2006) Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* 313(5790):1126-1130.
- [50] Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA (2008) Mucins in the mucosal barrier to infection. *Mucosal Immunol* 1(3):183-197.
- [51] Hollingsworth MA & Swanson BJ (2004) Mucins in cancer: protection and control of the cell surface. *Nat Rev Cancer* 4(1):45-60.
- [52] Allen A, Hutton DA, Pearson JP (1998) The MUC2 gene product: a human intestinal mucin. *Int J Biochem Cell Biol* 30(7):797-801.
- [53] Andrianifahanana M, Moniaux N, Batra SK (2006) Regulation of mucin expression: mechanistic aspects and implications for cancer and inflammatory diseases. *Biochim Biophys Acta* 1765(2):189-222.
- [54] Audie JP, Janin A, Porchet N, Copin MC, Gosselin B, Aubert JP (1993) Expression of human mucin genes in respiratory, digestive, and reproductive tracts ascertained by in situ hybridization. *J Histochem Cytochem* 41(10):1479-1485.
- [55] Buisine MP, Devisme L, Degand P, Dieu MC, Gosselin B, Copin MC, Aubert JP, Porchet N (2000) Developmental mucin gene expression in the gastroduodenal tract and accessory digestive glands. II. Duodenum and liver, gallbladder, and pancreas. *J Histochem Cytochem* 48(12):1667-1676.
- [56] Chang SK, Dohrman AF, Basbaum CB, Ho SB, Tsuda T, Toribara NW, Gum JR, Kim YS (1994) Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology* 107(1):28-36.

- [57] Corfield AP, Carroll D, Myerscough N, Probert CS (2001) Mucins in the gastrointestinal tract in health and disease. *Front Biosci* 6:D1321-1357.
- [58] Gum JR Jr., Crawley SC, Hicks JW, Szymkowski DE, Kim YS (2002) MUC17, a novel membrane-tethered mucin. *Biochem Biophys Res Commun* 291(3):466-475.
- [59] Pratt WS, Crawley S, Hicks J, Ho J, Nash M, Kim YS, Gum JR, Swallow DM (2000) Multiple transcripts of MUC3: evidence for two genes, MUC3A and MUC3B. *Biochem Biophys Res Commun* 275(3):916-923.
- [60] Reis CA, David L, Correa P, Carneiro F, de Bolos C, Garcia E, Mandel U, Clausen H, Sobrinho-Simoes M (1999) Intestinal metaplasia of human stomach displays distinct patterns of mucin (MUC1, MUC2, MUC5AC, and MUC6) expression. *Cancer Res* 59(5):1003-1007.
- [61] Desseyn JL, Tetaert D, Gouyer V (2008) Architecture of the large membrane-bound mucins. *Gene* 410(2):215-222.
- [62] Kato K, Lillehoj EP, Park YS, Umehara T, Hoffman NE, Madesh M, Kim KC (2012) Membrane-tethered MUC1 mucin is phosphorylated by epidermal growth factor receptor in airway epithelial cells and associates with TLR5 to inhibit recruitment of MyD88. *J Immunol* 188(4):2014-2022.
- [63] Li Y, Kuwahara H, Ren J, Wen G, Kufe D (2001) The c-Src tyrosine kinase regulates signaling of the human DF3/MUC1 carcinoma-associated antigen with GSK3 beta and beta-catenin. *J Biol Chem* 276(9):6061-6064.
- [64] Wang H, Lillehoj EP, Kim KC (2004) MUC1 tyrosine phosphorylation activates the extracellular signal-regulated kinase. *Biochem Biophys Res Commun* 321(2):448-454.
- [65] Levitin F, Stern O, Weiss M, Gil-Henn C, Ziv R, Prokocimer Z, Smorodinsky NI, Rubinstein DB, Wreschner DH (2005) The MUC1 SEA module is a self-cleaving domain. *J Biol Chem* 280(39):33374-33386.
- [66] Palmi-Pallag T, Khodabukus N, Kinarsky L, Leir SH, Sherman S, Hollingsworth MA, Harris A (2005) The role of the SEA (sea urchin sperm protein, enterokinase and agrin) module in cleavage of membrane-tethered mucins. *FEBS J* 272(11):2901-2911.
- [67] Perez-Vilar J & Hill RL (1999) The structure and assembly of secreted mucins. *J Biol Chem* 274(45):31751-31754.
- [68] Gum JR, Jr., Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. *J Biol Chem* 269(4):2440-2446.
- [69] Sadler JE (1998) Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem* 67:395-424.
- [70] Bansil R, Stanley E, LaMont JT (1995) Mucin biophysics. *Annual review of physiology* 57:635-657.
- [71] Van den Steen P, Rudd PM, Dwek RA, Opdenakker G (1998) Concepts and principles of O-linked glycosylation. *Critical reviews in biochemistry and molecular biology* 33(3):151-208.

- [72] Brockhausen I (1997) Biosynthesis and functions of *O*-glycans and regulation of mucin antigen expression in cancer. *Biochem Soc Trans* 25(3):871-874.
- [73] Filipe IM RS (1995) The histochemistry of intestinal mucins; changes in disease. *Gastrointestinal and oesophageal pathology*, ed R W (Churchill Livingstone, Edinburgh, United Kingdom), pp 73-95.
- [74] Morita H, Kettlewell MG, Jewell DP, Kent PW (1993) Glycosylation and sulphation of colonic mucus glycoproteins in patients with ulcerative colitis and in healthy subjects. *Gut* 34(7):926-932.
- [75] Oriol R, Le Pendu J, Mollicone R (1986) Genetics of ABO, H, Lewis, X and related antigens. *Vox Sang* 51(3):161-171.
- [76] Oriol R (1995) ABO, Hh, Lewis, and secretion serology, genetics, and tissue distribution. . *Blood Cell Biochemistry* ed Cartron J RP (Plenum press, New York, NY, USA), Vol 6, pp 37-73.
- [77] Roth J, Taatjes DJ, Weinstein J, Paulson JC, Greenwell P, Watkins WM (1986) Differential subcompartmentation of terminal glycosylation in the Golgi apparatus of intestinal absorptive and goblet cells. *J Biol Chem* 261(30):14307-14312.
- [78] Van Klinken BJ, Van der Wal JW, Einerhand AW, Buller HA, Dekker J (1999) Sulphation and secretion of the predominant secretory human colonic mucin MUC2 in ulcerative colitis. *Gut* 44(3):387-393.
- [79] Larsson JM, Karlsson H, Crespo JG, Johansson ME, Eklund L, Sjovall H, Hansson GC (2011) Altered *O*-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm Bowel Dis* 17(11):2299-2307.
- [80] Gendler SJ & Spicer AP (1995) Epithelial mucin genes. *Annual review of physiology* 57:607-634.
- [81] Toribara NW, Ho SB, Gum E, Gum JR, Jr., Lau P, Kim YS (1997) The carboxyl-terminal sequence of the human secretory mucin, MUC6. Analysis Of the primary amino acid sequence. *J Biol Chem* 272(26):16398-16403.
- [82] Asker N, Axelsson MA, Olofsson SO, Hansson GC (1998) Dimerization of the human MUC2 mucin in the endoplasmic reticulum is followed by a *N*-glycosylation-dependent transfer of the mono- and dimers to the Golgi apparatus. *J Biol Chem* 273(30):18857-18863.
- [83] Ho JJ, Jaituni RS, Crawley SC, Yang SC, Gum JR, Kim YS (2003) *N*-glycosylation is required for the surface localization of MUC17 mucin. *Int J Oncol* 23(3):585-592.
- [84] Parry S, Hanisch FG, Leir SH, Sutton-Smith M, Morris HR, Dell A, Harris A (2006) *N*-Glycosylation of the MUC1 mucin in epithelial cells and secretions. *Glycobiology* 16(7):623-634.
- [85] McGuckin MA, Linden SK, Sutton P, Florin TH (2011) Mucin dynamics and enteric pathogens. *Nat Rev Microbiol* 9(4):265-278.

- [86] Lan MS, Batra SK, Qi WN, Metzgar RS, Hollingsworth MA (1990) Cloning and sequencing of a human pancreatic tumor mucin cDNA. *J Biol Chem* 265(25):15294-15299.
- [87] Patton S, Gendler SJ, Spicer AP (1995) The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim Biophys Acta* 1241(3):407-423.
- [88] Pandey P, Kharbanda S, Kufe D (1995) Association of the DF3/MUC1 breast cancer antigen with Grb2 and the Sos/Ras exchange protein. *Cancer Res* 55(18):4000-4003.
- [89] Zrihan-Licht S, Baruch A, Elroy-Stein O, Keydar I, Wreschner DH (1994) Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins. Cytokine receptor-like molecules. *FEBS Lett* 356(1):130-136.
- [90] Macao B, Johansson DG, Hansson GC, Hard T (2006) Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nat Struct Mol Biol* 13(1):71-76.
- [91] Wreschner DH, McGuckin MA, Williams SJ, Baruch A, Yoeli M, Ziv R, Okun L, Zaretsky J, Smorodinsky N, Keydar I, Neophytou P, Stacey M, Lin HH, Gordon S (2002) Generation of ligand-receptor alliances by "SEA" module-mediated cleavage of membrane-associated mucin proteins. *Protein Sci* 11(3):698-706.
- [92] Hanisch FG & Muller S (2000) MUC1: the polymorphic appearance of a human mucin. *Glycobiology* 10(5):439-449.
- [93] Thathiah A, Blobel CP, Carson DD (2003) Tumor necrosis factor-alpha converting enzyme/ADAM 17 mediates MUC1 shedding. *J Biol Chem* 278(5):3386-3394.
- [94] Thathiah A & Carson DD (2004) MT1-MMP mediates MUC1 shedding independent of TACE/ADAM17. *Biochem J* 382(Pt 1):363-373.
- [95] Baeckstrom D, Hansson GC, Nilsson O, Johansson C, Gendler SJ, Lindholm L (1991) Purification and characterization of a membrane-bound and a secreted mucin-type glycoprotein carrying the carcinoma-associated sialyl-Lea epitope on distinct core proteins. *J Biol Chem* 266(32):21537-21547.
- [96] Cao Y, Blohm D, Ghadimi BM, Stosiek P, Xing PX, Karsten U (1997) Mucins (MUC1 and MUC3) of gastrointestinal and breast epithelia reveal different and heterogeneous tumor-associated aberrations in glycosylation. *J Histochem Cytochem* 45(11):1547-1557.
- [97] Corfield AP, Myerscough N, Gough M, Brockhausen I, Schauer R, Paraskeva C (1995) Glycosylation patterns of mucins in colonic disease. *Biochem Soc Trans* 23(4):840-845.
- [98] Lloyd KO, Burchell J, Kudryashov V, Yin BW, Taylor-Papadimitriou J (1996) Comparison of O-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells. *J Biol Chem* 271(52):33325-33334.

