

**BIOSYNTHESIS OF MICROBIAL GLYCANS AND THEIR RECOGNITION BY
HUMAN INTELECTIN-1**

By

Darryl A. Wesener

A dissertation submitted in partial fulfillment of
the requirements for the degree of

**Doctor of Philosophy
(Biochemistry)**

at the

UNIVERSITY OF WISCONSIN–MADISON

2016

Date of final oral examination: 06/28/2016

The dissertation is approved by the following members of the Final Oral Committee:
Laura L. Kiessling, Professor, Chemistry and Biochemistry
Ronald T. Raines, Professor, Biochemistry and Chemistry
John M. Denu, Professor, Biomolecular Chemistry
Douglas B. Weibel, Professor, Biochemistry, Chemistry, and Biomedical Engineering
Eric R. Strieter, Associate Professor, Chemistry

© Copyright by Darryl A. Wesener
All Rights Reserved
2016

**BIOSYNTHESIS OF MICROBIAL GLYCANS AND THEIR RECOGNITION BY
HUMAN INTELECTIN-1**

Darryl A. Wesener

Under the supervision of Professor Laura L. Kiessling

At the University of Wisconsin–Madison

The human body is challenged with constantly surveying and managing microbial guests, while preventing infection and colonization by pathogenic microbes. Both of these goals require the accurate assessment of a cell's origin, and the ability to effect a function. Virtually all cells are covered in the coat of carbohydrates. Cell surface glycans could serve as cell identification codes when recognized by lectins. Interestingly, the carbohydrates utilized in the assembly of microbial cell surface glycans differ substantially from those found on mammalian cells. My thesis work has taken an interdisciplinary approach using protein biochemistry, chemical biology, microbiology, and immunology to study both the biosynthesis of microbial glycans and the recognition of microbial glycans by the human lectin intelectin-1 (hIntL-1).

Using protein biochemistry and chemical biology I characterized the enzyme responsible for uridine 5'-diphosphate- α -D-galactofuranose (UDP-Galf) biosynthesis, UDP-galactopyranose mutase (UGM), in the nematode *Caenorhabditis elegans*. These studies revealed conservation between the structure and mechanism of prokaryotic and eukaryotic UGM enzymes despite significant sequence divergence (Chapter 2). While exploring the biological function of Galf-containing glycans, I became interested in how the human immune system may interact with microbe specific carbohydrates. Focusing our efforts on hIntL-1, a soluble lectin proposed to function in innate immunity, we identified several microbe specific carbohydrate ligands of the

lectin that include β -D-Galp-, D-phospho-glycerol-, heptose-, D-glycero-D-talo-oct-2-ulosonic acid (KO)-, and 3-deoxy-D-manno-oct-2-ulosonic acid (KDO)-containing glycans (Chapter 3). These microbial carbohydrate residues all share an exocyclic 1,2-diol epitope that is recognized by hIntL-1. Recognition of the exocyclic diol by a *Xenopus laevis* intelectin suggests that intelectin:diol binding may be a general mechanism for microbe detection by chordates (Chapter 4). Lastly, hIntL-1 binding to microbial communities representative of human gastrointestinal microbiomes suggests that exocyclic 1,2-diols are abundantly present on microbial symbionts and hIntL-1 interacts judiciously and tunably with binding species present within communities (Chapter 5). I anticipate future work will reveal a role for intelectin in recognition and regulation of microbial guests at mucosal surfaces.

Acknowledgements

I would like to thank many people who have helped me achieve this goal today for their longstanding encouragement and support. First I would like to thank my family. My parents, Carmen and Jerome, who have been incredibly supportive during this entire process and have always challenged me to pursue my goals fearlessly, I thank you for your love and support. I would also like to thank my sister, Eileen, for her support and the meals we shared while we both resided in Madison. I thank my extended family, my grandparents, aunts and uncles, and cousins who have been supportive and challenged me to effectively communicate my science to the general public. Lastly, I want to thank Edna Trujillo for her enthusiastic and wonderful support during this hectic time.

I would like to thank my labmates and classmates. Working in the Kiessling Group has been a wonderful experience because of the amazing people I get to interact with everyday. I first want to thank my baymates, Adam Courtney and Sayaka Masuko, for tolerating me. I want to thank Cleaved Paul Wrighton, Danny Rack'em Zwick, and Josh Fishman for their friendship and counseling. I am thankful to Matt Levensgood, Becca Splain, Adam Courtney, Joe Klim, Rachael Sheridan, and Lingyin Li mentoring early in my career and challenging me to think critically about my work. I am thankful to the many people I have worked with including Kittikhun Wangkanont, Lucas Zarling, Heather Hodges, Ginny Kincaid, Alex Justen, Becca Splain, Matt Kraft, Betsy Huffman, Christine Isabella, and Amanda Dugan. Their encouragement, intelligence, and drive have made collaborating fruitful and enjoyable. Lastly, I want to thank my IPiB and Chemistry classmates for their friendship and camaraderie.

I am thankful for the people that make UW–Madison such a wonderful place to learn and grow. I am thankful to the Biochemistry front office. I thank Darrelle McCaslin and Dan Stevens of the Biophysics Instrumentation Facility for their patience and support. I thank Dagna Sheerar and Rachael Sheridan of the Flow Cytometry Core, Greg Sabat and the Biotechnology Center Mass Spectrometry Facility. I am especially thankful for Sheila, Kate, and Kris Turkow for their support of the entire Group and helping ensure we remained sane. I am thankful to the UW–Madison Chemistry and Biology Interface Training Grant and the National Science Foundation Graduate Research Fellowship for financial support and a stimulating training environment that has provided me countless opportunities to advance my training.

I am thankful for the many collaborators at UW and around the world that I have had the privilege of working with during my training; including Matt Mead and Christina Hull, Robert Kerby and the group of Federico Rey, Mats Johnsson and Deane Moshers group, Tsokyi Choera and Nancy Keller, Katrina Forest and her entire Group, Ryan McBride and Jim Paulson, Xuezheng Song, David Smith and Richard Cummings. I am thankful to Amgen and many great mentors I met there while participating in an industrial internship. Specifically I want to thank Bram Estes and Jennitte Stevens for their past and continued support.

I am thankful to my thesis committee members for their support and mentorship. I thank Eric Strieter for his support, thoughtful discussion, and career advice. I thank Ron Raines for challenging me to think molecularly, his inspiring instruction in the classroom, and for mentoring me in both the classroom and the laboratory. I thank Brian Fox for thoughtful discussion about enzyme mechanism and function and for his wonderful teaching in multiple classes. I thank John Denu for stimulating scientific conversation and mentorship on navigating

graduate school and my career. I thank Doug Weibel for his forthcomingness with scientific assistance, mentorship, teaching, and his passion for education and research.

Lastly, I would like to thank my thesis advisor Laura L. Kiessing for her enthusiastic support, mentoring, and scientific training. Laura has created a stimulating work environment that has allowed me to fearlessly explore and follow the science. She has challenged me to think big picture about my work in the broader context and to effectively communicate with my peers. Laura has also forced me to become a better writer. I hope this comes through in this thesis, but please excuse a few spelling errors. Laura, I thank you for taking me into your lab and mentoring me during my development as an independent scientist.

Table of Contents

Abstract.....	i
Acknowledgements.....	iii
Table of Contents.....	vi
List of Figures.....	xiii
List of Tables.....	xvii
List of Abbreviations.....	xviii
Chapter 1: Biosynthesis of Microbial Glycans and Their Recognition by Soluble	
Human Lectins.....	1
1.1 Abstract.....	2
1.2 Introduction.....	3
1.3 Bacterial Glycoconjugates.....	3
1.3.1 Peptidoglycan.....	3
1.3.2 Lipopolysaccharide.....	4
1.3.3 Capsular Polysaccharide.....	8
1.3.4 Teichoic Acids.....	13
1.3.5 Others.....	15
1.4 Fungal Glycoconjugates.....	17
1.5 Lectins.....	20
1.5.1 Ficolins.....	20
1.5.2 Collectins.....	25

1.5.3	Intelectins.....	31
1.5.4	Reg Proteins.....	34
1.5.5	Peptidoglycan Recognition Proteins.....	39
1.5.6	Galectins.....	42
1.5.7	Pentraxins.....	47
1.5.8	Lysozyme.....	53
1.5.9	Human C-type Lectin Domain Family (CLEC) Proteins.....	53
1.5.10	ZG16p.....	54
1.5.11	Mindin.....	56
1.5.12	Soluble CD14 and Lipopolysaccharide Binding Protein.....	58
1.6	Conclusions.....	58
1.7	Acknowledgements.....	61
1.8	References.....	62
Chapter 2:	UDP-Galactopyranose Mutase in Nematodes.....	82
2.1	Abstract.....	83
2.2	Introduction.....	85
2.3	Purification of CeUGM and Initial Velocity Kinetic Analysis.....	89
2.4	Covalent Catalysis via FAD.....	91
2.5	Proposed Structure and Active Site of CeUGM.....	93
2.6	Chemical Inhibition of CeUGM.....	96
2.7	Conclusions.....	99

2.8 Methods.....	100
2.8.1 Cloning, Expression, and Purification of CeUGM.....	100
2.8.2 Enzymatic Activity.....	101
2.8.3 UV–Visible Spectroscopy.....	102
2.8.4 Far–UV Circular Dichroism (CD).....	102
2.8.5 Clustal W Alignment.....	103
2.8.6 Homology Model.....	103
2.8.7 Intermediate Trapping.....	103
2.8.8 Chemical Inhibition.....	104
2.9 Contributions.....	104
2.10 Acknowledgements.....	105
2.11 References.....	106
Chapter 3: Recognition of Microbial Glycans by Human Intelectin-1.....	111
3.1 Abstract.....	112
3.2 Introduction.....	113
3.3 hIntL-1 Binds β-Galf.....	115
3.4 hIntL-1 Binding to Microbial Glycans.....	118
3.5 Structure of hIntL-1.....	122
3.6 Structural Basis for hIntL-1 Selectivity.....	128
3.7 hIntL-1 Comparison with Ficolins.....	130
3.8 hIntL-1 Binding to <i>S. pneumoniae</i>.....	131

3.9 Effect of pH on hIntL-1 Binding.....	135
3.10 Murine IntL-1 Binding to GalF.....	136
3.11 Discussion.....	137
3.12 Methods.....	143
3.12.1 Chemical Synthesis of Glycan Ligands.....	143
3.12.2 Native Human Intelectin-1 Expression and Purification.....	143
3.12.3 Expression and Purification of Strep-tag [®] II hIntL-1.....	144
3.12.4 hIntL-1 Carbohydrate Binding ELISA-like Assay.....	145
3.12.5 Surface Plasmon Resonance (SPR).....	146
3.12.6 Construction of the Furanoside Glycan Array.....	147
3.12.7 Assay of hIntL-1 on Furanoside and CFG Mammalian Glycan Array.....	148
3.12.8 Assay of hIntL-1 on the Bacterial Glycan Array.....	149
3.12.9 Protein X-ray Crystallography.....	150
3.12.10 hIntL-1 Binding to <i>S. pneumoniae</i>	151
3.12.11 pH Profile of hIntL-1 Ligand Binding.....	153
3.12.12 Expression of Murine Intelectin-1.....	154
3.13 Accession Codes.....	156
3.14 Contributions.....	156
3.15 Acknowledgements.....	156
3.16 References.....	158

Chapter 4: Conserved Structure and Ligand Binding Mechanism Between Human Intelectin-1 and <i>Xenopus</i> Embryonic Epidermal Lectin.....	164
4.1 Abstract.....	165
4.2 Introduction.....	166
4.3 Expression and Purification of XEEL.....	168
4.4 Three-Dimensional Structure of XEEL_{CRD}.....	171
4.5 XEEL_{CRD} Binds Exocyclic 1,2-diol Containing Ligands.....	174
4.6 Human and Xenopus Intelectins use Similar Ligand-recognition Modes.....	175
4.7 Intelectin Structures Provide Insight into Residue Conservation and Function.....	177
4.8 XEEL_{CRD} is Trimeric in Solution.....	178
4.9 Discussion.....	180
4.10 Methods.....	186
4.10.1 Cloning, Expression, and Purification of XEEL.....	186
4.10.2 Protein X-ray Crystallography.....	188
4.10.3 SPR.....	189
4.10.4 Chemical Crosslinking.....	190
4.10.5 Sedimentation Equilibrium Analytical Ultracentrifugation.....	190
4.11 Accession Codes.....	192
4.12 Contributions.....	192

4.13 Acknowledgements.....	192
4.14 References.....	194
Chapter 5: Human Intelectin-1 Interactions with the Human Microbiome.....	198
5.1 Abstract.....	199
5.2 Introduction.....	201
5.3 hIntL-1 Recognition of Microbial Strains.....	205
5.4 hIntL-1 Binding to Microbial Communities.....	208
5.5 hIntL-1 Binding Fitness Depends on Community and Context.....	213
5.6 hIntL-1 Binding to the Human Fecal Microbiome.....	216
5.7 Microbial Carbohydrate Antigens on the Fecal Microbiota.....	220
5.8 hIntL-1 Interactions with Immune Cells.....	222
5.9 Discussion.....	225
5.10 Methods.....	229
5.10.1 hIntL-1 Binding top Bacterial Strains.....	229
5.10.2 hIntL-1 Binding to Synthetic Communities.....	230
5.10.3 hIntL-1 Binding to Human Fecal Samples.....	230
5.10.4 Sorting hIntL-1 Bound Bacteria From a Human Fecal Sample....	231
5.10.5 Sequencing the 16S rRNA Gene from hIntL-1 Bound Sorted Bacteria.....	231
5.10.6 Lectin Binding to Fecal Microbiota.....	232
5.10.7 hIntL-1 Binding to Human Immune Cells.....	232

5.11 Contributions.....	233
5.12 Acknowledgements.....	233
5.13 References.....	235
Appendix 1: Small Molecule Inhibition of UDP-Galactopyranose Mutase in <i>C. elegans</i>.....	239
A1.1 Abstract.....	240
A1.2 Introduction.....	241
A1.3 <i>glf-1</i> Knockdown Reduces Cuticle Stability.....	242
A1.4 Treatment of <i>C. elegans</i> with Compounds 2.3 and 2.5.....	243
A1.5 Lectin Staining of <i>C. elegans</i> Treated with Compound 2.5.....	245
A1.6 Staining <i>C. elegans</i> with hIntL-1.....	246
A1.7 Conclusions.....	248
A1.8 Methods.....	251
A1.8.1 <i>C. elegans</i> Culture.....	251
A1.8.2 Small Molecule Treatment of <i>C. elegans</i>	251
A1.8.3 Sodium Hypochlorite Sensitivity Assay.....	252
A1.8.4 Proteinase K Sensitivity Assay.....	252
A1.8.5 Lectin Staining.....	252
A1.9 Contributions.....	253
A1.10 Acknowledgements.....	253
A1.11 References.....	254

List of Figures

Figure 1-1.	Chemical structures of cell surface glycoconjugates found on Gram-negative bacteria.....	7
Figure 1-2.	Chemical structures of several LPS O-antigen and capsular polysaccharides.....	11
Figure 1-3.	Schematic representation of the cell surface of Gram-negative and Gram-positive bacteria.....	12
Figure 1-4.	Chemical structures of teichoic acids.....	15
Figure 1-5.	Schematic representation of a fungal cell wall.....	19
Figure 1-6.	Schematic of ficolin/collectin assembly into a function oligomer.....	24
Figure 1-7.	Expression and localization of human humoral immune lectins.....	27
Figure 1-8.	Structures of two peptidoglycan binding lectins.....	39
Figure 1-9.	Mechanisms by which soluble human immune lectins exert antimicrobial activity.....	47
Figure 1-10.	Structural classes of human soluble immune lectins.....	52
Figure 2-1.	Enzymatic biosynthesis of UDP-Galp.....	86
Figure 2-2.	Clustal W analysis of eukaryotic and prokaryotic UGM proteins.....	88
Figure 2-3.	A generalized view of the proposed mechanism of UGM, depicting a covalent flavin intermediate.....	89
Figure 2-4.	Steady-state kinetic analysis of CeUGM.....	91
Figure 2-5.	HPLC-MS analysis of the products of trapping the CeUGM-catalyzed reaction of UDP-Galp with sodium cyanoborohydride.....	93

Figure 2-6.	Proposed structure and active site of CeUGM.....	95
Figure 2-7.	Chemical inhibition of CeUGM.....	97
Figure 2-8.	Competitive inhibition of CeUGM by compound 2.3	98
Figure 3-1.	Expression, purification, and carbohydrate binding activity of hIntL-1.....	117
Figure 3-2.	Glycan selectivity of hIntL-1 assessed by glycan microarrays.....	119
Figure 3-3.	Human IntL-1 binding specificity as determined from the microbial glycan microarray (MGMv2).....	121
Figure 3-4.	Structure of hIntL-1 bound to allyl- β -D-Galf.....	124
Figure 3-5.	Structural alignment of hIntL-1 and human L-ficolin (PDB: 2J3U)...	126
Figure 3-6.	hIntL-1 bound to allyl- β -D-Galf.....	127
Figure 3-7.	hIntL-1 exhibits specificity for microbial glycan epitopes bearing terminal 1,2-diols.....	129
Figure 3-8.	Models for hIntL-1 interacting with relevant saccharide epitopes from humans (α -Neu5Ac) or microbes (α -KDO).....	130
Figure 3-9.	hIntL-1 binding to the surface of <i>S. pneumoniae</i>	134
Figure 3-10.	hIntL-1 binding to the surface of <i>S. pneumoniae</i> monitored by flow cytometry.....	135
Figure 3-11.	pH profile of hIntL-1 binding to immobilized β -Galf.....	136
Figure 3-12.	Mouse intelectin-1 binding to immobilized carbohydrates.....	137
Figure 3-13.	Structures of the 20 most prevalent monosaccharides that are unique to bacterial glycans.....	140

Figure 4-1.	Purification of XEEL.....	171
Figure 4-2.	Three-dimensional structure of XEEL _{CRD} trimer.....	173
Figure 4-3.	Binding of XEEL _{CRD} to immobilized carbohydrates as measured by SPR.....	175
Figure 4-4.	Crystal structure of XEEL _{CRD} -GroP complex.....	176
Figure 4-5.	Sequence alignment of XEEL and other intelectins.....	178
Figure 4-6.	Analysis of XEEL _{CRD} oligomeric state.....	179
Figure 4-7.	Comparison of lectin structures.....	182
Figure 5-1.	Schematic representation of the mucosal immune system functioning within the mammalian intestinal mucosa.....	205
Figure 5-2.	Binding of hIntL-1 to fixed bacterial strains.....	207
Figure 5-3.	Binding of hIntL-1 to synthetic microbial communities.....	212
Figure 5-4.	Competitive inhibition of hIntL-1 binding by microbial communities.....	215
Figure 5-5.	hIntL-1 binding to bacteria isolated from a human fecal slurry.....	219
Figure 5-6.	Genomic characterization of hIntL-1 bound bacteria.....	220
Figure 5-7.	Lectin binding to the surface of fecal microbiota.....	221
Figure 5-8.	hIntL-1 binding to human immune cells.....	224
Figure 5-9.	Hypothesis for the enhanced recognition of Gram-positive bacteria by hIntL-1 and its interaction with human immune cells.....	225
Figure A1-1.	Depletion of <i>glf-1</i> decreases <i>C. elegans</i> resistance to environmental stress.....	243

Figure A1-2.	Small molecule treatment of wild-type <i>C. elegans</i> does not phenocopy the results using RNAi and genetic deletion.....	245
Figure A1-3.	Surface staining of intact <i>C. elegans</i> with the plant lectin Con A.....	246
Figure A1-4.	Surface staining of intact <i>C. elegans</i> with hIntL-1.....	248
Figure A1-5.	Addition of detergents attenuated hIntL-1 binding to the surface of <i>glf-1</i> knockdown <i>C. elegans</i>	248

List of Tables

Table 2-1.	Kinetic Parameters of UGM Homologs.....	91
Table 2-2.	Inhibition and Binding Constants of 2-aminothiazoles Against CeUGM.....	97
Table 2-3.	Oligonucleotide Primer Sequences Used to Generate Point Mutants of CeUGM.....	101
Table 3-1.	The List of the Top 15 Microbial Glycan Ligands, Sorted by Average Fluorescence Intensity (RFU).....	121
Table 3-2.	Data Collection and Refinement Statistics (Molecular Replacement).	142
Table 4-1.	X-ray Crystallographic Data Collection and Refinement Statistics....	185
Table 5-1.	Summary of hIntL-1 Binding to Microbes.....	207
Table 5-2.	Example of Synthetic Microbial Mixtures Used to Assay hIntL-1 Binding.....	209

List of Abbreviations

Araf	arabinofuranose
APC	antigen presenting cell
ATCC	American Tissue Culture Collection
AUC	analytical ultracentrifugation
BAC	bacterial artificial chromosome
BCG	<i>Mycobacterium bovis</i> bacillus Calmette-Guérin
BCSCB	Bacterial Carbohydrate Structure Data Base
BIF	Biophysics Instrumentation Facility
Bis-Tris	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BSA	bovine serum albumin
C α	alpha carbon of an amino acid
CD	circular dichroism
CFG	Consortium for Functional Glycomics
CeUGM	<i>Ceanorhabditis elegans</i> UGM
CL-K1	collectin kidney 1
CL-L1	collectin liver 1
CL-P1	collectin placenta 1
CLEC	human C-type lectin domain family
Con A	concanavalin A
CPS	capsular polysaccharide
CRD	carbohydrate recognition domain

CRP	C-reactive protein
DAG	diacylglycerol
DC	dendritic cell
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ECL	<i>Erythrina cristagalli</i> lectin
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EM	electronmicroscopy
EPS	exopolysaccharide
EST	expressed sequence tag
FACS	fluorescence activated cell sorting
FAD	flavin adenine dinucleotide
FBD	fibrinogen-like domain
FBG	fibrinogen
FBS	fetal bovine serum
FMN	flavin mononucleotide
FPLC	fast protein liquid chromatography
Gal	galactose
Gal f	galactofuranose
GalNAc	N-acetylgalactosamine

Galp	galactopyranose
GI	gastrointestinal
Glc	glucose
GlcNAc	N-acetylglucosamine
GPI	glycosylphosphatidylinositol
Gro	glycerol
Gro-P	D-glycerol-1-phosphate
GWAS	genome-wide association study
HEK293T	human embryonic kidney 293T cells
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid
hIntL-1	human intelectin-1
HIP/PAP	hepatocarcinoma intestine-pancreas/pancreatic associated protein
His ₆	hexahistidine
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
IBD	irritable bowel disease
IC ₅₀	50% inhibitory concentration
IgA	immunoglobulin A
IgG	immunoglobulin G
IPTG	isopropyl-β-D-thiogalactopyranoside
K-antigen	capsular antigen (polysaccharide)
KDO	3-deoxy-D-manno-oct-2-ulosonic acid

KO	<i>D-glycero-D-talo-oct-2-ulosonic acid</i>
L-type	legumes plant-type
LAM	lipoarabinomannan
LB	Luria-Bertani
IntL	intelectin
LacNAc	N-acetyl-D-lactosamine or Gal- β (1-4)-GlcNAc
LOS	lipooligosaccharide
LBP	human lipopolysaccharide binding protein
LPS	lipopolysaccharide
LTA	lipoteichoic acid
mAGP	mycolyl-arabinogalactan-peptidoglycan
MALT	mucosa-associated lymphoid tissue
Man	mannose
ManNAc	N-acetylmannosamine
MASP	mannose-binding lectin-associated serine protease
MBL	human mannose-binding lectin
MGM	microbial glycan microarray
mIntL-1	mouse intelectin-1
mP	millipolarization
MS	mass spectrometry
MurNAc	N-acetylmuramic acid
Neu5Ac	N-acetylneuraminic acid

NHS	N-hydroxysuccinimidyl
NK	natural killer
NMR	nuclear magnetic resonance
O-antigen	repeating oligosaccharide attached to LPS
OD ₆₀₀	optical density at 600 nm
OPS	outer polysaccharide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PC	phosphatidylcholine
PEG	polyethylene glycol
PGLYRP	human peptidoglycan recognition proteins
PGRP	peptidoglycan recognition proteins
PP	Peyer's patches
PRC	phosphorylcholine
PI	phosphatidylinositol
PIgA	pentameric IgA
PCR	polymerase chain reaction
PDB	Protein Data Bank
PG	peptidoglycan
PMSF	phenylmethylsulfonyl fluoride
PTX3	pentraxin-3
RCA-I	<i>Ricinus communis</i> agglutinin I lectin

Reg	regenerating
RFU	relative fluorescence units
Rha	rhamnopyranose
Rib f	ribofuranose
Rib-P	ribitol-phosphate
RMSD	root-mean-square deviation
RNAi	RNA interference
rRNA	ribosomal ribonucleic acid
S	Svedberg
s.d.	standard deviation of the mean
S-layer	surface layer
SAP	serum amyloid P component
SDS	sodium dodecyl sulfate
SeMet	selenomethione
sIgA	soluble immunoglobulin A
sIgM	soluble immunoglobulin M
SP-A	surfactant proteins A
SP-D	surfactant proteins D
SPR	surface plasmon resonance
TLR	toll-like receptor
TMB	3,3',5,5'-tetramethylbenzidine
Tris	tris(hydroxymethyl)aminomethane

UDP	uridine 5'-diphosphate
UDP-Galf	UDP- α -D-galactofuranose
UDP-Galp	UDP- α -D-galactopyranose
UGM	uridine 5'-diphosphate galactopyranose mutase
UV	ultraviolet
WGA	wheat germ agglutinin
WTA	wall teichoic acid
X-type lectins	intelectins
XEEL	<i>Xenopus laevis</i> embryonic epidermal lectin
ZG16p	human zymogen granule proteins

Chapter 1

Biosynthesis of Microbial Glycans and Their Recognition by Soluble Human Lectins

1.1 Abstract

Microbes are covered with a layer of carbohydrate antigens. These glycoconjugates perform essential functions including resistance to turgor pressure, cation homeostasis, and immune evasion. The human immune system has evolved diverse mechanisms for detecting these molecules and effecting an immune response. One mechanism is the production of soluble carbohydrate binding proteins, or lectins. Some soluble lectins bind their cognate ligand and activate mechanisms resulting in microbial cell death, others stimulate phagocytosis of microbes by antigen presenting cells, and others function by mechanisms we yet do not fully understand. Within this chapter I introduce many of the bacterial glycoconjugates that populate the microbial surface. Special attention will be paid to their biosynthesis and physiological role. Next, I will review human soluble lectins. I will focus on their structure, ligands, potential mechanisms of action, and results from human and mouse genetics studies that provide details about their biological function. It is my hope that this can serve as a guide to future studies of human innate immune lectins that function in microbial surveillance and regulation.

1.2 Introduction

All cells are covered with a layer of carbohydrates. For microbes, this carbohydrate coat functions analogously to metazoan skin; it is a barrier between the microbe and its environment (1,2). Cell surface glycans are essential for protecting microbes from osmotic stress, desiccation, toxic small molecules, metals, and proteins. These attributes are critical for survival in a complex, dynamic, and competitive growth environment. Cell surface glycans however serve as much more than just a protective sugar coat. They can function as molecular mimics for immune evasion, immunomodulators, or virulence factors. To detect and survey microbial guests, multicellular organisms have evolved lectins that can recognize microbial cell surface glycans. In this overview, the biosynthesis of representative microbial cell surface glycopolymers is discussed. With this background, we focus on the recognition of microbial glycoconjugates by human immune lectins. Of particular interest are soluble, or humoral lectins, and their role in immunity.

1.3 Bacterial Glycoconjugates

1.3.1 Peptidoglycan

Bacteria are enveloped in a glycopeptide polymer termed the peptidoglycan (PG) is central to the cell wall (3). It forms an extended mesh framework around the cell that gives the cell shape, strength, and resistance to turgor pressure. The PG is assembled via polymerization of lipid II monomers which are composed of a disaccharide and pentapeptide (**Figure 1-1A**) (1,3). The disaccharide consists of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc). The peptidyl component is linked to the disaccharide by an ether bond to the C(3)

hydroxyl of MurNAc. Within the peptide portion of PG are D-amino acids, which render it more resistant to many extracellular proteases.

Lipid II is polymerized into an extended $\beta(1-4)$ polymer via extracellular penicillin-binding proteins which catalyze transglycosylation reactions. In exponentially growing *E. coli*, PG strands are between 25 and 35 disaccharide units in length (4). Penicillin-binding proteins then install peptide crosslinks using a transpeptidase domain. In general, the peptide crosslinks are between the carboxyl group of the D-alanine located at position four of the peptide, and the amino group of diaminopimelic acid or L-lysine (**Figure 1-1B**). Most variation in PG structure arises in the peptide stem and the composition of the crosslinks. The polymerized glycan portion of PG, however, can be modified to generate increased epitope diversity (5). In Gram-negative bacteria, the PG is anchored to the outer membrane via linkage to Braun's lipoprotein (6). Because PG is an essential and conserved feature amongst Gram-negative and Gram-positive bacteria, it is the target of numerous small molecule antibiotics and is a ligand for multiple human immune proteins.

1.3.2 Lipopolysaccharide

Lipopolysaccharide (LPS), or endotoxin, decorated the surface of Gram-negative bacteria (7,8). True to its name, LPS is an amphiphilic molecule that contains lipid and carbohydrate portions (**Figures 1-1C & D**). It is assembled in the cytosol and periplasm but ultimately is localized to the outer leaflet of the outer membrane of Gram-negative bacterium.

The molecule can schematically be broken down into three parts: lipid A, the portion of the molecule that is lipidated and is inserted into the bacterial outer membrane, this portion is recognized by toll-like receptors (TLRs) and is responsible for the toxicity of endotoxin to

eukaryotic cells (9,10); a carbohydrate based core region, this can be divided conceptually into an inner (more proximal to lipid A) and outer region, this is somewhat variable; and a highly variable repeating O-antigen (**Figure 1-1C & D**). The presence or absence of O-antigen polysaccharide evokes a bacterium as classified “smooth” or “rough”, respectively. In many Gram-negative bacteria, the lipid A and 3-deoxy-D-manno-oct-2-ulosonic acid (KDO) portions of LPS are absolutely required for growth, (7) though that has recently been challenged via generation of an *E. coli* strain lacking KDO attached to its LPS (11).

In addition to LPS, many mucosal pathogens, including *Campylobacter jejuni*, *Neisseria spp.*, and *Yersinia pestis*, synthesize lipooligosaccharides (LOS), which structurally are similar to LPS but lack an O-antigen polysaccharide (12). The LPS/LOS, along with a capsular polysaccharide (K-antigen), occupy much of the accessible surface of Gram-negative bacteria, and are targets of many human innate and adaptive immune proteins. In an attempt to evade immune detection and shield themselves, some bacteria will incorporate mammalian like sugars into the O-antigen as a form of molecular mimicry (13).

The structural variability of LPS increases as one moves away from Lipid A. In general, the Lipid A and the inner core are very similar. For example, the monosaccharide KDO is universally found in LPS (14). Thus, the majority of glycan diversity in LPS molecules resides in differences in the repeating structure of the O-antigen (**Figure 1-2**). Indeed, 180 distinct O-antigens have been described for *E. coli* (14).

Most O-antigens are assembled by the carbohydrate polymerase Wzy. Wzy-dependent O-antigen arises from an undecaprenyl diphosphate oligosaccharide assembled in the cytoplasm via the action of several glycosyltransferases using nucleotide-activated donor sugars. Each

undecaprenyl diphosphate oligosaccharide contains between two and seven saccharide residues. The oligosaccharide is polymerized into a fully mature O-antigen bound to undecaprenyl diphosphate by en bloc transfer of the growing chain onto the nonreducing end of the “new” lipid linked oligosaccharide within the periplasmic space (15). The polysaccharide polymerase Wzy typically terminates after 10-25 en bloc transfers. The final length of the polysaccharide is controlled by the chain length regulator protein Wzz (16). Once fully assembled, the O-antigen is ligated to the Lipid-A core by a WaaL enzyme.

Other mechanisms by which the O-antigen can be assembled is via the activity of an ABC transporter that is responsible for translocation of O-antigen oligosaccharide across the inner membrane, and by the activity of a synthase. O-antigens synthesized in an ABC-transporter-dependent manner are typically homopolymeric and are assembled and terminated within the cytosol. Specific examples of *E. coli* strain O-antigen assembled in this manner include O8, O9, and O9a; homopolymers of mannose (17). Synthase-dependent O-antigen biosynthesis is rarely observed and typically generates homopolymeric or O-antigen with little diversity as compared to O-antigens assembled in a Wzy-dependent mechanism.

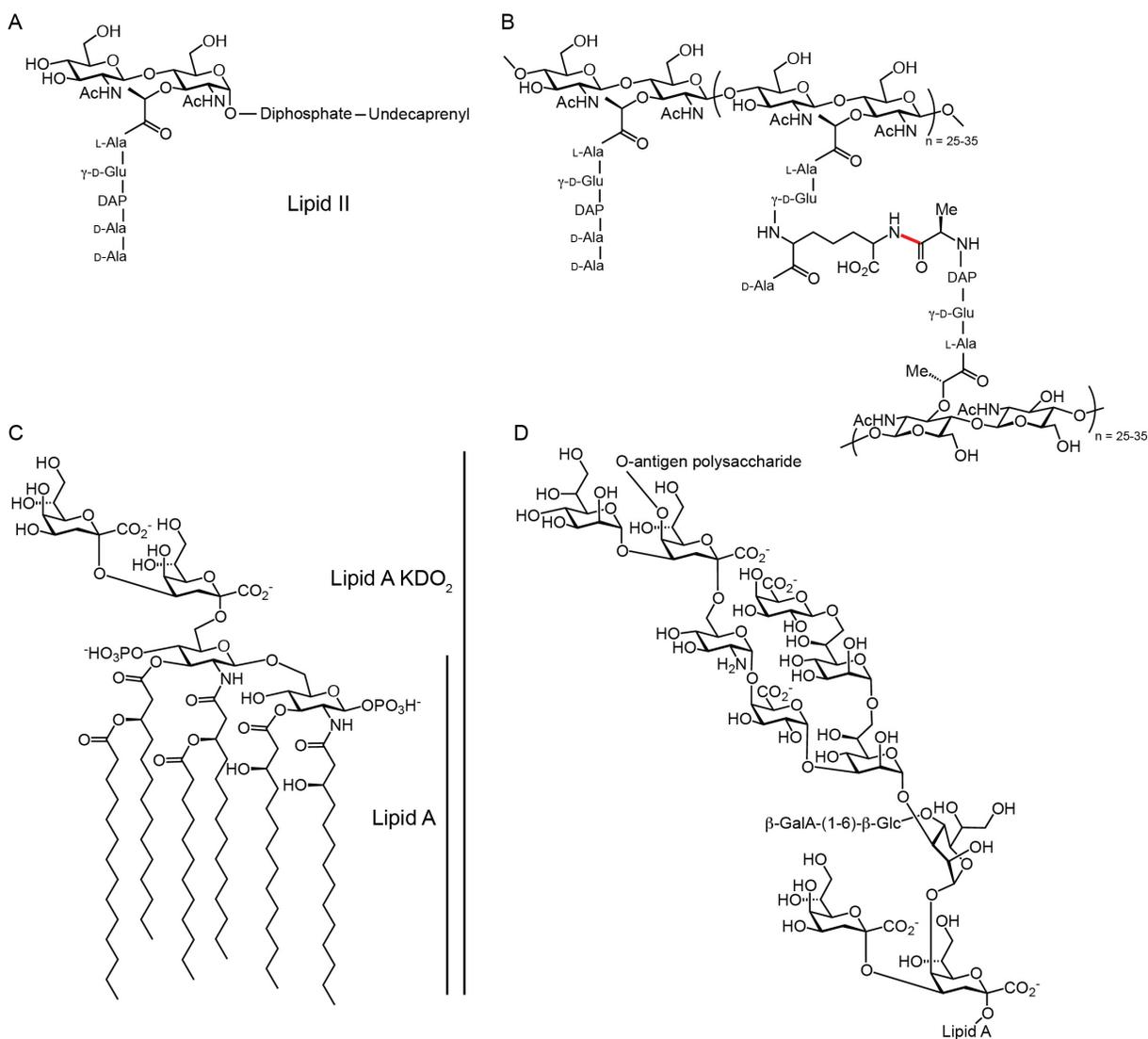


Figure 1-1. Chemical structures of cell surface glycoconjugates present on Gram-negative bacteria. (A) Structure of a generic lipid II monomer used in the polymerization on PG. Most of the species variation in the structure of PG arises from variation in the peptide stem. (B) Chemical structure of the product of a transpeptidation reaction between two adjacent strands of PG. The new bond formed between D-alanine and diaminopimelic acid is shown in red. The terminal D-alanine residues are often removed as a product of the crosslinking reaction. (C) Structure of lipid A and lipid A KDO₂ that form the base of all LPS molecules. Lipid A KDO₂ has been shown to be the minimal portion of LPS that is essential for cell viability. (D) Chemical structure of LSP core oligosaccharide from *K. pneumoniae*. The two KDO residues are included for clarity. This structure was chosen because *K. pneumoniae* uses only one LPS core structure in the assembly of full LPS.

1.3.3 Capsular Polysaccharide

Capsular and exo- polysaccharides (CPS and EPS, respectively) comprise another important class of cell surface glycopolymers found on bacteria. CPS and EPS can generally be differentiated based on the strength of their association with the bacterial cell wall. Commonly, CPS molecules are more adherent, often covalently bound to the cell, while EPS is loosely associated or secreted. Unlike LPS, CPS/EPS are found on Gram-negative and Gram-positive species (**Figure 1-3**). To differentiate these glycopolymers from the O-antigen, the CPS is referred to as K-antigen in Gram-negative species. CPS structures and LPS O-antigen are the major surface accessible glycans antigens in Gram-negative bacteria. In Gram-positive species CPS are the major antigenic glycan.

The capsule generated from capsular- and exo-polysaccharides is an important virulence factor for bacteria as it generates a physical barrier to small molecules, is antiphagocytic, and can prevent complement-mediated opsonization. Still, some capsules are extremely antigenic (18). The vast structural diversity of capsules is proposed to have arisen through pressure to evade the human immune system (19). The recognition of CPS can be exploited, evidenced by the development of a protein-conjugate polyvalent pneumococcal CPS vaccine that is protective against streptococcus infection (18). Similar to LPS O-antigen, some bacteria attempt to shield their capsule from the immune system. In an attempt to evade immune detection, *Bacteroidetes* spp. scavenge human L-fucose and incorporate it into CPS and surface proteins (20).

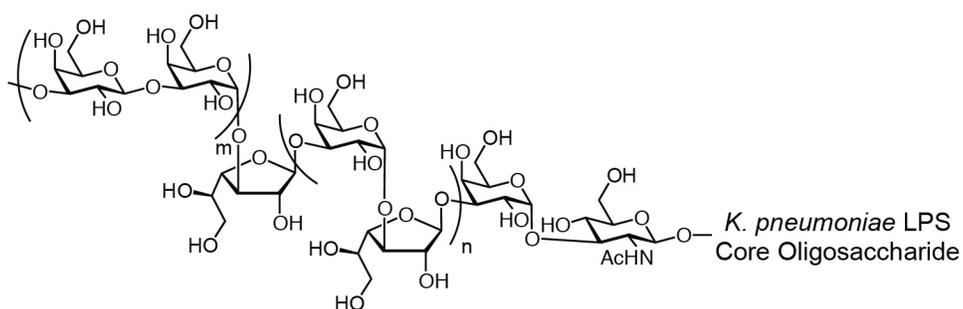
Within CPS molecules there is an incredible amount of carbohydrate diversity that is generated (**Figure 1-2**) (19). Each cell typically expresses one CPS serotype, although *Bacteroides fragilis* is known to transition between multiple distinct capsules (21,22). The

number of distinct CPS glycopolymers found amongst the members of a species can range from one, as is the case *Streptotoccus pyogenes*, to 80 identified K-antigens in *E. coli* (23), to 93 distinct capsule serotypes in *S. pneumoniae* (24).

Similar to O-antigen biosynthesis, Gram-negative bacteria can synthesize CPS in a Wzy-, ABC transporter-, and synthase-dependent mechanism (25). Only the Wzy- and synthase-dependent mechanisms have been described in Gram-positive bacteria (**Figure 1-3**). The genetic loci for Wzy-dependent capsule biosynthesis is similarly localized amongst closely related bacteria, this facilitates the rapid identification and characterization of CPS biosynthetic machinery. For example, in *S. pneumoniae* the capsular polysaccharide synthesis genes lie on the chromosome between *dexB* and *aliA* and are proposed to be transcribed as a single operon. Biosynthesis is initiated on the cytoplasmic face of the membrane on an undecaprenyl phosphate acceptor by the transfer of a sugar phosphate molecule, making a glycosylated undecaprenyl diphosphate. In all, 342 glycosyltransferases are encoded by the combined serotypes of *S. pneumoniae* to fully assemble each CPS oligosaccharide acceptor using NDP-sugars as donor molecules and additional enzymes for glycan tailoring (26). Each undecaprenyl diphosphate oligosaccharide is flipped across the membrane by Wzx and polymerized via en bloc transfer by the polymerase Wzy. Once fully polymerized, the capsule is transferred to the peptidoglycan or to another membrane acceptor (27).

In Gram-negative bacteria Wzy-dependent capsule is assembled similarly, although the capsule must be transported to the outer leaflet of the outer membrane and covalently linked to an acceptor. ABC transporter-dependent capsule synthesis in Gram-negative species has been shown to occur via assembly onto a poly-KDO linker covalently attached to a lyso-

phosphatidylglycerol molecule (28). After translocation to the outer membrane, the lipid is proposed to anchor the CPS to the cell. Capsules generated in a synthase-dependent manner use a single enzyme for acceptor priming, elongation, and transfer; the capsule from *S. pneumoniae* type 3 is generated in this fashion. It is polymer of alternating $\beta(1-3)$ -Glc and $\beta(1-4)$ -glucuronic acid that is polymerized onto a phosphatidylglycerol acceptor molecule and flipped to the outer leaflet of the cell membrane (29).



K. pneumoniae O1 LPS

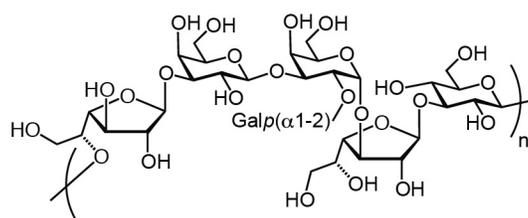
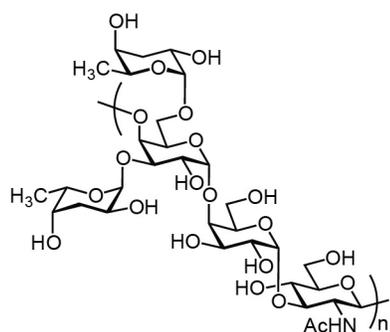
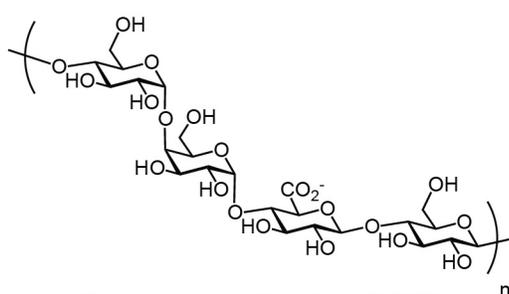
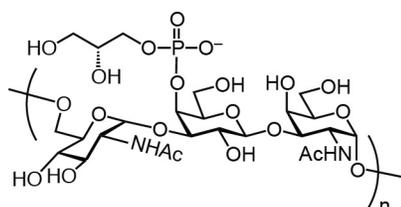


Figure 1-2. Chemical structure of several LPS O-antigens and capsular polysaccharides.

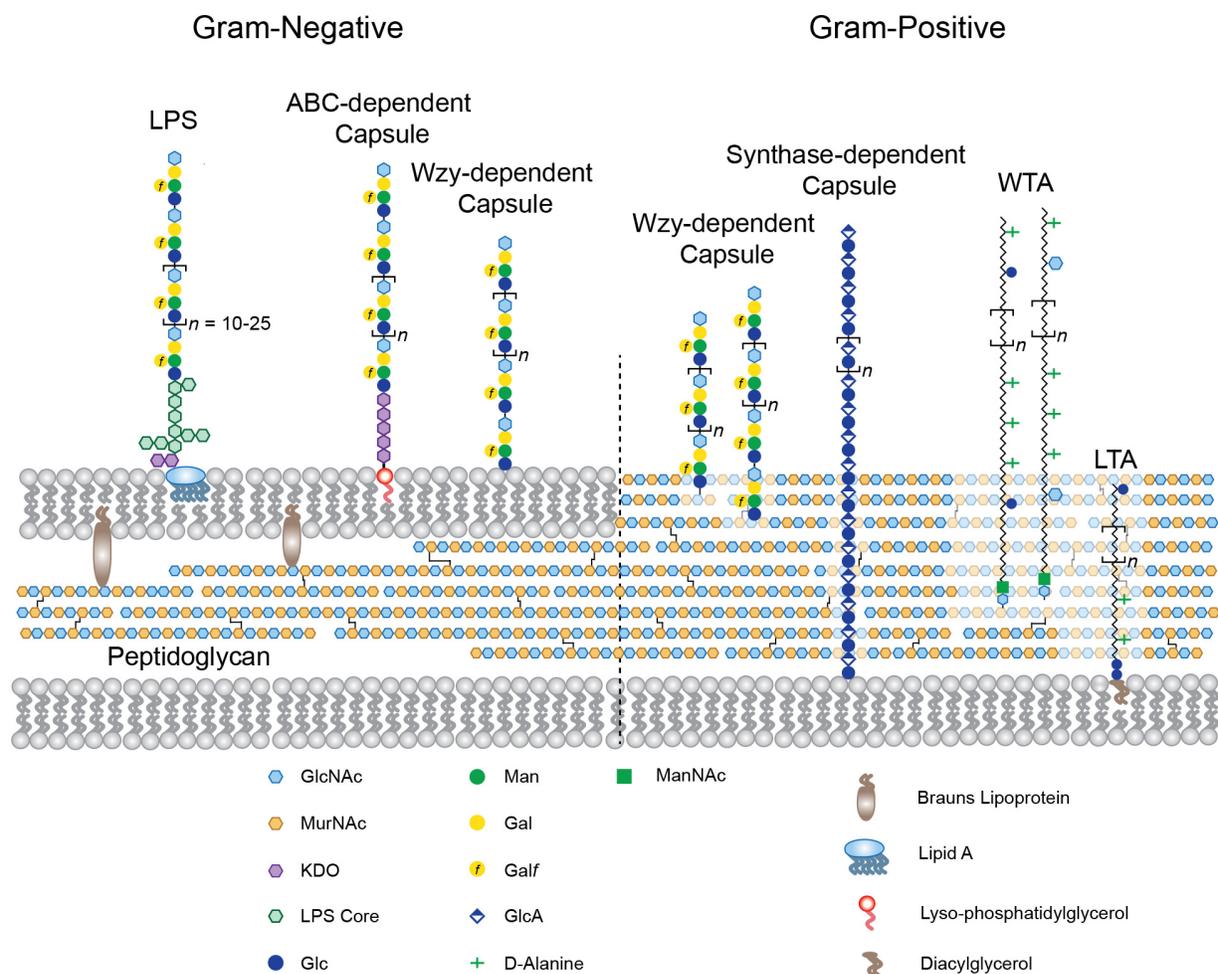


Figure 1-3. Schematic representation of the cell surface of Gram-negative and Gram-positive bacteria. Special attention has been paid to carbohydrate antigens. LPS, lipopolysaccharide; WTA, wall teichoic acid; LTA, lipoteichoic acid.

1.3.4 Teichoic Acids

Teichoic acids (TAs) are restricted to Gram-positive bacteria. They function in binding to cations and cation homeostasis, regulation of autolysin activity, placement of cell division machinery, repulsion of antimicrobial small molecules, peptides, and proteins, and general alteration of cell wall physiochemical properties (30). TAs were first described in the 1950's as glycerol- (Gro-) or ribitol-phosphate (Rib-P) glycopolymers elaborated with carbohydrates and amino acids, most notably D-alanine (**Figure 1-4**) (31). When TA fully functionalized with D-alanine the molecule possesses an high charge density. This property is critical in TA function.

TAs can be separated into distinct subtypes based on how they are anchored to the cell, and their charge. TAs covalently linked to the PG are referred to as wall teichoic acids (WTAs), while those attached to lipid and anchored into the cell membrane are referred to as lipoteichoic acids (LTAs) (32). LTAs generally do not protrude much past the PG, while WTA are much more exposed on the surface of the cell. Under laboratory conditions, WTA appear dispensable for bacterial growth, while LTA have been shown to be essential for viability of *S. aureus* (33).

Because of their localization on the surface of bacteria, TA glycopolymers are the target of multiple immune proteins and immune functions (30). Specifically they are important for bacterial adhesion to epithelial and endothelial cells and are ligands for scavenger receptors, several human lectins, and TLR-2 (34,35). Interestingly, zwitterionic polysaccharides such as TA have recently been shown to be CD4⁺ T-cell dependent antigens capable of giving rise to IgG antibodies specific for TA (36,37).

TAs are highly variable in their structure. Structural differences exist between WTAs and LTAs, between molecules synthesized by different bacterial species, and even between bacteria

grown under different conditions. Much of what is known about TA biosynthesis comes from work using *B. subtilis* 168, *B. subtilis* W23, and *S. aureus* (38). WTAs synthesis is initiated on the cytoplasmic leaflet of the membrane with the enzyme TagO catalyzing transfer of GlcNAc phosphate to an undecaprenyl phosphate acceptor. A N-acetylmannose (ManNAc) is added to the C(4) hydroxyl in a beta linkage by TagA to form a disaccharide primed acceptor. This acceptor is shared by WTA that contain a Gro-P polymer, or a combination of Gro-P and Rib-P. Once polymerized, the WTAs are elaborated by glycosyltransferases, flipped to the outer leaflet of the membrane by the ABC transporter TagGH, and attached to the C(6) hydroxyl of PG MurNAc residues. Upon extracellular localization, D-alanine residues are added in an ester linkage to generate the complete zwitterionic polysaccharide.

The presence of the lipid substituent in LTAs dictates that the biosynthesis pathways for WTA and LTA differ. *S. aureus* LTA is polymerized onto a Glc₂-diacylglycerol (Glc₂-DAG) acceptor glycolipid using phosphatidylglycerol as a donor for Gro-P by the enzyme LtaS (33). In the absence of Glc₂-DAG, LTA can be polymerized onto DAG. This entire biosynthesis of LTA, including subsequent glycosylation and D-alanyl esterification, is proposed to happen on the outer leaflet of the membrane (39). The above descriptions provide only a general overview of the structure of TAs. Species-specific variations are common, so much so that glycosylation of TAs is used as a taxonomic trait in some bacteria species (1). In addition, the structure is remarkably responsive to growth conditions as D-alanyl esterification is altered by pH, temperature, ionic strength, and phosphate availability (40).

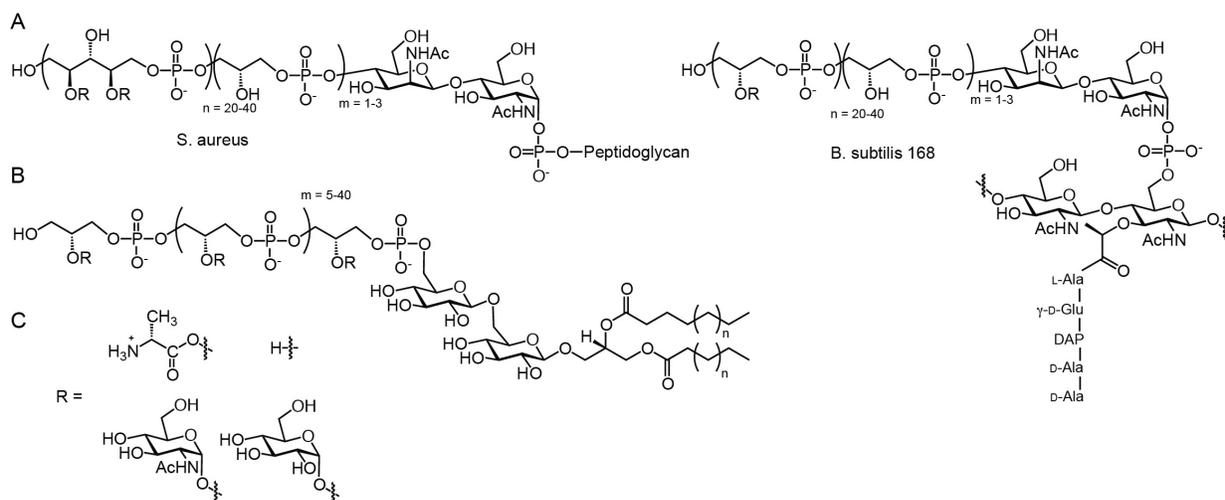


Figure 1-4. Chemical structures of TAs. (A) Structure of WTAs from *S. aureus* and *B. subtilis* 168. Examples from species are depicted to highlight the backbone can consist of TA polymerized of Rib-P or Gro-P. The *B. subtilis* 168 TA is shown linked to the C(6) hydroxyl of PG MurNAc. (B) Representative chemical structure of TA appended to a Glc₂-DAG tail. (C) Examples of the chemical substituents attached to the hydroxyl groups of TAs. The addition of D-alanine results in a zwitterionic structure that is important for many biological functions.

1.3.5 Others

Many other bacterially derived glycoconjugates are recognized by human immune lectins. In addition to those that I have described, I would like to introduce some that are more narrowly distributed amongst bacteria. Lipoglycans are part of the *Actinomyces* cell envelope. They are an excellent example of glycopolymers that are structurally unique, not generally distributed amongst bacteria, but extremely important to human health (41,42). The actinobacteria *Mycobacterium tuberculosis*, the causative agent of tuberculosis, synthesizes a complex cell envelope proposed to enhance the difficulty in treating infections by this pathogen (41).

An important mycobacterial lipoglycan is a heteroglycopolymer termed the mycolyl-arabinogalactan (mAG). The AG serves as a covalent linker between the PG and the waxy mycolic acids. The AG is covalently linked to the PG via a phosphate-GlcNAc-rhamnose

disaccharide. The disaccharide is extended by a linear polymer of Gal f residues termed the galactan (43). Branching from the Gal f polymer are large furcated polymers of arabinofuranose (Ara f). Some of the terminal Ara f residues are ultimately capped with long chain hydrocarbon mycolic acids (C70-C90), or mannose residues.

In addition to the mAG, a second carbohydrate-based large molecule is present in the cell envelope of *M. tuberculosis*, lipoarabinomannan (LAM). LAM is an amphiphile that contains a large carbohydrate portion composed of mannose and Ara f carbohydrates, linked to a phosphatidylinositol tail used in anchoring it to the cell membrane or the mycolic acids (44). The recognition of both of these cell wall components by the human immune system has been studied extensively to better understand the pathogenesis of *Actinomyces* bacteria (45).

Some bacteria and archaea synthesize an additional antigenic determinant localized outside of their traditional cell wall. This layer is called the surface layer, or S-layer. The S-layer is generated via self-assembly of a single monomeric (glyco)protein into a crystalline-like protein shell that encapsulates the cell (46). In many Archaea and Gram-positive bacteria, the S-layer protein is glycosylated (47). In bacteria, a large (50-150 monosaccharide units) repeating polysaccharide (15-50 repeats) that structurally resembles Gram-negative O-antigen is appended to one site on the protein. This glycan is typically attached to the protein via an O-glycosidic bond. In Archaea, the S-layer protein is typically glycosylated at multiple sites, although the glycans are typically shorter and are often attached via N-glycosylation (asparagine linked). The enzymes and glycosyltransferases responsible for synthesizing the S-layer glycan are sometimes encoded for in a cluster called the “S-layer glycosylation cluster (*slg*)”, aiding in prediction of S-layer glycans (47). Depending on the organism, the S-layer may maintain contact with the cell by

interacting with LPS, secondary cell wall polymers such as TA, or by direct interaction with the plasma membrane.

1.4 Fungal Glycoconjugates

Similar to bacteria, fungi are surrounded by a cell wall that is largely composed of glycoproteins and polysaccharides. The fungal cell wall performs many of the same functions as its bacterial counterpart; including resistance to osmotic pressure changes, resistance to antifungal agents, masking other antigenic epitopes, and mediation of cell:cell interactions (48). Alteration of the fungal cell wall can have a drastic impact on the morphology, virulence, and viability of cells.

Many fungal pathogens are associated with humans, but become virulent only when an individual is immunocompromised. This suggests that the innate and adaptive immune systems have evolved efficient mechanisms for detecting and controlling these pathogens (49). Owing to their localization on the surface of fungi, many of the antigens proposed to be important in immune surveillance and fungal virulence are cell wall glycopolymers. Significant effort in describing the structure of the fungal cell wall has focused on *Saccharomyces cerevisiae*, *Candida albicans*, *Cryptococcus neoformans*, and *Aspergillus fumigatus*, the following discussion will represent this work.

The fungal cell wall is composed primarily of three distinct glycans; chitin, beta-glucan, and glycoproteins (**Figure 1-5**). Many of the glycoproteins contain large polymers of mannose and are referred to as yeast mannans. Proximal to the cell membrane is chitin, a linear polymer of $\beta(1-4)$ linked GlcNAc. Chitin is synthesized and secreted simultaneously via vectorial synthesis by chitin synthases. Once extracellular, adjacent chitin polysaccharides form strong

intermolecular hydrogen bonding patterns that render chitin crystalline-like, and localize them parallel to the membrane.

By mass, the majority of the cell wall is composed of a polysaccharide termed beta-glucan. This polymer typically constitutes 50-60% of the entire cell wall by weight (50). Generally, beta-glucan is composed entirely of glucose residues joined via $\beta(1-3)$ bonds. Additional linkages, both alpha and beta, are present in small molar ratios and have been reported from different fungal species. Similar to chitin, beta-glucan is synthesized by membrane bound complexes in a vectorial manner. However unlike chitin, the beta-glucan polymer is branched. Beta-glucan synthases polymerize an extended linear polymer of roughly 1,500 glucose units. Throughout that polymer, additional glucose polymers are added to the C(6) hydroxyl of various glucose residues (51). Mature beta-glucan is the major structural component of the fungal cell wall and serves as a point of attachment for other cell wall components, hence it is indispensable for cell growth.

The last major component of the fungal cell wall is glycoproteins. In many species, glycoproteins represent 20-50% of the cell wall by dry weight (48). This collection of biomolecules includes large structural proteins anchored into the membrane by addition of a glycosylphosphatidylinositol (GPI) anchor, some that are covalently attached to chitin or beta-glucan, and others that noncovalently associate with cell wall polysaccharides. Fungal cell wall glycoproteins are highly N- and O-glycosylated. The N-glycans contain an extreme amount of mannose, often 50-200 α -Man residues. It is these highly mannosylated glycoproteins that are referred to as yeast mannan or yeast mannoproteins. The appended O-glycans are typically much smaller, usually consisting of 1-5 monosaccharide residues.

The overall structure of the fungal cell wall is remarkably similar across species. One example of variation is the inclusion of a significant amount of D-Galf in cell wall galactomannan and galactomannan modified glycoproteins from *A. fumigatus* (52). Some species express an additional polysaccharide layer termed a capsule that is found directly outside of their traditional cell wall. This complex is important for the infectivity and virulence of the human pathogen *C. neoformans* (53,54).

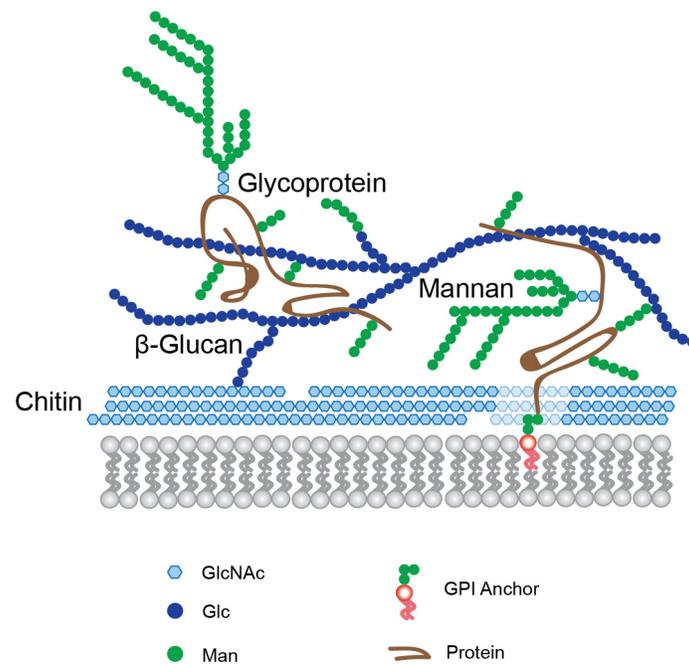


Figure 1-5. Schematic representation of a fungal cell wall. The major surface accessible carbohydrate antigens are depicted.

1.5 Lectins

With the many aforementioned glycopolymers localized to the surface of microbes, humans have evolved diverse mechanisms for detecting microbial carbohydrate antigens. These mechanisms include the adaptive immune response and antibody generation, cellular surface bound receptors such as TLRs, innate immune cells such as neutrophils and dendritic cells, and lectins both soluble and cell-associated. Here, I will discuss human humoral immune lectins. Specific attention will be paid to their structure, expression, localization, ligand selectivity, and proposed immune functions (55).

1.5.1 Ficolins

Ficolins are soluble immune lectins that share a carbohydraterecognition domain (CRD) and are similar in overall architecture. Ficolins contain an N-terminal region that contributes to disulfide-mediated oligomerization, a central collagen-like domain, and a C-terminal fibrinogen-like domain (FBG domain) that contains the CRD (**Figure 1-6**) (56). The FBG domain CRD of ficolins is related to the gamma chain of fibrinogen in sequence through roughly 50 amino acids that include 24 invariant residues, and 4 highly conserved cysteine residues.

In addition to the three human ficolins, ficolin proteins are found in many other animals, including mammals such as porcine and mouse, amphibians, and ascidians. The FBG domain CRD of ficolins is evolutionarily related to the tachylectin lectin from horseshoe crab, suggesting a primitive role for FBG domain-containing lectins in innate immunity (57). Ficolins are functionally and structurally related to collectins in that they both contain a central collagen-like domain and C-terminal CRD, a major distinction between these two groups of lectins is that collectins use a C-type CRD, as opposed to the FBG domain of ficolins (**Figure 1-6**) (58).

Functional ficolin is a homopolymer composed of three roughly 35 kDa polypeptides which associate via their collagen-like domain. Trimers assemble into a higher oligomeric structure through additional interactions, including N-terminal inter-trimer disulfide bonds. Electron microscopy indicates that native ficolin oligomers are bouquet-like in structure, with all the CRDs located in a single direction (59). Efficient ficolin oligomerization is important for multivalent binding to carbohydrate ligands, and for activation of the lectin pathway of complement through binding to serum mannose-binding lectin-associated serine proteases (MASPs) (60). Specifically, larger ficolin oligomers have been shown to more efficiently bind MASPs and activate complement (61). Complement is important to the immune function of ficolin proteins as it initiates multiple immune response including cell killing and opsonization, enhanced phagocytosis, immune cell chemotaxis, and cytokine/chemokine production.

Humans encode three ficolins; ficolin-1 or M-ficolin, ficolin-2 or L-ficolin, and ficolin-3 or H-ficolin. The three human ficolins are similar, but have nonidentical expression localization profiles, ligand specificities, and biological functions. All three ficolins contain an N-terminal signal peptide, suggesting they are either secreted or destined for the cell surface. L- and H-ficolin are expressed almost exclusively in the liver and are found largely in serum (**Figure 1-7**). The concentration of L-ficolin in human serum ranges from 3-5 $\mu\text{g/mL}$, while H-ficolin is present at higher concentrations, 18-33 $\mu\text{g/mL}$ (62). In addition to high expression in the liver, H-ficolin is also expressed in the lung. In contrast to serum ficolins, M-ficolin serum concentrations are low and have been reported between 0.06 and 1 $\mu\text{g/mL}$ (63). Instead of robust liver expression, M-ficolin is expressed in the lung, bone marrow, and by cells including peripheral leukocytes, monocytes, neutrophils, and type II alveolar epithelial cells (64,65).

Within these cells types M-ficolin has been localized to the cell surface and to secretory granules.

The carbohydrate binding of ficolins have been explored extensively and have revealed both overlapping and unique profiles. All three human ficolins bind N-acetylated carbohydrates, such as GlcNAc and GalNAc. Most ficolin ligand binding occurs near a protein bound calcium ion, although the calcium ion is not directly involved in ligand coordination. In addition to the shared affinity for acetylated ligands, unique specificities exist (66). For example, M-ficolin has been shown to bind the sialic acid N-acetylneuraminic acid (Neu5Ac), an important human self-antigen. L-ficolin possesses affinity toward sulfated ligands, including heparin. H-ficolin has been reported to bind D-fucose and galactose (67). It should be noted that the lectin activity of H-ficolin is less well understood than the other ficolins as glycan microarray screening has failed to yield ligands (66).

Ficolin ligand specificity has been probed by assaying binding to microbially derived glycopolymers and to whole cells. L-ficolin binds several microbial components including PG, LTA, and beta-glucan (60). Extensive profiling of ficolin binding to intact microbes has revealed affinity toward Gram-positive and Gram-negative bacteria, encapsulated bacterial strains, viruses such as influenza, and eukaryotic protozoa (68,69). Structural biology has informed about the mechanism of ligand recognition and self-non-self discrimination (70). Two interesting properties of ficolin:ligand interactions include the discovery of multiple ligand binding sites in L-ficolin (71), and a pH-dependent conformational change that can modulate M-ficolin ligand binding (72).

Ficolins perform biological roles in addition to complement activation via MASP binding. These biological tasks are carried out through protein:protein interactions as well as ligand recognition by the CRD. Interestingly, L- and H-ficolin, and mannose-binding lectin (MBL) have been shown to be important for clearance of apoptotic and necrotic human cells (73,74), and are able to bind released mitochondria (75). Ficolin clearance of cellular debris is proposed to function through ficolin binding C1q receptor/calreticulin on immune cells via the ficolin collagen-like domain (76).

A human patient deficient in H-ficolin due to a homozygous frameshift mutation was recently identified (77). This patient has frequent lower respiratory tract infections and has battled several episodes of bacterial pneumonia. Several ficolin single nucleotide polymorphisms have been identified and associated with serum protein concentration and disease (68,78). In summary, the findings suggest a decrease in serum ficolin concentrations is associated with infection and inflammatory disease, while upregulation of ficolins are linked to autoimmunity. Mouse models deficient in either ficolin-A (a mouse serum ficolin), ficolin-B (a mouse non-serum ficolin), or both ficolin-A and -B, have been generated (79). While there were no abnormalities in appearance, weight, and reproduction, all three mouse strains were more susceptible to intranasal infection by *Streptococcus pneumoniae* strain D39, suggesting an important role for both serum and non-serum ficolins in immunity against pneumococcal infection (79). Indeed, ficolins play an important role in innate immunity through recognition of carbohydrate ligands.

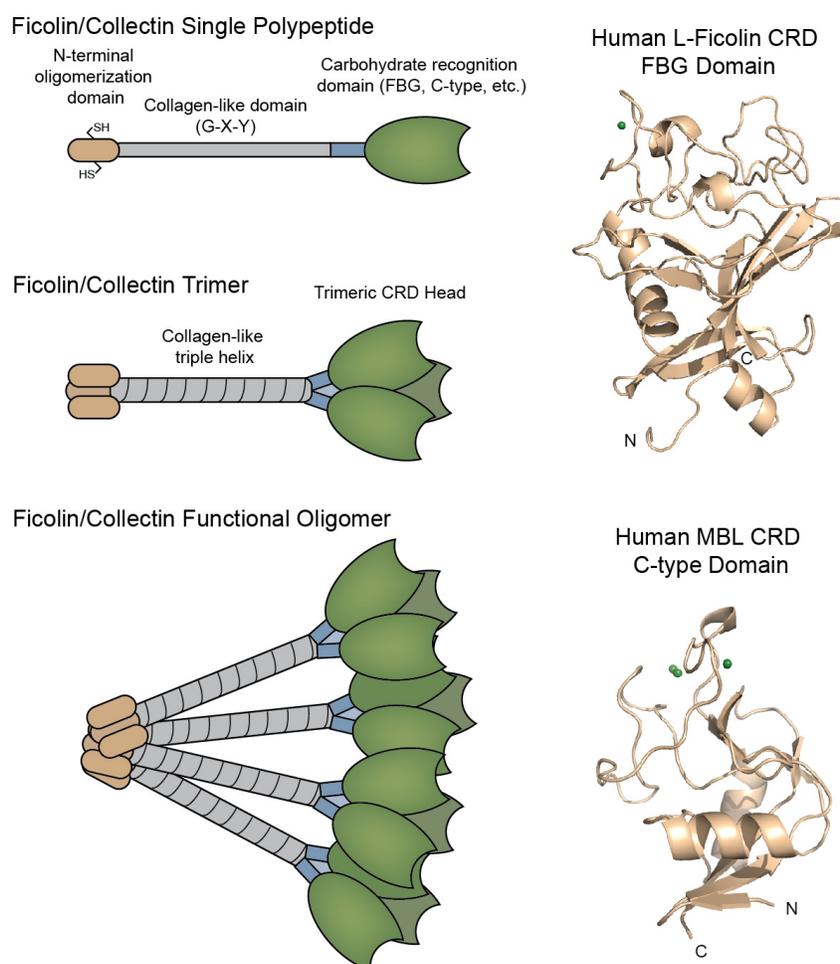


Figure 1-6. Schematic of the assembly of function ficolin/collectin oligomers. An oligomer of trimers is needed for MASP binding and complement activation. On the right are structural examples of the CRDs used by ficolins and collectins; FBG domain CRD from L-ficolin (PDB ID: 2J3U) (71), and the C-type lectin domain from MBL (PDB ID: 1HUP) (80). Protein bound calcium ions are shown as green spheres and the N- and C-termini are highlighted with a capital N and C, respectively.

