

# Selectivity and engineering of the sialoglycan-binding spectrum in Siglec-like adhesins

Barbara A. Bensing<sup>a,&</sup>, Lioudmila V. Loukachevitch<sup>b,&</sup>, Rupesh Agarwal<sup>c,d,%</sup>, Izumi Yamakawa<sup>b,%</sup>, Kelvin Luong<sup>b</sup>, Azadeh Hadadianpour<sup>e</sup>, Hai Yu<sup>f</sup>, Kevin P. Fialkowski<sup>b,†</sup>, Manuel A Castro<sup>g</sup>, Zdzislaw Wawrzak<sup>h</sup>, Xi Chen<sup>f</sup>, Jerome Baudry<sup>i</sup>, Jeremy C. Smith<sup>d,l</sup>, Paul M Sullam<sup>a</sup>, T M Iverson<sup>b,e,j,k,\*</sup>

From the <sup>a</sup>Division of Infectious Diseases, Veterans Affairs Medical Center, Department of Medicine, University of California, San Francisco, and the Northern California Institute for Research and Education, San Francisco, California 94121, USA; Departments of <sup>b</sup>Pharmacology, <sup>e</sup>Microbiology, Pathology, and Immunology, and <sup>g</sup>Biochemistry, <sup>i</sup>Center for Structural Biology and <sup>k</sup>Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37232, USA; <sup>c</sup>UT/ORNL Center for Molecular Biophysics, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6309, USA; <sup>d</sup>Graduate School of Genome Science and Technology, <sup>l</sup>Department of Biochemistry and Cellular and Molecular Biology, 37996 University of Tennessee, Knoxville, Tennessee, 37996, USA; <sup>f</sup>Department of Chemistry, University of California, Davis, CA 95616, USA, USA; <sup>h</sup>LS-CAT Synchrotron Research Center, Northwestern University, Argonne, Illinois, 60439, USA; <sup>i</sup>University of Alabama at Huntsville, Department of Biological Sciences. SST, Huntsville, Alabama 35899, USA

<sup>&</sup>These authors contributed equally to this manuscript.

<sup>%</sup>These authors contributed equally to this manuscript.

<sup>†</sup>Present address: College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

<sup>\*</sup>To whom correspondence may be addressed: ([tina.iverson@vanderbilt.edu](mailto:tina.iverson@vanderbilt.edu)).

**The Siglec-like Serine-Rich Repeat (SRR) adhesins mediate bacterial attachment to mammalian hosts via sialoglycan receptors. Here, we combine structural, computational, biochemical, and phylogenetic approaches to elucidate the determinants of the sialoglycan-binding spectrum across the family of Siglec-like SRR adhesins. We further identified mutable positions that disproportionately affect sialoglycan selectivity, as demonstrated by increases in binding to alternative ligands of 2- to 3- orders of magnitude. Biologically, these studies highlight how bacteria nimbly modulate the receptor interaction during coevolution of host and pathogen. These studies additionally created binding proteins specific for sialyl-T antigen or 6S-sialyl Lewis<sup>X</sup> that can recognize glycosylation of human plasma proteins. The engineered binding proteins can facilitate the characterization of normal cellular glycan modifications or may be used as diagnostic tools in disease states with altered glycosylation.**

**Significance:** The ability of bacteria to bind selectively to host receptors underlies both commensalism and pathogenesis. Here, we identify the molecular basis for receptor selectivity in streptococci that bind to sialoglycan receptors. This revealed how to convert these adhesins into selective probes that measure tri- and tetrasaccharides within the context of larger glycosylations. These probes that can be used in a laboratory with no specialized equipment and can be used to address biological questions relating to sialoglycan-dependent signaling and adhesion.

**T**he decoration of proteins with sialoglycans is functionally important in numerous mammalian signaling pathways. However, a wide array of bacterial and viral adhesive proteins exploit these sialoglycans as host receptors during infection. In these cases, sialoglycan selectivity determines whether a pathogen can adhere to a preferred anatomical niche or can infect a particular host (1, 2).

Sialic acid binding immunoglobulin like lectin (Siglec)-like adhesins are found within the larger family of Serine-rich repeat (SRR) adhesins (3-14), which form fibril-like protrusions on streptococci and staphylococci (15). The Siglec-like adhesins (3, 7, 16-23) always include two adjacent modules: a “Siglec” domain and a “Unique” domain (21). In oral streptococci, Siglec-like adhesins bind to carbohydrates containing a terminal Sia $\alpha$ 2-3Gal (Sia = Neu5Ac or Neu5Gc) (3, 4, 7-9, 16-19, 22) at a YTRY

sequence motif (16, 18, 21, 24). In humans, Neu5Ac $\alpha$ 2-3Gal is commonly found at the termini of the complex *O*-linked sialoglycans that modify the MUC7 salivary mucin (9, 16, 25) or glycoproteins in both blood plasma (17) and on platelets (8, 23). Binding to  $\alpha$ 2-3-linked sialoglycans may therefore allow colonization of the oral cavity, or can lead to endovascular infection (4, 21, 22, 26).