- [99] Reis CA, David L, Seixas M, Burchell J, Sobrinho-Simoes M (1998) Expression of fully and under-glycosylated forms of MUC1 mucin in gastric carcinoma. *Int J Cancer* 79(4):402-410.
- [100] Fernandez-Rodriguez J, Dwir O, Alon R, Hansson GC (2001) Tumor cell MUC1 and CD43 are glycosylated differently with sialyl-Lewis a and x epitopes and show variable interactions with E-selectin under physiological flow conditions. *Glycoconj J* 18(11-12):925-930.
- [101] Saeland E, Belo AI, Mongera S, van Die I, Meijer GA, van Kooyk Y (2012) Differential glycosylation of MUC1 and CEACAM5 between normal mucosa and tumour tissue of colon cancer patients. *Int J Cancer* 131(1):117-128.
- [102] Eckhardt AE, Timpl CS, DeLuca AW, Hill RL (1997) The complete cDNA sequence and structural polymorphism of the polypeptide chain of porcine submaxillary mucin. *J Biol Chem* 272(52):33204-33210.
- [103] Tytgat KM, Klomp LW, Bovelandt FJ, Opdam FJ, Van der Wurff A, Einerhand AW, Buller HA, Strous GJ, Dekker J (1995) Preparation of anti-mucin polypeptide antisera to study mucin biosynthesis. *Anal Biochem* 226(2):331-341.
- [104] Gum JR, Jr., Hicks JW, Toribara NW, Rothe EM, Lagace RE, Kim YS (1992) The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. *J Biol Chem* 267(30):21375-21383.
- [105] Godl K, Johansson ME, Lidell ME, Morgelin M, Karlsson H, Olson FJ, Gum JR, Jr., Kim YS, Hansson GC (2002) The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *J Biol Chem* 277(49):47248-47256.
- [106] Ambort D, van der Post S, Johansson ME, Mackenzie J, Thomsson E, Krenzel U, Hansson GC (2011) Function of the CysD domain of the gel-forming MUC2 mucin. *Biochem J* 436(1):61-70.
- [107] Ambort D, Johansson ME, Gustafsson JK, Ermund A, Hansson GC (2012) Perspectives on mucus properties and formation--lessons from the biochemical world. *Cold Spring Harbor perspectives in medicine* 2(11).
- [108] Ambort D, Johansson ME, Gustafsson JK, Nilsson HE, Ermund A, Johansson BR, Koeck PJ, Hebert H, Hansson GC (2012) Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proc Natl Acad Sci U S A* 109(15):5645-5650.
- [109] Huang RH, Wang Y, Roth R, Yu X, Purvis AR, Heuser JE, Egelman EH, Sadler JE (2008) Assembly of Weibel-Palade body-like tubules from N-terminal domains of von Willebrand factor. *Proc Natl Acad Sci U S A* 105(2):482-487.
- [110] Zhou YF, Eng ET, Nishida N, Lu C, Walz T, Springer TA (2011) A pH-regulated dimeric bouquet in the structure of von Willebrand factor. *EMBO J* 30(19):4098-4111.
- [111] Axelsson MA, Asker N, Hansson GC (1998) O-glycosylated MUC2 monomer and dimer from LS 174T cells are water-soluble, whereas larger MUC2 species formed early

- during biosynthesis are insoluble and contain nonreducible intermolecular bonds. *J Biol Chem* 273(30):18864-18870.
- [112] Gum JR, Byrd JC, Hicks JW, Toribara NW, Lampion DT, Kim YS (1989) Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence for genetic polymorphism. *J Biol Chem* 264(11):6480-6487.
- [113] Larsson JM, Karlsson H, Sjoval H, Hansson GC (2009) A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* 19(7):756-766.
- [114] Robbe C, Capon C, Coddeville B, Michalski JC (2004) Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* 384(Pt 2):307-316.
- [115] Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, Michalski JC (2003) Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *J Biol Chem* 278(47):46337-46348.
- [116] Holmen-Larsson JM, Thomsson KA, Rodriguez-Pineiro AM, Karlsson H, Hansson GC (2013) Gastrointestinal Muc5ac and Muc2 mucin O-glycan pattern reveal a regio-specific distribution - 3. Studies of mucus in mouse stomach, small intestine, and colon. *Am J Physiol Gastrointest Liver Physiol*.
- [117] Ismail MN, Stone EL, Panico M, Lee SH, Luu Y, Ramirez K, Ho SB, Fukuda M, Marth JD, Haslam SM, Dell A (2011) High-sensitivity O-glycomic analysis of mice deficient in core 2 {beta}1,6-N-acetylglucosaminyltransferases. *Glycobiology* 21(1):82-98.
- [118] Capon C, Maes E, Michalski JC, Leffler H, Kim YS (2001) Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochem J* 358(Pt 3):657-664.
- [119] Podolsky DK (1985) Oligosaccharide structures of isolated human colonic mucin species. *J Biol Chem* 260(29):15510-15515.
- [120] Podolsky DK (1985) Oligosaccharide structures of human colonic mucin. *J Biol Chem* 260(14):8262-8271.
- [121] Corfield AP, Wagner SA, Clamp JR, Kriaris MS, Hoskins LC (1992) Mucin degradation in the human colon: production of sialidase, sialate O-acetyltransferase, N-acetylneuraminidase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infect Immun* 60(10):3971-3978.
- [122] Robertson AM & Wright DP (1997) Bacterial glycosulphatases and sulphomucin degradation. *Can J Gastroenterol* 11(4):361-366.
- [123] Amerongen AVN, Bolscher JGM, Bloemena E, Veerman ECI (1998) Sulfomucins in the human body. *Biological Chemistry* 379(1):1-18.
- [124] Derrien M, Collado MC, Ben-Amor K, Salminen S, de Vos WM (2008) The Mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. *Appl Environ Microbiol* 74(5):1646-1648.

- [125] Marcobal A, Southwick AM, Earle KA, Sonnenburg JL (2013) A refined palate: Bacterial consumption of host glycans in the gut. *Glycobiology* 23(9):1038-1046.
- [126] Png CW, Linden SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin TH (2010) Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* 105(11):2420-2428.
- [127] Pokusaeva K, Fitzgerald GF, van Sinderen D (2011) Carbohydrate metabolism in Bifidobacteria. *Genes & nutrition* 6(3):285-306.
- [128] Turrone F, Milani C, van Sinderen D, Ventura M (2011) Genetic strategies for mucin metabolism in Bifidobacterium bifidum PRL2010: an example of possible human-microbe co-evolution. *Gut microbes* 2(3):183-189.
- [129] An G, Wei B, Xia B, McDaniel JM, Ju T, Cummings RD, Braun J, Xia L (2007) Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *J Exp Med* 204(6):1417-1429.
- [130] Fu J, Wei B, Wen T, Johansson ME, Liu X, Bradford E, Thomsson KA, McGee S, Mansour L, Tong M, McDaniel JM, Sferra TJ, Turner JR, Chen H, Hansson GC, Braun J, Xia L (2011) Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in mice. *J Clin Invest* 121(4):1657-1666.
- [131] Kline KA, Falker S, Dahlberg S, Normark S, Henriques-Normark B (2009) Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* 5(6):580-592.
- [132] Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, Van Maele L, Sirard JC, Mueller AJ, Heikenwalder M, Macpherson AJ, Strugnell R, von Mering C, Hardt WD (2010) The Microbiota Mediates Pathogen Clearance from the Gut Lumen after Non-Typhoidal Salmonella Diarrhea. *Plos Pathogens* 6(9).
- [133] Hooper LV, Midtvedt T, Gordon JI (2002) How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr* 22:283-307.
- [134] Sansonetti PJ (2004) War and peace at mucosal surfaces. *Nat Rev Immunol* 4(12):953-964.
- [135] Savage DC (1986) Gastrointestinal microflora in mammalian nutrition. *Annu Rev Nutr* 6:155-178.
- [136] Ayres JS, Trinidad NJ, Vance RE (2012) Lethal inflammasome activation by a multidrug-resistant pathobiont upon antibiotic disruption of the microbiota. *Nat Med* 18(5):799-806.
- [137] Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* 104(34):13780-13785.
- [138] Round JL & Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9(5):313-323.
- [139] Sommer F & Backhed F (2013) The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol* 11(4):227-238.

- [140] Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, Parameswaran P, Crowell MD, Wing R, Rittmann BE, Krajmalnik-Brown R (2009) Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A* 106(7):2365-2370.
- [141] Lebeer S, Vanderleyden J, De Keersmaecker SC (2010) Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nat Rev Microbiol* 8(3):171-184.
- [142] Tannock GW (1995) *Normal microflora: an introduction to microbes inhabiting the human body* (Chapman and Hall, London, UK).
- [143] Hopkins MJ, Sharp R, Macfarlane GT (2001) Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* 48(2):198-205.
- [144] Mackie RI, Sghir A, Gaskins HR (1999) Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* 69(5):1035S-1045S.
- [145] Mariat D, Firmesse O, Levenez F, Guimaraes V, Sokol H, Dore J, Corthier G, Furet JP (2009) The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol* 9:123.
- [146] Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312(5778):1355-1359.
- [147] Hao WL & Lee YK (2004) Microflora of the gastrointestinal tract: a review. *Methods Mol Biol* 268:491-502.
- [148] Sekirov I, Russell SL, Antunes LC, Finlay BB (2010) Gut microbiota in health and disease. *Physiological reviews* 90(3):859-904.
- [149] Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) Diversity of the human intestinal microbial flora. *Science* 308(5728):1635-1638.
- [150] Ley RE, Peterson DA, Gordon JI (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124(4):837-848.
- [151] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Meta HITC, Bork P, Ehrlich SD, Wang J (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59-65.
- [152] Zoetendal EG, Booijink CC, Klaassens ES, Heilig HG, Kleerebezem M, Smidt H, de Vos WM (2006) Isolation of RNA from bacterial samples of the human gastrointestinal tract. *Nat Protoc* 1(2):954-959.

- [153] Turrone F, Ribbera A, Foroni E, van Sinderen D, Ventura M (2008) Human gut microbiota and bifidobacteria: from composition to functionality. *Antonie Van Leeuwenhoek* 94(1):35-50.
- [154] Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP, Ugarte E, Munoz-Tamayo R, Paslier DL, Nalin R, Dore J, Leclerc M (2009) Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* 11(10):2574-2584.
- [155] Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI (2009) A core gut microbiome in obese and lean twins. *Nature* 457(7228):480-484.
- [156] Hold GL, Pryde SE, Russell VJ, Furrie E, Flint HJ (2002) Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. *FEMS Microbiol Ecol* 39(1):33-39.
- [157] Wang X, Heazlewood SP, Krause DO, Florin TH (2003) Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J Appl Microbiol* 95(3):508-520.
- [158] Walter J (2005) The microecology of lactobacilli in the gastrointestinal tract. *Probiotics and prebiotics: scientific aspects*, ed Tannock GW (Horizon Scientific Press, Wymondham), pp 51-82.
- [159] Molin G, Jeppsson B, Johansson ML, Ahrne S, Nobaek S, Stahl M, Bengmark S (1993) Numerical taxonomy of *Lactobacillus* spp. associated with healthy and diseased mucosa of the human intestines. *J Appl Bacteriol* 74(3):314-323.
- [160] Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM (2002) Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* 68(7):3401-3407.
- [161] Fioramonti J, Theodorou V, Bueno L (2003) Probiotics: what are they? What are their effects on gut physiology? *Best Pract Res Clin Gastroenterol* 17(5):711-724.
- [162] Walter J (2008) Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol* 74(16):4985-4996.
- [163] Araya M, Morelli L, Reid G, Sanders ME, Stanton C (2002) Guidelines for the Evaluation of Probiotics in Food. (Food and Agriculture Organization of the United Nations and World Health Organization, London, Ontario, Canada).
- [164] Collado MC, Grzeskowiak L, Salminen S (2007) Probiotic strains and their combination inhibit in vitro adhesion of pathogens to pig intestinal mucosa. *Curr Microbiol* 55(3):260-265.
- [165] Collado MC, Gueimonde M, Hernandez M, Sanz Y, Salminen S (2005) Adhesion of selected *Bifidobacterium* strains to human intestinal mucus and the role of adhesion in enteropathogen exclusion. *J Food Prot* 68(12):2672-2678.