1.5.2 Collectins

An independent family of proteins that share some structural and functional characteristics with ficolins are the collectins. Human collectins include MBL, surfactant proteins A and D (SP-A and SP-D), collectin-10 or collectin liver 1 (CL-L1), collectin-11 or collectin kidney 1 (CL-K1), and collectin-12 or collectin placenta 1 (CL-P1). The collectins are all soluble, fluid-phase lectins, with the exception of CL-P1, which contains a transmembrane helix. Collectins are important for vertebrate innate immunity (81). While some collectins such as MBL are distributed across species, other collectins such as the bovine protein conglutinin (82) are more restricted.

Collectins are structurally similar to one another and to ficolins in that they contain an N-terminal oligomerization domain, a central collagen-like domain, and a C-terminal CRD (**Figure 1-6**). Collectin monomers assemble via their collagen-like domain into trimers of three identical (or nearly identical in the case of SP-A) polypeptides. One major distinction between collectins and ficolins is the nature of their CRD. Ficolins use a FBG domain and the C-terminal CRD of collectin proteins is a traditional C-type lectin domain (83). More subtle structural differences include that collectins typically have shorter collagen-like domains than ficolins and that most collectins contain a small coiled-coil region between their collagen-like domain and CRD. The native structure of some collectins, specifically MBL and SP-A, resembles the bouquet-like arrangement visualized for ficolins, while SP-D assembles into more of a crucifix or wheel shape (84).

Collectins perform discrete biological functions based on their expression localization (**Figure 1-7**). MBL, which is produced predominantly by the liver and is found in serum at

concentrations between 0.5 and 4 $\mu\text{g}/\text{mL}$ (85). Similar to ficolins, serum MBL associates with MASPs and is able to activate the lectin pathway of complement. The lung is the predominant site of SP-A and SP-D expression. SPs are produced by several resident cell types, largely alveolar type II cells, and secreted onto the epithelial surface. While SP-A production is restricted to the lung, SP-D is also expressed by epithelial cells of the small intestine and by other cells at mucosal surfaces (86). SP-A and -D associate with surfactant proteins B and C to generate lung surfactant. This mixture coats microbes resulting in agglutination and opsonization.

Only recently has CL-K1 been identified and studied (87). CL-K1 is expressed by the kidney and liver, and is present in human plasma at 1-3 $\mu\text{g}/\text{mL}$. Like MBL, CL-K1 interacts with MASPs and can activate complement on the surface of *Candida albicans* (88,89). Similar to CL-K1, CL-L1 has only recently been identified (90). CL-L1 is produced predominantly by the liver and is secreted and localized to human plasma at concentrations from 1-5 $\mu\text{g}/\text{mL}$ (91). Recently, a CL-L1 and CL-K1 heterotrimeric complex termed CL-LK, was identified and shown to be complement competent (92).

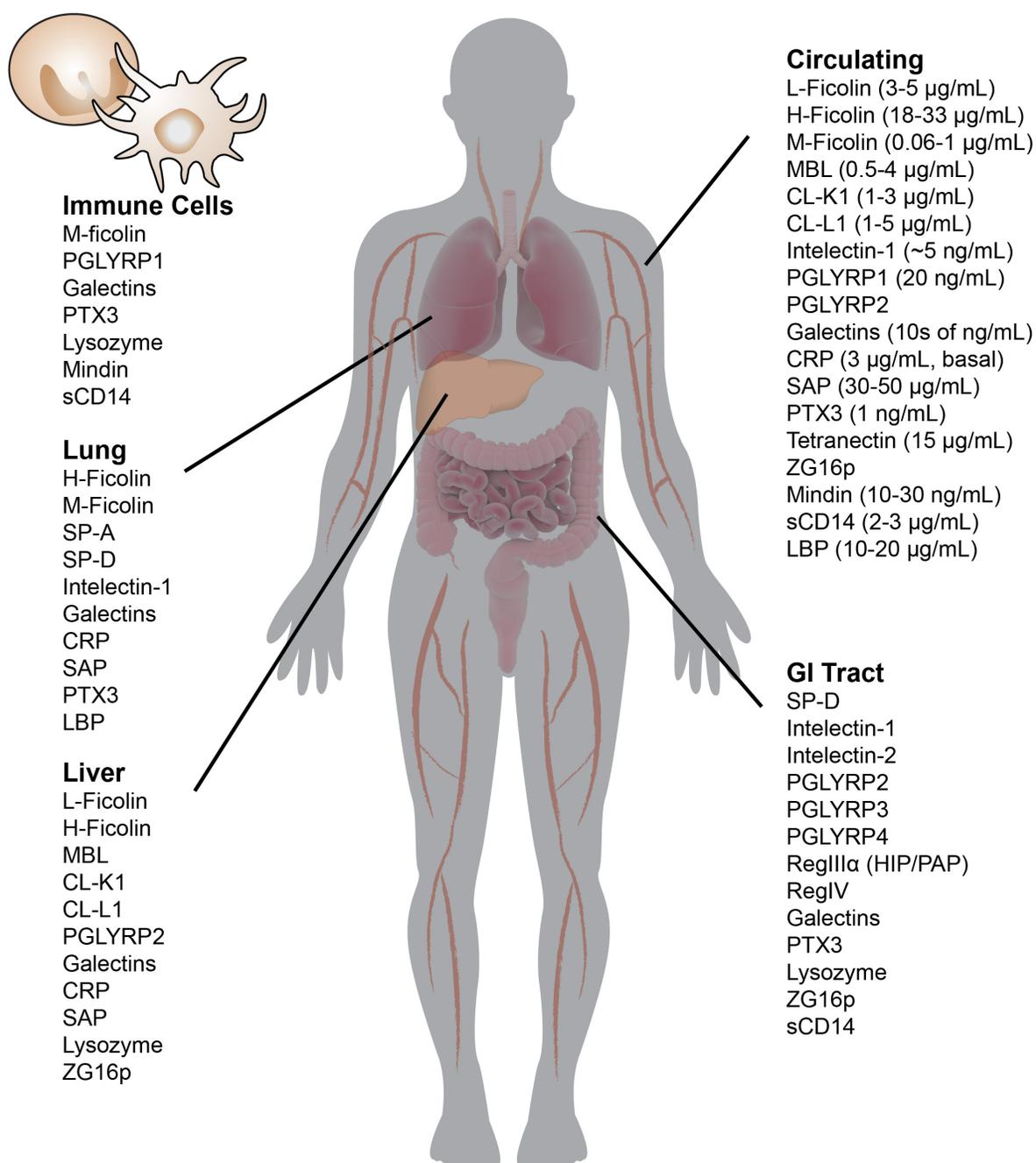


Figure 1-7. Expression and localization of human humoral immune lectins. Lectins expressed by multiple tissues and cell types are included in all locations. The reported circulating concentrations are taken from within this review.

All five human collectin CRDs discussed here contain an EPN motif that is hallmark of mannose-binding C-type lectin domains (93). The EPN motif in collectins necessitates carbohydrate ligands that contain di-equatorial hydroxyl substituents at positions C(3) and C(4) of a ligand ring. Those two hydroxyls displace ordered water molecules and coordinate directly to a protein bound calcium ion. Traditionally, this type of ligand binding was described as occurring in calcium binding site 2. Examples of ligands that bind in this manner include D-mannose, L-fucose, Glc, GlcNAc, and ManNAc. Importantly, residues outside of the calcium-binding pocket of each protein have been shown to influence ligand binding affinity and specificity. These contribute to differences in the monomeric affinity of each protein:ligand complex, but generally they can be approximated in the millimolar range. An alternative recognition mechanism for L,D-heptose by SP-D was recently revealed in which the exocyclic C(6) and C(7) hydroxyl groups coordinate the protein-bound calcium ion (94). These hydroxyls adopt a conformation almost superimposable with the vicinal equatorial hydroxyls of traditional collectin ligands.

The carbohydrate binding specificity of collectins is responsible for binding to many self, and non-self antigens. Many studies have demonstrated MBL binding to clinically important microbes such as *E. coli*, group A streptococci, *Staphylococcus aureus*, *Aspergillus*, *Candida*, and *Leishmania* (95). Binding occurs through recognition of microbial glycoconjugates such as LPS, LAM, yeast mannans, and enveloped virus high mannose glycans (81). Many microbes also functions as specific ligands for surfactant proteins-A and -D through recognition of LPS core antigens, high-mannose O-antigens polysaccharides, LTA, and PG (96). Surfactant proteins-A and -D are proposed to perform vital roles in anti-viral immunity (97). This is proposed to occur

through binding to viral high mannose oligosaccharides, leading to enhanced agglutination. Surfactant proteins also have affinity toward lipids; SP-D prefers to bind phosphatidylinositol (PI) while SP-A has an affinity toward phosphatidylcholine (PC) (98,99). The affinity toward PI can be rationalized via its polar carbohydrate head group, while the mechanism of PC binding is not fully understood (100).

Considerably less is known about CL-K1 and CL-L1 ligand recognition. Fucose and mannose glycosides have been reported as monomeric ligands for CL-K1, while immobilized rough *E. coli* LPS, yeast mannans, and intact *C. albicans* are examples of more biologically complex ligands (87). It should be noted however, that the lectin activity of CL-K1 appears much less robust than that of other collectins. CL-LK heteromeric complexes showed appreciable binding to immobilized deoxyribonucleic acid (DNA) and to yeast mannan, with binding leading to complement activation (92).

Another important function of collectins is facilitated through binding to self-epitopes (101). Specifically, collectins bind aberrant mammalian glycosylation, such as upregulated Lewis A and Lewis B antigens on cancer and apoptotic cells (102,103). Deletion of MBL in mice was shown to result in decreased clearance of apoptotic cells, likely through binding aberrant glycan ligands (104). Within the lung, SP-A has been shown to stimulate phagocytosis of apoptotic neutrophils by alveolar macrophages (103). Similar to ficolins, collectins are proposed to facilitate enhanced phagocytosis through cC1q receptor/calreticulin binding to their collagen-like domain (105).

Early analysis of human data revealed that roughly 5% of the population expresses low levels of MBL. These individuals display defects in complement activation and opsonization,

and are at risk of recurrent infections (106,107). Circulating MBL levels are strongly influenced by genomic variation within the population (85). When additional immunodeficiencies are coupled with low basal serum levels of MBL, individuals are at an increased risk of infection (108). Mouse models have been used to demonstrate that mice deficient in MBL (mice contain two copies of MBL, MBL-A and MBL-C) are highly susceptible to infection by multiple microbial species including *S. aureus*, *P. aeruginosa*, and HSV-2 (109).

Circulating levels of SP-A and -D are also influenced by genetic variation within the population (110). Association with specific disease states, and increased risk of infection have been reported (111). For example, polymorphisms in the SP-D gene have been associated with susceptibility to chronic obstructive pulmonary disease (COPD) (112). Mice deficient in both SP-A and -D have been generated, and studies suggest major defects in immune cell regulation and maturation. SP-D deficient mice were also found to have altered surfactant lipid profiles, defects in immune cell regulation and activation, and altered alveolar macrophage chemotactic activity (111,113,114). Interestingly, mice devoid of SP-A or SP-D exhibited differential responses during bacterial infection, suggesting key differences in each proteins role during an immune response (115).

Much less is known about the genetic variability of CL-L1 and CL-K1 and at this point, mice deficient in either of these proteins have not been generated. One interesting exception to this is a recent study that identified several mutations in human CL-K1 (116). The authors report that mutations in the CL-K1 and MASP1 genes cause 3MC syndrome, a rare autosomal recessive disease that affects development and cognition. Using zebrafish as a model, the authors demonstrated that both proteins were important for development and likely function as

chemoattractants in neuron migration. How, and if, the lectin activity of CL-K1 is driving this process remains an open question.

1.5.3 *Intelectins*

The intelectin family of lectins is a recent addition to human innate immune lectins, relatively speaking. One defining feature of intelectins that was noticed soon after their discovery was their use of a FBG domain CRD. This resulted in intelectin proteins being assimilated with ficolins, although several substantial differences in structure and function were soon thereafter noted. These differences resulted in the proposal of intelectins comprising their own class of lectins, X-type lectins (117). Within this thesis, substantial contributions to our understanding of the structure, ligand specificity, ligand recognition properties, and clues toward the biological function of intelectins will be presented.

Intelectin proteins were first identified in *Xenopus laevis* oocytes as helping to generate the jelly that coats a fertilized egg and prevents polyspermy (118). Since then, homologous proteins have been identified in species ranging from tunicates to humans (119). Most mammals, including humans, encode for two or more intelectins. Both human intelectin-1 (hIntL-1) and -2 (hIntL-2) are encoded on chromosome 1, and they share greater than 80% sequence identity. It is important to note that within the scientific literature, hIntL-1 is also referred to as omentin, and as the intestinal lactoferrin receptor. This is based on reports suggesting a role for hIntL-1 as an adipokine (120), and its reported affinity toward the iron chelator lactoferrin (121), respectively.

Intelectin proteins were originally reported to be similar to ficolins based on the presence of a FBG domain in their C-termini (122). However, this region makes up only a small portion of the protein and significant structural and sequence differences exist. For example, intelectins do

not contain the N-terminal oligomerization or collagen like domain required for MASP binding and complement activation. Additionally, and the proposed C-terminal CRD of intelectins is significantly larger than ficolins. Similar to ficolins and collectins, multivalency is likely key for intelectin ligand binding and biological activity. hIntL-1 has been shown to assemble into a disulfide linked trimer via three intermolecular disulfides (123). Interestingly, these disulfides are not strictly conserved; mouse intelectin-1 is lacking the required cysteine amino acid residues. How this affects trimerization remains an open question.

Mammalian intelectins are expressed by mucosal tissues such as the lung and intestine. They are predominantly expressed by lung and intestinal goblet cells, and intestinal paneth cells (124,125). These cells are responsible for secreting other antimicrobial innate immune proteins that help generate mucosal immunity. In addition, hIntL-1 expression has been reported in the heart, lung, small intestine and adipose tissue, while hIntL-2 expression appears restricted to the small intestine (119,126). Within the diabetes and nutrition field, where hIntL-1 is referred to as omentin, elevated serum levels are suggested to be predictive of metabolic disease. While initial omentin studies reported serum concentrations in the hundreds of ng/mL, recent reports typically are single digit ng/mL for serum omentin (120,127). The low levels of hIntL-1 in serum argue against a robust immune function for intelectin proteins within the circulatory system.

In contrast to serum levels, infection of mouse intestine with intestinal parasitic nematodes results in significant upregulation in mouse intelectinal tissue (128). Additionally, experimental models of antigen challenge or asthma have demonstrated intelectin upregulation in mouse lung (129,130). These observations suggest a role for intelectin proteins in innate

immunity at mucosal surfaces. However, direct evidence and a mechanism for this is still lacking.

Intelectins are calcium ion-dependent lectins that are not canonical C-type lectins (83,126). Instead, intelectins contain a FBG domain. hIntL-1 has been shown to exist as a disulfide linked homotrimer, suggesting multivalent binding to carbohydrate ligands. Using soluble carbohydrates, hIntL-1 binding to galactose-sepharose was assayed in a competitive manner. The results of this work suggested affinity for furanoside carbohydrates such as D-ribose and a D-Galf containing disaccharide, although the affinity and specificity of carbohydrate binding was low (126). Galf is the thermodynamically disfavored, five-membered ring isomer of galactose. Galf containing cell surface glycans are found on many bacteria, fungi and protozoans, but are absent in mammals (131,132).

Actinomycetes bacteria were used to assay hIntL-1 binding to biological ligands. hIntL-1 was reported to bind purified AG from *Nocardia rubra* (126), and to bind directly to *Mycobacterium bovis* bacillus Calmette-Guérin (BCG) cells (133). A similar monosaccharide ligand specificity was reported for mouse intelectin-1 (123). Considerably less is known about the binding specificity of hIntL-2 and mouse intelectin-2. The significant upregulation of mouse intelectin-2 upon mouse infection with intestinal nematodes suggests affinity toward nematode glycans, although evidence for this is lacking (134). Later, Mammalian intelectins were shown to interact with the surface of mammalian cells, albeit not through a traditional GPI anchor as had been reported previously, and stimulate phagocytosis of BCG by mouse alveolar macrophages (133).

Defining a role for intelectins in mammalian physiology, and specifically immune function, is an area of active research. One clue toward a function for IntLs comes from its upregulation in mice upon intestinal colonization by parasitic nematodes (128,135). However, transgenic mice with increased intelectin expression in the lung failed to increase the clearance rate of the helminth parasite *Nippostrongylus brasiliensis*, or *M. tuberculosis* (136). Revisiting similar experiments using mice devoid of a specific IntL, or both, may help reveal a potential role in immunity. hIntL-1 has been identified in a genome-wide association study (GWAS) for genetic loci correlated with susceptibility to Crohn's disease (137). While it is early to speculate on the biological meaning of this, the data suggests a role in microbiome regulation and/or maintenance of intestinal barrier function (138).

1.5.4 Reg Proteins

Multiple human proteins recognize peptidoglycan, the conserved glycopolymer found on all bacterial cells. One family of peptidoglycan binding proteins is the Regenerating proteins, or Reg proteins. In total, humans encode five Reg proteins, RegI α , RegI β , RegIII α , and RegIII γ are found on chromosome two, while RegIV is encoded on chromosome 1. Human RegIII α , also known as hepatocarcinoma intestine-pancreas/pancreatic associated protein or HIP/PAP, is most similar to murine RegIII γ . Studies using murine RegIII γ have provided foundational results elucidating toward the biological function of Reg proteins.

Reg proteins were originally identified in a screen for proteins upregulated during pancreatic islet regeneration (139). Interestingly, Reg proteins were also identified in pancreatic juice as a small glycoprotein able to inhibit the growth of calcium carbonate crystals that can result in pancreatic stones, in this context they were given the name lithostathine (140). Current

literature refers to lithostathine as RegI α . The Reg proteins have been best studied in humans and mice, although they are probably found generally in mammals based on available genetic information and physiology (141).

While RegI proteins are upregulated in the pancreas during islet regeneration, other Reg proteins are involved in microbial recognition and function in immunity. These proteins are expressed in the gastrointestinal (GI) tract. Human RegIII α is expressed in the small intestine, and during infection, expression is increased in the large intestine. Specifically, Reg proteins are secreted by intestinal enterocytes and are found in the secretory granules of Paneth cells (142,143). Human RegIV is also expressed in the GI tract (144). Interestingly, it has been shown that symbiotic bacteria are required for the expression of mouse RegIII γ , the homolog of human RegIII α (143). In fact, multiple antimicrobial peptides and proteins were found to be significantly upregulated upon microbial colonization. Mechanistically, stimulation of TLR-MYD88 by PG and LPS were found to drive this (145,146).

Structurally, Reg proteins are small soluble lectins composed almost entirely of a C-type lectin domain (83,147,148). In determining their three-dimensional structure, it was quickly noted that Reg proteins appear to have lost their ability to coordinate calcium ions in the canonical calcium ion binding site. Thus, glycan binding by Reg proteins has been shown to be independent of calcium ions and occur at sites distinct from where traditional C-type lectins bind ligands (148,149). Much of the literature on the lectin and immune function of Reg proteins has focused on the RegIII family and RegIV. Thus, the remainder of this discussion will focus on human RegIII α (HIP/PAP), and human RegIV.

Reg proteins exist largely as soluble monomeric proteins when not engaged with a target ligand. They contain a small N-terminal peptide that is sensitive to the protease trypsin, and recently this cleavage event was found to regulate the bactericidal activity of the protein (150). Another interesting structural property reported early in Reg protein research was the ability of the protein to undergo a pH-dependent reversible transformation from a soluble globule, into a fibril (141,151). Recently, the fibril formation by RegIII α was examined by cryo-electronmicroscopy (EM) (152). Modeling the structure of monomeric RegIII α into the EM density map revealed a series of protein rings generated from hexameric oligomerization. Hexameric oligomerization has subsequently been shown to be important for the bactericidal activity of RegIII α .

Early clues toward the lectin function of RegI α were realizing it contained a C-type lectin domain, and its ability to aggregate both Gram-positive and Gram-negative bacteria (140). A mechanistic understanding of what, and the how, RegIII α bound its carbohydrate ligand was not realized until its structure was solved in complex with PG using protein nuclear magnetic resonance (NMR) (148). This study highlighted that RegIII proteins lack canonical carbohydrate and calcium ion binding motifs, but interact with the carbohydrate portion of PG. The authors described an essential EPN motif that is used for ligand binding, and is essential for bactericidal activity (**Figure 1-8**). Interestingly, the amino acid tripeptide EPN is used in canonical C-type lectins for calcium and Man/Glc binding (93), but within RegIII proteins the motif is found outside of the traditional calcium binding site.

Human RegIV also recognizes glycosylated ligands in the absence of calcium ions and a recently solved three-dimensional structure suggested a recognition mechanism (149). RegIV

was shown to interact with yeast mannan in two independent sites, one near the mutated traditional calcium ion binding site, and one at a site similar to where RegIII α binds PG. The authors propose that Reg proteins lost the requirement for calcium to prevent ligand loss in the low pH environment of the mammalian GI tract.

A significant result alluding to the function of RegIII proteins came with the discovery that they are directly bactericidal (143). Mouse RegIII γ and human RegIII α were shown to damage the cell surface of multiple Gram-positive bacteria, resulting in cytoplasmic leaking and cell death. It was postulated that the outer membrane of Gram-negative species prevents access to the PG, and thus attenuated the killing of Gram-negative cells. Recently, structural biology, protein biochemistry, and biophysics revealed that RegIII α kills bacteria through formation of a pore that depolarizes the cell membrane (152). Interestingly, RegIII α needs to be devoid of its N-terminal peptide to form the bactericidal ring structure. Binding to acid lipids was shown to be essential for hexamer intercalation into the membrane.

This mechanism is similar to how other antibacterial peptides function. Such a general mechanism of killing requires distinct modes of regulation to prevent damage to host eukaryotic cells. One unique example from antibacterial peptides is the use of disulfide bond reduction potential to render oxidized peptide less toxic (153). Instead, RegIII proteins employ the trypsin cleavable N-terminal peptide (150) and a requirement for acidic lipids. The N-terminal peptide prevents premature oligomerization, and asymmetric eukaryotic membranes are devoid of acidic lipids in the outer leaflet.

Most of what is known about the function of Reg proteins comes from genetic deletion experiments in mice. The ability of RegIII γ to prevent intestinal colonization by multiple

microbial pathogens was first explored using a homozygous deletion of MYD88, an adaptor protein for TLR signaling. Abrogation of Toll-like receptor signaling severely attenuates intestinal RegIII γ expression. Under this condition, intestinal colonization by *Listeria monocytogenes* and VRE was significantly higher (146,154). Exogenous RegIII γ , or blocking antibodies, were used to demonstrate the specific immune function of RegIII γ in preventing intestinal infection. A transgenic mouse devoid of RegIII γ revealed that the lectin performs an essential role in spatially segregating the mammalian intestinal microbiota from host epithelium (145). By collaborating with the mucus layer of the intestine, RegIII γ helps maintain the segregation that is essential for intestinal health. This function may prove important to the pathology and progression of irritable bowel disease (IBD) as Reg protein expression, and increased microbial interaction with the intestine, have been associated with the disease (155,156).

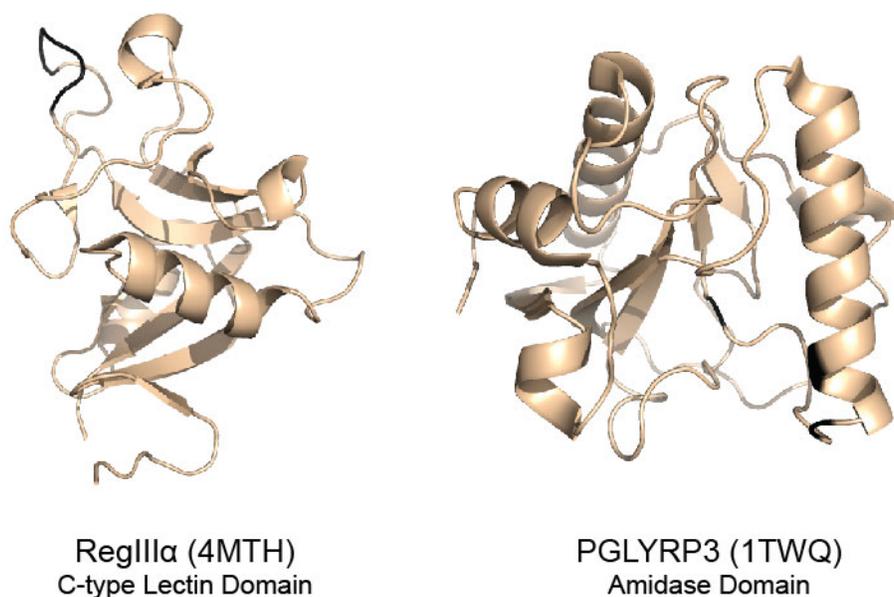


Figure 1-8. Structures of two PG binding lectins. Each image represents a single monomer from its respective lectin. The EPN motif in loop 2 of RegIIIα (PDB ID: 4MTH) (152) where PG binds is highlighted in black. Peptidoglycan binding by PGLYRP3 (PDB ID: 1TWQ) (157) is also shown highlighted in black.

1.5.5 Peptidoglycan Recognition Proteins

A second family of lectins that recognizes bacterial PG are the peptidoglycan recognition proteins, or PGRPs. Humans encode four unique PRGPs, PGLYRP1 and PGLYRP2 are encoded on chromosome 19, while PGLYRP3 and PGLYRP4 are encoded next to each other on chromosome 1 (158,159). PRGPs were originally discovered in insects (160), and consequently have been found highly conserved from insect to mammals.

Each human PGRP has a unique expression localization, and collectively they reside in all areas of the body that interact with microorganisms. PGLYRP1 is largely localized to secretory granules of neutrophils and eosinophils, but to a lesser extent is also found at mucosal surfaces. Because of expression by immune cells, PGLYRP1 is found in human blood at a

concentration of roughly 20 ng/mL (161). PGLYRP2 is largely expressed in the liver where it is secreted and found predominantly in the blood. PGLYRP2 is expressed by the skin and other epithelial cells, including those of the GI tract, and is inducible by the presence of bacteria. PGLYRP3 and -4 are expressed by the skin and mucus membranes (158,162).

Structurally, human PGLYRP proteins are similar. They contain at least one PGRP domain, a roughly 165 amino acid domain structurally similar to bacteriophage type 2 amidases (lysozymes) (157). This domain is used to bind MurNAc-peptide within PG with nanomolar affinity. While the PG-binding site is highly conserved between proteins, the opposite side of human PGLYRPs displays significantly more structural diversity (**Figure 1-8**). This region of the proteins is often highly hydrophobic and contains a deep groove that is thought to bind other ligands derived from microbes, including LPS and TA (163). Interestingly, all human PGRPs contain a type 2 amidase domain, only PGLYRP2 is catalytically competent and can degrade PG. Specifically, PGLYRP2 hydrolyzes the amide bond connecting MurNAc and the peptide stem (164). Most of the human PGRPs are secreted, some form disulfide-linked homo- and heterodimers.

As human PGRPs bind to PG, binding is likely limited to bacteria. Interestingly, the three PGLYRPs that lack amidase activity were found to be directly bactericidal to both Gram-positive and Gram-negative bacteria (165,166). This results was surprising as the PG of Gram-negative bacteria is concealed by the outer membrane of the organism.

Recently, the mechanism by which PGLYRP proteins kill bacteria was revealed (163). PGLYRP-1, -3, -4, and a 3:4 heterodimer, bind to the PG, or the outer membrane for Gram-negative bacteria, and activate a protein-sensing two-component system. In the Gram-positive

bacterium *Bacillus subtilis*, the lectins bind at sites of daughter cell septation and activate the CsrR-CsrS system (163,167). Septation sites undergo PG remodeling and thinning and this likely allow access to the receptors. The CsrR-CsrS two-component system is used by bacteria to sense extracytoplasmic misfolded and aggregated proteins, and then initiates protein repair processes. However, membrane depolarization and production of reactive oxygen species [OH]• also result from CsrR-CsrS activation. If signaling is too intense or too prolonged, activation ultimately results in cell death. Using *E. coli*, an analogous two-component system, CpxA-CpxR, was identified as the PGLYRP ligand. Upon lectin binding to the outer membrane, a similar mechanism of cell death was observed. This mechanism of cell killing is unique in that it does not function through direct membrane depolarization, a mechanism that has been described for multiple antimicrobial proteins such as antimicrobial peptides (168) and the lectins RegIII α (152) and galectins (169).

Human PGLYRPs are suggested to contribute to multiple disease states based on their divergent expression localization. Human GWAS for predisposition to Crohn's disease and ulcerative colitis have identified susceptibility locus that contain PGLYRP2, -3, and -4 (170,171). These reports highlight the importance of host-microbe interactions, and the potential role for PG in the pathogenesis of these diseases. While large, genome wide-studies can identify disease risk loci across the genome, a more targeted approach may identify specific polymorphisms. Recently, all four PGLYRP genes were sequenced from a large consortium of irritable bowel disease (IBD) patients in an effort to identify specific gene polymorphisms (172). From this, multiple variants that may affect protein function were identified.

In agreement with the human genetic data are a series of mouse deletion experiments where four mouse lines were established, each devoid of one mouse *Pglyrp* gene. Interestingly, all four mouse strains were more susceptible to dextran sulfate sodium-induced colitis (173). *Pglyrp* deletion was also shown to alter the composition of the mouse gut flora and cause defects in intestinal barrier function. Compromise of barrier function was postulated to occur via increased interferon- γ production by natural killer (NK) cells recruited during epithelial barrier breaching, enhancing tissue damage. A more thorough understanding of how PGLYRPs manipulate microbial populations, and alter cellular immune responses, may provide opportunities for the discovery of novel points for therapeutic intervention (174).

1.5.6 Galectins

Galectins are a large family of lectins that in humans have been implicated in diverse biological processes ranging from development to innate immunity (175). Galectin proteins are similar in their amino acid sequences and share a conserved affinity for β -galactose. In mammals, 15 galectin subtypes have been identified, and named chronologically based on their discovery (176). Galectins can be categorized into three families based on sequence, and how their CRDs assemble. Human galectin-1, -2, -7, -10, -13, and -14, are classified as prototype galectins and are expressed as a single CRD per polypeptide chain. They can, and often do, assemble into noncovalent homodimers. Human galectin-3 is the only human galectin that falls within the chimera type. Chimera galectins contain an N-terminal oligomerization domain and a single C-terminal CRD. Galectin-3 can assemble into functional trimers and tetramers through its N-terminal domain (177). Galectins-4, -8, -9, and -12 all belong to a class known as the tandem-repeat type. These galectins are expressed as two independent CRDs located within the same

polypeptide chain (178). Galectin proteins are highly conserved evolutionarily and similar proteins are found across the animal kingdom down to fungi and primitive tunicate species (176). This suggests an important role in innate immunity as invertebrates rely entirely on innate immunity for resistance to infection.

Most galectins are ubiquitously expressed by many cells types, including cells from the innate and adaptive immune system. Despite the absence of a classical signal peptide, many galectins are found within the cytosol and extracellular (179). They are secreted via a noncanonical mechanism as soluble, aglycosylated proteins. With respect to this review, we will focus on the extracellular galectins.

Multiple galectins, including galectins-1, -3, and -8, are found at double digit ng/mL concentrations in human blood. Infact, serum galectin-3 levels are used clinically as a diagnostic of myocardial infarction because of association with inflammation and fibrotic pathways (180). Multiple galectins are also reported to be expressed and found extracellularly in the GI tract (181). Within the GI, galectin secretion results in binding to heavily glycosylated mucin proteins that contain N-acetyl-D-lactosamine (Gal- β (1-4)-GlcNAc or LacNAc). Galectin:mucin binding may assist in establishing the mucus barrier that segregates the intestinal microbiota and host cells, or it may function to slow galectin diffusion and increase local galectin concentrations within the mucus.

When the first structure of a galectin protein was solved, it revealed a β -sandwich or jellyroll-like domain structure that completely lacks α -helices (182). A similar structure is observed in the legumes plant-type (L-type) family of lectins that are often isolated from legumes plants. All known human galectins share this. Another hallmark of galectin proteins is

their calcium ion-independent affinity toward β -galactose or LacNAc containing glycans.

Independence from divalent cations for carbohydrate binding was quickly realized, and used to differentiate galectins from C-type lectins (183).

Initially, a reducing environment was proposed to be essential for galectin ligand binding, and thus galectins were referred to as S-type lectins to contrast C-type lectins (183).

Mechanistically, galectin oxidation sensitivity is the result of a cysteine residue proximal to the ligand binding site. Cysteine oxidation results in a loss of activity and/or protein oligomerization. Consequently, this characteristic has been found to be not general to all galectins, but it does remain a careful consideration for some galectins, such as galectin-1. Thus, oxidation sensitive galectins are a rheostat of the environment reducing potential, as it can drastically affect galectin structure, ligand binding, and biology.

Galectin ligand specificity is highly dependent on the concentration of protein used during analysis. At high concentrations, ligand binding is broad, and galectins interact with multiple epitopes including LacNAc, O- and N-linked glycans, Neu5Ac, ABO(H) blood group antigens, and sulfated glycans (178,184). However, as the protein concentration is reduced to $< 1 \mu\text{M}$, the ligand binding scope contracts and centers around LacNAc containing glycans. Through assaying ligand binding at multiple concentrations, the identification and ligand preferences of multiple potential subsites within galectins were identified and mapped (184).

Traditionally, galectins were considered “self” glycan binders and thought to regulate developmental processes, cancer, and immune activation/suppression (175,176,185). Galectins are postulated to perform these biological functions through initiating signaling events via receptor crosslinking, mediated by their multivalent assembly and through receptor

glycosylation. Using this mechanism and disparate expression, different galectins result in divergent biological outcomes. Galectin-1 is associated with an anti-inflammatory response, while galectin-3 is pro-inflammatory (186).

Only recently has a potential role for galectins in innate microbial immunity become appreciated. An early report of galectin binding to microbial antigens was binding of LPS by galectin-3 (187). Since then, binding to many microbial species including bacteria, viruses, fungi, protozoans and helminthes has been reported (188). Recently, multiple galectins were found to selectively bind ABO(H) blood group antigens when assayed under stringent conditions, less than 0.5 μ M protein (169). Galectins-3, -4, and -8 bound blood group B antigen, while galectin-4, and -8 also bound to blood group A antigen. The authors postulated that galectins may assist in neutralizing microbes that synthesize glycans structurally similar to host glycans. For example, bacteria that express blood group antigens may be able to evade the immune system because of deletion of self-reactive immune cells (189). Interestingly, galectins-4 and -8 were shown to have direct bactericidal activity when bound to Gram-positive or Gram-negative bacteria that express blood group B containing glycans, and α -galactose epitopes (169). While the mechanism remains unknown, drastic alteration of the microbial membrane and inhibition of microbial motility were observed. How substantial this observation is remains to be seen as natural galectin concentrations are low and bacteria expressing self-like mimicry glycans are not especially prevalent (190).

Later, galectins-3, -4, and -8 were assayed using a glycan microarray assembled from microbial glycans (191). This experiment revealed that at high concentrations, galectin binding was not limited to blood group B antigen, and galectin-4 bound to multiple microbial glycans

that were immobilized on the array (191). How galectin toxicity is selective against microbes, especially when assayed at concentrations up to 5 μM ($\sim 200 \mu\text{g/mL}$), remains an important question that will require future attention (178,192). This is especially important as galectins are known to interact with antigens present on mammalian cells, and certain galectins have been shown to induce loss of mammalian membrane asymmetry and exposure of phosphatidylserine (178). Because galectins perform myriad roles in immune activation and suppression, definitive studies using animals or human genetic data are challenging to interpret.

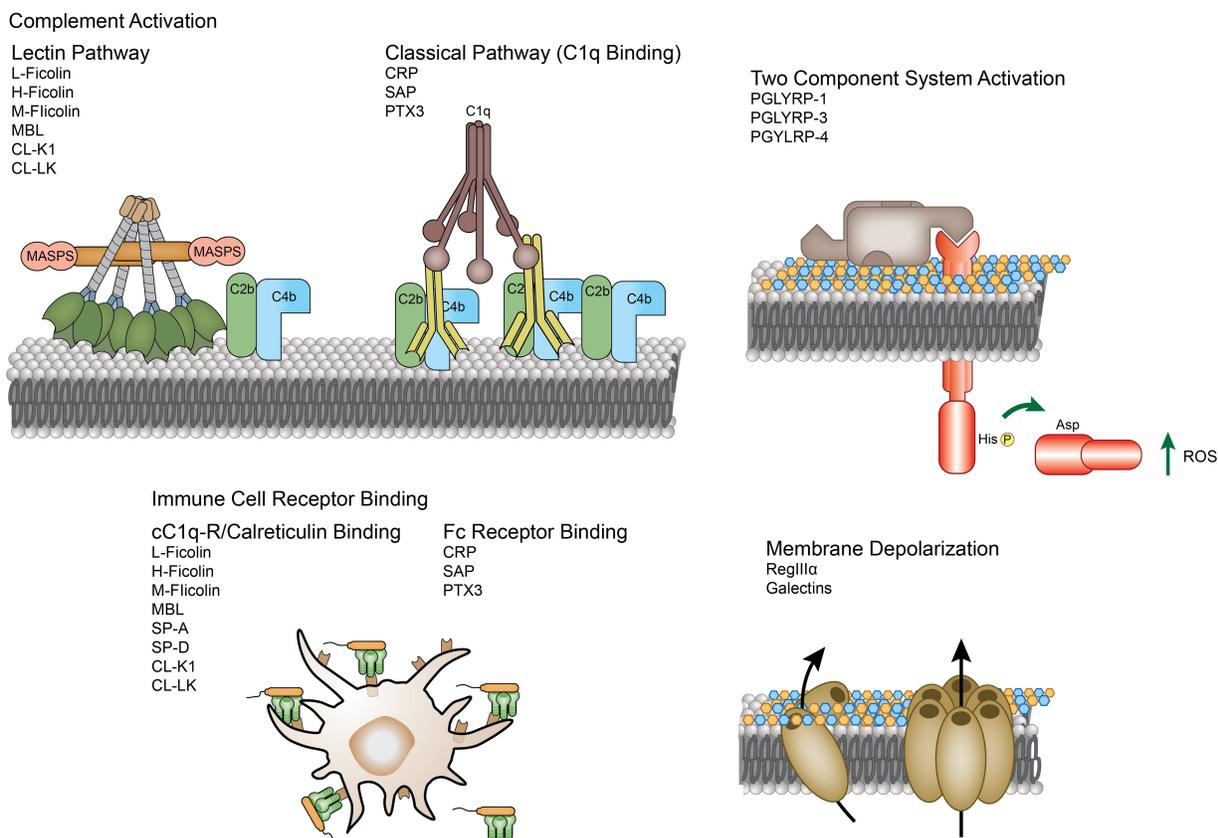


Figure 1-9. Mechanisms by which soluble human immune lectins exert antimicrobial activity. Below each short description are examples from proteins reviewed within this thesis that function through that mechanism.

1.5.7 Pentraxins

The pentraxin family of lectins is characterized by a cyclic multimeric structure and the presence of a roughly 200 amino acid pentraxin domain (193). Within the pentraxin domain lies an 8 amino acid pentraxin motif, HxCxS/TWxS, where x represents any amino acid. Within this review, I will limit the discussion to pentraxins postulated to be involved in immunity. Thus we will not discuss neuronal pentraxins-1, and -2, and the newly identified pentraxin, pentraxin-4 (194).

The first pentraxin identified was C-reactive protein (CRP) (195), followed by the identification of serum amyloid P component (SAP), which shares 51% sequence similarity with CRP. These two proteins are known collectively as the short pentraxins as their structure is composed almost exclusively of the pentraxin domain. CRP and SAP are both encoded on human chromosome 1. The other pentraxin reviewed here is pentraxin-3 (PTX3), which is encoded for on chromosome 3. Unlike the short pentraxins, PTX3 is referred to as a long pentraxin because it is almost 400 amino acids in length and contains an N-terminal coiled-coil region.

Pentraxin proteins assemble into higher order oligomeric structures to execute their biological functions. Short pentraxins form noncovalent pentamers, while PTX3 assembles into a disulfide linked octamer of two tetrameric assemblies (196). The pentraxin domain folds into a globular flattened jellyroll structure composed almost entirely of β -sheets, structurally similar to galectins and L-type plant lectins (197,198). However, unlike galectins which use a jellyroll structure and bind ligands independent of divalent cations, pentraxins require calcium ions (195). This is a distinguishing characteristic of pentraxins and hallmark of this lectin class. The presence of pentraxin proteins is conserved in evolution and can be found in diverse species ranging from arachnids and insects to humans. It is important to note that significant divergence in the short pentraxins is observed amongst mammals, while PTX3 sequence and function is more strictly conserved.

CRP was originally identified as a robustly induced, acute-phase immune protein localized in the blood. While oft referred to as human plasma proteins, pentraxin expression occurs in many other tissues and cell types. Human CRP is expressed predominately by liver

hepatocytes, and is secreted into blood where it is present at concentration $< 3 \mu\text{g/mL}$. However CRP levels are highly inducible and may increase up to 1,000 fold during infection, injury, or inflammation (199). Expression has also been reported at mucosal surfaces such as the lung, where concentrations can reach $42 \mu\text{g/mL}$ in human respiratory tract secretions, a level sufficient for antimicrobial effects (200). Similar to CRP, SAP is expressed by liver hepatocytes and is found in the serum at concentrations of $30\text{-}50 \mu\text{g/mL}$. Unlike CRP, SAP levels do not sharply increase during immune stimulation.

Human PTX3 is expressed by multiple immune cell types in response to proinflammatory cytokines or stimulation by immunogenic microbial cell wall components. Serum levels in healthy patients are often very low, averaging 1 ng/mL in healthy donors, but can increase substantially up to 250 ng/mL during sepsis (201). Additionally, PTX3 is produced by human lung epithelial cells and performs critical roles in lung immunity (202,203), and expression is significantly upregulated in the small intestine and colon of IBD patients (204).

One defining feature of pentraxins is calcium cation-dependent carbohydrate binding. CRP was originally identified as an acute phase protein highly upregulated upon infection with *S. pneumoniae* or challenge with purified Streptococcal C-polysaccharide (195,205), a teichoic acid heavily modified with phosphocholine (206,207). Phosphocholine is also present within the capsular polysaccharides of multiple *S. pneumoniae* serotypes (19). When the three dimensional structure of CRP was solved, the mechanism of phosphocholine recognition was revealed (198,208). Each CRP monomer binds two calcium ions, the phosphate group of phosphocholine coordinates directly to those calcium ions and the trimethylammonium fits snugly into a hydrophobic and aromatic box. CRP has also been reported to bind phosphoethanolamine,

traditionally a ligand of SAP, and to multiple microbial species including *S. pneumoniae* and fungi. CRP binding to human immune proteins such as Fc γ receptors (Fc γ R) and C1q has also been reported (193,209). One of the mechanisms pentraxins are thought to function in immunity is through C1q recruitment and activation of the classical complement pathway (210). Lastly, CRP has been shown to be important for binding to self-antigens and assisting in the clearance of apoptotic or dying mammalian cells (211).

The overall topology and calcium binding sites of SAP are similar to CRP (197). Like CRP, SAP coordinates two calcium ions that interact with the phosphate head group of phosphoethanolamine (212). Additionally, SAP binds to a 4,6-pyruvate acetal of β -D-galactopyranose through direct calcium ion coordination (213,214), the lipid A moiety of LPS (215), multiple Gram-positive and Gram-negative bacteria (216), and a similar collection of immune proteins (199). The mechanism by which pentraxins facilitate phagocytosis of microbes through Fc γ R engagement was recently revealed using an SAP:: Fc γ R protein co-crystal structure (217). Pentraxins bind to the IgG binding site on Fc γ R using the face opposite of pentraxins ligand binding.

Compared to CRP and SAP, less is known about the monovalent ligands of PTX3, however PTX3 has been shown to interact with outer membrane protein A from *K. pneumoniae*, multiple bacterial, fungal, viral species, and immune receptors that include C1q and Fc γ R (218). PTX3 glycosylation maybe be responsible for some of these interactions as it has been shown that influenza virus type A binds PTX3 through viral hemagglutinin binding to sialylated PTX3, not through PTX3 binding to viral glycans (219).

The role of pentraxins in immunity is still an active area of research. Pentraxin mediated C1q binding and complement activation can mark microbes for other innate immune effector cells, but interestingly CRP has been shown to poorly generate the membrane attack complex (211). Additionally, microbe binding by pentraxins can have disparate immunological outcomes as has been demonstrated using a combination of PTX3 deletion mice, microbiology, and chemical biology (216). In some cases, SAP binding inhibited infection, while for others binding enhanced infection, and in still others, SAP appears to alter infection in the absence of microbe binding. This work highlights the complexities of pentraxins in the innate immune system and the need for caution when trying to define function.

One of the best examples is the anti-fungal role of PTX3 in mice. Using a genetic deletion of PTX3, the authors demonstrate that PTX3 protects mice from pulmonary invasive aspergillosis upon challenge from *A. fumigatus* (203). To accomplish this, PTX3 enhances ligand recognition and phagocytosis by macrophages and dendritic cells (DCs), likely through Fc γ R binding. In this example, PTX3 mediated immunity appeared to function independent of C1q binding and complement activation. This mouse experiment is possible because mouse and human PTX3 are highly conserved, unlike the short pentraxins. Interestingly, soluble CRP and PTX3 administration have been shown to partially protect against pneumococcal and *A. fumigatus* infection, suggesting the potential utility of soluble pentraxins as therapeutics (220).

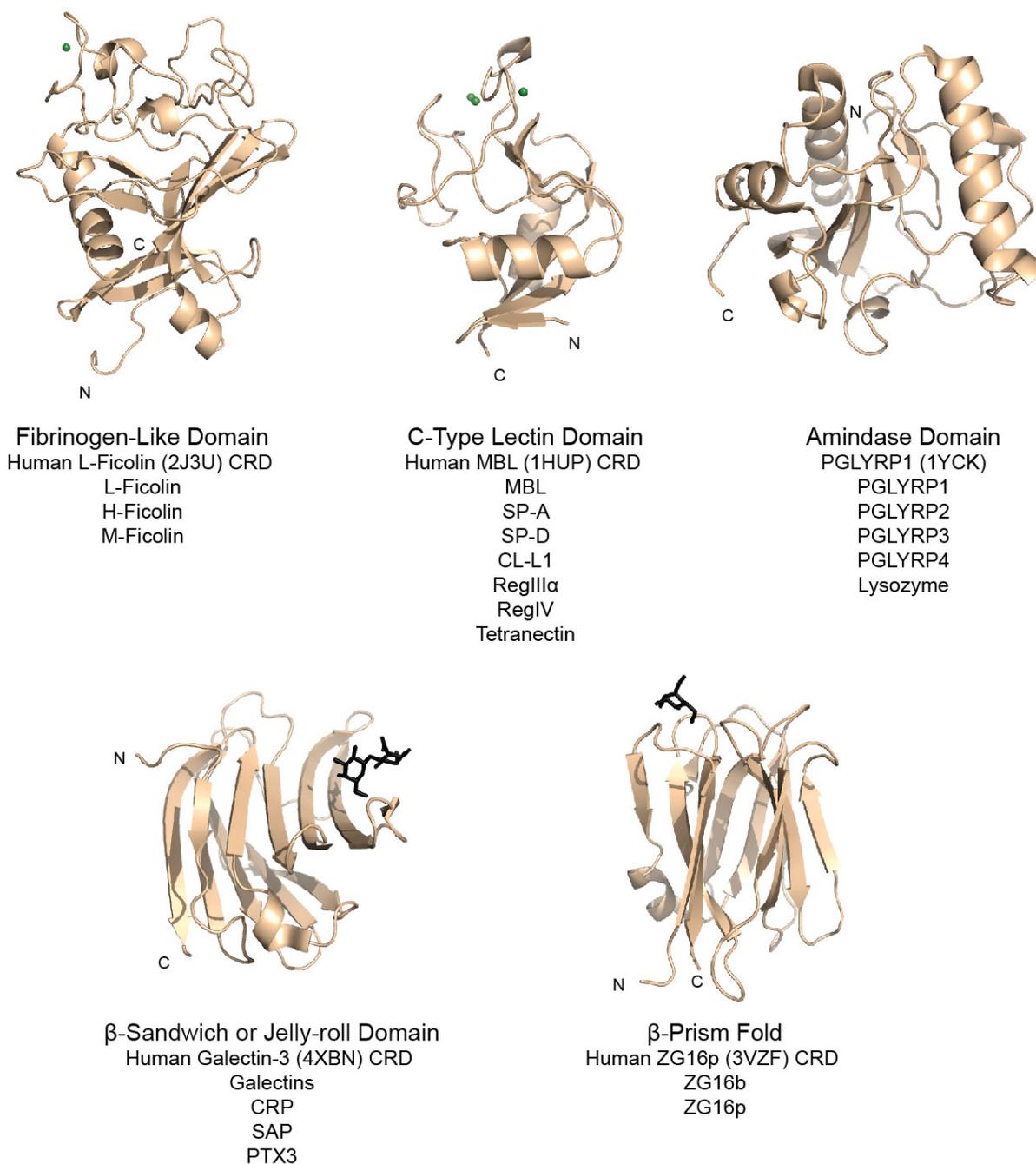


Figure 1-10. Structural classes of human soluble immune lectins. Each structure represents a monomeric CRD used by human soluble immune lectins. Calcium ions involved in ligand binding are shown as green spheres. The N- and C-termini are highlighted by an N and C, respectively. The ligands are included as black sticks in some images for clarity. Human L-

Ficolin (PDB ID: 2J3U) (71), MBL (PDB ID: 1HUP) (80), PGLYRP1 (PDB ID: 1YCK) (221), galectin-3 (PDB ID: 4XBN) (222), and ZG16p (PDB ID: 3VZF) (223) are included.

1.5.8 Lysozyme

Albeit not the focus of this review, human lysozyme is an innate immune protein that binds to and hydrolyzes microbial PG (142). Human lysozyme, or lysozyme C is encoded by the LYZ gene on the twelfth chromosome (224). Similar to other lysozymes, human lysozyme is a small, 130 amino acid, soluble secreted protein. Lysozyme is expressed by a variety of human tissue and cell types including intestinal paneth cells, lactating mammary tissue, salivary glands, and multiple immune cell types (224-226).

The most widely appreciated antimicrobial quality of lysozyme is its ability to degrade PG via hydrolysis of the $\beta(1-4)$ linkage between MurNAc and GlcNAc. While muramidase activity is likely important for antimicrobial function, it is not essential. Inhibition of muramidase activity via reduction of essential disulfide bonds, or addition of the competitive inhibitor chitotriose, still results in robust cell killing (227). Another physical characteristic of lysozyme that may be important in its antimicrobial function is its cationic nature, $pI > 9$. This is similar to other antimicrobial immune proteins such as lactoferrin, which has a pI of roughly 8.5.

1.5.9 Human C-type Lectin Domain Family (CLEC) Proteins

Within the human genome there are many small C-type lectin domain containing proteins whose functions are not yet well understood. Within this group lies multiple C-type Lectin Domain Families, including 1, 2, 3, 4, and 11; CLEC1, CLEC2, CLEC3, CLEC4, and CLEC11, respectively. These families collectively contain many transmembrane proteins with well documented lectin and immune function, examples include dendritic cell-specific ICAM-3-grabbing non-integrin, DC-SIGN, and its related protein, DC-SIGNR, CLEC4L and CLEC4M,

respectively (228,229). While many human CLEC genes encode for predicted membrane bound proteins, several are predicted to be soluble. These include cartilage-derived C-type lectin or CLEC3A, and tetranectin or CLEC3B.

Tetranectin was originally identified in human plasma as a plasminogen binding protein where it is found at concentrations of roughly 15 $\mu\text{g/mL}$ (230). Consequently, tetranectin has been localized in many tissues and cell types. When the structure of tetranectin was solved, its resemblance to other C-type lectins was apparent (231). Tetranectin contains two calcium ions that are bound in positions analogous to the calcium ions bound by bacterial mannose binding protein. The structure also revealed a trimeric assembly that was mediated by an N-terminal α -helical coiled coil region. In addition to plasminogen binding, tetranectin has been reported to bind sulfated polysaccharides in a calcium ion-dependent fashion (232). Outside of this, little is known about the potential carbohydrate binding activity of this protein. Revisiting tetranectin, employing advances in glycobiology research such as glycan microarrays may prove useful for identifying human or microbial carbohydrate ligands, and may elude to this protein's biological function.

1.5.10 ZG16p

Humans express a small group of soluble lectins structurally related to the Jacalin-related β -prism fold lectin family (233) termed the human zymogen granule proteins, ZG16p and ZG16b. Both human ZG16 proteins are encoded for on chromosome 16 by the ZG16 and ZG16B gene, respectively. Comparatively, more is known about the lectin properties of ZG16p, and will thus be the focus of this review, but much work is yet to be done with these two proteins.

ZG16p is a roughly 150 amino acid protein, of which 130 amino acids make up the Jacalin-like lectin domain. Human ZG16p was initially identified as a glycosaminoglycan binding lectin important for protein sorting to the zymogen granule membrane in pancreatic cells (234,235). Pancreatic acinar cell zymogen granules contain digestive enzymes and are secreted into the duodenum to aid in digestion and nutrient uptake. In addition to being expressed in the pancreas, ZG16p is expressed in the liver and localized to the serum, and is expressed in the small and large intestine (236). ZG16p expression in the GI is localized to goblet cells. The concentrated expression in the digestive system suggests a likely function there, although a clear biological role remains unknown.

Similar to other Jacalin-related lectins, ZG16p binds carbohydrate ligands in a calcium ion-dependent manner. It is suggested to function as a monomer based on the apparent absence of intermolecular disulfide bonds and the elution from a size exclusion column (237), though that would contradict other β -prism fold lectins. The structure of ZG16p was recently solved using protein X-ray crystallography and revealed the first mammalian lectin containing a Jacalin-related β -prism fold (237).

Glycan microarray screening has been employed to assay the ligand binding properties of ZG16p. These experiments revealed a preference for short α -Man oligosaccharides and *O*-linked mannose residues (223). Surprisingly, ZG16p did not interact with high mannose *N*-glycans. Monovalent affinity measurements were made using NMR titration experiments and suggest low affinity binding to α -mannose and Glc- β (1-3)-Glc (2-20 mM). This is in contrast to the micromolar affinities reported with a multivalent display, suggesting quaternary structure.

Ligand co-crystal structures revealed that ZG16p binds carbohydrate ligands at a site similar to other β -prism fold lectins and does so through interaction with ligand features distal from the anomeric hydroxyl (**Figure 1-10**) (223). In the same study, the glycosaminoglycan binding site was identified and shown to function independent of mannose binding. Based on these results, binding to microbial pathogens and microbially derived glycans was assayed. A microbial glycan microarray and NMR experiments were used to identify mycobacterial PI mannosides as ligands of ZG16p (238). Lastly, binding to multiple fungal pathogens has been demonstrated using flow cytometry, likely occurring through yeast mannan recognition (236). Further in vitro experiments, or experiments in model organisms, should help elucidate the biological function of this lectin, which expression data suggests could be in the mammalian GI.

1.5.11 Mindin

The mindin/F-spondin family of proteins is a small collection of extracellular matrix proteins where one member, mindin, has been implicated to function as a lectin. Mindin, also referred to as spondin-2, is a roughly 300 amino acid protein encoded for on human chromosome 4. Mindin/F-spondin proteins were initially identified in zebrafish as secreted proteins that accumulate in the basal lamina during embryonic development (239). Mindin contains two domains, a 200 amino acid N-terminal spondin domain and a C-terminal thrombospondin-type 1 repeat; this architecture is shared by other members of the mindin-F-spondin family.

Mindin contains eight conserved cysteine residues that are involved in intra- and inter-molecular disulfide bond formation. Immunoblotting suggests that under nonreducing conditions mindin exists exclusively as a disulfide-linked dimer at low concentrations, but will oligomerize into higher order structures at increasing concentrations. The structure of the human mindin

spondin domain was recently solved and surprisingly bared structural resemblance to a C2 domain (240). C2 domains are used by proteins for localization to the cell membrane as they can bind membrane phospholipids in an calcium ion-dependent, and independent manner (241). The mindin structure revealed a single protein-bound calcium ion coordinated to eight oxygen atoms made up by amino acids and protein bound water molecules (240). Much of what is known about human mindin comes from studies of mouse mindin, the two share 85% sequence identity. Mouse mindin is highly expressed in the lung, spleen, heart, lymph nodes, ilium, and colon (242,243). Mindin is also expressed by multiple immune cell types including macrophages and mast cells, and has been measured in mouse and human sera at 10-30 ng/mL (244).

Mindin has been shown to bind glycosylated ligands in both a calcium ion-dependent, and -independent manner (242,245). Recombinant mindin was shown to bind Gram-positive group B streptococcus (GBS), *Salmonella typhimurium*, *H. influenzae*, and influenza A virus (242,245). *S. typhimurium* agglutination was dependent on calcium ions, and inhibited by soluble glucose, while binding to influenza was not.

Mice devoid of mindin were used to reveal disparate biological functions of the protein. Interestingly, mindin null mice are resistant to LPS-induced septic shock. Mechanistically, this is postulated to be the result of mindin deficient macrophages exhibiting attenuated activation and phagocytosis in response to multiple microbe derived antigens such as LPS, PGN, LTA, and mannan. Mindin deficient mice were also shown to be impaired in their ability to clear intratracheally inoculated lung pathogens such as GBS or *H. influenzae* (242), and had higher lung viral titers after intranasal influenza virus infection (245). One common result from these mindin deficient mouse studies is attenuated immune cell activation. One hypothesis for the

attenuated cellular response is the ability of mindin to interact with multiple integrins through its spondin domain (246). Further defining how mindin interacts with carbohydrate ligands, and the functional outcome of immune cell binding or integrin engagement, should illuminate the function of mindin in LPS sensing.

1.5.12 Soluble CD14 and Lipopolysaccharide Binding Protein

While not traditionally considered lectins, human lipopolysaccharide binding protein (LBP) and CD14 are essential for binding to, and recognition of LPS. CD14, which functions on immune cells as a receptor for LPS and LBP (247), is attached to the membrane via a GPI anchor, but also can be secreted as soluble CD14 (248). Human CD14 is encoded on chromosome 5 and results in expression of a roughly 350 amino acid protein. CD14 was initially identified as a differentiation marker of monocytes, but is also found on the surface of macrophages. Soluble CD14 is found in conditioned culture media from human intestinal epithelial cells (249), is found at a concentration of 2-3 $\mu\text{g}/\text{mL}$ in human plasma, and 10-20 $\mu\text{g}/\text{mL}$ in breast milk (250). LBP consists of 450 amino acids and is primarily expressed in the liver and is secreted into human plasma to concentrations of 10-20 $\mu\text{g}/\text{mL}$ for recognition of the lipid A portion of LPS(250,251). Together, LBP and soluble CD14 collaborate with other receptors such as TLR4 to effectively recognize LPS and facilitate an immune response (252). While these proteins are not going to be discussed further here (253), it is important to remember them when considering the recognition of glycosylated ligands by soluble human lectins.

1.6 Conclusions

Within this review I have highlighted soluble human immune lectins involved in the recognition of microbial glycopolymers. As research in host:microbe interactions continues,

recognition of cell surface glycans will likely be centrally important. Specifically, I hypothesize that host lectins will be critical within the mammalian GI tract, where lectins may be able to modulate the resident microbial community. A major recent advance has been the generation of a microbial glycan microarray (191). This resource has been used to reveal the *in vitro* ligand specificity of multiple lectins, including work described within this thesis. Results from using this array already spark ones imagination concerning ligands to add to a next generation array. For example, surface glycans from symbiotic bacteria, specifically from the lung and gut, would be a useful tool for probing mammalian mucosal immune lectins. The individual spots afforded with glycan microarray printing make it especially attractive for assaying binding to bacterial strains, something that is currently difficult when working with complex environmental communities (254).

In addition to glycan microarrays, other tools may become equally important for assaying lectin interactions with microbial communities. The use of fluorescence activated cell sorting (FACS) and DNA sequencing can be used to simultaneously quantify lectin binding and genetically characterize bound microbes. A similar strategy has recently been applied to characterize IgA binding to bacteria isolated from human fecal samples (255,256). This analysis is made possible by the dramatic increase in sequence capabilities and a reduction in cost. While 16S ribosomal ribonucleic acid (rRNA) sequencing for phylogeny assignment is now routine (257), extracting *de novo* information about cell surface glycosylation based on genetic sequencing is still a challenge. Advances in analytical chemistry may also be integrated into such a workflow. Specifically, as mass spectrometry sensitivity, and computational power continues to accelerate, elucidation of microbial glycan structures may become customary. What I hope

this review highlights is the ability to integrate new technology into lectin:microbe research, while concurrently highlighting lectins as useful tools in biomedical research.

Moving forward, there is significant work to be done in understanding how soluble lectins mechanistically function in immunity. Centrally important to answering these questions will be an interdisciplinary research approach that employs protein biochemistry, chemical biology, microbiology, immunology, and mouse models. As the Kiessling Group continues work in this area, I cannot emphasize enough the importance of expression localization and physiological lectin concentrations. Lectins are fickle creatures that require careful attention. Lectins are often glycosylated, and thus themselves a target of lectins when assayed in vitro or in combination. Aggregated or misfolded lectins often retain carbohydrate binding activity, but have an altered ligand specificity and affinity. I hope this review provides a foundational background on lectins and glycobiology, and can help guide other researchers into this fascinating and essential field.

1.7 Acknowledgements

I would like to acknowledge the entire Kiessling lab group for their enthusiastic support of our labs research into human soluble immune lectins. I acknowledge Christine Isabella and Lucas Zarling for feedback in preparing this section.