Sequences of the Siglec-like adhesins are quite variable, as are the host sialoglycans to which they bind. For example, GspB from *Streptococcus gordonii* strain M99 binds with narrow selectivity to the sialyl-T antigen (sTa) trisaccharide (3, 16, 19) (for carbohydrate structures, see **Fig. S1**). In contrast, other SRR adhesins, such as Hsa from *S. gordonii* strain Challis, bind with high avidity to multiple glycans (3, 16, 19).

Here, we evaluated sialoglycan binding and selectivity of Siglec-like adhesins using structural, computational, and biochemical approaches. We then used this information to engineer adhesins with altered binding properties and showed that this affected the preferred host receptor. Our findings provide insights into the molecular basis for sialoglycan selectivity by Siglec-like adhesins and suggest a route for developing these adhesins into a broad array of tools to characterize sialoglycan distribution.

## Results

### Selection of representative adhesins

We began by correlating phylogenetic analysis of sialoglycan-binding Siglec and Unique domains (**Fig. S2**) with reported sialoglycan selectivity (3, 16, 17, 19, 20). This identified that evolutionary relatedness is a moderate, but not strong, predictor of glycan selectivity. In short, most of the adhesins of the first major branch of the tree (blue in **Fig. S2**) bound two or more related tri- or tetrasaccharides, albeit without a clear glycan preference (3, 16, 17, 19, 20). In contrast, the four characterized adhesins of the second major branch (green in **Fig. S2**) exhibit narrow selectivity for sTa (3, 16, 17, 19, 20).

From the first branch of the tree (blue in **Fig. S2**), we selected the Siglec and Unique domains of Hsa (termed Hsa<sub>Siglec+Unique</sub>), and the equivalent domains from *Streptococcus sanguinis* strain SK678 and *Streptococcus mitis* strain NCTC10712 for further study. These three adhesins are >80% identical but exhibit different receptor selectivity. Hsa<sub>Siglec+Unique</sub> binds detectably to a broad range of Sia $\alpha$ 2-3Gal $\beta$ 1-3/4HexNAc glycans but not to fucosylated derivatives (16, 19). In comparison, SK678<sub>Siglec+Unique</sub> exhibits narrow selectivity for 3'-sialyl-*N*-acetylglucosamine (3'sLn) and 6-*O*-sulfo-sialyl Lewis X (6S-sLe<sup>X</sup>), while NCTC10712<sub>Siglec+Unique</sub> binds strongly to a range of 3'sLn-related structures (16). The combination of high sequence identity and distinct binding spectrum

suggests that we will be able to pinpoint the origins of sialoglycan selectivity with these comparators.

The second major branch of the evolutionary tree (green in **Fig. S2**) includes GspB from *S. gordonii* strain M99 (3, 8, 9, 21, 27). GspB<sub>Siglec+Unique</sub> exhibits narrow selectivity for the sTa trisaccharide, as do the other previously-characterized members of this evolutionary branch (3, 16, 17, 19, 20). In seeking comparators of GspB, we performed binding studies on additional homologs. We identified that the Siglec and Unique domains of the adhesin from *S. gordonii* strain SK150 (termed SK150<sub>Siglec+Unique</sub>) are 62% identical to the corresponding regions of GspB but exhibit broader carbohydrate selectivity (**Fig. S3**). The distinct binding properties make these good comparators for understanding sialoglycan selectivity.

### Structures of the Hsa-like and GspB-like adhesins

Using these five comparators, we evaluated how sequence differences affect the structure. As determined by crystallography (**Fig. 1A-1D Table S1, S2**), all five adhesins exhibited similar folds of the individual domains (**Fig. 1A, 1B**). However, the interdomain angle differed between the Hsa-like and GspB-like adhesins in a way that correlates with phylogeny (**Fig. 1C, 1D, S4A**).

Notably, even in closely-related adhesins, the Siglec domain contains conformational differences in three loops of the V-set Ig fold: the CD loop, the EF loop, and the FG loop (**Fig. 1C, 1D**). Sequence variation of these adhesins disproportionately maps to these loops (**Fig. S4D**). Taken together, these studies of unliganded adhesins identify features that correlate with phylogeny and reveal regions of disproportionate variation in closely related adhesins.