- [166] Lee YK, Puong KY, Ouwehand AC, Salminen S (2003) Displacement of bacterial pathogens from mucus and Caco-2 cell surface by lactobacilli. *J Med Microbiol* 52(Pt 10):925-930.
- [167] Ma YL, Guo T, Xu ZR, You P, Ma JF (2006) Effect of *Lactobacillus* isolates on the adhesion of pathogens to chicken intestinal mucus in vitro. *Lett Appl Microbiol* 42(4):369-374.
- [168] Tuomola EM, Ouwehand AC, Salminen SJ (1999) The effect of probiotic bacteria on the adhesion of pathogens to human intestinal mucus. *FEMS Immunol Med Microbiol* 26(2):137-142.
- [169] Marteau PR, de Vrese M, Cellier CJ, Schrezenmeir J (2001) Protection from gastrointestinal diseases with the use of probiotics. *Am J Clin Nutr* 73(2 Suppl):430S-436S.
- [170] Van Niel CW, Feudtner C, Garrison MM, Christakis DA (2002) *Lactobacillus* therapy for acute infectious diarrhea in children: a meta-analysis. *Pediatrics* 109(4):678-684.
- [171] Lebeer S, Vanderleyden J, De Keersmaecker SC (2008) Genes and molecules of lactobacilli supporting probiotic action. *Microbiol Mol Biol Rev* 72(4):728-764, Table of Contents.
- [172] Hayashi H, Takahashi R, Nishi T, Sakamoto M, Benno Y (2005) Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol* 54(Pt 11):1093-1101.
- [173] Reuter G (2001) The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Current issues in intestinal microbiology* 2(2):43-53.
- [174] Savage DC (1977) Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* 31:107-133.
- [175] Walter J, Britton RA, Roos S (2011) Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. *Proc Natl Acad Sci U S A* 108 Suppl 1:4645-4652.
- [176] Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, Walter J (2010) Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *ISME J* 4(3):377-387.
- [177] Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, Oh PL, Heng NC, Patil PB, Juge N, Mackenzie DA, Pearson BM, Lapidus A, Dalin E, Tice H, Goltsman E, Land M, Hauser L, Ivanova N, Kyrpides NC, Walter J (2011) The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genet* 7(2):e1001314.
- [178] Frese SAH, R.W; Watler J. (2012) Comparison of the colonization ability of autochthonous and allochthonous strains of lactobacilli in th human gastrointestinal tract. *Advances in Microbiology* 2:399-409.

- [179] Derrien M, van Passel MW, van de Bovenkamp JH, Schipper RG, de Vos WM, Dekker J (2010) Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut microbes* 1(4):254-268.
- [180] El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* 11(7):497-504.
- [181] Kleerebezem M & Vaughan EE (2009) Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity. *Annu Rev Microbiol* 63:269-290.
- [182] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* 489(7415):220-230.
- [183] Marco ML, Peters TH, Bongers RS, Molenaar D, van Hemert S, Sonnenburg JL, Gordon JI, Kleerebezem M (2009) Lifestyle of *Lactobacillus plantarum* in the mouse caecum. *Environ Microbiol* 11(10):2747-2757.
- [184] Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. *J Biol Chem* 284(37):24673-24677.
- [185] Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* 9(12):e1001221.
- [186] Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299(5615):2074-2076.
- [187] Xu J & Gordon JI (2003) Honor thy symbionts. *Proc Natl Acad Sci U S A* 100(18):10452-10459.
- [188] Crost EH, Tailford LE, Le Gall G., Fons M., Henrissat B., Juge N. (2013) Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *Plos One* (in press).
- [189] Derrien M, Vaughan EE, Plugge CM, de Vos WM (2004) *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* 54(Pt 5):1469-1476.
- [190] Cherrington CA, Hinton M, Pearson GR, Chopra I (1991) Short-chain organic acids at pH 5.0 kill *Escherichia coli* and *Salmonella* spp. without causing membrane perturbation. *J Appl Bacteriol* 70(2):161-165.
- [191] Gantois I, Ducatelle R, Pasmans F, Haesebrouck F, Hautefort I, Thompson A, Hinton JC, Van Immerseel F (2006) Butyrate specifically down-regulates salmonella pathogenicity island 1 gene expression. *Appl Environ Microbiol* 72(1):946-949.
- [192] Leatham MP, Banerjee S, Autieri SM, Mercado-Lubo R, Conway T, Cohen PS (2009) Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157:H7 growth in the streptomycin-treated mouse intestine. *Infect Immun* 77(7):2876-2886.

- [193] Naaber P, Smidt I, Stsepetova J, Brilene T, Annuk H, Mikelsaar M (2004) Inhibition of *Clostridium difficile* strains by intestinal *Lactobacillus* species. *J Med Microbiol* 53(Pt 6):551-554.
- [194] Pacheco AR, Curtis MM, Ritchie JM, Munera D, Waldor MK, Moreira CG, Sperandio V (2012) Fucose sensing regulates bacterial intestinal colonization. *Nature* 492(7427):113-117.
- [195] Makivuokko H, Lahtinen SJ, Wacklin P, Tuovinen E, Tenkanen H, Nikkila J, Bjorklund M, Aranko K, Ouwehand AC, Matto J (2012) Association between the ABO blood group and the human intestinal microbiota composition. *BMC Microbiol* 12:94.
- [196] Rausch P, Rehman A, Kunzel S, Hasler R, Ott SJ, Schreiber S, Rosenstiel P, Franke A, Baines JF (2011) Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci U S A* 108(47):19030-19035.
- [197] Staubach F, Kunzel S, Baines AC, Yee A, McGee BM, Backhed F, Baines JF, Johnsen JM (2012) Expression of the blood-group-related glycosyltransferase B4galnt2 influences the intestinal microbiota in mice. *ISME J* 6(7):1345-1355.
- [198] Wacklin P, Makivuokko H, Alakulppi N, Nikkila J, Tenkanen H, Rabina J, Partanen J, Aranko K, Matto J (2011) Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. *PLoS One* 6(5):e20113.
- [199] Corfield AP, Myerscough N, Longman R, Sylvester P, Arul S, Pignatelli M (2000) Mucins and mucosal protection in the gastrointestinal tract: new prospects for mucins in the pathology of gastrointestinal disease. *Gut* 47(4):589-594.
- [200] Longman RJ, Poulosom R, Corfield AP, Warren BF, Wright NA, Thomas MG (2006) Alterations in the composition of the supramucosal defense barrier in relation to disease severity of ulcerative colitis. *J Histochem Cytochem* 54(12):1335-1348.
- [201] Tytgat KM, van der Wal JW, Einerhand AW, Buller HA, Dekker J (1996) Quantitative analysis of MUC2 synthesis in ulcerative colitis. *Biochem Biophys Res Commun* 224(2):397-405.
- [202] Weiss AA, Babyatsky MW, Ogata S, Chen A, Itzkowitz SH (1996) Expression of MUC2 and MUC3 mRNA in human normal, malignant, and inflammatory intestinal tissues. *J Histochem Cytochem* 44(10):1161-1166.
- [203] Bry L, Falk PG, Midtvedt T, Gordon JI (1996) A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273(5280):1380-1383.
- [204] Li JD, Feng W, Gallup M, Kim JH, Gum J, Kim Y, Basbaum C (1998) Activation of NF-kappaB via a Src-dependent Ras-MAPK-pp90rsk pathway is required for *Pseudomonas aeruginosa*-induced mucin overproduction in epithelial cells. *Proc Natl Acad Sci U S A* 95(10):5718-5723.
- [205] Lievin-Le Moal V, Servin AL, Coconnier-Polter MH (2005) The increase in mucin exocytosis and the upregulation of MUC genes encoding for membrane-bound mucins

- induced by the thiol-activated exotoxin listeriolysin O is a host cell defence response that inhibits the cell-entry of *Listeria monocytogenes*. *Cell Microbiol* 7(7):1035-1048.
- [206] Linden SK, Florin TH, McGuckin MA (2008) Mucin dynamics in intestinal bacterial infection. *PLoS One* 3(12):e3952.
- [207] Linden SK, Sheng YH, Every AL, Miles KM, Skoog EC, Florin TH, Sutton P, McGuckin MA (2009) MUC1 limits *Helicobacter pylori* infection both by steric hindrance and by acting as a releasable decoy. *PLoS Pathog* 5(10):e1000617.
- [208] Celli JP, Turner BS, Afdhal NH, Keates S, Ghiran I, Kelly CP, Ewoldt RH, McKinley GH, So P, Erramilli S, Bansil R (2009) *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. *Proc Natl Acad Sci U S A* 106(34):14321-14326.
- [209] Lidell ME, Moncada DM, Chadee K, Hansson GC (2006) *Entamoeba histolytica* cysteine proteases cleave the MUC2 mucin in its C-terminal domain and dissolve the protective colonic mucus gel. *Proc Natl Acad Sci U S A* 103(24):9298-9303.
- [210] Subramani DB, Johansson ME, Dahlen G, Hansson GC (2010) *Lactobacillus* and *Bifidobacterium* species do not secrete protease that cleaves the MUC2 mucin which organises the colon mucus. *Benef Microbes* 1(4):343-350.
- [211] Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR, Eng JK, Akira S, Underhill DM, Aderem A (2001) The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* 410(6832):1099-1103.
- [212] Pizarro-Cerda J & Cossart P (2006) Bacterial adhesion and entry into host cells. *Cell* 124(4):715-727.
- [213] Smits HH, Engering A, van der Kleij D, de Jong EC, Schipper K, van Capel TM, Zaat BA, Yazdanbakhsh M, Wierenga EA, van Kooyk Y, Kapsenberg ML (2005) Selective probiotic bacteria induce IL-10-producing regulatory T cells in vitro by modulating dendritic cell function through dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin. *J Allergy Clin Immunol* 115(6):1260-1267.
- [214] Voltan S, Martines D, Elli M, Brun P, Longo S, Porzionato A, Macchi V, D'Inca R, Scarpa M, Palu G, Sturniolo GC, Morelli L, Castagliuolo I (2008) *Lactobacillus crispatus* M247-derived H<sub>2</sub>O<sub>2</sub> acts as a signal transducing molecule activating peroxisome proliferator activated receptor-gamma in the intestinal mucosa. *Gastroenterology* 135(4):1216-1227.
- [215] Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA (2012) Structural basis of TLR5-flagellin recognition and signaling. *Science* 335(6070):859-864.
- [216] Zeng H, Carlson AQ, Guo Y, Yu Y, Collier-Hyams LS, Madara JL, Gewirtz AT, Neish AS (2003) Flagellin is the major proinflammatory determinant of enteropathogenic *Salmonella*. *J Immunol* 171(7):3668-3674.
- [217] Bos R, van der Mei HC, Busscher HJ (1999) Physico-chemistry of initial microbial adhesive interactions--its mechanisms and methods for study. *FEMS Microbiol Rev* 23(2):179-230.