1.8 References

1. Tytgat, H. L., and Lebeer, S. (2014) The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol. Mol. Biol. Rev.* **78**, 372-417
2. Cress, B. F., Englaender, J. A., He, W., Kasper, D., Linhardt, R. J., and Koffas, M. A. (2014) Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol. Rev.* **38**, 660-697
3. Vollmer, W., Blanot, D., and de Pedro, M. A. (2008) Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **32**, 149-167
4. Glauner, B. (1988) Separation and quantification of mucopeptides with high-performance liquid chromatography. *Anal. Biochem.* **172**, 451-464
5. Vollmer, W. (2008) Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol. Rev.* **32**, 287-306
6. Braun, V., and Rehn, K. (1969) Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. The specific effect of trypsin on the membrane structure. *Eur. J. Biochem.* **10**, 426-438
7. Raetz, C. R., and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635-700
8. Schnaitman, C. A., and Klena, J. D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**, 655-682
9. Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* **282**, 2085-2088
10. Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A., Jr. (1997) A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394-397
11. Meredith, T. C., Aggarwal, P., Mamat, U., Lindner, B., and Woodard, R. W. (2006) Redefining the requisite lipopolysaccharide structure in *Escherichia coli*. *ACS Chem. Biol.* **1**, 33-42
12. Preston, A., Mandrell, R. E., Gibson, B. W., and Apicella, M. A. (1996) The lipooligosaccharides of pathogenic gram-negative bacteria. *Crit. Rev. Microbiol.* **22**, 139-180
13. Brandtzaeg, P., Bjerre, A., Ovstebo, R., Brusletto, B., Joo, G. B., and Kierulf, P. (2001) *Neisseria meningitidis* lipopolysaccharides in human pathology. *J. Endotoxin. Res.* **7**, 401-420
14. Stenutz, R., Weintraub, A., and Widmalm, G. (2006) The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol. Rev.* **30**, 382-403
15. Robbins, P. W., Bray, D., Dankert, B. M., and Wright, A. (1967) Direction of chain growth in polysaccharide synthesis. *Science* **158**, 1536-1542
16. Woodward, R., Yi, W., Li, L., Zhao, G., Eguchi, H., Sridhar, P. R., Guo, H., Song, J. K., Motari, E., Cai, L., Kelleher, P., Liu, X., Han, W., Zhang, W., Ding, Y., Li, M., and Wang, P. G. (2010) In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. *Nat. Chem. Biol.* **6**, 418-423

17. Hagelueken, G., Clarke, B. R., Huang, H., Tuukkanen, A., Danciu, I., Svergun, D. I., Hussain, R., Liu, H., Whitfield, C., and Naismith, J. H. (2015) A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. *Nat. Struct. Mol. Biol.* **22**, 50-56
18. Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J. R., Elvin, L., Ensor, K. M., Hackell, J., Siber, G., Malinoski, F., Madore, D., Chang, I., Kohberger, R., Watson, W., Austrian, R., and Edwards, K. (2000) Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr. Infect. Dis. J.* **19**, 187-195
19. Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kalltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**, e31
20. Coyne, M. J., Reinap, B., Lee, M. M., and Comstock, L. E. (2005) Human symbionts use a host-like pathway for surface fucosylation. *Science* **307**, 1778-1781
21. Krinos, C. M., Coyne, M. J., Weinacht, K. G., Tzianabos, A. O., Kasper, D. L., and Comstock, L. E. (2001) Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555-558
22. Patrick, S., Blakely, G. W., Houston, S., Moore, J., Abratt, V. R., Bertalan, M., Cerdeno-Tarraga, A. M., Quail, M. A., Corton, N., Corton, C., Bignell, A., Barron, A., Clark, L., Bentley, S. D., and Parkhill, J. (2010) Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology* **156**, 3255-3269
23. Orskov, I., Orskov, F., Jann, B., and Jann, K. (1977) Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol. Rev.* **41**, 667-710
24. Yother, J. (2011) Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. *Annu. Rev. Microbiol.* **65**, 563-581
25. Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **75**, 39-68
26. Aanensen, D. M., Mavroidi, A., Bentley, S. D., Reeves, P. R., and Spratt, B. G. (2007) Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J. Bacteriol.* **189**, 7856-7876
27. Xayarath, B., and Yother, J. (2007) Mutations blocking side chain assembly, polymerization, or transport of a Wzy-dependent *Streptococcus pneumoniae* capsule are lethal in the absence of suppressor mutations and can affect polymer transfer to the cell wall. *J. Bacteriol.* **189**, 3369-3381
28. Willis, L. M., Stupak, J., Richards, M. R., Lowary, T. L., Li, J., and Whitfield, C. (2013) Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7868-7873
29. Cartee, R. T., Forsee, W. T., and Yother, J. (2005) Initiation and synthesis of the *Streptococcus pneumoniae* type 3 capsule on a phosphatidylglycerol membrane anchor. *J. Bacteriol.* **187**, 4470-4479

30. Weidenmaier, C., and Peschel, A. (2008) Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nat. Rev. Microbiol.* **6**, 276-287
31. Armstrong, J. J., Baddiley, J., Buchanan, J. G., Davision, A. L., Kelemen, M. V., and Neuhaus, F. C. (1959) Composition of teichoic acids from a number of bacterial walls. *Nature* **184**, 247-248
32. Fischer, W. (1988) Physiology of lipoteichoic acids in bacteria. *Adv. Microb. Physiol.* **29**, 233-302
33. Grundling, A., and Schneewind, O. (2007) Synthesis of glycerol phosphate lipoteichoic acid in *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8478-8483
34. Dunne, D. W., Resnick, D., Greenberg, J., Krieger, M., and Joiner, K. A. (1994) The type I macrophage scavenger receptor binds to gram-positive bacteria and recognizes lipoteichoic acid. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1863-1867
35. Travassos, L. H., Girardin, S. E., Philpott, D. J., Blanot, D., Nahori, M. A., Werts, C., and Boneca, I. G. (2004) Toll-like receptor 2-dependent bacterial sensing does not occur via peptidoglycan recognition. *EMBO Rep.* **5**, 1000-1006
36. Cobb, B. A., Wang, Q., Tzianabos, A. O., and Kasper, D. L. (2004) Polysaccharide processing and presentation by the MHCII pathway. *Cell* **117**, 677-687
37. Tzianabos, A. O., Wang, J. Y., and Lee, J. C. (2001) Structural rationale for the modulation of abscess formation by *Staphylococcus aureus* capsular polysaccharides. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9365-9370
38. Swoboda, J. G., Campbell, J., Meredith, T. C., and Walker, S. (2010) Wall teichoic acid function, biosynthesis, and inhibition. *ChemBiochem : a European journal of chemical biology* **11**, 35-45
39. Fischer, W. (1994) Lipoteichoic acid and lipids in the membrane of *Staphylococcus aureus*. *Med. Microbiol. Immunol.* **183**, 61-76
40. Neuhaus, F. C., and Baddiley, J. (2003) A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **67**, 686-723
41. Brennan, P. J. (2003) Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis* **83**, 91-97
42. Mennink-Kersten, M. A., Ruegebrink, D., Klont, R. R., Warris, A., Gavini, F., Op den Camp, H. J., and Verweij, P. E. (2005) Bifidobacterial lipoglycan as a new cause for false-positive platelia *Aspergillus* enzyme-linked immunosorbent assay reactivity. *J. Clin. Microbiol.* **43**, 3925-3931
43. Daffe, M., Brennan, P. J., and McNeil, M. (1990) Predominant structural features of the cell wall arabinogalactan of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by ¹H and ¹³C NMR analyses. *J. Biol. Chem.* **265**, 6734-6743
44. Chatterjee, D., Lowell, K., Rivoire, B., McNeil, M. R., and Brennan, P. J. (1992) Lipoarabinomannan of *Mycobacterium tuberculosis*. Capping with mannosyl residues in some strains. *J. Biol. Chem.* **267**, 6234-6239
45. Birch, H. L., Alderwick, L. J., Appelmelk, B. J., Maaskant, J., Bhatt, A., Singh, A., Nigou, J., Eggeling, L., Geurtsen, J., and Besra, G. S. (2010) A truncated lipoglycan from

- mycobacteria with altered immunological properties. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2634-2639
46. Sara, M., and Sleytr, U. B. (2000) S-Layer proteins. *J. Bacteriol.* **182**, 859-868
 47. Messner, P., Steiner, K., Zarschler, K., and Schaffer, C. (2008) S-layer nanoglycobiology of bacteria. *Carbohydr. Res.* **343**, 1934-1951
 48. Bowman, S. M., and Free, S. J. (2006) The structure and synthesis of the fungal cell wall. *BioEssays : news and reviews in molecular, cellular and developmental biology* **28**, 799-808
 49. Netea, M. G., Brown, G. D., Kullberg, B. J., and Gow, N. A. (2008) An integrated model of the recognition of *Candida albicans* by the innate immune system. *Nat. Rev. Microbiol.* **6**, 67-78
 50. Lipke, P. N., and Ovalle, R. (1998) Cell wall architecture in yeast: new structure and new challenges. *J. Bacteriol.* **180**, 3735-3740
 51. Manners, D. J., Masson, A. J., and Patterson, J. C. (1973) The structure of a beta-(1 leads to 3)-D-glucan from yeast cell walls. *Biochem. J.* **135**, 19-30
 52. Bernard, M., and Latge, J. P. (2001) *Aspergillus fumigatus* cell wall: composition and biosynthesis. *Med. Mycol.* **39 Suppl 1**, 9-17
 53. Doering, T. L. (2009) How sweet it is! Cell wall biogenesis and polysaccharide capsule formation in *Cryptococcus neoformans*. *Annu. Rev. Microbiol.* **63**, 223-247
 54. Heiss, C., Skowrya, M. L., Liu, H., Klutts, J. S., Wang, Z., Williams, M., Srikanta, D., Beverley, S. M., Azadi, P., and Doering, T. L. (2013) Unusual galactofuranose modification of a capsule polysaccharide in the pathogenic yeast *Cryptococcus neoformans*. *J. Biol. Chem.* **288**, 10994-11003
 55. Solis, D., Bovin, N. V., Davis, A. P., Jimenez-Barbero, J., Romero, A., Roy, R., Smetana, K., and Gabius, H. J. (2015) A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *BBA-Gen. Subjects* **1850**, 186-235
 56. Matsushita, M., Endo, Y., Taira, S., Sato, Y., Fujita, T., Ichikawa, N., Nakata, M., and Mizuochi, T. (1996) A novel human serum lectin with collagen- and fibrinogen-like domains that functions as an opsonin. *J. Biol. Chem.* **271**, 2448-2454
 57. Kairies, N., Beisel, H. G., Fuentes-Prior, P., Tsuda, R., Muta, T., Iwanaga, S., Bode, W., Huber, R., and Kawabata, S. (2001) The 2.0-Å crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and the blood coagulation systems. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13519-13524
 58. Holmskov, U., Malhotra, R., Sim, R. B., and Jensenius, J. C. (1994) Collectins: collagenous C-type lectins of the innate immune defense system. *Immunol. Today* **15**, 67-74
 59. Ohashi, T., and Erickson, H. P. (2004) The disulfide bonding pattern in ficolin multimers. *J. Biol. Chem.* **279**, 6534-6539
 60. Ma, Y. G., Cho, M. Y., Zhao, M., Park, J. W., Matsushita, M., Fujita, T., and Lee, B. L. (2004) Human mannose-binding lectin and L-ficolin function as specific pattern recognition proteins in the lectin activation pathway of complement. *J. Biol. Chem.* **279**, 25307-25312
 61. Degn, S. E., Kjaer, T. R., Kidmose, R. T., Jensen, L., Hansen, A. G., Tekin, M., Jensenius, J. C., Andersen, G. R., and Thiel, S. (2014) Complement activation by ligand-

- driven juxtaposition of discrete pattern recognition complexes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13445-13450
62. Krarup, A., Sorensen, U. B., Matsushita, M., Jensenius, J. C., and Thiel, S. (2005) Effect of capsulation of opportunistic pathogenic bacteria on binding of the pattern recognition molecules mannan-binding lectin, L-ficolin, and H-ficolin. *Infect. Immun.* **73**, 1052-1060
 63. Honore, C., Rorvig, S., Munthe-Fog, L., Hummelshoj, T., Madsen, H. O., Borregaard, N., and Garred, P. (2008) The innate pattern recognition molecule Ficolin-1 is secreted by monocytes/macrophages and is circulating in human plasma. *Mol. Immunol.* **45**, 2782-2789
 64. Teh, C., Le, Y., Lee, S. H., and Lu, J. (2000) M-ficolin is expressed on monocytes and is a lectin binding to N-acetyl-D-glucosamine and mediates monocyte adhesion and phagocytosis of *Escherichia coli*. *Immunology* **101**, 225-232
 65. Liu, Y., Endo, Y., Iwaki, D., Nakata, M., Matsushita, M., Wada, I., Inoue, K., Munakata, M., and Fujita, T. (2005) Human M-ficolin is a secretory protein that activates the lectin complement pathway. *J. Immunol.* **175**, 3150-3156
 66. Gout, E., Garlatti, V., Smith, D. F., Lacroix, M., Dumestre-Perard, C., Lunardi, T., Martin, L., Cesbron, J. Y., Arlaud, G. J., Gaboriaud, C., and Thielens, N. M. (2010) Carbohydrate recognition properties of human ficolins: glycan array screening reveals the sialic acid binding specificity of M-ficolin. *J. Biol. Chem.* **285**, 6612-6622
 67. Sugimoto, R., Yae, Y., Akaiwa, M., Kitajima, S., Shibata, Y., Sato, H., Hirata, J., Okochi, K., Izuhara, K., and Hamasaki, N. (1998) Cloning and characterization of the Hakata antigen, a member of the ficolin/opsonin p35 lectin family. *J. Biol. Chem.* **273**, 20721-20727
 68. Endo, Y., Matsushita, M., and Fujita, T. (2015) New insights into the role of ficolins in the lectin pathway of innate immunity. *Int. Rev. Cell. Mol. Biol.* **316**, 49-110
 69. Kjaer, T. R., Hansen, A. G., Sorensen, U. B., Holm, A. T., Sorensen, G. L., Jensenius, J. C., and Thiel, S. (2013) M-ficolin binds selectively to the capsular polysaccharides of *Streptococcus pneumoniae* serotypes 19B and 19C and of a *Streptococcus mitis* strain. *Infect. Immun.* **81**, 452-459
 70. Tanio, M., Kondo, S., Sugio, S., and Kohno, T. (2007) Trivalent recognition unit of innate immunity system: crystal structure of trimeric human M-ficolin fibrinogen-like domain. *J. Biol. Chem.* **282**, 3889-3895
 71. Garlatti, V., Belloy, N., Martin, L., Lacroix, M., Matsushita, M., Endo, Y., Fujita, T., Fontecilla-Camps, J. C., Arlaud, G. J., Thielens, N. M., and Gaboriaud, C. (2007) Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J.* **26**, 623-633
 72. Garlatti, V., Martin, L., Gout, E., Reiser, J. B., Fujita, T., Arlaud, G. J., Thielens, N. M., and Gaboriaud, C. (2007) Structural basis for innate immune sensing by M-ficolin and its control by a pH-dependent conformational switch. *J. Biol. Chem.* **282**, 35814-35820
 73. Kuraya, M., Ming, Z., Liu, X., Matsushita, M., and Fujita, T. (2005) Specific binding of L-ficolin and H-ficolin to apoptotic cells leads to complement activation. *Immunobiology* **209**, 689-697

74. Jensen, M. L., Honore, C., Hummelshoj, T., Hansen, B. E., Madsen, H. O., and Garred, P. (2007) Ficolin-2 recognizes DNA and participates in the clearance of dying host cells. *Mol. Immunol.* **44**, 856-865
75. Brinkmann, C. R., Jensen, L., Dagnaes-Hansen, F., Holm, I. E., Endo, Y., Fujita, T., Thiel, S., Jensenius, J. C., and Degn, S. E. (2013) Mitochondria and the lectin pathway of complement. *J. Biol. Chem.* **288**, 8016-8027
76. Lacroix, M., Dumestre-Perard, C., Schoehn, G., Houen, G., Cesbron, J. Y., Arlaud, G. J., and Thielens, N. M. (2009) Residue Lys57 in the collagen-like region of human L-ficolin and its counterpart Lys47 in H-ficolin play a key role in the interaction with the mannan-binding lectin-associated serine proteases and the collectin receptor calreticulin. *J. Immunol.* **182**, 456-465
77. Munthe-Fog, L., Hummelshoj, T., Honore, C., Madsen, H. O., Permin, H., and Garred, P. (2009) Immunodeficiency associated with FCN3 mutation and ficolin-3 deficiency. *N. Engl. J. Med.* **360**, 2637-2644
78. Hummelshoj, T., Munthe-Fog, L., Madsen, H. O., and Garred, P. (2008) Functional SNPs in the human ficolin (FCN) genes reveal distinct geographical patterns. *Mol. Immunol.* **45**, 2508-2520
79. Endo, Y., Takahashi, M., Iwaki, D., Ishida, Y., Nakazawa, N., Kodama, T., Matsuzaka, T., Kanno, K., Liu, Y., Tsuchiya, K., Kawamura, I., Ikawa, M., Waguri, S., Wada, I., Matsushita, M., Schwaeble, W. J., and Fujita, T. (2012) Mice deficient in ficolin, a lectin complement pathway recognition molecule, are susceptible to *Streptococcus pneumoniae* infection. *J. Immunol.* **189**, 5860-5866
80. Sheriff, S., Chang, C. Y., and Ezekowitz, R. A. (1994) Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple alpha-helical coiled-coil. *Nat. Struct. Biol.* **1**, 789-794
81. Holmskov, U., Thiel, S., and Jensenius, J. C. (2003) Collectins and ficolins: humoral lectins of the innate immune defense. *Annu. Rev. Immunol.* **21**, 547-578
82. Dec, M., and Wernicki, A. (2006) Conglutinin, CL-43 and CL-46--three bovine collectins. *Pol. J. Vet. Sci.* **9**, 265-275
83. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19-34
84. Crouch, E., Persson, A., Chang, D., and Heuser, J. (1994) Molecular structure of pulmonary surfactant protein D (SP-D). *J. Biol. Chem.* **269**, 17311-17319
85. Madsen, H. O., Garred, P., Thiel, S., Kurtzhals, J. A., Lamm, L. U., Ryder, L. P., and Svejgaard, A. (1995) Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. *J. Immunol.* **155**, 3013-3020
86. Madsen, J., Kliem, A., Tornoe, I., Skjodt, K., Koch, C., and Holmskov, U. (2000) Localization of lung surfactant protein D on mucosal surfaces in human tissues. *J. Immunol.* **164**, 5866-5870
87. Keshi, H., Sakamoto, T., Kawai, T., Ohtani, K., Katoh, T., Jang, S. J., Motomura, W., Yoshizaki, T., Fukuda, M., Koyama, S., Fukuzawa, J., Fukuoh, A., Yoshida, I., Suzuki, Y., and Wakamiya, N. (2006) Identification and characterization of a novel human collectin CL-K1. *Microbiol. Immunol.* **50**, 1001-1013

88. Hansen, S., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M. O., Nielsen, O., Hartshorn, K., Jorgensen, T. J., Skjoedt, K., and Holmskov, U. (2010) Collectin 11 (CL-11, CL-K1) is a MASP-1/3-associated plasma collectin with microbial-binding activity. *J. Immunol.* **185**, 6096-6104
89. Ma, Y. J., Skjoedt, M. O., and Garred, P. (2013) Collectin-11/MASP complex formation triggers activation of the lectin complement pathway--the fifth lectin pathway initiation complex. *J. Innate Immun.* **5**, 242-250
90. Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T., and Wakamiya, N. (1999) Molecular cloning of a novel human collectin from liver (CL-L1). *J. Biol. Chem.* **274**, 13681-13689
91. Axelgaard, E., Jensen, L., Dyrland, T. F., Nielsen, H. J., Enghild, J. J., Thiel, S., and Jensenius, J. C. (2013) Investigations on collectin liver 1. *J. Biol. Chem.* **288**, 23407-23420
92. Henriksen, M. L., Brandt, J., Andrieu, J. P., Nielsen, C., Jensen, P. H., Holmskov, U., Jorgensen, T. J., Palarasah, Y., Thielens, N. M., and Hansen, S. (2013) Heteromeric complexes of native collectin kidney 1 and collectin liver 1 are found in the circulation with MASPs and activate the complement system. *J. Immunol.* **191**, 6117-6127
93. Drickamer, K. (1992) Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* **360**, 183-186
94. Wang, H., Head, J., Kosma, P., Brade, H., Muller-Loennies, S., Sheikh, S., McDonald, B., Smith, K., Cafarella, T., Seaton, B., and Crouch, E. (2008) Recognition of heptoses and the inner core of bacterial lipopolysaccharides by surfactant protein D. *Biochemistry* **47**, 710-720
95. Neth, O., Jack, D. L., Dodds, A. W., Holzels, H., Klein, N. J., and Turner, M. W. (2000) Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. *Infect. Immun.* **68**, 688-693
96. van de Wetering, J. K., van Eijk, M., van Golde, L. M., Hartung, T., van Strijp, J. A., and Batenburg, J. J. (2001) Characteristics of surfactant protein A and D binding to lipoteichoic acid and peptidoglycan, 2 major cell wall components of gram-positive bacteria. *J. Infect. Dis.* **184**, 1143-1151
97. Hartshorn, K. L., White, M. R., Shepherd, V., Reid, K., Jensenius, J. C., and Crouch, E. C. (1997) Mechanisms of anti-influenza activity of surfactant proteins A and D: comparison with serum collectins. *Am. J. Physiol.* **273**, L1156-1166
98. Ogasawara, Y., Kuroki, Y., and Akino, T. (1992) Pulmonary surfactant protein D specifically binds to phosphatidylinositol. *J. Biol. Chem.* **267**, 21244-21249
99. Kuroki, Y., and Akino, T. (1991) Pulmonary surfactant protein A (SP-A) specifically binds dipalmitoylphosphatidylcholine. *J. Biol. Chem.* **266**, 3068-3073
100. Seaton, B. A., Crouch, E. C., McCormack, F. X., Head, J. F., Hartshorn, K. L., and Mendelsohn, R. (2010) Structural determinants of pattern recognition by lung collectins. *Innate. Immun.* **16**, 143-150
101. Ip, W. K., Takahashi, K., Ezekowitz, R. A., and Stuart, L. M. (2009) Mannose-binding lectin and innate immunity. *Immunol. Rev.* **230**, 9-21

102. Muto, S., Sakuma, K., Taniguchi, A., and Matsumoto, K. (1999) Human mannose-binding lectin preferentially binds to human colon adenocarcinoma cell lines expressing high amount of Lewis A and Lewis B antigens. *Biol. Pharm. Bull.* **22**, 347-352
103. Schagat, T. L., Wofford, J. A., and Wright, J. R. (2001) Surfactant protein A enhances alveolar macrophage phagocytosis of apoptotic neutrophils. *J. Immunol.* **166**, 2727-2733
104. Stuart, L. M., Takahashi, K., Shi, L., Savill, J., and Ezekowitz, R. A. (2005) Mannose-binding lectin-deficient mice display defective apoptotic cell clearance but no autoimmune phenotype. *J. Immunol.* **174**, 3220-3226
105. Ogden, C. A., deCathelineau, A., Hoffmann, P. R., Bratton, D., Ghebrehiwet, B., Fadok, V. A., and Henson, P. M. (2001) C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells. *J. Exp. Med.* **194**, 781-795
106. Super, M., Thiel, S., Lu, J., Levinsky, R. J., and Turner, M. W. (1989) Association of low levels of mannan-binding protein with a common defect of opsonisation. *Lancet* **2**, 1236-1239
107. Turner, M. W. (1991) Deficiency of mannan binding protein--a new complement deficiency syndrome. *Clin. Exp. Immunol.* **86 Suppl 1**, 53-56
108. Frakking, F. N., van de Wetering, M. D., Brouwer, N., Dolman, K. M., Geissler, J., Lemkes, B., Caron, H. N., and Kuijpers, T. W. (2006) The role of mannose-binding lectin (MBL) in paediatric oncology patients with febrile neutropenia. *Eur. J. Cancer.* **42**, 909-916
109. Shi, L., Takahashi, K., Dundee, J., Shahroor-Karni, S., Thiel, S., Jensenius, J. C., Gad, F., Hamblin, M. R., Sastry, K. N., and Ezekowitz, R. A. (2004) Mannose-binding lectin-deficient mice are susceptible to infection with *Staphylococcus aureus*. *J. Exp. Med.* **199**, 1379-1390
110. Sorensen, G. L., Husby, S., and Holmskov, U. (2007) Surfactant protein A and surfactant protein D variation in pulmonary disease. *Immunobiology* **212**, 381-416
111. Haczku, A. (2008) Protective role of the lung collectins surfactant protein A and surfactant protein D in airway inflammation. *J. Allergy Clin. Immunol.* **122**, 861-879; quiz 880-861
112. Foreman, M. G., Kong, X., DeMeo, D. L., Pillai, S. G., Hersh, C. P., Bakke, P., Gulsvik, A., Lomas, D. A., Litonjua, A. A., Shapiro, S. D., Tal-Singer, R., and Silverman, E. K. (2011) Polymorphisms in surfactant protein-D are associated with chronic obstructive pulmonary disease. *Am. J. Respir. Cell. Mol. Biol.* **44**, 316-322
113. Botas, C., Poulain, F., Akiyama, J., Brown, C., Allen, L., Goerke, J., Clements, J., Carlson, E., Gillespie, A. M., Epstein, C., and Hawgood, S. (1998) Altered surfactant homeostasis and alveolar type II cell morphology in mice lacking surfactant protein D. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11869-11874
114. Yoshida, M., and Whitsett, J. A. (2006) Alveolar macrophages and emphysema in surfactant protein-D-deficient mice. *Respirology* **11 Suppl**, S37-40
115. LeVine, A. M., Whitsett, J. A., Gwozdz, J. A., Richardson, T. R., Fisher, J. H., Burhans, M. S., and Korfhagen, T. R. (2000) Distinct effects of surfactant protein A or D deficiency during bacterial infection on the lung. *J. Immunol.* **165**, 3934-3940

116. Rooryck, C., Diaz-Font, A., Osborn, D. P., Chabchoub, E., Hernandez-Hernandez, V., Shamseldin, H., Kenny, J., Waters, A., Jenkins, D., Kaissi, A. A., Leal, G. F., Dallapiccola, B., Carnevale, F., Bitner-Glindzicz, M., Lees, M., Hennekam, R., Stanier, P., Burns, A. J., Peeters, H., Alkuraya, F. S., and Beales, P. L. (2011) Mutations in lectin complement pathway genes COLEC11 and MASP1 cause 3MC syndrome. *Nat. Genet.* **43**, 197-203
117. Lee, J. K., Baum, L. G., Moremen, K., and Pierce, M. (2004) The X-lectins: a new family with homology to the *Xenopus laevis* oocyte lectin XL-35. *Glycoconj. J.* **21**, 443-450
118. Lee, J. K., Buckhaults, P., Wilkes, C., Teilhet, M., King, M. L., Moremen, K. W., and Pierce, M. (1997) Cloning and expression of a *Xenopus laevis* oocyte lectin and characterization of its mRNA levels during early development. *Glycobiology* **7**, 367-372
119. Lee, J. K., Schnee, J., Pang, M., Wolfert, M., Baum, L. G., Moremen, K. W., and Pierce, M. (2001) Human homologs of the *Xenopus* oocyte cortical granule lectin XL35. *Glycobiology* **11**, 65-73
120. Batista, C. M. D. S., Yang, R. Z., Lee, M. J., Glynn, N. M., Yu, D. Z., Pray, J., Ndubuizu, K., Patil, S., Schwartz, A., Kligman, M., Fried, S. K., Gong, D. W., Shuldiner, A. R., Pollin, T. I., and McLenithan, J. C. (2007) Omentin plasma levels and gene expression are decreased in obesity. *Diabetes* **56**, 1655-1661
121. Suzuki, Y. A., Shin, K., and Lonnerdal, B. (2001) Molecular cloning and functional expression of a human intestinal lactoferrin receptor. *Biochemistry* **40**, 15771-15779
122. Thomsen, T., Schlosser, A., Holmskov, U., and Sorensen, G. L. (2011) Ficolins and FIBCD1: soluble and membrane bound pattern recognition molecules with acetyl group selectivity. *Mol. Immunol.* **48**, 369-381
123. Tsuji, S., Yamashita, M., Nishiyama, A., Shinohara, T., Zhongwei, U., Myrvik, Q. N., Hoffman, D. R., Henriksen, R. A., and Shibata, Y. (2007) Differential structure and activity between human and mouse intelectin-1: Human intelectin-1 is a disulfide-linked trimer, whereas mouse homologue is a monomer. *Glycobiology* **17**, 1045-1051
124. Komiya, T., Tanigawa, Y., and Hirohashi, S. (1998) Cloning of the novel gene intelectin, which is expressed in intestinal paneth cells in mice. *Biochem. Biophys. Res. Comm.* **251**, 759-762
125. French, A. T., Bethune, J. A., Knight, P. A., McNeilly, T. N., Wattedegera, S., Rhind, S., Miller, H. R., and Pemberton, A. D. (2007) The expression of intelectin in sheep goblet cells and upregulation by interleukin-4. *Vet. Immunol. Immunopathol.* **120**, 41-46
126. Tsuji, S., Uehori, J., Matsumoto, M., Suzuki, Y., Matsuhisa, A., Toyoshima, K., and Seya, T. (2001) Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *J. Biol. Chem.* **276**, 23456-23463
127. Lesna, J., Ticha, A., Hyspler, R., Musil, F., Blaha, V., Sobotka, L., Zadak, Z., and Smahelova, A. (2015) Omentin-1 plasma levels and cholesterol metabolism in obese patients with diabetes mellitus type 1: impact of weight reduction. *Nutr. Diabetes.* **5**, e183
128. Pemberton, A. D., Knight, P. A., Wright, S. H., and Miller, H. R. (2004) Proteomic analysis of mouse jejunal epithelium and its response to infection with the intestinal nematode, *Trichinella spiralis*. *Proteomics* **4**, 1101-1108

129. Kuperman, D. A., Lewis, C. C., Woodruff, P. G., Rodriguez, M. W., Yang, Y. H., Dolganov, G. M., Fahy, J. V., and Erle, D. J. (2005) Dissecting asthma using focused transgenic modeling and functional genomics. *J. Allergy Clin. Immunol.* **116**, 305-311
130. Kerr, S. C., Carrington, S. D., Oscarson, S., Gallagher, M. E., Solon, M., Yuan, S. P., Ahn, J. N., Dougherty, R. H., Finkbeiner, W. E., Peters, M. C., and Fahy, J. V. (2014) Intelectin-1 Is a Prominent Protein Constituent of Pathologic Mucus Associated with Eosinophilic Airway Inflammation in Asthma. *Am. J. Resp. Crit. Care* **189**, 1005-1007
131. Pedersen, L. L., and Turco, S. J. (2003) Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. *Cell. Mol. Life Sci.* **60**, 259-266
132. Nassau, P. M., Martin, S. L., Brown, R. E., Weston, A., Monsey, D., McNeil, M. R., and Duncan, K. (1996) Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *J. Bacteriol.* **178**, 1047-1052
133. Tsuji, S., Yamashita, M., Hoffman, D. R., Nishiyama, A., Shinohara, T., Ohtsu, T., and Shibata, Y. (2009) Capture of heat-killed *Mycobacterium bovis* bacillus Calmette-Guerin by intelectin-1 deposited on cell surfaces. *Glycobiology* **19**, 518-526
134. Pemberton, A. D., Knight, P. A., Gamble, J., Colledge, W. H., Lee, J. K., Pierce, M., and Miller, H. R. (2004) Innate BALB/c enteric epithelial responses to *Trichinella spiralis*: inducible expression of a novel goblet cell lectin, intelectin-2, and its natural deletion in C57BL/10 mice. *J. Immunol.* **173**, 1894-1901
135. Datta, R., deSchoolmeester, M. L., Hedeler, C., Paton, N. W., Brass, A. M., and Else, K. J. (2005) Identification of novel genes in intestinal tissue that are regulated after infection with an intestinal nematode parasite. *Infect. Immun.* **73**, 4025-4033
136. Voehringer, D., Stanley, S. A., Cox, J. S., Completo, G. C., Lowary, T. L., and Locksley, R. M. (2007) *Nippostrongylus brasiliensis*: Identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. *Exp. Parasitol.* **116**, 458-466
137. Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M., Bitton, A., Dassopoulos, T., Datta, L. W., Green, T., Griffiths, A. M., Kistner, E. O., Murtha, M. T., Regueiro, M. D., Rotter, J. I., Schumm, L. P., Steinhart, A. H., Targan, S. R., Xavier, R. J., Libioulle, C., Sandor, C., Lathrop, M., Belaiche, J., Dewit, O., Gut, I., Heath, S., Laukens, D., Mni, M., Rutgeerts, P., Van Gossum, A., Zelenika, D., Franchimont, D., Hugot, J. P., de Vos, M., Vermeire, S., Louis, E., Cardon, L. R., Anderson, C. A., Drummond, H., Nimmo, E., Ahmad, T., Prescott, N. J., Onnie, C. M., Fisher, S. A., Marchini, J., Ghori, J., Bumpstead, S., Gwilliam, R., Tremelling, M., Deloukas, P., Mansfield, J., Jewell, D., Satsangi, J., Mathew, C. G., Parkes, M., Georges, M., Daly, M. J., Consortium, N. I. G., Consortium, B.-F. I., and Control, W. T. C. (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955-962
138. Fasano, A., and Shea-Donohue, T. (2005) Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. *Nat. Rev. Gastroenterol. Hepatol.* **2**, 416-422
139. Terazono, K., Yamamoto, H., Takasawa, S., Shiga, K., Yonemura, Y., Tochino, Y., and Okamoto, H. (1988) A novel gene activated in regenerating islets. *J. Biol. Chem.* **263**, 2111-2114

140. Iovanna, J., Frigerio, J. M., Dusetti, N., Ramare, F., Raibaud, P., and Dagorn, J. C. (1993) Lithostathine, an inhibitor of CaCO₃ crystal growth in pancreatic juice, induces bacterial aggregation. *Pancreas* **8**, 597-601
141. Gross, J., Brauer, A. W., Bringhurst, R. F., Corbett, C., and Margolies, M. N. (1985) An unusual bovine pancreatic protein exhibiting pH-dependent globule-fibril transformation and unique amino acid sequence. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5627-5631
142. Gallo, R. L., and Hooper, L. V. (2012) Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12**, 503-516
143. Cash, H. L., Whitham, C. V., Behrendt, C. L., and Hooper, L. V. (2006) Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126-1130
144. Hartupee, J. C., Zhang, H., Bonaldo, M. F., Soares, M. B., and Dieckgraefe, B. K. (2001) Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: Reg IV. *Biochim. Biophys. Acta.* **1518**, 287-293
145. Vaishnava, S., Yamamoto, M., Severson, K. M., Ruhn, K. A., Yu, X., Koren, O., Ley, R., Wakeland, E. K., and Hooper, L. V. (2011) The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255-258
146. Brandl, K., Plitas, G., Mihu, C. N., Ubeda, C., Jia, T., Fleisher, M., Schnabl, B., DeMatteo, R. P., and Pamer, E. G. (2008) Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature* **455**, 804-807
147. Bertrand, J. A., Pignol, D., Bernard, J. P., Verdier, J. M., Dagorn, J. C., and Fontecilla-Camps, J. C. (1996) Crystal structure of human lithostathine, the pancreatic inhibitor of stone formation. *EMBO J.* **15**, 2678-2684
148. Lehotzky, R. E., Partch, C. L., Mukherjee, S., Cash, H. L., Goldman, W. E., Gardner, K. H., and Hooper, L. V. (2010) Molecular basis for peptidoglycan recognition by a bactericidal lectin. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 7722-7727
149. Ho, M. R., Lou, Y. C., Wei, S. Y., Luo, S. C., Lin, W. C., Lyu, P. C., and Chen, C. (2010) Human RegIV protein adopts a typical C-type lectin fold but binds mannan with two calcium-independent sites. *J. Mol. Biol.* **402**, 682-695
150. Mukherjee, S., Partch, C. L., Lehotzky, R. E., Whitham, C. V., Chu, H., Bevins, C. L., Gardner, K. H., and Hooper, L. V. (2009) Regulation of C-type lectin antimicrobial activity by a flexible N-terminal prosegment. *J. Biol. Chem.* **284**, 4881-4888
151. Gross, J., Carlson, R. I., Brauer, A. W., Margolies, M. N., Warshaw, A. L., and Wands, J. R. (1985) Isolation, characterization, and distribution of an unusual pancreatic human secretory protein. *J. Clin. Invest.* **76**, 2115-2126
152. Mukherjee, S., Zheng, H., Derebe, M. G., Callenberg, K. M., Partch, C. L., Rollins, D., Propheter, D. C., Rizo, J., Grabe, M., Jiang, Q. X., and Hooper, L. V. (2014) Antibacterial membrane attack by a pore-forming intestinal C-type lectin. *Nature* **505**, 103-107
153. Schroeder, B. O., Wu, Z., Nuding, S., Groscurth, S., Marcinowski, M., Beisner, J., Buchner, J., Schaller, M., Stange, E. F., and Wehkamp, J. (2011) Reduction of disulphide bonds unmasks potent antimicrobial activity of human beta-defensin 1. *Nature* **469**, 419-423

154. Brandl, K., Plitas, G., Schnabl, B., DeMatteo, R. P., and Pamer, E. G. (2007) MyD88-mediated signals induce the bactericidal lectin RegIII gamma and protect mice against intestinal *Listeria monocytogenes* infection. *J. Exp. Med.* **204**, 1891-1900
155. Ogawa, H., Fukushima, K., Naito, H., Funayama, Y., Unno, M., Takahashi, K., Kitayama, T., Matsuno, S., Ohtani, H., Takasawa, S., Okamoto, H., and Sasaki, I. (2003) Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. *Inflamm. Bowel Dis.* **9**, 162-170
156. Swidsinski, A., Weber, J., Loening-Baucke, V., Hale, L. P., and Lochs, H. (2005) Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J. Clin. Microbiol.* **43**, 3380-3389
157. Guan, R., Roychowdhury, A., Ember, B., Kumar, S., Boons, G. J., and Mariuzza, R. A. (2004) Structural basis for peptidoglycan binding by peptidoglycan recognition proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17168-17173
158. Liu, C., Xu, Z., Gupta, D., and Dziarski, R. (2001) Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. *J. Biol. Chem.* **276**, 34686-34694
159. Dziarski, R., and Gupta, D. (2010) Review: Mammalian peptidoglycan recognition proteins (PGRPs) in innate immunity. *Innate Immun.* **16**, 168-174
160. Yoshida, H., Kinoshita, K., and Ashida, M. (1996) Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. *J. Biol. Chem.* **271**, 13854-13860
161. Rohatgi, A., Ayers, C. R., Khera, A., McGuire, D. K., Das, S. R., Matulevicius, S., Timaran, C. H., Rosero, E. B., and de Lemos, J. A. (2009) The association between peptidoglycan recognition protein-1 and coronary and peripheral atherosclerosis: Observations from the Dallas Heart Study. *Atherosclerosis* **203**, 569-575
162. Royet, J., Gupta, D., and Dziarski, R. (2011) Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat. Rev. Immunol.* **11**, 837-851
163. Kashyap, D. R., Wang, M., Liu, L. H., Boons, G. J., Gupta, D., and Dziarski, R. (2011) Peptidoglycan recognition proteins kill bacteria by activating protein-sensing two-component systems. *Nat. Med.* **17**, 676-683
164. Wang, Z. M., Li, X. N., Cocklin, R. R., Wang, M. H., Wang, M., Fukase, K., Inamura, S., Kusumoto, S., Gupta, D., and Dziarski, R. (2003) Human peptidoglycan recognition protein-L is an N-acetylmuramoyl-L-alanine amidase. *J. Biol. Chem.* **278**, 49044-49052
165. Lu, X., Wang, M., Qi, J., Wang, H., Li, X., Gupta, D., and Dziarski, R. (2006) Peptidoglycan recognition proteins are a new class of human bactericidal proteins. *J. Biol. Chem.* **281**, 5895-5907
166. Wang, M. H., Liu, L. H., Wang, S. Y., Li, X. N., Lu, X. F., Gupta, D., and Dziarski, R. (2007) Human peptidoglycan recognition proteins require zinc to kill both Gram-positive and Gram-negative bacteria and are synergistic with antibacterial peptides. *J. Immunol.* **178**, 3116-3125
167. Kohanski, M. A., Dwyer, D. J., Wierzbowski, J., Cottarel, G., and Collins, J. J. (2008) Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death. *Cell* **135**, 679-690

168. Ganz, T. (2003) Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* **3**, 710-720
169. Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J. P., Heimburg-Molinaro, J., Ju, T., Molinaro, R. J., Rivera-Marrero, C., Xia, B., Smith, D. F., and Cummings, R. D. (2010) Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* **16**, 295-301
170. Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, L. P., Sharma, Y., Anderson, C. A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S. L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J. P., Ahmad, T., Amininejad, L., Ananthakrishnan, A. N., Andersen, V., Andrews, J. M., Baidoo, L., Balschun, T., Bampton, P. A., Bitton, A., Boucher, G., Brand, S., Buning, C., Cohain, A., Cichon, S., D'Amato, M., De Jong, D., Devaney, K. L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L. R., Franchimont, D., Fransen, K., Gearry, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X. L., Karlsen, T. H., Kupcinskis, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I. C., Lees, C. W., Louis, E., Mahy, G., Mansfield, J., Morgan, A. R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C. Y., Potocnik, U., Prescott, N. J., Regueiro, M., Rotter, J. I., Russell, R. K., Sanderson, J. D., Sans, M., Satsangi, J., Schreiber, S., Simms, L. A., Sventoraityte, J., Targan, S. R., Taylor, K. D., Tremelling, M., Verspaget, H. W., De Vos, M., Wijmenga, C., Wilson, D. C., Winkelmann, J., Xavier, R. J., Zeissig, S., Zhang, B., Zhang, C. K., Zhao, H. Y., Silverberg, M. S., Annese, V., Hakonarson, H., Brant, S. R., Radford-Smith, G., Mathew, C. G., Rioux, J. D., Schadt, E. E., Daly, M. J., Franke, A., Parkes, M., Vermeire, S., Barrett, J. C., Cho, J. H., and IIBDGC, I. I. G. C. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124
171. Lakatos, P. L., Fischer, S., Lakatos, L., Gal, I., and Papp, J. (2006) Current concept on the pathogenesis of inflammatory bowel disease-crosstalk between genetic and microbial factors: Pathogenic bacteria and altered bacterial sensing or changes in mucosal integrity take "toll"? *World J. Gastroentero.* **12**, 1829-1841
172. Zulfiqar, F., Hozo, I., Rangarajan, S., Mariuzza, R. A., Dziarski, R., and Gupta, D. (2013) Genetic Association of Peptidoglycan Recognition Protein Variants with Inflammatory Bowel Disease. *Plos One* **8**
173. Saha, S., Jing, X. F., Park, S. Y., Wang, S. Y., Li, X. N., Gupta, D., and Dziarski, R. (2010) Peptidoglycan Recognition Proteins Protect Mice from Experimental Colitis by Promoting Normal Gut Flora and Preventing Induction of Interferon-gamma. *Cell Host Microbe* **8**, 147-162
174. Saha, S., Qi, J., Wang, S. Y., Wang, M. H., Li, X. N., Kim, Y. G., Nunez, G., Gupta, D., and Dziarski, R. (2009) PGLYRP-2 and Nod2 Are Both Required for Peptidoglycan-Induced Arthritis and Local Inflammation. *Cell Host Microbe* **5**, 137-150
175. Rabinovich, G. A., and Toscano, M. A. (2009) Turning 'sweet' on immunity: galectin-glycan interactions in immune tolerance and inflammation. *Nat. Rev. Immunol.* **9**, 338-352

176. Vasta, G. R., Ahmed, H., Nita-Lazar, M., Banerjee, A., Pasek, M., Shridhar, S., Guha, P., and Fernandez-Robledo, J. A. (2012) Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. *Front. Immunol.* **3**
177. Brewer, C. F., Miceli, M. C., and Baum, L. G. (2002) Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Curr. Opin. Struct. Biol.* **12**, 616-623
178. Stowell, S. R., Arthur, C. M., Slanina, K. A., Horton, J. R., Smith, D. F., and Cummings, R. D. (2008) Dimeric Galectin-8 induces phosphatidylserine exposure in leukocytes through polylectosamine recognition by the C-terminal domain. *J. Biol. Chem.* **283**, 20547-20559
179. Cho, M., and Cummings, R. D. (1995) Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. II. Localization and biosynthesis. *J. Biol. Chem.* **270**, 5207-5212
180. Lok, D. J., Van Der Meer, P., de la Porte, P. W., Lipsic, E., Van Wijngaarden, J., Hillege, H. L., and van Veldhuisen, D. J. (2010) Prognostic value of galectin-3, a novel marker of fibrosis, in patients with chronic heart failure: data from the DEAL-HF study. *Clin. Res. Cardiol.* **99**, 323-328
181. Gitt, M. A., Colnot, C., Poirier, F., Nani, K. J., Barondes, S. H., and Leffler, H. (1998) Galectin-4 and galectin-6 are two closely related lectins expressed in mouse gastrointestinal tract. *J. Biol. Chem.* **273**, 2954-2960
182. Liao, D. I., Kapadia, G., Ahmed, H., Vasta, G. R., and Herzberg, O. (1994) Structure of S-lectin, a developmentally regulated vertebrate beta-galactoside-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1428-1432
183. Drickamer, K. (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Biol. Chem.* **263**, 9557-9560
184. Stowell, S. R., Arthur, C. M., Mehta, P., Slanina, K. A., Blixt, O., Leffler, H., Smith, D. F., and Cummings, R. D. (2008) Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. *J. Biol. Chem.* **283**, 10109-10123
185. Liu, F. T., and Rabinovich, G. A. (2005) Galectins as modulators of tumour progression. *Nat. Rev. Cancer* **5**, 29-41
186. Toscano, M. A., Bianco, G. A., Ilarregui, J. M., Croci, D. O., Correale, J., Hernandez, J. D., Zwirner, N. W., Poirier, F., Riley, E. M., Baum, L. G., and Rabinovich, G. A. (2007) Differential glycosylation of TH1, TH2 and TH-17 effector cells selectively regulates susceptibility to cell death. *Nat. Immunol.* **8**, 825-834
187. Mey, A., Leffler, H., Hmama, Z., Normier, G., and Revillard, J. P. (1996) The animal lectin galectin-3 interacts with bacterial lipopolysaccharides via two independent sites. *J. Immunol.* **156**, 1572-1577
188. Vasta, G. R. (2009) Roles of galectins in infection. *Nat. Rev. Microbiol.* **7**, 424-438
189. Rowe, J. A., Handel, I. G., Thera, M. A., Deans, A. M., Lyke, K. E., Kone, A., Diallo, D. A., Raza, A., Kai, O., Marsh, K., Plowe, C. V., Doumbo, O. K., and Moulds, J. M. (2007) Blood group O protects against severe Plasmodium falciparum malaria through the mechanism of reduced rosetting. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17471-17476

190. Knirel, Y. A., Gabius, H. J., Blixt, O., Rapoport, E. M., Khasbiullina, N. R., Shilova, N. V., and Bovin, N. V. (2014) Human tandem-repeat-type galectins bind bacterial non-betaGal polysaccharides. *Glycoconj. J.* **31**, 7-12
191. Stowell, S. R., Arthur, C. M., McBride, R., Berger, O., Razi, N., Heimburg-Molinaro, J., Rodrigues, L. C., Gourdine, J. P., Noll, A. J., von Gunten, S., Smith, D. F., Knirel, Y. A., Paulson, J. C., and Cummings, R. D. (2014) Microbial glycan microarrays define key features of host-microbial interactions. *Nat. Chem. Biol.* **10**, 470-476
192. Arthur, C. M., Patel, S. R., Mener, A., Kamili, N. A., Fasano, R. M., Meyer, E., Winkler, A. M., Sola-Visner, M., Josephson, C. D., and Stowell, S. R. (2015) Innate immunity against molecular mimicry: Examining galectin-mediated antimicrobial activity. *BioEssays : news and reviews in molecular, cellular and developmental biology* **37**, 1327-1337
193. Bottazzi, B., Doni, A., Garlanda, C., and Mantovani, A. (2010) An integrated view of humoral innate immunity: pentraxins as a paradigm. *Annu. Rev. Immunol.* **28**, 157-183
194. Martinez de la Torre, Y., Fabbri, M., Jaillon, S., Bastone, A., Nebuloni, M., Vecchi, A., Mantovani, A., and Garlanda, C. (2010) Evolution of the pentraxin family: the new entry PTX4. *J. Immunol.* **184**, 5055-5064
195. Abernethy, T. J., and Avery, O. T. (1941) The Occurrence during Acute Infections of a Protein Not Normally Present in the Blood : I. Distribution of the Reactive Protein in Patients' Sera and the Effect of Calcium on the Flocculation Reaction with C Polysaccharide of Pneumococcus. *J. Exp. Med.* **73**, 173-182
196. Inforzato, A., Riviaccio, V., Morreale, A. P., Bastone, A., Salustri, A., Scarchilli, L., Verdoliva, A., Vincenti, S., Gallo, G., Chiapparino, C., Pacello, L., Nucera, E., Serlupi-Crescenzi, O., Day, A. J., Bottazzi, B., Mantovani, A., De Santis, R., and Salvatori, G. (2008) Structural characterization of PTX3 disulfide bond network and its multimeric status in cumulus matrix organization. *J. Biol. Chem.* **283**, 10147-10161
197. Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B., and Wood, S. P. (1994) Structure of pentameric human serum amyloid P component. *Nature* **367**, 338-345
198. Shrive, A. K., Cheetham, G. M., Holden, D., Myles, D. A., Turnell, W. G., Volanakis, J. E., Pepys, M. B., Bloomer, A. C., and Greenhough, T. J. (1996) Three dimensional structure of human C-reactive protein. *Nat. Struct. Biol.* **3**, 346-354
199. Du Clos, T. W. (2013) Pentraxins: structure, function, and role in inflammation. *ISRN Inflamm.* **2013**, 379040
200. Gould, J. M., and Weiser, J. N. (2001) Expression of C-reactive protein in the human respiratory tract. *Infect. Immun.* **69**, 1747-1754
201. Muller, B., Peri, G., Doni, A., Torri, V., Landmann, R., Bottazzi, B., and Mantovani, A. (2001) Circulating levels of the long pentraxin PTX3 correlate with severity of infection in critically ill patients. *Crit. Care Med.* **29**, 1404-1407
202. Han, B., Mura, M., Andrade, C. F., Okutani, D., Lodyga, M., dos Santos, C. C., Keshavjee, S., Matthay, M., and Liu, M. (2005) TNFalpha-induced long pentraxin PTX3 expression in human lung epithelial cells via JNK. *J. Immunol.* **175**, 8303-8311
203. Garlanda, C., Hirsch, E., Bozza, S., Salustri, A., De Acetis, M., Nota, R., Maccagno, A., Riva, F., Bottazzi, B., Peri, G., Doni, A., Vago, L., Botto, M., De Santis, R., Carminati,

- P., Siracusa, G., Altruda, F., Vecchi, A., Romani, L., and Mantovani, A. (2002) Non-redundant role of the long pentraxin PTX3 in anti-fungal innate immune response. *Nature* **420**, 182-186
204. Kato, S., Ochiai, M., Sakurada, T., Ohno, S., Miyamoto, K., Sagara, M., Ito, M., Takeuchi, K., Imaki, J., Itoh, K., and Yakabi, K. (2008) Increased expression of long pentraxin PTX3 in inflammatory bowel diseases. *Dig. Dis. Sci.* **53**, 1910-1916
205. Tillett, W. S., and Francis, T. (1930) Serological Reactions in Pneumonia with a Non-Protein Somatic Fraction of Pneumococcus. *J. Exp. Med.* **52**, 561-571
206. Gisch, N., Kohler, T., Ulmer, A. J., Muthing, J., Pribyl, T., Fischer, K., Lindner, B., Hammerschmidt, S., and Zahringer, U. (2013) Structural reevaluation of *Streptococcus pneumoniae* Lipoteichoic acid and new insights into its immunostimulatory potency. *J. Biol. Chem.* **288**, 15654-15667
207. Schneewind, O., and Missiakas, D. (2014) Lipoteichoic acids, phosphate-containing polymers in the envelope of gram-positive bacteria. *J. Bacteriol.* **196**, 1133-1142
208. Thompson, D., Pepys, M. B., and Wood, S. P. (1999) The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* **7**, 169-177
209. Agrawal, A., Shrive, A. K., Greenhough, T. J., and Volanakis, J. E. (2001) Topology and structure of the C1q-binding site on C-reactive protein. *J. Immunol.* **166**, 3998-4004
210. Kaplan, M. H., and Volanakis, J. E. (1974) Interaction of C-reactive protein complexes with the complement system. I. Consumption of human complement associated with the reaction of C-reactive protein with pneumococcal C-polysaccharide and with the choline phosphatides, lecithin and sphingomyelin. *J. Immunol.* **112**, 2135-2147
211. Gershov, D., Kim, S., Brot, N., and Elkon, K. B. (2000) C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity. *J. Exp. Med.* **192**, 1353-1364
212. Mikolajek, H., Kolstoe, S. E., Pye, V. E., Mangione, P., Pepys, M. B., and Wood, S. P. (2011) Structural basis of ligand specificity in the human pentraxins, C-reactive protein and serum amyloid P component. *J. Mol. Recognit.* **24**, 371-377
213. Thompson, D., Pepys, M. B., Tickle, I., and Wood, S. (2002) The structures of crystalline complexes of human serum amyloid P component with its carbohydrate ligand, the cyclic pyruvate acetal of galactose. *J. Mol. Biol.* **320**, 1081-1086
214. Hind, C. R., Collins, P. M., Baltz, M. L., and Pepys, M. B. (1985) Human serum amyloid P component, a circulating lectin with specificity for the cyclic 4,6-pyruvate acetal of galactose. Interactions with various bacteria. *Biochem. J.* **225**, 107-111
215. de Haas, C. J., van der Tol, M. E., Van Kessel, K. P., Verhoef, J., and Van Strijp, J. A. (1998) A synthetic lipopolysaccharide-binding peptide based on amino acids 27-39 of serum amyloid P component inhibits lipopolysaccharide-induced responses in human blood. *J. Immunol.* **161**, 3607-3615
216. Noursadeghi, M., Bickerstaff, M. C., Gallimore, J. R., Herbert, J., Cohen, J., and Pepys, M. B. (2000) Role of serum amyloid P component in bacterial infection: protection of the host or protection of the pathogen. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14584-14589

217. Lu, J., Marnell, L. L., Marjon, K. D., Mold, C., Du Clos, T. W., and Sun, P. D. (2008) Structural recognition and functional activation of FcγR by innate pentraxins. *Nature* **456**, 989-992
218. Mantovani, A., Valentino, S., Gentile, S., Inforzato, A., Bottazzi, B., and Garlanda, C. (2013) The long pentraxin PTX3: a paradigm for humoral pattern recognition molecules. *Ann. N.Y. Acad. Sci.* **1285**, 1-14
219. Reading, P. C., Bozza, S., Gilbertson, B., Tate, M., Moretti, S., Job, E. R., Crouch, E. C., Brooks, A. G., Brown, L. E., Bottazzi, B., Romani, L., and Mantovani, A. (2008) Antiviral activity of the long chain pentraxin PTX3 against influenza viruses. *J. Immunol.* **180**, 3391-3398
220. Gang, T. B., Hammond, D. J., Singh, S. K., Ferguson, D. A., Mishra, V. K., and Agrawal, A. (2012) The Phosphocholine-binding Pocket on C-reactive Protein Is Necessary for Initial Protection of Mice against Pneumococcal Infection. *J. Biol. Chem.* **287**, 43116-43125
221. Guan, R., Wang, Q., Sundberg, E. J., and Mariuzza, R. A. (2005) Crystal structure of human peptidoglycan recognition protein S (PGRP-S) at 1.70 Å resolution. *J. Mol. Biol.* **347**, 683-691
222. Hsieh, T. J., Lin, H. Y., Tu, Z., Huang, B. S., Wu, S. C., and Lin, C. H. (2015) Structural Basis Underlying the Binding Preference of Human Galectins-1, -3 and -7 for Galβ1-3/4GlcNAc. *Plos One* **10**, e0125946
223. Kanagawa, M., Liu, Y., Hanashima, S., Ikeda, A., Chai, W., Nakano, Y., Kojima-Aikawa, K., Feizi, T., and Yamaguchi, Y. (2014) Structural basis for multiple sugar recognition of Jacalin-related human ZG16p lectin. *J. Biol. Chem.* **289**, 16954-16965
224. Chung, L. P., Keshav, S., and Gordon, S. (1988) Cloning the human lysozyme cDNA: inverted Alu repeat in the mRNA and in situ hybridization for macrophages and Paneth cells. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6227-6231
225. Ghos, Y., and Vantrappen, G. (1971) The cytochemical localization of lysozyme in Paneth cell granules. *Histochem. J.* **3**, 175-178
226. Mason, D. Y., and Taylor, C. R. (1975) The distribution of muramidase (lysozyme) in human tissues. *J. Clin. Pathol.* **28**, 124-132
227. Laible, N. J., and Germaine, G. R. (1985) Bactericidal activity of human lysozyme, muramidase-inactive lysozyme, and cationic polypeptides against *Streptococcus sanguis* and *Streptococcus faecalis*: inhibition by chitin oligosaccharides. *Infect. Immun.* **48**, 720-728
228. Feinberg, H., Mitchell, D. A., Drickamer, K., and Weis, W. I. (2001) Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **294**, 2163-2166
229. Pohlmann, S., Zhang, J., Baribaud, F., Chen, Z., Leslie, G. J., Lin, G., Granelli-Piperno, A., Doms, R. W., Rice, C. M., and McKeating, J. A. (2003) Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR. *J. Virol.* **77**, 4070-4080
230. Clemmensen, I., Petersen, L. C., and Kluft, C. (1986) Purification and characterization of a novel, oligomeric, plasminogen kringle 4 binding protein from human plasma: tetranectin. *Eur. J. Biochem.* **156**, 327-333

231. Nielsen, B. B., Kastrup, J. S., Rasmussen, H., Holtet, T. L., Graversen, J. H., Etzerodt, M., Thogersen, H. C., and Larsen, I. K. (1997) Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an alpha-helical coiled coil. *FEBS Lett.* **412**, 388-396
232. Clemmensen, I. (1989) Interaction of tetranectin with sulphated polysaccharides and trypan blue. *Scand. J. Clin. Lab. Invest.* **49**, 719-725
233. Meagher, J. L., Winter, H. C., Ezell, P., Goldstein, I. J., and Stuckey, J. A. (2005) Crystal structure of banana lectin reveals a novel second sugar binding site. *Glycobiology* **15**, 1033-1042
234. Kleene, R., Dartsch, H., and Kern, H. F. (1999) The secretory lectin ZG16p mediates sorting of enzyme proteins to the zymogen granule membrane in pancreatic acinar cells. *Eur. J. Cell. Biol.* **78**, 79-90
235. Schmidt, K., Schrader, M., Kern, H. F., and Kleene, R. (2001) Regulated apical secretion of zymogens in rat pancreas. Involvement of the glycosylphosphatidylinositol-anchored glycoprotein GP-2, the lectin ZG16p, and cholesterol-glycosphingolipid-enriched microdomains. *J. Biol. Chem.* **276**, 14315-14323
236. Tateno, H., Yabe, R., Sato, T., Shibazaki, A., Shikanai, T., Gono, T., Narimatsu, H., and Hirabayashi, J. (2012) Human ZG16p recognizes pathogenic fungi through non-self polyvalent mannose in the digestive system. *Glycobiology* **22**, 210-220
237. Kanagawa, M., Satoh, T., Ikeda, A., Nakano, Y., Yagi, H., Kato, K., Kojima-Aikawa, K., and Yamaguchi, Y. (2011) Crystal structures of human secretory proteins ZG16p and ZG16b reveal a Jacalin-related beta-prism fold. *Biochem. Biophys. Res. Commun.* **404**, 201-205
238. Hanashima, S., Gotze, S., Liu, Y., Ikeda, A., Kojima-Aikawa, K., Taniguchi, N., Varon Silva, D., Feizi, T., Seeberger, P. H., and Yamaguchi, Y. (2015) Defining the Interaction of Human Soluble Lectin ZG16p and Mycobacterial Phosphatidylinositol Mannosides. *Chembiochem : a European journal of chemical biology* **16**, 1502-1511
239. Higashijima, S., Nose, A., Eguchi, G., Hotta, Y., and Okamoto, H. (1997) Mindin/F-spondin family: novel ECM proteins expressed in the zebrafish embryonic axis. *Dev. Biol.* **192**, 211-227
240. Li, Y., Cao, C., Jia, W., Yu, L., Mo, M., Wang, Q., Huang, Y., Lim, J. M., Ishihara, M., Wells, L., Azadi, P., Robinson, H., He, Y. W., Zhang, L., and Mariuzza, R. A. (2009) Structure of the F-spondin domain of mindin, an integrin ligand and pattern recognition molecule. *EMBO J.* **28**, 286-297
241. Cho, W., and Stahelin, R. V. (2005) Membrane-protein interactions in cell signaling and membrane trafficking. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 119-151
242. He, Y. W., Li, H., Zhang, J., Hsu, C. L., Lin, E., Zhang, N., Guo, J., Forbush, K. A., and Bevan, M. J. (2004) The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. *Nat. Immunol.* **5**, 88-97
243. Guleng, B., Lian, Y. M., and Ren, J. L. (2010) Mindin is upregulated during colitis and may activate NF-kappa B in a TLR-9 mediated manner. *World J. Gastroenterol.* **16**, 1070-1075
244. Kahvecioglu, S., Guclu, M., Ustundag, Y., Gul, C. B., Dogan, I., Dagal, T., Esen, B., Esen, S. A., Celik, H., and Esen, I. (2015) Evaluation of serum Spondin 2 levels in the different stages of Type 2 diabetic nephropathy. *Nephrology* **20**, 721-726

245. Jia, W., Li, H., and He, Y. W. (2008) Pattern recognition molecule mindin promotes intranasal clearance of influenza viruses. *J. Immunol.* **180**, 6255-6261
246. Jia, W., Li, H., and He, Y. W. (2005) The extracellular matrix protein mindin serves as an integrin ligand and is critical for inflammatory cell recruitment. *Blood* **106**, 3854-3859
247. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J., and Mathison, J. C. (1990) CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science* **249**, 1431-1433
248. Bazil, V., Baudys, M., Hilgert, I., Stefanova, I., Low, M. G., Zbrozek, J., and Horejsi, V. (1989) Structural Relationship between the Soluble and Membrane-Bound Forms of Human Monocyte Surface Glycoprotein-Cd14. *Mol. Immunol.* **26**, 657-662
249. Funda, D. P., Tuckova, L., Farre, M. A., Iwase, T., Moro, I., and Tlaskalova-Hogenova, H. (2001) CD14 is expressed and released as soluble CD14 by human intestinal epithelial cells in vitro: Lipopolysaccharide activation of epithelial cells revisited. *Infect. Immun.* **69**, 3772-3781
250. Labeta, M. O., Vidal, K., Nores, J. E. R., Arias, M., Vita, N., Morgan, B. P., Guillemot, J. C., Loyaux, D., Ferrara, P., Schmid, D., Affolter, M., Borysiewicz, L. K., Donnet-Hughes, A., and Schiffrin, E. J. (2000) Innate recognition of bacteria in human milk is mediated by a milk-derived highly expressed pattern recognition receptor, soluble CD14. *J. Exp. Med.* **191**, 1807-1812
251. Schumann, R. R., Leong, S. R., Flaggs, G. W., Gray, P. W., Wright, S. D., Mathison, J. C., Tobias, P. S., and Ulevitch, R. J. (1990) Structure and function of lipopolysaccharide binding protein. *Science* **249**, 1429-1431
252. Pugin, J., Schurermaly, C. C., Leturcq, D., Moriarty, A., Ulevitch, R. J., and Tobias, P. S. (1993) Lipopolysaccharide Activation of Human Endothelial and Epithelial-Cells Is Mediated by Lipopolysaccharide-Binding Protein and Soluble Cd14. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2744-2748
253. Triantafilou, M., and Triantafilou, K. (2002) Lipopolysaccharide recognition: CD14, TLRs and the LPS-activation cluster. *Trends Immunol.* **23**, 301-304
254. Faith, J. J., Colombel, J. F., and Gordon, J. I. (2015) Identifying strains that contribute to complex diseases through the study of microbial inheritance. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 633-640
255. Palm, N. W., de Zoete, M. R., Cullen, T. W., Barry, N. A., Stefanowski, J., Hao, L., Degnan, P. H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J. H., Goodman, A. L., and Flavell, R. A. (2014) Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000-1010
256. Kau, A. L., Planer, J. D., Liu, J., Rao, S., Yatsunenkov, T., Trehan, I., Manary, M. J., Liu, T. C., Stappenbeck, T. S., Maleta, K. M., Ashorn, P., Dewey, K. G., Houpt, E. R., Hsieh, C. S., and Gordon, J. I. (2015) Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Sci. Transl. Med.* **7**
257. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J.,

Yatsunenko, T., Zaneveld, J., and Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335-336

Chapter 2

UDP-Galactopyranose Mutase in Nematodes

Portions of this work have been published in:

Wesener, D. A., May, J. F., Huffman E. M., and Kiessling, L. L. UDP-Galactopyranose Mutase in Nematodes. *Biochemistry*. **2013**, 52(25): 4391-4398.

2.1 Abstract

Nematodes represent a diverse phylum of both free living and parasitic species. While the species *Caenorhabditis elegans* (*C. elegans*) is a valuable model organism, parasitic nematodes or helminths pose a serious threat to human health. Indeed, helminths cause many neglected tropical diseases that afflict humans. Nematode glycoconjugates have been implicated in evasive immunomodulation, a hallmark of nematode infections. One monosaccharide residue present in the glycoconjugates of several human pathogens is galactofuranose (Gal_f). This five-membered ring isomer of galactose has not been detected in mammals, making Gal_f metabolic enzymes attractive therapeutic targets. The only known pathway for biosynthetic incorporation of Gal_f into glycoconjugates depends upon generation of the glycosyl donor UDP-Gal_f by the flavoenzyme uridine 5'-diphosphate (UDP) galactopyranose mutase (UGM or Glf). A putative UGM encoding gene (*glf-1*) was recently identified in *C. elegans*. Because Gal_f has yet to be identified in any nematode glycan, the function of the *glf-1* gene product was an open question. Here, we examine the catalytic activity of the *C. elegans glf-1* gene product, CeUGM. We report that CeUGM catalyzes the isomerization of UDP-Gal_f and UDP-galactopyranose (UDP-Gal_p). In the presence of enzyme, substrate and a hydride source, a galactose-N5-FAD adduct was isolated, suggesting the CeUGM flavin adenine dinucleotide (FAD) cofactor serves as a nucleophile in covalent catalysis. The data indicate that CeUGM possesses an active site similar to that of prokaryotic enzymes, despite the low sequence identity (~15%) between eukaryotic and prokaryotic UGM proteins. Additionally, heterocyclic UGM inhibitors developed against prokaryotic proteins also inhibit CeUGM activity. We postulate that these inhibitors can serve as chemical probes of Gal_f in nematodes and as anthelmintic leads. Together, our data suggest that

CeUGM facilitates the biosynthetic incorporation of *Gal*f into nematode glycoconjugates through generation of the glycosyl donor UDP-*Gal*f.

2.2 Introduction

Nematodes are a serious threat to agriculture (1), livestock (2), and human health (3). Plant parasitic nematodes cause estimated crop losses of US\$100 billion annually (4), and some 4 billion people worldwide are infected or at risk of nematode infection (5). Parasitic nematode infections place tremendous stress on national economies and healthcare systems. Helminth infection and modulation of the host immune response are areas of intense research (6,7). Glycoconjugates displayed on the surface of helminths are thought to be major contributors to the observed immunomodulation (8-11). Indeed, several nematode glycoconjugates have been structurally characterized (12) and shown to alter human immune responses (13-16). Additionally, glycoconjugate-based vaccines that target parasitic pathogens are in preclinical or clinical development (17,18). Thus, a more thorough understanding of nematode glycoconjugate biosynthesis can lead to new strategies for combating human helminth infections. The value of such studies is mounting as many helminth strains are becoming increasingly resistant to current chemotherapeutics (2,19,20).

A recently described gene in *C. elegans*, *glf-1*, is intriguing as it may encode UGM enzyme (21). UGM flavoproteins catalyze the production of the glycosyl donor UDP-Galf from UDP-Galp (**Figure 2-1**) (22). The monosaccharide D-Galf is the thermodynamically disfavored, five-membered ring isomer of galactose. Galf residues are absent in mammals, yet they are a prominent component of glycoconjugates from several bacterial, fungal, and protozoan pathogens (23-25). *C. elegans glf-1* deletion mutant experiments suggest that the enzyme is essential (21,26). Still, Galf residues have yet to be identified in a nematode glycan, and the putative UGM has not been shown to be catalytically active (25,27).

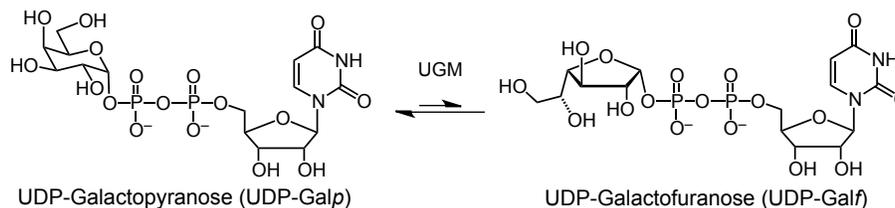


Figure 2-1. Enzymatic biosynthesis of UDP-Galf. The precursor to Galf-containing glycans is UDP-Galf, which is generated by the enzyme UGM, or Glf. UGM catalyzes the isomerization of UDP-Galp and UDP-Galf.

Specific examples of UGM or Galf-containing glycoconjugates contributing to viability and virulence in both prokaryotic and eukaryotic pathogens are emerging. Within the mycobacterial cell wall, a polymer composed of Galf residues, termed the galactan, anchors the mycolic acids to the PG. Genetic deletion (28) or chemical inhibition (29) of UGM prevents mycobacterial growth. UGM deletion in the opportunistic fungus *Aspergillus fumigatus* abates virulence and decreases cell wall thickness, thereby enhancing sensitivity to antifungal agents (30). Genetic disruption of a putative galactofuranosyl transferase gene (LPG1) in *Leishmania major* leads to attenuated infectivity of the parasite and increased susceptibility to human complement and oxidants (31,32). These reports suggest that Galf-containing glycans are essential for the viability of several human pathogens. In the case of nematodes, however, Galf metabolism is largely unexplored. Presumably, this disconnect is due to the absence of characterized Galf-containing glycoconjugates from nematodes (25) and the low sequence identity (~15%) between prokaryotic and eukaryotic UGM proteins (**Figure 2-2**) (21,33).

All organisms that incorporate Galf into their glycans use UDP-Galf as a building block, which is produced by UGM catalysis. To determine whether nematodes possess a catalytically active UGM, we investigated the protein encoded by the *C. elegans* gene *glf-1*. Genes homologous to *glf-1* have been identified in several pathogenic nematode species, and are likely

conserved in the phylum. Our results indicate that the protein encoded by *glf-1* (CeUGM) can catalyze the interconversion of UDP-Galp and UDP-Galf—it is a UGM. In addition, we provide biochemical data that the flavin cofactor of CeUGM participates in covalent catalysis (**Figure 2-3**), a result that is consistent with previous mechanistic studies of UGM homologs from other species (34-36). These data prompted us to devise a homology model for CeUGM, and its validity is supported by the catalytic activities of CeUGM variants. Consistent with the conservation of catalytic mechanism and structural features, small molecule inhibitors developed against prokaryotic UGMs also block CeUGM. Thus, we have identified the first inhibitors of a eukaryotic UGM (29,37).

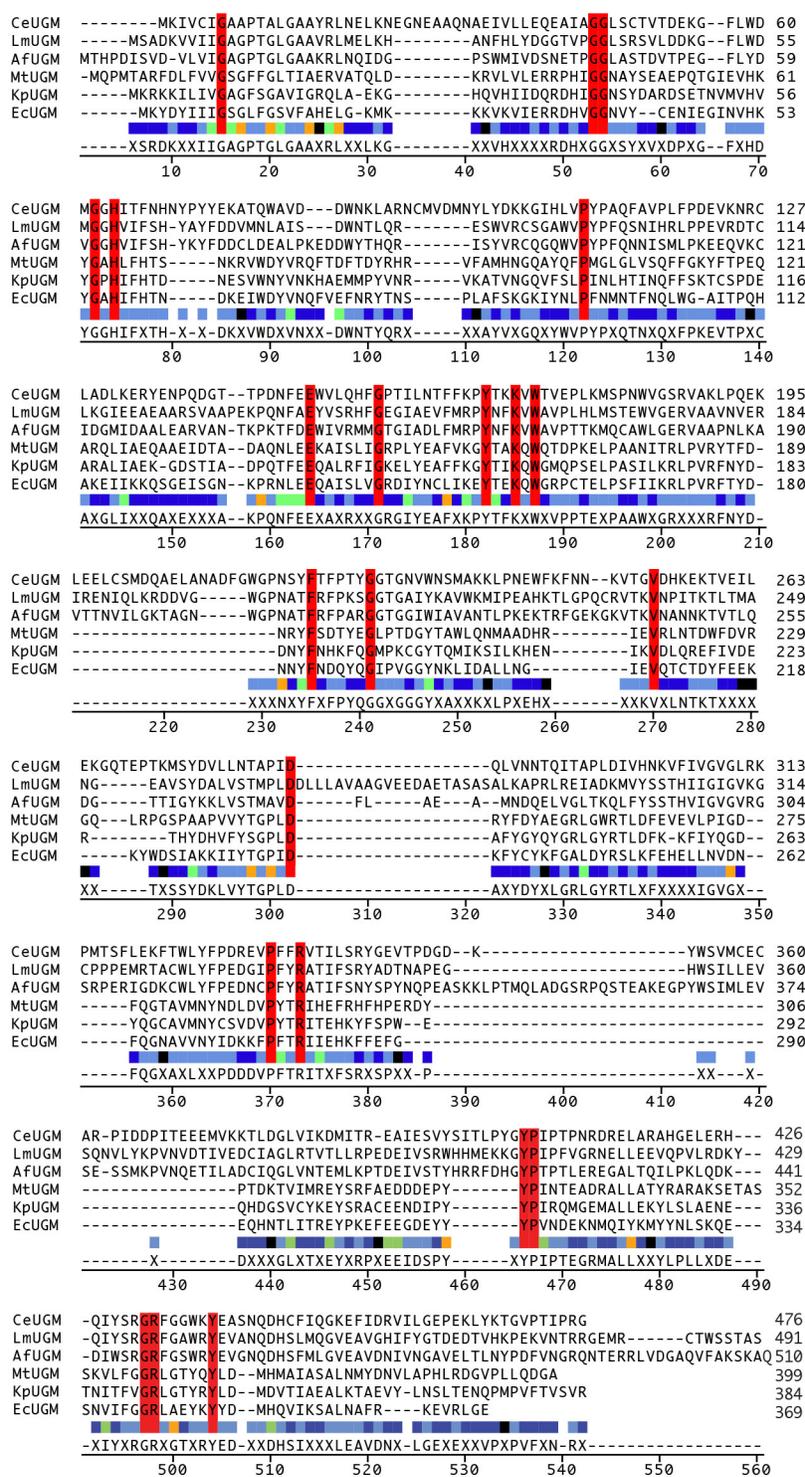


Figure 2-2. Clustal W analysis of eukaryotic and prokaryotic UGM proteins. Residues conserved

in every species are highlighted in red. A complete description of each protein, including its accession number, can be found in section 2.8.5.