### Sialoglycan binding and conformational selection

We next determined costructures of sTa with Hsa<sub>Siglec+Unique</sub> (**Fig. 2A**) or the Siglec domain of GspB (GspB<sub>Siglec</sub>) (**Fig. 2B**). In both costructures, sTa binds in a defined pocket of the Siglec domain (**Fig. 2A, 2B**). This pocket is analogous to the sTa-binding site identified in the SRR adhesin SrpA from *S. sanguinis* strain SK36 (18, 20) (**Fig. 2C**), which phylogenetically groups with Hsa (**Fig. S2**). Interactions between sTa and each adhesin involves the YTRY sialic acid-binding motif (Hsa<sup>338-341</sup> or GspB<sup>482-485</sup>) (**Fig. 2D, Fig. S5**) (16) and three inserts of the V-set Ig fold: the CD loop (Hsa<sup>284-296</sup> or GspB<sup>440-453</sup>), the EF loop (Hsa<sup>330-336</sup> or GspB<sup>475-481</sup>), and the FG loop (Hsa<sup>352-364</sup> or GspB<sup>499-511</sup>) (**Fig. 2, Fig. S5**). These same regions vary disproportionately in both sequence and conformation in the unliganded structures (**Fig. 1, S4D**).

The YTRY motif is located on the F-strand of the V-set Ig fold and contributes to binding the invariant terminal Sia $\alpha$ 2-3Gal of the target *O*-linked sialoglycans (18, 21, 24). However, the role of the three loops in glycan affinity and

selectivity is unknown. We queried whether these loops exhibited inherent flexibility, a property believed to correlate with the ability to evolve binding to new ligands (28-30). Temperature factor analysis suggests that these loops have high flexibility in the absence of ligand (**Fig. S6**). Moreover, these loops exhibit conformational differences between the ligand-bound and ligand-free structures (**Fig. 1, 3**). In the GspB<sup>Siglec</sup> structure, the helix of the FG loop rotates 10° in response to sTa which results in a maximal displacement of 1.3 Å (**Fig. 3A**) while in the Hsa<sup>Siglec+Unique</sup> structure, the EF loop moves 5.9 Å (**Fig. 3B**) and allows the Hsa<sup>K335</sup> carbonyl to form hydrogen-bonding interactions to the Neu5Ac C5 nitrogen and C4 hydroxyl.

To explore the conformations available to these loops, we performed MD simulations of unliganded Hsa<sup>Siglec+Unique</sup> and GspB<sup>Siglec+Unique</sup>. The loops surrounding the glycan binding pocket exhibited considerably more flexibility than other parts of the protein (**Fig. S7A–D**). Moreover, the ligand-bound conformation is among the predicted conformations sampled in the absence of ligand (**Fig. 3, S7**). Of particular note is the main chain carbonyl of Hsa<sup>K335</sup>, which forms a hydrogen bond to sTa in the experimental costructure and samples both the bound and unbound states in the apo form (**Fig. 3B, 3C**). These calculations predict that sTa shifts the equilibrium of the EF loop to the position observed in the crystal structure of the bound state (**Fig. 2D, 3B, 3C, S5A**). Together, these analyses support a conformational selection mechanism over an induced fit mechanism, a property that may allow adaptation to changes of the host *O*-glycan receptors.

To experimentally assess whether conformational selection could contribute to ligand binding, we focused on the broadly selective Hsa<sup>Siglec+Unique</sup>. We introduced rigidifying prolines or replaced glycines at predicted hinges (Hsa<sup>N333P</sup>, Hsa<sup>G287A/G288P</sup>), both of which are predicted to reduce the flexibility required for conformational selection. As controls, we developed variants that introduced glycines (Hsa<sup>L363G</sup>, Hsa<sup>S253G</sup>) (**Fig. S8A**). Hsa<sup>N333P</sup> was associated with substantially reduced sialoglycan binding for all ligands tested; Hsa<sup>G287A/G288P</sup> also exhibited reduced binding, but the effect was less pronounced (**Fig. S8B – S8D**). In contrast, glycine-substituted Hsa<sup>L363G</sup> and Hsa<sup>S253G</sup> exhibited binding similar to wild-type (**Fig. S8B – S8D**). These experiments provide support for a conformational selection mechanism.