- [218] Mozes N, Marchal F, Hermesse MP, Van Haecht JL, Reuliaux L, Leonard AJ, Rouxhet PG (1987) Immobilization of microorganisms by adhesion: interplay of electrostatic and nonelectrostatic interactions. *Biotechnol Bioeng* 30(3):439-450.
- [219] van der Mei HC, de Vries J, Busscher HJ (1993) Hydrophobic and electrostatic cell surface properties of thermophilic dairy streptococci. *Appl Environ Microbiol* 59(12):4305-4312.
- [220] van Loosdrecht MC, Lyklema J, Norde W, Schraa G, Zehnder AJ (1987) The role of bacterial cell wall hydrophobicity in adhesion. *Appl Environ Microbiol* 53(8):1893-1897.
- [221] Busscher HJ, Bos R, van der Mei HC, Handley PS (2000) Physicochemistry of microbial adhesion from an overall approach to the limits. *Physical chemistry of biological surfaces.*, ed Baszkin A. NWNNew York, USA), pp 431-458.
- [222] Jucker BA, Harms H, Zehnder AJ (1996) Adhesion of the positively charged bacterium *Stenotrophomonas (Xanthomonas) maltophilia* 70401 to glass and Teflon. *J Bacteriol* 178(18):5472-5479.
- [223] Vadillo-Rodriguez V, Busscher HJ, Norde W, de Vries J, van der Mei HC (2004) Dynamic cell surface hydrophobicity of Lactobacillus strains with and without surface layer proteins. *J Bacteriol* 186(19):6647-6650.
- [224] Beachey EH (1981) Bacterial adherence: adhesin-receptor interactions mediating the attachment of bacteria to mucosal surface. *J Infect Dis* 143(3):325-345.
- [225] Navarre WW & Schneewind O (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63(1):174-229.
- [226] Patti JM, Allen BL, McGavin MJ, Hook M (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* 48:585-617.
- [227] Proft T & Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66(4):613-635.
- [228] Tasteyre A, Barc MC, Collignon A, Boureau H, Karjalainen T (2001) Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun* 69(12):7937-7940.
- [229] Arora SK, Ritchings BW, Almira EC, Lory S, Ramphal R (1998) The *Pseudomonas aeruginosa* flagellar cap protein, FliD, is responsible for mucin adhesion. *Infect Immun* 66(3):1000-1007.
- [230] Lillehoj EP, Kim BT, Kim KC (2002) Identification of *Pseudomonas aeruginosa* flagellin as an adhesin for Muc1 mucin. *Am J Physiol Lung Cell Mol Physiol* 282(4):L751-756.
- [231] Erdem AL, Avelino F, Xicohtencatl-Cortes J, Giron JA (2007) Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. *J Bacteriol* 189(20):7426-7435.
- [232] Sajjan SU & Forstner JF (1990) Characteristics of binding of *Escherichia coli* serotype O157:H7 strain CL-49 to purified intestinal mucin. *Infect Immun* 58(4):860-867.

- [233] Sajjan SU & Forstner JF (1990) Role of the putative "link" glycopeptide of intestinal mucin in binding of piliated *Escherichia coli* serotype O157:H7 strain CL-49. *Infect Immun* 58(4):868-873.
- [234] Bouckaert J, Berglund J, Schembri M, De Genst E, Cools L, Wuhrer M, Hung CS, Pinkner J, Slattegard R, Zavialov A, Choudhury D, Langermann S, Hultgren SJ, Wyns L, Klemm P, Oscarson S, Knight SD, De Greve H (2005) Receptor binding studies disclose a novel class of high-affinity inhibitors of the *Escherichia coli* FimH adhesin. *Mol Microbiol* 55(2):441-455.
- [235] Hung CS, Bouckaert J, Hung D, Pinkner J, Widberg C, DeFusco A, Auguste CG, Strouse R, Langermann S, Waksman G, Hultgren SJ (2002) Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Mol Microbiol* 44(4):903-915.
- [236] Kukkonen M, Raunio T, Virkola R, Lahteenmaki K, Makela PH, Klemm P, Clegg S, Korhonen TK (1993) Basement membrane carbohydrate as a target for bacterial adhesion: binding of type I fimbriae of *Salmonella enterica* and *Escherichia coli* to laminin. *Mol Microbiol* 7(2):229-237.
- [237] Buts L, Bouckaert J, De Genst E, Loris R, Oscarson S, Lahmann M, Messens J, Brosens E, Wyns L, De Greve H (2003) The fimbrial adhesin F17-G of enterotoxigenic *Escherichia coli* has an immunoglobulin-like lectin domain that binds N-acetylglucosamine. *Mol Microbiol* 49(3):705-715.
- [238] Mouricout M, Milhavet M, Durie C, Grange P (1995) Characterization of glycoprotein glycan receptors for *Escherichia coli* F17 fimbrial lectin. *Microb Pathog* 18(4):297-306.
- [239] Bumbaca D, Littlejohn JE, Nayakanti H, Lucas AH, Rigden DJ, Galperin MY, Jedrzejak MJ (2007) Genome-based identification and characterization of a putative mucin-binding protein from the surface of *Streptococcus pneumoniae*. *Proteins-Structure Function and Bioinformatics* 66(3):547-558.
- [240] Du Y, He YX, Zhang ZY, Yang YH, Shi WW, Frolet C, Guilmi AM, Vernet T, Zhou CZ, Chen Y (2010) Crystal structure of the mucin-binding domain of Spr1345 from *Streptococcus pneumoniae*. *J Struct Biol*.
- [241] Ilver D, Arnqvist A, Ogren J, Frick IM, Kersulyte D, Incecik ET, Berg DE, Covacci A, Engstrand L, Boren T (1998) *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279(5349):373-377.
- [242] Linden S, Nordman H, Hedenbro J, Hurtig M, Boren T, Carlstedt I (2002) Strain- and blood group-dependent binding of *Helicobacter pylori* to human gastric MUC5AC glycoforms. *Gastroenterology* 123(6):1923-1930.
- [243] Linden SK, Wickstrom C, Lindell G, Gilshenan K, Carlstedt I (2008) Four modes of adhesion are used during *Helicobacter pylori* binding to human mucins in the oral and gastric niches. *Helicobacter* 13(2):81-93.
- [244] Mahdavi J, Sonden B, Hurtig M, Olfat FO, Forsberg L, Roche N, Angstrom J, Larsson T, Teneberg S, Karlsson KA, Altraja S, Wadstrom T, Kersulyte D, Berg DE, Dubois A,

- Petersson C, Magnusson KE, Norberg T, Lindh F, Lundskog BB, Arnqvist A, Hammarstrom L, Boren T (2002) *Helicobacter pylori* SabA adhesin in persistent infection and chronic inflammation. *Science* 297(5581):573-578.
- [245] Bhowmick R, Ghosal A, Das B, Koley H, Saha DR, Ganguly S, Nandy RK, Bhadra RK, Chatterjee NS (2008) Intestinal Adherence of *Vibrio cholerae* Involves a Coordinated Interaction between Colonization Factor GbpA and Mucin. *Infection and Immunity* 76(11):4968-4977.
- [246] Chadee K, Petri WA, Jr., Innes DJ, Ravdin JI (1987) Rat and human colonic mucins bind to and inhibit adherence lectin of *Entamoeba histolytica*. *J Clin Invest* 80(5):1245-1254.
- [247] Petri WA, Jr., Smith RD, Schlesinger PH, Murphy CF, Ravdin JI (1987) Isolation of the galactose-binding lectin that mediates the in vitro adherence of *Entamoeba histolytica*. *J Clin Invest* 80(5):1238-1244.
- [248] Ravdin JI, Murphy CF, Salata RA, Guerrant RL, Hewlett EL (1985) N-Acetyl-D-galactosamine-inhibitable adherence lectin of *Entamoeba histolytica*. I. Partial purification and relation to amoebic virulence in vitro. *J Infect Dis* 151(5):804-815.
- [249] Linden SK, Bierne H, Sabet C, Png CW, Florin TH, McGuckin MA, Cossart P (2008) *Listeria monocytogenes* internalins bind to the human intestinal mucin MUC2. *Archives of Microbiology* 190(1):101-104.
- [250] Josenhans C & Suerbaum S (2002) The role of motility as a virulence factor in bacteria. *Int J Med Microbiol* 291(8):605-614.
- [251] Moens S & Vanderleyden J (1996) Functions of bacterial flagella. *Critical reviews in microbiology* 22(2):67-100.
- [252] Pouttu R, Puustinen T, Virkola R, Hacker J, Klemm P, Korhonen TK (1999) Amino acid residue Ala-62 in the FimH fimbrial adhesin is critical for the adhesiveness of meningitis-associated *Escherichia coli* to collagens. *Mol Microbiol* 31(6):1747-1757.
- [253] Wu XR, Sun TT, Medina JJ (1996) In vitro binding of type 1-fimbriated *Escherichia coli* to uroplakins Ia and Ib: relation to urinary tract infections. *Proc Natl Acad Sci U S A* 93(18):9630-9635.
- [254] Choudhury D, Thompson A, Stojanoff V, Langermann S, Pinkner J, Hultgren SJ, Knight SD (1999) X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* 285(5430):1061-1066.
- [255] Chessa D, Dorsey CW, Winter M, Baumler AJ (2008) Binding specificity of *Salmonella* plasmid-encoded fimbriae assessed by glycomics. *J Biol Chem* 283(13):8118-8124.
- [256] Amano J & Oshima M (1999) Expression of the H type 1 blood group antigen during enterocytic differentiation of Caco-2 cells. *J Biol Chem* 274(30):21209-21216.
- [257] Chessa D, Winter MG, Jakomin M, Baumler AJ (2009) *Salmonella enterica* serotype Typhimurium Std fimbriae bind terminal alpha(1,2)fucose residues in the cecal mucosa. *Mol Microbiol* 71(4):864-875.

