nature structural & molecular biology

Recognition of microbial glycans by human intelectin-1

Darryl A Wesener[1], Kittikhun Wangkanont[2], Ryan McBride[3,4], Xuezheng Song[5,6], Matthew B Kraft[2,8], Heather L Hodges[2], Lucas C Zarling[1], Rebecca A Splain[2,8], David F Smith[5,6], Richard D Cummings[5,6], James C Paulson[3,4], Katrina T Forest[7] & Laura L Kiessling[1,2]

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 does interact with multiple glycan epitopes found exclusively on microbes: β-linked D-galactofuranose (β-Gal*f*), D-phosphoglycerol–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 complexed with 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, has an exocyclic 1,2-diol but does not bind hIntL-1, probably owing to unfavorable steric and electronic effects. hIntL-1 marks only *Streptococcus pneumoniae* serotypes that display surface glycans with terminal 1,2-diol groups. This ligand selectivity suggests that hIntL-1 functions in microbial surveillance.

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 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 its being self or nonself[5]. Cell-surface glycans can be distinguished by carbohydrate-binding proteins, or lectins[6], which are typically categorized on the basis of their monosaccharide selectivity[7]. These lectins can be exploited for host defense, as in the case of innate immune lectins, such as mannose-binding lectin[8]. In serum, mannose-binding lectin is precomplexed with mannose-binding lectin–associated serine proteases, and interaction of this complex with a cell surface results in activation of the lectin pathway of complement and ultimately leads 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 has been unclear is the intelectins (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 homo-oligomers 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 have been proposed to be most similar to ficolins, a class of FBD-containing innate immune lectins[11].

Several observations have implicated 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]. Moreover, hIntL-1 has been reported to be the intestinal lactoferrin receptor[24] and to function as a tumor marker[25]. It also has been suggested to be involved in metabolic disorders including diabetes, in which it is known as omentin[26]. Given these diverse potential functions, we set out to examine the ligand specificity of hIntL-1.

hIntL-1 has been reported to bind furanose residues (five-membered-ring saccharide isomers), including ribofuranose (Rib*f*) 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 but not human glycans is an example of phylogenetic glycan differences[31]. This is just one example, and collectively the surface glycans of microbes are generated from more than 700 unique building blocks, whereas fewer than 35 carbohydrate residues are needed to assemble mammalian glycans[32,33]. In principle, targeting of monosaccharide residues unique to microbes could be used by the innate immune system to differentiate mammalian cells from microbes.

[1]Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin, USA. [2]Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin, USA. [3]Department of Cell and Molecular Biology, Scripps Research Institute, La Jolla, California, USA. [4]Department of Chemical Physiology, Scripps Research Institute, La Jolla, California, USA. [5]Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia, USA. [6]Glycomics Center, Emory University School of Medicine, Atlanta, Georgia, USA. [7]Department of Bacteriology, University of Wisconsin–Madison, Madison, Wisconsin, USA. [8]Present Addresses: Gilead Sciences, Foster City, California, USA (M.B.K.) and Global API Chemistry, GlaxoSmithKline, King of Prussia, Pennsylvania, USA (R.A.S.). Correspondence should be addressed to L.L.K. (kiessling@chem.wisc.edu).

Received 17 March; accepted 2 June; published online 6 July 2015; doi:10.1038/nsmb.3053

We reasoned that clues to hIntL-1 function would emerge from determining the glycans to which hIntL-1 binds and the molecular basis for its recognition selectivity. Here, we use glycan microarrays to demonstrate that hIntL-1 preferentially binds microbial over human glycans. Given the diversity of microbial glycans, a lectin that binds a single microbial saccharide epitope (for example, galactofuranose) would be expected to have specialized functions. It is therefore striking that hIntL-1 does not engage a single monosaccharide or even related saccharides but instead interacts with multiple structurally divergent microbial monosaccharide residues. We have used X-ray crystallography to reveal the molecular mechanism by which hIntL-1 recognizes its targets: 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 because hIntL-1 targets *S. pneumoniae* serotypes that display its glycan ligands.

RESULTS

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 can be purified with an immobilized–β-Gal*f* column (**Supplementary Fig. 1a**,**b**). Because lectin-carbohydrate interactions often depend on multivalent binding[34,35] we postulated that hIntL-1 trimers might bind avidly to multivalent carbohydrate displays. Hence, we evaluated hIntL-1 carbohydrate binding specificity by using immobilized biotinylated carbohydrates (β-D-Gal*f*, β-D-galactopyranose (β-Gal*p*) and β-D-ribofuranose (β-Rib*f*)) in an enzyme-linked immunosorbent-like assay (ELISA) (**Fig. 1a** and **Supplementary Fig. 1c**,**d**). We chose the monosaccharide epitopes that we tested on the basis of a previous study in which a small carbohydrate panel was evaluated for inhibition of hIntL-1 binding to a polysaccharide immobilized on a polymeric resin[17]. In those studies, ribose was the most effective competitor (half-maximal inhibitory concentration (IC_{50}) <5 mM), and it was followed by Gal*f*-β(1,4)-GlcNAc (IC_{50} of 9 mM), with galactose being less potent (IC_{50} of 66 mM)[17]. Our data indicate that hIntL-1 does not bind ribofuranose or galactopyranose, but it does engage the β-Gal*f*–substituted surface avidly with a functional affinity (apparent affinity) of 85 ± 14 nM (**Fig. 1b**).

Our results contrast with those of the previous study[17] because we did not detect binding to the pyranose form of galactose or to ribofuranose. The apparent discrepancies could arise because inhibition was obtained in the previous investigation with high concentrations of 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 that we observed for hIntL-1 binding to immobilized β-D-Gal*f* suggests that the protein binds tightly to a ligand, but the previous IC_{50} for the β-D-Gal*f*–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 because it mimics key aspects of the multivalent display of carbohydrate ligands on a cell surface[34]. Nonetheless, the differences between the reported hIntL-1 binding specificities and those we observed prompted us to examine hIntL-1 binding with another assay. We used surface plasmon resonance (SPR) and monitored hIntL-1 interaction with surfaces to which the aforementioned saccharides or β-D-arabinofuranose (β-Ara*f*) 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-Gal*f* (**Fig. 1c** and **Supplementary Fig. 1e**).

hIntL-1 binding to microbial glycans

Glycan microarray technology can provide a more comprehensive assessment of hIntL-1 ligand recognition[36]. Therefore, we prepared a focused array that included furanosides (**Supplementary Table 1**), with the methods used to generate the Consortium for Functional Glycomics (CFG; http://www.functionalglycomics.org/) mammalian glycan v5.1 array, and we tested both arrays for hIntL-1 binding. In the focused array, we included lacto-*N*-neotetraose (LNnT) and asialo, galactosylated biantennary N-linked glycan (NA2) to ascertain the efficiency of carbohydrate immobilization. Data from the focused array were consistent with those obtained from the ELISA and SPR assays, thus indicating that, of the carbohydrates displayed, hIntL-1 bound only to carbohydrates with β-Gal*f* residues (**Fig. 2a** and **Supplementary Table 1**). We attribute the small amount of binding to β-Gal*p* to its hydrophobic, alkyl anomeric linker. In contrast to the furanoside array, the CFG v5.1 array yielded no validated interactions