### Sialoglycan binding spectrum

All characterized ligands of the Siglec-like SRR adhesins contain a Sia $\alpha$ 2-3Gal disaccharide at the non-reducing terminus (16, 19). However, the identity of, and linkage to, the adjacent sub-terminal sugar varies. Analysis of the contacts in the costructures of Hsa<sup>Siglec+Unique</sup> and GspB<sup>Siglec</sup> with sTa identified that the sub-terminal sugar predominantly contacts the CD loop and the FG loop of the

Siglec domain (**Fig. 2, S5**). In contrast, the Neu5Ac $\alpha$ 2-3Gal interacts with the YTRY motif and residues in the EF loop (**Fig. 2, S5**).

To test how these loops affect sialoglycan selectivity, we engineered chimeras with the backbone of one adhesin and the loops of a closely-related adhesin. In the SK678<sup>Hsa-loops</sup> and NCTC10712<sup>Hsa-loops</sup> chimeras, selectivity became more similar to that of Hsa than the parent adhesin (**Fig. 4, Table S3**). This indicates that a major determinant of selectivity in Hsa-like adhesins is the combined contribution of the CD, EF, and FG loops. We next assessed whether one loop dominates this effect using individual substitutions. SK678<sup>Hsa-CD-loop</sup> exhibited substantially decreased affinity for 3'sLn and 6S-sLe<sup>X</sup>, while SK678<sup>Hsa-EF-loop</sup> had somewhat increased binding for all of the glycans tested, and SK678<sup>Hsa-FG-loop</sup> had moderately decreased binding for 3'sLn and a substantially decreased affinity for 6S-sLe<sup>X</sup> (**Fig. S9A**). NCTC10712<sup>Hsa-CD-loop</sup> and NCTC10712<sup>Hsa-FG-loop</sup> exhibited differential changes in binding for the sialoglycan ligands, while NCTC10712<sup>Hsa-EF-loop</sup> increased the range of bound ligands, but left binding unchanged for the preferred ligands (**Fig. S9B**).

One interpretation of the chimeragenesis data takes into consideration the position of each loop with respect to the ligand (**Fig. 2, S5**). The residues of the EF loop only interact with sialic acid (**Fig. S5**) and may act in concert with the YTRY motif to support binding of the invariant region of the ligands, i.e. Sia $\alpha$ 2-3Gal. Yet substitution of the EF loop of the more promiscuous Hsa into SK678 and NCTC10712 resulted in a somewhat broader binding spectrum (**Fig. S9A, S9B**). We posit that flexibility of the EF loop (**Fig. 3B, 3C, S6, S7**) adjusts the orientation of the entire sialoglycan to optimize the interaction between the variant position of the ligand and the CD and FG loops (**Fig. 3**). If the EF loop controls the ligand orientation, then the CD and FG loops may act in synergy to select the glycan. In particular, the FG loop of Hsa restricts the binding pocket and inhibits accommodation of Fuc $\alpha$ 1-3GlcNAc, as reflected by the lower binding of sLe<sup>X</sup> and 6S-sLe<sup>X</sup> to SK678<sup>Hsa-FG-loop</sup> and NCTC10712<sup>Hsa-FG-loop</sup> (**Fig. 5D, 5E, S7A, S7B**).

We next evaluated chimeras of the GspB-like adhesins. GspB<sup>SK150-CD-loop</sup> and GspB<sup>SK150-FG-loop</sup> substantially decreased glycan affinity; as with the Hsa-like adhesins, GspB<sup>SK150-EF-loop</sup> had little impact (**Fig. S9C**). However, in the GspB<sup>SK150-loops</sup> chimera, which substituted all three loops, the binding affinity remained low (**Fig. S9C**). One explanation for the uneven success of chimeragenesis is that the Hsa-like chimeras used starting adhesins with more flexible loops that could better adjust to the non-native scaffold. It is also possible that the Hsa-like adhesins benefitted from a better starting match between the sequences. To evaluate these possibilities, we engineered

GspB-SK150 “mini-chimeras” swapping only residues that directly contact the ligand (**Fig. S5**). The GspB<sup>L442Y/Y443N</sup> mini-chimera had increased binding to 3’sLn and sLe<sup>c</sup> and was overall more similar in selectivity to SK150 than to GspB (**Fig. S3, S10A, S10B**); the converse mini-chimera of SK150 still exhibited reduced binding (**Fig. S10C**). The incomplete success of the mini-chimeras suggests that both the sequence match of the starting adhesins and the limited loop flexibility impacted the ability to alter selectivity via chimeragenesis.