- [258] Danne C & Dramsi S (2012) Pili of gram-positive bacteria: roles in host colonization. *Res Microbiol* 163(9-10):645-658.
- [259] Evans CM & Koo JS (2009) Airway mucus: the good, the bad, the sticky. *Pharmacology & therapeutics* 121(3):332-348.
- [260] Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, Kojadinovic M, Lalioui L, Poyart C, Trieu-Cuot P (2006) Assembly and role of pili in group B streptococci. *Mol Microbiol* 60(6):1401-1413.
- [261] Krishnan V, Gaspar AH, Ye N, Mandlik A, Ton-That H, Narayana SV (2007) An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion. *Structure* 15(8):893-903.
- [262] Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol* 16(1):33-40.
- [263] Manetti AG, Zingaretti C, Falugi F, Capo S, Bombaci M, Bagnoli F, Gambellini G, Bensi G, Mora M, Edwards AM, Musser JM, Graviss EA, Telford JL, Grandi G, Margarit I (2007) *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Mol Microbiol* 64(4):968-983.
- [264] Magalhaes A & Reis CA (2010) *Helicobacter pylori* adhesion to gastric epithelial cells is mediated by glycan receptors. *Braz J Med Biol Res* 43(7):611-618.
- [265] Marcos NT, Magalhaes A, Ferreira B, Oliveira MJ, Carvalho AS, Mendes N, Gilmartin T, Head SR, Figueiredo C, David L, Santos-Silva F, Reis CA (2008) *Helicobacter pylori* induces beta3GnT5 in human gastric cell lines, modulating expression of the SabA ligand sialyl-Lewis x. *J Clin Invest* 118(6):2325-2336.
- [266] Ravdin JI & Guerrant RL (1981) Role of adherence in cytopathogenic mechanisms of *Entamoeba histolytica*. Study with mammalian tissue culture cells and human erythrocytes. *J Clin Invest* 68(5):1305-1313.
- [267] Attridge SR, Manning PA, Holmgren J, Jonson G (1996) Relative significance of mannose-sensitive hemagglutinin and toxin-coregulated pili in colonization of infant mice by *Vibrio cholerae* El Tor. *Infect Immun* 64(8):3369-3373.
- [268] Yamamoto T, Kamano T, Uchimura M, Iwanaga M, Yokota T (1988) *Vibrio cholerae* O1 adherence to villi and lymphoid follicle epithelium: in vitro model using formalin-treated human small intestine and correlation between adherence and cell-associated hemagglutinin levels. *Infect Immun* 56(12):3241-3250.
- [269] Bierne H & Cossart P (2007) *Listeria monocytogenes* surface proteins: from genome predictions to function. *Microbiol Mol Biol Rev* 71(2):377-397.
- [270] Gaillard JL, Berche P, Frehel C, Gouin E, Cossart P (1991) Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. *Cell* 65(7):1127-1141.

- [271] Salminen S, Isolauri E, Salminen E (1996) Clinical uses of probiotics for stabilizing the gut mucosal barrier: successful strains and future challenges. *Antonie Van Leeuwenhoek* 70(2-4):347-358.
- [272] Kirjavainen PV, Ouwehand AC, Isolauri E, Salminen SJ (1998) The ability of probiotic bacteria to bind to human intestinal mucus. *FEMS Microbiol Lett* 167(2):185-189.
- [273] Ouwehand AC, Kirjavainen PV, Shortt C, Salminen S (1999) Probiotics: mechanisms and established effects. *Int Dairy J* 9(1):43-52.
- [274] Ouwehand AC, Tuomola EM, Tolkkio S, Salminen S (2001) Assessment of adhesion properties of novel probiotic strains to human intestinal mucus. *Int J Food Microbiol* 64(1-2):119-126.
- [275] Van Tassell ML & Miller MJ (2011) *Lactobacillus* adhesion to mucus. *Nutrients* 3(5):613-636.
- [276] Dague E, Le DT, Zanna S, Marcus P, Loubiere P, Mercier-Bonin M (2010) Probing in vitro interactions between *Lactococcus lactis* and mucins using AFM. *Langmuir* 26(13):11010-11017.
- [277] Huang JY, Lee SM, Mazmanian SK (2011) The human commensal *Bacteroides fragilis* binds intestinal mucin. *Anaerobe* 17(4):137-141.
- [278] Uchida H, Fujitani K, Kawai Y, Kitazawa H, Horii A, Shiiba K, Saito K, Saito T (2004) A new assay using surface plasmon resonance (SPR) to determine binding of the *Lactobacillus acidophilus* group to human colonic mucin. *Biosci Biotechnol Biochem* 68(5):1004-1010.
- [279] Heavens D, Tailford LE, Crossman L, Jeffers F, Mackenzie DA, Caccamo M, Juge N (2011) Genome sequence of the vertebrate gut symbiont *Lactobacillus reuteri* ATCC 53608. *J Bacteriol* 193(15):4015-4016.
- [280] Kleerebezem M, Hols P, Bernard E, Rolain T, Zhou M, Siezen RJ, Bron PA (2010) The extracellular biology of the lactobacilli. *FEMS Microbiol Rev* 34(2):199-230.
- [281] Ventura M, O'Flaherty S, Claesson MJ, Turrone F, Klaenhammer TR, van Sinderen D, O'Toole PW (2009) Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat Rev Microbiol* 7(1):61-71.
- [282] MacKenzie DA, Tailford LE, Hemmings AM, Juge N (2009) Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. *J Biol Chem* 284(47):32444-32453.
- [283] Roos S & Jonsson H (2002) A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. *Microbiology-Sgm* 148:433-442.
- [284] Miyoshi Y, Okada S, Uchimura T, Satoh E (2006) A mucus adhesion promoting protein, MapA, mediates the adhesion of *Lactobacillus reuteri* to Caco-2 human intestinal epithelial cells. *Biosci Biotechnol Biochem* 70(7):1622-1628.

- [285] Rojas M, Ascencio F, Conway PL (2002) Purification and characterization of a surface protein from *Lactobacillus fermentum* 104R that binds to porcine small intestinal mucus and gastric mucin. *Appl Environ Microbiol* 68(5):2330-2336.
- [286] Aleljung P, Shen W, Rozalska B, Hellman U, Ljungh A, Wadstrom T (1994) Purification of collagen-binding proteins of *Lactobacillus reuteri* NCIB 11951. *Curr Microbiol* 28(4):231-236.
- [287] Roos S, Aleljung P, Robert N, Lee B, Wadstrom T, Lindberg M, Jonsson H (1996) A collagen binding protein from *Lactobacillus reuteri* is part of an ABC transporter system? *FEMS Microbiol Lett* 144(1):33-38.
- [288] Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P, Satokari R, Vesterlund S, Hendrickx AP, Lebeer S, De Keersmaecker SC, Vanderleyden J, Hamalainen T, Laukkanen S, Salovuori N, Ritari J, Alatalo E, Korpela R, Mattila-Sandholm T, Lassig A, Hatakka K, Kinnunen KT, Karjalainen H, Saxelin M, Laakso K, Surakka A, Palva A, Salusjarvi T, Auvinen P, de Vos WM (2009) Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proc Natl Acad Sci U S A* 106(40):17193-17198.
- [289] von Ossowski I, Reunanen J, Satokari R, Vesterlund S, Kankainen M, Huhtinen H, Tynkkynen S, Salminen S, de Vos WM, Palva A (2010) Mucosal adhesion properties of the probiotic *Lactobacillus rhamnosus* GG SpaCBA and SpaFED pilin subunits. *Appl Environ Microbiol* 76(7):2049-2057.
- [290] von Ossowski I, Satokari R, Reunanen J, Lebeer S, De Keersmaecker SC, Vanderleyden J, de Vos WM, Palva A (2011) Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* 77(13):4465-4472.
- [291] Meyrand M, Guillot A, Goin M, Furlan S, Armalyte J, Kulakauskas S, Cortez-Perez NG, Thomas G, Chat S, Pechoux C, Dupres V, Hols P, Dufrene YF, Trugnan G, Chapot-Chartier MP (2013) Surface proteome analysis of a natural isolate of *Lactococcus lactis* reveals the presence of pili able to bind human intestinal epithelial cells. *Mol Cell Proteomics*.
- [292] Buck BL, Altermann E, Svingerud T, Klaenhammer TR (2005) Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 71(12):8344-8351.
- [293] Pretzer G, Snel J, Molenaar D, Wiersma A, Bron PA, Lambert J, de Vos WM, van der Meer R, Smits MA, Kleerebezem M (2005) Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J Bacteriol* 187(17):6128-6136.
- [294] Kinoshita H, Wakahara N, Watanabe M, Kawasaki T, Matsuo H, Kawai Y, Kitazawa H, Ohnuma S, Miura K, Horii A, Saito T (2008) Cell surface glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of *Lactobacillus plantarum* LA 318 recognizes human A and B blood group antigens. *Res Microbiol* 159(9-10):685-691.

- [295] Watanabe M, Kinoshita H, Huang IN, Eguchi K, Tsurumi T, Kawai Y, Kitazawa H, Kimura K, Taketomo N, Kikuchi D, Sase T, Miura K, Ogawa H, Shibata C, Horii A, Saito T (2012) An adhesin-like protein, Lam29, from *Lactobacillus mucosae* ME-340 binds to histone H3 and blood group antigens in human colonic mucus. *Biosci Biotechnol Biochem* 76(9):1655-1660.
- [296] Watanabe M, Kinoshita H, Nitta M, Yukishita R, Kawai Y, Kimura K, Taketomo N, Yamazaki Y, Tateno Y, Miura K, Horii A, Kitazawa H, Saito T (2010) Identification of a new adhesin-like protein from *Lactobacillus mucosae* ME-340 with specific affinity to the human blood group A and B antigens. *J Appl Microbiol*.
- [297] Bergonzelli GE, Granato D, Pridmore RD, Marvin-Guy LF, Donnicola D, Corthesy-Theulaz IE (2006) GroEL of *Lactobacillus johnsonii* La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen *Helicobacter pylori*. *Infect Immun* 74(1):425-434.
- [298] Granato D, Bergonzelli GE, Pridmore RD, Marvin L, Rouvet M, Corthesy-Theulaz IE (2004) Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins. *Infect Immun* 72(4):2160-2169.
- [299] van Pijkeren JP, Canchaya C, Ryan KA, Li Y, Claesson MJ, Sheil B, Steidler L, O'Mahony L, Fitzgerald GF, van Sinderen D, O'Toole PW (2006) Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 72(6):4143-4153.
- [300] Macias-Rodriguez ME, Zagorec M, Ascencio F, Vazquez-Juarez R, Rojas M (2009) *Lactobacillus fermentum* BCS87 expresses mucus- and mucin-binding proteins on the cell surface. *J Appl Microbiol* 107(6):1866-1874.
- [301] Marraffini LA, Dedent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* 70(1):192-221.
- [302] MacKenzie DA, Jeffers F, Parker ML, Vibert-Vallet A, Bongaerts RJ, Roos S, Walter J, Juge N (2010) Strain-specific diversity of mucus-binding proteins in the adhesion and aggregation properties of *Lactobacillus reuteri*. *Microbiology-Sgm* 156:3368-3378.
- [303] Coic YM, Baleux F, Poyraz O, Thibeaux R, Labruyere E, Chretien F, Sobhani I, Lazure T, Wyplosz B, Schneider G, Mulard L, Sansonetti PJ, Marteyn BS (2012) Design of a specific colonic mucus marker using a human commensal bacterium cell surface domain. *J Biol Chem* 287(19):15916-15922.
- [304] Boekhorst J, Helmer Q, Kleerebezem M, Siezen RJ (2006) Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria. *Microbiology-Sgm* 152:273-280.
- [305] Walshaw J, Etzold S, Hemmings AM, MacKenzie DA, Juge N (2013) Mucus-binding proteins in human microbiota: a sequence-based characterisation (The Roslin Institute, Edinburgh, UK ).