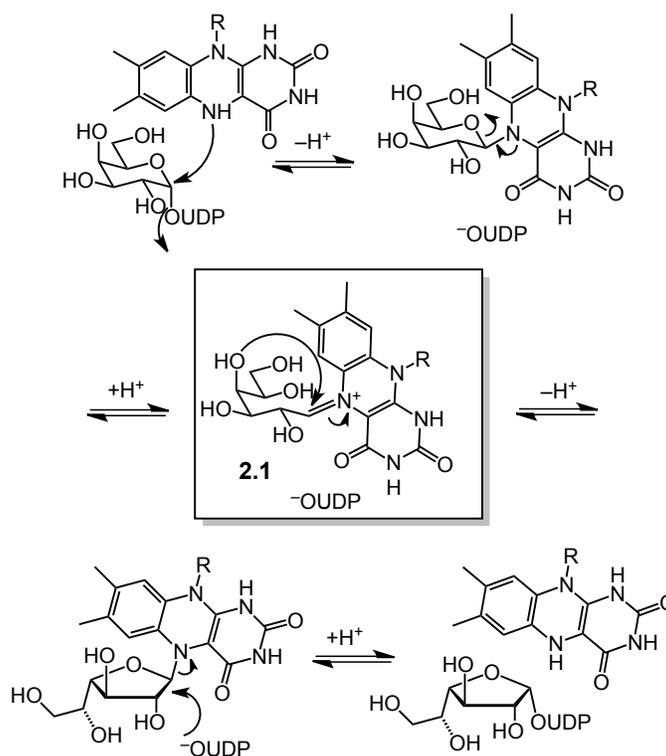


Figure 2-3. A generalized view of the proposed mechanism of UGM, depicting a covalent flavin intermediate. The arrows shown depict changes in covalent bond formation, but whether the mechanism proceeds via an S_N1 or S_N2 reaction is not known nor is the protonation state of the flavin cofactor. A key intermediate in this proposal is iminium ion **2.1**.

2.3 Purification of CeUGM and Initial Velocity Kinetic Analysis

Based on sequence analysis and a LPS synthesis assay (21), the *C. elegans* genome appears to encode a putative UGM, *glf-1*, but the catalytic activity of the *glf-1* gene product, CeUGM, had not been assessed directly. We therefore produced the putative UGM as a His-tagged protein in *E. coli*. The resulting protein gave rise to a UV-visible absorbance spectrum with maximal absorbance near 380 and 450 nm, a spectral signature indicative of a flavoprotein. We then evaluated the enzymatic activity of recombinant CeUGM using an HPLC-based assay

(38,39). Recombinant protein was incubated with synthetic UDP-Galf as the substrate under reducing conditions (40), because reduction of the flavin cofactor is essential for catalytic activity (41). Recombinant CeUGM catalyzed the isomerization reaction.

Steady-state kinetic parameters were determined from the initial velocities of UDP-Galp production over a range of UDP-Galf concentrations (**Figure 2-4**). The K_M and k_{cat} values are approximately ten-fold lower for CeUGM than those reported for prokaryotic UGMs (**Table 2-1**). The catalytic efficiency (k_{cat}/K_M) of CeUGM, however, is similar to that of the UGM from *L. major*, the nearest homolog of CeUGM characterized to date (42). Prior protein localization studies using a CeUGM::GFP fusion revealed that the enzyme is produced in *C. elegans* seam cells (26). Seam cells are involved in nematode surface glycoconjugate biosynthesis. The available data, therefore, suggest that the biological function of CeUGM is to generate UDP-Galf. This building block then can serve as a glycosyl donor for yet unidentified *C. elegans* galactofuranosyl transferases, which biosynthesize Galf-containing glycans. These findings indicate that a catalytically competent UGM is produced in a multicellular organism. We predict that generation of UDP-Galf is not restricted to *C. elegans*, but is general to nematodes with a gene homologous to *glf-1*, such as the human pathogen *Brugia malayi*.

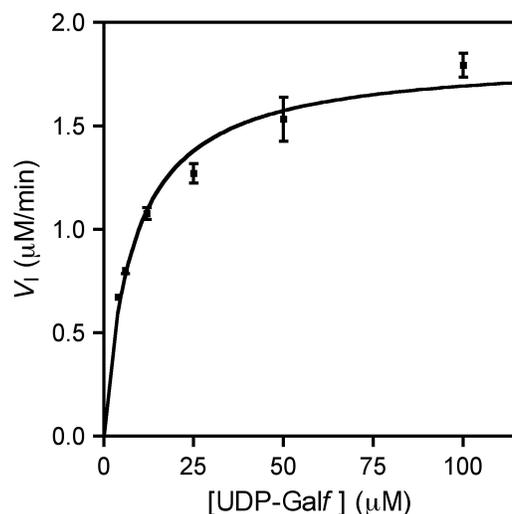


Figure 2-4. Steady-state kinetic analysis of CeUGM. Initial velocities were calculated by measuring the rate of UDP-Galp formation at increasing concentrations of UDP-Galf. Data were fit to the Michaelis-Menten equation. Error bars represent the standard deviation (s.d.) of measurements performed in triplicate.

Table 2-1. Kinetic Parameters of UGM Homologs

Species	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($10^4 \text{ M}^{-1} \text{ s}^{-1}$)
<i>C. elegans</i>	8 ± 0.8	$0.61 \pm .08$	7.6 ± 1.2
<i>L. major</i> (42)	87 ± 11	5 ± 0.2	5.7 ± 0.6
<i>A. fumigatus</i> (43)	110 ± 15	72 ± 4	65 ± 9
<i>K. pneumoniae</i> (44)	43 ± 6	5.5 ± 0.7	13 ± 2
<i>E. coli</i> (39)	27	22	81

^a All constants were determined with UDP-Galf as the substrate.

2.4 Covalent Catalysis via FAD

In 2004, our group provided direct evidence that prokaryotic UGM enzymes use their flavin cofactor to facilitate covalent catalysis (**Figure 2-3**) (34). Since then, additional results have been described that are consistent with such a mechanism, including those using NMR (35), FAD analogs (45,46), other spectrophotometric methods (34,36), and computational approaches (47). We were interested in whether results supporting this unique mechanism could be obtained

using CeUGM. Mechanistic conservation would support the development of highly selective irreversible covalent inhibitors.

Carbohydrate ring contraction *via* covalent catalysis by FAD is predicted to proceed through a galactose–N5-flavin iminium ion intermediate (**Figure 2-3**). If CeUGM catalyzes isomerization by this nucleophilic mechanism, the addition of exogenous reductant should trap the iminium ion to form a covalent adduct, compound **2.2** (**Figure 2-5**) (34). Isolation and characterization of this adduct would suggest flavin engages in covalent catalysis in evolutionarily distant eukaryotic and prokaryotic proteins. To trap the iminium ion intermediate, a mixture of CeUGM, UDP-Galp and sodium dithionite was exposed to the hydride donor sodium cyanoborohydride. After the reaction was quenched with the hydride donor, the protein was precipitated and the supernatant analyzed using high-performance liquid chromatography mass spectrometry (HPLC-MS). In reactions containing cyanoborohydride, a second peak of near equal abundance elutes prior to FAD (**Figure 2-5A**). Analysis of the species that gives rise to this peak using electrospray mass spectrometry indicates that the predominant ion corresponds to the covalent adduct (**Figure 2-5B**).

The proposal that UGM uses its flavin cofactor in covalent catalysis was initially controversial, as it invokes a new catalytic role for flavin (34). Our data from CeUGM provides additional support that UGM-catalyzed ring isomerization proceeds *via* covalent catalysis. A related adduct was recently reported to be trapped from the *Trypanosoma cruzi* UGM, although the mass and structure correspond to a C4a hydroxylated species (36). Position C4a of the flavin isoalloxazine is the site of molecular oxygen addition during flavin reoxidation, a step not consistent with the non-redox UGM mechanism we have proposed (34). Our isolation of a

galactose–N5-flavin covalent adduct suggests that CeUGM uses a nucleophilic flavin cofactor to catalyze isomerization of UDP-Galp and UDP-Galf (**Figure 2-3**). This reactivity mode further expands the catalytic breadth of flavoprotein catalyzed reactions (48). We and others (49) postulate that the nucleophilic character of the UGM flavin can be targeted by small molecules for extremely potent and selective inhibition of UDP-Galf biosynthesis.

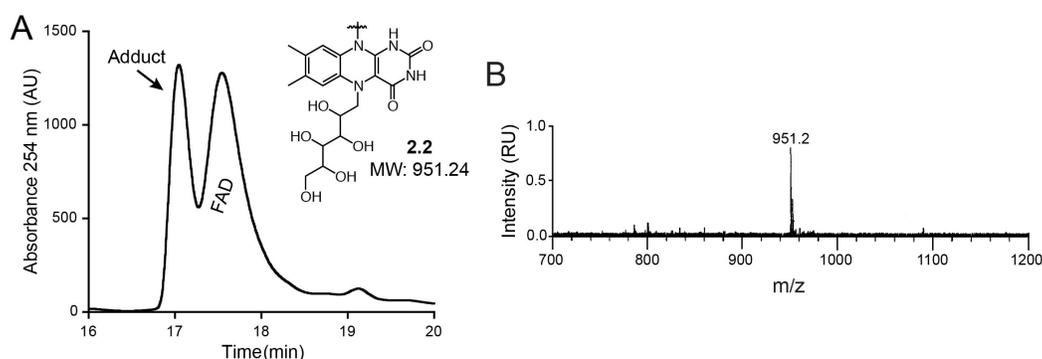


Figure 2-5. HPLC-MS analysis of the products of trapping the CeUGM-catalyzed reaction of UDP-Galp with sodium cyanoborohydride. (A) A reverse phase HPLC chromatogram obtained from analysis of the soluble reaction products monitored at 254 nm. Inset shows the structure and mass of the predicted galactose–N5-FAD adduct, **2.2**, which is the product of reduction of iminium ion **2.1** (**Figure 2-3**). (B) Mass spectral analysis of product eluting as adduct in the chromatograph above. The predicted mass of unmodified FAD^{ox} is 785.16 Da.

2.5 Proposed Structure and Active Site of CeUGM

Structural data have been invaluable for understanding substrate binding and the catalytic mechanism of prokaryotic UGM proteins (35,41,50-52). Recently, two independent groups described the application of protein x-ray crystallography to determine the structure and biological reducing agent of the *A. fumigatus* UGM (53-55). These studies revealed that despite low sequence identity and several insertions into the *A. fumigatus* gene, the overall structure and folds of prokaryotic and eukaryotic UGMs are similar (53,54). Additionally, our ability to trap a catalytic intermediate (compound **2.2**) further suggested that the active site of CeUGM is similar

to that of its prokaryotic homologs. We therefore postulated that small molecules previously shown to inhibit prokaryotic UGMs (29,37) would also block CeUGM. To assess this possibility, we generated a homology model (56) of CeUGM using the UGM from *M. tuberculosis* (MtUGM) (PDB code: 1V0J) (57) as a template. The model predicts that the overall architecture of CeUGM and MtUGM is similar. It also indicates that the locations of many residues involved in substrate binding are structurally conserved (**Figure 2-6A & B**). The primary sequence alignment and the homology model collectively suggest that arginines 187 and 336 from CeUGM correspond to the two arginine residues essential for substrate binding by prokaryotic proteins (44). To test the accuracy of the homology model, two CeUGM variants were generated in which either Arg187 or Arg336 was substituted with alanine. The proper folding and flavin binding of the variants was assessed by circular dichroism (CD) and UV-visible absorbance spectroscopy, respectively. The resulting spectra were nearly identical to those from wild-type enzyme (**Figure 2-6C**).

The homology model led us to predict that both arginine variants would exhibit significantly diminished catalytic activity. Indeed, under standard conditions, replacement of either arginine residue drastically hampers catalysis. To quantify the activity of these enzyme variants, a high substrate concentration (approximately 12 fold above the K_M for the wild-type enzyme) and extended periods of incubation were employed. Because the catalytic rate of wild-type CeUGM was not linear during the extended incubations, we chose to measure bulk catalytic activity by integration of the substrate and product HPLC trace peaks. The total fraction of UDP-Galp converted to UDP-Galp by the R187A variant was roughly one tenth of that catalyzed by the wild-type enzyme (**Figure 2-6D**). This diminished, but detectable and quantifiable activity at

a high substrate concentration indicates that our model is useful for identifying active site residues. Because the high concentration of UDP-Galf restores activity in the R187A variant, we postulate that arginine 187 is involved in substrate binding. Under the same conditions, no activity was detected from the R336A variant, suggesting a more critical role in enzyme catalysis. Our data are consistent with results from the *A. fumigatus* UGM (53,54); the collective experiments suggest a role for these arginine residues in coordinating the negatively charged pyrophosphoryl group of the substrate. Cumulatively, the kinetic and spectrophotometric data we present here supports our model of CeUGM. The similarities of the enzyme active sites of our homology model and crystallized prokaryotic enzymes provided impetus to explore whether previously identified inhibitors of the prokaryotic homologs might also inhibit CeUGM.

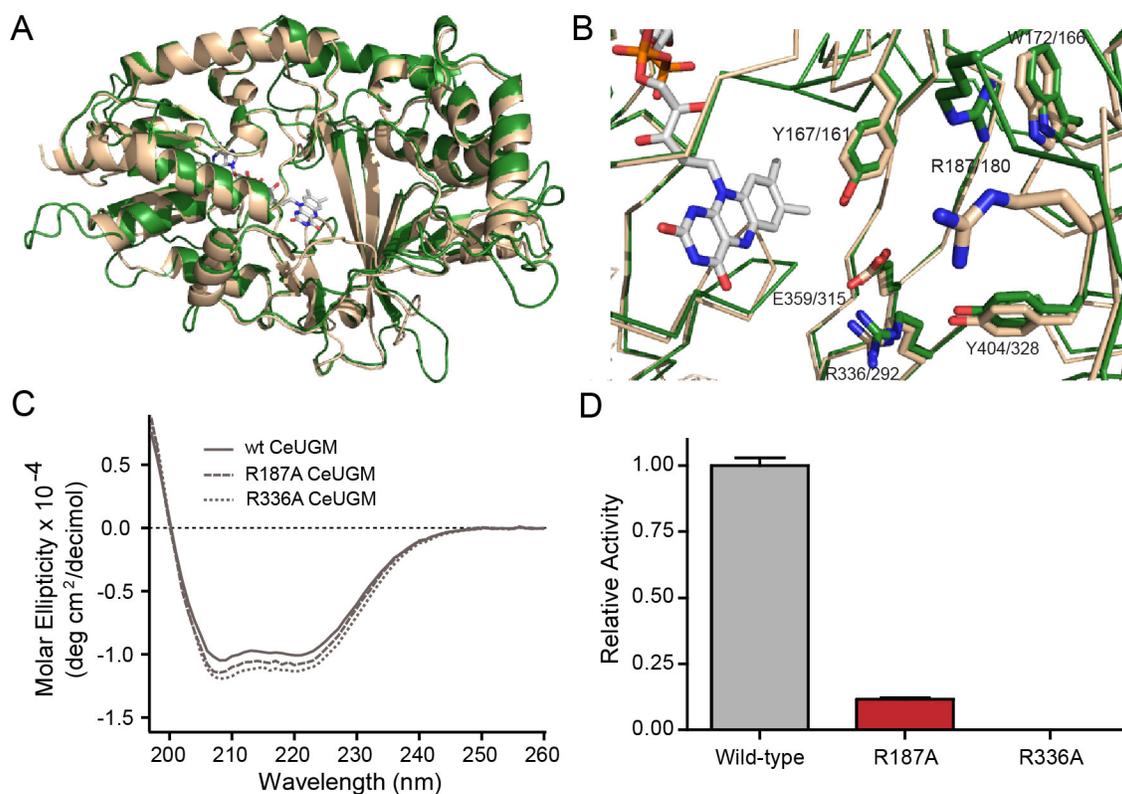


Figure 2-6. Proposed structure and active site of CeUGM. (A) CeUGM homology model (green)

generated using SWISS-MODEL superimposed with the structure of *M. tuberculosis* UGM (PDB Code: 1V0J; wheat). (B) A comparison of residues in the active site. Select conserved residues predicted (CeUGM) or known (*M. tuberculosis* UGM) involved in substrate binding are highlighted, with *C. elegans* residue numbers denoted first. (C) CD spectra for wild-type CeUGM, R187A CeUGM variant, and R336A CeUGM variant. Baseline scans of buffer solution were subtracted from all spectra, and data were converted to molar ellipticity for accurate comparison between samples. (D) Relative activity of wild-type, the R187A variant, and the R336A variant CeUGM at a UDP-Galp concentration of approximately 12-fold above the K_m of the wild-type enzyme. Error bars represent the s.d. of triplicate measurements. Relative activity is derived from normalizing to wild-type enzyme.

2.6 Chemical Inhibition of CeUGM

We tested two known inhibitors of prokaryotic UGMs, compounds **2.3** and **2.4**, with CeUGM (**Figure 2-7A**). For comparison, a compound that is inactive against prokaryotic enzymes, compound **2.5**, was also assessed. Using the HPLC assay that monitors the production of UDP-Galp, compounds **2.3** and **2.4** were shown to be potent inhibitors of CeUGM. The IC_{50} values of compounds **2.3** and **2.4** were 3.3 μ M and 1.8 μ M respectively (**Figure 2-7B, Table 2-2**). When tested at 10 μ M, compound **2.5** had only a modest effect on CeUGM activity. The mode of inhibition was determined by monitoring the kinetics of UDP-Galp formation by CeUGM in the presence of varying concentrations of **2.3**. Analysis of the double reciprocal linear regression plots indicates that the 2-aminothiazole inhibitors are competitive for the active site with UDP-Galp (**Figure 2-8**).

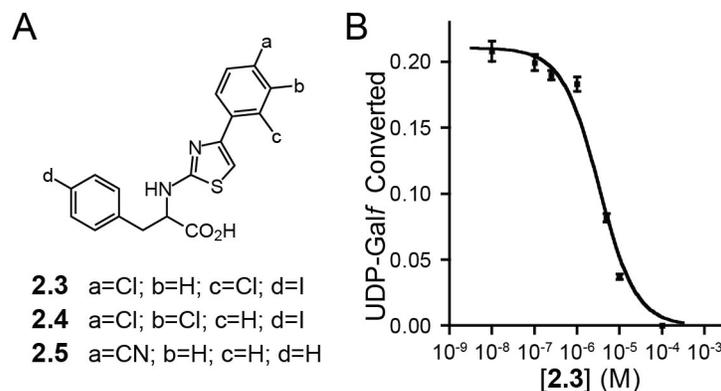


Figure 2-7. Chemical inhibition of CeUGM. (A) Structures of 2-aminothiazole inhibitors used in this study. (B) Inhibition of UDP-Galp isomerization by CeUGM with increasing concentrations of compound 2.3. No UDP-Galp formation could be detected at 100 μM 2.3. Error bars represent the s.d. of triplicate measurements.

Table 2-2. Inhibition and Binding Constants of 2-aminothiazoles Against CeUGM^a

Compound	% Inhibition at 10 μM	IC ₅₀ (μM)	K _i (μM)
2.3	83 ± 6	3.3 ± 1.1	1.3 ± 0.4
2.4	95 ± 1	1.8 ± 1.2	0.7 ± 0.5
2.5	16 ± 3	n/d	n/d

^an/d represents not determined. K_i values were calculated using the Cheng-Prusoff equation (58).

The identification of small molecule inhibitors of CeUGM suggests new research directions. First, the data indicate that inhibitors of eukaryotic UGMs can be found. Of the small panel of potential 2-aminothiazole based compounds we tested, the most effective inhibitor, compound 2.4, had an IC₅₀ of 1.8 μM. These compounds could be used to illuminate the effects of perturbing nematode glycans. Indeed, *C. elegans* serves as a useful model to investigate helminth biology (59-61) — specifically the biosynthesis of glycoconjugates, their structure, and their physiological roles (15,62-64). Glycomic analysis supports this approach as glycoconjugates in *C. elegans* and parasitic species are similar (27,65,66). Inhibitors of nematode UGM enzymes,

therefore, may prove to be valuable for probing *Galf* biology in *C. elegans* and parasitic nematodes. For example, a putative UGM was recently identified in a protein mass spectrometry study of the parasitic nematode *Brugia malayi*, the causative agent of human lymphatic filariasis (67). We therefore anticipate that the inhibitors identified in this study can be used to probe the physiological role of *Galf* in *B. malayi*.

Eukaryotic UGM inhibitors may have other useful roles. It has been previously demonstrated that *Galf* contributes to virulence (30) of the human fungal pathogen *A. fumigatus*, and remarkably, *Galf* constitutes ~5% of the dry weight of the fungus(68). The temporal control afforded with chemical genetics would allow researchers to probe the role of *Galf* during specific stages of infection or host colonization. We therefore anticipate that these compounds will serve as chemical probes and chemotherapeutic leads in nematodes and other eukaryotes that utilize *Galf*.

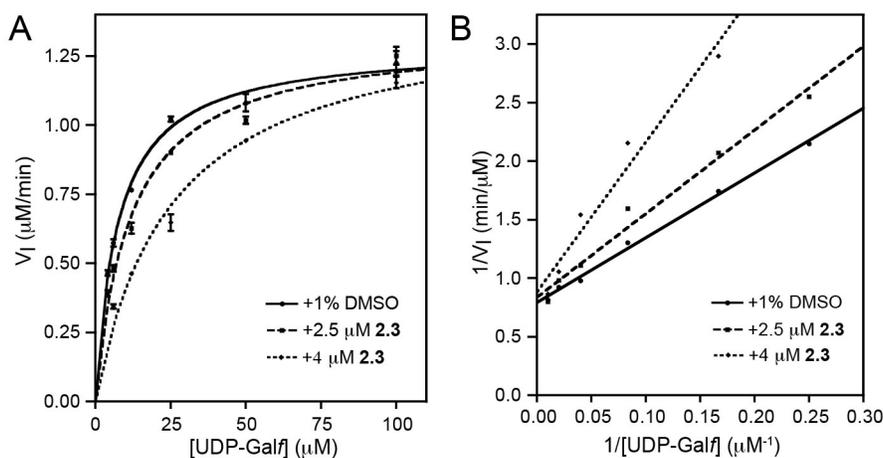


Figure 2-8. Competitive inhibition of CeUGM by compound 2.3. (A) Initial velocity analysis of CeUGM in the presence of increasing concentrations of compound 2.3. Error bars represent the s.d. of triplicate measurements. (B) Double reciprocal analysis of the data presented in (A).

2.7 Conclusions

In summary, the *glf-1* gene from *C. elegans* encodes an enzyme that catalyzes the isomerization of UDP-Galp and UDP-Galf. Our ability to trap a galactose–N5-flavin adduct implicates a nucleophilic flavin in the catalytic mechanism. Conservation of this flavoprotein mechanism supports the development of highly specific activity-based covalent inhibitors of UGM proteins. Our homology model suggests that CeUGM possesses an active site similar to those described for other UGM proteins, a conclusion supported by results from site-directed mutagenesis. The high level of structural similarity between the predicted active sites from the CeUGM homology model and crystallized prokaryotic UGMs prompted us to test previously described inhibitors of prokaryotic proteins as inhibitors of CeUGM (29). Intriguingly, 2-aminothiazole based compounds are competitive inhibitors of CeUGM, indicating eukaryotic UGMs are amenable to small molecule inhibition. Finally, our data suggest that nematodes use CeUGM to catalyze UDP-Galf biosynthesis. It is likely that Galf is then incorporated into critical nematode glycans *via* yet unidentified galactofuranosyl transferases. Identification of Galf in a nematode glycan or the glycosyltransferases responsible for addition of Galf remains critical for understanding and capitalizing on nematode glycoconjugate biosynthesis.

2.8 Methods

2.8.1 Cloning, Expression, and Purification of CeUGM

The *glf-1* ORF was amplified by PCR using PfuTurbo DNA Polymerase (Stratagene) from a pET3a:*glf-1* construct (21) generously provided by Professor Stephen Beverley (Washington University in St. Louis School of Medicine) using the forward primer 5'-GACCACAACGGTTTCCCTCTAGAAATAATTTTG-3' and the reverse primer 5'-GCAGCCGGATCCGCGGCCGCTCCCCGTGGAATAGTTGG-3'. The forward and reverse primers added an *XbaI* and *NotI* restriction site, respectively. The purified PCR product and pET-24a vector were digested with *XbaI* and *NotI* restriction endonucleases (Promega). The resulting products were purified using the QIAquick PCR Purification Kit (Qiagen). Digested pET-24a vector and *glf-1* insert were ligated using T4 DNA Ligase (Promega). The presence of *glf-1* with a vector encoded C-terminal hexahistidine (His₆) tag was confirmed by DNA sequence analysis. The pET-24a:*glf-1* construct was used as template DNA for generating point mutants *via* site-directed mutagenesis using the QuickChange Kit (Stratagene). The oligonucleotide primer sequences used for generating each mutant can be found in **Table 2-3**.

The pET-24a:*glf-1* construct was transformed into competent *E. coli* BL21(DE3) cells (Novagen). Cultures were grown in LB medium supplemented with 50 µg/L kanamycin at 37 °C until OD₆₀₀ = 0.6 were reached. Cells were cooled to 20 °C and protein overexpression was induced upon the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to 0.1 mM. Cells were grown for 18 hours at 20 °C then harvested by centrifugation (5,000g). Pellets were resuspended in buffer containing 20 mM potassium phosphate (pH 7.4), 25 mM imidazole, 300 mM sodium chloride and 15% glycerol. Cells were disrupted by lysozyme, 0.1% Triton X-100, and

sonication (Branson 450 sonifier). Lysates were cleared with centrifugation (22,000g) and filtration (0.45 μM). Cleared lysate was purified with immobilized nickel-ion chromatography using a HisTrap HP column (GE Healthcare) on an AKTA FPLC (Amersham Biosciences). Protein was eluted with a linear gradient of 25 to 500 mM imidazole in potassium phosphate (pH 7.4), 300 mM sodium chloride, and 15% glycerol. Typical yields were 4 mg/L. Fractions of CeUGM, >90% pure, were pooled and dialyzed against a solution of 50 mM potassium phosphate (pH 7.4), 500 mM sodium chloride, 2 mM dithiothreitol (DTT) and 15% glycerol. Protein concentration was determined by absorbance of the flavin cofactor at 450 nm ($\epsilon_{450} = 11,300 \text{ M}^{-1}\text{cm}^{-1}$). Protein was aliquoted, vitrified in liquid nitrogen, and stored at $-80 \text{ }^{\circ}\text{C}$.

Table 2-3. Oligonucleotide Primer Sequences Used to Generate Point Mutants of CeUGM^a

Mutant	Primer Sequence
R187A	5'-CTCCAAATTGGGTTGGATCT <u>GCT</u> GTTGCTAAGCTTCCAC-3'
R336A	5'-CCCAGATCGTGAAGTCCATTCTT <u>CGC</u> IGTCACAATTCTCAGC-3'

^aThe primer described above and its reverse complement (sequence not shown) were used together in a single PCR reaction to generate each mutant. The codon changed during the reaction is underlined.

2.8.2 Enzymatic Activity

The activity of CeUGM was measured using an HPLC-based assay with minor modifications (38,39). Reactions were performed in 50 mM potassium phosphate (pH 7.0) and 10 mM fresh sodium dithionite at $22 \text{ }^{\circ}\text{C}$. The reaction was initiated with the rapid addition of chemically synthesized UDP-Galf to the enzyme solution (40). The time that each reaction was allowed to proceed was adjusted to ensure the conversion of UDP-Galf to UDP-Galp was under 10%. Reaction mixtures were quenched with the addition of an equal volume of 1:1

chloroform:methanol. The aqueous portion was removed and analyzed using a CarboPac PA-100 column (Dionex) on a Waters HPLC system. Reaction substrate (e.g. UDP-Galf) and product (e.g. UDP-Galp) were separated *via* isocratic elution using 200 mM ammonium acetate (pH 8.0), and the eluate was monitored at 262 nm. Initial reaction rates were calculated based on the concentration of substrate and the percentage converted as determined by integration of the HPLC chromatograph. Kinetic parameters were determined by nonlinear regression analysis using Graphpad Prism 4. Quantified error represents the standard error of the mean. The activities of CeUGM variants R187A and R336A were measured at 100 μ M UDP-Galf, or \sim 12 times the K_M value for the wild-type enzyme. The activity of the variants was decreased to a level in which the kinetic and binding constants could not be reliably determined. Data were normalized to the activity measured from the wild-type sample.

2.8.3 UV-Visible Spectroscopy

UV-visible absorbance spectra of wild-type, R187A, and R336A CeUGM, or free FAD were taken in a Cary 50 Bio UV-Visible Spectrophotometer (Varian) using a 1 cm cuvette. As a blank, 50 mM potassium phosphate (pH 7.4), 500 mM sodium chloride, and 15% glycerol was used. Spectra were normalized to 450 nm.

2.8.4 Far-UV Circular Dichroism (CD)

Wild-type, R187A, and R336A CeUGM were dialyzed into 20 mM potassium phosphate (pH 7.0) and 15% glycerol. Protein concentrations were determined using the BCA Protein Assay (Pierce). Samples were diluted to roughly 3 μ M. Spectra were collected in a Model 202SF Circular Dichroism Spectrophotometer (AVIV Instruments) using a quartz cuvette with a 0.1 cm path length. Data were collected every 1 nm from 197-300 nm using seven shot averages at 22

°C. A baseline scan of a solution of 20 mM potassium phosphate (pH 7.0) and 15% glycerol was subtracted from each sample. Molar ellipticities (ME; deg cm²/decimol) were calculated using the equation $ME = MRW\theta/(10lC)$ where MRW is the mean residue weight (114.36 for CeUGM-His₆), θ is the ellipticity in degrees, l is the pathlength in centimeters, and C is the concentration in g/mL. Data were plotted using Igor Pro (WaveMetrics).

2.8.5 Clustal W Alignment

Putative UGM proteins from three eukaryotes and three prokaryotes were selected for Clustal W analysis using MegAlign in the Lasergene 8 Suite (DNASTAR). Eukaryotic UGMs include *Caenorhabditis elegans* (CeUGM), H04M03.4; *Leishmania major* (LmUGM), AAX09638; *Aspergillus fumigatus* (AfUGM), CAI38754. Prokaryotic UGMs analyzed consist of *Mycobacterium tuberculosis* (MtUGM), NP_218326; *Klebsiella pneumoniae* (KpUGM), KP1_3695; *Escherichia coli* (EcUGM), ACD37140. Proteins were aligned using the default Clustal W Method parameters on the slow and accurate mode. Residues that were conserved in every species are highlighted in red.

2.8.6 Homology Model

The homology model of CeUGM was generated using the “Automated Mode” of SWISS-MODEL with default parameters (56). Full length protein was used as the query. As a target, UGM from *M. tuberculosis* (PDB Code: 1V0J) was selected (57). The model, residue 2-473 of CeUGM, was viewed and analyzed using PyMOL.

2.8.7 Intermediate Trapping

A solution of CeUGM, sodium dithionite, and UDP-Galp was allowed to equilibrate for one minute. Solid sodium cyanoborohydride was added to a concentration of 1 M. After thirty

minutes, sodium chloride was added to a final concentration of 1 M to facilitate the extraction of the FAD cofactor. Protein was precipitated with heat and the soluble components were collected. The buffer was exchanged using a Sep-Pak C18 cartridge (Waters). To elute the sample, 10 mM ammonium acetate (pH 6.4) in 1:1 water:methanol was added. The eluate was concentrated to dryness using a SpeedVac SC100 (Varian) under vacuum. Samples were analyzed sequentially using a C18 column on a Shimadzu electrospray ionization HPLC-MS with the mass analyzer in both positive and negative ion mode. For mass assignment, positive ion mode was used.

2.8.8 Chemical Inhibition

Compounds **2.3**, **2.4**, and **2.5** were synthesized as previously reported (29). For percentage inhibition and IC_{50} determination, solutions of CeUGM, sodium dithionite, 12 μ M UDP-Galf, and compound (at various concentrations from dimethyl sulfoxide (DMSO_ stocks) were assessed in the aforementioned HPLC-based product formation assay for catalytic activity. Compound stocks were adjusted such that 1% DMSO was present in each reaction. As a vehicle control, 1% DMSO was used. The IC_{50} values were calculated using the One Site Competition Model from Graphpad Prism 4. Quantified error represents standard error of the mean. To assess the mode of inhibition of the 2-aminothiazoles described above, reaction kinetics were determined in the presence of varying concentrations of compound **2.3**. Data were fit using nonlinear regression analysis. The double reciprocal plot was generated and fit using Graphpad Prism 4 (GraphPad Software).

2.9 Contributions

John F. May performed initial experiments using CeUGM. Elizabeth M. Huffman synthesized and characterized the small molecules **2.3**, **2.4**, and **2.5**.

2.10 Acknowledgements

This research was supported by National Institutes of Health grants R01-AI063596 to L.L.K. I am thankful for support by a National Science Foundation Graduate Research Fellowship and by the NIH-supported Chemistry–Biology Interface Training Program (T32 GM008505) at the University of Wisconsin–Madison. Circular dichroism spectra were obtained at the UW-Madison Biophysics Instrumentation Facility, which is supported by UW-Madison, NSF Grant BIR-9512577, and NIH Grant S10 RR13790. LCMS data was collected in the UW-Madison Chemistry Instrumentation Center Mass Spectrometry Facility on a Shimadzu LCMS-2010A that was purchased with funds from a Keck grant, a Shimadzu grant, and the Chemistry Department.

I thank Drs. Matthew R. Levensgood, Rebecca A. Splain and Adam H. Courtney for experimental advice and feedback in preparing the manuscript. I would also like to acknowledge Dr. Jonathon Hudon for assistance with the collection of LCMS data.

2.11 References

1. Sasser, J. (1980) Root-knot nematodes - A global menace to crop production. *Plant. Dis.* **64**, 36-41
2. Kaplan, R. (2004) Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol.* **20**, 477-481
3. Hotez, P. J., and Kamath, A. (2009) Neglected Tropical Diseases in Sub-Saharan Africa: Review of Their Prevalence, Distribution, and Disease Burden. *PLoS Neglected Tropical Diseases* **3**, e412
4. Jasmer, D. P., Goverse, A., and Smart, G. (2003) Parasitic nematode interactions with mammals and plants. *Annu. Rev. Phytopathol.* **41**, 245-270
5. Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., and Savioli, L. (2007) Control of neglected tropical diseases. *N. Engl. J. Med.* **357**, 1018-1027
6. Maizels, R. M., Hewitson, J. P., and Grainger, J. R. (2009) Helminth immunoregulation: The role of parasite secreted proteins in modulating host immunity. *Mol Biochem Parasit* **167**, 1-11
7. Harnett, W., and Harnett, M. M. (2006) Molecular basis of worm-induced immunomodulation. *Parasite Immunol.* **28**, 535-543
8. Riddle, D. L. (1997) *C. elegans II*, Cold Spring Harbor Laboratory Press, Plainview, N.Y.
9. van Die, I., and Cummings, R. D. (2010) Glycan gimmickry by parasitic helminths: a strategy for modulating the host immune response? *Glycobiology* **20**, 2-12
10. Harn, D. A., McDonald, J., Atochina, O., and Da'dara, A. A. (2009) Modulation of host immune responses by helminth glycans. *Immunol. Rev.* **230**, 247-257
11. Davies, K. G., and Curtis, R. H. (2011) Cuticle surface coat of plant-parasitic nematodes. *Annu. Rev. Phytopathol.* **49**, 135-156
12. Mendonca-Previato, L., Todeschini, A. R., Heise, N., Agrellos, O. A., Dias, W. B., and Previato, J. O. (2008) Chemical structure of major glycoconjugates from parasites. *Curr Org Chem* **12**, 926-939
13. van Die, I., and Cummings, R. D. (2006) Glycans modulate immune responses in helminth infections and allergy. *Chem. Immunol. Allergy.* **90**, 91-112
14. Hewitson, J. P., Grainger, J. R., and Maizels, R. M. (2009) Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol Biochem Parasit* **167**, 1-11
15. Tawill, S., Le Goff, L., Ali, F., Blaxter, M., and Allen, J. E. (2004) Both free-living and parasitic nematodes induce a characteristic Th2 response that is dependent on the presence of intact glycans. *Infect. Immun.* **72**, 398-407
16. van Stijn, C. M., Meyer, S., van den Broek, M., Bruijns, S. C., van Kooyk, Y., Geyer, R., and van Die, I. (2010) *Schistosoma mansoni* worm glycolipids induce an inflammatory phenotype in human dendritic cells by cooperation of TLR4 and DC-SIGN. *Mol. Immunol.* **47**, 1544-1552
17. Astronomo, R. D., and Burton, D. R. (2010) Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nat. Rev. Drug Discovery* **9**, 308-324

18. Hein, W. R., and Harrison, G. B. (2005) Vaccines against veterinary helminths. *Vet. Parasitol.* **132**, 217-222
19. Osei-Atweneboano, M. Y., Eng, J. K. L., Boakye, D. A., Gyapong, J. O., and Prichard, R. K. (2007) Prevalence and intensity of *Onchocerca volvulus* infection and efficacy of ivermectin in endemic communities in Ghana: a two-phase epidemiological study. *Lancet* **369**, 2021-2029
20. Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S. S., Wenger, A., Wieland-Berghausen, S., Goebel, T., Gauvry, N., Pautrat, F., Skripsky, T., Froelich, O., Komoin-Oka, C., Westlund, B., Sluder, A., and Mäser, P. (2008) A new class of anthelmintics effective against drug-resistant nematodes. *Nature* **452**, 176-180
21. Beverley, S. M., Owens, K. L., Showalter, M., Griffith, C. L., Doering, T. L., Jones, V. C., and McNeil, M. R. (2005) Eukaryotic UDP-galactopyranose mutase (GLF gene) in microbial and metazoal pathogens. *Eukaryotic Cell* **4**, 1147-1154
22. Weston, A., Stern, R. J., Lee, R. E., Nassau, P. M., Monsey, D., Martin, S. L., Scherman, M. S., Besra, G. S., Duncan, K., and McNeil, M. R. (1997) Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. *Tuber. Lung Dis.* **78**, 123-131
23. Pedersen, L. L., and Turco, S. J. (2003) Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. *Cell. Mol. Life Sci.* **60**, 259-266
24. Richards, M. R., and Lowary, T. L. (2009) Chemistry and biology of galactofuranose-containing polysaccharides. *ChemBioChem* **10**, 1920-1938
25. Tefsen, B., Ram, A. F., van Die, I., and Routier, F. H. (2012) Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology* **22**, 456-469
26. Novelli, J. F., Chaudhary, K., Canovas, J., Benner, J. S., Madinger, C. L., Kelly, P., Hodgkin, J., and Carlow, C. K. S. (2009) Characterization of the *Caenorhabditis elegans* UDP-galactopyranose mutase homolog glf-1 reveals an essential role for galactofuranose metabolism in nematode surface coat synthesis. *Dev. Biol.* **335**, 340-355
27. Paschinger, K., Gutternigg, M., Rendic, D., and Wilson, I. B. (2008) The N-glycosylation pattern of *Caenorhabditis elegans*. *Carbohydr. Res.* **343**, 2041-2049
28. Pan, F., Jackson, M., Ma, Y., and McNeil, M. (2001) Cell wall core galactofuran synthesis is essential for growth of mycobacteria. *J. Bacteriol.* **183**, 3991-3998
29. Dykhuizen, E. C., May, J. F., Tongpenyai, A., and Kiessling, L. L. (2008) Inhibitors of UDP-galactopyranose mutase thwart mycobacterial growth. *J. Am. Chem. Soc.* **130**, 6706-6707
30. Schmalhorst, P. S., Krappmann, S., Vervecken, W., Rohde, M., Müller, M., Braus, G. H., Contreras, R., Braun, A., Bakker, H., and Routier, F. H. (2008) Contribution of galactofuranose to the virulence of the opportunistic pathogen *Aspergillus fumigatus*. *Eukaryotic Cell* **7**, 1268-1277
31. Spath, G., Epstein, L., Leader, B., Singer, S., Avila, H., Turco, S., and Beverley, S. (2000) Lipophosphoglycan is a virulence factor distinct from related glycoconjugates in the protozoan parasite *Leishmania major*. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9258-9263
32. Spath, G., Garraway, L., Turco, S., and Beverley, S. (2003) The role(s) of lipophosphoglycan (LPG) in the establishment of *Leishmania major* infections in mammalian hosts. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9536-9541

33. Bakker, H., Kleczka, B., Gerardy-Schahn, R., and Routier, F. H. (2005) Identification and partial characterization of two eukaryotic UDP-galactopyranose mutases. *Biol. Chem.* **386**, 657-661
34. Soltero-Higgin, M., Carlson, E. E., Gruber, T. D., and Kiessling, L. L. (2004) A unique catalytic mechanism for UDP-galactopyranose mutase. *Nat. Struct. Mol. Biol.* **11**, 539-543
35. Gruber, T. D., Westler, W. M., Kiessling, L. L., and Forest, K. T. (2009) X-ray crystallography reveals a reduced substrate complex of UDP-galactopyranose mutase poised for covalent catalysis by flavin. *Biochemistry* **48**, 9171-9173
36. Oppenheimer, M., Valenciano, A. L., Kizjakina, K., Qi, J., and Sobrado, P. (2012) Chemical Mechanism of UDP-Galactopyranose Mutase from Trypanosoma cruzi: A Potential Drug Target against Chagas' Disease. *PLoS One* **7**, e32918
37. Soltero-Higgin, M., Carlson, E. E., Phillips, J. H., and Kiessling, L. L. (2004) Identification of inhibitors for UDP-galactopyranose mutase. *J. Am. Chem. Soc.* **126**, 10532-10533
38. Lee, R., Monsey, D., Weston, A., Duncan, K., Rithner, C., and McNeil, M. (1996) Enzymatic synthesis of UDP-galactofuranose and an assay for UDP-galactopyranose mutase based on high-performance liquid chromatography. *Anal. Biochem.* **242**, 1-7
39. Zhang, Q., and Liu, H. (2000) Studies of UDP-galactopyranose mutase from Escherichia coli: An unusual role of reduced FAD in its catalysis. *J. Am. Chem. Soc.* **122**, 9065-9070
40. Marlow, A. L., and Kiessling, L. L. (2001) Improved chemical synthesis of UDP-galactofuranose. *Org. Lett.* **3**, 2517-2519
41. Sanders, D. A., Staines, A. G., McMahon, S. A., McNeil, M. R., Whitfield, C., and Naismith, J. H. (2001) UDP-galactopyranose mutase has a novel structure and mechanism. *Nat. Struct. Mol. Biol.* **8**, 858-863
42. Oppenheimer, M., Valenciano, A. L., and Sobrado, P. (2011) Isolation and characterization of functional Leishmania major virulence factor UDP-galactopyranose mutase. *Biochem. Biophys. Res. Commun.* **407**, 552-556
43. Oppenheimer, M., Poulin, M. B., Lowary, T. L., Helm, R. F., and Sobrado, P. (2010) Characterization of recombinant UDP-galactopyranose mutase from Aspergillus fumigatus. *Arch. Biochem. Biophys.* **502**, 31-38
44. Chad, J. M., Sarathy, K. P., Gruber, T. D., Addala, E., Kiessling, L. L., and Sanders, D. A. R. (2007) Site-directed mutagenesis of UDP-Galactopyranose mutase reveals a critical role for the active-site, conserved arginine residues. *Biochemistry* **46**, 6723-6732
45. Huang, Z., Zhang, Q., and Liu, H. (2003) Reconstitution of UDP-galactopyranose mutase with 1-deaza-FAD and 5-deaza-FAD: analysis and mechanistic implications. *Bioorg. Chem.* **31**, 494-502
46. Sun, H. G., Rusczycky, M. W., Chang, W. C., Thibodeaux, C. J., and Liu, H. W. (2012) Nucleophilic participation of reduced flavin coenzyme in mechanism of UDP-galactopyranose mutase. *J. Biol. Chem.* **287**, 4602-4608
47. Huang, W. J., and Gauld, J. W. (2012) Tautomerization in the UDP-Galactopyranose Mutase Mechanism: A DFT-Cluster and QM/MM Investigation. *J. Phys. Chem. B* **116**, 14040-14050

48. Ghisla, S., and Massey, V. (1989) Mechanisms of flavoprotein-catalyzed reactions. *Eur. J. Biochem.* **181**, 1-17
49. Ansiaux, C., N'go, I., and Vincent, S. P. (2012) Reversible and Efficient Inhibition of UDP-Galactopyranose Mutase by Electrophilic, Constrained and Unsaturated UDP-Galactitol Analogues. *Chem.--Eur. J.* **18**, 14860-14866
50. Partha, S. K., van Straaten, K. E., and Sanders, D. A. R. (2009) Structural basis of substrate binding to UDP-galactopyranose mutase: crystal structures in the reduced and oxidized state complexed with UDP-galactopyranose and UDP. *J. Mol. Biol.* **394**, 864-877
51. Yuan, Y., Wen, X., Sanders, D. A., and Pinto, B. M. (2005) Exploring the mechanism of binding of UDP-galactopyranose to UDP-galactopyranose mutase by STD-NMR spectroscopy and molecular modeling. *Biochemistry* **44**, 14080-14089
52. Gruber, T. D., Borrok, M. J., Westler, W. M., Forest, K. T., and Kiessling, L. L. (2009) Ligand binding and substrate discrimination by UDP-galactopyranose mutase. *J. Mol. Biol.* **391**, 327-340
53. van Straaten, K. E., Routier, F. H., and Sanders, D. A. (2012) Structural insight into the unique substrate binding mechanism and flavin redox state of UDP-galactopyranose mutase from *Aspergillus fumigatus*. *J. Biol. Chem.* **287**, 10780-10790
54. Dhatwalia, R., Singh, H., Oppenheimer, M., Karr, D. B., Nix, J. C., Sobrado, P., and Tanner, J. J. (2012) Crystal Structures and Small-angle X-ray Scattering Analysis of UDP-galactopyranose Mutase from the Pathogenic Fungus *Aspergillus fumigatus*. *J. Biol. Chem.* **287**, 9041-9051
55. Dhatwalia, R., Singh, H., Solano, L. M., Oppenheimer, M., Robinson, R. M., Ellerbrock, J. F., Sobrado, P., and Tanner, J. J. (2012) Identification of the NAD(P)H Binding Site of Eukaryotic UDP-Galactopyranose Mutase. *J. Am. Chem. Soc.* **134**, 18132-18138
56. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201
57. Beis, K., Srikannathan, V., Liu, H., Fullerton, S. W. B., Bamford, V. A., Sanders, D. A. R., Whitfield, C., McNeil, M. R., and Naismith, J. H. (2005) Crystal structures of *Mycobacteria tuberculosis* and *Klebsiella pneumoniae* UDP-galactopyranose mutase in the oxidised state and *Klebsiella pneumoniae* UDP-galactopyranose mutase in the (active) reduced state. *J. Mol. Biol.* **348**, 971-982
58. Cheng, Y., and Prusoff, W. (1973) Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099-3108
59. Burglin, T., Lobos, E., and Blaxter, M. (1998) *Caenorhabditis elegans* as a model for parasitic nematodes. *Int. J. Parasitol.* **28**, 395-411
60. Politz, S. M., and Philipp, M. (1992) *Caenorhabditis elegans* as a model for parasitic nematodes: a focus on the cuticle. *Parasitol. Today* **8**, 6-12
61. Jones, A. K., Buckingham, S. D., and Sattelle, D. B. (2005) Chemistry-to-gene screens in *Caenorhabditis elegans*. *Nat. Rev. Drug Discovery* **4**, 321-330

62. Gerdt, S., Lochnit, G., Dennis, R. D., and Geyer, R. (1997) Isolation and structural analysis of three neutral glycosphingolipids from a mixed population of *Caenorhabditis elegans* (Nematoda:Rhabditida). *Glycobiology* **7**, 265-275
63. Griffiths, J. S., Whitacre, J. L., Stevens, D. E., and Aroian, R. V. (2001) Bt toxin resistance from loss of a putative carbohydrate-modifying enzyme. *Science* **293**, 860-864
64. Griffiths, J. S., Haslam, S. M., Yang, T. L., Garczynski, S. F., Mulloy, B., Morris, H., Cremer, P. S., Dell, A., Adang, M. J., and Aroian, R. V. (2005) Glycolipids as receptors for *Bacillus thuringiensis* crystal toxin. *Science* **307**, 922-925
65. Cipollo, J. F., Costello, C. E., and Hirschberg, C. B. (2002) The fine structure of *Caenorhabditis elegans* N-glycans. *J. Biol. Chem.* **277**, 49143-49157
66. Cipollo, J. F., Awad, A. M., Costello, C. E., and Hirschberg, C. B. (2005) N-Glycans of *Caenorhabditis elegans* are specific to developmental stages. *J. Biol. Chem.* **280**, 26063-26072
67. Bennuru, S., Meng, Z. J., Ribeiro, J. M. C., Semnani, R. T., Ghedin, E., Chan, K., Lucas, D. A., Veenstra, T. D., and Nutman, T. B. (2011) Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9649-9654
68. Latge, J. P. (2009) Galactofuranose containing molecules in *Aspergillus fumigatus*. *Med. Mycol.* **47**, S104-S109

Chapter 3

Recognition of Microbial Glycans by Human Intelectin-1

Portions of this work have been published in:

Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zharling, L. C., Splain, R. S., Smith D. F., Cummings, R. D., Paulson, J. C., Forest, K. T., and Kiessling, L. L. Recognition of microbial glycans by human intelectin-1. *Nature Structural and Molecular Biology*. **2015**. 22, 603-610.

3.1 Abstract

The glycans displayed on mammalian cells can differ markedly from those on microbes. Such differences could, in principle, be read by carbohydrate-binding proteins, or lectins. We used glycan microarrays to show that human intelectin-1 (hIntL-1) does not bind known human glycan epitopes but interacts with multiple glycan epitopes found exclusively on microbes: β -linked D-galactofuranose (β -Gal f), D-phospho-glycerol-modified glycans, heptoses, D-*glycero*-D-*talo*-oct-2-ulosonic acid (KO) and 3-deoxy-D-*manno*-oct-2-ulosonic acid (KDO). The 1.6 Å resolution crystal structure of hIntL-1 bound to β -Gal f revealed that hIntL-1 uses a bound calcium ion to coordinate terminal exocyclic 1,2-diols. N-Acetylneuraminic acid (Neu5Ac), a sialic acid widespread in human glycans, possesses an exocyclic 1,2-diol but does not bind hIntL-1, likely due to unfavorable steric and electronic effects. Human IntL-1 marks only *Streptococcus pneumoniae* serotypes that display surface glycans with terminal 1,2-diol groups. This ligand selectivity suggests hIntL-1 functions in microbial surveillance.

3.2 Introduction

Organisms that serve as hosts for microbes must distinguish microbial cells from those of their own (1,2). A mechanism of differentiation is especially important at sites in which host tissues contact the environment, such as in the lung, intestine, and skin (3,4). Differences in cellular surface glycosylation can serve as markers of a cell's identity—its developmental state, its tissue type, or whether it is self or non-self (5). Cell surface glycans can be distinguished by carbohydrate binding proteins or lectins (6), which are typically categorized based on their monosaccharide selectivity (7). These lectins can be exploited for host defense, as in the case of innate immune lectins, such as MBL (8). In the serum, MBL is precomplexed with MASPs, and interaction of this complex with a cell surface results in activation of the lectin pathway of complement, ultimately leading to pathogen opsonization and clearance (9,10). Other humoral lectins implicated in immunity include ficolins, collectins, galectins, and HIP/PAP (1,11-13).

One group of lectins whose specificity remains unclear is that composed of IntLs. The first IntL protein was reported in *Xenopus laevis* oocytes (14). Homologs have since been identified in many other chordates; including other amphibians, fishes, and many mammals. IntLs belong to a family of lectins termed X-type lectins (15) and have been shown to exist as homooligomers of 35 kDa monomers. They are reported to function as calcium ion-dependent lectins; however, they do not contain the calcium-dependent C-type lectin sequence motif (16) present in many human lectins. IntLs instead contain a fibrinogen like domain (FBD, residues 37-82 in hIntL-1 ((ref. 17)) and are proposed to be most similar to ficolins, a class of FBD-containing innate immune lectins (11).

Several observations implicate IntLs in innate immunity. Mammalian IntLs are predominantly produced by lung and intestinal goblet cells, and intestinal paneth cells (17-19). In sheep and mice, IntL expression increases upon infection with intestinal parasitic nematodes (20,21). In humans, the mucus induced by allergic reactions is enriched in IntLs (22,23). Still, hIntL-1 has been reported to be the intestinal lactoferrin receptor (24), to function as a tumor marker (25). It also be suggested to be involved in metabolic disorders including diabetes, where it is known as omentin (26). Given these diverse potential functions, we set out to examine the ligand specificity of hIntL-1.

Human IntL-1 has been reported to bind furanose residues (5-membered ring saccharide isomers), including ribofuranose (Ribf) and a β -Gal f -containing disaccharide (17,27). The monosaccharide Gal f is present in the cell surface glycans produced by a number of microbes, but the biosynthetic enzymes that mediate Gal f incorporation are absent in humans (28-30). The presence of Gal f in microbial glycans but not in those of humans is an example of phylogenetic glycan differences (31). This is just one example, as collectively the surface glycans of microbes are generated from more than 700 unique building blocks, while less than 35 carbohydrate residues are needed to assemble mammalian glycans (32,33). In principle, targeting monosaccharide residues unique to microbes could be used by the innate immune system to differentiate mammalian cells from microbes.

We reasoned that clues to hIntL-1 function would emerge from determining the glycans it binds and the molecular basis for its recognition selectivity. Here, we use glycan microarrays to demonstrate that hIntL-1 binds microbial over human glycans. Given the diversity of microbial glycans, a lectin that binds a single microbial saccharide epitope (e.g., Gal f) would be expected

to have specialized function. It is therefore striking that hIntL-1 does not engage a single monosaccharide or even related saccharides; rather, hIntL-1 interacts with multiple, structurally divergent microbial monosaccharide residues. The molecular mechanism by which hIntL-1 recognizes its targets was revealed by X-ray crystallography: hIntL-1 binds its carbohydrate ligands through calcium ion-dependent coordination of a conserved exocyclic, terminal 1,2-diol. The functional group selectivity observed in the glycan arrays is manifested in the context of cells, as hIntL-1 targets *S. pneumoniae* serotypes that display its glycan ligands.

3.3 hIntL-1 Binds β -Gal f

Native hIntL-1 has been shown to exist as a disulfide-linked trimer (17,27). Therefore, we first developed a robust expression system that yields the protein as a disulfide-linked trimer that could be purified using an immobilized β -Gal f column (**Figures 3-1A & 3-1B**). Because lectin–carbohydrate interactions often depend on multivalent binding (34,35) we postulated that hIntL-1 trimers might bind avidly to multivalent carbohydrate displays. Human IntL-1 carbohydrate-binding specificity hence was evaluated using immobilized biotinylated carbohydrates (β -D-Gal f , β -D-Gal p , β -D-ribofuranose (β -Rib f)) in an enzyme-linked immunoabsorbent-like assay (ELISA) (**Figures 3-1C—E**). The monosaccharide binding epitopes we tested were chosen based upon a previous study in which a small carbohydrate panel was evaluated for inhibition of hIntL-1 binding to an immobilized carbohydrate (17). In those studies, ribose was the most effective competitor ($IC_{50} < 5$ mM) followed by Gal f - β (1,4)-GlcNAc ($IC_{50} = 9$ mM) with galactose being less potent ($IC_{50} = 66$ mM) (17). Our data indicate that hIntL-1 does not bind Rib f nor Gal p , but it does engage the β -Gal f -substituted surface avidly with a functional affinity (apparent affinity) of 85 ± 14 nM (**Figure 3-1F**).

Our results contrast with those of the previous study (17), as we did not detect binding to the pyranose form of galactose nor to ribofuranose. The apparent discrepancies could arise because the previous investigation required high concentrations free carbohydrate. Under those conditions, competition could arise from protein modification or from the less prevalent open chain form of the saccharide. The apparent binding constant we observed for hIntL-1 binding to immobilized β -D-Galf suggests that the protein binds tightly to a ligand, but the previous IC_{50} for the β -D-Galf-containing disaccharide (9 mM) suggests the interaction is weak. This difference presumably stems from the distinct assay formats. We postulated that the presentation of glycosides from a surface is a more relevant assessment of hIntL-1 activity as it mimics key aspects of the multivalent display of carbohydrate ligands on a cell surface (34). Still, the differences between the reported hIntL-1 binding specificities and those we observed prompted us to examine hIntL-1 binding using another assay. We used surface plasmon resonance (SPR) and monitored hIntL-1 interaction with surfaces to which the aforementioned saccharides or β -D-Araf or α -L-rhamnopyranose (α -L-Rha) were appended. Even at high concentrations of hIntL-1, we observed only selective hIntL-1 binding to β -Galf (**Figure 3-1G**).

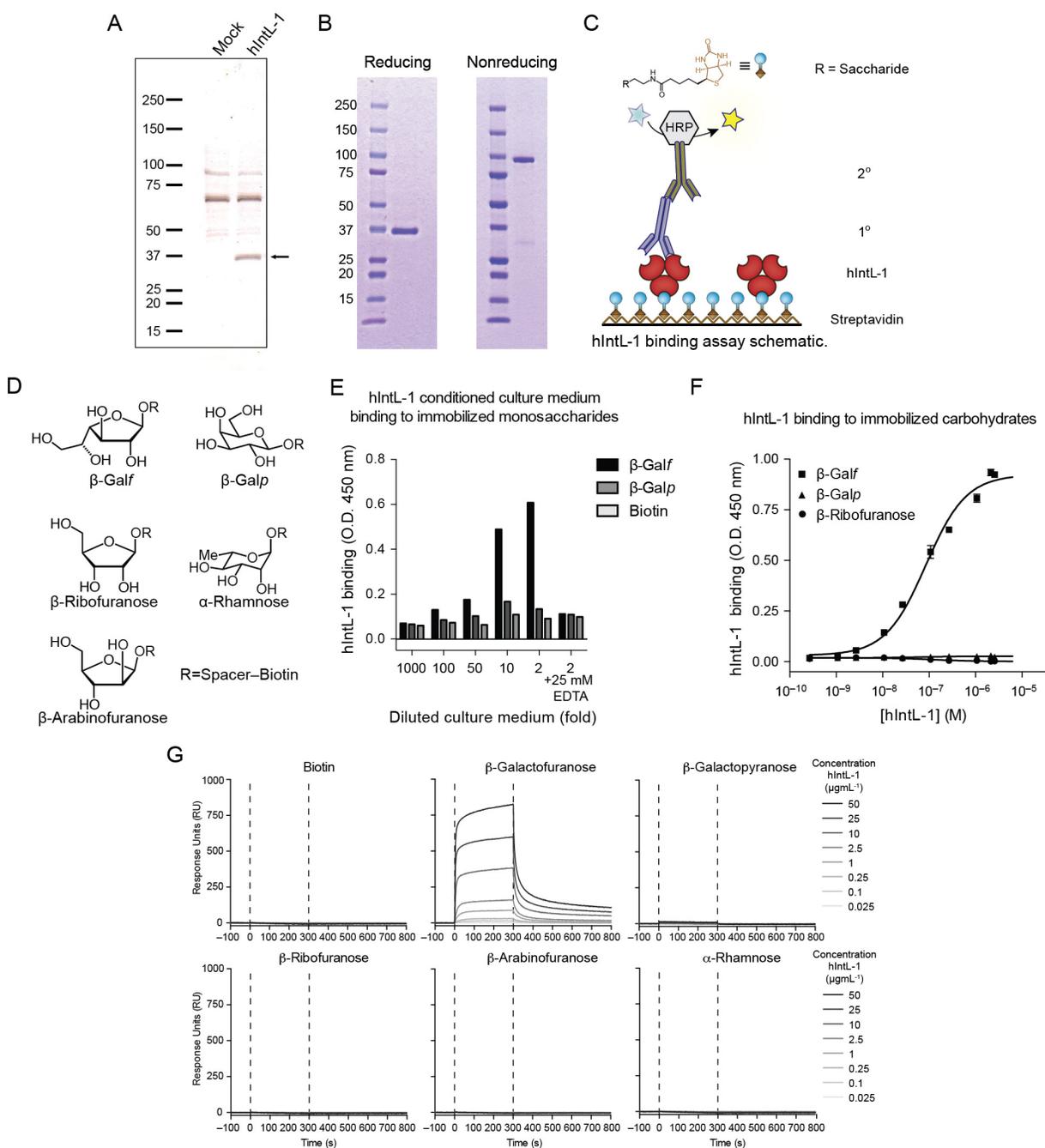


Figure 3-1. Expression, purification, and carbohydrate binding activity of hIntL-1. (A) Reducing SDS-PAGE analysis of HEK 293T culture medium from hIntL-1 transfected cells. Samples were analyzed by silver stain 48 hours post transfection. An arrow indicates the band corresponding to the molecular weight of a hIntL-1 reduced monomer. (B) Coomassie stained gels of samples subjected to reducing and nonreducing SDS-PAGE analysis of hIntL-1 purified on an

immobilized β -Gal f column. The molecular weight of the sample analyzed under non-reducing conditions corresponds to that of a disulfide-linked hIntL-1 homotrimer. (C) Schematic of streptavidin-based, ELISA-like carbohydrate binding assay developed for assessing hIntL-1 ligand specificity. Biotin-functionalized carbohydrate is immobilized. Bound hIntL-1 is detected the enzyme horseradish peroxidase (HRP) conjugated to an antibody (either a secondary or directly conjugated primary), and a chromogenic HRP substrate. (D) Carbohydrate-binding activity of HEK 293T cell conditioned culture medium following transfection with hIntL-1 expression plasmid. The calcium ion dependence was assayed with addition of 25 mM ethylenediaminetetraacetic acid (EDTA). Data are presented as the mean ($n=2$ of a technical replicate and is representative of >3 independent experiments). (E) Structures of saccharides used for characterization of hIntL-1 by ELISA and SPR. (F) The specificity of hIntL-1 for β -Gal f , β -Rib f and β -Gal p evaluated by ELISA (See Supplementary Fig. 1b for schematic). Data are presented as the mean \pm s.d. ($n=3$ technical replicates, data are representative of >3 independent experiments). Data were fit to a single site binding equation (solid lines) and therefore represent the apparent (App) affinity of trimeric hIntL-1. Values for hIntL-1 bound to immobilized β -Gal f ($K_{d(\text{App, Trimer})} \pm$ s.d.) are 85 ± 14 nM or 8.0 ± 1.3 $\mu\text{g/mL}$. (G) Representative real-time SPR sensorgrams of hIntL-1 binding to immobilized carbohydrates. Biotin served as a control.

3.4 hIntL-1 Binding to Microbial Glycans

Glycan microarray technology afforded a more comprehensive assessment of hIntL-1 ligand recognition (36). We prepared a focused array that included furanosides using the methods employed in generating the Consortium for Functional Glycomics (CFG; <http://www.functionalglycomics.org>) mammalian glycan v5.1 array, and both arrays were tested for hIntL-1 binding. In the focused array, lacto-N-neotetraose (LNnT) and asialo, galactosylated bi-antennary N-linked glycan (NA2) were included to ascertain the efficiency of carbohydrate immobilization. Data from the focused array were consistent with those obtained from the ELISA and SPR assays, indicating that of the carbohydrates displayed, hIntL-1 bound only to those with β -Gal f residues (**Figure 3-2A**). We attribute the small amount of binding to β -Gal p to its hydrophobic, alkyl anomeric linker. In contrast to the furanoside array, testing of the CFG v5.1 array yielded no validated interactions with mammalian glycans (**Figure 3-2A**). Increasing the protein concentration yielded similarly low signals suggesting the modest residual binding

detected arose from nonspecific interactions (Data not shown). Thus, none of the human glycans examined are ligands of hIntL-1.

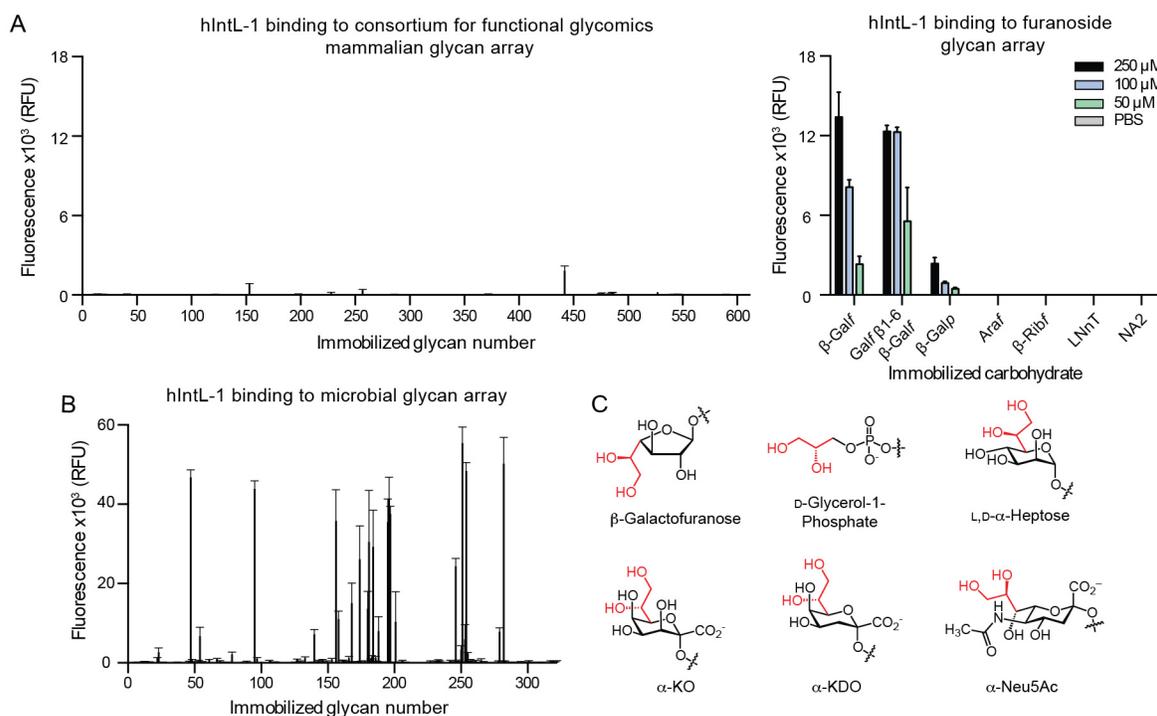


Figure 3-2. Glycan selectivity of hIntL-1 assessed by glycan microarrays. (A) Recombinant hIntL-1 (50 μ g/mL) binding to mammalian glycan microarray CFG v5.1 and a furanoside array. The concentrations given for the furanoside array represent those used in the carbohydrate immobilization reaction. Data are presented as the mean \pm s.d. ($n=4$ technical replicates). The full data set can be found in Supplementary Table 1 of reference (37). (B) Recombinant *Strep*-hIntL-1 (50 μ g/mL) binding to microbial glycan array. For glycan array data organized by genus, see **Figure 3-3A**. Data are presented as the mean \pm s.d. ($n=4$ technical replicates). The full data set can be found in Supplementary Table 2 of reference (37). (C) Structural representation of the putative key binding epitopes for hIntL-1 and the non-binding α -Neu5Ac. A terminal vicinal diol (red) is a common feature of α -Neu5Ac and all of the ligands identified.

The initial binding data revealing hIntL-1 robustly complexes β -Galnf residues but not human glycans prompted us to evaluate the lectin's specificity for a more diverse collection of microbial glycans. Though absent from mammals (28), Galnf residues occur in glycans from a number of human pathogens, including the bacteria *Mycobacterium tuberculosis* and *Klebsiella*

pneumoniae, and the fungal pathogen *Aspergillus fumigates* (29,38). The possibility that hIntL-1 interacts with microbial glycans was tested using a microarray displaying more than 300 oligosaccharides from bacterial species (39). Screening of this array revealed multiple glycan ligands for hIntL-1 (**Figure 3-2B and 3-3A**). These ligands encompassed glycans from Gram-negative and Gram-positive bacteria; including *S. pneumoniae*, *Proteus mirabilis*, *Proteus vulgaris*, *Yersinia pestis*, and *K. pneumoniae* (**Table 3-1**). Four of the top-fifteen ligands contained terminal β -Gal f epitopes, including outer polysaccharide (OPS) from *K. pneumoniae* and a CPS from *S. pneumoniae*. Surprisingly, the majority of the glycans identified did not possess Gal f residues. The top five hits had saccharide residues with Gro-P substituents. This epitope was the common feature, as the residue to which it was appended varied between glycans. Other common epitopes included either D/L-*manno*-heptose, KO, or KDO residues (**Figure 3-2C**). Each characterized glycan ligand from the top 15 hits contains at least one of the five aforementioned epitopes. Despite its ability to bind structurally diverse glycans, hIntL-1 exhibited selectivity. Conspicuously missing from hit microbial glycan ligands were those containing α -Gal f residues (**Figure 3-3B**). What was especially notable, however, was that none of the hIntL-1 ligands we identified on the microbial glycan array are found in mammalian glycans, but collectively these five residues are widely distributed in bacteria (32).

Table 3-1. The List of the Top 15 Microbial Glycan Ligands, Sorted by Average Fluorescence Intensity (RFU).

Rank	Microbial sample	Proposed ligand
1	<i>S. pneumoniae</i> type 43	Glycerol-phosphate
2	<i>P. mirabilis</i> O54ab	Glycerol-phosphate
3	<i>S. pneumoniae</i> type 56	Glycerol-phosphate
4	<i>P. mirabilis</i> O54a, 54b	Glycerol-phosphate
5	<i>P. vulgaris</i> O54a, 54c	Glycerol-phosphate
6	<i>K. pneumoniae</i> O2a OPS	β -Gal f
7	<i>K. pneumoniae</i> O2ac OPS	β -Gal f
8	<i>Y. pestis</i> KM260(11)- Δ O187 ^a	
9	<i>K. pneumoniae</i> O1 OPS	β -Gal f
10	<i>Y. pestis</i> 11M-37	Heptose, KO, KDO
11	<i>Y. pestis</i> KM260(11)-6C ^a	
12	<i>Y. pestis</i> KM260(11)- Δ waal	Heptose, KO, KDO
13	<i>S. pneumoniae</i> type 20	β -Gal f
14	<i>Y. pestis</i> KM260(11)- Δ pmrF	Heptose, KO, KDO
15	<i>Y. pestis</i> 11M-25	Heptose, KO, KDO

^a These glycans are currently structurally uncharacterized.