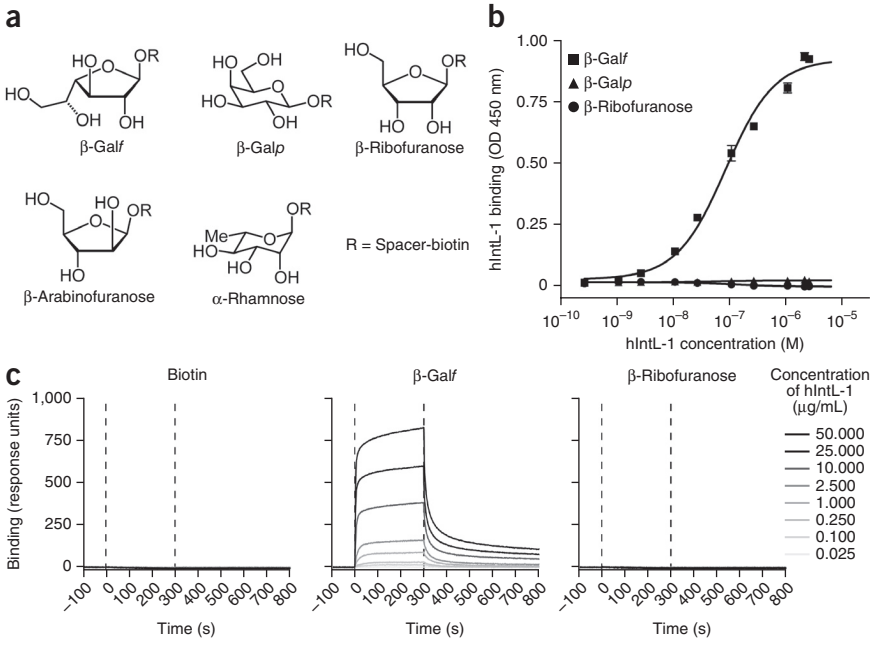

Figure 1 hIntL-1 selectivity for monosaccharides. (**a**) Structures of saccharides used for characterization of hIntL-1 by ELISA and SPR. (**b**) Specificity of hIntL-1 binding to immobilized β-Gal*f*, β-ribofuranose (β-Rib*f*) and β-galactopyranose (β-Gal*p*), evaluated by ELISA (schematic in **Supplementary Fig. 1b**). Data are shown as mean ± s.d. (n = 3 technical replicates)). Data were fit to a single-site binding equation (solid lines) and therefore represent the apparent affinity of trimeric hIntL-1. Values for hIntL-1 bound to immobilized β-Gal*f* ($K_{d(apparent, trimer)}$ ± s.d.) are 85 ± 14 nM or 8.0 ± 1.3 μg/ml. OD, optical density. (**c**) Representative real-time SPR sensorgrams (from three independent experiments) of hIntL-1 binding to immobilized carbohydrates. Biotin served as a control. (The complete SPR data set is in **Supplementary Fig. 1e**.)

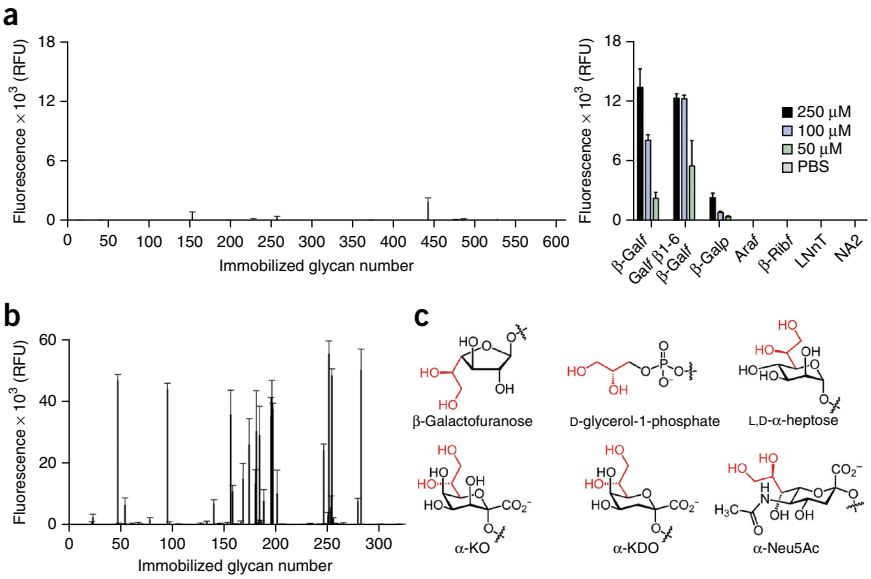

Figure 2 Glycan selectivity of hIntL-1, assessed by glycan microarrays. (**a**) Recombinant hIntL-1 (50 μg/ml) binding to mammalian glycan microarray CFG v5.1 (left) and to a furanoside array (right). The concentrations given for the furanoside array represent those used in the carbohydrate-immobilization reaction. Data are shown as mean ± s.d. ($n = 4$ technical replicates). (The full data set is in **Supplementary Tables 1** and **2**.) RFU, relative fluorescence units. (**b**) Binding of recombinant *Strep*-tagged hIntL-1 (50 μg/ml) to microbial glycan array. Data are shown as mean ± s.d. ($n = 4$ technical replicates). (Glycan array data organized by genus are in **Supplementary Fig. 2a**, and the full data set is in **Supplementary Table 3**.) (**c**) Structural representation of the putative key binding epitopes for hIntL-1 and the nonbinding *N*-acetylneuraminic acid (α-Neu5Ac). A terminal vicinal diol (red) is a common feature of α-Neu5Ac and all of the ligands identified.

with mammalian glycans (**Fig. 2a**). Increasing the protein concentration yielded similarly low signals, a result suggesting that the modest residual binding that we detected arose from nonspecific interactions (**Supplementary Table 2**). Thus, none of the human glycans examined are ligands of hIntL-1.

The initial binding data revealing that hIntL-1 robustly complexes β-Gal*f* 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], Gal*f* residues occur in glycans from a number of human pathogens, including the bacteria *Mycobacterium tuberculosis* and *Klebsiella pneumoniae*, and the fungus *Aspergillus fumigatis*[29,37]. The possibility that hIntL-1 interacts with microbial glycans was tested with a microarray displaying more than 300 oligosaccharides from bacterial species[38]. Screening of this array revealed multiple glycan ligands for hIntL-1 (**Fig. 2b**, **Supplementary Fig. 2a** and **Supplementary Table 3**). These ligands comprised glycans from Gram-negative and Gram-positive bacteria, including *S. pneumoniae*, *Proteus mirabilis*, *Proteus vulgaris*, *Yersinia pestis* and *K. pneumoniae* (**Table 1**). Four of the top 15 ligands contained terminal β-Gal*f* epitopes, including the outer polysaccharide from *K. pneumoniae* and a capsular polysaccharide from *S. pneumoniae*. Surprisingly, the majority of the glycans identified did not possess Gal*f* residues. The top five hits had saccharide residues with D-glycerol-1-phosphate substituents. This epitope was the common feature because the residue to which it was appended varied between glycans. Other common epitopes included either D/L-*manno*-heptose, KO or KDO residues (**Fig. 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 (**Supplementary Fig. 2b**). What was especially notable, however, was that none of the hIntL-1 ligands that we identified on the microbial glycan array are found in mammalian glycans, but collectively these five residues are widely distributed in bacteria[32].

Table 1 Top 15 microbial glycan ligands, sorted by average fluorescence intensity

Rank	Microbial sample	Proposed ligand
1	*S. pneumoniae* type 43	Glycerol phosphate
2	*P. mirabilis* O54ab	Glycerol phosphate
3	*S. pneumoniae* type 56	Glycerol phosphate
4	*P. mirabilis* O54a, 54b	Glycerol phosphate
5	*P. vulgaris* O54a, 54c	Glycerol phosphate
6	*K. pneumoniae* O2a OPS	β-Gal*f*
7	*K. pneumoniae* O2ac OPS	β-Gal*f*
8	*Y. pestis* KM260(11)-Δ0187[a]	
9	*K. pneumoniae* O1 OPS	β-Gal*f*
10	*Y. pestis* 11M-37	Heptose, KO, KDO
11	*Y. pestis* KM260(11)-6C[a]	
12	*Y. pestis* KM260(11)-Δwaal	Heptose, KO, KDO
13	*S. pneumoniae* type 20	Heptose, KO, KDO
14	*Y. pestis* KM260(11)-ΔpmrF	Heptose, KO, KDO
15	*Y. pestis* 11M-25	Heptose, KO, KDO

[a]These glycans are currently structurally uncharacterized.