- [306] Avall-Jaaskelainen S & Palva A (2005) *Lactobacillus* surface layers and their applications. *FEMS Microbiol Rev* 29(3):511-529.
- [307] Forde BM, Neville BA, O'Donnell MM, Riboulet-Bisson E, Claesson MJ, Coghlan A, Ross RP, O'Toole PW (2011) Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. *Microb Cell Fact* 10 Suppl 1:S13.
- [308] Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, Pittet AC, Zwahlen MC, Rouvet M, Altermann E, Barrangou R, Mollet B, Mercenier A, Klaenhammer T, Arigoni F, Schell MA (2004) The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci U S A* 101(8):2512-2517.
- [309] Mandlik A, Swierczynski A, Das A, Ton-That H (2007) *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol Microbiol* 64(1):111-124.
- [310] Kinoshita H, Uchida H, Kawai Y, Kitazawa H, Miura K, Shiiba K, Horii A, Saito T (2007) Quantitative evaluation of adhesion of lactobacilli isolated from human intestinal tissues to human colonic mucin using surface plasmon resonance (BIAcore assay). *Journal of Applied Microbiology* 102(1):116-123.
- [311] Adlerberth I, Ahrne S, Johansson ML, Molin G, Hanson LA, Wold AE (1996) A mannose-specific adherence mechanism in *Lactobacillus plantarum* conferring binding to the human colonic cell line HT-29. *Appl Environ Microbiol* 62(7):2244-2251.
- [312] Gross G, Snel J, Boekhorst J, Smits MA, Kleerebezem M (2010) Biodiversity of mannose-specific adhesion in *Lactobacillus plantarum* revisited: strain-specific domain composition of the mannose-adhesin. *Beneficial Microbes* 1(1):61-66.
- [313] Williams AF & Barclay AN (1988) The immunoglobulin superfamily--domains for cell surface recognition. *Annual review of immunology* 6:381-405.
- [314] Drickamer K (1993) Recognition of complex carbohydrates by Ca(2+)-dependent animal lectins. *Biochem Soc Trans* 21(2):456-459.
- [315] Imberty A, Mitchell EP, Wimmerova M (2005) Structural basis of high-affinity glycan recognition by bacterial and fungal lectins. *Curr Opin Struct Biol* 15(5):525-534.
- [316] Graille M, Stura EA, Housden NG, Beckingham JA, Bottomley SP, Beale D, Taussig MJ, Sutton BJ, Gore MG, Charbonnier JB (2001) Complex between *Peptostreptococcus magnus* protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. *Structure* 9(8):679-687.
- [317] Ebbes M, Blyemuller WM, Cernescu M, Nolker R, Brutschy B, Niemann HH (2011) Fold and function of the InIB B-repeat. *J Biol Chem*.
- [318] Imberty A & Varrot A (2008) Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* 18(5):567-576.
- [319] Niemann HH, Schubert WD, Heinz DW (2004) Adhesins and invasins of pathogenic bacteria: a structural view. *Microbes Infect* 6(1):101-112.

- [320] Bork P, Holm L, Sander C (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* 242(4):309-320.
- [321] Schubert WD, Gobel G, Diepholz M, Darji A, Kloer D, Hain T, Chakraborty T, Wehland J, Domann E, Heinz DW (2001) Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain. *J Mol Biol* 312(4):783-794.
- [322] Cossart P (2007) Listeriology (1926-2007): the rise of a model pathogen. *Microbes Infect* 9(10):1143-1146.
- [323] Schubert WD & Heinz DW (2003) Structural aspects of adhesion to and invasion of host cells by the human pathogen *Listeria monocytogenes*. *Chembiochem* 4(12):1285-1291.
- [324] Jeffers F (2009) Investigating the molecular mechanisms of the interactions between *Lactobacillus reuteri* strains and intestinal mucus. Ph.D. (University of East Anglia, Norwich, Norfolk, UK).
- [325] Gerharz EW, Turner WH, Kalble T, Woodhouse CR (2003) Metabolic and functional consequences of urinary reconstruction with bowel. *BJU Int* 91(2):143-149.
- [326] Nabi G, N'Dow J, Hasan TS, Booth IR, Cash P (2005) Proteomic analysis of urine in patients with intestinal segments transposed into the urinary tract. *Proteomics* 5(6):1729-1733.
- [327] Backstrom M, Link T, Olson FJ, Karlsson H, Graham R, Picco G, Burchell J, Taylor-Papadimitriou J, Noll T, Hansson GC (2003) Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. *Biochem J* 376(Pt 3):677-686.
- [328] Olson FJ, Backstrom M, Karlsson H, Burchell J, Hansson GC (2005) A MUC1 tandem repeat reporter protein produced in CHO-K1 cells has sialylated core 1 O-glycans and becomes more densely glycosylated if coexpressed with polypeptide-GalNAc-T4 transferase. *Glycobiology* 15(2):177-191.
- [329] Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* 35(6):e45.
- [330] Provencher SW & Glockner J (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20(1):33-37.
- [331] Laue TM, Shah BD; Ridgeway TM, Pelletier SL (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, Royal Society of Chemistry (Royal Society of Chemistry).
- [332] Demeler B (2005) UltraScan: A comprehensive data analysis software package for analytical ultracentrifugation experiments. *Modern analytical ultracentrifugation: Techniques and methods*, (Royal Society of Chemistry, UK), pp 210-229.

- [333] Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):271-281.
- [334] Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Pt 1):72-82.
- [335] Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43(1):186-190.
- [336] Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760-763.
- [337] Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58(Pt 11):1948-1954.
- [338] Read RJ (1986) Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Crystallographica Section A* 42:140-149.
- [339] Matthews BW (1968) Solvent Content of Protein Crystals. *Journal of Molecular Biology* 33(2):491-&.
- [340] Vagin A & Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *J Appl Crystallogr* 30:1022-1025.
- [341] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
- [342] Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* 41:641-643.
- [343] Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
- [344] Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240-255.
- [345] Terwilliger T (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* 11(Pt 1):49-52.
- [346] Sauter NK, Grosse-Kunstleve RW, Adams PD (2004) Robust indexing for automatic data collection. *J Appl Crystallogr* 37(Pt 3):399-409.
- [347] Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.
- [348] Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):61-69.

- [349] Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36:1277-1282.
- [350] Semenyuk AV & Svergun DI (1991) Gnom - a Program Package for Small-Angle Scattering Data-Processing. *J Appl Crystallogr* 24:537-540.
- [351] Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946-2953.
- [352] Volkov VV & Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36:860-864.
- [353] Birmanns S, Rusu M, Wriggers W (2011) Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *J Struct Biol* 173(3):428-435.
- [354] Wriggers W (2010) Using Situs for the integration of multi-resolution structures. *Biophysical reviews* 2(1):21-27.
- [355] Svergun D, Barberato C, Koch MHJ (1995) CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768-773.
- [356] Tack DM, Karem KL, Montgomery JR, Collins L, Bryant-Genevier MG, Tiernan R, Cano M, Lewis P, Engler RJ, Damon IK, Reynolds MG (2013) Unintentional transfer of vaccinia virus associated with smallpox vaccines: ACAM2000 ((R)) compared with Dryvax ((R)). *Hum Vaccin Immunother* 9(7).
- [357] Ehrig K, Kilinc MO, Chen NG, Stritzker J, Buckel L, Zhang Q, Szalay AA (2013) Growth inhibition of different human colorectal cancer xenografts after a single intravenous injection of oncolytic vaccinia virus GLV-1h68. *J Transl Med* 11:79.
- [358] Holm L & Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545-549.
- [359] Horsington J, Lynn H, Turnbull L, Cheng D, Braet F, Diefenbach RJ, Whitchurch CB, Karupiah G, Newsome TP (2013) A36-dependent Actin Filament Nucleation Promotes Release of Vaccinia Virus. *PLoS Pathog* 9(3):e1003239.
- [360] Gilbert SC (2013) Clinical development of Modified Vaccinia virus Ankara vaccines. *Vaccine*.
- [361] Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.
- [362] Kollman JM, Pandi L, Sawaya MR, Riley M, Doolittle RF (2009) Crystal structure of human fibrinogen. *Biochemistry* 48(18):3877-3886.
- [363] Erickson HP (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological procedures online* 11:32-51.

- [364] Dell A, Galadari A, Sastre F, Hitchen P (2010) Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *International journal of microbiology* 2010:148178.
- [365] Nothaft H & Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 8(11):765-778.
- [366] Schmidt MA, Riley LW, Benz I (2003) Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol* 11(12):554-561.
- [367] Fredriksen L, Mathiesen G, Moen A, Bron PA, Kleerebezem M, Eijsink VG, Egge-Jacobsen W (2012) The major autolysin Acm2 from *Lactobacillus plantarum* undergoes cytoplasmic O-glycosylation. *J Bacteriol* 194(2):325-333.
- [368] Lebeer S, Claes IJ, Balog CI, Schoofs G, Verhoeven TL, Nys K, von Ossowski I, de Vos WM, Tytgat HL, Agostinis P, Palva A, Van Damme EJ, Deelder AM, De Keersmaecker SC, Wuhler M, Vanderleyden J (2012) The major secreted protein Msp1/p75 is O-glycosylated in *Lactobacillus rhamnosus* GG. *Microb Cell Fact* 11:15.
- [369] Turner MS, Hafner LM, Walsh T, Giffard PM (2004) Identification and characterization of the novel LysM domain-containing surface protein Sep from *Lactobacillus fermentum* BR11 and its use as a peptide fusion partner in *Lactobacillus* and *Lactococcus*. *Appl Environ Microbiol* 70(6):3673-3680.
- [370] Sanchez B, Bressollier P, Urdaci MC (2008) Exported proteins in probiotic bacteria: adhesion to intestinal surfaces, host immunomodulation and molecular cross-talking with the host. *FEMS Immunol Med Microbiol* 54(1):1-17.
- [371] Mobili P, Serradell Mde L, Trejo SA, Aviles Puigvert FX, Abraham AG, De Antoni GL (2009) Heterogeneity of S-layer proteins from aggregating and non-aggregating *Lactobacillus kefir* strains. *Antonie Van Leeuwenhoek* 95(4):363-372.
- [372] Stepper J, Shastri S, Loo TS, Preston JC, Novak P, Man P, Moore CH, Havlicek V, Patchett ML, Norris GE (2011) Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett* 585(4):645-650.
- [373] Zhou M & Wu H (2009) Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* 155(Pt 2):317-327.
- [374] Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue):D290-301.
- [375] Zwart PH, Grosse-Kunstleve RW, Lebedev AA, Murshudov GN, Adams PD (2008) Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):99-107.
- [376] Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125-132.
- [377] Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247(4):536-540.