### Engineered adhesins with altered selectivity

If we are correct that selectivity is largely conferred by the CD and FG loops, we should be able to engineer the binding spectrum through mutation of these loops. We selected the Hsa-like adhesins, where chimeragenesis had greater success (**Fig. 4**), possibly as a result of increased loop flexibility. Hsa<sup>E286</sup> (in the CD loop) and Hsa<sup>D356</sup> (in the FG loop) each directly contact the GalNAc of sTa (**Fig. 2D**). In these positions and the equivalent positions of SK678 and NCTC10712, we substituted residues predicted to alter hydrogen-bonding characteristics.

As measured by ELISA, our engineered adhesins exhibited altered binding spectra (**Table S3, Fig. 5, Fig. S11**). The most striking results were for variants of the CD loop of NCTC10712<sup>Siglec+Unique</sup> and SK678<sup>Siglec+Unique</sup>, which become highly selective for 6S-sLe<sup>X</sup> via a substantial increase in binding for this sulfated tetrasaccharide and a decrease in binding to other glycans (**Fig. 5A, 5B**). Variants of the FG loop lost binding to fucosylated ligands but had little increase in binding to alternative ligands (**Fig. 5C, 5D**). As a result, the NCTC10712<sup>Q345D</sup> variant became more selective for 3’sLn while the SK678<sup>Q371D</sup> variant exhibited low binding to all tested ligands. The observed loss of binding to the fucose-containing sLe<sup>X</sup> and 6S-sLe<sup>X</sup> by FG loop variants is consistent with the chimeragenesis showing that the FG loop is particularly important for accommodation of fucosylated sialoglycans (**Fig. S9A, S9B**).

We also found that Hsa<sup>E286R</sup>, Hsa<sup>D356R</sup>, and Hsa<sup>D356Q</sup> (**Fig. S11**) increased binding to 3’sLn, sLe<sup>c</sup>, sLe<sup>X</sup>, and 6S-sLe<sup>X</sup> as compared to wild-type (**Fig. 4A**), but showed different degrees of discrimination between 3’sLn, sLe<sup>c</sup>, and sTa. Hsa<sup>E286R</sup> showed similar binding to all ligands tested and thus is even more broadly selective than wild-type (**Fig. S11A**). In contrast, Hsa<sup>D356R</sup> and Hsa<sup>D356Q</sup> each had an increase in 3’sLn, sLe<sup>X</sup>, and 6S-sLe<sup>X</sup> binding, but a decrease in sTa binding. Or in other words, these variants bind to a broad range of ligands with a distinct ligand preference from wild-type via a gain-of-function mechanism.

### Engineered adhesins show differential recognition of human plasma glycoproteins

We used these engineered adhesins to assess glycosylation of plasma proteins. Our prior studies identified that Hsa<sup>Siglec+Unique</sup> preferentially binds proteoglycan 4 (460 kD) in human plasma while NCTC10712<sup>Siglec+Unique</sup> binds GPIb $\alpha$  (150 kD). These adhesins also bind different glycoforms of the C1-esterase inhibitor (100 kD) (17). Here, Far Western analysis showed that the NCTC10712<sup>Hsa-loops</sup> and SK678<sup>Hsa-loops</sup> chimeras recognized proteoglycan 4 rather than the preferred receptors for wild-type SK678<sup>Siglec+Unique</sup> and NCTC10712<sup>Siglec+Unique</sup> (**Fig. 6A**). We also found that the 6S-sLe<sup>X</sup>-selective SK678<sup>E302R</sup> variant binds both GPIb $\alpha$  and the C1-esterase inhibitor (**Fig. 6B**), a binding pattern similar to that of wild-type NCTC10712<sup>Siglec+Unique</sup> (**Fig. 6A**). This latter finding suggests that 6S-sLe<sup>X</sup> is present on both GPIb $\alpha$  and C1-esterase inhibitor. It also identifies that SK678<sup>E302R</sup> will be useful as a probe for detecting this modification.

### Discussion

Individual Siglec-like adhesins recognize sialoglycans with as few as three and possibly more than six linked sugars (16, 17, 19, 20). Many of these adhesins bind to a preferred ligand with narrow selectivity, and many, like Hsa, bind strongly to multiple ligands (16, 17, 19, 20). Our results suggest that for the Siglec-like adhesins that recognize trisaccharides, the binding pockets contain two distinct recognition regions. The first region interacts with the sialic acid-containing non-reducing terminus of the sialoglycan, i.e. Sia $\alpha$ 2-3Gal (18, 20). This region is formed from both the YTRY motif on the F-strand (16, 18, 21, 24) and the EF loop (**Fig. 2**). The second region selects for the reducing end sugar and is tuned by the CD and FG loops of the V-set Ig fold (**Fig. 4, 5, S5–S11**). One advantage of this architecture is that the likely flexible trisaccharides can productively interact with the binding pocket via multiple approaches, i.e. binding the sialic acid first or by binding the reducing terminus of the glycan first. The concept of a binding site with multiple independent recognition regions can be extrapolated to adhesins that recognize larger sialoglycans. For example, the Siglec-like adhesin SrpA may biologically recognize a hexasaccharide (17) but can bind to partial ligands, albeit with low affinity (16, 18, 20) (**Fig. 2C**).