Structure of hIntL-1

To understand the molecular mechanisms underlying glycan recognition by hIntL-1, we determined its structure by X-ray crystallography. Apo–hIntL-1 crystals diffracted to 1.8-Å resolution, and we solved the structure of the protein with molecular replacement by using the structure of a selenomethione-labeled *Xenopus laevis* IntL as a search model (**Table 2**) (PDB 4WMO). hIntL-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 (**Fig. 3a**). The second of these β-sheet structures closes on itself to form a very short stretch of unusually flattened β-ribbons (amino acids 221–226 and 248–278). A Dali search[39] with the hIntL-1 structure yielded several weak fibrinogen and ficolin structure hits (r.m.s. deviation values of ~4 Å). The secondary structures of L-ficolin[40] 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 (**Supplementary Fig. 3**). Indeed, removal of the first 150 residues from the hIntL-1 Dali input yielded no hits. These data indicate that 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 nonidentical (Cα r.m.s. deviation of 0.65 Å), disulfide-linked trimers, each

Table 2 Data collection and refinement statistics

	Apo–hIntL-1	Gal*f*-bound hIntL-1
Data collection		
Space group	$P2_13$	$P2_13$
Cell dimensions		
a, *b*, *c* (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)
R_{sym}	0.119 (0.495)	0.078 (0.773)
$I / \sigma I$	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	48,784	68,256
R_{work} / R_{free}	0.133 / 0.164	0.155 / 0.180
No. atoms		
Protein	4,551	4,606
Ca^{2+}	6	6
Allyl-β-d-Gal*f*	–	30
Water	658	616
B 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
r.m.s. deviation		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.107	1.119

Each data set was collected from one crystal.
[a]Values in parentheses are for highest-resolution shell.

arranged around a crystallographic three-fold axis. In one trimer, the peptide chain that connects each monomer to the adjacent monomer is resolved, so that the intermolecular disulfide bond between residues C31 and C48 is apparent (**Fig. 3a**). These data are consistent with SDS-PAGE analysis indicating that hIntL-1 exists as a trimer. Each hIntL-1 monomer has three calcium ions, and each cation is chelated 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.

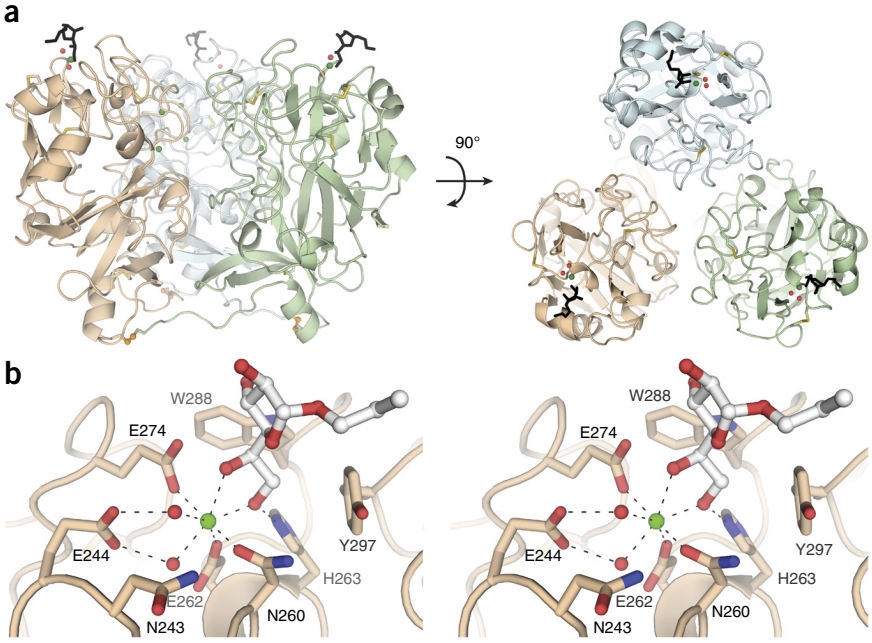

Figure 3 Structure of hIntL-1 bound to allyl-β-D-Gal*f*. (**a**) Complex of hIntL-1 disulfide-linked trimer and allyl-β-D-Gal*f*. Each monomer unit is depicted in green, wheat or gray; the β-allyl Gal*f* in black; calcium ions in green; the intermonomer disulfides in orange; and ordered water molecules 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 ($F_o - F_c$, 3σ) of the allyl-β-D-Gal*f* ligand is in **Supplementary Fig. 4b**.)

To determine how hIntL-1 binds its ligands, we solved a structure of the complex of allyl-β-D-Gal*f* bound to hIntL-1, to 1.6-Å resolution. The Cα r.m.s. deviation between the asymmetric unit of apo and Gal*f*–bound structures (0.118 Å) suggested that no substantial 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. Protein side chains are poised for hydrogen-bonding (i.e., H263 to the Gal*f* O(6) hydroxyl group; **Fig. 3b** and **Supplementary Fig. 4a**), enhance calcium coordination. As they chelate the calcium, the carbohydrate vicinal exocyclic hydroxyl groups adopt a *gauche* conformation, with dihedral angles of 45° and 51° for chains 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 (**Fig. 2b** and **Supplementary Table 3**). 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 contribute to affinity.

The high resolution of the structure of the hIntL-1 complex allows unambiguous assignment of the β-Gal*f* ring conformation[41,42] in each monomer (**Supplementary Fig. 4b**). Using the Altona-Sundaralingam pseudorotational model[43], we calculated the pseudorotational phase angle, *P*, of each furanoside to assign its conformation. In hIntL-1 chain A, the furanoside is in the twist conformation with C_1 above the plane of the ring and the ring O below, and the torsion angles around the C_4-C_5 bond are both in *gauche* conformation, while those around the C_5-C_6 bond are *gauche-trans* (^1T_O-gg-gt) (calculated *P* of 105°) conformation, whereas the β-Gal*f* shown in **Figure 3b** adopts the envelope ^{4}E-gg-gt (calculated *P* of 57°) conformation (**Supplementary Fig. 4c**,**d**). The presence of conformational differences within the structures is consistent with the flexibility of furanosides[42].

Structural basis for hIntL-1 selectivity

The structure of the lectin–Gal*f* complex reveals why the acyclic 1,2-diol moiety is critical: the vicinal hydroxyl groups engage in calcium-ion coordination. However, other glycan properties contribute to hIntL-1 recognition. For example, hIntL-1 does not bind α-Gal*f*–substituted glycans (**Supplementary Fig. 2b**). A cursory assessment of the β-Gal*f*

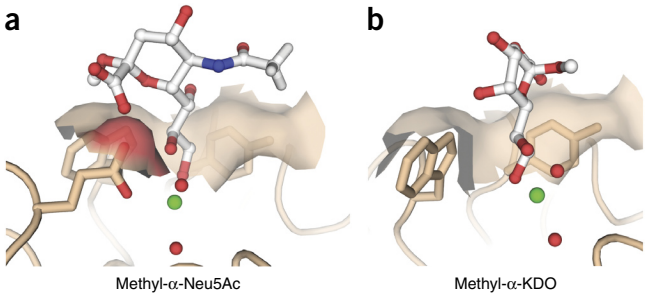

Figure 4 Models for hIntL-1 interacting with relevant saccharide epitopes from humans (α-Neu5Ac) or microbes (α-KDO). (**a**) Docking of methyl-α-Neu5Ac into the hIntL-1 structure. The conformation shown is similar to that observed in other protein structures with a methyl-α-Neu5Ac ligand (PDB 2BAT, 2P3I, 2P3J, 2P3K, 2I2S, 1KQR, 1HGE and 1HGH (refs. 56–60)). All models in this figure were generated from the allyl-β-D-Gal*f*–bound structure by docking the relevant diol of each compound into the Gal*f* diol electron density in Coot without further refinement. Calcium ions are shown in green and ordered water molecules in red. (**b**) Docking of methyl-α-KDO into the hIntL-1 structure. Comparison with methyl-α-Neu5Ac docked into the hIntL-1 structure reveals differences in the steric requirements for binding for each molecule.

complex suggests that hIntL-1 might accommodate α-Gal*f* 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-Gal*f* have revealed that the anomers have dramatically different conformational preferences[42]. The β-Gal*f* ^{4}E-gg-gt conformer that we find in hIntL-1 chain B is predicted to be the second lowest in energy (0.4 kcal/mol)[42]. That conformation for methyl-α-Gal*f* is destabilized by 3.2 kcal/mol. As a result, the expected Boltzmann population for methyl-α-Gal*f* in a ^{4}E-gg-gt conformation is less than 0.2%, and it is thus ranked 25th out of the 90 conformations examined[42]. These data suggest that α-Gal*f* 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 (**Fig. 2a**). A saccharide epitope widespread in human glycans, α-Neu5Ac has a terminal 1,2-diol and shares similarity with KDO, which are common in microbial glycans and do function as hIntL-1 ligands[44]. We used a biotinylated glycoside to confirm that hIntL-1 fails to interact with surfaces displaying α-Neu5Ac (**Supplementary Fig. 5a**). Moreover, compounds identified as hIntL-1 ligands—glycerol and glycerol-1-phosphate—competitively inhibit the lectin from binding to β-Gal*f*, but methyl-α-mannopyranoside and methyl-α-Neu5Ac do not (**Supplementary Fig. 5b**). 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.