- [378] Burroughs AM, Balaji S, Iyer LM, Aravind L (2007) Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. *Biol Direct* 2:18.
- [379] Deivanayagam CC, Rich RL, Carson M, Owens RT, Danthuluri S, Bice T, Hook M, Narayana SV (2000) Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8(1):67-78.
- [380] Budzik JM, Poor CB, Faull KF, Whitelegge JP, He C, Schneewind O (2009) Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc Natl Acad Sci U S A* 106(47):19992-19997.
- [381] Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* 318(5856):1625-1628.
- [382] Kang HJ, Paterson NG, Gaspar AH, Ton-That H, Baker EN (2009) The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci U S A* 106(40):16967-16971.
- [383] Paterson NG & Baker EN (2011) Structure of the full-length major pilin from *Streptococcus pneumoniae*: implications for isopeptide bond formation in gram-positive bacterial pili. *PLoS One* 6(7):e22095.
- [384] Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, Ward NJ, Yankovskaya V, Oliver KM, Cecchini G, Sulikowski GA, Tyska MJ, Sullam PM, Iverson TM (2011) A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. *PLoS Pathog* 7(7):e1002112.
- [385] Vengadesan K & Narayana SV (2011) Structural biology of Gram-positive bacterial adhesins. *Protein Sci* 20(5):759-772.
- [386] Petrey D, Fischer M, Honig B (2009) Structural relationships among proteins with different global topologies and their implications for function annotation strategies. *Proc Natl Acad Sci U S A* 106(41):17377-17382.
- [387] Zhang QC, Petrey D, Norel R, Honig BH (2010) Protein interface conservation across structure space. *Proc Natl Acad Sci U S A* 107(24):10896-10901.
- [388] Nayal M & Honig B (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins* 63(4):892-906.
- [389] Romero, Obradovic, Dunker K (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family. *Genome informatics. Workshop on Genome Informatics* 8:110-124.
- [390] Liu Q, Ponnuraj K, Xu Y, Ganesh VK, Sillanpaa J, Murray BE, Narayana SV, Hook M (2007) The *Enterococcus faecalis* MSCRAMM ACE binds its ligand by the Collagen Hug model. *J Biol Chem* 282(27):19629-19637.
- [391] Ramboarina S, Garnett JA, Zhou M, Li Y, Peng Z, Taylor JD, Lee WC, Bodey A, Murray JW, Alguet Y, Bergeron J, Bardiaux B, Sawyer E, Isaacson R, Tagliaferri C, Cota E, Nilges M, Simpson P, Ruiz T, Wu H, Matthews S (2010) Structural insights into serine-rich fimbriae from Gram-positive bacteria. *J Biol Chem* 285(42):32446-32457.

- [392] Voltan S, Castagliuolo I, Elli M, Longo S, Brun P, D'Inca R, Porzionato A, Macchi V, Palu G, Sturniolo GC, Morelli L, Martines D (2007) Aggregating phenotype in *Lactobacillus crispatus* determines intestinal colonization and TLR2 and TLR4 modulation in murine colonic mucosa. *Clin Vaccine Immunol* 14(9):1138-1148.
- [393] Kos B, Suskovic J, Vukovic S, Simpraga M, Frece J, Matosic S (2003) Adhesion and aggregation ability of probiotic strain *Lactobacillus acidophilus* M92. *J Appl Microbiol* 94(6):981-987.
- [394] Ledder RG, Timperley AS, Friswell MK, Macfarlane S, McBain AJ (2008) Coaggregation between and among human intestinal and oral bacteria. *FEMS Microbiol Ecol* 66(3):630-636.
- [395] Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. *Science* 318(5856):1625-1628.
- [396] Nelson AL, Ries J, Bagnoli F, Dahlberg S, Falker S, Rounioja S, Tschop J, Morfeldt E, Ferlenghi I, Hilleringmann M, Holden DW, Rappuoli R, Normark S, Barocchi MA, Henriques-Normark B (2007) RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. *Mol Microbiol* 66(2):329-340.
- [397] Patti JM, House-Pompeo K, Boles JO, Garza N, Gurusiddappa S, Hook M (1995) Critical residues in the ligand-binding site of the *Staphylococcus aureus* collagen-binding adhesin (MSCRAMM). *J Biol Chem* 270(20):12005-12011.
- [398] Patti JM & Hook M (1994) Microbial adhesins recognizing extracellular matrix macromolecules. *Curr Opin Cell Biol* 6(5):752-758.
- [399] Ponnuraj K, Bowden MG, Davis S, Gurusiddappa S, Moore D, Choe D, Xu Y, Hook M, Narayana SV (2003) A "dock, lock, and latch" structural model for a staphylococcal adhesin binding to fibrinogen. *Cell* 115(2):217-228.
- [400] Symersky J, Patti JM, Carson M, House-Pompeo K, Teale M, Moore D, Jin L, Schneider A, DeLucas LJ, Hook M, Narayana SV (1997) Structure of the collagen-binding domain from a *Staphylococcus aureus* adhesin. *Nat Struct Biol* 4(10):833-838.
- [401] Deivanayagam CC, Wann ER, Chen W, Carson M, Rajashankar KR, Hook M, Narayana SV (2002) A novel variant of the immunoglobulin fold in surface adhesins of *Staphylococcus aureus*: crystal structure of the fibrinogen-binding MSCRAMM, clumping factor A. *EMBO J* 21(24):6660-6672.
- [402] Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in gram-positive pathogens. *Nat Rev Microbiol* 4(7):509-519.
- [403] Solovyova AS, Pointon JA, Race PR, Smith WD, Kehoe MA, Banfield MJ (2010) Solution structure of the major (Spy0128) and minor (Spy0125 and Spy0130) pili subunits from *Streptococcus pyogenes*. *Eur Biophys J* 39(3):469-480.
- [404] El Mortaji L, Contreras-Martel C, Moschioni M, Ferlenghi I, Manzano C, Vernet T, Dessen A, Di Guilmi AM (2012) The full-length *Streptococcus pneumoniae* major pilin RrgB crystallizes in a fibre-like structure, which presents the D1 isopeptide bond and provides details on the mechanism of pilus polymerization. *Biochem J* 441(3):833-841.

- [405] Karlsson H, Holmen Larsson, J. M, Thomsson, K. A., Hard, I., Backstrom, M., Hansson, G. C. (2009) High-Throughput and High-Sensitivity Nano-LC/MS and MS/MS for *O*-Glycan Profiling. *Methods in Molecular Biology, Glycomics: Methods and Protocols*, ed Packer NH, Karlsson, N. G. (Humana Press), Vol 534, pp 117-131.
- [406] Afrough B, Dwek MV, Greenwell P (2007) Identification and elimination of false-positives in an ELISA-based system for qualitative assessment of glycoconjugate binding using a selection of plant lectins. *Biotechniques* 43(4):458, 460, 462 passim.
- [407] Gunning AP, Kirby AR, Fuell C, Pin C, Tailford LE, Juge N (2013) Mining the "glycocode"--exploring the spatial distribution of glycans in gastrointestinal mucin using force spectroscopy. *Faseb J* 27(6):2342-2354.
- [408] Wu AM, Wu JH, Song SC, Tsai MS, Herp A (1998) Studies on the binding of wheat germ agglutinin (*Triticum vulgare*) to *O*-glycans. *FEBS Lett* 440(3):315-319.
- [409] Kilcoyne M, Gerlach JQ, Gough R, Gallagher ME, Kane M, Carrington SD, Joshi L (2012) Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. *Anal Chem* 84(7):3330-3338.
- [410] Kilcoyne M, Gerlach JQ, Kane M, Joshi L (2012) Surface chemistry and linker effects on lectin-carbohydrate recognition for glycan microarrays. *Anal Methods-Uk* 4(9):2721-2728.
- [411] Smith DF, Song X, Cummings RD (2010) Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods Enzymol* 480:417-444.
- [412] Moller I, Marcus SE, Haeger A, Verhertbruggen Y, Verhoef R, Schols H, Ulvskov P, Mikkelsen JD, Knox JP, Willats W (2008) High-throughput screening of monoclonal antibodies against plant cell wall glycans by hierarchical clustering of their carbohydrate microarray binding profiles. *Glycoconj J* 25(1):37-48.
- [413] Gosalia N, Leir SH, Harris A (2013) Coordinate regulation of the gel-forming mucin genes at chromosome 11p15.5. *J Biol Chem* 288(9):6717-6725.
- [414] van Klinken BJ, Oussoren E, Weenink JJ, Strous GJ, Buller HA, Dekker J, Einerhand AW (1996) The human intestinal cell lines Caco-2 and LS174T as models to study cell-type specific mucin expression. *Glycoconj J* 13(5):757-768.
- [415] Mammen M, Choi SK, Whitesides GM (1998) Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew Chem Int Edit* 37(20):2755-2794.
- [416] Edwards AM, Potter U, Meenan NA, Potts JR, Massey RC (2011) *Staphylococcus aureus* keratinocyte invasion is dependent upon multiple high-affinity fibronectin-binding repeats within FnBPA. *PLoS One* 6(4):e18899.
- [417] Edwards AM, Potts JR, Josefsson E, Massey RC (2010) *Staphylococcus aureus* host cell invasion and virulence in sepsis is facilitated by the multiple repeats within FnBPA. *PLoS Pathog* 6(6):e1000964.

- [418] Lin YP, McDonough SP, Sharma Y, Chang YF (2010) The terminal immunoglobulin-like repeats of LigA and LigB of *Leptospira* enhance their binding to gelatin binding domain of fibronectin and host cells. *PLoS One* 5(6):e11301.
- [419] Marottoli RA & Drickamer MA (1993) Psychomotor mobility and the elderly driver. *Clinics in geriatric medicine* 9(2):403-411.
- [420] Gerken TA, Gupta R, Jentoft N (1992) A novel approach for chemically deglycosylating O-linked glycoproteins. The deglycosylation of submaxillary and respiratory mucins. *Biochemistry* 31(3):639-648.
- [421] Koutsoulis D, Landry D, Guthrie EP (2008) Novel endo-alpha-N-acetylgalactosaminidases with broader substrate specificity. *Glycobiology* 18(10):799-805.
- [422] Plummer TH, Jr. & Tarentino AL (1991) Purification of the oligosaccharide-cleaving enzymes of *Flavobacterium meningosepticum*. *Glycobiology* 1(3):257-263.
- [423] Katouli M (2010) Population structure of gut *Escherichia coli* and its role in development of extra-intestinal infections. *Iranian journal of microbiology* 2(2):59-72.
- [424] Esko JD, Lindahl U (2008) Proteoglycans and Sulfate Glycosaminoglycans. *Essential of Glycobiology*, ed Varki A CR, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA), Vol 2, pp 229-248.
- [425] Nickel JC, Emerson L, Cornish J (1993) The bladder mucus (glycosaminoglycan) layer in interstitial cystitis. *J Urol* 149(4):716-718.
- [426] Parsons CL & Mulholland SG (1978) Bladder surface mucin. Its antibacterial effect against various bacterial species. *Am J Pathol* 93(2):423-432.
- [427] Parsons KF, O'Boyle PJ, Gibbon NO (1977) A further assessment of bladder transection in the management of adult enuresis and allied conditions. *Br J Urol* 49(6):509-514.
- [428] Jones CH, Pinkner JS, Roth R, Heuser J, Nicholes AV, Abraham SN, Hultgren SJ (1995) FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc Natl Acad Sci U S A* 92(6):2081-2085.
- [429] Kuehn MJ, Heuser J, Normark S, Hultgren SJ (1992) P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature* 356(6366):252-255.
- [430] Lintermans PF, Pohl P, Bertels A, Charlier G, Vandekerckhove J, Van Damme J, Schoup J, Schlicker C, Korhonen T, De Greve H, et al. (1988) Characterization and purification of the F17 adhesin on the surface of bovine enteropathogenic and septicemic *Escherichia coli*. *Am J Vet Res* 49(11):1794-1799.
- [431] Roberts JA, Marklund BI, Ilver D, Haslam D, Kaack MB, Baskin G, Louis M, Mollby R, Winberg J, Normark S (1994) The Gal(alpha 1-4)Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proc Natl Acad Sci U S A* 91(25):11889-11893.