Mutagenesis (**Fig. S8**), chimeragenesis (**Fig. 4, S9, S10**), and computer simulations (**Fig. 3, S7**) all suggest that flexibility of these loops controls the breadth of the binding spectrum via a conformational selection mechanism. Binding promiscuity correlates with the identity of the EF loop (**Fig. S9**), which suggests a mechanism where the EF loop adjusts ligand orientation.

The variable region of the ligand can then approach the CD and FG selectivity loops at different angles in order to optimize interactions with the myriad of positions of hydrogen-bonding donors and acceptors that decorate the diverse glycans recognized. If the ligand binds to the CD and FG loops first, the order of events would be reversed, but the mechanism unchanged.

Chimeragenesis and mutagenesis also indicate that the CD and FG loops are particularly important in determining the preferred ligand (**Fig. 4, 5, S9 – S11**). The use of loops to control selectivity has previously been observed in other sialoglycan-binding systems. For example, the mammalian Siglec proteins are built upon a V-set Ig-fold but are not detectably related in sequence to the SRR adhesins (20, 21, 31, 32). In Siglec-7, the CC' loop (33) controls sialoglycan selectivity. From an evolutionary standpoint, having a mutable loop control selectivity makes particular sense for oral bacteria because it allows facile alteration of ligand preference in response to a changing environment. Indeed, mutation of loops is unlikely to impact protein stability.

The more promiscuous Hsa-like adhesins appeared to be particularly amenable to engineering (**Fig. 4, 5, S11**) and mutants exhibited binding increases of 2- to 3- orders of magnitude for non-native ligands. These increases exceed those reported for dedicated engineering studies (34-42), where the maximum enhancement in binding to a non-native glycan is ~20-fold (34-39) but selectivity is often achieved via a decrease in affinity to non-desired ligands in a promiscuous starting lectin (40-42). One intriguing interpretation of the unusually facile engineering of these Siglec-like adhesins is that their biological role necessitates adjusting to changes in host environment. An easily mutable adhesin may confer a survival advantage by allowing a bacterial strain to adapt to changes in the glycan modifications on salivary MUC7, adapt to binding to distinct receptors in a new anatomical location, or even to adapt to a new host. One impact of adaptation to different preferred receptors could be the ability of these bacteria to convert from commensals to pathogens (26).

An exciting outcome is our engineering of adhesins selective for sTa (**Fig. 4E**) and 6S-sLe<sup>x</sup> (**Fig. 5A, 5B**) on human proteins (**Fig. 6**). Adhesins with novel sialoglycan selectivity have multiple potential applications. The inherent challenges associated with characterizing *O*-glycans leave many biological questions arising from knowledge of sialoglycan distribution under-addressed. One strategy for mapping the glycome has been to repurpose naturally-occurring glycan-binding proteins as probes (43). Engineered probes could expand the range of detectable glycans. A second application is in detecting altered glycosylation in disease. Overexpression of sialoglycans is a biomarker for many types of cancers and commonly associated with poor prognosis (44-49). Robust antibodies to many sialoglycans, in particular sialyl-

Thompson-nouvelle antigen (sTn) have proven a challenge to develop (50). One could envision highly-selective lectins being used for detection of sialoglycans via lectin-based microarrays (43) or ELISAs. These may also be used in histological mapping or affinity purification of specific protein glycoforms. Additional work could develop probes selective for other Sia $\alpha$ 2-3Gal-linked sialoglycans or for sialoglycans with other linkages and is a future goal.

## Methods

Detailed procedures for protein expression, purification, crystallization, structure determination, computational analyses, and binding assays are described in the SI.

## Accession numbers

Atomic coordinates and structure factors have been deposited into the RCSB Protein Data Bank and raw data have been deposited into SGrid with accession codes listed in **Tables S1 and S2**.