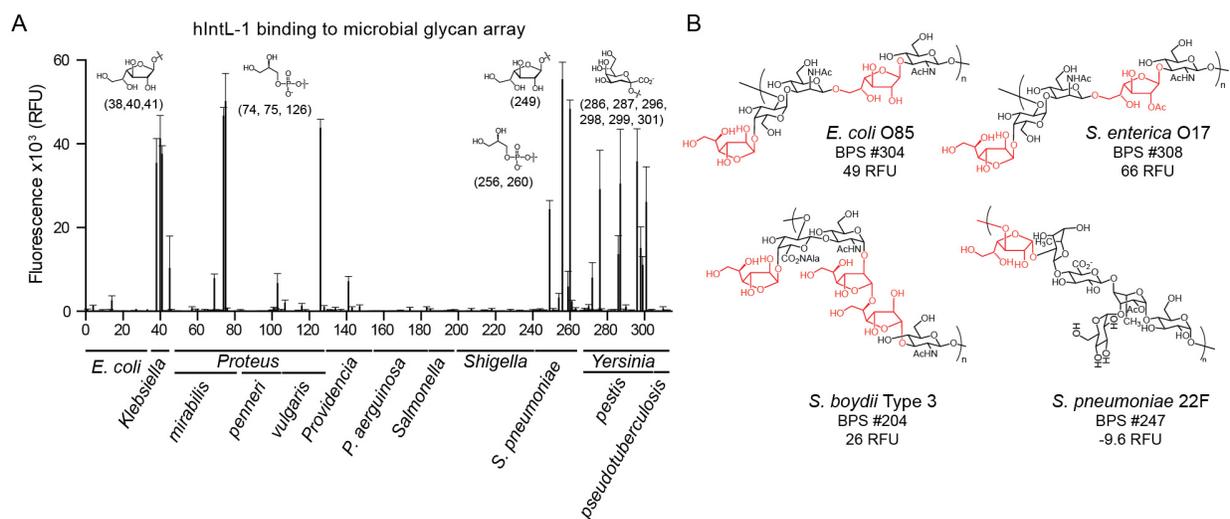


Figure 3-3. Human IntL-1 binding specificity as determined from the microbial glycan microarray (MGMv2). (A) Results of the MGM organized by genus and species, alphabetically. The fluorescence values are identical to those presented in **Figure 3-2B**. The chemical epitope that is proposed to be a hIntL-1 ligand is depicted. The chart identification number from this graph is provided in parenthesis below the graphically depicted ligand. Data are presented as the mean \pm s.d. ($n=4$ of a technical replicate for each immobilized glycan). The complete data for this experiment are available in Supplementary Table 3 of reference (37). (B) Chemical structures of terminal α -Gal f containing glycans that failed to bind hIntL-1. The Gal f residues in each glycan are depicted in red. The BPS number (BPS #) that references each glycan is derived

from reference (39), and the hIntL-1 signal (from **Figure 3-2B**) are shown.

3.5 Structure of hIntL-1

To understand the molecular mechanisms underlying glycan recognition by hIntL-1, we determined its structure using X-ray crystallography. Apo-hIntL-1 crystals diffracted to 1.8 Å resolution, and the structure of the protein was solved by molecular replacement using the structure of a selenomethione (SeMet)-labeled *Xenopus laevis* IntL as a search model (**Table 3-2**) (PDB ID: 4WMO). Human IntL-1 possesses an oblong, globular structure containing two highly twisted β sheet-containing structures surrounded by seven short α helices and extensive random coil regions (**Figure 3-4A**). The second of these β sheets structures closes on itself to form a very short stretch of unusually flattened β-ribbons (amino acids 221-226 + 248-278). A Dali search (40) using the hIntL-1 structure yielded several weak FBG and ficolin structure hits (RMSD values ~4 Å). The secondary structures of L-ficolin (41) and hIntL-1 are related up to residue 150, although the sequence conservation is limited to the FBD. The remaining residues diverge substantially in sequence and structure (**Figure 3-5**). Indeed, removal of the first 150 residues from the hIntL-1 Dali input yielded no hits. These data indicate hIntL-1 has a composite fold not previously reported.

Two hIntL-1 monomers are present in the asymmetric unit (Chain A and Chain B), and they represent two similar, though non-identical ($C\alpha$ RMSD=0.65 Å), disulfide-linked trimers, each arranged around a crystallographic threefold axis. In one trimer, the peptide chain that connects each monomer to the adjacent monomer is resolved, such that the intermolecular disulfide bond between residues C31 and C48 is apparent (**Figure 3-4A**). These data are consistent with SDS-PAGE analysis indicating the hIntL-1 exists as a trimer. Each hIntL-1

monomer has three calcium ions, and each cation is chelated exclusively by hard protein or water ligands (bond distance 2.3–2.5 Å). Two of these cations are embedded within the protein while one is surface exposed.

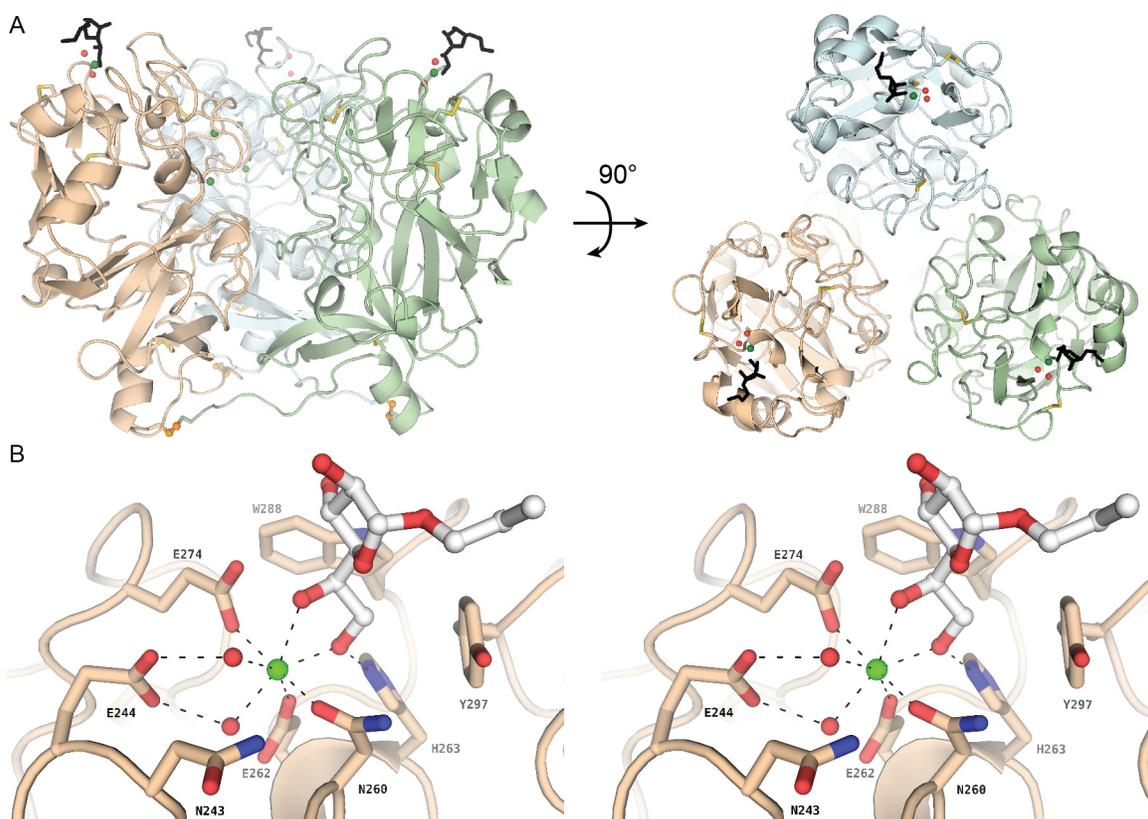


Figure 3-4. Structure of hIntL-1 bound to allyl- β -D-Galf. (A) Complex of hIntL-1 disulfide-linked trimer and allyl- β -D-Galf. Each monomer unit is depicted in green, wheat, or grey, the β -allyl Galf is shown in black, calcium ions in green, the inter-monomer disulfides in orange, and ordered water molecules in the binding site in red. The two orientations indicate the positioning of all three ligand-binding sites within the trimer. The trimeric structure is produced from Chain A in the asymmetric unit by a three-fold crystallographic operation. (B) Stereo image of the carbohydrate-binding site. Residues involved in calcium coordination and ligand binding are noted. Dashed lines are included to show the heptavalent coordination of the calcium ion and to highlight functional groups important for ligand and calcium ion binding. Difference density map (Fo-Fc, 3σ) of the allyl- β -D-Galf ligand is provided in **Figure 3-6A**.

To determine how hIntL-1 binds its ligands, we solved a structure of the complex of allyl- β -D-Galf bound to hIntL-1 to 1.6 Å resolution. The C α RMSD between the asymmetric unit of apo- and Galf-bound structures (0.118 Å) suggested no significant structural changes occur

upon ligand binding. The Gal f O(5) and O(6) hydroxyl groups displace ordered water molecules and serve as coordinating ligands for the surface accessible calcium ion, with protein side chains poised for hydrogen bonding, (i.e., H263 to the Gal f O(6) hydroxyl group, **Figure 3-4B & Figure 3-6A**) thereby enhancing calcium coordination. The carbohydrate vicinal exocyclic hydroxyl groups adopt a gauche conformation as they chelate the calcium, with dihedral angles of 45° and 51° for Chain A and B, respectively. As anticipated from the structure, glycans containing Gal f residues with substituents at either the O(5) or O(6) fail to bind hIntL-1 (**Figure 3-4B**). This portion of the saccharide also fits well into a binding pocket formed by W288 and Y297. The presence of these aromatic groups suggests that CH- π bonds may contribute to affinity.

The high resolution of the structure of the hIntL-1 complex allows unambiguous assignment of the β -Gal f ring conformation in each monomer (**Figure 3-6B**) (42,43). Using the Altona-Sundaralingam pseudorotational model, we calculated the pseudorotational phase angle, P , of each furanoside to assign its conformation (44). In hIntL-1 Molecule A, the furanoside is in the 1T_O -gg-gt (calculated $P=105^\circ$) conformer, while the β -Gal f shown in **Figure 3-4b** adopts the 4E -gg-gt (calculated $P=57^\circ$) conformation (**Figure 3-6C and 3-6D**). The presence of conformational differences within the structures are consistent with the flexibility of furanosides (43).

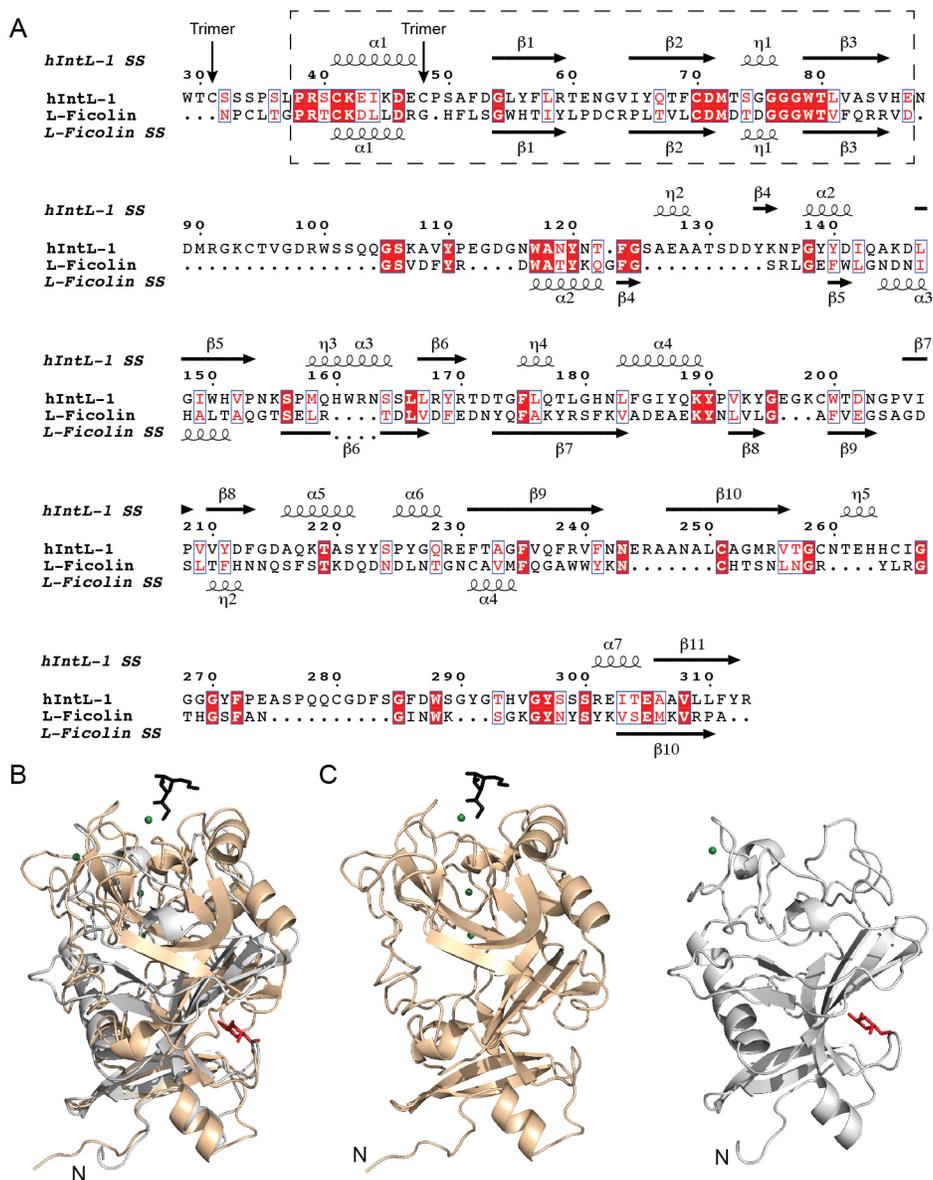


Figure 3-5. Structural alignment of hIntL-1 and human L-ficolin (PDB 2J3U) (41). (A) Primary protein sequence and secondary structure comparison of hIntL-1 and L-ficolin (PDB: 2J3U) generated using ESPrnt 3.0 (ref. (45)). The figure was produced from a Clustal W alignment of hIntL-1 (residues 29-313) and L-ficolin (Residues 96-313). The residues depicted correspond to those that were resolvable in each protein structure. This alignment omits the collagen-like domain of L-ficolin. The box denotes the proposed FBD of each molecule. A red box highlights identical residues. The cysteine residues from hIntL-1 that are involved in intermolecular trimerization are identified with an arrow. (B) A hIntL-1 monomer (wheat) aligned to a L-ficolin monomer (PDB: 2J3U) (grey) using Gesamt v6.4 (ref. (46)). Reported RMSD=3.6 Å for 165 superimposable C α atoms between the two structures. After the first 165 C α atoms, the structures

are too divergent to assign C α atoms as superimposable, and they are not included in this calculation. The co-crystallized carbohydrate ligands are depicted to highlight differences in ligand binding sites. The hIntL-1 ligand is shown in black and the L-ficolin ligand is shown in red. Calcium ions are shown in green. Human IntL-1 binds three calcium ions, while L-ficolin binds one. The N-termini are highlighted with an N. (C) The alignment shown in panel B, except that L-ficolin is translated by 45 Å for clarity. The N-terminus of each monomer is denoted with an N.

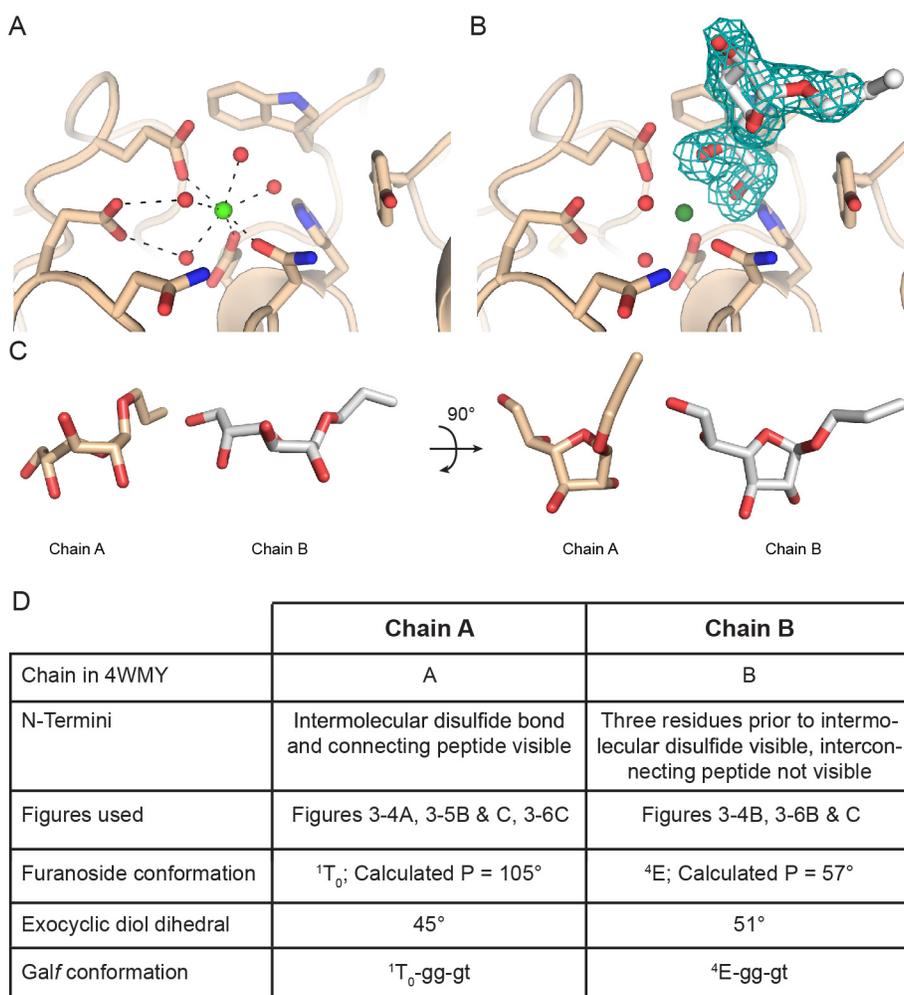


Figure 3-6. hIntL-1 bound to allyl- β -D-Galf. (A) Structure of the ligand-binding site in Apo-hIntL-1 (4WMQ). Calcium ions are shown in green, and ordered water molecules in red. Dashed lines highlight functional groups important for the heptavalent coordination of the ligand binding site calcium ion. (B) Close-up view of the ligand-binding site of the β -Galf-hIntL-1 protein structure (4WMY). This image is the same as depicted in **Figure 3-4B**, although surface mesh is depicted around the β -Galf ligand to highlight the ligand electron density. Mesh represents an

difference density map (mFo-DFc, 3σ). Calcium ions are depicted in green and ordered waters are shown in red. The ligand O(5) and O(6) hydroxyl groups coordinate to the calcium ion and displace two ordered water molecules. (C) Structural comparison of the crystallized allyl- β -D-Galf ligands. The molecule from Chain A is shown in wheat, while the molecule shown in Chain B is shown in grey. The furanosides were overlaid using the C(2)-C(3) bond and translated apart by 8 Å. (D) Table summarizing Chain A and Chain B in the β -Galf-hIntL-1 protein structure (4WMY).

3.6 Structural Basis for hIntL-1 Selectivity

The structure of the lectin-Galf complex reveals why the acyclic 1,2-diol moiety is critical – the vicinal hydroxyl groups engage in calcium ion coordination. Still, other glycan properties contribute to hIntL-1 recognition. For example, hIntL-1 does not bind α -Galf substituted glycans (**Figure 3-3B**). A cursory assessment of the β -Galf complex suggests hIntL-1 might accommodate α -Galf linkages. An alteration in anomeric configuration for furanosides, however, can drastically change conformational preferences. Although the low energetic barrier of furanoside ring pseudorotation complicates definitive analysis, experimental and computational studies of the isomeric methyl glycosides of D-Galf have revealed that the anomers have dramatically different conformational preferences (43). The β -Galf 4E -gg-gt conformer that we find in hIntL-1 Chain B is predicted to be the second lowest in energy (0.4 kcal/mol) (43). That conformation for methyl- α -Galf is destabilized by 3.2 kcal/mol. As a result, the expected Boltzmann population for methyl- α -Galf in a 4E -gg-gt conformation is less than 0.2%, ranking it 25th out of the 90 conformations examined (43). These data suggest that α -Galf residues adopt a conformation incompatible with favorable hIntL-1 interactions.

One of the most striking findings from the binding data is that the lectin failed to interact with any of the 148 α -Neu5Ac-containing glycans in the mammalian glycan array (**Figure 3-2A**). A saccharide epitope widespread in human glycans, α -Neu5Ac residues have a terminal

1,2-diol and share similarity with KDO, which are common in microbial glycans and do function as hIntL-1 ligands (47). We used a biotinylated glycoside to confirm that hIntL-1 fails to interact with surfaces displaying α -Neu5Ac (Figure 3-7A). Moreover, compounds identified as hIntL-1 ligands—Gro and Gro-P— competitively inhibit the lectin from binding to β -Gal f , but methyl- α -mannopyranoside and methyl- α -Neu5Ac do not (Figure 3-7B). These results indicate that hIntL-1 uses a single site to bind disparate sterically unhindered 1,2-diol epitopes within microbial glycans, yet the lectin evades interaction with human carbohydrate epitopes.

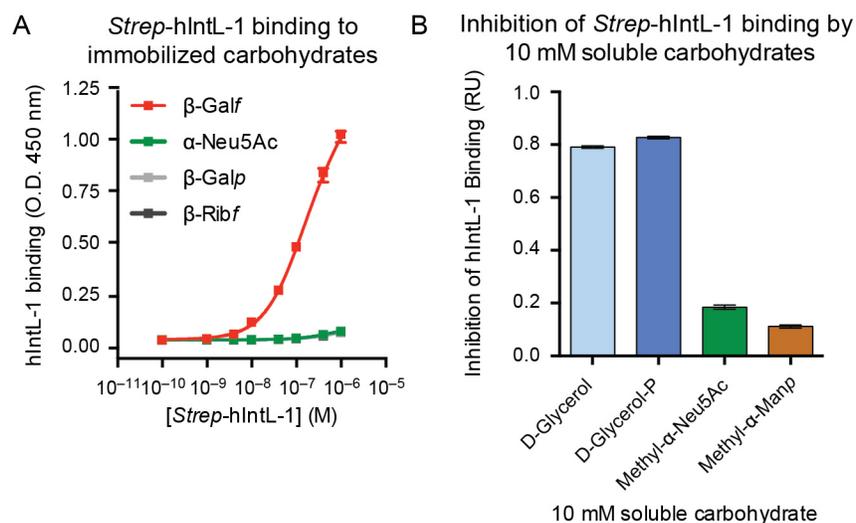


Figure 3-7. hIntL-1 exhibits specificity for microbial glycan epitopes bearing terminal 1,2-diols. (A) hIntL-1 binding to immobilized α -Neu5Ac assayed by the ELISA-like carbohydrate-binding assay (Supplementary Figure 1-1C). Data are fit to a one site binding equation (solid lines). Data are presented as the mean and error bars represent the standard deviation (n=2 of a technical replicate and is representative of three independent experiments). (B) Inhibition of hIntL-1 binding to immobilized β -Gal f . Four compounds (Gro, Gro-P, the methyl- α -glycoside of Neu5Ac, and methyl- α -mannopyranose) were dissolved in binding buffer and included during the hIntL-1 incubation. Binding is relative to a control where no competitor was added to the binding buffer. Data are presented as the mean and error bars represent the standard deviation (n=2 of a technical replicate and is representative of three independent experiments).

To understand the ability of hIntL-1 to discriminate between methyl- α -Neu5Ac and bacterial carboxylic acid-containing sugars such as KDO and KO, we docked methyl- α -Neu5Ac

and methyl- α -KDO into the hIntL-1 structure. The KDO glycoside is readily accommodated, but the α -Neu5Ac glycoside is not (**Figure 3-8A and B**). Anion – anion repulsion between the α -Neu5Ac anomeric exocyclic carboxylate and the carboxylate side chains in the binding site should destabilize binding. Additionally, steric interactions between the methyl group of the anomeric oxygen, and the bulky C(5) N-acetyl group with the protein surface should disfavor α -Neu5Ac complexation (**Figure 3-8A**). The destabilizing interactions with α -Neu5Ac cannot be mitigated by rotating bonds or by adopting accessible low energy conformations. Future experiments using protein variants and ligand analogs will be useful in testing this proposed evasion mechanism.

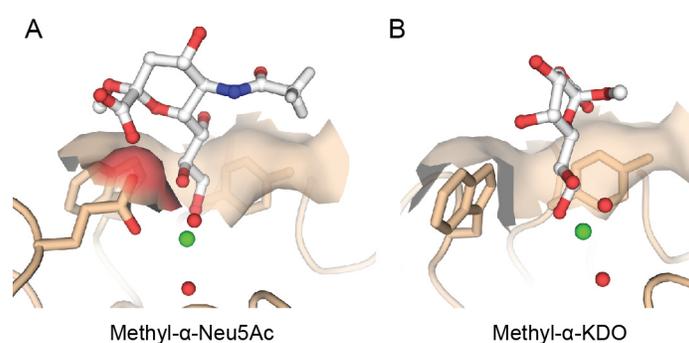


Figure 3-8. Models for hIntL-1 interacting with relevant saccharide epitopes from humans (α -Neu5Ac) or microbes (α -KDO). (A) Docking of methyl- α -Neu5Ac into the hIntL-1 structure. The conformation shown is similar to that observed in other protein structures with a methyl- α -Neu5Ac ligand (PDB: 2BAT, 2P3I, 2P3J, 2P3K, 2I2S, 1KQR, 1HGE, 1HGH (refs. (48-52))). All models in this figure were generated from the allyl- β -D-Galf-bound structure by docking the relevant diol of each compound into the Galf diol electron density using Coot without further refinement. Calcium ions are shown in green and ordered water molecules are depicted in red. (B) Docking of methyl- α -KDO into the hIntL-1 structure. Comparison with methyl- α -Neu5Ac docked into the hIntL-1 structure reveal differences in the steric requirements for binding for each molecule.

3.7 hIntL-1 Comparison with Ficolins

The FBD of hIntL-1 suggested it would be related to the ficolins. With the structure of an X-type lectin complex, it is now apparent that, outside the FBD, intelectins and the ficolins

deviate extensively. IntLs lack the collagen-like domain that mediates complement activation. Additionally, the hIntL-1 CRD is larger than that of the ficolins, and hIntL-1 coordinates three calcium ions, two of which are buried, while the ficolins bind only a single calcium ion. Finally, the carbohydrate binding site and mode of recognition differ. The ficolin calcium ion is not found in the glycan-binding site; in contrast, a surface-exposed calcium ion in hIntL-1 participates directly in glycan binding (**Figure 3-5A**). Together, the data suggest that X-type lectins, of which the hIntL-1 structure serves as the founding member, constitute a distinct protein structural class.

3.8 hIntL-1 Binding to *S. pneumoniae*

Based on the expression of hIntL-1 in mucosal tissues, we examined binding to immunologically distinct serotypes of the encapsulated human lung pathogen *S. pneumoniae*, the causative agent of several diseases, including pneumonia, meningitis, and septicemia (53). The surface exposed pneumococcal CPS is among the first microbial antigens encountered by the immune system upon challenge (54). This capsule is important for pathogen survival and is associated with virulence. Antibodies targeting the capsule have been shown to be protective against pneumococcal diseases, an observation that was leveraged to develop a polysaccharide-based vaccine that is protective against streptococcus infections (55). The serotypes that we selected possess glycans that were present on the microbial glycan array: serotype 8 displays a glycan that lacks a terminal diol, serotype 43 displays a phosphoglycerol unit, and serotypes 20 and 70 possess β -Gal_f residues (**Figure 3-9A**) (54). The data indicate that hIntL-1 binds to the surface of serotypes 20, 70 and 43, each of which displays cell surface glycans with an exocyclic, terminal 1,2-diol (**Figures 3-9B & C and Figure 3-10A & B**). As predicted by the β -

Galf-hIntL-1 complex structure, binding to these strains depends on calcium ion-mediated coordination, and Gro functions as a competitive ligand (**Figures 3-9B & C and Figure 3-10B**). The relative fluorescence intensity of hIntL-1 binding to whole bacteria is generally consistent with the results predicted by the microbial glycan array. Specifically, hIntL-1 bound to strains that display β -Galf (i.e., hit 13 from the microbial array, **Table 3-1**), but it interacted most avidly with the serotype displaying the Gro-P-modified saccharide that was the top hit from the microbial glycan array (**Figure 3-2B**). These data suggest that the relative ligand ranking from the array analysis can provide information about how effectively a lectin can target cells displaying those glycans. Moreover, the results demonstrate that hIntL-1 specifically recognizes structurally diverse exocyclic 1,2-diol containing glycans on the surface of bacteria.

Human IntL-1 has been reported to bind lactoferrin (24), a protein that appears to have antimicrobial activity (56). These observations suggest that hIntL-1 could recruit lactoferrin to microbial cell surfaces for cell killing. To examine the interaction between these proteins, we immobilized human lactoferrin and assayed hIntL-1 binding by ELISA. As reported, we detected an interaction between lactoferrin and hIntL-1, but in contrast to the previous reports, in our assay, this interaction did not require calcium ions. The apparent affinity we measured for the hIntL-1 trimer is rather weak for a specific protein-protein interaction ($K_d \sim 500$ nM). The isoelectric points (pI) of the proteins, pI \sim 5.5 for hIntL-1 and pI \sim 8.5 for lactoferrin, suggest the interaction may be mediated by bulk Coulombic interactions. We were unable to detect any killing of *S. pneumoniae* by human lactoferrin (up to 100 μ g/mL) in a buffer that would be compatible with hIntL-1 binding the cell surface (HEPES buffered saline, pH 7.4, with 2 mM CaCl_2). Our results were consistent with those of others who noted that under similar conditions,

the bactericidal activity of lactoferrin is abolished (57,58). These initial results are inconsistent with a central role for lactoferrin–intelectin complexes mediating microbial cell killing, and they suggest other functional roles for hIntL-1 should be explored.

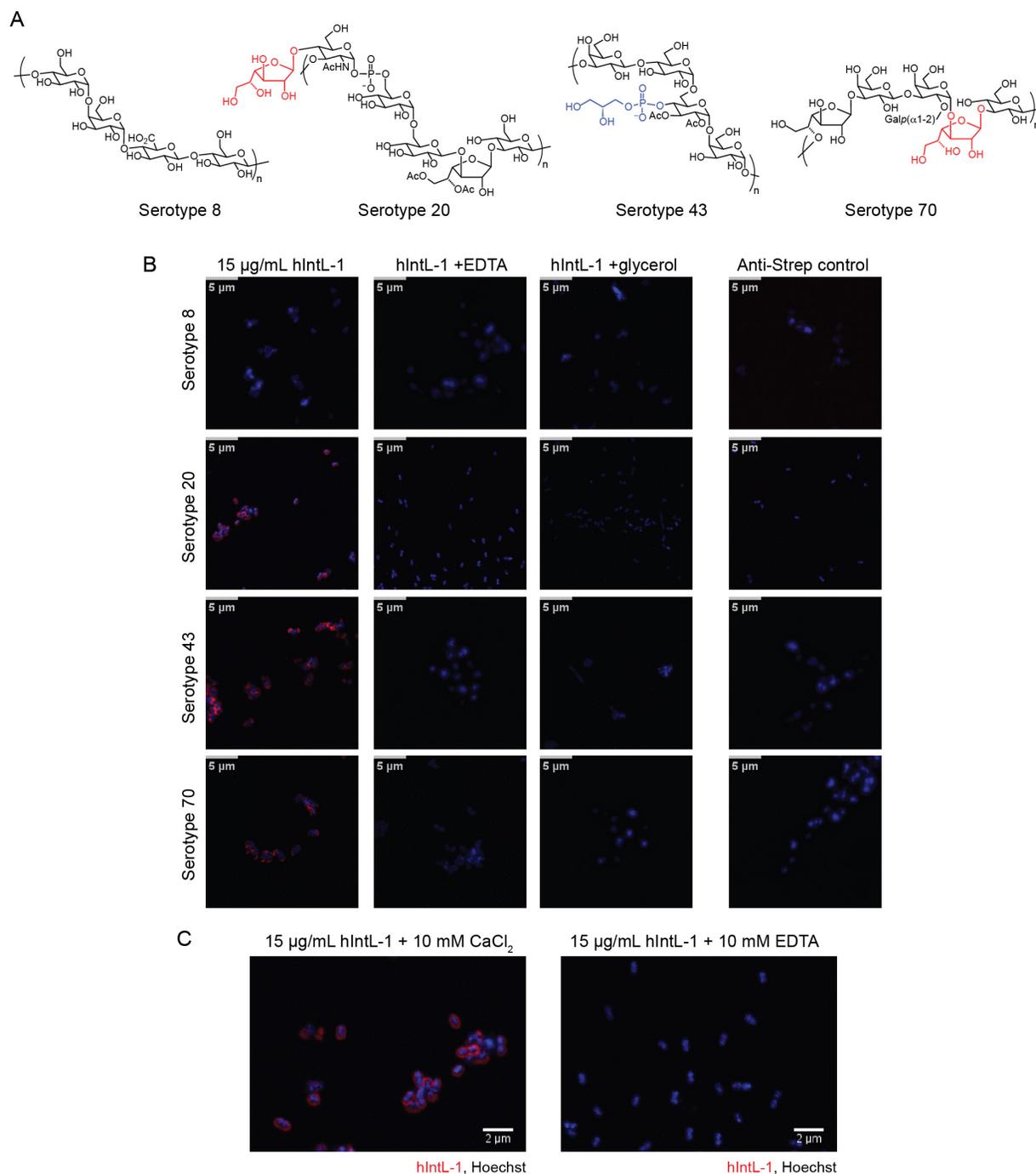


Figure 3-9. hIntL-1 binding to the surface of *S. pneumoniae*. (A) Chemical structure of the capsular polysaccharides displayed on the *S. pneumoniae* serotypes (8, 20, 43, 70) tested. The Galf residues assumed to mediate hIntL-1 cell binding are shown in red and the phosphoglycerol moiety is shown in blue. (B) Fluorescence microscopy of hIntL-1 binding to *S. pneumoniae*.

Bacteria were treated with Strep-tagged hIntL-1 (15 $\mu\text{g}/\text{mL}$) and an anti-Strep-tag antibody conjugate (red). Cellular DNA was visualized with Hoechst (blue). Images are representative of >5 fields of view per sample. Scale bar, 5 μm . (C) Enhanced magnification of hIntL-1 binding the surface of serotype 20 bacteria. Right: EDTA addition abrogates hIntL-1 binding to the bacterial surface, supporting the role for Ca^{2+} . Images are representative of >5 fields of view per sample. Scale bar, 2 μm .

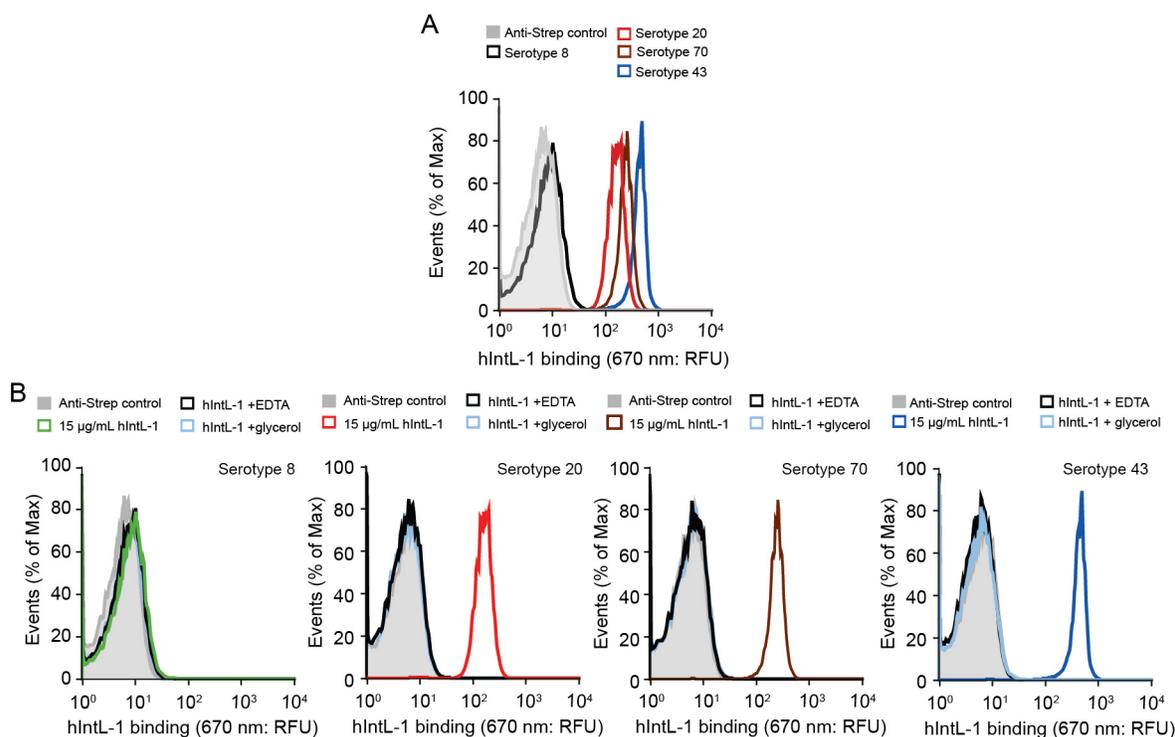


Figure 3-10. hIntL-1 binding to the surface of *S. pneumoniae* monitored by flow cytometry. (A) Flow cytometry analysis of serotypes 8, 20, 43, and 70; data were collected consecutively with identical instrument settings. (B) The dependence of the hIntL-1–carbohydrate interaction on Ca^{2+} was tested by adding 10 mM EDTA and ligand selectivity was tested by adding 100 mM glycerol. Data are representative of two independent experiments.

3.9 Effect of pH on hIntL-1 Binding

The structure of hIntL-1 bound to β -Gal β ligand suggested that ligand binding would be sensitive to pH. Specifically, we hypothesized that as the pH approaches the pKa of the carboxylate side chains used to coordinate the calcium ion involved in ligand binding, they

would become protonated and ligand binding would be attenuated. Using a label-free direct binding experiment on a *Forté*BIO Octet, the effect of pH on binding to immobilized β -D-Galf was assayed. When the pH of the assay buffer fell below 5.5, a reduction in hIntL-1 binding was observed (**Figure 3-11**). At pH=4.5, binding was completely inhibited. We hypothesize that the increased apparent binding at pH=9.0 is due to protein aggregation and enhanced multivalency. The pH within the human lung is typically 7.3 – 7.4 (59), and the pH within the GI tract rarely falls below pH=6 after the stomach (60). These results suggest that the pH of the human lung and GI tract, where hIntL-1 is expressed and localized, is conducive to lectin binding.

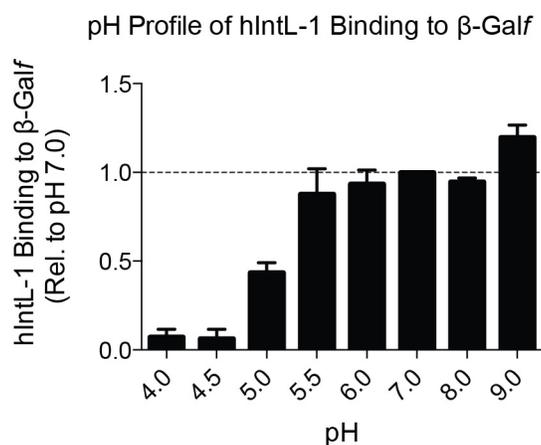


Figure 3-11. pH profile of hIntL-1 binding to immobilized β -Galf. Data is presented as the total amount of hIntL-1 bound to a *Forté*BIO Octet streptavidin labeled tip saturated with biotinylated β -Galf ligand. Data is normalized to the sample assayed at pH=7.0. Error bars represent the s. d. of the mean. n = 5 measurements, on a total of two independently loaded sensors. Sensor surface was saturated with biotinylated ligand.

3.10 Murine IntL-1 Binding to Galf

If the role of intelectins is to participate in defense against microbes, the recognition specificity of intelectins from other mammals should be preserved. We therefore produced murine IntL-1 (mIntL-1), which is the mouse homolog (27) of hIntL-1. When mIntL-1 was tested using the SPR assay used with the human homolog, its glycan recognition properties were

analogous: It failed to interact with β -Ribf, β -Araf, α -Rha, or β -Galp, but it did interact with β -Galf (Figure 3-12). These data support the prospect that IntLs from different species have evolved to bind widely distributed 1,2-diol-containing epitopes unique to microbes.

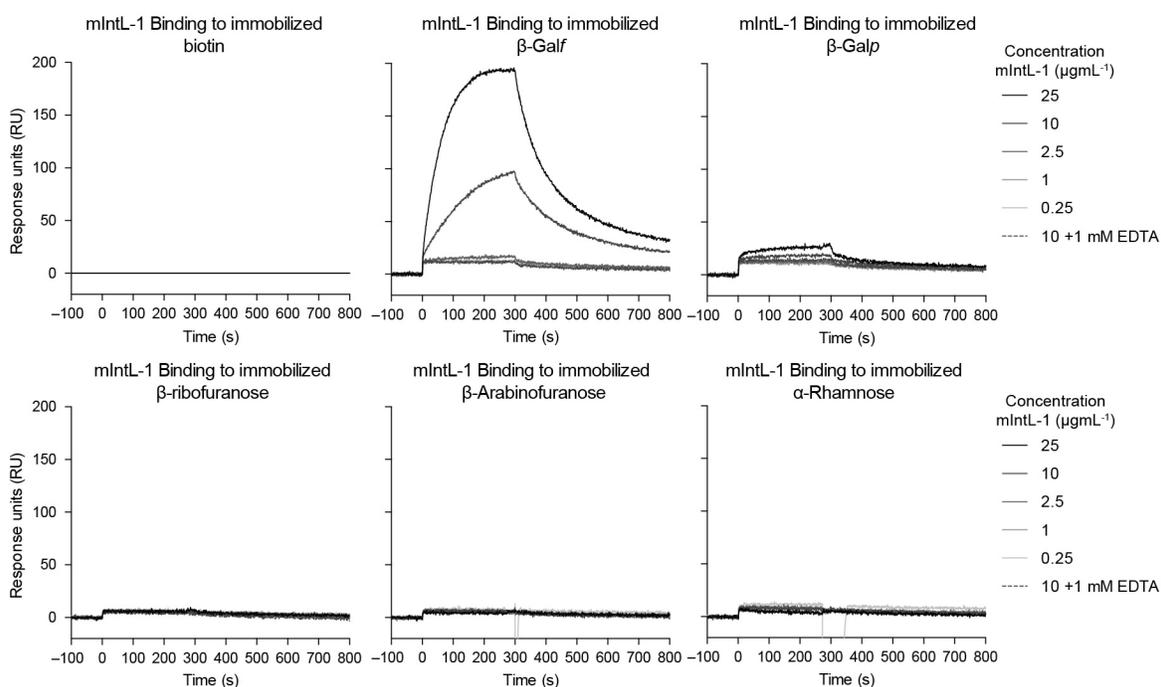


Figure 3-12. Mouse intelectin-1 binding to immobilized carbohydrates. Purified Strep-mIntL-1 binding to immobilized carbohydrates monitored using SPR. Addition of EDTA prevents carbohydrate binding, supporting a role for calcium ions in carbohydrate binding. Data are referenced to the biotin channel.

3.11 Discussion

Data from glycan microarrays reveal that hIntL-1 recognizes multiple microbial glycan epitopes yet paradoxically can discriminate between microbial and mammalian glycans. By determining the structure of this X-type lectin bound to Gal f , this apparent contradiction was resolved. The five common saccharide epitopes identified as recognition motifs (Gal f , Gro-P, glycerol-D-manno-heptose, KDO, and KO) share a common feature: a terminal acyclic 1,2-diol

group. The hIntL-1 X-ray structure indicates that these terminal vicinal hydroxyl groups can coordinate to a protein-bound calcium ion. This binding mode has similarities to that employed by another major class of mammalian carbohydrate-binding proteins: the C-type lectins (16). C-Type lectins also recognize glycans through calcium ions in the binding site to which carbohydrate hydroxyl groups coordinate (7). In the case of C-type lectins, however, the hydroxyl groups employed are typically those on the pyranose ring of a mannose or fucose residue. The hIntL-1 binding pocket requires that any 1,2-diol motifs possess a primary hydroxyl group, as the aromatic substituents W288 and Y297 act as walls to preclude the binding of more substituted diols. These aromatic substituents presumably not only contribute to specificity but also to affinity. Their positioning may allow them to participate in CH- π interactions (61) which would enhance binding.

Although the terminal 1,2-diol is necessary for hIntL-1 recognition, it is not sufficient. The lectin is unable to bind human glycans, including those with an α -Neu5Ac residue. This result was confusing as glycans with α -Neu5Ac residues were prevalent on the mammalian glycan microarray, and although many glycans in this array present a terminal 1,2 diol, none were bound by hIntL-1. We were unable to model methyl- α -Neu5Ac in the hIntL-1 binding site without incurring Coulombic repulsion or severe steric interactions. These observations suggest a molecular basis for hIntL-1's ability to avoid interaction with human glycans. With a structure that identifies the glycan-binding site, the proposed rationale for hIntL-1's selectivity for microbial glycans can be tested further.

We anticipate our structure will also provide insight into the physiological roles of the intelectins. The upregulation of intelectins upon infection suggests they may function in innate

immunity. Although existing data from GWAS do not directly link intelectin mutations and increased susceptibility to infection, there are studies linking hIntL-1 to asthma (62) and Crohn's disease (63). These diseases arise from defects at mucosal surfaces where intelectins are secreted. In relation to asthma, the amino acid variant V109D was identified with an increased risk (62). Our structure reveals that this residue is not centrally important for binding, but it is located at a monomer–monomer interface.

We postulate that the trimeric form of hIntL-1 is important for the lectin's function. The presence of three binding sites on one face of the hIntL-1 trimer (**Figure 3-4A**) suggested the protein could exploit multivalency to recognize relevant terminal 1,2 diol motifs and bind avidly to microbes. We therefore tested whether hIntL-1's selectivity for glycans would be manifested in a proclivity to engage only those *S. pneumoniae* serotypes whose capsular polysaccharides possess hIntL-1 recognition motifs. Our finding that hIntL-1 bound to strains bearing Gal β (serotypes 20 or 70) or Gro-P (serotype 43) but not those lacking the requisite terminal 1,2-diol (serotype 8) highlights the advantages of using a simple binding epitope: Human IntL is not restricted to binding solely one glycan building block, rather it can interact with bacterial cells that present glycans composed of very different components (Gal β versus Gro-P).

Because it engages a small epitope found within microbial glycans, hIntL-1 should be capable of recognizing a wide variety of microbes. Analysis of the twenty most common glycan building blocks unique to microbes indicates that half of these possess an acyclic 1,2-diol that could, in principle, be recognized by intelectins (**Figure 3-13**) (32). The potential that a given microbe generates glycan ligands for hIntL-1 can be inferred from genetic sequence data. For example, organisms bearing Gal β residues harbor a *glf* gene (29). D-Glycerol-1-phosphate-

modified glycans are generated using CDP-D-glycerol as an activated donor and therefore will encode functional homologs of the *S. pneumoniae gct* gene (54). Pathways that lead to the incorporation of heptose, KO, and KDO are known, as these residues are found in LPS (64) and K-antigen of Gram-negative bacteria (65). The orientation of the saccharide binding sites on a single face of the hIntL-1 trimer not only can enhance the avidity of cell-surface binding, it also provides a surface for recruitment of other immune proteins or effectors to a hIntL-1-bound microbe. The remarkable selectivity of hIntL-1 for microbial over human cell surface glycans raises the intriguing possibility that IntLs function as microbial detectors. It is possible that this selective microbial recognition can be harnessed to deliver cargo to microbes, to detect them, or to target them for destruction.

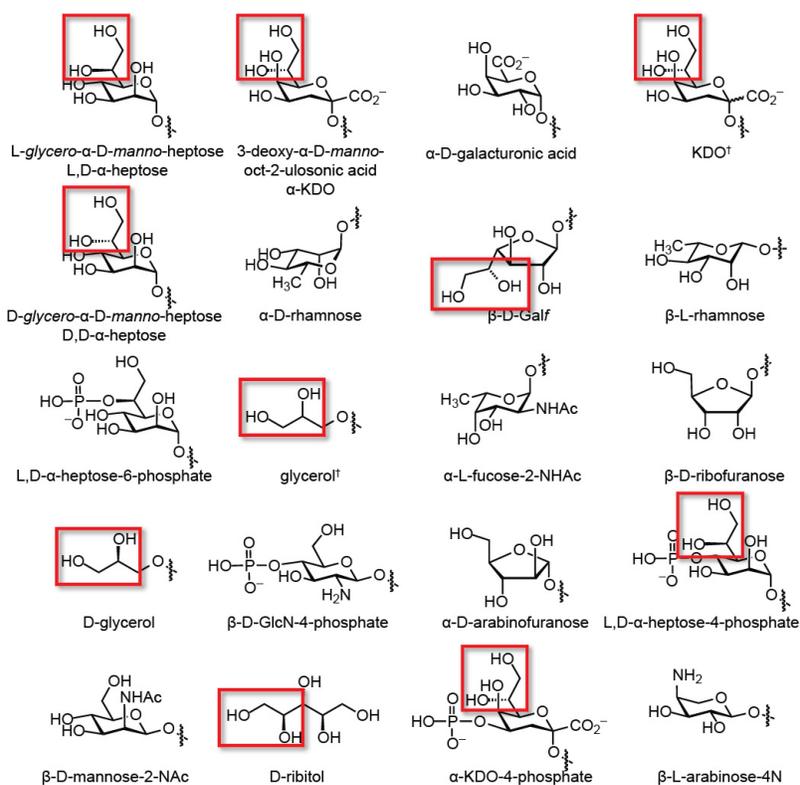


Figure 3-13. Structures of the 20 most prevalent monosaccharides that are unique to bacterial glycans. The most common, L,D- α -heptose, is shown in the top left corner and number twenty, β -

L-arabinose-4-N, is shown in the bottom right. This figure is derived from data in reference (32). Terminal acyclic 1,2-diol epitopes that could serve as ligands of hIntL-1 are highlighted with a red box.

Table 3-2. Data Collection and Refinement Statistics (Molecular Replacement).

	Apo-hIntL-1	Gal f -bound hIntL-1
Data collection		
Space group	P 2 ₁ 3	P 2 ₁ 3
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	118.4, 118.4, 118.4	117.9, 117.9, 117.9
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	22.00-1.80 (1.86-1.80) ^a	28.59-1.60 (1.66-1.60)
<i>R</i> _{sym}	0.119 (0.495)	0.078 (0.773)
<i>I</i> / σ <i>I</i>	19.6 (3.7)	29.4 (3.0)
Completeness (%)	100 (100)	100 (100)
Redundancy	11.2 (10.1)	11.1 (10.9)
Refinement		
Resolution (Å)	22.00-1.80 (1.86-1.80)	28.59-1.60 (1.68-1.60)
No. reflections	48784	68256
<i>R</i> _{work} / <i>R</i> _{free}	0.133/ 0.164	0.155/0.180
No. atoms		
Protein	4551	4606
Ca ²⁺	6	6
Allyl- β -D-Gal f	-	30
Water	658	616
<i>B</i> -factors (Å ²)		
Protein	14.0	20.2
Ca ²⁺	10.1	14.5
Allyl- β -D-Gal f	-	33.8
Water	26.0	32.6
rms deviation		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.107	1.119

Each data set was collected from one crystal.

^aValues in parentheses are for highest-resolution shell.

3.12 Methods

3.12.1 Chemical Synthesis of Glycan Ligands

Procedures for glycan synthesis are included and described in detail in the Supplementary Note of reference (37).

3.12.2 Native Human Intelectin-1 Expression and Purification

The cDNA for hIntL-1 (Accession Number: NM_017625) was obtained from Open Biosystems Clone LIFESEQ2924416 as a glycerol stock (GE Healthcare). The full coding sequence, residues 1-313, was amplified using PCR with the forward primer 5'-CGTGGGATCCTGGAGGGAGGGAGTGAAGGAGC-3' and the reverse primer 5'-GCCAGCTCGAGACCTTGGGATCTCATGGTTGGGAGG-3'. The primers installed sites for the restriction endonucleases *Bam*HI and *Xho*I, respectively. The doubly digested PCR fragment encoding hIntL-1 was ligated into a doubly digested pcDNA4/*myc*-HisA vector backbone (Life Technologies). Correct insertion was confirmed with DNA sequencing (UW-Madison Biotechnology Center).

The hIntL-1 gene was expressed via transient transfection of suspension adapted HEK 293T cells obtained from the American Tissue Culture Collection (ATCC). Cells were transfected in Opti-mem I Reduced Serum Medium (Life Technologies) at ~2E6 cells/mL using Lipofectamine 2000 (Life Technologies), according to the manufacturers protocol. Six hours post transfection, the culture medium was exchanged to FreeStyle F17 expression medium (Life Technologies) supplemented with 50 U/mL penicillin-streptomycin, 4 mM L-glutamine, 1x nonessential amino acids, 0.1% fetal bovine serum (FBS) and 0.1% Pluronic F-68 (Life Technologies). Cells expressing hIntL-1 were cultured for up to 6 days, or until viability

decreased below 60%, at which point the conditioned expression medium was harvested by centrifugation and sterile filtration.

Conditioned medium was adjusted to pH 7.4 by slow addition of a 0.1 M solution of sodium hydroxide (NaOH), and calcium chloride (CaCl₂) was added from a 1 M stock solution to achieve a final concentration of 10 mM. Recombinant hIntL-1 was purified by binding to a β -Gal column generated from reaction of a β -Gal glycoside bearing an anomeric linker and an amine to UltraLink Biosupport (Piere). The resulting resin was washed with a solution of 20 mM HEPES (7.4), 150 mM sodium chloride (NaCl), and 10 mM CaCl₂. Human IntL-1 was eluted with a solution of 20 mM HEPES (7.4), 150 mM NaCl, and 10 mM ethylenediaminetetraacetic acid (EDTA), and the protein was concentrated using a 10,000 molecular weight cut-off (MWCO) Amicon Ultra Centrifugal Filter. The buffer was exchanged to 20 mM HEPES (7.4), 150 mM NaCl, and 1 mM EDTA. Protein purity was assessed by SDS-PAGE electrophoresis and Coomassie blue staining, and was often >95%. The concentration of hIntL-1 was determined using absorbance at 280 nm with a calculated $\epsilon=237,400 \text{ cm}^{-1}\text{M}^{-1}$ for the trimer, and an estimated trimer molecular mass of 101,400 Da (to account for glycosylation). Typical yields from a 30 mL transfection were 400 μg .

3.12.3 Expression and Purification of *Strep-tag*[®] II hIntL-1

An N-terminal *Strep-tag*[®] II was cloned into the hIntL-1::pcDNA4 vector using site-directed mutagenesis and a primer set comprised of 5'-ACCACCAGAGGATGGAGTACAGATTGGAGCCATCCGCAGTTTGAAAAGTCTACAGATGAGGCTAATACTTACTTCAAGGA-3' and its reverse complement. The correct insertion was confirmed with DNA sequencing. *Strep*-hIntL-1 was expressed identically to hIntL-1. For

purification, conditioned *Strep*-hIntL-1 medium was adjusted to pH=7.4 using NaOH, avidin was added per the IBA GmbH protocol (IBA GmbH, cat. no. 2-0205-050), CaCl₂ was added to 10 mM, and the solution was cleared with centrifugation (15,000g for 15 minutes). Protein was captured onto 2 mL of *Strep*-Tactin Superflow resin (IBA GmbH, cat. no. 2-1206-002). The resulting resin was washed with a solution of 20 mM HEPES (7.4), 150 mM NaCl, and 10 mM CaCl₂ and then 20 mM HEPES (7.4), 150 mM NaCl, and 1 mM EDTA. The protein was eluted with 5 mM d-desthiobiotin (Sigma) in 20 mM HEPES (7.4), 150 mM NaCl, and 1 mM EDTA and concentrated using a 10,000 MWCO Amicon Ultra Centrifugal Filter. The concentration of *Strep*-hIntL-1 was determined using absorbance at 280 nm with a calculated $\epsilon=237,400 \text{ cm}^{-1}\text{M}^{-1}$ for the trimer, and an estimated trimer molecular mass of 101,400 Da. Typical yields were similar to what was measured with untagged hIntL-1.

For protein X-ray crystallography, *Strep*-hIntL-1 was purified following culture medium dialysis against 20 mM BIS-TRIS (6.7), 150 mM NaCl, and 1 mM EDTA. The pH of the culture medium was adjusted to 6.7, avidin was added per the IBA GmbH protocol, CaCl₂ was added to 10 mM and the solution was cleared with centrifugation. Protein was purified by capture onto *Strep*-Tactin Superflow resin. Resin was washed with 20 mM BIS-TRIS (6.7), 150 mM NaCl, 10 mM CaCl₂ and then 20 mM BIS-TRIS (6.7), 150 mM NaCl, 0.5 mM EDTA. Protein was eluted with 5 mM D-desthiobiotin (Sigma) in 20 mM BIS-TRIS (6.7), 150 mM NaCl, 0.5 mM EDTA and concentrated using a 10,000 MWCO Amicon Ultra Centrifugal Filter.

3.12.4 hIntL-1 Carbohydrate Binding ELISA-like Assay

To fabricate carbohydrate-displaying surfaces, 0.5 μg of streptavidin (Prozyme, cat. no. SA20) was adsorbed onto a Maxisorp (Nunc) flat bottom 96 well plate in phosphate-buffered

saline (PBS). Wells were washed with PBS and then coated with 5 μ M of carbohydrate–biotin ligand in PBS for 1 hour at 22 °C. Wells were blocked with bovine serum albumin (BSA) in ELISA buffer (20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂, and 0.1% Tween-20). Samples containing hIntL-1 were prepared by serial dilution into ELISA buffer + 0.1% bovine serum albumin (BSA) and added to wells for 2 hours at 22 °C. Wells were washed four times with ELISA buffer. Bound hIntL-1 was detected using 0.75 μ g/mL of a sheep polyclonal IgG hIntL-1 antibody (R&D Systems, cat. no. AF4254) in ELISA buffer + 0.1% BSA for 2 hours at 22 °C. This primary antibody has been validated by the company for detecting intelectin by Western blot, immunohistochemistry, and direct ELISA. Wells were washed with ELISA buffer. A donkey anti-sheep IgG HRP conjugate (Jackson ImmunoResearch Laboratories) was added at a 1:5,000 dilution in ELISA buffer + 0.1% BSA for 1 hour at 22 °C. When *Strep*-hIntL-1 was assayed, StrepMAB-Classic HRP conjugate (IBA GmbH, cat. no. 2-1509-001) was used to specifically recognize the Strep-tag[®] II of bound hIntL-1. StrepMAB-Classic HRP conjugate was diluted 1:10,000 in ELISA buffer + 0.1% BSA and incubated for 2 hours at 22 °C. Wells were washed. Human IntL-1 was detected colorimetrically with addition of 1-Step Ultra TMB-ELISA (Pierce). Once sufficient signal was achieved (typically < 2 minutes), the reaction was quenched with addition of equal volume of 2 M sulphuric acid (H₂SO₄). Plates were read at 450 nm on an ELx800 plate reader (Bio-Tek). When testing the calcium ion dependency of hIntL-1, 1 mM EDTA replaced 10 mM CaCl₂ in all steps. Data were analyzed on Prism6 (GraphPad). Data were fit to the one site binding equation.

3.12.5 SPR Analysis

Analysis of intelectins using SPR was conducted on a ProteOn XPR36 (Bio-Rad) at the University of Wisconsin–Madison Department of Biochemistry Biophysics Instrumentation Facility (BIF). To measure intelectin binding, ProteOn NLC sensor chips (NeutrAvidin coated sensor chip) (Bio-Rad, cat. no. 176-5021) were used to capture the biotinylated carbohydrate ligand. All experiments presented here were conducted at surface saturated levels of ligand, ~200 response units (RU). In all experiments, captured biotin was used in flow cell one as a control. Samples containing purified intelectin were prepared by serial dilution into intelectin SPR running buffer (20 mM HEPES (7.4), 150 mM NaCl, 1 mM CaCl₂, and 0.005 % Tween-20). Surfaces were regenerated with short injections of solutions of 10 mM hydrochloric acid (HCl). Data were referenced using either the interspots or the biotin reference channel, and processed using the Bio-Rad ProteOn software package.

3.12.6 Construction of the Furanoside Glycan Array

The microarray of furanoside containing glycans was printed as previously described(66,67). Briefly, the amine functionalized glycans shown in Fig. s6A were dissolved in 100 mM sodium phosphate (8.0) and printed as 14 arrays on N-hydroxysuccinimidyl (NHS) ester-activated slides (Shott Nexterion, Louisville, KY). Arrays were printed in replicates of n=4 at different glycan concentrations (as indicated in Fig. s6B) using a Piezorray printer (Perkin Elmer, Waltham, MA) that delivered 0.33 nL per spot. The 2-amino(N-aminoethyl) benzamine (AEAB) derivatives of lacto-N-neotetraose (LNnT) and asialo, galactosylated bi-antennary N-linked glycan (NA2) were printed as controls to confirm glycan immobilization. After printing, covalent coupling of glycans to the surface was facilitated by incubation at 55 °C in an atmosphere of >80% humidity for 1 hour. Slides were dried in a desiccator overnight and

blocked using a solution of 50 mM ethanolamine in 50 mM borate buffer (8.0). Prior to interrogating with glycan binding proteins (GBPs), the arrays are rehydrated in binding buffer.

3.12.7 Assay of hIntL-1 on Furanoside and CFG Mammalian Glycan Array

GBPs at various concentrations were applied to separate furanoside arrays in 70 μ L of binding buffer (20 mM HEPES (7.4), 150 mM NaCl, 1 mM EDTA, 10 mM CaCl₂, 1% BSA and 0.05% Tween-20) in the wells formed on the slide with a silicon grid (14 wells per slide). After incubation for 1 hr at RT, the slides were washed with wash buffer (20 mM HEPES (7.4), 150 mM NaCl, 1 mM EDTA and 10 mM CaCl₂, 0.05% Tween-20). The biotinylated lectins *Erythrina cristagalli* lectin (ECL) and *Ricinus communis* agglutinin I lectin (RCA-I) were detected using Alexa Fluor[®] 488-labeled streptavidin (10 μ g/ml) in binding buffer (Fig. s6C and D). hIntL-1 was detected with a sheep polyclonal IgG antibody specific for hIntL-1 (5 μ g/ml) (R&D Systems) and an Alexa Fluor[®] 488-labeled donkey anti-sheep IgG secondary antibody (5 μ g/ml) (Life Technologies). Bound protein was detected using a ProScanArray Scanner (Perkin Elmer) equipped with 4 lasers covering an excitation range from 488 to 633 nm. The data from the furanoside glycan array were analyzed with the ScanArray Express software (Perkin Elmer) as the average of the 4 replicates.

For the analysis of the CFG glycan array(36), hIntL-1 was applied in 70 μ l at a concentration of 50 and 200 μ g/ml in binding buffer under a coverslip to distribute the solution evenly over the large array of 610 glycans printed in replicates of $n=6$ (Array v5.1). After washing and scanning, the data from the CFG glycan microarray were analyzed using ImaGene software (BioDiscovery, Hawthorne, CA) as the average of 4 values after removing the high and low values of the 6 replicates. With both the furanoside and mammalian glycan array, the images

were converted to Excel files, and the data are reported as histograms of average Relative Fluorescence Units (RFU) versus print identification number that identified the glycan targets. Figures were made using Prism6 (GraphPad) or Excel (Microsoft).

3.12.8 Assay of *hIntL-1* on the Bacterial Glycan Array

Strep-hIntL-1 was used to interrogate the MGMv2. Construction of the MGMv2 is previously described (39). Briefly, bacterial polysaccharide samples were dissolved and diluted to 0.5 mg/mL in printing buffer (150 mM sodium phosphate buffer (8.4) + 0.005% Tween-20). Samples were immobilized on NHS-activated glass slides (SlideH, Schott/Nexterion) using a MicroGrid II (Digilab) contact microarray printer equipped with SMP-4B printing pins (Telechem). Six replicates of each bacterial glycan sample were printed. Covalent coupling of glycans to the surface was facilitated by incubation for 1 hour post-print at 100% relative humidity. The remaining reactive NHS-moieties were quenched using a blocking solution (50 mM ethanolamine in 50 mM borate buffer (9.2)). Blocked slides were stored at -20 °C until assays were performed.

To interrogate the MGMv2, *Strep-hIntL-1* was diluted to 50 µg/mL in binding buffer (20 mM Tris-HCl (7.4), 150 mM NaCl, 2 mM CaCl₂, 2 mM magnesium chloride (MgCl₂) 1% BSA, and 0.05% Tween-20) and applied directly to the array surface for 1 hour. Following incubation, the array was washed by dipping into binding buffer four times. The *Strep-tag*[®] II on bound *hIntL-1* was detected using StrepMAB-Classic Chromeo647 nm (10 µg /mL, IBA GmbH Lifesciences) diluted in binding buffer applied directly to the array surface and allowed to incubate for 1 hour. The array was washed in binding buffer (4 dips), binding buffer minus BSA and Tween-20 (4 dips) and de-ionized water (4 dips). Finally, the array was dried by

centrifugation and scanned. Interrogated arrays were scanned for Chromeo647 signal using a ProScanArray Express scanner (Perkin Elmer) and resultant images were processed to extract signal data using Imagene (v6.0, Biodiscovery). Signal data was calculated as the average of 4 values after removing the high and low values of the 6 replicates. Data were plotted using Excel (Microsoft) as average RFU versus print identification number. Figures were made using Prism6 (GraphPad).

3.12.9 Protein X-ray Crystallography

The *Strep*-hIntL-1 protein that was purified using 20 mM BIS-TRIS (6.7) buffers, was concentrated to 1.5 mg/mL, 1 M CaCl₂ was added to a final concentration of 10 mM, and crystallization (hanging-drop vapor-diffusion) was achieved by mixing 1 μ L of the protein solution and 1 μ L of well solution (100 mM BIS-TRIS (6.0) and 25% PEG 3350). Crystals grew to full size in two weeks. Protein crystals of Apo-hIntL-1 were cryoprotected via transfer to well solution supplemented to a total concentration of 35% PEG 3350 for one minute and then vitrified in liquid nitrogen. The allyl- β -Gal-f-hIntL-1 complex was formed by soaking apo-hIntL-1 crystals in cryoprotection solution supplemented with 50 mM allyl- β -D-galactofuranose for two weeks.

Single crystal X-ray diffraction experiments were performed at beamline 21-ID-D (Life Sciences Collaborative Access Team, LS-CAT), Advanced Photon Source, Argonne National Laboratory. The wavelength for data collection was 0.97924 Å for the Apo-hIntL-1 structure and 1.00394 Å for Gal-f-Bound hIntL-1. Integration, scaling, and merging were performed with HKL2000 (68). The structure was solved using the PHENIX suite(69). The *Xenopus laevis* intelectin structure recently solved in our lab was used as a search model to determine the

structure of apo-hIntL-1 by molecular replacement using Phaser (70). Because the apo-hIntL-1 and β -Gal β -bound hIntL-1 data are isomorphous, the structure of β -Gal β -bound hIntL-1 was solved by a difference Fourier method using apo-hIntL-1 as a starting model for rigid-body refinement with phenix.refine (71). The chemical restraint for β -Gal β was generated by PRODRG(72). Model adjustment and refinement were performed in Coot and phenix.refine, respectively (Supplemental Table 1) (73). The model was validated using MolProbity(74). Crystal structure figures were generated with PyMOL (<http://www.pymol.org>).

3.12.10 hIntL-1 Binding to *Streptococcus pneumoniae*

Streptococcus pneumoniae (Klein) Chester serotype 8 (ATCC[®] 6308[™]), 20 (ATCC[®] 6320[™]), 43 (ATCC[®] 10343[™]) and 70 (ATCC[®] 10370[™]) were obtained from the ATCC. The structure of the capsular polysaccharide from each of these serotypes has been previously determined(54). Cells were revived in tryptic soy broth containing 5% defibrinated sheep blood. Cells were grown on plates of tryptic soy agar containing 5% defibrinated sheep or in suspension in Luria Broth (LB). Cells were grown at 37 °C supplemented with 5% carbon dioxide gas. During liquid culture, cells were shaken at 100 RPM. To analyze hIntL-1 binding to the bacterial cell surface, cells were harvested by centrifugation, washed with PBS and fixed in 1% formaldehyde in PBS for 30 minutes on ice. Cells were stained with 15 μ g/mL *Strep*-hIntL-1 with a 1:250 dilution of StrepMAB-Classic Oyster 645 conjugate (IBA GmbH, cat. no. 2-1555-050) in 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂, 0.1% BSA and 0.05% Tween-20 for 2 hours at 4 °C. To test the calcium ion dependency of binding, 20 mM HEPES (7.4), 150 mM NaCl, 10 mM EDTA, 0.1% BSA and 0.05% Tween-20 was used as the buffer. To assay for competitive inhibition by soluble glycerol, 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂,

100 mM glycerol, 0.1% BSA and 0.05% Tween-20 was used as the buffer. Cells were washed with 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂, 0.1% BSA and 0.05% Tween-20, aggregates were removed with a flow cytometry cell-strainer cap (Falcon), and propidium iodide (Life Technologies) was added to a 1:500 dilution. Cells were analyzed on a BD FACSCalibur (Becton, Dickinson and Company) at the University of Wisconsin–Madison Carbone Center Center Flow Cytometry Laboratory. Propidium iodide was used to differentiate fixed *S. pneumoniae* cells from debris. Data was analyzed using FlowJo (FlowJo, LLC, Ashland, OR).