- [432] Varki A & Angata T (2006) Siglecs--the major subfamily of I-type lectins. *Glycobiology* 16(1):1R-27R.
- [433] Alpey MS, Attrill H, Crocker PR, van Aalten DM (2003) High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family. *J Biol Chem* 278(5):3372-3377.
- [434] May AP, Robinson RC, Vinson M, Crocker PR, Jones EY (1998) Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution. *Mol Cell* 1(5):719-728.
- [435] Crocker PR, Mucklow S, Bouckson V, McWilliam A, Willis AC, Gordon S, Milon G, Kelm S, Bradfield P (1994) Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. *EMBO J* 13(19):4490-4503.
- [436] Engel P, Wagner N, Miller AS, Tedder TF (1995) Identification of the ligand-binding domains of CD22, a member of the immunoglobulin superfamily that uniquely binds a sialic acid-dependent ligand. *J Exp Med* 181(4):1581-1586.
- [437] Law CL, Aruffo A, Chandran KA, Doty RT, Clark EA (1995) Ig domains 1 and 2 of murine CD22 constitute the ligand-binding domain and bind multiple sialylated ligands expressed on B and T cells. *J Immunol* 155(7):3368-3376.
- [438] Reunanen J, von Ossowski I, Hendrickx AP, Palva A, de Vos WM (2012) Characterization of the SpaCBA pilus fibers in the probiotic *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* 78(7):2337-2344.
- [439] Oyelaran O & Gildersleeve JC (2009) Glycan arrays: recent advances and future challenges. *Curr Opin Chem Biol* 13(4):406-413.
- [440] Oyelaran O, Li Q, Farnsworth D, Gildersleeve JC (2009) Microarrays with varying carbohydrate density reveal distinct subpopulations of serum antibodies. *J Proteome Res* 8(7):3529-3538.
- [441] Parthasarathy N, DeShazer D, England M, Waag DM (2006) Polysaccharide microarray technology for the detection of *Burkholderia pseudomallei* and *Burkholderia mallei* antibodies. *Diagn Microbiol Infect Dis* 56(3):329-332.
- [442] Wang D, Liu S, Trummer BJ, Deng C, Wang A (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nat Biotechnol* 20(3):275-281.
- [443] Willats WG, Rasmussen SE, Kristensen T, Mikkelsen JD, Knox JP (2002) Sugar-coated microarrays: a novel slide surface for the high-throughput analysis of glycans. *Proteomics* 2(12):1666-1671.
- [444] [Anon] (2005) Intestinal mucin gene modulation in vivo using orally administered probiotic bacteria. *J Pediatr Gastr Nutr* 41(4):558-559.
- [445] Ban L & Mrksich M (2008) On-chip synthesis and label-free assays of oligosaccharide arrays. *Angew Chem Int Ed Engl* 47(18):3396-3399.

- [446] Blixt O, Allin K, Bohorov O, Liu X, Andersson-Sand H, Hoffmann J, Razi N (2008) Glycan microarrays for screening sialyltransferase specificities. *Glycoconj J* 25(1):59-68.
- [447] Godula K & Bertozzi CR (2012) Density variant glycan microarray for evaluating cross-linking of mucin-like glycoconjugates by lectins. *J Am Chem Soc* 134(38):15732-15742.
- [448] Padler-Karavani V, Song X, Yu H, Hurtado-Ziola N, Huang S, Muthana S, Chokhawala HA, Cheng J, Verhagen A, Langereis MA, Kleene R, Schachner M, de Groot RJ, Lasanajak Y, Matsuda H, Schwab R, Chen X, Smith DF, Cummings RD, Varki A (2012) Cross-comparison of protein recognition of sialic acid diversity on two novel sialoglycan microarrays. *J Biol Chem* 287(27):22593-22608.
- [449] Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D* 65:582-601.
- [450] Holm L & Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* 16(6):566-567.
- [451] Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR (2004) The Pfam protein families database. *Nucleic Acids Research* 32:D138-D141.
- [452] Vazquez-Boland JA, Kuhn M, Berche P, Chakraborty T, Dominguez-Bernal G, Goebel W, Gonzalez-Zorn B, Wehland J, Kreft J (2001) *Listeria* pathogenesis and molecular virulence determinants. *Clin Microbiol Rev* 14(3):584-640.
- [453] Ebbes M, Blyemuller WM, Cernescu M, Nolker R, Brutschy B, Niemann HH (2011) Fold and function of the InIB B-repeat. *J Biol Chem* 286(17):15496-15506.
- [454] Chen BY & Honig B (2010) VASP: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. *PLoS computational biology* 6(8).
- [455] Mengaud J, Ohayon H, Gounon P, Mege RM, Cossart P (1996) E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells. *Cell* 84(6):923-932.
- [456] Parida SK, Domann E, Rohde M, Muller S, Darji A, Hain T, Wehland J, Chakraborty T (1998) Internalin B is essential for adhesion and mediates the invasion of *Listeria monocytogenes* into human endothelial cells. *Mol Microbiol* 28(1):81-93.
- [457] Shen Y, Naujokas M, Park M, Ireton K (2000) InIB-dependent internalization of *Listeria* is mediated by the Met receptor tyrosine kinase. *Cell* 103(3):501-510.
- [458] Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couve E, de Daruvar A, Dehoux P, Domann E, Dominguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, Garcia-del Portillo F, Garrido P, Gautier L, Goebel W, Gomez-Lopez N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G,

- Novella S, de Pablos B, Perez-Diaz JC, Purcell R, Rimmel B, Rose M, Schlueter T, Simoes N, Tierrez A, Vazquez-Boland JA, Voss H, Wehland J, Cossart P (2001) Comparative genomics of *Listeria* species. *Science* 294(5543):849-852.
- [459] Karlsson NG, Herrmann A, Karlsson H, Johansson ME, Carlstedt I, Hansson GC (1997) The glycosylation of rat intestinal Muc2 mucin varies between rat strains and the small and large intestine. A study of *O*-linked oligosaccharides by a mass spectrometric approach. *J Biol Chem* 272(43):27025-27034.
- [460] Linden SK, Bierne H, Sabet C, Png CW, Florin TH, McGuckin MA, Cossart P (2008) *Listeria monocytogenes* internalins bind to the human intestinal mucin MUC2. *Arch Microbiol* 190(1):101-104.
- [461] Renier S, Hebraud M, Desvaux M (2011) Molecular biology of surface colonization by *Listeria monocytogenes*: an additional facet of an opportunistic Gram-positive foodborne pathogen. *Environ Microbiol* 13(4):835-850.
- [462] Bierne H, Sabet C, Personnic N, Cossart P (2007) Internalins: a complex family of leucine-rich repeat-containing proteins in *Listeria monocytogenes*. *Microbes Infect* 9(10):1156-1166.
- [463] Hamon M, Bierne H, Cossart P (2006) *Listeria monocytogenes*: a multifaceted model. *Nat Rev Microbiol* 4(6):423-434.
- [464] Freiberg A, Machner MP, Pfeil W, Schubert WD, Heinz DW, Seckler R (2004) Folding and stability of the leucine-rich repeat domain of internalin B from *Listeria monocytogenes*. *J Mol Biol* 337(2):453-461.
- [465] Lecuit M, Ohayon H, Braun L, Mengaud J, Cossart P (1997) Internalin of *Listeria monocytogenes* with an intact leucine-rich repeat region is sufficient to promote internalization. *Infect Immun* 65(12):5309-5319.
- [466] Copp J, Marino M, Banerjee M, Ghosh P, van der Geer P (2003) Multiple regions of internalin B contribute to its ability to turn on the Ras-mitogen-activated protein kinase pathway. *J Biol Chem* 278(10):7783-7789.
- [467] Jonquieres R, Pizarro-Cerda J, Cossart P (2001) Synergy between the N- and C-terminal domains of InlB for efficient invasion of non-phagocytic cells by *Listeria monocytogenes*. *Mol Microbiol* 42(4):955-965.
- [468] Cummings RD MR (2009) Glycan binding proteins *Essentials of Glycobiology* ed Varki A CR, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA), Vol 2, pp 439-474.
- [469] Drickamer K & Taylor ME (1993) Biology of animal lectins. *Annual review of cell biology* 9:237-264.
- [470] Tripathi P, Dupres V, Beaussart A, Lebeer S, Claes IJ, Vanderleyden J, Dufrene YF (2012) Deciphering the nanometer-scale organization and assembly of *Lactobacillus rhamnosus* GG pili using atomic force microscopy. *Langmuir* 28(4):2211-2216.
- [471] Bingham RJ, Rudino-Pinera E, Meenan NA, Schwarz-Linek U, Turkenburg JP, Hook M, Garman EF, Potts JR (2008) Crystal structures of fibronectin-binding sites from

*Staphylococcus aureus* FnBPA in complex with fibronectin domains. *Proc Natl Acad Sci U S A* 105(34):12254-12258.

- [472] Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE (2009) A general *O*-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell* 137(2):321-331.
- [473] Yan F, Cao H, Cover TL, Whitehead R, Washington MK, Polk DB (2007) Soluble proteins produced by probiotic bacteria regulate intestinal epithelial cell survival and growth. *Gastroenterology* 132(2):562-575.
- [474] Lebeer S, Claes I, Tytgat HL, Verhoeven TL, Marien E, von Ossowski I, Reunanen J, Palva A, Vos WM, Keersmaecker SC, Vanderleyden J (2012) Functional analysis of *Lactobacillus rhamnosus* GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells. *Appl Environ Microbiol* 78(1):185-193.
- [475] van Baarlen P, Wells JM, Kleerebezem M (2013) Regulation of intestinal homeostasis and immunity with probiotic lactobacilli. *Trends Immunol* 34(5):208-215.