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. We found that the KDO glycoside is readily accommodated, but the α-Neu5Ac glycoside is not (**Fig. 4a**,**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 (**Fig. 4a**). The destabilizing interactions with α-Neu5Ac cannot be mitigated by rotating bonds or by adopting accessible low-energy conformations. Future experiments with protein variants and ligand analogs will be useful in testing this proposed evasion mechanism.

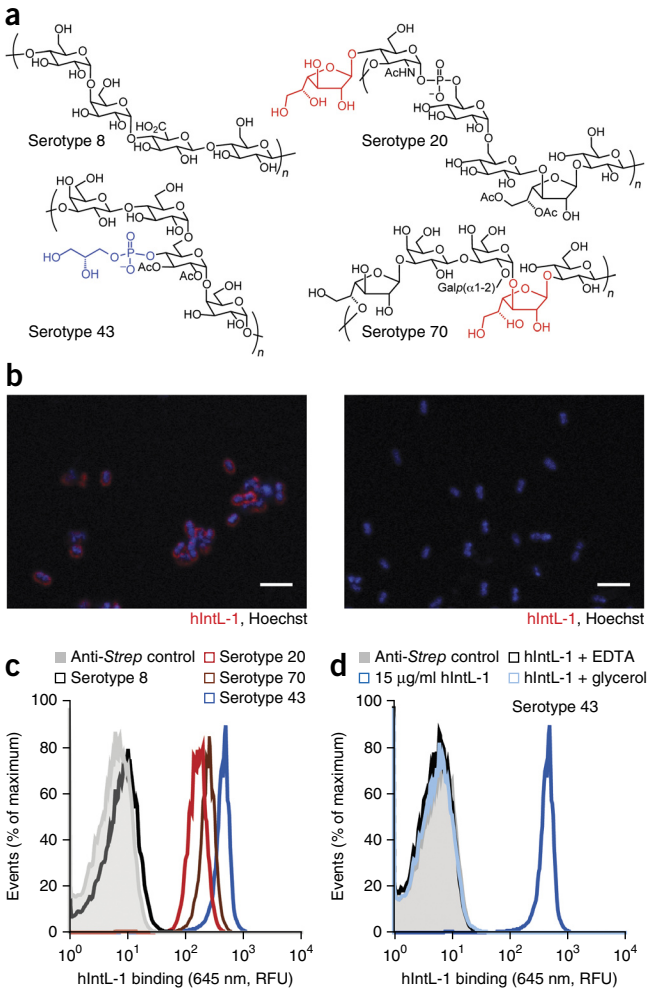

Figure 5 hIntL-1 binds to *S. pneumoniae* serotypes producing capsular polysaccharides with terminal vicinal diols. (**a**) Chemical structure of the capsular polysaccharides displayed on the *S. pneumoniae* serotypes (8, 20, 43 and 70) tested. The Gal*f* residues assumed to mediate hIntL-1 cell binding are shown in red, and the phosphoglycerol moiety is shown in blue. Ac, acetyl. (**b**) Fluorescence microscopy of hIntL-1 binding to *S. pneumoniae* serotype 20. Bacteria were treated with *Strep*-tagged hIntL-1 (15 μg/ml). Red, anti–*Strep*-tag antibody conjugate; blue, cellular DNA visualized with Hoechst. hIntL-1 at the surface of serotype 20 bacteria in the presence of Ca^{2+} (left) or EDTA (right). Images are representative of more than five fields of view per sample. Scale bars, 2 μm. (Results for serotypes 43, 70 and 8 are shown in **Supplementary Fig. 6a**.) (**c**,**d**) Flow cytometry analysis of *Strep*–hIntL-1 binding to *S. pneumoniae* serotypes with an anti–*Strep*-tag antibody conjugate. In the anti-*Strep* control sample, recombinant hIntL-1 was omitted. Cells were labeled with propidium iodide. (**c**) Flow cytometry analysis of serotypes 8, 20, 43 and 70. Data were collected consecutively with identical instrument settings. (**d**) Dependence of the hIntL-1–carbohydrate interaction on Ca^{2+}, tested by addition of 10 mM EDTA. Ligand selectivity was tested by addition of 100 mM glycerol. Data are representative of two independent experiments. (Analyses of serotypes 20, 70 and 8 are shown in **Supplementary Fig. 6b**.)

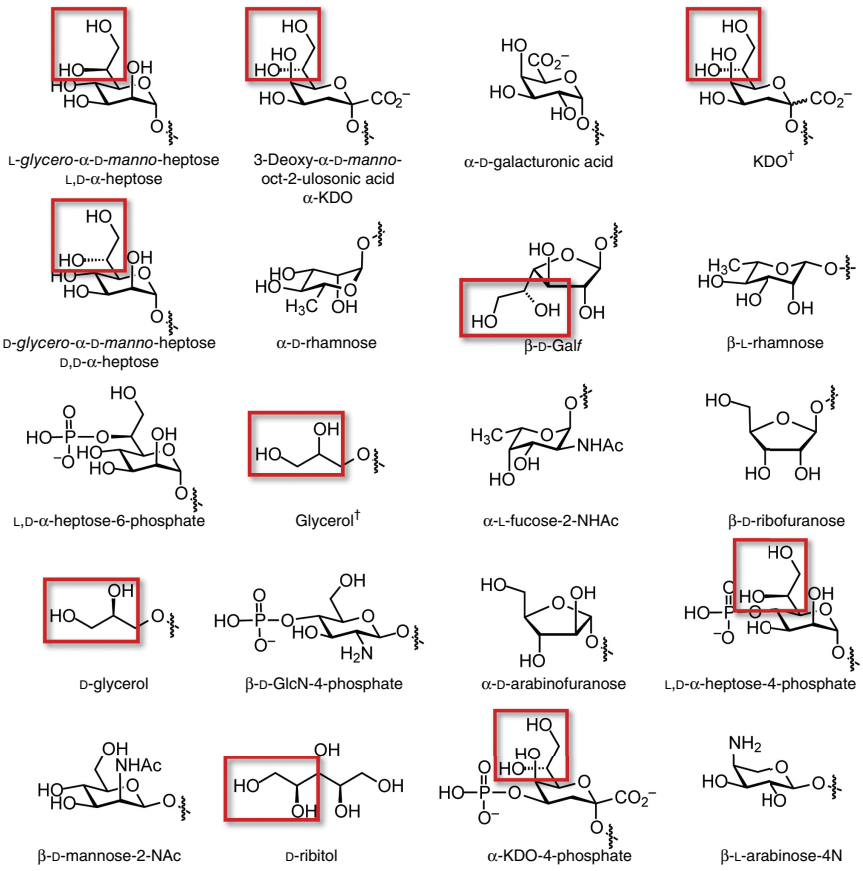

Figure 6 Structures of the 20 most prevalent monosaccharides that are unique to bacterial glycans. The most common, L,D-α-heptose, is shown in the top left corner, and number 20, β-L-arabinose-4N, is shown at the bottom right. This figure is derived from data in ref. 32. Terminal acyclic 1,2-diol epitopes that could serve as ligands of hIntL-1 are highlighted with a red box. Cross symbol designates monosaccharides for which no stereochemical information was provided.