## Acknowledgements

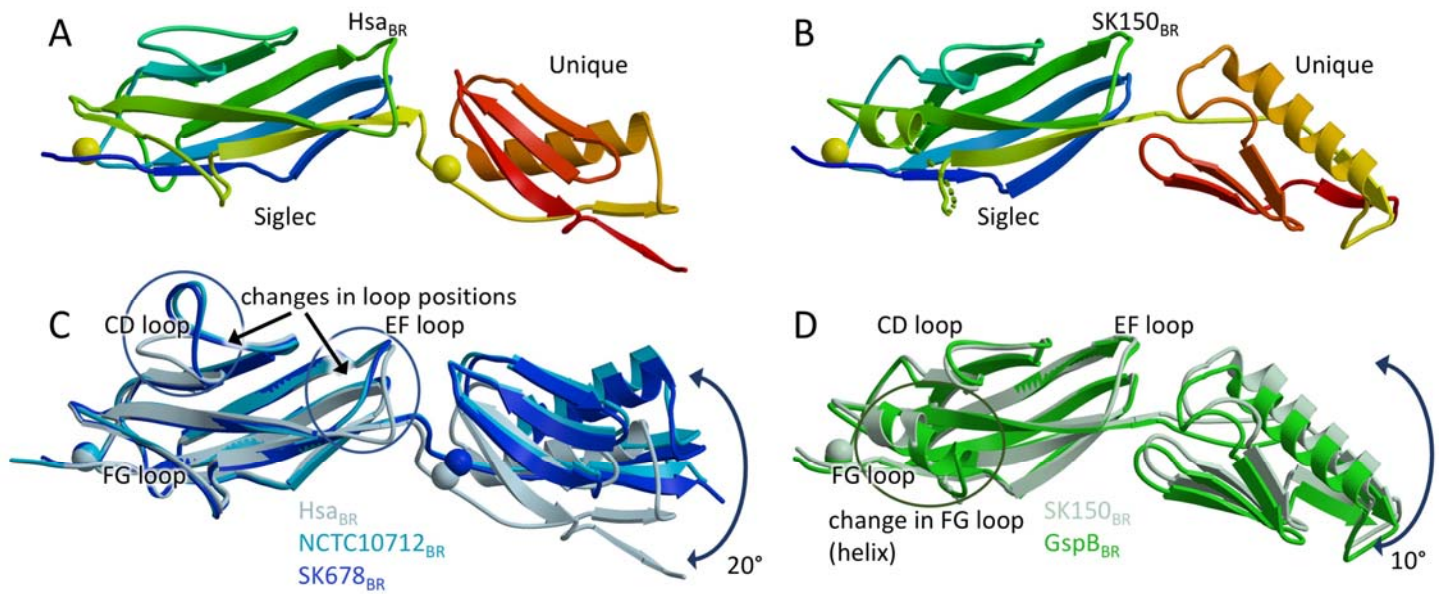
We thank S Bordenstein for assistance with phylogenetic analysis and B Bachmann, R Woods, and O Grant for helpful discussions. This work was supported by the Department of Veterans Affairs, the National Institutes of Health (AI41513 and U01CA221244 to PS, HD065122 to XC, AI106987 to TMI/PMS, and DE019807 to TMI), and the American Heart Association (14GRNT20390021 to TMI; 17SDG33660424 to BAB). AH was supported by the Vanderbilt International Scholars Program, and MAC was supported by a National Science Foundation Individual pre-doctoral fellowship (DGE-1445197). Use of the Stanford Synchrotron Radiation Lightsource, SLAC

National Accelerator Laboratory, is supported by the U.S. Department of Energy, 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 National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The Advanced Photon Source, a User Facility operated for the U.S. DOE Office of Science, was supported under Contract DE-AC02-06CH11357. LS-CAT Sector 21 is supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (085P1000817).