For analysis by microscopy, cell aliquots were taken directly from the flow cytometry samples prior to propidium iodide staining. Samples were subsequently stained with Hoechst 33342 (Life Technologies). Each sample was spotted onto a glass bottom microwell dish (MatTek corporation) and covered with a 1% (w/v) agarose pads prepared in a matched buffer. Images were collected at room temperature using a Nikon A1 laser scanning confocal microscope (Nikon Instruments Inc.). Images were acquired using a Nikon plan apo 100/1.4 oil objective using a 1.2 AU pinhole diameter and NIS-elements C software (Nikon Instruments Inc.). Laser setting were determined by imaging the brightest control sample, serotype 43 treated with 15 µg/mL *Strep-hIntL-1* and a 1:250 dilution of StrepMAB-Classic Oyster 645 conjugate in calcium buffer, to prevent pixel oversaturation. The pinhole diameter, offset, PMT gain, and laser power were then held constant for each prepared sample. Each image was taken at the Z-plane that provided maximal signal for the given section. For Hoechst 33258, illumination was performed using a 405 nm laser and emission was collected between 425 and 475 nm. For StrepMAB-Classic Oyster 645 conjugate, illumination was performed using a 638 nm laser and emission was collected between 663 and 738 nm. Images were prepared using the open source

Fiji distribution of ImageJ, brightness and contrast were adjusted in the control sample (serotype 43 treated with 15 $\mu\text{g}/\text{mL}$ *Strep*-hIntL-1 with a 1:250 dilution of StrepMAB-Classic Oyster 645 conjugate in calcium buffer) and propagated to all selected sample images for comparison. Images were then converted to an RGB format to preserve normalization and then assembled into panels.

3.12.11 pH Profile of hIntL-1 Ligand Binding

A *Forté*BIO Octet RED96 instrument was used to measure hIntL-1 binding to a biotin- β -Gal f immobilized on a streptavidin surface (37). Streptavidin coated biosensors (Pall Life Sciences; cat. no., 18-0519) were coated with β -Gal f by incubation in 200 μL PBS with 1 μM biotin: β -Gal f for 180 seconds. This was shown to result in a surface nearly saturated with ligand. To assay the effect of pH on hIntL-1 binding, 5 $\mu\text{g}/\text{mL}$ of hIntL-1 was added to a solution of 150 mM NaCl, 10 mM CaCl_2 , 0.1 % BSA, and 0.05 % Tween-20; the buffer component was 20 mM acetate for pH 4.0 – 5.5, 20 mM MES for pH=6.0, 20 mM HEPES for pH 7.0 and 8.0, and 20 mM glycine for pH=9.0. Each pH was measured in one well of the eight wells available, per run. hIntL-1 binding to the surface was monitored for 400 seconds and dissociation was monitored for 300 seconds. After each binding event, and before the first for a sensor, the surface was regenerated using six, 10 second exposures to 10 mM HCl and then PBS, iteratively. The first set of sensors was assayed twice. An independent set of sensors were generated and subjected to three binding experiments. The amount of hIntL-1 bound to the surface was determined by quantifying the signal of each tip after the 400 second association step. Data from each pH were normalized to the value obtained at pH=7.0, within each run. Data was graphed using GraphPad Prism 6.

3.12.12 Expression of Murine *IntL-1*

The cDNA for mIntL-1 (Accession Number: O88310) was obtained from Open Biosystems Clone 1095883 as a glycerol stock (GE Healthcare). The full coding sequence, residues 1-313, was amplified using PCR with 3 primers mIntL-1_a 5'-GCGCGGATCCATGACCCAACTGGGATTCTGCTGTTTATCATGGTTGCTACCAGAGGTTGCAGT-3' mIntL-1_b 5'-ATCATGGTTGCTACCAGAGGTTGCAGTGCAGCTGAATGGAGCCATCCGCAGTTTGAA AAGGCTGAAGAGAACCTGGACACCAACAGG-3' and mIntL-1_c 5'-TACCAAGCTTTCATTAGCGATAAAACAGAAGCACAGCTGCTTCAG -3'. The gene was first amplified with mIntL-1_b and mIntL-1_c, and then amplified with mIntL-1_a and mIntL-1_b. The resulting PCR product was digested with *Bam*HI and *Hind*III and ligated into a similarly digested pFastBac1 vector. The N-terminus of the resulting protein product is predicted to be MTQLGFLLFIMVATRGCSAAE**WSHPQFEK**AEEEN; where the underlined amino acids are the predicted secretion signal peptide and the bold amino acids denote the *Strep*-tag[®] II incorporated for purification. Following sequence verification, the vector was subjected to recombination into a baculovirus according to the manufacturers protocol (Bac-to-Bac Baculovirus Expression System, Life Technologies). The resulting baculovirus genomic DNA was transfected (Insect GeneJuice, Novagen) into Sf21 insect cells to produce the first generation (P1) of recombinant baculovirus. Amplified baculovirus (P2) was produced using P1 to infect suspension Sf21 cells grown in SF900-II-SFM (Life Technologies). Recombinant P2 baculovirus was isolated by centrifugation of the culture, and sterile filtration of the conditioned medium. Fetal bovine serum was added to 2% to stabilize the baculovirus.

Murine IntL-1 protein was produced as a secreted protein using High Five cells (Life Technologies), a derivative of *Trichoplusia ni*, cultured in Express Five SFM (Life Technologies) supplemented with 1x antibiotic-antimycotic, 10 µg/mL gentamicin, 4 mM glutamine. The cells were grown to ~2E6 cells/mL, and then infected with 1 µL baculovirus (P2) per 1E6 viable cells. After four days, the culture medium was harvested via centrifugation and filtration (0.22 µM). Conditioned medium was dialyzed extensively against 20mM BIS-TRIS (6.5), 150 mM NaCl, and 500 µM EDTA and then against 50 mM TRIS (7.5), 150 mM NaCl, and 500 uM EDTA. The medium was then slowly adjusted to pH=7.5 with addition of a solution of 0.1 M NaOH, CaCl₂ was added to 10 mM, and avidin was added per the IBA GmbH protocol (IBA GmbH, cat. no. 2-0205-050) to absorb excess biotin. The solution was cleared by centrifugation (20000g for 20 minutes).

Protein was purified by capture onto *Strep*-Tactin Superflow resin. The column was washed with 20 mM TRIS (7.4), 150 mM NaCl, and 10 mM CaCl₂ and then 20 mM TRIS (7.4), 150 mM NaCl, 1 mM EDTA. Protein was eluted with 5 mM d-desthiobiotin (Sigma) dissolved in 20 mM TRIS (7.4), 150 mM NaCl, 0.5 mM EDTA and concentrated using a 10,000 MWCO Amicon Ultra Centrifugal Filter. Buffer was exchanged to 20 mM TRIS (7.4), 150 mM NaCl, 0.5 mM EDTA. Protein purity was assessed by SDS-PAGE electrophoresis and Coomassie blue staining, and was >95 %. The concentration of mIntL-1 was determined using absorbance at 280 nm with a calculated $\epsilon=75,330 \text{ cm}^{-1}\text{M}^{-1}$ for a monomer, and an estimated monomer molecular mass of 33,951 Da. Typical yields from 500 mL of conditioned medium were 40 µg.

3.13 Accession Codes

Coordinates and structure factors have been deposited in the PDB under accession codes 4WMQ (apo-hIntL-1) and 4WMY (Gal β -bound hIntL-1).

3.14 Contributions

Kittikhun Wangkanont determined the crystallization conditions for hIntL-1 and solved the crystal structures reported here. In doing this, K.W. also solved the structure of a *X. laevis* intelectin. K.W. also modeled the 1,2-diol containing ligands into the hIntL-1 crystal structure. The carbohydrate ligands used in this study were synthesized and characterized by Rebecca A. Splain and Matthew B. Kraft. The mammalian glycan microarray was performed by Xuezheng Song, David F. Smith, and Richard D. Cummings at the Glycomics Center at Emory University. The microbial glycan microarray was performed by Ryan McBride and James C. Paulson at the Scripps Research Institute. Microscopy images were obtained by Heather L. Hodges.

3.15 Acknowledgements

This research was supported by the US National Institutes of Health (NIH) (R01GM55984 and R01AI063596 (LLK)). I am thankful to the US National Science Foundation (NSF) and the NIH Chemistry–Biology Interface Training Program (T32 GM008505) for fellowships. K.W. was supported by a fellowship from the Development and Promotion of Science and Technology Talents Project of Thailand. The glycan array experiments were made possible by the Consortium for Functional Glycomics (NIGMS GM062116 and GM98791 (J.C.P.)), which supported the Glycan Array Synthesis Core at The Scripps Research Institute, La Jolla, CA and the Protein-Glycan Interaction Resource (Emory University School of Medicine, Atlanta, GA). These resources assisted with analysis of samples on the array. Printing and processing the furanoside array was supported through the National Center for Functional

Glycomics supported by NIGMS (P41GM103694 (R.D.C.)). Surface plasmon resonance experiments were performed at the UW–Madison Biophysics Instrumentation Facility, which is supported by UW-Madison, NSF grant BIR-9512577, and NIH grant S10 RR13790. Flow cytometry data were obtained at the UW-Madison Carbone Cancer Center (P30 CA014520) and microscopy images were acquired at the UW–Madison W.M. Keck Laboratory for Biological Imaging (1S10RR024715). The UW–Madison Chemistry NMR facility is supported by the NSF (CHE-9208463, CHE-9629688) and NIH (1s10 RR08389). Use of the Advanced Photon Source at the Argonne National Laboratory, was supported by the US Department of Energy (Contract DE-AC02-06CH11357) and the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). I thank Josh M. Fishman for assistance preparing the synthetic methods reported here, his help was instrumental in this work being published. I also would like to acknowledge Matthew R. Levengood, Adam H. Courtney, and Darrell R. McCaslin for thoughtful discussions. I also am grateful to Michele R. Richards (Univ. Alberta) for helpful discussions on furanoside conformations.

3.16 References

1. Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J. P., Heimburg-Molinaro, J., Ju, T., Molinaro, R. J., Rivera-Marrero, C., Xia, B., Smith, D. F., and Cummings, R. D. (2010) Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* **16**, 295-301
2. Vaishnava, S., Yamamoto, M., Severson, K. M., Ruhn, K. A., Yu, X., Koren, O., Ley, R., Wakeland, E. K., and Hooper, L. V. (2011) The antibacterial lectin RegIII γ promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255-258
3. Gallo, R. L., and Hooper, L. V. (2012) Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12**, 503-516
4. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920
5. Varki, A. (2009) *Essentials of glycobiology*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
6. Lis, H., and Sharon, N. (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637-674
7. Weis, W. I., and Drickamer, K. (1996) Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441-473
8. Turner, M. W. (2003) The role of mannose-binding lectin in health and disease. *Mol. Immunol.* **40**, 423-429
9. Fujita, T. (2002) Evolution of the lectin-complement pathway and its role in innate immunity. *Nat. Rev. Immunol.* **2**, 346-353
10. Jack, D. L., Klein, N. J., and Turner, M. W. (2001) Mannose-binding lectin: targeting the microbial world for complement attack and opsonophagocytosis. *Immunol. Rev.* **180**, 86-99
11. Thomsen, T., Schlosser, A., Holmskov, U., and Sorensen, G. L. (2011) Ficolins and FIBCD1: soluble and membrane bound pattern recognition molecules with acetyl group selectivity. *Mol. Immunol.* **48**, 369-381
12. Holmskov, U., Thiel, S., and Jensenius, J. C. (2003) Collectins and ficolins: Humoral lectins of the innate immune defense. *Annu. Rev. Immunol.* **21**, 547-578
13. Lehotzky, R. E., Partch, C. L., Mukherjee, S., Cash, H. L., Goldman, W. E., Gardner, K. H., and Hooper, L. V. (2010) Molecular basis for peptidoglycan recognition by a bactericidal lectin. *P Natl Acad Sci USA* **107**, 7722-7727
14. Lee, J. K., Buckhaults, P., Wilkes, C., Teilhet, M., King, M. L., Moremen, K. W., and Pierce, M. (1997) Cloning and expression of a *Xenopus laevis* oocyte lectin and characterization of its mRNA levels during early development. *Glycobiology* **7**, 367-372
15. Lee, J. K., Baum, L. G., Moremen, K., and Pierce, M. (2004) The X-lectins: a new family with homology to the *Xenopus laevis* oocyte lectin XL-35. *Glycoconjugate J.* **21**, 443-450
16. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19-34

17. Tsuji, S., Uehori, J., Matsumoto, M., Suzuki, Y., Matsuhisa, A., Toyoshima, K., and Seya, T. (2001) Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *J. Biol. Chem.* **276**, 23456-23463
18. French, A. T., Bethune, J. A., Knight, P. A., McNeilly, T. N., Wattedegera, S., Rhind, S., Miller, H. R., and Pemberton, A. D. (2007) The expression of intelectin in sheep goblet cells and upregulation by interleukin-4. *Vet. Immunol. Immunopathol.* **120**, 41-46
19. Voehringer, D., Stanley, S. A., Cox, J. S., Completo, G. C., Lowary, T. L., and Locksley, R. M. (2007) Nippostrongylus brasiliensis: identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. *Exp. Parasitol.* **116**, 458-466
20. Pemberton, A. D., Knight, P. A., Wright, S. H., and Miller, H. R. (2004) Proteomic analysis of mouse jejunal epithelium and its response to infection with the intestinal nematode, *Trichinella spiralis*. *Proteomics* **4**, 1101-1108
21. Datta, R., deSchoolmeester, M. L., Hedeler, C., Paton, N. W., Brass, A. M., and Else, K. J. (2005) Identification of novel genes in intestinal tissue that are regulated after infection with an intestinal nematode parasite. *Infect. Immun.* **73**, 4025-4033
22. Kerr, S. C., Carrington, S. D., Oscarson, S., Gallagher, M. E., Solon, M., Yuan, S., Ahn, J. N., Dougherty, R. H., Finkbeiner, W. E., Peters, M. C., and Fahy, J. V. (2014) Intelectin-1 is a prominent protein constituent of pathologic mucus associated with eosinophilic airway inflammation in asthma. *Am. J. Respir. Crit. Care Med.* **189**, 1005-1007
23. Kuperman, D. A., Lewis, C. C., Woodruff, P. G., Rodriguez, M. W., Yang, Y. H., Dolganov, G. M., Fahy, J. V., and Erle, D. J. (2005) Dissecting asthma using focused transgenic modeling and functional genomics. *J. Allergy Clin. Immunol.* **116**, 305-311
24. Suzuki, Y. A., Shin, K., and Lonnerdal, B. (2001) Molecular cloning and functional expression of a human intestinal lactoferrin receptor. *Biochemistry* **40**, 15771-15779
25. Tsuji, S., Tsuura, Y., Morohoshi, T., Shinohara, T., Oshita, F., Yamada, K., Kameda, Y., Ohtsu, T., Nakamura, Y., and Miyagi, Y. (2010) Secretion of intelectin-1 from malignant pleural mesothelioma into pleural effusion. *Br. J. Cancer* **103**, 517-523
26. de Souza Batista, C. M., Yang, R. Z., Lee, M. J., Glynn, N. M., Yu, D. Z., Pray, J., Ndubuizu, K., Patil, S., Schwartz, A., Kligman, M., Fried, S. K., Gong, D. W., Shuldiner, A. R., Pollin, T. I., and McLenithan, J. C. (2007) Omentin plasma levels and gene expression are decreased in obesity. *Diabetes* **56**, 1655-1661
27. Tsuji, S., Yamashita, M., Nishiyama, A., Shinohara, T., Zhongwei, U., Myrvik, Q. N., Hoffman, D. R., Henriksen, R. A., and Shibata, Y. (2007) Differential structure and activity between human and mouse intelectin-1: Human intelectin-1 is a disulfide-linked trimer, whereas mouse homologue is a monomer. *Glycobiology* **17**, 1045-1051
28. Pedersen, L. L., and Turco, S. J. (2003) Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. *Cell. Mol. Life Sci.* **60**, 259-266
29. Nassau, P. M., Martin, S. L., Brown, R. E., Weston, A., Monsey, D., McNeil, M. R., and Duncan, K. (1996) Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *J. Bacteriol.* **178**, 1047-1052
30. Wesener, D. A., May, J. F., Huffman, E. M., and Kiessling, L. L. (2013) UDP-galactopyranose mutase in nematodes. *Biochemistry* **52**, 4391-4398

31. Bishop, J. R., and Gagneux, P. (2007) Evolution of carbohydrate antigens--microbial forces shaping host glycomes? *Glycobiology* **17**, 23R-34R
32. Herget, S., Toukach, P. V., Ranzinger, R., Hull, W. E., Knirel, Y. A., and von der Lieth, C. W. (2008) Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC Struct. Biol.* **8**, 35
33. Adibekian, A., Stallforth, P., Hecht, M. L., Werz, D. B., Gagneux, P., and Seeberger, P. H. (2011) Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chem. Sci.* **2**, 337-344
34. Mann, D. A., Kanai, M., Maly, D. J., and Kiessling, L. L. (1998) Probing low affinity and multivalent interactions with surface plasmon resonance: Ligands for concanavalin A. *J. Am. Chem. Soc.* **120**, 10575-10582
35. Kiessling, L. L., and Grim, J. C. (2013) Glycopolymer probes of signal transduction. *Chem. Soc. Rev.* **42**, 4476-4491
36. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C. H., and Paulson, J. C. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *P Natl Acad Sci USA* **101**, 17033-17038
37. Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zarling, L. C., Splain, R. A., Smith, D. F., Cummings, R. D., Paulson, J. C., Forest, K. T., and Kiessling, L. L. (2015) Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.* **22**, 603-610
38. Tefsen, B., Ram, A. F., van Die, I., and Routier, F. H. (2012) Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology* **22**, 456-469
39. Stowell, S. R., Arthur, C. M., McBride, R., Berger, O., Razi, N., Heimburg-Molinaro, J., Rodrigues, L. C., Gourdine, J. P., Noll, A. J., von Gunten, S., Smith, D. F., Knirel, Y. A., Paulson, J. C., and Cummings, R. D. (2014) Microbial glycan microarrays define key features of host-microbial interactions. *Nat. Chem. Biol.* **10**, 470-476
40. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-549
41. Garlatti, V., Belloy, N., Martin, L., Lacroix, M., Matsushita, M., Endo, Y., Fujita, T., Fontecilla-Camps, J. C., Arlaud, G. J., Thielens, N. M., and Gaboriaud, C. (2007) Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J.* **26**, 623-633
42. Taha, H. A., Richards, M. R., and Lowary, T. L. (2013) Conformational analysis of furanoside-containing mono- and oligosaccharides. *Chem. Rev.* **113**, 1851-1876
43. Richards, M. R., Bai, Y., and Lowary, T. L. (2013) Comparison between DFT- and NMR-based conformational analysis of methyl galactofuranosides. *Carbohydr. Res.* **374**, 103-114
44. Altona, C., and Sundaral.M. (1972) Conformational-Analysis of Sugar Ring in Nucleosides and Nucleotides - New Description Using Concept of Pseudorotation. *J. Am. Chem. Soc.* **94**, 8205-&

45. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-W324
46. Krissinel, E. (2012) Enhanced fold recognition using efficient short fragment clustering. *J. Mol. Biochem.* **1**, 76-85
47. Angata, T., and Varki, A. (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem. Rev.* **102**, 439-469
48. Varghese, J. N., Mckimmbreschkin, J. L., Caldwell, J. B., Kortt, A. A., and Colman, P. M. (1992) The Structure of the Complex between Influenza-Virus Neuraminidase and Sialic-Acid, the Viral Receptor. *Proteins* **14**, 327-332
49. Kraschnefski, M. J., Bugarcic, A., Fleming, F. E., Yu, X., von Itzstein, M., Coulson, B. S., and Blanchard, H. (2009) Effects on sialic acid recognition of amino acid mutations in the carbohydrate-binding cleft of the rotavirus spike protein. *Glycobiology* **19**, 194-200
50. Blanchard, H., Yu, X., Coulson, B. S., and von Itzstein, M. (2007) Insight into host cell carbohydrate-recognition by human and porcine rotavirus from crystal structures of the virion spike associated carbohydrate-binding domain (VP8*). *J. Mol. Biol.* **367**, 1215-1226
51. Dormitzer, P. R., Sun, Z. Y. J., Wagner, G., and Harrison, S. C. (2002) The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO. J.* **21**, 885-897
52. Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S. J., Skehel, J. J., and Wiley, D. C. (1992) Binding of Influenza-Virus Hemagglutinin to Analogs of Its Cell-Surface Receptor, Sialic-Acid - Analysis by Proton Nuclear-Magnetic-Resonance Spectroscopy and X-Ray Crystallography. *Biochemistry* **31**, 9609-9621
53. Cartwright, K. (2002) Pneumococcal disease in western Europe: burden of disease, antibiotic resistance and management. *Eur. J. Pediatr.* **161**, 188-195
54. Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kalltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**, 262-269
55. Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J. R., Elvin, L., Ensor, K. M., Hackell, J., Siber, G., Malinoski, F., Madore, D., Chang, I., Kohberger, R., Watson, W., Austrian, R., Edwards, K., and Vaccine, N. C. K. P. (2000) Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. *Pediatr. Infect. Dis. J.* **19**, 187-195
56. Arnold, R. R., Cole, M. F., and McGhee, J. R. (1977) A bactericidal effect for human lactoferrin. *Science* **197**, 263-265
57. Alexander, D. B., Iigo, M., Yamauchi, K., Suzui, M., and Tsuda, H. (2012) Lactoferrin: an alternative view of its role in human biological fluids. *Biochem. Cell. Biol.* **90**, 279-306
58. Arnold, R. R., Russell, J. E., Champion, W. J., and Gauthier, J. J. (1981) Bactericidal activity of human lactoferrin: influence of physical conditions and metabolic state of the target microorganism. *Infect. Immun.* **32**, 655-660
59. Effros, R. M., and Chinard, F. P. (1969) The in vivo pH of the extravascular space of the lung. *J. Clin. Invest.* **48**, 1983-1996

60. Evans, D. F., Pye, G., Bramley, R., Clark, A. G., Dyson, T. J., and Hardcastle, J. D. (1988) Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut* **29**, 1035-1041
61. Asensio, J. L., Arda, A., Canada, F. J., and Jimenez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46**, 946-954
62. Pemberton, A. D., Rose-Zerilli, M. J., Holloway, J. W., Gray, R. D., and Holgate, S. T. (2008) A single-nucleotide polymorphism in intelectin 1 is associated with increased asthma risk. *J. Allergy Clin. Immunol.* **122**, 1033-1034
63. Barrett, J. C., Hansoul, S., Nicolae, D. L., Cho, J. H., Duerr, R. H., Rioux, J. D., Brant, S. R., Silverberg, M. S., Taylor, K. D., Barmada, M. M., Bitton, A., Dassopoulos, T., Datta, L. W., Green, T., Griffiths, A. M., Kistner, E. O., Murtha, M. T., Regueiro, M. D., Rotter, J. I., Schumm, L. P., Steinhart, A. H., Targan, S. R., Xavier, R. J., Consortium, N. I. G., Libioulle, C., Sandor, C., Lathrop, M., Belaiche, J., Dewit, O., Gut, I., Heath, S., Laukens, D., Mni, M., Rutgeerts, P., Van Gossom, A., Zelenika, D., Franchimont, D., Hugot, J. P., de Vos, M., Vermeire, S., Louis, E., Belgian-French, I. B. D. C., Wellcome Trust Case Control, C., Cardon, L. R., Anderson, C. A., Drummond, H., Nimmo, E., Ahmad, T., Prescott, N. J., Onnie, C. M., Fisher, S. A., Marchini, J., Ghori, J., Bumpstead, S., Gwilliam, R., Tremelling, M., Deloukas, P., Mansfield, J., Jewell, D., Satsangi, J., Mathew, C. G., Parkes, M., Georges, M., and Daly, M. J. (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955-962
64. Schnaitman, C. A., and Klena, J. D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**, 655-682
65. Willis, L. M., Stupak, J., Richards, M. R., Lowary, T. L., Li, J., and Whitfield, C. (2013) Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. *P Natl Acad Sci USA* **110**, 7868-7873
66. Song, X., Lasanajak, Y., Xia, B., Smith, D. F., and Cummings, R. D. (2009) Fluorescent glycosylamides produced by microscale derivatization of free glycans for natural glycan microarrays. *ACS Chem. Biol.* **4**, 741-750
67. Heimburg-Molinaro, J., Song, X., Smith, D. F., and Cummings, R. D. (2011) Preparation and analysis of glycan microarrays. *Curr. Protoc. Protein Sci.* **Chapter 12**, Unit12 10
68. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. in *Methods Enzymol.* (C.W. Carter, J., and Sweet, R. M. eds.), Academic Press New York. pp 307-326
69. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213-221
70. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674

71. Afonine, P. V., Grosse-Kunstleve, R. W., and Adams, P. D. (2005) The Phenix refinement framework. *CCP4 Newsl.* **42**, 8
72. Schuttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 1355-1363
73. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 486-501
74. Chen, V. B., III, W. B. A., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 12-21

Chapter 4

Conserved Structure and Ligand Binding Mechanism Between Human Intelectin-1 and *Xenopus* Embryonic Epidermal Lectin

Portions of this chapter have been published in:

Wangkanont, W., Wesener, D. A., Vidani, J. A., Kiessling, L. L., and Forest, K. T. Structures of *Xenopus* Embryonic Epidermal Lectin Reveal a Conserved Mechanism of Microbial Glycan Recognition. *The Journal of Biological Chemistry*. **2016**. 291 (11), 5596-5610.

&

Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zharling, L. C., Splain, R. S., Smith D. F., Cummings, R. D., Paulson, J. C., Forest, K. T., and Kiessling, L. L. Recognition of microbial glycans by human intelectin-1. *Nature Structural and Molecular Biology*. **2015**. 22, 603-610.

4.1 Abstract

Intelectins (X-type lectins) are broadly distributed throughout chordates, and they have been implicated in innate immunity. *Xenopus laevis* embryonic epidermal lectin (XEEL), an intelectin secreted into environmental water by the *X. laevis* embryo, is postulated to function in defense against microbes. XEEL is homologous (64% identical) to hIntL-1, which is also implicated in innate immune defense. We previously demonstrated that hIntL-1 binds microbial glycans bearing exocyclic vicinal diol groups. It is unknown whether XEEL has the same ligand specificity. Also unclear is whether XEEL and hIntL-1 have similar quaternary structures, as XEEL lacks the corresponding cysteine residues in hIntL-1 that stabilize a disulfide-linked trimer. These observations prompted us to further characterize XEEL. We found that hIntL-1 and XEEL have similar structural features. Even without the corresponding intermolecular disulfide bonds present in hIntL-1, the CRD of XEEL (XEEL_{CRD}) forms a stable trimer in solution. The structure of XEEL_{CRD} in complex with Gro-P, a residue present in microbe-specific glycans, indicated the exocyclic vicinal diols coordinate to a protein-bound calcium ion. This ligand-binding mode is conserved between XEEL and hIntL-1. These data support a role for XEEL in innate immunity, and they highlight structural and functional conservation of X-type lectins among chordates.

4.2 Introduction

The carbohydrate coats of cells can be used by the immune system to distinguish self from non-self (1). Lectins can play a role in this discrimination because they have the ability to distinguish among carbohydrate residues. Several families of lectins, including the C-type lectins and ficolins, are classified as immune lectins. Their occurrence throughout the animal kingdom suggests carbohydrate recognition is critical in immunity (2,3). Recently, intelectins, also known as X-type lectins (4), have been proposed to function in immunity (5). This suggestion is intriguing, but little is known about the structure or function of this class of lectins.

Intelectins have been identified in diverse species from tunicates to humans (6). The limited functional studies conducted to date suggest a role in host defense. For example, tunicate intelectin can serve as an opsonin for phagocytosis by hemocytes (7). Amphioxus intelectin can agglutinate bacteria *in vitro*, suggesting it could sequester them from the organism (8,9). In some species, intelectin expression levels increase upon infection. Specifically, in rainbow trout and zebrafish, intelectins are upregulated upon microbial infection (10-12). Sheep and mice increase intelectin production upon infection with intestinal parasitic nematodes (13-15). The upregulation of intelectins upon exposure to microbes supports the hypothesis that these proteins function in animal immune defense.

The first intelectin was discovered in *Xenopus laevis* (16-18). At least five *X. laevis* intelectins have been subsequently described. *X. laevis* cortical granule lectins (XCGL-1 and XCGL-1) are found in oocytes and developing embryos (19), where they facilitate the formation of the fertilization envelope to block polyspermy (16,17). *X. laevis* serum lectins (XSL-1 and XSL-2) are transcribed in response to LPS exposure (20,21). *X. laevis* embryonic epidermal

lectin (XEEL) is secreted during various stages of development, but its level is highest around hatching (22,23). It is produced by the goblet cells of the larval epidermis, which has properties akin to those of the human mucosal epithelia (24). Human goblet cells are critical for preserving the epithelial barrier in mucosal tissue and therefore perform a critical role in microbial defense. These parallels suggest XEEL may function in innate immunity in *Xenopus*.

Humans encode two intelectins that have similarities to XEEL: hIntL-1 and hIntL-2 (25). These human lectins are expressed at mucosal barriers, including the lung and the intestine. The recognition properties of hIntL-1 are consistent with a role for this lectin in innate immunity. Specifically, it was suggested that hIntL-1 recognizes furanose carbohydrate residues, including Gal β and ribose (5). Our studies of hIntL-1 indicate that it does not bind ribose, but it does interact with β -Gal β residues (26) and is thus not a general furanose-binding lectin. Analysis using glycan microarrays revealed that hIntL-1 does not bind any of the tested human glycans, but rather binds diverse microbial glycan epitopes. Specifically, hIntL-1 can interact with β -Gal β , GroP, heptoses, KO, and KDO (26). None of these epitopes are present on human glycans. In accord with its glycan specificity, hIntL-1 binds microbial cell surfaces, supporting a role for hIntL-1 in the recognition of intact microbes (26). Given these recognition properties and the presence of intelectins in diverse species, an intriguing possibility is that the intelectins serve as microbial detectors.

XEEL and other intelectins share high sequence similarity, including a putative conserved FBD. Although this domain is also found in ficolins (6), it comprises a mere 45 of the 300 amino acid residues in intelectins. Experiments suggested that intelectin carbohydrate binding is calcium ion-dependent. C-type lectins typically use calcium ions to bind their

carbohydrate ligands, but the intelectins are not members of the C-type lectin family (16,25,27). We recently used protein X-ray crystallography to determine the structure of apo-hIntL-1 and hIntL-1 in complex with a β -Gal β glycoside (26). Our results indicate that hIntL-1 binds the exocyclic vicinal diol present in β -Gal β through direct calcium ion coordination. This exocyclic vicinal diol is the common feature of hIntL-1 ligands. Still, whether all hIntL-1 ligands bind similarly, and whether intelectins from different species use a similar recognition mode was not known. In addition, whether different intelectins adopt different quaternary states is unclear. Because XEEL is secreted into environmental water, we hypothesized that this protein is stable and therefore would be conducive to crystallographic studies.

Here, we describe the biophysical properties and three-dimensional structure of the CRD of XEEL (XEEL_{CRD}) with and without a bound GroP ligand. The structure of the complex of XEEL_{CRD} and GroP has striking similarities to that of hIntL-1 bound to β -Gal β . The mechanism of ligand recognition, direct calcium ion coordination, is conserved. Though it is missing the intermolecular disulfide bonds present in hIntL-1, the XEEL_{CRD} is trimeric in solution. In addition to highlighting that hIntL-1 and XEEL have a conserved structure and ligand-binding mechanism, the XEEL_{CRD} structure provides the means to identify and compare functional residues across species. We anticipate that the data will be useful as research into the ligand specificity and biological functions of intelectins expands.

4.3 Expression and Purification of XEEL

During our attempt to solve the protein crystal structure of hIntL-1 (**Chapter 3**) (26), the dearth of previously determined structures with a similar architecture became apparent because of difficulty in phasing crystal diffraction data collected from a hIntL-1 crystal using molecular

replacement. In an attempt to circumvent this problem, we began biochemically exploring the XEEL lectin from *X. laevis*. For XEEL, we employed an insect cell expression system, both because of its ability to yield the large quantities of protein needed for biophysical characterization, but also because it can be used to incorporate L-SeMet into the protein for crystallographic structure determination. The expression construct encoded full-length XEEL with an N-terminal *Strep-tag II*[®] purification tag sequence linked to the predicted signal peptide (28). Cleavage of the signal peptide should yield XEEL with a two amino acid linker followed by an eight amino acid Strep-tag II at the N-terminus. Secreted XEEL was expressed robustly in insects cells using both full culture medium, and methione dropout medium supplemented with SeMet every 24 hours.

During our attempts to purify XEEL, several important results led to the ultimate success of XEEL purification. The first was that a dialysis step was required to remove a small molecule precipitate from the conditioned culture medium that was interfering with the purification protocol. We believe the culprit was L-glutamine. During the dialysis step, we observed a significant amount of precipitation when the dialysis buffer was raised to a pH greater than 7.0. SDS-PAGE analysis revealed that a substantial portion of this precipitate was XEEL protein. When we attempted to purify XEEL at a pH greater than 7.0, a similar problem with precipitation was observed. This resulted in a low yield of soluble protein for biochemical experimentation. This two results prompted us to attempt the entire dialysis and purification at an acidic pH, 6.7. At pH = 6.7 there was notably less precipitate, and a large amount of protein could be measured in the elution from *Strep-Tactin*[®] purification resin (**Figure 4-1A**) (29). When the elution was cooled and concentrated via centrifugation, large sheet-like crystals formed

(Figure 4-1B). These crystals increase in size and quantity during incubation of the concentrated protein solution. The crystallized XEEL was collected and washed with buffer to yield protein > 95 % pure and an apparent molecular weight under reducing conditions of ~35 kDa by SDS-PAGE. Fortuitously, these crystals could be redissolved with the addition of calcium chloride. Both the native and SeMet labeled protein behaved this way. This finding was instrumental to the success of our intelectin project.

Analysis of recrystallized XEEL by mass spectrometry indicated a truncation of at least 30 N-terminal residues had occurred. N-terminal amino acid sequencing revealed secondary proteolysis yielded a protein product corresponding to residues 54 to 342 of XEEL, XEEL_{CRD}. The XEEL_{CRD} lacks the *Strep*-tag II purification tag, yet the protein could be purified using a *Strep*-tactin resin. We postulate that the lectin has modest affinity for agarose resin and was eluted by the EDTA included in the elution buffer.

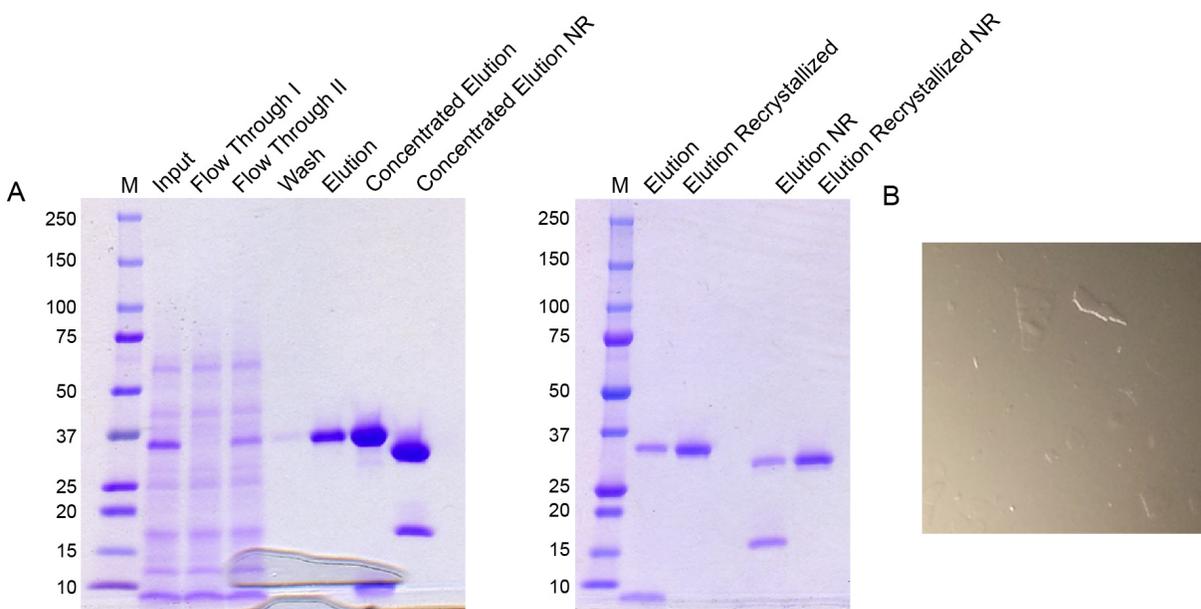


Figure 4-1. Purification of XEEL. (A) SDS-PAGE analysis of a typical XEEL purification performed at pH = 6.7. Elution from the column yields one major contaminant that can be removed with XEEL recrystallization and washing, yielding a protein > 95 % pure. M denotes protein molecular weight marker; NR denotes the sample was processed in the absence of reducing agents. (B) Image of the XEEL crystals that form during purification and concentration. Image was acquired at 5x magnification.

4.4 Three-dimensional Structure of XEEL_{CRD}

To analyze the ligand-binding site of XEEL and to compare the oligomeric state of XEEL_{CRD} to that of hIntL-1, we used X-ray crystallography to determine the structure of XEEL_{CRD}. For phasing purposes, we produced SeMet-labeled XEEL_{CRD}. MS analysis indicated that Se-Met was substituted at all seven methionine residues. Native and SeMet-labeled XEEL_{CRD} crystallized under similar conditions. The structure of SeMet-XEEL was solved by Se-SAD (with 67.5% solvent content) and refined to 2.3 Å resolution (**Table 4-1**). The six molecules in each asymmetric unit form two trimers (**Figure 4-2A**), related to one another by a

pseudotranslation vector of 55.6 Å approximately parallel to the b axis. While the C-terminus is clearly resolved in the electron density, the first resolvable residue on the N-terminus is G66. The trimeric carbohydrate recognition domains of XEEL and hIntL-1 are highly related, each possesses a globular subunit that does not fall into a well-known structural family. Both have a fibrinogen-like lobe that contains a split and twisted 7-stranded β -sheet. The intelectin-specific region is characterized by a large fraction of random coil and a 3-stranded curved β -sheet. The domains are encircled by 12 short solvent-exposed α -helical stretches (**Figure 4-2A**).

Three well-ordered calcium ions are a defining feature of the intelectin-specific region of XEEL_{CRD}. Two calcium ions are buried deep in the protein core and are presumed to play a structural role (**Figure 4-2B**). The structural calcium ion closer to the protein surface is heptacoordinated. Calcium ion ligands include two distal waters and five side chain interactions: the carboxylates of E116 and D127 (both oxygen atoms) and the backbone carbonyls of N118 and G121. N117 forms additional hydrogen bonds to calcium-bound water molecules through both its side chain and backbone amide nitrogen. The structural calcium ion located farther inside the trimeric protein is coordinated directly by the carboxylates of D162 (both oxygen atoms) and D311, backbone amide carbonyls of H115 and G126, and two water molecules. Finally, the H115 side chain also forms a hydrogen bond with a calcium-bound water molecule. No single amino acid or water molecule coordinates both calcium ions directly.

The remaining calcium ion is positioned analogously to that in the β -Gal β -bound hIntL-1 structure (PDB ID: 4WMY) (26) (**Figure 4-2C**). This congruence suggests that this ion also marks the carbohydrate recognition pocket in XEEL_{CRD}. This calcium ion is directly coordinated by N289, E291, E303, and four ordered water molecules. The pocket geometry is determined by

a network of 16 hydrogen bonds distributed amongst the four calcium-bound water molecules and the aforementioned residues, S272, E273, and H292. Side chains from W306 and W326 form the back walls of this conserved binding site. The structural similarity of this region of XEEL and hIntL-1 suggest that these lectins may recognize similar glycans.

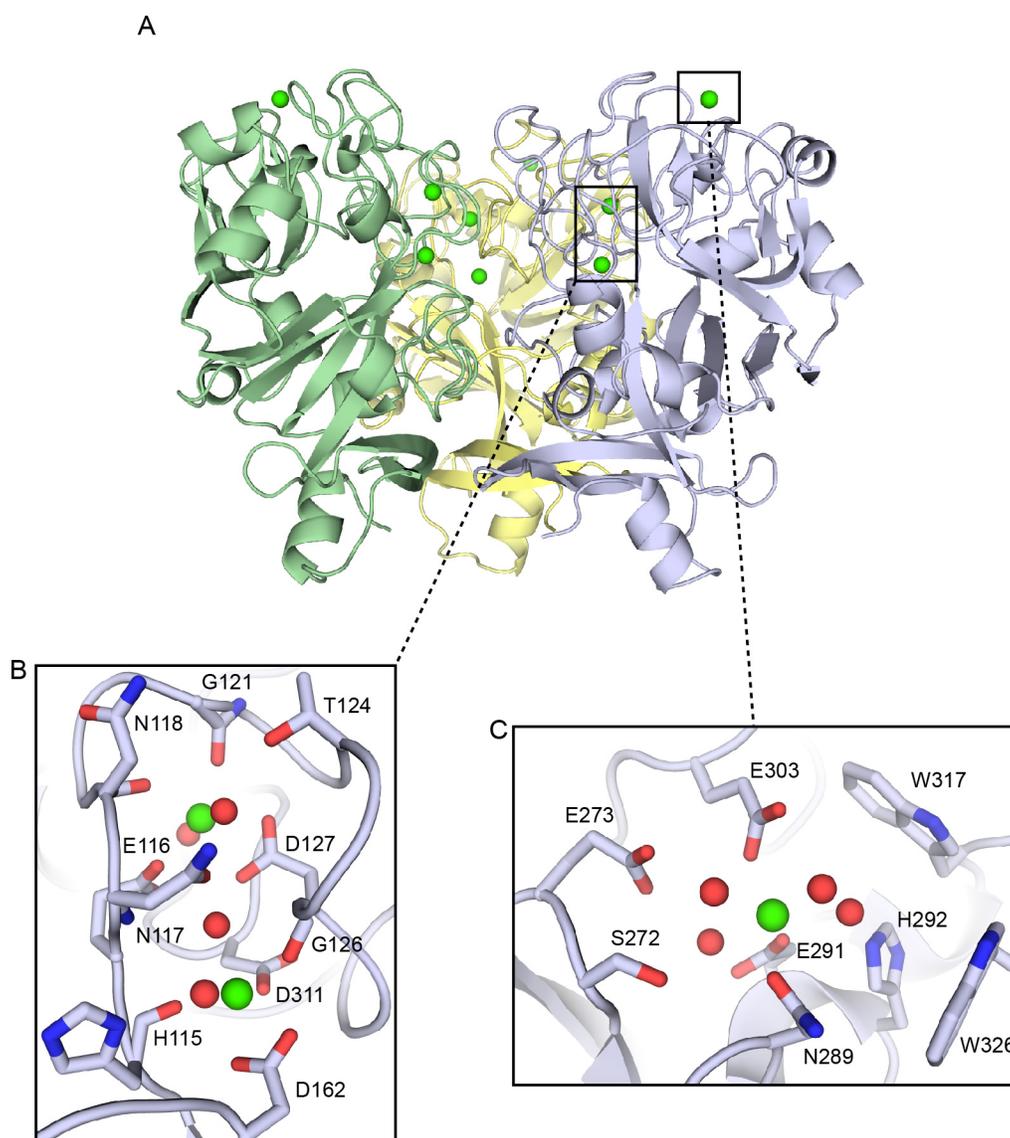


Figure 4-2. Three-dimensional structure of XEEL_{CRD} trimer. (A) Crystal structure of SeMet-labeled XEEL_{CRD} trimer oriented with the intelectin-specific domains toward the top of the figure and the FBG-like lobes below. The second trimer in the asymmetric unit is removed for clarity. (B) Structural calcium site with two calcium ions (green) and four ordered water

molecules (red). (C) Ligand binding site with one calcium ion (green) and four ordered water molecules (red).

4.5 XEEL_{CRD} Binds Exocyclic 1,2-diol Containing Ligands

The human homolog of XEEL, hIntL-1, interacts with a range of glycan epitopes, each of which possesses a terminal 1,2-diol (i.e., a diol of the form RCH(OH)-CH₂OH) (26). To address whether this ligand specificity is conserved between XEEL and hIntL-1, we used SPR to monitor XEEL_{CRD} ligand binding properties (**Figure 4-3A**). XEEL_{CRD} displays concentration-dependent binding to immobilized β -Gal f . The on- and off-rates are rapid, suggesting minimal structural reorganization of XEEL_{CRD} occurs upon ligand binding. As a measure of the affinity of XEEL_{CRD} for immobilized β -Gal f , we used an equilibrium binding model to analyze the SPR data (**Figure 4-3B**). The apparent affinity of the XEEL_{CRD} trimer for β -Gal f is $4.1 \pm 0.5 \mu\text{M}$. No specific binding to β -Gal f or β -Rib f was observed. These results suggest that, like hIntL-1, XEEL does not bind generally to furanoside ligands, but instead recognizes the exocyclic 1,2-diol epitope of β -Gal f .

To test whether the exocyclic vicinal diol on β -Gal f is the epitope recognized by XEEL_{CRD}, we conducted competition binding studies with the simplest 1,2-terminal diol, glycerol. Gro inhibits XEEL_{CRD} binding to immobilized β -Gal f , indicating it is a competitive inhibitor (**Figure 4-3C**). These experiments suggest that XEEL binds both β -Gal f and GroP through their common exocyclic vicinal diol epitope. From the inhibition data, we determined the IC₅₀ of glycerol to be $0.6 \pm 0.6 \text{ mM}$. The Gro inhibition data fit well to a one-site competition model. Thus, the determined IC₅₀ value serves as an estimate of the K_d of an exocyclic vicinal diol-containing compound for a single XEEL binding site.

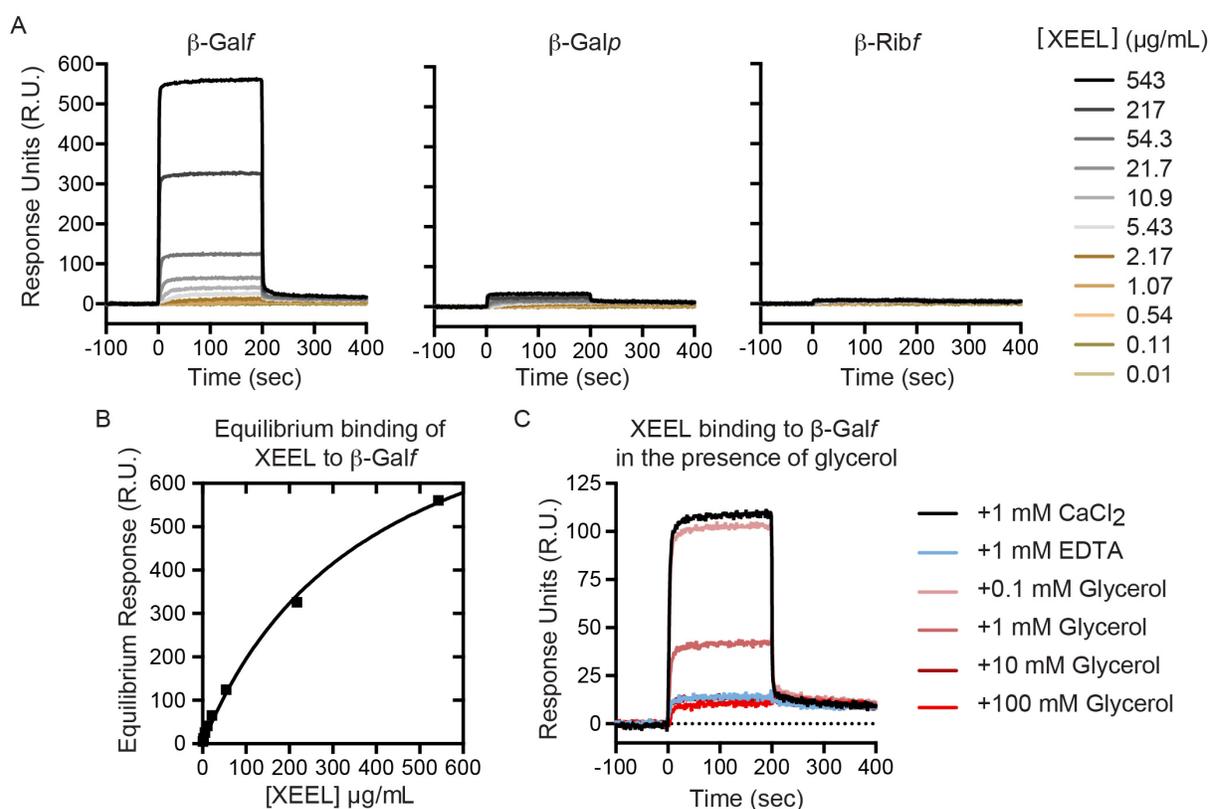


Figure 4-3. Binding of XEEL_{CRD} to immobilized carbohydrates as measured by SPR. (A) SPR sensorgrams of XEEL_{CRD} binding to immobilized carbohydrates. (B) Equilibrium binding analysis of XEEL_{CRD} to immobilized β -Gal f . These data were used to determine the apparent affinity of XEEL_{CRD} for β -Gal f . (C) Gro was assayed for its ability to inhibit XEEL_{CRD} binding to β -Gal f . These data were used to determine the IC₅₀ of Gro. This experiment is representative of two independently purified protein samples.

4.6 Human and *Xenopus* Intelectins use Similar Ligand-recognition Modes

To compare ligand recognition of XEEL and hIntL-1, we co-crystallized the XEEL with GroP. Alignment of Se-Met XEEL_{CRD} and GroP-bound XEEL_{CRD} yields an RMSD value of 0.17 Å over 276 C α atoms, suggesting no drastic structural changes occur upon either SeMet labeling or ligand binding. Specifically, no significant movement of residues in the vicinity of the ligand-binding site is observed upon XEEL complexation to GroP.

The structure reveals two hydroxyl groups from GroP coordinate directly to a protein bound calcium ion (**Figure 4-4A**). Upon binding, these hydroxyls displace two ordered water molecules in the ligand-binding site. This mode of ligand binding is nearly identical to what was observed in the structure of the hIntL-1- β -Gal β complex (**Figure 4-4B**). In XEEL, the diol moiety of the GroP binds in a pocket formed by the side chains of W317 and W326. W326 serves as an aromatic wall: it is positioned such that it allows binding of a diol terminating with a hydroxymethylene (-CH₂OH) group, but not a more substituted diol. A similar binding mode is found for hIntL-1, but the human protein possesses a tyrosine residue in place of W326 in XEEL.

The indole NH of W326 forms a hydrogen bond with the phosphate group of GroP. Electron density around the phosphate group reveals elongation toward the indole NH of W317, suggesting dynamic equilibration of hydrogen bonding between W317 and W326, and the phosphate group. The primary hydroxyl of GroP forms an additional hydrogen bond with H292. In addition to calcium coordination, E303 forms a hydrogen bond with the secondary hydroxyl group of GroP. These additional hydrogen-bonding interactions enhance the ability of exocyclic diol hydroxyl groups to serve as ligands for the calcium ion.

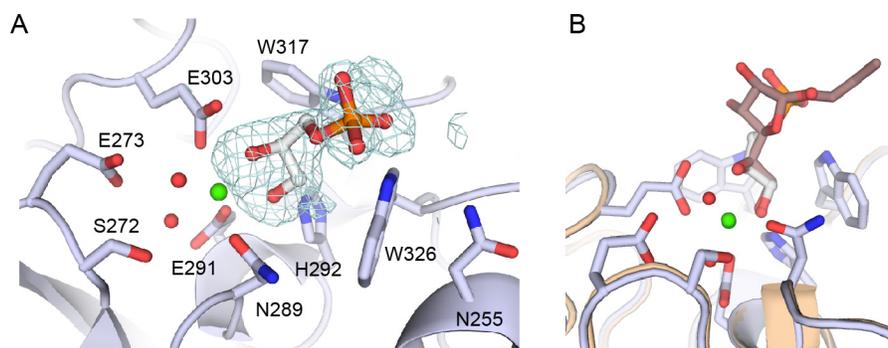


Figure 4-4. Crystal structure of XEEL_{CRP}-GroP complex. (A) Calcium ion and ordered water molecules are shown as green and red spheres, respectively. The mesh represents the 3 σ contour in a difference electron density map ($mF_o - DF_c$, where m is figure of merit, and D is σA weight)

after refinement without ligand. (B) Alignment of XEEL_{CRD}-GroP complex with hIntL-1- β -Gal α complex (PDB ID: 4WMY) (26). hIntL-1 is shown in wheat and the β -Gal α is shown with carbon atoms colored brown. Both ligand binding calcium ions are shown as green spheres. The ordered water molecules and proteins side chains are from the XEEL_{CRD} structure.

4.7 Intelectin Structures Provide Insight into Residue Conservation and Function

The availability of the XEEL and hIntL-1 structures allow assignment of residue function to the broader class of intelectins. We aligned the sequences of intelectins from humans and a range of model organisms, using Clustal W2 (**Figure 4-5**) (30). The structural calcium sites are highly conserved. The only exception is the amino acid corresponding to N118 of XEEL, which coordinates the calcium ion through its backbone amide carbonyl. The majority of the structural calcium site residues cluster between amino acids 115-127, with the consensus sequence of HENXXXGXCTXGD. Cysteine C123 is conserved in the intelectin fold but has no role in calcium ion binding. Though not present in the consensus sequence, aspartates 162 and 311 are also conserved residues that ligate the structural calcium ion. The high conservation of amino acids in this region suggests that most, if not all, intelectins contain two structural calcium ions.

Compared to the structural calcium ion site, more amino acid variation is found in the ligand binding site (**Figure 4-5**). Among the five *X. laevis* intelectins, residues that directly coordinate the ligand binding site calcium ion are completely conserved. W317 and W326, which form a box around the vicinal diol ligand, are conserved among XEEL, XSL-1, and XSL-2, all of which are proposed to participate in innate immunity. These observations suggest that XSL-1 and XSL-2 also can recognize the terminal 1,2-diols prevalent in microbial glycans. By contrast, the corresponding residues in XCGL-1 and XCGL-2 are phenylalanine and asparagine, respectively. XCGL-1 and XCGL-2 are involved in fertilization membrane formation (16,17), and they therefore likely recognize self carbohydrate epitopes. Indeed, glycan array screening

data suggests that XCGL-1 binds Gal- α (1-3)-GalNAc (CFG, primscreen_758). This disaccharide is not specific to microbes.

		115	116	117	118	121	124	126	127	162	272	273	289	291	292	303	311	317	326																																						
XEEL	(BC087616)	H	E	N	N	M	F	G	K	C	T	V	G	D	D	D	Y	N	S	E	R	C	N	V	E	H	H	P	E	G	S	P	R	Q	C	G	D	F	A	A	L	D	W	D	G	Y	G	T	N	L	G	W	S
XCGL-1	(X82626)	H	E	N	N	M	A	G	K	C	T	I	G	D	D	D	Y	N	T	E	K	C	N	V	E	H	V	P	E	A	D	P	R	Q	C	G	D	F	A	A	Y	D	F	N	G	Y	G	T	K	K	F	N	S
XCGL-2	(BF232570)	H	E	N	Y	L	A	G	K	C	T	V	G	D	D	D	Y	N	T	E	K	C	N	A	E	H	V	P	E	A	D	P	R	Q	C	G	D	F	A	A	Y	D	F	N	G	Y	G	T	K	K	F	N	S
XSL-1	(AB061238)	H	E	N	N	M	F	G	K	C	T	V	G	D	D	D	Y	N	A	E	K	C	N	T	E	H	H	A	E	G	N	P	K	Q	C	G	D	F	T	G	F	D	W	D	G	Y	G	T	H	Q	D	W	S
XSL-2	(AB061239)	H	E	N	N	M	F	G	K	C	T	V	G	D	D	D	Y	N	T	E	R	C	N	A	E	H	H	P	E	G	N	P	V	Q	C	G	D	F	A	A	F	D	W	N	G	Y	G	T	G	Y	A	W	S
hIntL-1	(BC020664)	H	E	N	D	M	R	G	K	C	T	V	G	D	D	D	Y	N	N	E	R	C	N	T	E	H	H	P	E	A	S	P	Q	Q	C	G	D	F	S	G	F	D	W	S	G	Y	G	T	H	V	G	Y	S
hIntL-2	(AY358905)	H	E	N	D	M	R	G	K	C	T	V	G	D	D	D	Y	N	N	E	R	C	N	T	E	H	H	P	Q	G	K	P	R	Q	C	G	D	F	S	A	F	D	W	D	G	Y	G	T	H	V	K	S	S
mIntL-1	(AAU88049)	H	E	N	N	M	R	G	K	C	T	V	G	D	D	D	Y	N	N	E	R	C	N	T	E	H	H	P	E	G	N	P	V	Q	C	G	D	F	A	S	F	D	W	D	G	Y	G	T	H	N	G	Y	S
mIntL-2	(AAO60215)	H	E	N	N	L	R	G	R	C	T	V	G	D	D	D	Y	N	N	E	R	C	N	T	E	H	H	P	E	F	D	P	E	E	C	G	D	F	A	A	F	D	A	N	G	Y	G	T	H	I	R	Y	S
DrIntL-1	(ACC62155)	H	E	N	N	I	H	G	S	C	T	V	G	D	D	D	Y	N	N	Y	Q	C	Y	A	Q	Y	Y	P	N	-	-	-	T	F	C	S	D	F	A	Y	L	N	A	N	-	G	A	T	N	G	Y	G	
DrIntL-2	(ACC62157)	H	E	N	N	M	Y	G	K	C	T	V	G	D	D	D	Y	N	T	E	K	C	H	T	E	Y	F	P	E	G	V	P	K	Q	C	G	D	F	T	G	F	D	W	N	G	Y	G	T	N	T	G	N	S

Figure 4-5. Sequence alignment of XEEL and other intelectins. Intelectins proteins include proteins from humans (hIntL-1 and hIntL-2), common model organisms, and *Xenopus laevis* (XCGL-1, XCGL-2, XSL-1, and XSL-2). mIntL-1 and mIntL-2 are derived from mouse, DrIntL-1 and DrIntL-2 from zebrafish. Corresponding GenBank accession numbers are included in the second column. Residues in the structural calcium site and ligand-binding site are highlighted in blue and green respectively.

4.8 XEEL_{CRD} is Trimeric in Solution

Despite lacking the N-terminal region that engages in disulfide bond formation, the structure of XEEL_{CRD} indicates it forms a non-covalent trimer. Two head-to-tail trimers are observed in the asymmetric unit of the XEEL_{CRD} structure, yet there are few contacts between them. This observation is consistent with the fragility of XEEL_{CRD} crystals, and their tendency to separate into thin sheets. These properties led us to suspect that this particular crystallographic packing arrangement is not biologically relevant. To examine the oligomerization state of XEEL in solution, we performed chemical cross-linking using bis(sulfosuccinimidyl)suberate (**Figure 4-6A**). Trimers predominated at high cross-linker concentrations, consistent with the intimate trimer resolved in the crystallographic asymmetric unit. No species larger than trimeric XEEL_{CRD} were observed.

We also employed equilibrium analytical ultracentrifugation to assess the oligomeric state of XEEL_{CRD}. The absence of curvature in the plots of $\ln(Abs)$ versus radial position squared indicated the presence of a single molecular species (**Figure 4-6B**). The reduced mass (M_r) obtained from a global fit was 27,120 daltons over all of the XEEL_{CRD} concentrations tested. The calculated molecular mass was 96,811 Da, which returns an aggregation number of 3.08, based on the molecular mass of the N-terminally truncated XEEL_{CRD}. This result is consistent with chemical crosslinking data indicating that XEEL_{CRD} is trimeric. The analytical ultracentrifugation (AUC) data show that the XEEL_{CRD} trimer is the preferred species because no free monomeric XEEL_{CRD} was observed at any point during the experiment. Thus, the XEEL_{CRD} forms a highly stable trimer in solution, despite lacking intermolecular disulfide bonds.

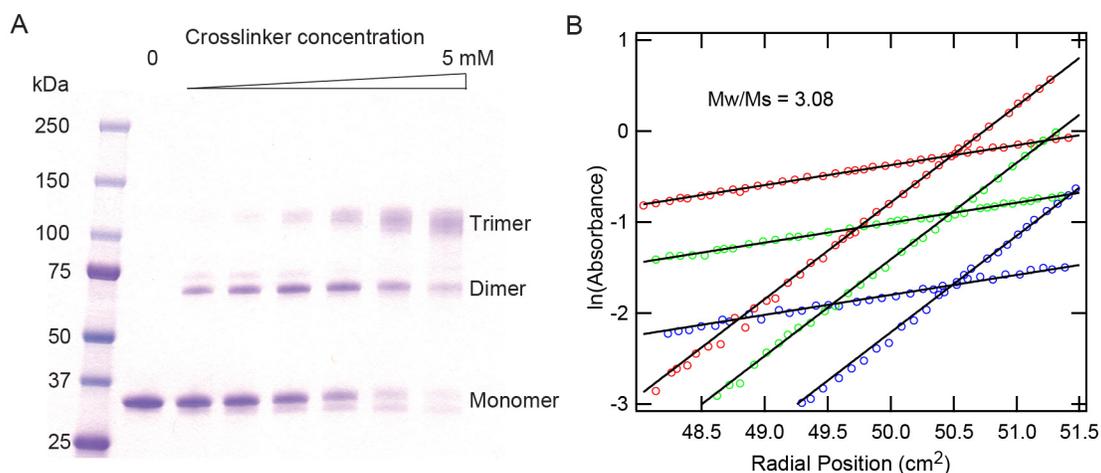


Figure 4-6. Analysis of XEEL_{CRD} oligomeric state. (A) Chemical crosslinking of XEEL_{CRD} with bis(sulfosuccinimidyl)suberate. The final concentrations employed of this cross-linker are 0, 0.1, 0.25, 0.5, 1, 2.5 and 5 mM. The expected masses are 31.6, 63.2, and 94.8 kDa for monomer, dimer, and trimer, respectively. (B) Sedimentation equilibrium AUC of XEEL_{CRD}. $\ln(\text{absorbance})$ vs. radial position squared (cm^2) plot of XEEL_{CRD} at 6,000 RPM (lower slope) or 13,200 RPM (higher slope) for each XEEL_{CRD} concentration (2.6 μM (blue), 5.2 μM (green), 9.0 μM (red)). Only every third raw data point is shown for clarity. The solid lines represent the single species fit of the data that was used to calculate the reduced mass.

Analysis of the interfaces in the XEEL_{CRD} structure indicates that each monomer has an average buried surface area of 1,571 Å² resulting in a total of 4,715 Å² buried upon trimerization. The interface surfaces are polar, but not highly charged. Interactions between each monomer include 18-19 hydrogen bonds and 1-3 salt bridges. Superposition of trimeric hIntL-1 (PDB ID: 4WMY) (26) and GroP-bound trimeric XEEL_{CRD} using all C α atoms yields an RMSD value of 0.63 Å, suggesting a similar trimeric arrangement. The ligand binding sites of both trimers occupy the same space. Therefore, the trimeric architecture of XEEL_{CRD} and hIntL-1 are conserved in 3-dimensional space. Amino acid sequences at the interface are not highly conserved as determined by ConSurf (31) using both automatic sequence alignment of 150 homologs or intelectins from model organisms (**Figure 4-5**). With the large buried surface area and only two unique intelectin structures available to date, the relationship between sequence variation at the monomer interface, and the oligomeric state of intelectins remains to be explored.

4.9 Discussion

Intelectins are soluble lectins produced by organisms ranging from tunicates to humans, and they have been suggested to participate in innate immunity. Despite this proposal and the high sequence similarity amongst intelectins, little is known regarding intelectin function. An understanding of structure and carbohydrate-binding specificity can illuminate intelectin functional roles. The human intelectin, hIntL-1, binds a range of epitopes found only in microbial glycans (26). To examine whether other homologs share a ligand recognition mechanism, we determined the structure of the carbohydrate-recognition domain of the *Xenopus* homolog, XEEL, alone and complexed to GroP.

The X-ray crystal structures of XEEL_{CRD} and hIntL-1 indicate that the intelectins are a discrete lectin class. Though the intelectins and ficolins share an FBD, our X-ray structures show that they are structurally divergent. Intelectins, ficolins, and many C-type lectins possess structural calcium ions. Still, the structural calcium site of XEEL bears no resemblance to the single structural calcium ion site present in ficolins or the C-type lectin DC-SIGN (**Figure 4-7**). The structural calcium ions of intelectins are buried, which contrasts with the solvent-exposed site in L-ficolin. The calcium binding region specific to intelectins is distinct from the two non-conserved structural calcium ions in the C-type lectin DC-SIGN, which are solvent-exposed and coordinated by two protein loops (32,33).

The ligand-binding site of intelectin proteins is also different from that of other lectins. In the intelectin structures solved to date, a calcium ion directly coordinates an exocyclic vicinal diol present in the carbohydrate epitopes. Ficolin structures show no direct interactions between the calcium ion and the ligand (34). Alternatively, C-type lectins, such as DC-SIGN, recognize their carbohydrate ligands through calcium coordination to adjacent secondary hydroxyl groups within a pyranose ring (**Figure 4-7**). As would be anticipated from its ability to coordinate to two secondary hydroxyl groups, the ligand-binding site of DC-SIGN contains no electron-rich aromatic residues that surround the calcium-coordinated ligand, as observed in hIntL-1 and the XEEL_{CRD}. These observations underscore that intelectins are a distinct lectin class.

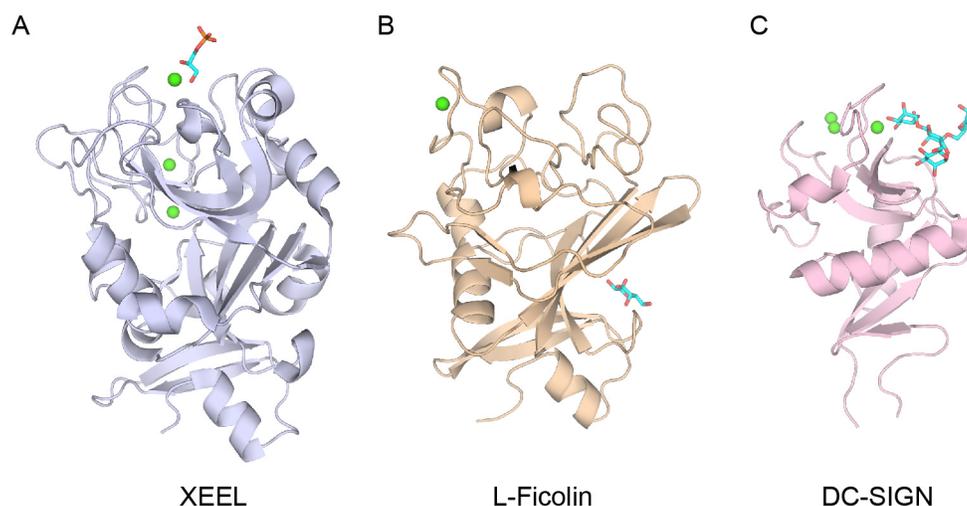


Figure 4-7. Comparison of lectin structures. A) XEEL from this work (PDB ID 4WN0). B) L-ficolin CRD (PDB ID 2J3U) (34). C) DC-SIGN (PDB ID 1SL4) (32). Calcium ions are represented by green spheres and the carbohydrate ligands are shown in cyan.

The structures of XEEL_{CRD} and hIntL-1 reveal conserved features: the intelectin fold, the trimeric CRD core, and the ligand-binding mode. We attribute the ligand-binding specificity of XEEL to W317 and W326, which form a tight box around the exocyclic vicinal diol and imposes steric constraints during binding (**Figure 4-4A**). The structure of GroP-bound XEEL_{CRD} confirms that GroP does indeed bind in the same site as β -Gal β binds to hIntL-1 (**Figure 4-4B**). Both hIntL-1 and XEEL contain aromatic residue (W326 in XEEL, Y297 in hIntL-1) in the binding site that serve as an aromatic wall, a steric barrier that prevents the binding of more substituted diols. Moreover, the placement of the aromatic residue suggests it can engage in a CH- π interaction with the methylene group of 1,2-diol-containing ligands (35). We anticipate that intelectins with this dual aromatic residue signature will show similar ligand-binding modes and specificities.

The structures of XEEL and hIntL-1 provide the means to further interpret intelectin sequence alignment data. We compared human intelectins to those from common laboratory

model organisms. The amino acid sequence of the structural calcium site is exquisitely conserved. Moreover, the overall tertiary structure observed for hIntL-1 and XEEL_{CRD} is conserved throughout chordate species. At least one intelectin in each model organism shows a high degree of conservation in the ligand-binding site with XEEL and hIntL-1, suggesting that a microbe-binding intelectin likely exists in most animals.