hIntL-1 comparison with ficolins

The FBD of hIntL-1 suggested that hIntL-1 would be structurally related to the ficolins. With the structure of an X-type lectin complex, it is now apparent that, outside the FBD, intelectins and ficolins deviate extensively. IntLs lack the collagen-like domain that mediates complement activation. Additionally, the hIntL-1 carbohydrate-recognition domain is larger than that of the ficolins, and hIntL-1 coordinates three calcium ions, two of which are buried, whereas 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 (**Supplementary Fig. 3c**). 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.

hIntL-1 binding to *S. pneumoniae*

Because hIntL-1 is expressed in mucosal tissues, we examined its binding to immunologically distinct serotypes of the encapsulated human lung pathogen *S. pneumoniae*, the causative agent of several diseases, including pneumonia, meningitis and septicemia[45]. The surface-exposed pneumococcal capsular polysaccharide is among the first microbial antigens encountered by the immune system upon challenge[46]. 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 previously leveraged to develop a polysaccharide-based vaccine that is protective against streptococcus infections[47]. 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[46] (chemical structures in **Fig. 5a**). The data indicate that hIntL-1 binds to the surfaces of serotypes 20, 70 and 43, each of which displays cell-surface glycans with an exocyclic, terminal 1,2-diol (**Fig. 5b–d** and **Supplementary Fig. 6**). As predicted by the structure of the β-Gal*f* –hIntL-1 complex, binding to these strains depends on calcium ion–mediated coordination, and glycerol functions as a competitive ligand (**Fig. 5b,d**). 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-Gal*f* (i.e., hit 13 from the microbial array, **Table 1**), but it interacted most avidly with the serotype displaying the D-glycerol-1-phosphate–modified saccharide that was the top hit from the microbial glycan array (**Fig. 5c**). 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 bacterial cell surfaces.

hIntL-1 has been reported to bind lactoferrin[24], a protein that appears to have antimicrobial activity[48]. 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 our assay, in contrast to the previous reports, this interaction did not require calcium ions. The apparent affinity that we measured for the hIntL-1 trimer is rather weak for a specific protein-protein interaction (K_d of ~500 nM). The isoelectric points (pI) of the proteins (~5.5 for hIntL-1 and ~8.5 for lactoferrin) suggest that 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 μg/ml) in a buffer that would be compatible with hIntL-1 binding to 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 the bactericidal activity of lactoferrin is abolished under similar conditions[49,50]. These initial data are inconsistent with a central role for lactoferrin–intelectin complexes in mediating microbial cell killing, and they suggest that other functional roles for hIntL-1 should be explored.

Mouse IntL-1 binding to Gal*f*

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 mouse IntL-1, which is the

mouse homolog[27] of hIntL-1. When we tested mouse IntL-1 with the same SPR assay used with the human homolog, its glycan-recognition properties were analogous: it failed to interact with β-ribofuranose, β-arabinofuranose, α-rhamnopyranose or β-Gal*p*, but it did interact with β-Gal*f* (**Supplementary Fig. 7**). These data support the prospect that IntLs from different species have evolved to bind widely distributed 1,2-diol–containing epitopes unique to microbes.

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*, we have resolved this apparent contradiction. The five saccharide epitopes identified as recognition motifs (Gal*f*, phosphoglycerol, *glycero*-D-*manno*-heptose, KDO and KO) share a common feature: a terminal acyclic 1,2-diol. 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 used 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 are typically those on the pyranose ring of a mannose, fucose or galactose residue. The hIntL-1–binding pocket requires that any 1,2-diol motifs possess a primary hydroxyl group because the aromatic substituents W288 and Y297 act as walls to preclude the binding of more substituted diols. These aromatic substituents could contribute not only to specificity but also to affinity. The positioning of Y297 could allow it to participate in a CH-π interaction[51], 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 because 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 that our structure will also provide insight into the physiological roles of the intelectins. The upregulation of intelectins upon infection suggests that they may function in innate immunity. Although existing data from genome-wide association studies have not directly linked intelectin mutations and increased susceptibility to infection, there are studies that have linked hIntL-1 to asthma[52] and Crohn's disease[53], both of which arise from defects at mucosal surfaces where intelectins are secreted. Moreover, the amino acid variant V109D is associated with an increased risk of asthma[52]. 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 (**Fig. 3a**) suggested that 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*f* (serotypes 20 or 70) or phosphoglycerol (serotype 43) but not those lacking the requisite terminal 1,2-diol (serotype 8) highlights the advantages of using a simple binding epitope: hIntL-1 is not restricted to binding solely one glycan building block; instead, it can interact with bacterial cells that present glycans composed of very different components (Gal*f* versus phosphoglycerol).

Because it engages a small epitope found within microbial glycans, hIntL-1 should be capable of recognizing a wide variety of microbes. We analyzed the 20 most common glycan building blocks unique to microbes[32] and found that half of these possess an acyclic 1,2-diol that could, in principle, be recognized by intelectins (structures in **Fig. 6**). The potential that a given microbe generates glycan ligands for hIntL-1 can be inferred from genetic sequence data. For example, organisms bearing Gal*f* residues have a *glf* gene[29]. D-glycerol-1-phosphate–modified glycans are generated from the activated donor CDP-D-glycerol and therefore will encode functional homologs of the *S. pneumoniae gct* gene[46]. Pathways that lead to the incorporation of heptose, KO and KDO are known, because these residues are found in lipopolysaccharide[54] and in the capsular (K) antigen of Gram-negative bacteria[55]. The orientation of the saccharide-binding sites on a single face of the hIntL-1 trimer not only enhances the avidity of cell-surface binding but 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.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4WMQ (apo–hIntL-1) and 4WMY (Gal*f*-bound hIntL-1).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This research was supported by the US National Institutes of Health (NIH) (R01GM55984 and R01AI063596 (L.L.K.)). D.A.W. thanks 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. M.B.K. and H.L.H. were supported by the NIH (F32 GM100729 to M.B.K. and T32 GM008505 to H.L.H.). L.C.Z. was supported by a UW–Madison Hilldale Fellowship. R.A.S. thanks the American Chemical Society Division of Medicinal Chemistry for a fellowship. The glycan array experiments were made possible by the Consortium for Functional Glycomics (NIH NIGMS GM062116 and GM98791 (J.C.P.)), which supported the Glycan Array Synthesis Core at The Scripps Research Institute and the Protein-Glycan Interaction Resource (Emory University School of Medicine). These resources assisted with analysis of samples on the array. Printing and processing the furanoside array was supported through the US National Center for Functional Glycomics supported by NIH NIGMS (P41GM103694 (R.D.C.)). SPR experiments were performed at the University of Wisconsin (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 and 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). We thank J.M. Fishman for assistance in

preparing the synthetic methods and M.R. Levengood, A.H. Courtney and D.R. McCaslin for thoughtful discussions. We thank M.R. Richards (University of Alberta) for helpful discussions.