The conservation of the ligand-binding site in some, but not all, members of the intelectin family across the evolutionary tree may distinguish those intelectins that recognize microbial surface glycan epitopes from those involved in other functions. Interestingly, intelectin paralogs within the same species (e.g., XEEL and XCGL-2) show appreciable variation in ligand-binding site residues (**Figure 4-5**). The corresponding gene sequences of XCGL-1 and XCGL-2 encode binding site residue substitutions (W317F and W326N), and these lectins appear to have altered ligand specificity, as mentioned earlier. Thus, one intriguing hypothesis is that these two residues control carbohydrate specificity to discriminate between different carbohydrate epitopes. Indeed, further examination of sequence alignment predicts intelectins with novel ligand binding properties. For example, mouse intelectin-2 does not have a predicted aromatic box as it has a W317A substitution. In addition, there are marked variations in the ligand-binding site of the zebrafish intelectins (**Figure 4-5**). The differences between XEEL and hIntL-2 are especially intriguing. For example, the equivalent of XEEL W326 (or hIntL-1 Y297) in hIntL-2 is a serine residue. As a result, hIntL-2 is unlikely to recognize terminal 1,2-diols. Furthermore, E303, which directly coordinates the ligand binding site calcium ion in XEEL and hIntL-1, is a glutamine in hIntL-2. This change may alter the ability of hIntL-2 to chelate a calcium ion. The

consequences of these changes on intelectin structure and ligand specificity warrant careful study.

The oligomeric state of intelectins appears to vary among family members; hIntL-1 is a disulfide-linked trimer, whereas XEEL was reported to be a disulfide-linked hexamer. Here we show that the CRD of XEEL is trimeric in solution despite lacking intermolecular disulfide bonds. The intermolecular disulfide bonds that do exist in full length XEEL can be mapped to residues C24 and C42. These are missing from XEEL_{CRD}. Thus in XEEL, intermolecular disulfide bonds may be required to form stable high-order hexamers, but are not required for trimerization of the carbohydrate binding core. The two intelectin X-ray structures suggest this twice-observed trimer arrangement, which orients three ligand binding sites on one face, is a basic unit. Other portions of the protein may be used to mediate further assembly into higher-order oligomers, or to recruit protein binding partners. Our findings that XEEL_{CRD} exists as a non-covalent trimer highlight the utility of using additional techniques such as cross-linking and equilibrium analytical ultracentrifugation to establish the oligomeric state of an intelectin.

In conclusion, structure and ligand recognition mechanism of intelectins are unique among lectins. With structures of the XEEL_{CRD} and hIntL-1, features conserved between intelectins can be identified. Our results also lay a foundation for further structural analysis of the intelectin family, which will provide insight into the biological function of individual intelectins. Because both XEEL_{CRD} and hIntL-1 bind microbe-specific carbohydrate residues, these intelectin structures provide blueprints for engineering intelectin variants for microbe recognition and targeting.

Table 4-1. X-ray Crystallographic Data Collection and Refinement Statistics.

Data collection statistics	SeMet XEEL _{CRD}	XEEL _{CRD} -GroP complex
Wavelength (Å)	0.97924	0.97856
Resolution range (Å)*	30.10 - 2.30 (2.38 - 2.30)	30.2 - 2.20 (2.28 - 2.20)
Space group	P 2 ₁	P 6
Unit cell (Å)	123.6 111.1 123.6	124.6 124.6 55.6
	90 119.7 90	90 90 120
Total reflections	991049	285681
Unique reflections	129043 (12885)	25366 (2526)
Multiplicity	7.7 (7.7)	11.3 (11.1)
Completeness (%)	100 (100)	100 (100)
Mean I/σ(I)	13.7 (2.6)	19.4 (3.2)
Wilson B-factor (Å ²)	21.0	29.6
R-merge	0.168 (0.625)	0.127 (0.755)
R-meas	0.180 (0.670)	0.133 (0.792)
R-pim	0.065 (0.241)	0.040 (0.237)
Refinement Statistics		
Resolution range (Å)	30.10 - 2.30 (2.38 - 2.30)	29.93 - 2.20 (2.25 - 2.20)
R-factor	0.1597 (0.2152)	0.1515 (0.1977)
R-free (5%)	0.1881 (0.2673)	0.1692 (0.2363)
Number of atoms		
Protein	12927	2151
Calcium	18	3
GroP	-	10
Water	1470	166
Protein residues	1662	277
RMSD (bonds, Å)	0.011	0.010
RMSD (angles, °)	1.170	1.089
Est. coordinate error (ML, Å)	0.21	0.17
Ramachandran favored (%)	96.1	95.7
Ramachandran outliers (%)	0	0
Average B-factor (Å ²)		
Protein	13.0	24.2
Calcium	7.8	17.3
GroP	-	38.0
Solvent	24.4	33.7

*Statistics for the highest-resolution shell are shown in parentheses. Coordinates and structure factors of SeMet XEEL_{CRD} and XEEL_{CRD}-GroP complex were deposited at the PDB under accession code 4WMO and 4WN0, respectively.

4.10 Methods

4.10.1 Cloning, Expression, and Purification of XEEL

The cDNA for XEEL (accession number BC087616) was purchased (Source BioScience) and amplified with primer A (5'-TTTTTGCACCTGCATTTCC-AGCAGGGCACGCTGGTTCATGGAGCCATCCGCAGTTTGAAAAGGGTTCATGTGAACAAGCTTCAATTTCTG-3') and primer B (5'-GGTACCAAGCTCATTAACGGTAGAAGAGCATCACAGCTGCC-3'). The PCR product was then amplified with primers B and C (5'-GCGCGGATCCATGTTGTCATATAGCCTGTTGCTTTTTGCACTTGCATTTCCAGCAGGG-3'). The resulting PCR product was digested with *Bam*HI and *Kpn*I and ligated into a similarly digested pFastBac1 vector. The N-terminus of the protein product is expected to be (residue 4) MLSYSLLLFALAFPAGHA (residue 22) **GSWSHPQFEK** (residue 22) GSCEQASIS; where the underlined amino acids are the predicted secretion signal peptide and the bold amino acids denote the *Strep*-tag[®] II incorporated for purification. Following sequence verification, the vector was subjected to recombination into a baculovirus according to the manufacturer's protocol (Bac-to-Bac Baculovirus Expression System, Life Technologies). The resulting baculovirus genomic DNA was transfected (Insect GeneJuice, Novagen) into Sf21 insect cells to produce the first generation (P1) of recombinant baculovirus. Amplified baculovirus (P2) was produced using P1 to infect suspension Sf21 cells grown in SF900-II-SFM (Life Technologies). Virus-containing supernatant was harvested by centrifugation. Fetal bovine serum was added to 2% to stabilize the baculovirus.

XEEL was produced as a secreted protein using High Five cells (Life Technologies), a derivative of *Trichopulsia ni*. For expression of SeMet labeled XEEL used for experimental phasing, High Five cells were suspension cultured in 921 Delta Series, methionine deficient medium (Expression Systems, cat. no. 96-200, 200 mL) supplemented with 1x antibiotic-antimycotic (Life Technologies) and 10 µg/mL gentamicin (Life Technologies). Cells were infected when they reached a density $\geq 2 \times 10^6$ cells/mL by the addition of 0.5 µL of baculovirus P2 conditioned media per 1×10^6 viable cells. After the first addition of L-SeMet (10 mg, Acros Organics) at 12 hours post infection, additional portions (10 mg) were added every 24 hours until medium harvest. Cells producing XEEL were cultured for 5 days at 22 °C in a baffled flask with shaking at 90 rpm. Conditioned culture medium was harvested by centrifugation and filtered through a 0.22 µm filter; the medium was stored at 4 °C for at least one week. This incubation period was crucial for the proteolytic cleavage that precedes successful crystallization of XEEL residues 54-342. Unlabeled XEEL was produced using the same conditions except Express Five SFM (Life Technologies) supplemented with 1x antibiotic-antimycotic, 10 µg/mL gentamicin, 4 mM glutamine was used as a culture medium, and no Se-Met was added.

Conditioned medium was dialyzed extensively against 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, and 1 mM EDTA. The medium was slowly adjusted to pH 6.7, and a 1 M solution of CaCl₂ was added to a final concentration of 10 mM. In addition, avidin (Calbiochem) was added to the conditioned medium (28 µg/mL) to absorb excess biotin, per the manufacturers protocol (IBA GmbH). The solution was then cleared by centrifugation. *Strep-tagII* XEEL was purified by binding to *Strep-Tactin*[®] Superflow resin (IBA GmbH, cat. no. 2-1206-002). The column was washed with 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, 10 mM CaCl₂ and then 20 mM Bis-Tris

(pH 6.7), 150 mM NaCl, 0.5 mM EDTA. Protein was eluted with 5 mM d-desthiobiotin (Sigma Aldrich) in 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, 0.5 mM EDTA and concentrated using a 10,000 MWCO Amicon Ultra Centrifugal Filter. During the concentration process, large sheet-like crystals began to form. Crystals were harvested by centrifugation at 2,000 RPM and washed 2 times using 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, 0.5 mM EDTA. The crystals were resuspended in 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, 0.5 mM EDTA and CaCl_2 was added to 5 mM. Within one minute, the crystals completely dissolved. Protein purity of the resulting solution was assessed by SDS-PAGE electrophoresis and Coomassie blue staining to be >95 %. The concentration of XEEL was determined using absorbance at 280 nm with an estimated $\epsilon=75,455 \text{ cm}^{-1}\text{M}^{-1}$ for the monomer and a calculated molecular mass of 36,258 Da, post signal peptide removal. Typical yields were 0.5 mg per 50 mL of conditioned medium.

Although the expected mass of the glycosylated protein produced is 36,258 Da, MALDI and ESI-TOF analysis returned masses of 32884 Da and 32802 Da, respectively. The observed mass difference indicated a truncation of at least 30 residues. N-Terminal sequencing (ABI 494, Tufts University Core Facility) revealed the first five amino acids to be RSGGS. Therefore, the XEEL construct used in this study corresponds to residue 54 to 342. We refer to this as XEEL_{CRD} to denote the carbohydrate recognition domain.

4.10.2 Protein X-ray Crystallography

SeMet-labeled XEEL_{CRD} was redissolved in 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, 0.5 mM EDTA and 5 mM CaCl_2 at a concentration of 2 mg/mL and crystallization (hanging-drop vapor-diffusion) was achieved by mixing 1 μL of the protein solution and 1 μL of well solution (100 mM Tris, pH 7.0, 20-24% PEG 400). Crystals appeared in 2-3 hours and matured

to full size within 2-3 days. For cryoprotection, crystals were briefly dipped in well solution containing 30% PEG 400 and vitrified in liquid nitrogen.

Unlabeled XEEL_{CRD} protein was used to obtain the crystal structure with bound GroP (Sigma Aldrich cat. no. G7886). Crystals were grown under similar conditions to SeMet-labeled XEEL, with 50 mM GroP present during crystal growth and cryoprotection.

Single crystal X-ray diffraction experiments were performed on the Life Sciences Collaborative Access Team beamline 21-ID-D at the Advanced Photon Source, Argonne, IL. To determine the structure of SeMet XEEL_{CRD}, diffraction data were collected at the Se K-edge (12661.38 eV). Integration, scaling, and merging were performed with HKL2000 (36). The structure was solved using PHENIX by Se-SAD phasing with AutoSol (37). 42 selenium sites were expected; the substructure search yielded 39. Figures of merit were 0.28 and 0.66 before and after density modification, respectively. The GroP-bound XEEL_{CRD} structure was solved by molecular replacement using Phaser with monomeric Se-Met XEEL_{CRD} as a search model (38). Both structures were adjusted and refined with Coot and phenix.refine respectively (39,40). The coordinate restraint set for GroP was generated using PRODRG (41). MolProbity was used for validation (42). Data collection and refinement statistics are presented in Table 1. PyMOL was used to generate figures depicting the protein structures (43). Surface properties were calculated using PISA (44).

4.10.3 SPR

Analysis of XEEL_{CRD} using SPR was conducted on a ProteOn XPR36 (Bio-Rad) at the University of Wisconsin–Madison Department of Biochemistry BIF. To measure XEEL_{CRD} binding to carbohydrate, ProteOn NLC sensor chips (NeutrAvidin coated sensor chip) (Bio-Rad,

cat. no. 176-5021) were used to capture biotinylated carbohydrate ligands. All experiments presented here were conducted at surface saturated levels of ligand, ~200 response units (R.U.). In all experiments, captured biotin was used in flow cell one as a control. Samples containing purified XEEL_{CRD} were prepared by serial dilution into SPR running buffer (20 mM HEPES (pH 7.4), 150 mM NaCl, 1 mM CaCl₂, and 0.005 % Tween-20). Surfaces were regenerated with short injections of solutions of 10 mM hydrochloric acid (HCl). Data were referenced using the biotin reference channel and processed using the Bio-Rad ProteOn software package. Competition binding experiments were conducted by adding glycerol to the SPR running buffer in which purified XEEL_{CRD} was diluted. Data were analyzed on Prism6 (GraphPad). For determination of the glycerol IC₅₀, SPR data were fit to a one-site competition model in Prism6. The top of the fit was constrained to 110 R.U. based on the equilibrium response in the presence of 1 mM CaCl₂.

4.10.4 Chemical Crosslinking

XEEL_{CRD} aliquots (1 mg/mL, 2 μ L) were mixed with 1 μ L of bis(sulfosuccinimidyl)suberate crosslinker (Pierce) stock solutions to achieve final crosslinker concentrations ranging from 0 – 5 mM. The buffer for all components was 20 mM Bis-Tris (pH 6.7), 150 mM NaCl, 0.5 mM EDTA, and 5 mM CaCl₂. Crosslinking was performed at room temperature for 30 minutes. Each reaction was diluted by adding 7 μ L of the reaction buffer and denatured by adding 2 μ L of 6X SDS loading buffer (350 mM Tris pH 6.8, 30% glycerol, 10% SDS, 9.3% DTT, and 0.06% bromophenol blue). Samples were heated at 95 °C for 3 minutes prior to analysis by SDS-PAGE stained with Coomassie blue.

4.10.5 Sedimentation Equilibrium AUC

Recrystallized XEEL_{CRD} was dissolved in 20 mM BIS-TRIS (pH 6.7), 150 mM NaCl, 0.5 mM EDTA, and 5 mM CaCl₂ and the concentration was determined using UV-visible absorbance spectroscopy. Samples were prepared by dilution to concentrations of 9.0 μM (trimer), 5.2 μM, and 2.6 μM for analysis by sedimentation equilibrium analytical ultracentrifugation. Equilibrium data were collected at 20 °C in a Beckman Optima XLA Analytical Ultracentrifuge using 1.2-cm double sector charcoal-filled Epon centerpieces. Protein gradients were recorded at 276 nm every 2-3 h until two or more were superimposable. Equilibrium data were collected at speeds of 6,000, 8,000, 9,600, 11,500, 13,200, and 15,900 rotations per minute (rpm). After the 15,900 rpm run, the contribution from non-sedimenting absorbance was determined by high speed depletion of the protein; this absorbance was ≤ 0.006 for the three samples. After depletion, subsequent re-equilibration at 11,500 rpm resulted in a gradient essentially superimposable on the original gradient at this speed, indicating no significant loss of protein to irreversible aggregation during the course of the experiment.

The solvent density (ρ) of 1.004 g/mL was computed from density increments (45), except that the BIS-TRIS buffer was not included as it is not in the available table. The calculated partial specific volumes (v) of intact XEEL and XEEL_{CRD} are the same and based on the amino acid sequences, are 0.719 mL/g. The masses based on sequence (M_s) are 36,258 and 31,652, respectively.

Analysis of the equilibrium data followed an approach similar to that previously described (46) and was performed by Darrell R. McCaslin (University of Wisconsin–Madison, Department of Biochemistry, BIF). The analysis utilized programs developed in Igor Pro (Wavemetrics Inc., Lake Oswega, OR) by Dr. D. R. McCaslin. The measured non-sedimenting

absorbance for each sample was included as a fixed parameter. The data from three concentrations and five speeds were globally fit to models consisting of one or two macromolecular species. The single species model described the complete data set adequately, and thus was employed. The reduced molecular weight (M_r) was used as the fitting parameter as it removes the impact of ambiguities in v and ρ on the fit. The reduced molecular weight is defined as $M_r = M_w(1-v\rho)$, where M_w is, in the single species case, the molecular weight of the macromolecular complex. M_r must be an integral multiple of the sequence weight M_s ; therefore, $M_r = nM_s(1-v\rho)$, where n is equal to the oligomeric state of the protein.

4.11 Accession Codes

Coordinates and structure factors have been deposited in the PDB under accession codes 4WMO (Se-Met XEEL_{CRD}) and 4WN0 (XEEL_{CRD}-GroP).

4.12 Contributions

Kittikhun Wangkanont initiated our lab's interest in *Xenopus* intelectin homologs. K. W. generated the baculovirus, identified XEEL_{CRD} crystallization conditions, solved the XEEL protein structures presented in this chapter, and performed the chemical crosslinking. Darrell R. McCaslin collected the equilibrium AUC data and performed analysis.

4.13 Acknowledgements

This work was supported by grants from the National Institutes of Health (NIH) R01GM55984 and R01AI063596 (to L. L. K.) and Chemistry-Biology Interface Training Grant T32 GM008505 (to D. A. W.) and the National Science Foundation IOS1353674 (to K. T. F.) and NSF Graduate Research Fellowship (to D. A. W.). K. W. was supported by a fellowship from the Development and Promotion of Science and Technology Talents Project of Thailand.

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). Use of the Stanford Synchrotron Radiation Lightsource (SLAC National Accelerator Laboratory) is supported by the DOE Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the NIH NIGMS (grant P41GM103393). SPR and analytical ultracentrifugation experiments were conducted at the UW–Madison Biophysics Instrumentation Facility supported by UW-Madison, NSF BIR-9512577, and NIH S10 RR13790.

We would like to acknowledge several people for access to equipment and technical assistance. We acknowledge E. Kondrashkina (LS-CAT), M. M. Benning (Bruker AXS, Madison, WI), D. R. McCaslin (UW–Madison, BIF), G. Sabat (UW-Madison, Biotechnology Center), M. Rajendarm, G. Auer, and D. B. Weibel (UW–Madison), and L. C. Zarling. We thank R. A. Brown and A.W. Baker (UW–Madison) for critical feedback.

4.14 References

1. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., and Dwek, R. A. (2001) Glycosylation and the immune system. *Science* **291**, 2370-2376
2. Holmskov, U., Thiel, S., and Jensenius, J. C. (2003) Collections and ficolins: humoral lectins of the innate immune defense. *Annual review of immunology* **21**, 547-578
3. Fujita, T., Matsushita, M., and Endo, Y. (2004) The lectin-complement pathway--its role in innate immunity and evolution. *Immunological reviews* **198**, 185-202
4. Lee, J. K., Baum, L. G., Moremen, K., and Pierce, M. (2004) The X-lectins: a new family with homology to the *Xenopus laevis* oocyte lectin XL-35. *Glycoconjugate journal* **21**, 443-450
5. Tsuji, S., Uehori, J., Matsumoto, M., Suzuki, Y., Matsuhisa, A., Toyoshima, K., and Seya, T. (2001) Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *J. Biol. Chem.* **276**, 23456-23463
6. Yan, J., Xu, L., Zhang, Y., Zhang, C., Zhang, C., Zhao, F., and Feng, L. (2013) Comparative genomic and phylogenetic analyses of the intelectin gene family: implications for their origin and evolution. *Developmental and comparative immunology* **41**, 189-199
7. Abe, Y., Tokuda, M., Ishimoto, R., Azumi, K., and Yokosawa, H. (1999) A unique primary structure, cDNA cloning and function of a galactose-specific lectin from ascidian plasma. *European journal of biochemistry / FEBS* **261**, 33-39
8. Yan, J., Wang, J., Zhao, Y., Zhang, J., Bai, C., Zhang, C., Zhang, C., Li, K., Zhang, H., Du, X., and Feng, L. (2012) Identification of an amphioxus intelectin homolog that preferably agglutinates gram-positive over gram-negative bacteria likely due to different binding capacity to LPS and PGN. *Fish & shellfish immunology* **33**, 11-20
9. Yan, J., Zhang, C., Zhang, Y., Li, K., Xu, L., Guo, L., Kong, Y., and Feng, L. (2013) Characterization and comparative analyses of two amphioxus intelectins involved in the innate immune response. *Fish & shellfish immunology* **34**, 1139-1146
10. Lin, B., Cao, Z., Su, P., Zhang, H., Li, M., Lin, Y., Zhao, D., Shen, Y., Jing, C., Chen, S., and Xu, A. (2009) Characterization and comparative analyses of zebrafish intelectins: highly conserved sequences, diversified structures and functions. *Fish & shellfish immunology* **26**, 396-405
11. Russell, S., Hayes, M. A., and Lumsden, J. S. (2009) Immunohistochemical localization of rainbow trout ladderlectin and intelectin in healthy and infected rainbow trout (*Oncorhynchus mykiss*). *Fish & shellfish immunology* **26**, 154-163
12. Russell, S., Young, K. M., Smith, M., Hayes, M. A., and Lumsden, J. S. (2008) Identification, cloning and tissue localization of a rainbow trout (*Oncorhynchus mykiss*) intelectin-like protein that binds bacteria and chitin. *Fish & shellfish immunology* **25**, 91-105
13. Datta, R., deSchoolmeester, M. L., Hedeler, C., Paton, N. W., Brass, A. M., and Else, K. J. (2005) Identification of novel genes in intestinal tissue that are regulated after infection with an intestinal nematode parasite. *Infection and immunity* **73**, 4025-4033
14. French, A. T., Knight, P. A., Smith, W. D., Brown, J. K., Craig, N. M., Pate, J. A., Miller, H. R., and Pemberton, A. D. (2008) Up-regulation of intelectin in sheep after infection with *Teladorsagia circumcincta*. *International journal for parasitology* **38**, 467-475

15. Pemberton, A. D., Knight, P. A., Gamble, J., Colledge, W. H., Lee, J. K., Pierce, M., and Miller, H. R. (2004) Innate BALB/c enteric epithelial responses to *Trichinella spiralis*: inducible expression of a novel goblet cell lectin, intelectin-2, and its natural deletion in C57BL/10 mice. *Journal of immunology (Baltimore, Md. : 1950)* **173**, 1894-1901
16. Roberson, M. M., and Barondes, S. H. (1982) Lectin from embryos and oocytes of *Xenopus laevis*. Purification and properties. *J. Biol. Chem.* **257**, 7520-7524
17. Nishihara, T., Wyrick, R. E., Working, P. K., Chen, Y. H., and Hedrick, J. L. (1986) Isolation and characterization of a lectin from the cortical granules of *Xenopus laevis* eggs. *Biochemistry* **25**, 6013-6020
18. Lee, J. K., Buckhaults, P., Wilkes, C., Teilhet, M., King, M. L., Moremen, K. W., and Pierce, M. (1997) Cloning and expression of a *Xenopus laevis* oocyte lectin and characterization of its mRNA levels during early development. *Glycobiology* **7**, 367-372
19. Shoji, H., Ikenaka, K., Nakakita, S., Hayama, K., Hirabayashi, J., Arata, Y., Kasai, K., Nishi, N., and Nakamura, T. (2005) *Xenopus* galectin-VIIa binds N-glycans of members of the cortical granule lectin family (xCGL and xCGL2). *Glycobiology* **15**, 709-720
20. Ishino, T., Kunieda, T., Natori, S., Sekimizu, K., and Kubo, T. (2007) Identification of novel members of the *Xenopus* Ca²⁺-dependent lectin family and analysis of their gene expression during tail regeneration and development. *Journal of biochemistry* **141**, 479-488
21. Nagata, S., Nishiyama, S., and Ikazaki, Y. (2013) Bacterial lipopolysaccharides stimulate production of XCL1, a calcium-dependent lipopolysaccharide-binding serum lectin, in *Xenopus laevis*. *Developmental and comparative immunology* **40**, 94-102
22. Nagata, S. (2005) Isolation, characterization, and extra-embryonic secretion of the *Xenopus laevis* embryonic epidermal lectin, XEEL. *Glycobiology* **15**, 281-290
23. Nagata, S., Nakanishi, M., Nanba, R., and Fujita, N. (2003) Developmental expression of XEEL, a novel molecule of the *Xenopus* oocyte cortical granule lectin family. *Development genes and evolution* **213**, 368-370
24. Dubaissi, E., Rousseau, K., Lea, R., Soto, X., Nardeosingh, S., Schweickert, A., Amaya, E., Thornton, D. J., and Papalopulu, N. (2014) A secretory cell type develops alongside multiciliated cells, ionocytes and goblet cells, and provides a protective, anti-infective function in the frog embryonic mucociliary epidermis. *Development (Cambridge, England)* **141**, 1514-1525
25. Lee, J. K., Schnee, J., Pang, M., Wolfert, M., Baum, L. G., Moremen, K. W., and Pierce, M. (2001) Human homologs of the *Xenopus* oocyte cortical granule lectin XL35. *Glycobiology* **11**, 65-73
26. Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zarling, L. C., R.S., S., Smith, D. F., Cummings, R. D., Paulson, J. C., Forest, K. T., and Kiessling, L. L. (2015) Recognition of Microbial Glycan Epitopes by Human Intelectin. *Nat. Struct. Mol. Biol.* **22**, 603-610
27. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-type lectin superfamily in the immune system. *Immunological reviews* **163**, 19-34
28. Kall, L., Krogh, A., and Sonnhammer, E. L. (2007) Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* **35**, W429-432

29. Schmidt, T. G. M., and Skerra, A. (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protocols* **2**, 1528-1535
30. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948
31. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529-533
32. Guo, Y., Feinberg, H., Conroy, E., Mitchell, D. A., Alvarez, R., Blixt, O., Taylor, M. E., Weis, W. I., and Drickamer, K. (2004) Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* **11**, 591-598
33. Feinberg, H., Powlesland, A. S., Taylor, M. E., and Weis, W. I. (2010) Trimeric structure of langerin. *J. Biol. Chem.* **285**, 13285-13293
34. Garlatti, V., Belloy, N., Martin, L., Lacroix, M., Matsushita, M., Endo, Y., Fujita, T., Fontecilla-Camps, J. C., Arlaud, G. J., Thielens, N. M., and Gaboriaud, C. (2007) Structural insights into the innate immune recognition specificities of L- and H-ficolins. *The EMBO journal* **26**, 623-633
35. Asensio, J. L., Arda, A., Canada, F. J., and Jimenez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46**, 946-954
36. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. in *Macromolecular Crystallography, part A* (C.W. Carter, J., and Sweet, R. M. eds.), Academic Press New York. pp 307-326
37. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica. Section D, Biological crystallography* **65**, 582-601
38. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
39. Afonine, P. V., Grosse-Kunstleve, R. W., and Adams, P. D. (2005) The Phenix refinement framework. *CCP4 Newsl.* **42**, 8
40. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501
41. Schuttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica. Section D, Biological crystallography* **60**, 1355-1363
42. Chen, V. B., III, W. B. A., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12-21
43. The PyMOL Molecular Graphics System, Version 1.3. Schrödinger, LLC
44. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797

45. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided Interpretation of Analytical Sedimentation Data for Proteins. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C. eds.), Royal Society of Chemistry, Cambridge. pp 90-125
46. Laue, T. M. (1995) Sedimentation equilibrium as thermodynamic tool. *Methods in enzymology* **259**, 427-452

Chapter 5

Human Intelectin-1 Interactions with the Human Microbiome

5.1 Abstract

The human body creates unique ecological niches that populate with diverse communities of microbes collectively referred to as the microbiome. The composition of the human microbiome can influence and function as an indicator of human health and disease. To evaluate or manipulate the microbiome composition requires the ability to target subsets of microbes within the context of the larger population. Microbial surfaces are covered in a coat of strain-specific glycans. These glycans are poised to serve as identification codes, which could in principle, be interpreted by lectins functioning as readers. The selectivity of lectins functioning within microbial communities is not known. We focused on human intelectin-1 (hIntL-1), a lectin expressed predominantly in lung and intestinal mucosal epithelial tissues, to examine lectin recognition of bacteria within microbial communities. Human IntL-1 binds microbial but not human glycans by recognizing an exocyclic 1,2-diol epitope present in several microbe-specific carbohydrates. Still not all microbes produce hIntL-1 binding glycans. We began by profiling hIntL-1 binding to bacteria resident in the human gastrointestinal (GI) tract. 45 bacterial strains from diverse taxa were individually screened to yield 12 novel hIntL-1 binding strains. Actinobacteria and Firmicutes have an increased propensity to bind the lectin. When hIntL-1 was assayed against synthetic microbial communities assembled from these strains, however, it was surprisingly more discerning in microbe binding. Specific members within the community were recognized preferentially at the expense of other binding strains. We found that the bound population from within a community could be altered using different hIntL-1 concentrations, a property we attribute to competition amongst hIntL-1 binding strains. Our analysis of lectin binding within microbial communities indicates that community composition and lectin

expression tune native lectin–carbohydrate interactions in a manner not revealed using traditional one lectin—one microbe experiments.

5.2 Introduction

Mammals continuously interact with a complex milieu of microbial companions. While many microbes exist in the external environment, some take up residence on or within humans and generate complex microbial communities (1-3). Centrally important to human health is the ability to survey and sense invading pathogens. The paradox of safely harboring and regulating microbial symbionts, while constantly surveying for pathogens, creates a major challenge for the human immune system.

To accomplish effective surveillance and regulation of microbes, the human body relies on humoral (soluble) and cellular immunity, as well as innate and adaptive immunity (4,5). At mucosal sites where humans are constantly interacting with both symbiont and pathogen, many molecules perform partially redundant functions to successfully supervise microbial communities. For example, mucin glycoproteins are expressed and secreted by intestinal goblet cells and associate into the gel-like substance called mucus. Mucus helps to generate a physical barrier between intestinal mammalian epithelial cells and intestinal microbes (6). Soluble lectins and soluble immunoglobulin A (SIgA) are expressed and secreted to help reinforce this spatial separation barrier (7). Other humoral molecules important in mucosal immunity include pentameric IgA (pIgA), soluble immunoglobulin M (SIgM), antimicrobial peptides, and various small molecules. Also essential to mucosal immunity are specialized mucosa-associated lymphoid tissue (MALT) such as Peyer's patches (PP) which allow antigen presenting cells (APCs) of the immune system access to mucosal associated microbes. Microbe binding and phagocytosis are important for immunoglobulin production, specifically IgA class switch recombination and secretion (8). Many immune cell types reside within MALT, including B- and

T-cells, eosinophils, mast cells, and APCs such as DCs, macrophages, and follicular DCs (9). Indeed, humoral and cellular components work in concert to survey and regulate mucosal resident microbes (**Figure 5-1**).

One soluble lectin secreted at mucosal sites is hIntL-1 (10,11). Human IntL-1 is a member of the recently discovered family of lectins termed intelectins or X-type lectins (12,13). Intelectins are calcium cation-dependent, non-C-type lectins (10,14) which are most similar to the ficolin family of lectins (15). Intelectins and ficolins share a small N-terminal FBG domain (roughly 55 amino acids), however outside of this region distinct structural differences exist. Multiple proteins X-ray crystal structures of intelectin proteins were recently solved to reveal how intelectin structure effects ligand recognition (13,16). Intelectin proteins bind three calcium ions and use a C-terminal CRD to bind microbe specific carbohydrates via direct coordination to a protein-bound calcium ion. Carbohydrate binding assays, glycan microarrays, and structural biology were used to reveal that hIntL-1 recognizes an exocyclic 1,2-diol epitope that is shared amongst several microbial specific carbohydrates such as β -linked D-galactofuranose (β -Gal_f), D-phospho-glycerol-modified glycans, heptoses, D-*glycero*-D-*talo*-oct-2-ulosonic acid (KO) and 3-deoxy-D-*manno*-oct-2-ulosonic acid (KDO) (13). Ficolin proteins coordinate one calcium ion in a surface exposed loop and bind acetylated sugars such as GlcNAc, GalNAc, ManNAc, and N-acetyl neuraminic acid through glycan–amino acid interactions (17). Importantly, intelectin's lack the N-terminal collagen-like domain that is essential for ficolin and collectin binding to MASP proteins for activation of the lectin pathway of complement (15,18). This structural difference suggests that intelectin's do not exert their biological activity through complement-dependent opsonization of bacteria.

Most chordates, including humans, encode multiple intelectin proteins (16). Humans encode two intelectins, hIntL-1 and hIntL-2, which share 84% sequence identity (19). Human IntL-1 is expressed in the lung and intestinal epithelial mucosal tissue as well heart and omental adipose tissue (10,20), while IntL-2 transcripts are found exclusively in the small intestine (19,21). Both IntL-1 and IntL-2 are responsive to immune system activation and their levels can be significantly upregulated in both the lung and intestine (22). When a transgenic mouse model for asthma was stimulated by an allergen or interleukin-13, mouse intelectin transcript levels increased up to 150-fold (23). Infection of mammalian intestine with parasitic nematodes results in substantial upregulation of intelectin proteins in both mice and sheep (11,24-26). Lastly, colonization of gnotobiotic mice by bacteria results in mouse intelectin-1 upregulation by intestinal Paneth cells (7). The increased expression of mammalian intelectin proteins upon immune stimulation or challenge combined with IntL-1's affinity for microbial glycans suggest a function in microbial detection or defense. Still, a biological role for intelectin proteins, and a rationale for intelectin upregulation are lacking.

Most bacteria are covered in a layer of strain-specific cell surface glycans (27-30). The glycopolymers localized to the surface of bacteria include but are not limited to lipopolysaccharide, lipopolysaccharide O-antigen, Gram-negative capsular antigen, Gram-positive capsular polysaccharides, lipoglycans, S-layer protein and other glycoproteins, peptidoglycan, and teichoic acids. Together, these molecules generate a unique glycan barcode indicative of the cell they cover (31). Cell surface glycan barcodes could be read by carbohydrate binding proteins such as lectins. Typically, lectin interactions with microbes are profiled in a one lectin, one microbe manner (32-35). While these studies are useful in demonstrating specific

microbe recognition, it is unlikely that microbes would ever interact with human immune lectins in a one lectin—one strain fashion naturally. Instead, humans interact with complex microbial communities that contain many unique members and significant taxonomical diversity (1,2). Of the limited examples where more than one bacteria was included in a lectin binding assay (are there others?) (35), we are not aware of a situation where more than one lectin binding bacteria was included or how the community effected recognition was examined.

Here, we envisioned profiling hIntL-1 interactions with microbes resident within the human gastrointestinal (GI) microbiome to illuminate potential functions of this lectin in organismal physiology. Using bacterial strains representative of natural isolates from human GI microbiomes (36), we measured hIntL-1 binding to 12 of the 45 strains we assayed. As a proxy of human microbiomes, we generated synthetic mixtures from the binding and non-binding strains to examine hIntL-1 recognition within the context of microbial communities. The results from analysis of multiple diverse communities suggest that hIntL-1 binding to bacteria is competitive and tunable. One surprising result from our community-based analysis was that some strains, often Gram-negative bacteria, are susceptible to binding inhibition when assayed within a community. Additionally, we demonstrate that modulation of hIntL-1 levels (i.e., protein upregulation) altered the bacteria within a community that were targeted by the lectin. The ability of hIntL-1 to differentially recognize “binding bacteria” whether examined in isolation or within a microbial community was striking and suggests that lectin recognition within microbial communities is malleable and dependent upon host lectin expression levels and the microbial community composition.

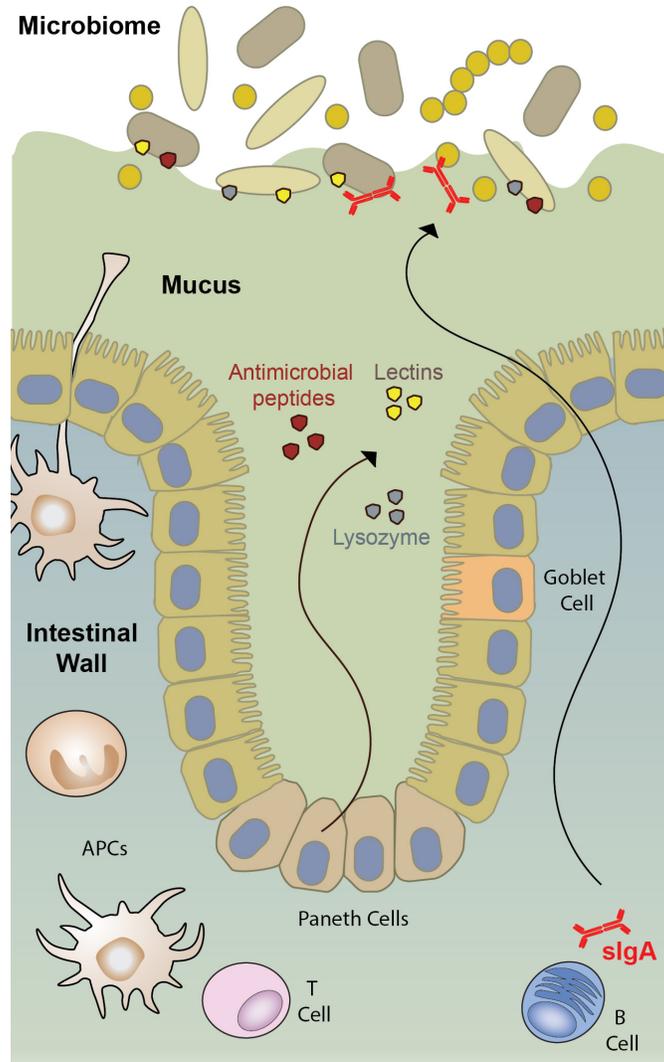


Figure 5-1. Schematic representation of the mucosal immune system functioning within the mammalian intestinal mucosa. This figure was generously provided by Amanda Dugan and adapted from reference (37).

5.3 hIntL-1 Recognition of Microbial Strains

We began by exploring hIntL-1 binding to bacterial strains that are similar to bacteria found within GI microbiomes (36). These strains occupy diverse taxa and have available genomic sequencing data. We were interested in what portion of these strains would be bound by

hIntL-1 and if certain taxa would be recognized preferentially. Binding of Strep-tagged hIntL-1 to freshly grown, fixed bacterial cells was assayed and quantified using flow cytometry (13). The calcium ion-dependence of hIntL-1 carbohydrate-binding was evaluated by also monitoring binding in the presence of EDTA (**Figure 5-2A**). In total, 45 strains from four taxa were assayed; hIntL-1 binding was measured to 12 (**Table 5-1**). The results suggest that binding is favored to Actinobacteria and Firmicutes (**Figure 5-2B**). This preference is supported by a recent examination of the Bacterial Carbohydrate Structure Data Base (BCSDB) that revealed increased utilization of hIntL-1 binding ligands in Actinobacteria and Firmicute bacteria (38). For example, Actinobacteria cell surface glyconjugates were shown to be enriched with the hIntL-1 ligand β -D-Galf.

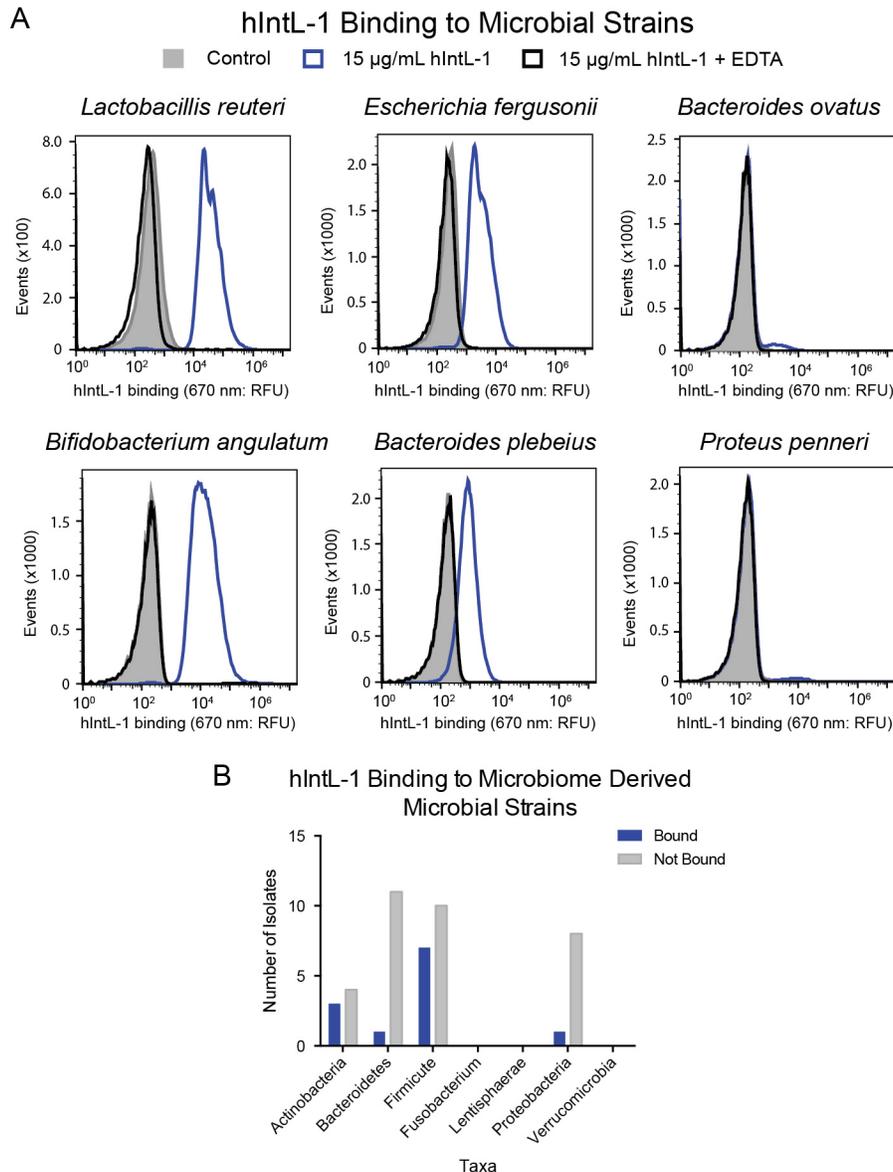


Figure 5-2. Binding of hIntL-1 to fixed bacterial strains. (A) Representative data of hIntL-1 binding to fixed bacterial cells. hIntL-1 was visualized using a fluorophore-labeled Anti-Strep-tag II antibody. For the EDTA treated sample, cells were stained in the presence of 5 mM EDTA. (B) Summary of 45 assayed strains, sorted by taxa. All stains have been confirmed by 16S rRNA sequencing. More information on these bacteria can be found in **Table 5-1**.

Table 5-1. Summary of hIntL-1 Binding to Microbes

<u>Binding Cells</u>		<u>Non-binding Cells</u>	
<i>Genus species</i>	Strain ^a	<i>Genus species</i>	Strain ^a

Actinobacteria		Actinobacteria	
<i>Bifidobacterium angulatum</i>	ATCC 27535	<i>Bifidobacterium adolescentis</i>	ATCC 15703
<i>Bifidobacterium bifidum</i>	ATCC 29521	<i>Bifidobacterium pseudocatenulatum</i>	ATCC 27919
<i>Bifidobacterium dentium</i>	ATCC 27678	<i>Collinsella aerofaciens</i>	ATCC 25986
		<i>Collinsella intestinalis</i>	DSMZ 13280
Bacteroidetes		Bacteroidetes	
<i>Bacteroides plebeius</i>	DSMZ 17135	<i>Alistipes indistinctus</i>	DSMZ 22520
		<i>Bacteroides caccae</i>	ATCC 43185
		<i>Bacteroides fingoldii</i>	DSMZ 17565
		<i>Bacteroides intestinalis</i>	DSMZ 17393
		<i>Bacteroides ovatus</i>	ATCC 8483
		<i>Bacteroides thetaiotaomicron</i>	7330
		<i>Bacteroides thetaiotaomicron</i>	VPI-5462 ^b
		<i>Bacteroides uniformis</i>	ATCC 8492
		<i>Bacteroides vulgatus</i>	ATCC 8482
		<i>Bacteroides xylanisolvens</i>	DSMZ 18836
		<i>Parabacteroides merdae</i>	ATCC 43184
Firmicute		Firmicute	
<i>Anaerococcus hydrogenalis</i>	ATCC 49630	<i>Blautia hansenii</i>	ATCC 27752
<i>Dorea longicatena</i>	DSMZ 13814	<i>Clostridium asparagiforme</i>	DSMZ 15981
<i>Eubacterium bifforme</i>	ATCC 27806	<i>Clostridium bloteae</i>	DSMZ 15670 ^c
<i>Lactobacillus reuteri</i>	DSMZ 20016	<i>Clostridium hylemonae</i>	DSMZ 15053
<i>Lactobacillus ruminis</i>	ATCC 27780	<i>Clostridium symbiosum</i>	ATCC 14940
<i>Roseburia intestinalis</i>	DSMZ 14610	<i>Coprococcus comes</i>	ATCC 27758
<i>Ruminococcus torques</i>	ATCC 27756	<i>Holdemania filiformis</i>	ATCC 51649
		<i>Mitsuokella multacida</i>	ATCC 27723
		<i>Ruminococcus gnavus</i>	ATCC 29149
		<i>Streptococcus infantarius</i>	ATCC BAA-102
Proteobacteria		Proteobacteria	
<i>Escherichia fergusonii</i>	ATCC 35469	<i>Citrobacter youngae</i>	ATCC 29200
		<i>Edwardsiella tarda</i>	ATCC 23685
		<i>Enterobacter cancerogenus</i>	ATCC 35316
		<i>Escherichia coli</i>	MS.200.1
		<i>Escherichia coli</i> K12	ATCC 47076
		<i>Proteus penneri</i>	ATCC 35198
		<i>Providencia rettgeri</i>	DSMZ 1131
		<i>Providencia stuartii</i>	ATCC 25827

^a Strain information is provided as the ATCC strain designation, or the Leibniz Institute DSMZ–German Collection of Microorganisms and Cell Cultures identification number.

^b ATCC 29148

^c ATCC BAA-613

5.4 hIntL-1 Binding to Microbial Communities

After profiling hIntL-1 binding to strains in isolation, we were interested in how the lectin would interact with bacteria present in complex mixtures. We envisioned using synthetic communities of known percentages of binding and non-binding bacteria to learn more about hIntL-1—community interactions. Using the strains identified in **Table 5-1**, we generated synthetic communities and assayed hIntL-1 binding. The communities were assembled in a manner representative of the human gut with respect to taxa. In the first community *L. reuteri* and *E. fergusonii* were included as hIntL-1 binding strains, and *B. ovatus* and *P. penneri* were included as non-binding strains. An example of this community can be seen in **Table 5-2**.

Table 5-2. Example of Synthetic Microbial Mixtures Used to Assay hIntL-1 Binding

Predicted 10% Binding	
<u>Binding</u>	<u>% Included</u>
<i>L. reuteri</i>	5
<i>E. fergusonii</i>	5
<u>Non-binding</u>	
<i>B. ovatus</i>	45
<i>P. penneri</i>	45

Predicted 50% Binding	
<u>Binding</u>	<u>% Included</u>
<i>L. reuteri</i>	25
<i>E. fergusonii</i>	25
<u>Non-binding</u>	
<i>B. ovatus</i>	25
<i>P. penneri</i>	25

Prior to assaying synthetic communities, strains were analyzed individually (**Figure 5-3A**). Each strain exhibited differences in forward scatter, side scatter, and the amount of hIntL-1 bound to its surface (**Figure 5-3A**). The properties (size, shape, amount of hIntL-1 bound, etc.) of each strain should facilitate their identification from within a community. For example, *B. ovatus* is the smallest microbe analyzed in the community represented in **Table 5-2**, and more

hIntL-1 binds the surface of *L. reuteri* as compared to *E. fergusonii* despite both strains appearing similar by forward and side scatter (**Figure 5-3A**). While analyzing strains in isolation, we found that a small number of the events within each sample were opposite the majority with respect to hIntL-1 binding; i.e., 95.2% of the *B. ovatus* events were hIntL-1 negative while 4.8% were hIntL-1 positive, and 1.6% of the *L. reuteri* events were hIntL-1 negative while 98.4% were hIntL-1 positive. This phenomenon was observed to a roughly equal extent in every strain tested. Thus, when calculating the predicted percentage of binding bacteria within a community we treated each strain as monomodal, either 0% or 100% of the event were bound.

Using the four strains described in **Table 5-2**, communities with an increasing amount of hIntL-1 binding bacteria were assembled. Each community was assayed using 0.187 μg of hIntL-1 per $1\text{E}6$ cells. The microbial community was washed once and analyzed by flow cytometry. The percentage of hIntL-1 binding cells was determined using gate established from same community stained with antibody only. Quantitative analysis of the percentage of bacteria within each community bound by hIntL-1 revealed that the number of bacteria bound by hIntL-1 was consistently less than what was predicted (**Figures 5-3B & C**). Based on the characteristics of each strain when assayed in isolation, we hypothesized that *L. reuteri* were faithfully quantified, while the *E. fergusonii* were underrepresented in the binding population. The apparent inhibition of hIntL-1 binding to *E. fergusonii* was striking and unexpected because when analyzed individually hIntL-1 bound the strain robustly (**Figure 5-2A & 5-3B**).

The results from analysis of this community suggested the intriguing possibility that the residents within a community influence hIntL-1 binding. Certain strains may be preferentially

bound and they may inhibit binding to other strains. Of note is the different taxonomy of the two hIntL-1 binding strains. *L. reuteri* is Gram-positive and may have capsular polysaccharide (CPS), exo-polysaccharide (EPS), and teichoic acid (TA) localized to its cell surface (39) while *E. fergusonii* is Gram-negative and hIntL-1 is likely binding to LPS O-antigen or capsular antigen.

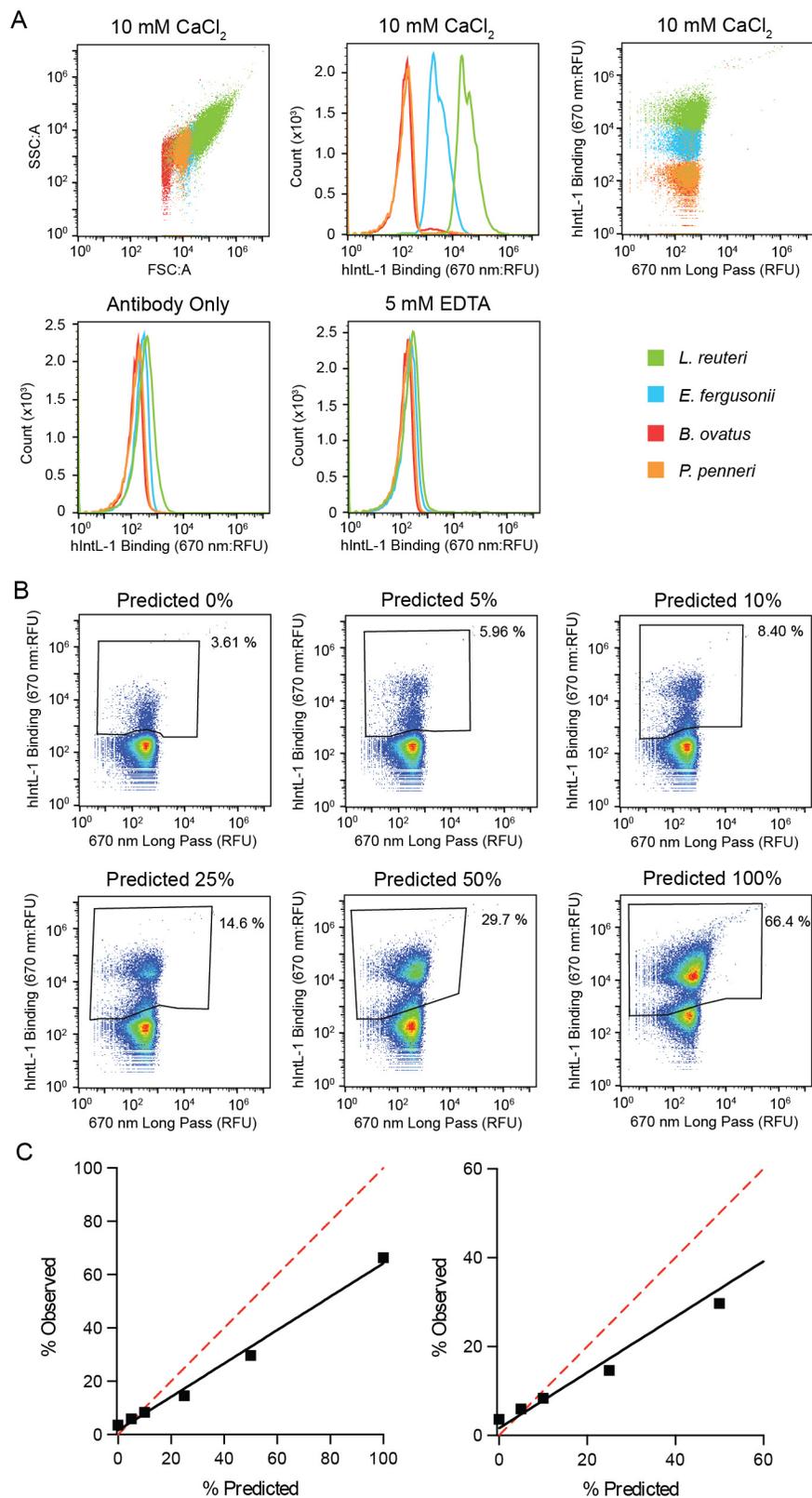


Figure 5-3. hIntL-1 binding to a synthetic microbial communities. (A) Four strains of bacteria were individually assayed for hIntL-1 binding to establish the attributes useful in distinguishing them within mixtures. The results from strains assayed in isolation are shown together here as different colors. The specificity of binding is demonstrated by binding inhibition in the presence of 5 mM EDTA. (B) Results of synthetic communities assayed for hIntL-1 binding. Cells were stained with 0.187 μg of hIntL-1 per $1\text{E}6$ cells. The percentage of bound events was quantified using a gate that was determined from an unstained sample, the gate is shown here as a black box. The percentage of bound events is included within each graph. The bacteria bound in the predicted 0% mixture are the result of a small number of events whose binding by hIntL-1 was opposite the majority of events. A total of 100,000 individual events were collected under each condition. (C) Summary of the results shown in section B. The red line represents a perfect correlation between predicted and observed with a slope of 1 and y-intercept of 0. The black line represents a linear regression of the data. The same graph is shown on the right with the axes reduced to more clearly highlight points near the origin.

5.5 hIntL-1 Binding Depends on Community and Context

We hypothesized that the taxa represented within a community, and thus the cell surface glycoconjugates that make up each bacteria's unique glycan barcode, may influence lectin binding. To test this, additional communities of hIntL-1 binding and non-binding strains were prepared. The first contained *L. reuteri* and *E. fergusonii* as the binding strains. The second community contained one Gram-positive binding strain, *L. reuteri*, and one Gram-negative binding strain, *B. plebeius*. The third community contained exclusively Gram-negative binding bacteria, *E. fergusonii* and *B. plebeius*, and the fourth contained exclusively Gram-positive binding bacteria, *L. reuteri* and *B. angulatum*. All four communities contained *B. ovatus* and *P. penneri* as the non-binding strains.

As intelectin expression levels have been shown to vary within the human population and after antigen/allergen challenge (11,22-26), each community was assayed using two different concentrations of lectin. To examine the functional stability of hIntL-1 binding to bacteria within microbial communities, each sample was assayed following a fourfold dilution and assayed following a wash step. We hypothesized that more transient interactions would be sensitive to

the wash, while stable interactions would remain largely unchanged. In all of the communities tested, the number of hIntL-1 binding cells was higher in samples stained with more lectin. This result is most evident at high percentages of predicted hIntL-1 binding cells, and in mixtures that contain at least one Gram-negative hIntL-1 binding strain (**Figures 5-4A - C**). When the stability of hIntL-1 binding cells was assayed by comparing washed and diluted samples, the lability of Gram-negative binding within a community was pronounced (**Figures 5-4A - C**). In all the communities that contain at least one Gram-negative bacterial strain, the percentage of cells bound by hIntL-1 is closer to the predicted percentage of binding cells when analyzed following the dilution.

Indeed, the pool of bacteria from within a community that were bound by hIntL-1 was sensitive to lectin levels and the context of the interaction. At the highest lectin concentration and analysis following dilution, greater than 90% of the cells from the predicted 100% binding mixtures of *L. reuteri* and *E. fergusonii*, and *L. reuteri* and *B. plebeius* were bound by hIntL-1 (**Figure 5-4A & B**). In the community composed entirely of Gram-positive hIntL-1 binding strains, all four assay conditions yielded hIntL-1 binding to greater than 99% of bacteria in the predicted 100% mixture (**Figure 5-4C**).

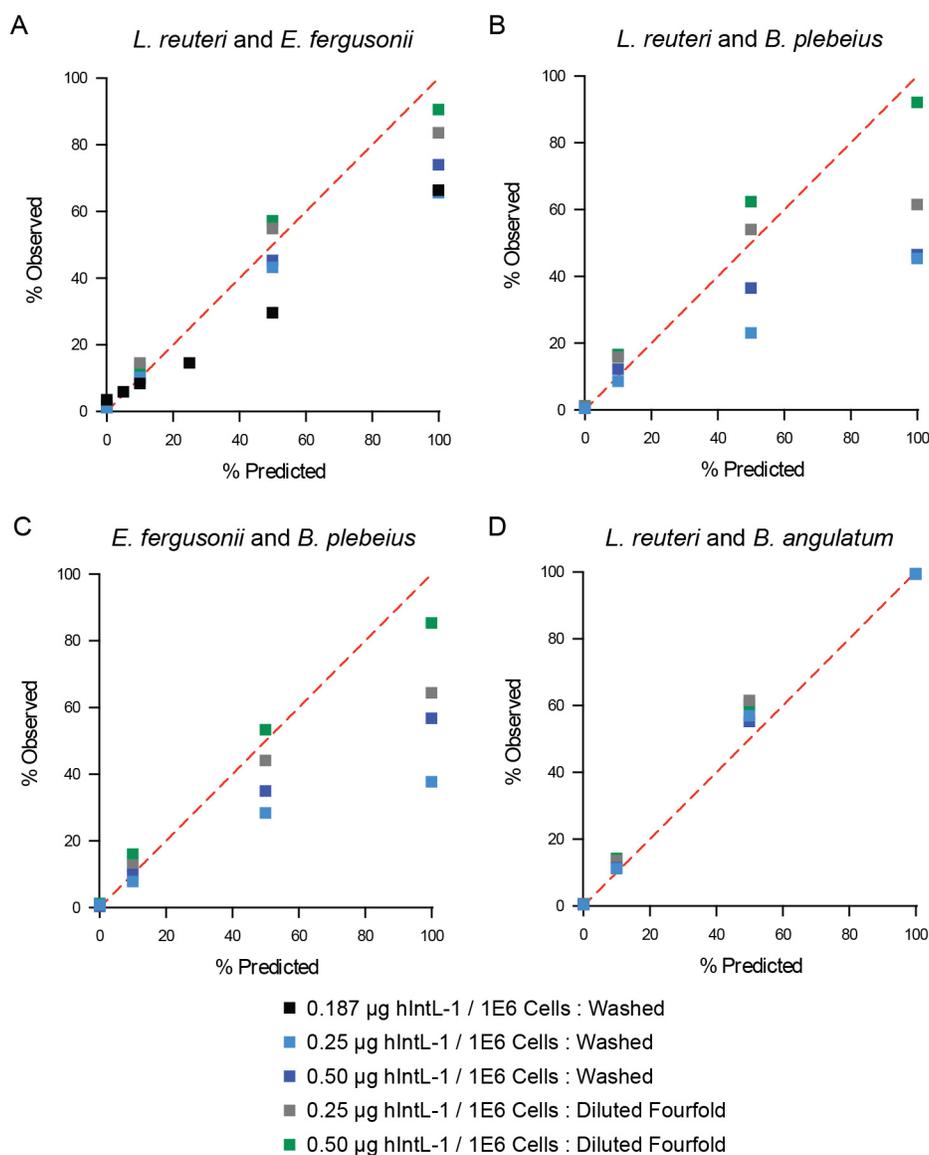


Figure 5-4. Competitive inhibition of hIntL-1 binding by microbial communities. Four communities were assayed for hIntL-1 binding by flow cytometry. The strains in each mixture that are known hIntL-1 ligands are denoted above each graph ((A) contained *L. reuteri* and *E. fergusonii*, (B) *L. reuteri* and *B. plebeius*, (C) *E. fergusonii* and *B. plebeius*, and (D) *L. reuteri* and *B. angulatum*) while all four share *B. ovatus* and *P. penneri* as non-binding strains. The effects of increased hIntL-1, and the removal of a wash step, were used to suggest competition between bound strains that is dependent on the composition of each community. The red line represents a perfect correlation between predicted and observed with a slope of 1, and y-intercept of 0. The data from 0.187 µg of hIntL-1 per 1e6 cells : Washed is the same data shown in **Figure 5-3C**.

5.6 hIntL-1 Binding to the Fecal Microbiome

After using synthetic communities to assay the effect of community and context on hIntL-1 binding, we sought to assay lectin binding to a native community. For this we chose the bacteria associated with a human fecal sample. To begin probing hIntL-1 binding to a fecal microbiome, we chose to use a fecal sample prepared from a single individual. Previous work has shown that this individual's sample is composed largely of Gram-positive bacteria; ~90 % Firmicutes. To assay for calcium ion-dependence, a sample containing 3 mM EDTA was included. When hIntL-1 binding was assayed in the presence of 10 mM calcium ions, between 20 and 30% of all events were bound by the lectin (**Figure 5-5**). Analysis of the EDTA treated sample revealed the majority of the events were dependent on calcium ions for binding (**Figure 5-5C**). Of the events not dependent on calcium, most were among the largest events in the sample as determined by forward and side scatter. We hypothesize some of these events may be yeast (40,41). In this experiment the bacteria:hIntL-1 mixture was washed once before analysis. The same sample has been analyzed using a fourfold dilution and similar results were obtained. We suspect the results were similar because the sample is predominated by Gram-positive Firmicutes. While this data is representative of greater than five independent experiments, we seek to repeat this using the assay conditions presented in section 5.5. We would like to assay the sample three times, and then analyze fecal samples from additional patients. This will provide a more firm understanding of the percentage of bacteria in a sample capable of hIntL-1 binding, and the diversity of hIntL-1 binding microbes within the human population.

We have begun developing the methodology to use hIntL-1 labeling in combination with FACS to further characterize hIntL-1 binding bacteria. We suspect that DNA sequencing, both

16S rRNA and metagenomic, will reveal much about hIntL-1 binding partners. Cell sorting and 16S rRNA sequencing was recently used to characterize bacteria bound by SIgA and revealed a correlation between SIgA bound bacteria and IBD (42,43). To our knowledge, these seminal examples are among the few where microbiota interacting with soluble immune proteins have been explored and characterized genetically.

Working with the UW-Madison Carbone Cancer Center Flow Cytometry Core and Robert Kerby from Professor Federico Rey's group in the Department of Bacteriology, we have developed a protocol to sort hIntL-1 binding bacteria and subject them to 16S rRNA sequencing. While these results are not complete, I would like to present some of the data we have obtained. Preliminary studies demonstrated that $1E6$ to $2E6$ cells were optimal for genomic DNA extraction, and PCR amplification of the V3-V4 region of the 16S rRNA gene (**Figure 5-6A**). For the analysis presented in this thesis, the hIntL-1 signal was slowly decreasing during the long duration of the sort (~3 hrs). This resulted in cells that are bound by hIntL-1 being sorted into the non-binding pool. Thus we suspect the true results will be more pronounced than what we report here. We have already addressed the problem of hIntL-1 binding stability by omitting the wash step and replacing it with a fourfold dilution (section 5.5). The V3-V4 region of the 16S rRNA gene was amplified and subjected to Illumina sequencing. Data were processed using QIIME software (44). When the data are processed using a standard QIIME protocol and presented based on taxa, all of the samples appeared similar (**Figure 5-6B**). We hypothesized that analyzing the operational taxonomic units (OTUs) would reveal differences masked by the standard procedure. When OTUs were assembled based on 97% sequence similarity and their appearance within each sample was quantified, enrichment of many OTUs to either binding or non-binding

cells was observed (**Figure 5-6C**). We hypothesize that some OTUs enrich in either the binding or non-binding pool because some strains are genetically distinct enough to be their own OTU and the hIntL-1 glycosylation analysis is sensitive to a strains unique glycan barcode. We suspect that glycan and OTU-based analysis may be uniquely equipped for differentiating bacterial strains. Analysis techniques that are sensitive to strain information will become increasingly important as correlation moves toward causation in the microbiota field. The glycosylation/sequencing based analysis we present here, which we refer to as GlycoSEQ, is uniquely suited for differentiating bacteria into strains as the combination of lectin binding and sequencing provide information not attainable through either method alone. We are in the process of repeating this analysis using our updated binding conditions, as well as assaying the reproducibility and robustness of the work flow. We would like to expand our analysis to fecal samples from other donors to better understand the variability of hIntL-1 bound bacteria within the human population. We suspect that this information, in conjunction with medical records, may provide clues toward the biological function of hIntL-1.

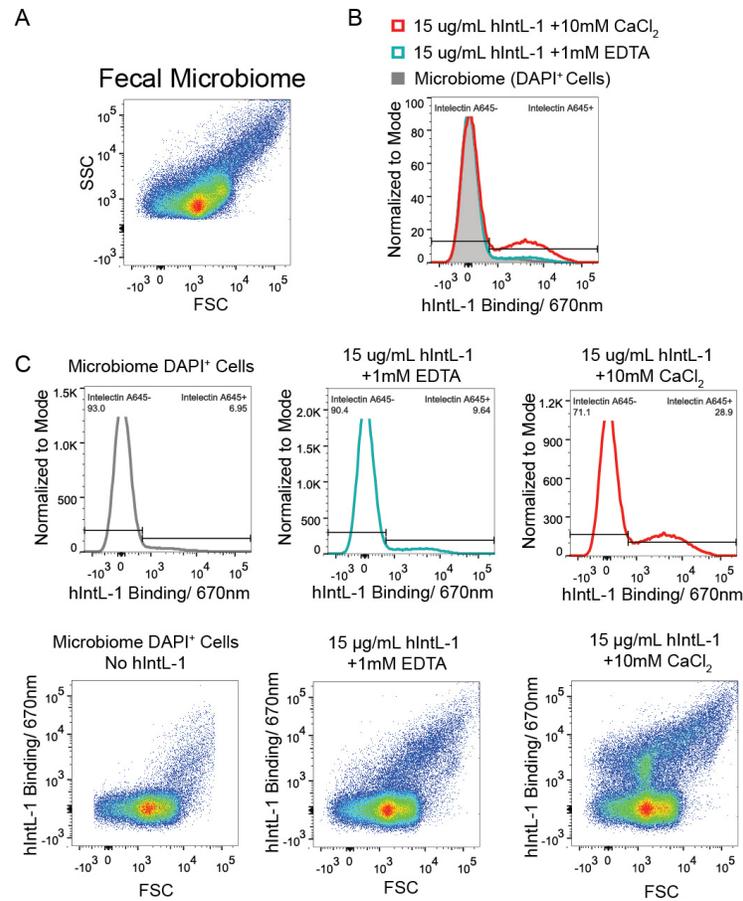


Figure 5-5. hIntL-1 binding to bacteria isolated from a human fecal slurry. (A) Analysis of the fecal microbiome isolated from a human fecal slurry. The complexity of the sample is revealed from the large disparity in the forward and side scatter of the sample. (B) Histogram representing the staining of a fecal microbiome in the presence of hIntL-1. Only with the inclusion of hIntL-1 and free calcium ions does the population become bimodal, with a portion bound by hIntL-1 (C) The samples shown collectively in panel (B) are depicted individually. The gate and quantification are included in the histogram representation of each sample. The same data is depicted below as a dot plot to show the forward and side scatter profile of binding and non-binding bacteria. Data is representative of greater than 5 independent experiments.