AUTHOR CONTRIBUTIONS

D.A.W. and L.L.K. conceived the project. D.A.W., K.W. and L.L.K. planned the experiments, analyzed the data and wrote the paper, with input from all the other authors. Cloning, protein expression and biochemical experiments were performed by D.A.W. and L.C.Z. Microscopy was performed by H.L.H. Baculovirus was made by K.W. The carbohydrate ligands were synthesized and characterized by M.B.K. and R.A.S. The furanoside glycan microarray was constructed and analyzed with the mammalian glycan microarray by X.S., D.F.S. and R.D.C. The microbial glycan array was constructed and analyzed by R.M. and J.C.P. Protein crystallization and structure determination were performed by K.W. and K.T.F.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Stowell, S.R. *et al.* Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* **16**, 295–301 (2010).
2. Vaishnava, S. *et al.* The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255–258 (2011).
3. Gallo, R.L. & Hooper, L.V. Epithelial antimicrobial defence of the skin and intestine. *Nat. Rev. Immunol.* **12**, 503–516 (2012).
4. Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A. & Gordon, J.I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
5. Varki, A. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2009).
6. Lis, H. & Sharon, N. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674 (1998).
7. Weis, W.I. & Drickamer, K. Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441–473 (1996).
8. Turner, M.W. The role of mannose-binding lectin in health and disease. *Mol. Immunol.* **40**, 423–429 (2003).
9. Fujita, T. Evolution of the lectin-complement pathway and its role in innate immunity. *Nat. Rev. Immunol.* **2**, 346–353 (2002).
10. Jack, D.L., Klein, N.J. & Turner, M.W. Mannose-binding lectin: targeting the microbial world for complement attack and opsonophagocytosis. *Immunol. Rev.* **180**, 86–99 (2001).
11. Thomsen, T., Schlosser, A., Holmskov, U. & Sorensen, G.L. Ficolins and FIBCD1: soluble and membrane bound pattern recognition molecules with acetyl group selectivity. *Mol. Immunol.* **48**, 369–381 (2011).
12. Holmskov, U., Thiel, S. & Jensenius, J.C. Collectins and ficolins: humoral lectins of the innate immune defense. *Annu. Rev. Immunol.* **21**, 547–578 (2003).
13. Lehotzky, R.E. *et al.* Molecular basis for peptidoglycan recognition by a bactericidal lectin. *Proc. Natl. Acad. Sci. USA* **107**, 7722–7727 (2010).
14. Lee, J.K. *et al.* Cloning and expression of a *Xenopus laevis* oocyte lectin and characterization of its mRNA levels during early development. *Glycobiology* **7**, 367–372 (1997).
15. Lee, J.K., Baum, L.G., Moremen, K. & Pierce, M. The X-lectins: a new family with homology to the *Xenopus laevis* oocyte lectin XL-35. *Glycoconj. J.* **21**, 443–450 (2004).
16. Weis, W.I., Taylor, M.E. & Drickamer, K. The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19–34 (1998).
17. Tsuji, S. *et al.* Human intelectin is a novel soluble lectin that recognizes galactofuranose in carbohydrate chains of bacterial cell wall. *J. Biol. Chem.* **276**, 23456–23463 (2001).
18. French, A.T. *et al.* The expression of intelectin in sheep goblet cells and upregulation by interleukin-4. *Vet. Immunol. Immunopathol.* **120**, 41–46 (2007).
19. Voehringer, D. *et al. Nippostrongylus brasiliensis*: identification of intelectin-1 and -2 as Stat6-dependent genes expressed in lung and intestine during infection. *Exp. Parasitol.* **116**, 458–466 (2007).
20. Pemberton, A.D., Knight, P.A., Wright, S.H. & Miller, H.R. Proteomic analysis of mouse jejunal epithelium and its response to infection with the intestinal nematode, *Trichinella spiralis*. *Proteomics* **4**, 1101–1108 (2004).
21. Datta, R. *et al.* Identification of novel genes in intestinal tissue that are regulated after infection with an intestinal nematode parasite. *Infect. Immun.* **73**, 4025–4033 (2005).
22. Kerr, S.C. *et al.* 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 (2014).
23. Kuperman, D.A. *et al.* Dissecting asthma using focused transgenic modeling and functional genomics. *J. Allergy Clin. Immunol.* **116**, 305–311 (2005).
24. Suzuki, Y.A., Shin, K. & Lonnerdal, B. Molecular cloning and functional expression of a human intestinal lactoferrin receptor. *Biochemistry* **40**, 15771–15779 (2001).
25. Tsuji, S. *et al.* Secretion of intelectin-1 from malignant pleural mesothelioma into pleural effusion. *Br. J. Cancer* **103**, 517–523 (2010).
26. de Souza Batista, C.M. *et al.* Omentin plasma levels and gene expression are decreased in obesity. *Diabetes* **56**, 1655–1661 (2007).
27. Tsuji, S. *et al.* 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 (2007).
28. Pedersen, L.L. & Turco, S.J. Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. *Cell. Mol. Life Sci.* **60**, 259–266 (2003).
29. Nassau, P.M. *et al.* Galactofuranose biosynthesis in *Escherichia coli* K-12: identification and cloning of UDP-galactopyranose mutase. *J. Bacteriol.* **178**, 1047–1052 (1996).
30. Wesener, D.A., May, J.F., Huffman, E.M. & Kiessling, L.L. UDP-galactopyranose mutase in nematodes. *Biochemistry* **52**, 4391–4398 (2013).
31. Bishop, J.R. & Gagneux, P. Evolution of carbohydrate antigens: microbial forces shaping host glycomes? *Glycobiology* **17**, 23R–34R (2007).
32. Herget, S. *et al.* 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 (2008).
33. Adibekian, A. *et al.* Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chem. Sci.* **2**, 337–344 (2011).
34. Mann, D.A., Kanai, M., Maly, D.J. & Kiessling, L.L. Probing low affinity and multivalent interactions with surface plasmon resonance: ligands for concanavalin A. *J. Am. Chem. Soc.* **120**, 10575–10582 (1998).
35. Kiessling, L.L. & Grim, J.C. Glycopolymer probes of signal transduction. *Chem. Soc. Rev.* **42**, 4476–4491 (2013).
36. Blixt, O. *et al.* Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. USA* **101**, 17033–17038 (2004).
37. Tefsen, B., Ram, A.F., van Die, I. & Routier, F.H. Galactofuranose in eukaryotes: aspects of biosynthesis and functional impact. *Glycobiology* **22**, 456–469 (2012).
38. Stowell, S.R. *et al.* Microbial glycan microarrays define key features of host-microbial interactions. *Nat. Chem. Biol.* **10**, 470–476 (2014).
39. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
40. Garlatti, V. *et al.* Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J.* **26**, 623–633 (2007).
41. Taha, H.A., Richards, M.R. & Lowary, T.L. Conformational analysis of furanoside-containing mono- and oligosaccharides. *Chem. Rev.* **113**, 1851–1876 (2013).
42. Richards, M.R., Bai, Y. & Lowary, T.L. Comparison between DFT- and NMR-based conformational analysis of methyl galactofuranosides. *Carbohydr. Res.* **374**, 103–114 (2013).
43. Altona, C. & Sundaralingam, M. Conformational analysis of sugar ring in nucleosides and nucleotides: a new description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94**, 8205–8212 (1972).
44. Angata, T. & Varki, A. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem. Rev.* **102**, 439–469 (2002).
45. Cartwright, K. Pneumococcal disease in western Europe: burden of disease, antibiotic resistance and management. *Eur. J. Pediatr.* **161**, 188–195 (2002).
46. Bentley, S.D. *et al.* Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**, e31 (2006).
47. Black, S. *et al.* Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. *Pediatr. Infect. Dis. J.* **19**, 187–195 (2000).
48. Arnold, R.R., Cole, M.F. & McGhee, J.R. A bactericidal effect for human lactoferrin. *Science* **197**, 263–265 (1977).
49. Alexander, D.B., Iigo, M., Yamauchi, K., Suzui, M. & Tsuda, H. Lactoferrin: an alternative view of its role in human biological fluids. *Biochem. Cell Biol.* **90**, 279–306 (2012).
50. Arnold, R.R., Russell, J.E., Champion, W.J. & Gauthier, J.J. Bactericidal activity of human lactoferrin: influence of physical conditions and metabolic state of the target microorganism. *Infect. Immun.* **32**, 655–660 (1981).
51. Asensio, J.L., Arda, A., Canada, F.J. & Jimenez-Barbero, J. Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46**, 946–954 (2013).
52. Pemberton, A.D., Rose-Zerilli, M.J., Holloway, J.W., Gray, R.D. & Holgate, S.T. A single-nucleotide polymorphism in intelectin 1 is associated with increased asthma risk. *J. Allergy Clin. Immunol.* **122**, 1033–1034 (2008).
53. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
54. Schnaitman, C.A. & Klena, J.D. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **57**, 655–682 (1993).
55. Willis, L.M. *et al.* Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. *Proc. Natl. Acad. Sci. USA* **110**, 7868–7873 (2013).
56. Varghese, J.N., Mckimmbreschkin, J.L., Caldwell, J.B., Kortt, A.A. & Colman, P.M. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* **14**, 327–332 (1992).
57. Kraschnefski, M.J. *et al.* Effects on sialic acid recognition of amino acid mutations in the carbohydrate-binding cleft of the rotavirus spike protein. *Glycobiology* **19**, 194–200 (2009).
58. Blanchard, H., Yu, X., Coulson, B.S. & von Itzstein, M. 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 (2007).
59. Dormitzer, P.R., Sun, Z.Y.J., Wagner, G. & Harrison, S.C. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **21**, 885–897 (2002).
60. Sauter, N.K. *et al.* 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 (1992).

ONLINE METHODS

Chemical synthesis of glycans. Procedures for glycan synthesis are described in detail in the **Supplementary Note**.

Native human intelectin-1 expression and purification. The cDNA for hIntL-1 (NM_017625) was obtained from Open Biosystems clone LIFESEQ2924416 as a glycerol stock (GE Healthcare). The full coding sequence, residues 1–313, was amplified with PCR with the forward primer 5′-CGTGGGATCCTGGAGGGAG GGAGTGAAGGAGC-3′ and the reverse primer 5′-GCCAGCTCGAGACCTTG GGATCTCATGGTTGGGAGG-3′. The primers included sites for the restriction endonucleases BamHI and *XhoI*, 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 by DNA sequencing (UW–Madison Biotechnology Center).

The gene encoding hIntL-1 was expressed via transient transfection of suspension-adapted HEK 293T cells obtained from the American Type Culture Collection (ATCC). Cells were transfected in Opti-MEM I Reduced Serum Medium (Life Technologies) at ~2×10^6 cells/mL with Lipofectamine 2000 (Life Technologies), according to the manufacturer's protocol. Six hours after transfection, the culture medium was exchanged to FreeStyle F17 expression medium (Life Technologies) supplemented with 50 U/mL penicillin-streptomycin, 4 mM L-glutamine, 1× nonessential amino acids, 0.1% FBS and 0.1% Pluronic F-68 (Life Technologies). Cells expressing hIntL-1 were cultured for up to 6 d, 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 ($CaCl_2$) 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*f* column generated from reaction of a β-Gal*f* glycoside bearing an anomeric linker and an amine to UltraLink Biosupport (Pierce). The resulting resin was washed with a solution of 20 mM HEPES, pH 7.4, 150 mM sodium chloride (NaCl) and 10 mM $CaCl_2$. hIntL-1 was eluted with a solution of 20 mM HEPES, pH 7.4, 150 mM NaCl and 10 mM EDTA, and the protein was concentrated with a 10,000 molecular-weight-cutoff (MWCO) Amicon Ultra centrifugal filter. The buffer was exchanged to 20 mM HEPES, pH 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 according to absorbance at 280 nm, with a calculated $\varepsilon = 237{,}400\ cm^{-1}\ M^{-1}$ 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.

Expression and purification of *Strep*-tag II hIntL-1. An N-terminal *Strep*-tag II was cloned into the hItnL-1::pcDNA4 vector with site-directed mutagenesis and a primer set composed of 5′-ACCACCAGAGGATGGAGTACAGATTG GAGCCATCCGCAGTTTGAAAAGTCTACAGATGAGGCTAATACTTACT TCAAGGA-3′ and its reverse complement. 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 with NaOH, avidin was added per the IBA protocol (IBA, cat. no. 2-0205-050), $CaCl_2$ was added to 10 mM and the solution was cleared with centrifugation (15,000*g* for 15 min). Protein was captured onto 2 mL of *Strep*-Tactin Superflow resin (IBA, cat. no. 2-1206-002). The resulting resin was washed with a solution of 20 mM HEPES, pH 7.4, 150 mM NaCl, and 10 mM $CaCl_2$ and then 20 mM HEPES, pH 7.4, 150 mM NaCl and 1 mM EDTA. The protein was eluted with 5 mM *d*-desthiobiotin (Sigma) in 20 mM HEPES, pH 7.4, 150 mM NaCl and 1 mM EDTA and concentrated with a 10,000-MWCO Amicon Ultra centrifugal filter. The concentration of *Strep*–hIntL-1 was determined with absorbance at 280 nm, with a calculated $\varepsilon = 237{,}400\ cm^{-1}M^{-1}$ 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 after culture-medium dialysis against 20 mM bis-Tris, pH 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 protocol, $CaCl_2$ 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, pH 6.7, 150 mM NaCl, and 10 mM $CaCl_2$ and then with 20 mM bis-Tris, pH 6.7, 150 mM NaCl and 0.5 mM EDTA. Protein was eluted with 5 mM *d*-desthiobiotin (Sigma) in 20 mM bis-Tris, pH 6.7, 150 mM NaCl, and 0.5 mM EDTA and concentrated with a 10,000-MWCO Amicon Ultra centrifugal filter.

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-bottomed 96-well plate in PBS. Wells were washed with PBS and then coated with 5 µM of carbohydrate-biotin ligand in PBS for 1 h at 22 °C. Wells were blocked with bovine serum albumin (BSA) in ELISA buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$ and 0.1% Tween-20). Samples containing hIntL-1 were prepared by serial dilution into ELISA buffer with 0.1% bovine serum albumin (BSA) and added to wells for 2 h at 22 °C. Wells were washed four times with ELISA buffer. Bound hIntL-1 was detected with 0.75 µg/mL of a sheep polyclonal IgG hIntL-1 antibody (R&D Systems, cat. no. AF4254) in ELISA buffer with 0.1% BSA for 2 h 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 horseradish peroxidase (HRP) conjugate (Jackson ImmunoResearch Laboratories) was added at a 1:5,000 dilution in ELISA buffer with 0.1% BSA for 1 h at 22 °C. When *Strep*–hIntL-1 was assayed, *Strep*MAB-Classic HRP conjugate (IBA, cat. no. 2-1509-001) was used to specifically recognize the *Strep*-tag II of bound hIntL-1. *Strep*MAB-Classic HRP conjugate was diluted 1:10,000 in ELISA buffer with 0.1% BSA and incubated for 2 h at 22 °C. Wells were washed. hIntL-1 was detected colorimetrically with addition of 1-Step Ultra TMB-ELISA (Pierce). Once sufficient signal was achieved (typically in <2 min), the reaction was quenched by addition of an equal volume of 2 M sulfuric acid (H_2SO_4). 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 $CaCl_2$ in all steps. Data were analyzed in Prism6 (GraphPad). Data were fit to a one-site binding equation.

Surface plasmon resonance (SPR). Analysis of intelectins with 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, pH 7.4, 150 mM NaCl, 1 mM $CaCl_2$ and 0.005% Tween-20). Surfaces were regenerated with short injections of solutions of 10 mM hydrochloric acid (HCl). Data were referenced with either the interspots or the biotin reference channel and processed with the Bio-Rad ProteOn software package.

Construction of the furanoside glycan array. The microarray of furanoside-containing glycans was printed as previously described[61,62]. Briefly, the amine functionalized glycans shown in **Supplementary Figure 6a** were dissolved in 100 mM sodium phosphate, pH 8.0, and printed as 14 arrays on *N*-hydroxysuccinimidyl (NHS) ester–activated slides (Shott Nexterion). Arrays were printed in quadruplicate at different glycan concentrations (as indicated in **Supplementary Fig. 6b**) with a Piezorray printer (PerkinElmer) that delivered 0.33 nL per spot. The 2-amino(*N*-aminoethyl) benzamine (AEAB) derivatives of lacto-*N*-neotetraose (LNnT) and asialo, galactosylated biantennary 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 h. Slides were dried in a desiccator overnight and blocked with a solution of 50 mM ethanolamine in 50 mM borate buffer, pH 8.0. Prior to interrogation with glycan-binding proteins (GBPs), the arrays were rehydrated in binding buffer.