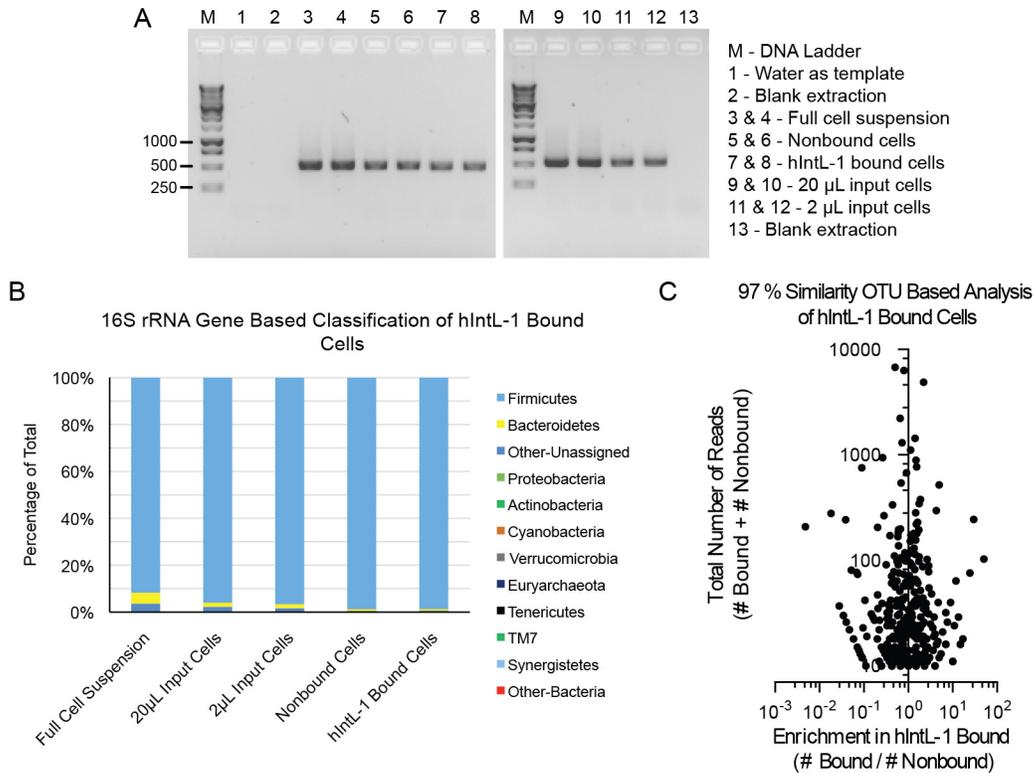


Figure 5-6. Genomic characterization of hIntL-1 bound bacteria. (A) PCR amplification of the V3-V4 region of the 16S rRNA gene from genomic DNA isolated from input, and sorted bacterial cells. (B) Taxa based classification of the results of the V3-V4 16S rRNA gene sequencing. This figure was generated using the standard QIIME protocol and 97% sequence similarity OTU assignment. (C) Analysis of OTU (97% sequence similarity) distribution between the binding and non-binding samples. OTUs with less than 10 combined reads in the binding and non-binding pool were omitted. OTUs are omitted from this graph if they were found exclusively within one pool; 12 were exclusive to the non-binding sample, 7 were exclusive to the hIntL-1 bound sample. Data was collected and processed by Robert Kerby (Federico Rey Group).

5.7 Microbial Carbohydrate Antigens Within the Fecal Microbiota

After profiling hIntL-1 binding to a fecal microbiota sample, we were interested in other carbohydrate antigens on the surface of the intestinal microbiota. To define additional surface accessible carbohydrate epitopes, I employed commercially available plant lectins with documented ligand specificities. I tested seven lectins with varying specificities; concanavalin A

(Con A), RCA-I, WGA, DBA, PNA, SBA, and UEA I. Substantial binding to bacteria from the fecal slurry was observed using three lectins, Con A, RCA-I, WGA (**Figure 5-7**). These three lectins have unique specificities: Con A interacts with Man and Glu residues, RCA-I has affinity toward Gal, and WGA binds to GlcNAc. Surprisingly, only a small percentage of cells were bound by GalNAc (DBA and SBA) and α -fucose binding lectins (UEA I). These results provide an initial fingerprint of the carbohydrates antigens displayed on microbes of the human intestinal microbiota and suggest that human lectins with affinity toward Man, Glc, Gal, and GlcNAc may bind a significant portion of the intestinal community. Interestingly, the membrane-bound C-type lectin dectin-1 (*CLEC7A*), which has affinity for fungal β -1,3-glucan, was recently found to bind intestinal fungi and influence intestinal colitis (40). Our flow cytometry analysis of the fecal microbiota includes bacteria and fungi. While we did not differentiate between fungi and bacteria in our experiment, our flow cytometry results with Con A suggest that a significant portion of intestinal microbes express surface Man- and Glu-containing glycoconjugates that may also interact with dectin-1. How other microbes within the microbiota community influence dectin-1—fungi binding remains unexplored.

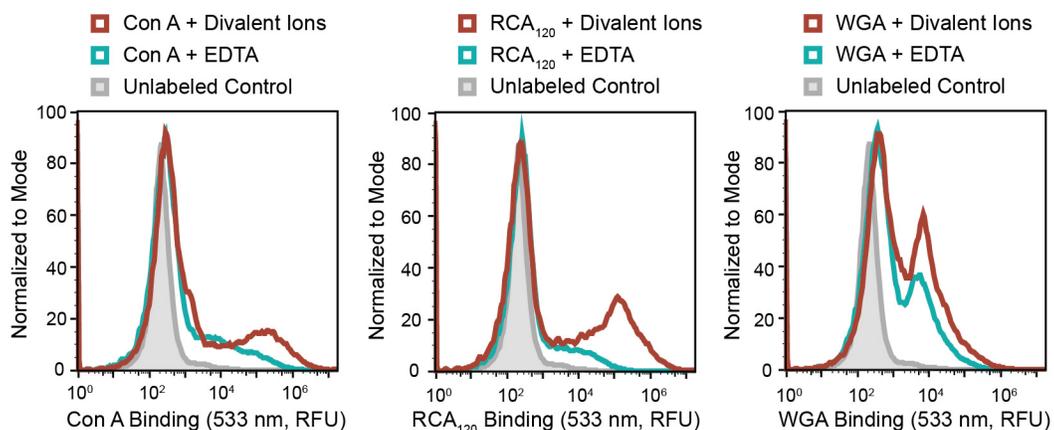


Figure 5-7. Lectin binding to the surface of fecal microbiota. Fluorescein conjugated lectins

were used to assay binding. Lectin binding was assayed in the presence 10 mM calcium, 1 mM magnesium, and 0.5 mM zinc; or 5 mM EDTA. Data is representative of two independent experiments.

5.8 hIntL-1 Interactions with Immune Cells

A biological role for hIntL-1 within mammals has yet to be determined. Thus far we have focused on the carbohydrate binding activity of hIntL-1 and lectin binding to the intestinal microbiota. In addition to microbe binding, there are many host factors within the gut and lung that hIntL-1 may interact with. One example is other proteins, potentially immune proteins, which could be identified using pull-down experiments and mass spectrometry. Another possibility is mammalian cells, specifically immune cells, which are known to reside within the mammalian GI tract (**Figure 5-1**).

Immune cells have specific receptors on their surface that allows them to recognize ligands, and respond differentially. We sought to assay for hIntL-1 interactions with human immune cells using whole human blood. This work has been largely performed by a talented undergraduate research assistant I have had the pleasure of working with for four years, Lucas C. Zarling, in collaboration with Deane Mosher's Group. Whole human blood was incubated with antibodies for differentiating immune cell populations, and untagged-hIntL-1 that was directly conjugated with Alexa Fluor 645 nm. Following red blood cell lysis in a hypotonic solution, the remaining cells were analyzed by flow cytometry (**Figure 5-8A**). Antibodies specific for the cell surface markers CD45 (leukocyte common antigen) and CD14 (monocyte differentiation marker CD14) were used to identify immune cell populations. Using a sample stained with hIntL-1 and anti-CD45, granulocytes, platelets, lymphocytes, monocytes, and eosinophils were identified (**Figure 5-8B**). This analysis revealed hIntL-1 binding to monocytes, eosinophils and granulocytes, but not to lymphocytes or platelets (**Figure 5-8C**). We next performed the

straining procedure in the presence 10 mM calcium ions, or 5 mM EDTA. The results were nearly identical. This suggests that hIntL-1 is bound by certain immune cell types in a calcium ion-independent manner, likely through a specific protein:protein interaction.

Future experiments will test the ability of monocyte derived innate immune antigen presenting cells such as DCs and macrophages to bind hIntL-1. Lastly, we seek to identify the receptor that is binding hIntL-1 on the surface of these cells. We suspect that pull-down experiments using human blood derived monocytes and mass spectrometry will characterize this interaction. Based on binding to innate immune APCs, and ability of hIntL-1 to interact robustly Gram-positive bacteria, we suspect that hIntL-1 may function in connecting soluble innate immunity and cellular immunity. Specifically, hIntL-1 may assist in recognition, phagocytosis, and antigen presentation of encapsulated Gram-positive bacteria by APCs. Thus, hIntL-1 may help expose bacteria hiding behind what is commonly thought of as their anti-phagocytic capsule (**Figure 5-10**). Supporting this hypothesis are preliminary experiments that suggest the majority of hIntL-1 binding bacteria within a fecal sample are also bound by SIgA (i.e., up to 80% of hIntL-1 binding bacteria are SIgA+, data not shown). Future experiments such as microbial cell killing and phagocytosis experiments will provide insight into this fascinating question (45).

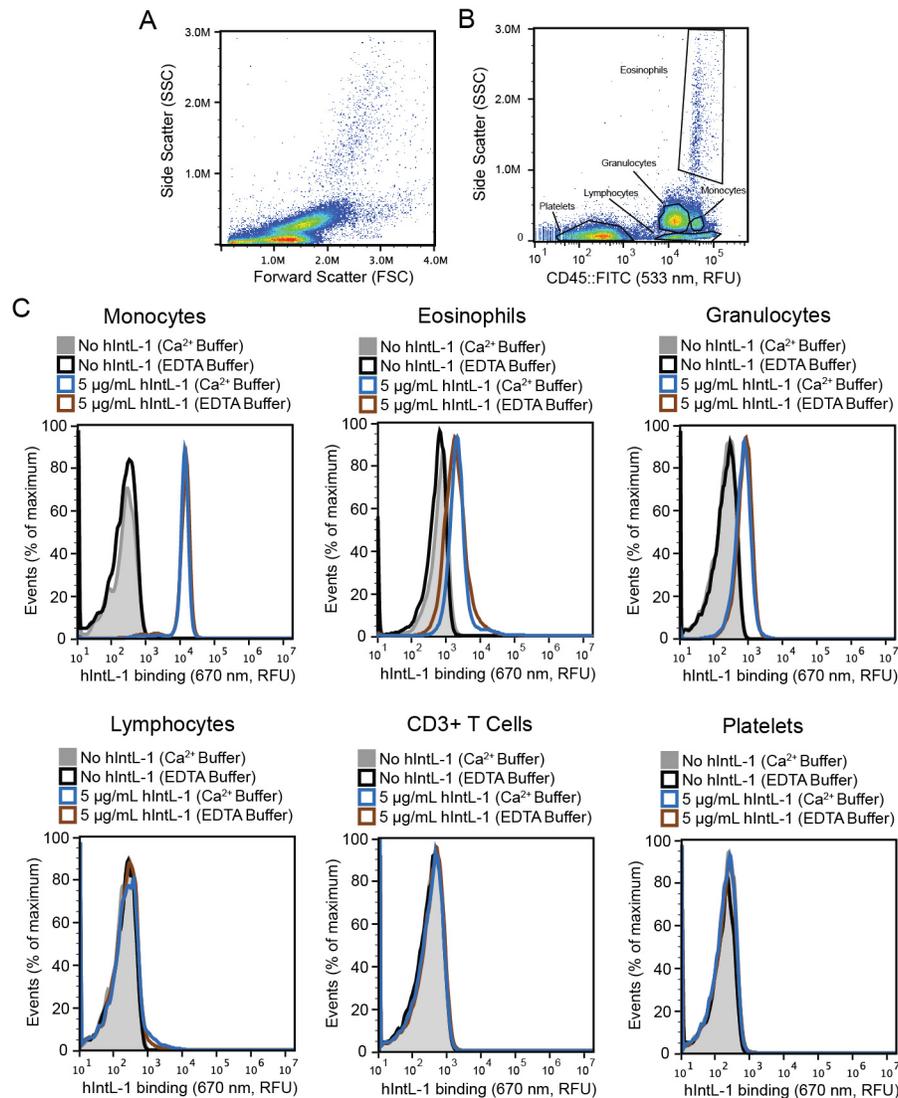


Figure 5-8. hIntL-1 binding to human immune cells. (A) Human immune cells isolated from whole blood were analyzed by flow cytometry to profile hIntL-1 binding. (B) An anti-CD45 antibody was used to identify cell populations. The gates and cell types used in this experiment are included in this dot plot. (C) Differential binding of hIntL-1 to different immune cell types. The interaction is not dependent on calcium ions. This is representative of greater than five independent experiments from five different human donors. This data was collected by my undergraduate research assistant Lucas C. Zarling.

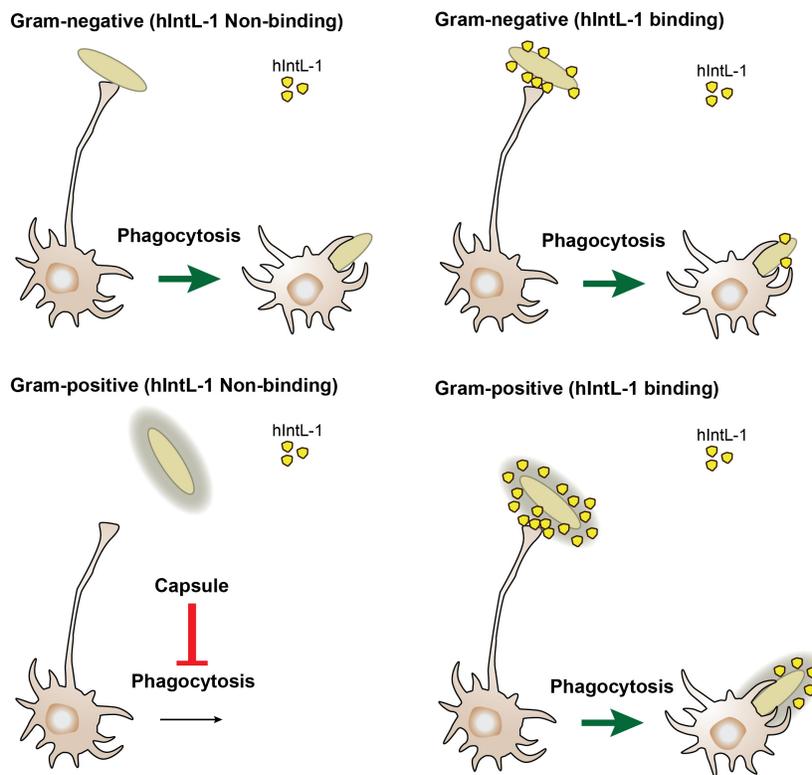


Figure 5-9. Hypothesis for the enhanced recognition of Gram-positive bacteria by hIntL-1 and its interaction with human immune cells. hIntL-1 mediated crosslinking of Gram-positive bacteria and immune APCs may be a mechanism for recognition and immune response to encapsulated bacteria.

5.9 Discussion

Here we describe experiments that provide clues toward the biological function of hIntL-1. 1. Among the most surprising results from this section was the amount of strains representative of human commensal bacteria that are bound by hIntL-1 and presumably express exocyclic 1,2-diols within cell surface glycoconjugates. We identified 12 strains from diverse taxa that are robustly bound by the lectin in a calcium ion-dependent manner. Another result from this strain analysis is the dearth of Gram-negative bacteria bound by hIntL-1 (i.e., Bacteroidetes and Proteobacteria). This was surprising as the LPS core contains KDO, KO, and heptoses, all

proposed ligands of hIntL-1. This suggests that the LPS core is often inaccessible to lectins. We hypothesize that the LPS core is shielded by O-antigen and capsular polysaccharides.

The results from our analysis of microbiome strains suggest multiple future research avenues. We envision leveraging the results of this strain analysis to search for genomic signatures useful in predicting hIntL-1 binding. They also provide a foundation for characterizing hIntL-1 binding to microbial communities and suggest strains useful in future experiments employing mouse models, both conventionally raised and gnotobiotic, to assay the effect of intelectin bound bacteria on mammalian health

We leveraged our identification of strains that are bound by hIntL-1 in isolation to explore lectin binding to synthetic microbial communities. This revealed that hIntL-1 binding to microbes can be strikingly different when assayed individually as a purified strain or assayed within a community. Our results suggest that within communities of bacteria representative of natural communities, competition exists for hIntL-1 binding. Specifically, Gram-negative bacteria appear susceptible to inhibition of hIntL-1 binding in the presence of competing Gram-negative and Gram-positive bacteria. We hypothesize that this is because LPS O-antigens are less avidly bound by hIntL-1 due of decreased epitope density and prevalence as compared to CPS.

We also demonstrated that hIntL-1 binding is sensitive to the context of the assay and the amount of lectin available. This result was surprising because when the strains were assayed individually, binding appeared monomodal, either all or none of the cells interacted with the lectin. In contrast to this, the binding within a community could be modulated through changing the concentration of hIntL-1. This feature is particularly evident when at least one Gram-negative

binding strain was included in the community (**Figure 5-4A - C**). We do not believe that free lectin was limiting in these experiments as quantitative detection of bacteria in an exclusively Gram-positive community was obtained down to 0.05 μg of hIntL-1 per $1\text{E}6$ cells. The ability to alter the bacteria targeted by hIntL-1 within a community based on amount of lectin (i.e., expression level) suggests a novel paradigm for bacteria:lectin interactions in vivo. Lectin interactions within a community may be more similar to a rheostat that is tunable, than to the digital yes:no decision that is observed when individual strains are assayed in vitro. It has not escaped us that the human body may use hIntL-1 expression levels, and lectins expression levels in general, to selectively target and shape microbial communities through alteration of lectin:microbe binding pools.

We also assayed binding to a natural microbial community, the bacteria isolated from a human fecal sample. This revealed that 20 – 30% of the bacteria resident in our donor's GI have the ability to be bound by hIntL-1. We are using FACS and sequencing, GlycoSEQ, to characterize this microbial population. We hypothesize GlycoSEQ experiments may reveal information about localized hIntL-1 expression, hIntL-1 interactions with microbial strains, and hIntL-1's role in intestinal homeostasis.

Lastly, we identified selective binding of hIntL-1 to human innate immune cells. This result suggests a plethora of potential biological functions for the lectin. One possibility we are especially excited about is that hIntL-1 may function as a complement-independent opsonin. In **Chapter 1** of this thesis I described a collagen-like domain that is present within some immune lectins (i.e., ficolins and collectins) that provides them the ability to bind serum proteases and activate the lectin pathway of complement. Intellectins do not contain a collagen-like domain.

Instead, intelectins may be secreted into the extracellular space where they can bind the surface of microbes. When a microbe begins to breach the mucus layer of mucosal tissues, antigen presenting cells may interact with microbe bound intelectin and phagocytose the microbial cell (**Figure 5-10**). While at this point the hypothesis is speculative, many of the tools and reagents to test it are now in place. A similar mechanism may function in the lung where alveolar macrophages are the predominant immune cell type. Stimulation of phagocytosis by hIntL-1 bound bacteria may enhance IgA production at mucosal tissue against hIntL-1 bound cells. Future work in this area will help to answer the interesting biological question of how intelectins function in mammalian immunity.

5.10 Methods

5.10.1 *hIntL-1 Binding to Bacterial Strains*

Bacteria were grown under anaerobic conditions by Robert Kerby in Federico Rey's group. 2 mL of overnight or saturated growth were pelleted at 5,000 RPM. Cells were washed with 5 mL of cold PBS, pelleted by centrifugation, and fixed in 5 mL of cold PBS + 1% formaldehyde for 30 min on ice. Fixation was quenched with addition of 5 mL of PBS + 1 M lysine for 30 min on ice. Fixed cells were pelleted by centrifugation and resuspended in 5 mL of 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂, 0.1 % BSA, and 0.05 % tween-20. 100 µL of the fixed bacteria solution was used for for each staining condition, ~ 3E6 – 30E6 cells. Cells were stained in a total volume of 250 µL. To assay hIntL-1 binding, cell were stained in 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂, 0.1 % BSA, 0.05 % tween-20, 15 µg/mL Strep-hIntL-1, and a 1:250 dilution of an Oyster 645 nm:Anti-Strep-tag II antibody (IBA Bioscience). To assay for calcium ion dependence, the calcium was omitted from the staining procedure and replaced with 5 mM EDTA. An antibody only control was performed with each strain by omitting hIntL-1 from the lectin staining conditions. After staining for two hours at 4 °C, cells were centrifuged, the supernatant removed, and resuspended in 1 mL of staining buffer. Stained cells were analyzed on a BD FACSCalibur or Accuri C6 flow cytometer. To help visualize cells, propidium iodide was added to some samples at a 1:250 dilution (Life Technologies). A minimum of 50,000 events were collected under each condition. The identity of each stain has been confirmed with 16S rRNA sequencing. Data were processed using FlowJo. Histograms were generated and strains that demonstrated a substantial increase in fluorescence in the

presence of hIntL-1, and were completely sensitive to EDTA addition were considered hIntL-1 bound.

5.10.2 hIntL-1 Binding to Synthetic Communities

Stains of bacteria were prepared identically to what is described above. Stains were first analyzed individually to profile each stain, and quantify the density of the cells on a BD Accuri C6 flow cytometer. A minimum of 50,000 events were collected under each condition. Mixtures of bacteria were generated using the quantified cell density reported during the strain analysis. 20E6, and later 10E6 cells were stained in a final volume of 250 μ L. Cells were stained under similar conditions as reported above except the amount of hIntL-1 per sample was varied. After staining for two hours at 4 °C, the cells were split into two samples. One was diluted fourfold with the addition of 400 μ L of staining buffer, these cells were the “diluted” sample, while the other half was centrifuged and resuspended in 500 μ L of staining buffer, these cells were the “washed” sample. Samples were analyzed immediately on a BD Accuri C6 flow cytometer. Data were processed using FlowJo software. Bound cells were quantified where a gate was derived from an unstained sample. Quantified data were graphed using Graphpad Prism 6.

5.11.3 hIntL-1 Binding to Human Fecal Samples

A slurry of human feces was generated by Robert Kerby in the Rey Group by resuspending 200 mg of fecal matter in 5 mL anaerobic PBS. The solution was vortexed occasionally and was left on ice for 2 hours. To generate the sample of fixed bacteria, 2 mL of the slurry was pelleted via centrifugation. The sample was washed once with PBS, and then fixed in PBS + 1 % formaldehyde for 30 min on ice. The reaction was quenched with equal volume addition of PBS + 1 M lysine. Fixed cells were pelleted by centrifugation and resuspended in 5

mL of 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl₂, 0.1 % BSA, and 0.05 % tween-20. To assay hIntL-1 by flow cytometry, 75 – 100 µL of the fixed bacterial solution was used for each staining condition. Cells were staining using an identical protocol to what was described above for assaying hIntL-1 binding to bacterial stains. Calcium ion dependence is used to demonstrate that the binding is carbohydrate mediated and to ensure the assay is working. Cells were analyzed on a BD Accuri C6 or a BD LSR II flow cytometer. A minimum of 50,000 events were collected under each condition. Data were processed using FlowJo software.

5.10.4 Sorting hIntL-1 Bound Bacteria From a Human Fecal Sample

An identical protocol to what has been described for staining a human fecal sample was employed when the cells were to be sorted. Cells were sorted using FACS on a BD Aria II by the staff at the UW-Madison Carbone Cancer Center Flow Cytometry Core Facility with a nozzle pressure of 70 pounds per square inch (PSI). The gating parameters were determined using an unstained sample, antibody only sample, and a sample stained in the presence of EDTA. The cells were gated in hIntL-1 fluorescence vs. a 710 nm filter with a 50 nm window excited by a 488 laser. The cells were spit two ways and 1.5E6 cells were collected per tube.

5.10.5 Sequencing the 16S rRNA gene from hIntL-1 bound sorted bacteria

Bacteria isolated by FACS, and a control sample of input cells, were immediately returned to the Rey lab to prepare the samples for sequencing. The sorted cells were transferred to a 1.5 mL centrifuge tube, the original tube was washed extensively. Cell pellets were stored at -80 °C until processed. The genomic DNA was extracted from each sample by bead beating using standard protocols. The V3-V4 region of the 16S rRNA gene was amplified and sequenced by an Illumina

HiSeq at the UW-Madison Biotechnology Center Next Generation DNA sequencing facility. The data were processed by QIIME using the standard protocol (44).

5.10.6 Lectin Binding to Fecal Microbiota

Fecal microbiota were prepared identically as described above. Fluorescein conjugated plant lectin were purchased from Vector Labs and assayed for binding at a concentration of 15 $\mu\text{g}/\text{mL}$. Cells were stained in 20 mM HEPES (7.4), 150 mM NaCl, 10 mM CaCl_2 , 1 mM MgCl_2 , and 0.5 mM ZnCl_2 , 0.1 % BSA, and 0.05 % tween-20, or removal of all divalent cations and addition of 3 mM EDTA, in a final volume of 250 μL . Cells were stained for two hours at 4 $^\circ\text{C}$, centrifuged, and resuspended in 1 mL of staining buffer. Cells were analyzed on a BD Accuri C6 flow cytometer. A minimum of 50,000 events were collected under each condition. Data were processed using FlowJo software.

5.10.7 hIntL-1 Binding to Human Immune Cells

Whole human blood was obtained through collaboration with Mats Johansson in the Deane Mosher research group. To assay lectin binding to the surface of immune cells, 100 μL of whole blood was diluted in 400 μL 20 mM HEPES (pH 7.4), 150 mM NaCl, 0.1% BSA, 0.05% Tween-20 and either 10mM CaCl_2 or 3 mM EDTA. For assaying hIntL-1 binding, untagged hIntL-1 was expressed in suspension HEK293T cells and secreted hIntL-1 was purified on an immobilized sorbitol resin. Purified hIntL-1 was labeled using an activated Alexa Fluor[®] 647 nm fluorophore according to the manufacturers guidelines (Life Technologies; cat. no., A20186) and assayed at 5 $\mu\text{g}/\text{mL}$. Cell populations were determined by staining with anti-human CD14:PE (BD Biosciences, cat. no., 555398) and anti-human CD45::FITC (BD Biosciences, cat. no., 555482) according to the manufacturers suggested protocol. Cell populations were gated using

side scatter vs. anti-human CD45::FITC staining. Staining was allowed to proceed for 1 hour at room temperature with occasional gentle vortexing. Cells were centrifuged at 1200 RPM and washed three times with the HEPES buffer. Cells were then resuspended in BD Lysing Solution (BD Biosciences; cat. no., 349202) to lyse the red blood cells and fix the remaining primary immune cells. Samples were centrifuged, resuspended in the HEPES buffer, and analyzed on a BD Accuri C6 at the University of Wisconsin-Madison Biophysics Instrumentation Facility. A minimum of 50,000 events were collected under each condition. Data was processed using FlowJo software.

5.11 Contributions

Bacterial strains were cultured and verified via 16S rRNA sequencing by Robert Kerby. Robert Kerby also developed the method for isolating genomic DNA from sorted bacteria and performed the library preparation for Illumina sequencing. The samples were sequenced by the UW–Madison Biotechnology Center. Sequencing data was processed by Robert Kerby with assistance from Julia Kreznar. Lucas C. Zarling analyzed hIntL-1 binding to immune cells.

5.12 Acknowledgements

I would first like to thank Robert Kerby, Julia Kreznar, and the entire group of Professor Federico Rey for our collaboration into the binding of hIntL-1 to bacteria associated with the human microbiome, and for thoughtful discussions about the project. I would also like to acknowledge Lucas C. Zarling, an undergraduate research associate that I had the pleasure of working with for the last four years. Lucas has taken a leadership role in our examination of hIntL-1 binding to immune cells and has been instrumental in planning and performing those experiments. I would also like to acknowledge Mats Johansson and Professor Deane Mosher for

our collaboration profiling hIntL-1 binding to immune cells. It has been a pleasure to work with them and we are thankful for their initial suggestions on potential immune cell interactions and their willingness to help us follow the science as the intelectin project has evolved. I would like to acknowledge Darrell R. McCaslin and Dan Stevens of the Department of Biochemistry BIF for their thoughtful discussion and assistance using the BD Accuri C6 and the *Forté*BIO Octet RED96. Flow cytometry and microbial sorting were performed at the UW-Madison Carbone Cancer Center Flow Cytometry Core (P30 CA014520). I am thankful for Dagna Sheerar and Rachael Sheridan with assistance and technical expertise in performing those experiments. And lastly I would like to thank the entire Kiessling Group for their thoughtful discussions on this project, but specifically Amanda Dugan and Christine Isabella as we are in the intelectin trenches together.

5.13 References

1. Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007) The human microbiome project. *Nature* **449**, 804-810
2. Human Microbiome Project, C. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214
3. Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., Meta, H. I. T. C., Antolin, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariáz, G., Dervyn, R., Foerster, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D., and Bork, P. (2011) Enterotypes of the human gut microbiome. *Nature* **473**, 174-180
4. Gallo, R. L., and Hooper, L. V. (2012) Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12**, 503-516
5. Belkaid, Y., and Hand, T. W. (2014) Role of the microbiota in immunity and inflammation. *Cell* **157**, 121-141
6. Johansson, M. E., Sjovall, H., and Hansson, G. C. (2013) The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 352-361
7. Cash, H. L., Whitham, C. V., Behrendt, C. L., and Hooper, L. V. (2006) Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126-1130
8. Macpherson, A. J., McCoy, K. D., Johansen, F. E., and Brandtzaeg, P. (2008) The immune geography of IgA induction and function. *Mucosal Immunol.* **1**, 11-22
9. Mowat, A. M., and Agace, W. W. (2014) Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14**, 667-685
10. Tsuji, S., Uehori, J., Matsumoto, M., Suzuki, Y., Matsuhisa, A., Toyoshima, K., and Seya, T. (2001) Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *J. Biol. Chem.* **276**, 23456-23463
11. French, A. T., Bethune, J. A., Knight, P. A., McNeilly, T. N., Wattedegera, S., Rhind, S., Miller, H. R., and Pemberton, A. D. (2007) The expression of intelectin in sheep goblet cells and upregulation by interleukin-4. *Vet. Immunol. Immunopathol.* **120**, 41-46
12. Lee, J. K., Baum, L. G., Moremen, K., and Pierce, M. (2004) The X-lectins: a new family with homology to the *Xenopus laevis* oocyte lectin XL-35. *Glycoconj. J.* **21**, 443-450
13. Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zarling, L. C., Splain, R. A., Smith, D. F., Cummings, R. D., Paulson, J. C., Forest, K. T.,

- and Kiessling, L. L. (2015) Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.* **22**, 603-610
14. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19-34
 15. Holmskov, U., Thiel, S., and Jensenius, J. C. (2003) Collectins and ficolins: Humoral lectins of the innate immune defense. *Annu. Rev. Immunol.* **21**, 547-578
 16. Wangkanont, K., Wesener, D. A., Vidani, J. A., Kiessling, L. L., and Forest, K. T. (2016) Structures of Xenopus Embryonic Epidermal Lectin Reveal a Conserved Mechanism of Microbial Glycan Recognition. *J. Biol. Chem.* **291**, 5596-5610
 17. Thomsen, T., Schlosser, A., Holmskov, U., and Sorensen, G. L. (2011) Ficolins and FIBCD1: soluble and membrane bound pattern recognition molecules with acetyl group selectivity. *Mol. Immunol.* **48**, 369-381
 18. Fujita, T. (2002) Evolution of the lectin-complement pathway and its role in innate immunity. *Nat. Rev. Immunol.* **2**, 346-353
 19. Lee, J. K., Schnee, J., Pang, M., Wolfert, M., Baum, L. G., Moremen, K. W., and Pierce, M. (2001) Human homologs of the Xenopus oocyte cortical granule lectin XL35. *Glycobiology* **11**, 65-73
 20. Schaffler, A., Neumeier, M., Herfarth, H., Furst, A., Scholmerich, J., and Buchler, C. (2005) Genomic structure of human omentin, a new adipocytokine expressed in omental adipose tissue. *Biochim. Biophys. Acta.* **1732**, 96-102
 21. Pemberton, A. D., Knight, P. A., Gamble, J., Colledge, W. H., Lee, J. K., Pierce, M., and Miller, H. R. (2004) Innate BALB/c enteric epithelial responses to *Trichinella spiralis*: inducible expression of a novel goblet cell lectin, intelectin-2, and its natural deletion in C57BL/10 mice. *J. Immunol.* **173**, 1894-1901
 22. Kerr, S. C., Carrington, S. D., Oscarson, S., Gallagher, M. E., Solon, M., Yuan, S. P., Ahn, J. N., Dougherty, R. H., Finkbeiner, W. E., Peters, M. C., and Fahy, J. V. (2014) Intelectin-1 Is a Prominent Protein Constituent of Pathologic Mucus Associated with Eosinophilic Airway Inflammation in Asthma. *Am. J. Respir. Crit. Care Med.* **189**, 1005-1007
 23. Kuperman, D. A., Lewis, C. C., Woodruff, P. G., Rodriguez, M. W., Yang, Y. H., Dolganov, G. M., Fahy, J. V., and Erle, D. J. (2005) Dissecting asthma using focused transgenic modeling and functional genomics. *J. Allergy Clin. Immunol.* **116**, 305-311
 24. Voehringer, D., Stanley, S. A., Cox, J. S., Completo, G. C., Lowary, T. L., and Locksley, R. M. (2007) *Nippostrongylus brasiliensis*: Identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. *Exp. Parasitol.* **116**, 458-466
 25. Datta, R., deSchoolmeester, M. L., Hedeler, C., Paton, N. W., Brass, A. M., and Else, K. J. (2005) Identification of novel genes in intestinal tissue that are regulated after infection with an intestinal nematode parasite. *Infect. Immun.* **73**, 4025-4033
 26. Pemberton, A. D., Knight, P. A., Wright, S. H., and Miller, H. R. P. (2004) Proteomic analysis of mouse jejunal epithelium and its response to infection with the intestinal nematode, *Trichinella spiralis*. *Proteomics* **4**, 1101-1108

27. (2009). in *Essentials of Glycobiology* (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. eds.), 2nd Ed., Cold Spring Harbor (NY). pp
28. Stenutz, R., Weintraub, A., and Widmalm, G. (2006) The structures of Escherichia coli O-polysaccharide antigens. *FEMS Microbiol. Rev.* **30**, 382-403
29. Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kalltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**, 262-269
30. Tytgat, H. L., and Lebeer, S. (2014) The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol. Mol. Biol. Rev.* **78**, 372-417
31. Tytgat, H. L., and de Vos, W. M. (2016) Sugar Coating the Envelope: Glycoconjugates for Microbe-Host Crosstalk. *Trends Microbiol.*
32. Kjaer, T. R., Hansen, A. G., Sorensen, U. B., Holm, A. T., Sorensen, G. L., Jensenius, J. C., and Thiel, S. (2013) M-ficolin binds selectively to the capsular polysaccharides of Streptococcus pneumoniae serotypes 19B and 19C and of a Streptococcus mitis strain. *Infect. Immun.* **81**, 452-459
33. Krarup, A., Sorensen, U. B., Matsushita, M., Jensenius, J. C., and Thiel, S. (2005) Effect of capsulation of opportunistic pathogenic bacteria on binding of the pattern recognition molecules mannan-binding lectin, L-ficolin, and H-ficolin. *Infect. Immun.* **73**, 1052-1060
34. Neth, O., Jack, D. L., Dodds, A. W., Holzel, H., Klein, N. J., and Turner, M. W. (2000) Mannose-binding lectin binds to a range of clinically relevant microorganisms and promotes complement deposition. *Infect. Immun.* **68**, 688-693
35. Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J. P., Heimburg-Molinaro, J., Ju, T., Molinaro, R. J., Rivera-Marrero, C., Xia, B., Smith, D. F., and Cummings, R. D. (2010) Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* **16**, 295-301
36. Rajilic-Stojanovic, M., and de Vos, W. M. (2014) The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38**, 996-1047
37. Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837-848
38. Herget, S., Toukach, P. V., Ranzinger, R., Hull, W. E., Knirel, Y. A., and von der Lieth, C. W. (2008) Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC Struct. Biol.* **8**, 35
39. Kleerebezem, M., Hols, P., Bernard, E., Rolain, T., Zhou, M., Siezen, R. J., and Bron, P. A. (2010) The extracellular biology of the lactobacilli. *FEMS Microbiol. Rev.* **34**, 199-230
40. Iliev, I. D., Funari, V. A., Taylor, K. D., Nguyen, Q., Reyes, C. N., Strom, S. P., Brown, J., Becker, C. A., Fleshner, P. R., Dubinsky, M., Rotter, J. I., Wang, H. L. L., McGovern, D. P. B., Brown, G. D., and Underhill, D. M. (2012) Interactions Between Commensal Fungi and the C-Type Lectin Receptor Dectin-1 Influence Colitis. *Science* **336**, 1314-1317

41. Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., Lewis, J. D., and Bushman, F. D. (2013) Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS One* **8**, e66019
42. Palm, N. W., de Zoete, M. R., Cullen, T. W., Barry, N. A., Stefanowski, J., Hao, L., Degnan, P. H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J. H., Goodman, A. L., and Flavell, R. A. (2014) Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000-1010
43. Kau, A. L., Planer, J. D., Liu, J., Rao, S., Yatsunenko, T., Trehan, I., Manary, M. J., Liu, T. C., Stappenbeck, T. S., Maleta, K. M., Ashorn, P., Dewey, K. G., Houpt, E. R., Hsieh, C. S., and Gordon, J. I. (2015) Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Science Transl. Med.* **7**
44. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335-336
45. Tsuji, S., Yamashita, M., Hoffman, D. R., Nishiyama, A., Shinohara, T., Ohtsu, T., and Shibata, Y. (2009) Capture of heat-killed *Mycobacterium bovis* bacillus Calmette-Guerin by intelectin-1 deposited on cell surfaces. *Glycobiology* **19**, 518-526

Appendix 1

Small Molecule Inhibition of UDP-Galactopyranose Mutase in *C. elegans*

Portions of this work have been published in:

Wesener, D. A., May, J. F., Huffman E. M., and Kiessling, L. L. UDP-Galactopyranose Mutase in Nematodes. *Biochemistry*. **2013**, 52(25): 4391-4398.

&

Wesener, D.A., Wangkanont, K., McBride, R., Song, X., Kraft, M.B., Hodges, H.L., Zharling, L.C., Splain, R.S., Smith D.F., Cummings, R.D., Paulson, J.C., Forest, K.T., and Kiessling, L.L. Recognition of microbial glycans by human intelectin-1. *Nature Structural and Molecular Biology*. **2015**. 22, 603-610.

A1.1 Abstract

In **Chapter 2** of this thesis I reported the identification of small molecule inhibitors of the UGM enzyme from the nematode *Caenorhabditis elegans*. Within the phylum Nematoda lie many parasitic species that represent pests of crops, livestock, and humans. Expressed sequence tag (EST) analysis from several parasitic species suggests that UGM enzymes are found in many, and may represent a novel point for therapeutic intervention. The free living nematode *C. elegans* has been used as a model organism for studying processes such as development, cell signaling, and neuron function. Here, I employed *C. elegans* to assay the effect of compound **2.3** on nematode viability and cuticle integrity. Using RNA interference (RNAi) technology, a sodium hypochlorite sensitivity assay, a proteinase K degradation assay, and lectin staining of whole nematodes, a minimal effect of compound addition to culture was observed. Based on the robust enzyme inhibition measured in vitro, I hypothesize the lack of in vivo efficacy is the result of low permeability of the small molecule, and/or rapid metabolism by *C. elegans* detoxification enzymes (i.e. oxidases and UDP-glucuronosyltransferases). Results described herein will be valuable in directing future chemical biology studies into nematode cell surface glycosylation.

A1.2 Introduction

The phylum Nematoda is composed of both free living and parasitic organisms. While many free living species reside in the soil where they consume and shape microbial communities (1), half of the nematode species are predicted to be parasitic. Parasitic nematodes infect plants, livestock, and humans, and collectively cause billions of dollars in economic losses and have the ability to infect an estimated 4 billion people worldwide (2,3). As resistance to current anthelmintics is increasing, interest in novel small molecule inhibitors that can modulate nematode viability, or recognition of nematodes by a host immune system are of intense interest (4-6). This has resulted in the approval and usage of ZOLVIX[®] (monepantel) in goats and sheep, the first broad spectrum anthelmintic brought to market since the 1980s (7).

In **Chapter 2** of this thesis, I described the identification of compounds **2.3** and **2.4** as potent in vitro inhibitors of *C. elegans* UGM (**Figure 2-7 & Table 2-2**). I next was interested in the potential utility of these compounds in whole animals. *C. elegans* is an attractive laboratory model because of its genetic tractability, it is amenable to RNAi gene knockdown (8,9), is easy and cost effective to culture, and is transparent (10). Additionally, *C. elegans* has been used extensively as a model of helminth biology (11,12). Prior work has explored the biosynthesis and biological roles of nematode glycoconjugates both analytically and biologically using *C. elegans* (13-15).

Recently, CeUGM gene, *glf-1*, was studied in *C. elegans* (16). This revealed many important clues toward the potential role of Galf in nematode biology, and provided a foundation for my chemical biology studies. First, CeUGM was visualized to hypodermal cells, with robust expression observed in seam cells. Hypodermal cells and seam cells are responsible for synthesis

of a thick, multilayered collagenous cuticle, and a collection of loosely associated glycoconjugates termed the glycocalyx (17,18). The cuticle and glycocalyx are responsible for protecting the worm from, and interacting with the environment. Interestingly, although not entirely surprising based on its expression localization, deletion or knock-down of *glf-1* resulted in structural defects in the cuticle, enhanced binding to the glycocalyx by *E. coli* cells, compromised cuticle integrity, increased surface staining by lectins, and a roughly 90 % decrease in viability. These results suggest an essential role for GalF in *C. elegans* biology, but to date, GalF has yet to be identified in any nematode glycan. We were interested in using compounds **2.3** and **2.4** as chemical probes of GalF in nematodes. We hypothesized that upon addition to *C. elegans* culture, we could recapitulate the defects first reported using genetic tools in Novelli *et al* (16). Compounds **2.3** and **2.5** were added to *C. elegans* culture and their effect on viability, resistance to proteases and sodium hypochlorite (bleach), and changes in surface glycosylation were assayed. While my results suggest that the 2-aminothiazole based small molecules described here may not be effective for inhibiting UDP-GalF biosynthesis in nematodes, they provide chemical leads for future small molecule development and describe assays useful in measuring their effectiveness.

A1.3 *glf-1* Knockdown Reduces Cuticle Stability

A potential role for GalF in nematode biology was first described in 2009 by Novelli *et al* (16). Using RNAi targeted to *glf-1* (19), and deletion of chromosomal *glf-1*, the authors suggest that CeUGM and its product, UDP-GalF, perform an essential role in nematode viability and resistance to chemical stresses. I first attempted to reproduce these results and assay the effect of *glf-1* knockdown in wild type N2 *C. elegans* using a similar sodium hypochlorite sensitivity, and

proteinase K sensitivity assays. For these assays, I employed the Ahringer Lab RNAi *E. coli* strain expressing *glf-1* from a T7 promoter, and a control RNAi *E. coli* strain. Adult worms were added to a solution of sodium hypochlorite and sodium hydroxide and the time required for worm lysis was recorded. While wild-type N2 and the RNAi control worms lasted roughly 90 seconds in the solution, *glf-1* RNAi *C. elegans* lysed after an average of 30 seconds (**Figure A1-1A**). As a second assay of cuticle stability and resistance to environmental stress, I optimized a proteinase K based protease resistance assay. Similar to the sodium hypochlorite sensitivity assay, knockdown of *glf-1* using RNAi substantially decreased resistance to degradation by proteinase K (**Figure A1-1B**). These data are in agreement with what has been published previously and suggest a role for Galf in nematode resistance to environmental stresses (16).

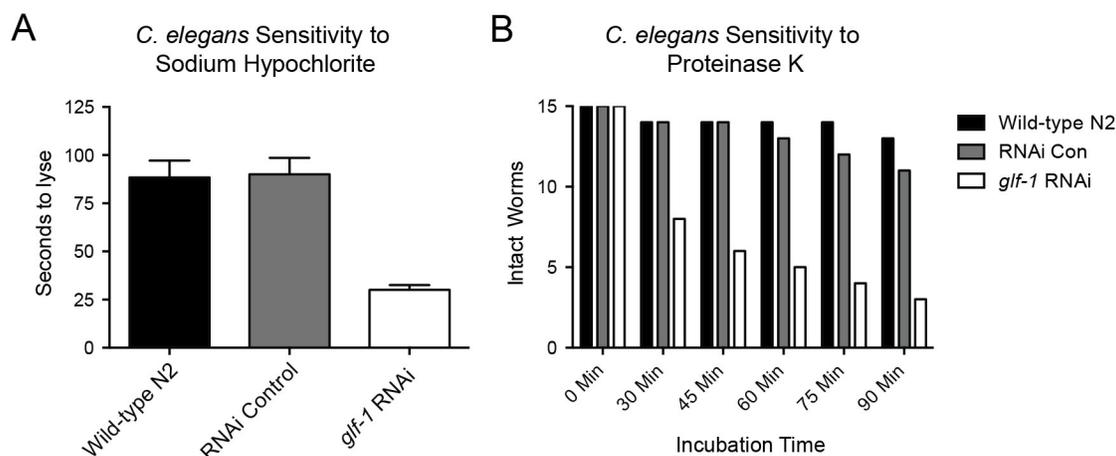


Figure A1-1. Depletion of *glf-1* decreases *C. elegans* resistance to environmental stress. (A) Resistance to oxidative stress is attenuated when *glf-1* is knocked down using RNAi by *E. coli* feeding. $n = 10$ adult worms for each condition. Error bars represent the s.d. of the mean. This result is representative of greater than three independent experiments. (B) Resistance to extracellular proteases is also decreased when *glf-1* mRNA is depleted. 15 adult worms were used at the initiation of the experiment for each condition. Worms were scored visually each 15 minutes using a microscope.

A1.4 Treatment of *C. elegans* with Compounds 2.3 and 2.5

Based on the in vitro results from **Chapter 2** and the assays described above that suggest an essential role for CeUGM in nematode biology, we were excited to assay compounds **2.3** and **2.5** in whole *C. elegans*. To do so, I added stock solutions of compound dissolved in DMSO to either *C. elegans* liquid culture or nematodes cultured on NGM agar. Initially, addition to liquid culture was used, although interference with the sodium hypochlorite and proteinase K assay under those growth conditions were observed. Ultimately, compound addition to molten NGM agar was used and the effects were assayed under standard growth conditions. The level of DMSO was limited to 0.5% for all the experiments as DMSO appeared to alter worm physiology at higher concentrations. Compound **2.3** was used as an inhibitor of CeUGM, and compound **2.5** was included as a scaffold control. Both were added to NGM agar plates at concentrations of 25 μM , 50 μM , 100 μM and 200 μM . Above those concentrations, precipitation of the compounds in the agar medium was observed. Addition of either compound, or the DMSO solvent control, did not appear to visually effect the viability, motility, or reproductive capability of *C. elegans* (data not shown). This lack of an apparent phenotype is in contrast of what we hypothesized based on the RNAi results (16).

I then characterized small molecule treated worms using the sodium hypochlorite sensitivity assay and the proteinase K degradation assay. Using both assays, the addition of compound **2.3** failed to phenocopy the results from knockdown of *glf-1* (**Figure A1-2**). The addition of DMSO appeared to alter the cuticle. I hypothesize this may have been through upregulation of collagen synthesis or collagen crosslinking. These results suggest that the 2-aminothiazole based inhibitors described in **Chapter 2** may not be useful in whole nematodes.

Also, these data suggest that DMSO as a small molecule vehicle may be problematic when assaying cuticle and glycocalyx stability in *C. elegans*.

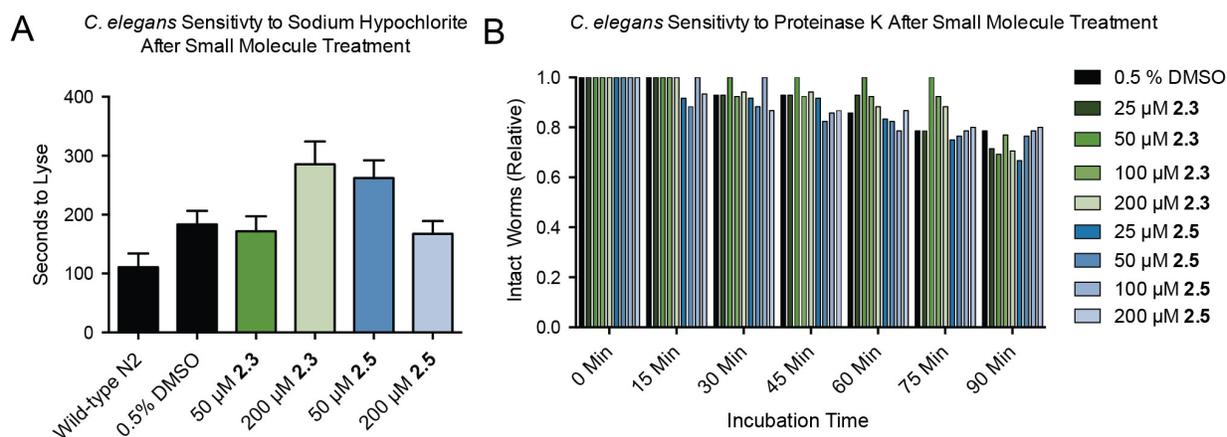


Figure A1-2. Small molecule treatment of wild-type *C. elegans* does not phenocopy the results using RNAi and genetic deletion. (A) Treatment of wild-type N2 *C. elegans* with compound **2.3** does not enhance sensitivity to oxidants. Compound **2.5** was used as a scaffold control. $n = 3-7$ adult worms for each condition. Error bars represent the s.d. of the mean. This result is representative of at least two independent experiments. (B) Resistance to proteinase K is not altered upon treatment of N2 *C. elegans* with compound **2.3**. 12-17 adult worms were used at the initiation of the experiment for each condition. The worms were scored visually each 15 minutes using a microscope. The relative portion of intact nematodes remaining at each time point is presented.

A1.5 Lectin Staining of *C. elegans* Treated with Compound 2.3

Another phenotype described with the deletion of *glf-1* was increased binding of soluble lectins to the surface of the nematode (16). We hypothesized that detection of changes in surface glycosylation may be a more direct output of small molecule inhibition of CeUGM. For this, I chose to stain intact *C. elegans* using ConA, an α -D-mannose and α -D-glucose specific plant lectin originally isolated from jack-beans. Wild-type *C. elegans* have very low binding by lectins and examples of specific binding are limited. In agreement with previous reports, when *glf-1* is knocked down, surface binding to Con A increases (**Figure A1-3**). This result suggests an

alteration to the surface coat or glycocalyx of the nematode. When wild-type *C. elegans* were treated with 200 μ M of compound **2.3**, only faint surface staining was observed (**Figure A1-3**). Phenotypically, this resembled more of a wild-type like stain than a nematode whose *glf-1* was depleted. These lectin staining results, combined with the above assays of cuticle integrity suggest that compound **2.3** cannot enter the necessary nematode cells and inhibit CeUGM to elicit a similar effect as RNAi knockdown.

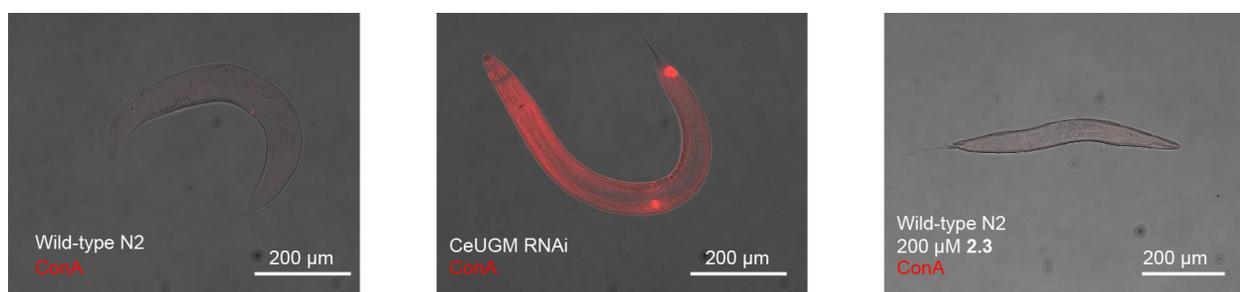


Figure A1-3. Surface staining of intact *C. elegans* with the plant lectin Con A. *C. elegans*, either wild-type, wild-type fed *glf-1* expressing RNAi *E. coli*, or wild-type treated with compound **2.3** were stained with the mannose/glucose specific lectin Con A. All images were processed identically using Image J. These images are representative of multiple analyzed nematodes.

A1.6 Staining *C. elegans* with hIntL-1

In **Chapter 2** I highlighted the proposed importance of *Galf* metabolism in nematodes. Based on the expression localization of *glf-1* and the observed phenotypes when *glf-1* was knocked down or deleted, we postulate that *Galf* is localized to surface glycoconjugates. Within **Chapter 4** of this thesis I described the characterization of the human lectin, hIntL-1 (20). Additionally, hIntL-1 has been reported to be upregulated in mammals upon intestinal nematode infection (21,22). If *Galf* is localized to the surface of nematodes, hIntL-1 should bind to the surface of *C. elegans*, and binding should specifically decrease after knockdown of *glf-1*.

Human IntL-1 conditioned mammalian cell culture medium was used to stain both wild-type N2 and N2 *C. elegans* subjected to *glf-1* knockdown via feeding. hIntL-1 did not bind to wild-type worms, but bound robustly when *glf-1* was decreased (**Figure A1-4**). Initially, we were surprised by this result, however examples of lectin binding to wild-type *C. elegans* are very limited. This is likely due to selection for decreased antigenicity of the nematode surface coat to assist in immune evasion. I hypothesized that the ligand of hIntL-1 seen when *glf-1* was depleted may be a surface glycolipid or protein. To test this, I also stained *glf-1* RNAi *C. elegans* in the presence of the detergent Triton™ X-100. Addition of Triton™ X-100 removes surface glycolipid ligands and proteins not covalently bound (16). Addition of the detergent decreased hIntL-1 binding to the surface of the nematode (**Figure A1-5**), although binding was still observed. This result suggests that in *glf-1* depleted cells, the surface of the nematode is altered which leads to hIntL-1 binding exocyclic diol ligands present in surface ligands both covalently and notcovalently attached to the cuticle. Another explanation is *glf-1* knockdown results in significant cuticle and glycocalyx remodeling that exposes neoepitope hIntL-1 ligands. It would be interesting to screen hIntL-1 binding to a nematode glycan microarray in an attempt to identify its molecular targets. A nematode array may be generally useful as nematode glycoconjugates are postulated to modulate the host immune response and effect immune evasion (23).

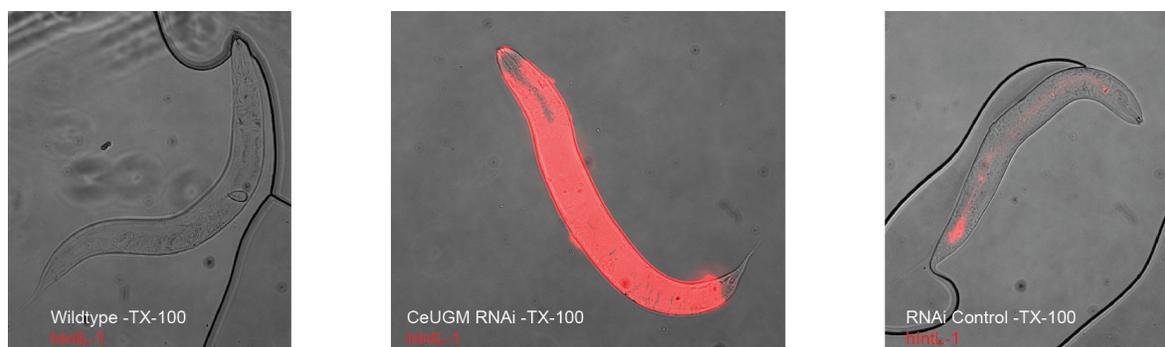


Figure A1-4. Surface staining of intact *C. elegans* with hIntL-1. *C. elegans*, wild-type N2, wild-type fed *glf-1* expressing RNAi *E. coli*, or wild-type fed a control RNAi were stained with hIntL-1 conditioned mammalian cell culture medium. Bound hIntL-1 was detected using a polyclonal antibody and a fluorescent secondary. No detergents were included under these staining conditions. All images were processed identically using Image J.

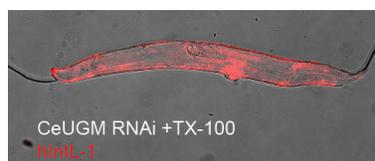


Figure A1-5. Addition of Triton X-100 detergent attenuated hIntL-1 binding to the surface of *glf-1* knockdown *C. elegans*. Triton X-100 was included to assay if a lipid is functioning as the hIntL-1 ligand in *glf-1* RNAi treated nematodes. The image was processed using Image J.

A1.7 Conclusions

While the in vitro protein activity assay suggested that compound **2.3** would be an effective inhibitor of CeUGM, I was unable to recapitulate the phenotype of genetic knockdown or deletion of *glf-1* (16). Despite this, several important results were derived from this work. First, the use of DMSO as a solvent appears to be incompatible with assays of cuticle and glycocalyx stability. This is evidenced by the increase in the time it takes the nematode to lyse when treated with DMSO vehicle and the increased error bars of all samples that include DMSO

(Figure A1-2A). Even visually, the nematodes appear altered. This suggests that in future chemical biology experiments to assay inhibition of enzymes localized to seam cells and involved in cuticle biosynthesis, the use of other solvents is advised (i.e., ethanol).

While the exact reason for the failure of **2.3** in whole animals is unknown, I hypothesize it is due to some of the functional groups of the molecule. Certain functional groups are known to have deleterious effects on small molecules efficacy. One example of this is the inclusion Michael-acceptors that are attacked by cellular thiols (24). Conjugate addition by nucleophiles within the reducing environment of a cells results in lowering the effective concentration of the small molecule inside the cell. Another example is the propensity of anionic groups, such as phosphates or carboxylates, to inhibit a compounds ability to cross a cell membrane and enter the cell (25,26). In the process of this work, a study on the effects of different functional groups on *C. elegans* small molecule uptake and efficacy was published (27). This study highlighted the difficulty in treating *C. elegans* with small molecules, as less than 10% of 1,000 drug-like compounds failed to accumulate inside of the nematode to appreciable concentrations. Another key finding was that the presence of a carboxylate functional group is the single best predictor of a compound that cannot accumulate within a nematode. The carboxylate is thought to 1) prevent efficient cell membrane permeability, and 2) be highly prone to glucuronidation and inactivation/export (27,28). We suspect that this is happening to compound **2.3**. We propose modifying the existing structure to test these possible mechanisms; the carboxylate could be protected via formation of an ester bond, removing the anionic nature of the molecule until uptake by the nematode, or a functional group of similar sterics and electronics could replace the carboxylate altogether. If the second is performed, maintaining and anionic nature is likely

essential for small molecule binding, but removal of the oxygen atoms is suggested to help prevent glycuronidation and export.

Lastly, I stained *C. elegans* with hIntL-1, a human lectin specific for exocyclic 1,2-diol containing ligands such as Gal β . Knockdown of *glf-1* actually resulted in enhanced binding of the lectin to the surface of the nematode. When detergent was included in the staining procedure, hIntL-1 binding was markedly decreased and became punctate. This suggests that a hIntL-1 ligand may be detergent sensitive, such as a glycolipid or glycoprotein only loosely associated with the glycocalyx. Future experiments toward identifying Gal β in a nematode glycan could employ hIntL-1 in a far-Western blot of *C. elegans* lysates, or use hIntL-1 to enrich ligands for mass spectrometry and analytical chemistry analysis (29).

A.1.8 Methods

A1.8.1 *C. elegans* Culture

Wild type N2 *C. elegans* were obtained from the laboratory of Judith Kimble (University of Wisconsin–Madison, Department of Biochemistry) and maintained on either Nematode Growth Medium (NGM) agar plates or in liquid culture at 22 °C (30). On plates, a small lawn of *E. coli* OP50 was used as a food source. When cultured in solution, a small amount of OP50 culture was added. Worms were maintained every 3-5 days by transfer of 5-10 hermaphrodites to a fresh NGM agar plate.

RNAi via *E. coli* feeding was used to knockdown *glf-1* (31). Bacteria expressing *glf-1* RNA (H04M03.4) and a control *E. coli* strain were obtained from the Ahringer Lab collection(8). The RNAi *E. coli* strain from the Vidal lab collection was also obtained and similar results were observed upon feeding (GE Healthcare; cat. # RCE1182-202297092). The *E. coli* were cultured in LB supplemented with 50 µg/mL carbenicillin and 12.5 µg/mL tetracycline. RNA transcription was induced with the addition of 1 mM IPTG to NGM agar growth plates. *C. elegans* were passaged consecutively onto RNAi plates at least three times prior to being used to assay for *glf-1* knockdown.

A1.8.2 Small Molecule Treatment of *C. elegans*

To treat *C. elegans* with compounds **2.3** and **2.5**, serial dilutions of compound dissolved in DMSO were used to maintain the level of DMSO constant at 0.5 %. 0.5 % DMSO was used as a solvent control. Compounds were added to liquid agar just prior to addition to a 24-well plate. The plates were allowed to cool and a small lawn of *E. coli* OP50 was added. Three hermaphrodite worms were added to each plate. Each compound was assayed at concentrations

of 25 μ M, 50 μ M, 100 μ M, and 200 μ M. Worms were cultured at least three days prior to assaying.

A1.8.3 Sodium Hypochlorite Sensitivity Assay

To assay for cuticle integrity, a sodium hypochlorite sensitivity assay was performed similarly as described (16). A fresh solution of sodium hypochlorite (available chlorine 10-15%; Sigma Aldrich; cat. # 425044), 1 M NaOH, and water was prepared immediately before each assay at a volume:volume ratio of 1:1:5. Ten μ L of this solution was spotted onto a glass cover slip, and to that one *C. elegans* was added. The reaction was monitored through a microscope and the time required for each nematode to lyse was recorded. For wild type N2 and RNAi control nematodes, lysis usually occurred at the mouth or anus, while CeUGM RNAi nematodes typically fractured in half, extruding their entire abdomen.

A1.8.4 Proteinase K Sensitivity Assay

To assay for cuticle integrity, a protease sensitivity assay was performed similarly as described(16). To a solution of 3 mM CaCl₂, proteinase K (New England Biolabs) was added to a final concentration of 2 mg/mL. Adult worms, 12-25, were added and the number of intact, living worms was monitored every 15 minutes.

A1.8.5 Lectin Staining

C. elegans were removed from plates by washing with 50 mM HEPES (pH 7.4), 150 mM NaCl, 10 mM CaCl₂, 1 % bovine serum albumin, and 0.5 % triton-X 100. Worms were stained in the HEPES solution with the addition of lectin, 50 μ g/mL ConA:biotin (Vector Labs) or conditioned hIntL-1 mammalian expression medium, at room temperature for two hours. The animals were washed and lectin was detected with Streptavidin::Alexa Fluor 555 nm (Life

Technologies) or a sheep anti-hIntL-1 polyclonal antibody (R&D Systems) and a donkey anti-sheep IgG Alexa Fluor 555 nm (Life Technologies). In the hIntL-1 stained sample lacking detergent, the Triton X-100 was omitted. Animals were washed two times and fixed in a solution of PBS plus 4 % formaldehyde for 30 minutes at room temperature. They samples were washed, mounted onto a 2 % agarose pad, and mounted to a cover slip with Fluoromount G. *C. elegans* were imaged on an Olympus IX81 microscope using a Hamamatsu digital camera. Images were processed identically using Image J.

A1.9 Contributions

Elizabeth M. Huffman synthesized and characterized the small molecules **2.3**, **2.4**, and **2.5**.

A1.10 Acknowledgements

I would like to thank several members of the Kimble lab for their assistance in getting the *C. elegans* up and running in our group. I would specifically like to thank Peggy Kroll-Conner and Clinton Morgan for help learning *C. elegans* culture and genetics. I would also like to acknowledge Danny “Rack’em” Zwick and Paul Wrighton with assistance with microscopy.

A1.11 References

1. Southey, J. F. (1986) *Laboratory methods for work with plant and soil nematodes*, 6th ed., HMSO, London
2. Kaplan, R. (2004) Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol.* **20**, 477-481
3. Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., and Savioli, L. (2007) Control of neglected tropical diseases. *N. Engl. J. Med.* **357**, 1018-1027
4. Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S. S., Wenger, A., Wieland-Berghausen, S., Goebel, T., Gauvry, N., Pautrat, F., Skripsky, T., Froelich, O., Komoin-Oka, C., Westlund, B., Sluder, A., and Mäser, P. (2008) A new class of anthelmintics effective against drug-resistant nematodes. *Nature* **452**, 176-180
5. Osei-Atweneboano, M. Y., Eng, J. K. L., Boakye, D. A., Gyapong, J. O., and Prichard, R. K. (2007) Prevalence and intensity of *Onchocerca volvulus* infection and efficacy of ivermectin in endemic communities in Ghana: a two-phase epidemiological study. *Lancet* **369**, 2021-2029
6. Epe, C., and Kaminsky, R. (2013) New advancement in anthelmintic drugs in veterinary medicine. *Trends Parasitol.* **29**, 129-134
7. Lecova, L., Stuchlikova, L., Prchal, L., and Skalova, L. (2014) Monepantel: the most studied new anthelmintic drug of recent years. *Parasitology* **141**, 1686-1698
8. Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325-330
9. Rual, J. F., Ceron, J., Koreth, J., Hao, T., Nicot, A. S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S. H., Hill, D. E., van den Heuvel, S., and Vidal, M. (2004) Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14**, 2162-2168
10. Hulme, S. E., and Whitesides, G. M. (2011) Chemistry and the worm: *Caenorhabditis elegans* as a platform for integrating chemical and biological research. *Angew. Chem. Int. Ed. Engl.* **50**, 4774-4807
11. Burglin, T., Lobos, E., and Blaxter, M. (1998) *Caenorhabditis elegans* as a model for parasitic nematodes. *Int. J. Parasitol.* **28**, 395-411
12. Jones, A. K., Buckingham, S. D., and Sattelle, D. B. (2005) Chemistry-to-gene screens in *Caenorhabditis elegans*. *Nat. Rev. Drug Discovery* **4**, 321-330
13. Griffiths, J. S., Haslam, S. M., Yang, T. L., Garczynski, S. F., Mulloy, B., Morris, H., Cremer, P. S., Dell, A., Adang, M. J., and Aroian, R. V. (2005) Glycolipids as receptors for *Bacillus thuringiensis* crystal toxin. *Science* **307**, 922-925
14. Gerdt, S., Lochnit, G., Dennis, R. D., and Geyer, R. (1997) Isolation and structural analysis of three neutral glycosphingolipids from a mixed population of *Caenorhabditis elegans* (Nematoda: Rhabditida). *Glycobiology* **7**, 265-275
15. Cipollo, J. F., Awad, A. M., Costello, C. E., and Hirschberg, C. B. (2005) N-Glycans of *Caenorhabditis elegans* are specific to developmental stages. *J. Biol. Chem.* **280**, 26063-26072

16. Novelli, J. F., Chaudhary, K., Canovas, J., Benner, J. S., Madinger, C. L., Kelly, P., Hodgkin, J., and Carlow, C. K. S. (2009) Characterization of the *Caenorhabditis elegans* UDP-galactopyranose mutase homolog *glf-1* reveals an essential role for galactofuranose metabolism in nematode surface coat synthesis. *Dev. Biol.* **335**, 340-355
17. Page, A. P., and Johnstone, I. L. (2007) The cuticle. *WormBook*, 1-15
18. Bird, A. F., and Bird, J. (1991) *The structure of nematodes*, 2nd ed., Academic Press, San Diego
19. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811
20. Wesener, D. A., Wangkanont, K., McBride, R., Song, X., Kraft, M. B., Hodges, H. L., Zarling, L. C., Splain, R. A., Smith, D. F., Cummings, R. D., Paulson, J. C., Forest, K. T., and Kiessling, L. L. (2015) Recognition of microbial glycans by human intelectin-1. *Nat. Struct. Mol. Biol.* **22**, 603-610
21. Voehringer, D., Stanley, S. A., Cox, J. S., Completo, G. C., Lowary, T. L., and Locksley, R. M. (2007) *Nippostrongylus brasiliensis*: identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. *Exp. Parasitol.* **116**, 458-466
22. Pemberton, A. D., Knight, P. A., Gamble, J., Colledge, W. H., Lee, J. K., Pierce, M., and Miller, H. R. (2004) Innate BALB/c enteric epithelial responses to *Trichinella spiralis*: inducible expression of a novel goblet cell lectin, intelectin-2, and its natural deletion in C57BL/10 mice. *J. Immunol.* **173**, 1894-1901
23. van Die, I., and Cummings, R. D. (2009) Glycomics in Unraveling Glycan-Driven Immune Responses by Parasitic Helminths. *Handbook of Glycomics*, 367-396
24. Carlson, E. E., May, J. F., and Kiessling, L. L. (2006) Chemical probes of UDP-galactopyranose mutase. *Chem. Biol.* **13**, 825-837
25. Parang, K., Wiebe, L. I., and Knaus, E. E. (2000) Novel approaches for designing 5'-O-ester prodrugs of 3'-azido-2',3'-dideoxythymidine (AZT). *Curr. Med. Chem.* **7**, 995-1039
26. Lavis, L. D. (2008) Ester bonds in prodrugs. *ACS Chem. Biol.* **3**, 203-206
27. Burns, A. R., Wallace, I. M., Wildenhain, J., Tyers, M., Giaever, G., Bader, G. D., Nislow, C., Cutler, S. R., and Roy, P. J. (2010) A predictive model for drug bioaccumulation and bioactivity in *Caenorhabditis elegans*. *Nat. Chem. Biol.* **6**, 549-557
28. Stupp, G. S., von Reuss, S. H., Izrayelit, Y., Ajredini, R., Schroeder, F. C., and Edison, A. S. (2013) Chemical detoxification of small molecules by *Caenorhabditis elegans*. *ACS Chem. Biol.* **8**, 309-313
29. Wu, Y., Li, Q., and Chen, X. Z. (2007) Detecting protein-protein interactions by Far western blotting. *Nat. Protoc.* **2**, 3278-3284
30. Stiernagle, T. (2006) Maintenance of *C. elegans*. *WormBook*, 1-11
31. Timmons, L., and Fire, A. (1998) Specific interference by ingested dsRNA. *Nature* **395**, 854