Assay of hIntL-1 on furanoside and CFG mammalian glycan array. GBPs at various concentrations were applied to separate furanoside arrays in 70 µL of binding buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 10 mM $CaCl_2$, 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 h at RT, the slides were washed with wash buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 10 mM $CaCl_2$ and 0.05% Tween-20). The biotinylated lectins *Erythrina cristagalli* lectin (ECL) and *Ricinus communis* agglutinin I lectin (RCA-I) were detected with Alexa Fluor 488–labeled streptavidin (10 µg/ml) in binding buffer (**Supplementary Fig. 6c**,**d**). hIntL-1 was detected with the same sheep polyclonal IgG antibody specific for hIntL-1 (5 µg/ml) (R&D Systems) and an Alexa Fluor 488–labeled donkey anti-sheep IgG secondary antibody (5 µg/ml) (Life Technologies). Bound protein was detected with a ProScanArray Scanner (PerkinElmer) equipped with four lasers covering an excitation range from 488 to 633 nm. The data from the furanoside glycan array were analyzed with ScanArray Express (PerkinElmer) as the average of the four replicates.

For the analysis of the CFG glycan array[36], hIntL-1 was applied in 70 µl at concentrations of 50 and 200 µg/ml in binding buffer under a coverslip to distribute the solution evenly over the large array of 610 glycans printed in sextuplicate (Array v5.1). After washing and scanning steps, the data from the CFG glycan microarray were analyzed with ImaGene software (BioDiscovery) as the average of four values after removal of the high and low values of the six 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 the print identification numbers that identified the glycan targets. Figures were made with Prism6 (GraphPad) or Excel (Microsoft).

Assay of hIntL-1 on the bacterial glycan array. *Strep*–hIntL-1 was used to interrogate the Microbial Glycan Microarray version 2 (MGMv2). Construction of the MGMv2 was as previously described[38]. Briefly, bacterial polysaccharide samples were dissolved and diluted to 0.5 mg/mL in printing buffer (150 mM sodium phosphate, pH 8.4, and 0.005% Tween-20). Samples were immobilized on NHS-activated glass slides (SlideH, Schott/Nexterion) with 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 h after printing at 100% relative humidity. The remaining reactive NHS moieties were quenched with a blocking solution (50 mM ethanolamine in 50 mM borate buffer, pH 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, pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM magnesium chloride ($MgCl_2$), 1% BSA and 0.05% Tween-20) and applied directly to the array surface for 1 h. After incubation, the array was washed by dipping into binding buffer four times. The *Strep*-tag II on bound hIntL-1 was detected with *Strep*MAB-Classic Chromeo 647 nm (10 µg /mL, IBA Lifesciences) diluted in binding buffer, applied directly to the array surface and allowed to incubate for 1 h. The array was washed in binding buffer (four dips), binding buffer without BSA and Tween-20 (four dips) and deionized water (four dips). Finally, the array was dried by centrifugation and scanned. Interrogated arrays were scanned for Chromeo 647 signal with a ProScanArray Express scanner (PerkinElmer), and resultant images were processed to extract signal data with Imagene (v6.0, Biodiscovery). Signal data were calculated as the average of four values after removal of the high and low values of the six replicates. Data were plotted with Excel (Microsoft) as average relative fluorescence units (RFU) versus print identification number. Figures were made with Prism6 (GraphPad).

Protein X-ray crystallography. The *Strep*–hIntL-1 protein that was purified with 20 mM bis-Tris, pH 6.7, was concentrated to 1.5 mg/mL, 1 M $CaCl_2$ was added to a final concentration of 10 mM, and crystallization (hanging-drop vapor diffusion) was achieved by mixture of 1 µL of the protein solution and 1 µL of well solution (100 mM bis-Tris, pH 6.0, and 25% PEG 3350). Crystals grew to full size in 2 weeks. Protein crystals of Apo–hIntL-1 were cryoprotected via transfer to well solution supplemented to a total concentration of 35% PEG 3350 for 1 min and then were vitrified in liquid nitrogen. The allyl-β-Gal*f*–hIntL-1 complex was formed by soaking of apo–hIntL-1 crystals in cryoprotection solution supplemented with 50 mM allyl-β-D-galactofuranose for 2 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 for Gal*f*-Bound hIntL-1. Integration, scaling, and merging were performed with HKL2000 (ref. 63). The structure was solved with the PHENIX suite[64]. The *Xenopus laevis* intelectin structure recently solved in our laboratory was used as a search model to determine the structure of apo–hIntL-1 by molecular replacement with Phaser[65]. Because the data for apo–hIntL-1– and β-Gal*f*–bound hIntL-1 are isomorphous, the structure of β-Gal*f*–bound hIntL-1 was solved by a difference Fourier method with apo–hIntL-1 as a starting model for rigid-body refinement with phenix.refine[66]. The chemical restraint for β-Gal*f* was generated by PRODRG[67]. Model adjustment and refinement were performed in Coot[68] and phenix.refine, respectively (**Supplementary Table 1**). The model was validated with MolProbity[69]. Crystal structure figures were generated with PyMOL (http://www.pymol.org/).

hIntL-1 binding to *S. pneumoniae*. *S. pneumoniae* (Klein) Chester serotypes 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[46]. 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 blood or in suspension in Luria Broth (LB). Cells were grown at 37 °C under 5% carbon dioxide gas. During liquid culture, cells were shaken at 100 r.p.m. 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 min on ice. Cells were stained with 15 µg/mL *Strep*–hIntL-1 with a 1:250 dilution of *Strep*MAB-Classic Oyster 645 conjugate (IBA, cat. no. 2-1555-050) in 20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$, 0.1% BSA and 0.05% Tween-20 for 2 h at 4 °C. To test the calcium-ion dependency of binding, 20 mM HEPES, pH 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, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$, 100 mM glycerol, 0.1% BSA and 0.05% Tween-20 was used as the buffer. Cells were washed with 20 mM HEPES, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$, 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 at the University of Wisconsin–Madison Carbone Canter Center (UWCCC) Flow Cytometry Laboratory. Propidium iodide was used to differentiate fixed *S. pneumoniae* cells from debris. Data were analyzed with FlowJo (http://www.flowjo.com/).

For analysis by microscopy, cell aliquots were taken directly from the flow cytometry samples before propidium iodide staining. Samples were subsequently stained with Hoechst 33342 (Life Technologies). Each sample was spotted onto a glass-bottomed microwell dish (MatTek corporation) and covered with a 1% (w/v) agarose pad prepared in a matched buffer. Images were collected at room temperature with a Nikon A1 laser scanning confocal microscope (Nikon Instruments). Images were acquired with a Nikon plan apo 100/1.4 oil objective with a 1.2-AU pinhole diameter and NIS-elements C software (Nikon Instruments). Laser settings were determined by imaging the brightest control sample, serotype 43 treated with 15 µg/mL *Strep*–hIntL-1 and a 1:250 dilution of *Strep*MAB-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 with a 405-nm laser, and emission was collected between 425 and 475 nm. For *Strep*MAB-Classic Oyster 645 conjugate, illumination was performed with a 638-nm laser, and emission was collected between 663 and 738 nm. Images were prepared with the open-source Fiji distribution of ImageJ, and brightness and contrast were adjusted in the control sample (serotype 43 treated with 15 µg/mL *Strep*–hIntL-1 with a 1:250 dilution of *Strep*MAB-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.

Expression of mouse intelectin-1. A detailed description of mIntL-1 expression is available in the **Supplementary Note**.

61. Song, X., Lasanajak, Y., Xia, B., Smith, D.F. & Cummings, R.D. Fluorescent glycosylamides produced by microscale derivatization of free glycans for natural glycan microarrays. *ACS Chem. Biol.* **4**, 741–750 (2009).

doi:10.1038/nsmb.3053

62. Heimburg-Molinaro, J., Song, X., Smith, D.F. & Cummings, R.D. Preparation and analysis of glycan microarrays. *Curr. Protoc. Protein Sci.* **64**, 12.10 (2011).
63. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
64. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
65. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
66. Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D. The Phenix refinement framework. *CCP4 Newsl.* **42**, 8 (2005).
67. Schüttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
68. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
69. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).