## Author contributions

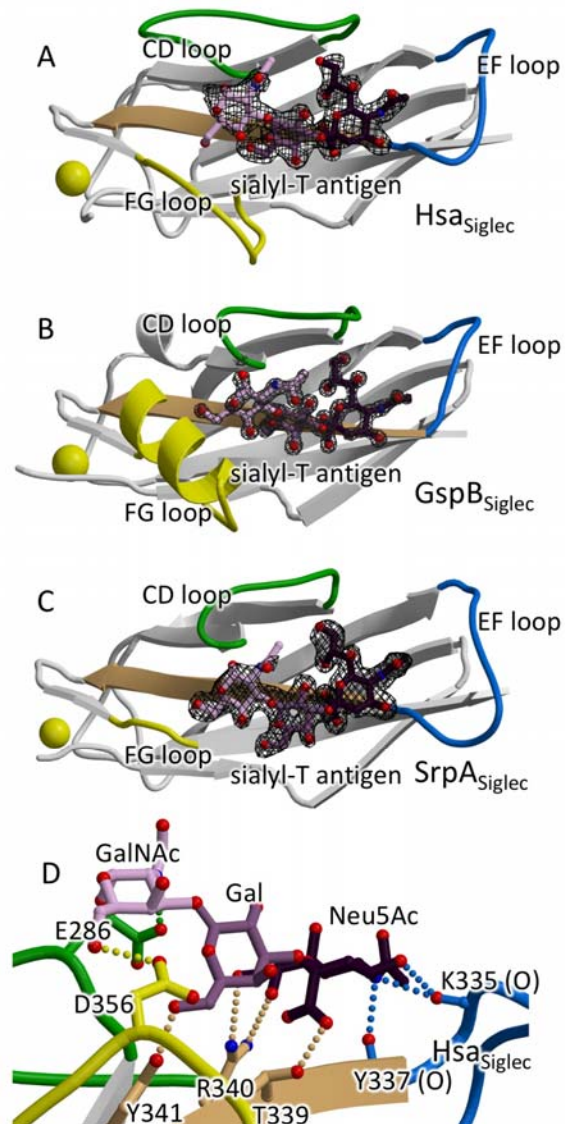
TMI and BAB designed research, BAB, LVL, IY, KL, AH, RA and ZW performed research, BAB, MAC, KL, IY, KF and TMI analyzed data, HY and XC contributed reagents/analytical tools. JB, JCS, PMS and TMI conceived and guided the study. TMI wrote the manuscript. All authors approved the final version of the manuscript.

**Figure 1.**



**Figure 1. Structures of sialoglycan binding regions of Siglec-like SRR adhesins.** A and B. Ribbon diagrams of A. Hsa-Siglec\_Unique and B. SK150<sub>Siglec+Unique</sub> with the N-terminus in *blue* and the C-terminus in *red*. Ions are shown as spheres. C. Hsa-like adhesins. Hsa<sub>Siglec+Unique</sub> is in *grey*, NCTC10712<sub>Siglec+Unique</sub> is in *cyan*, and SK678<sub>Siglec+Unique</sub> is in *blue*. D. GspB-like adhesins. GspB<sub>Siglec+Unique</sub> is in *green* and SK150<sub>Siglec+Unique</sub> is in *light green*.

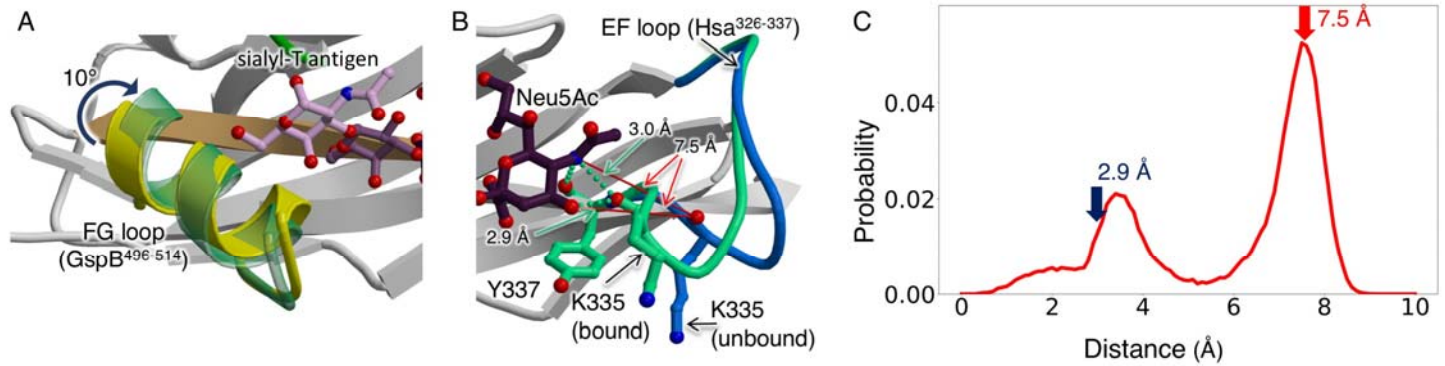
**Figure 2.**



**Figure 2. Costructures of sTa bound to Siglec-like SRR adhesins.** A-C. Siglec domain of A. Hsa, B. GspB, and C. SrpA (PDB entry 5IJ3 (18)) bound to sTa.  $F_o - F_c$  omit electron density contoured at  $3\sigma$  is shown as black mesh. In each panel, the F-strand harboring the “YTRY” motif is shown in *tan*, variable loops that interact with the ligand are shown in *green* (CD loop), *blue* (EF loop), and *yellow* (FG loop). D. Contacts between Hsa<sub>Siglec</sub> and sTa. The color of the hydrogen-bond reflects the structural element involved in the interaction.

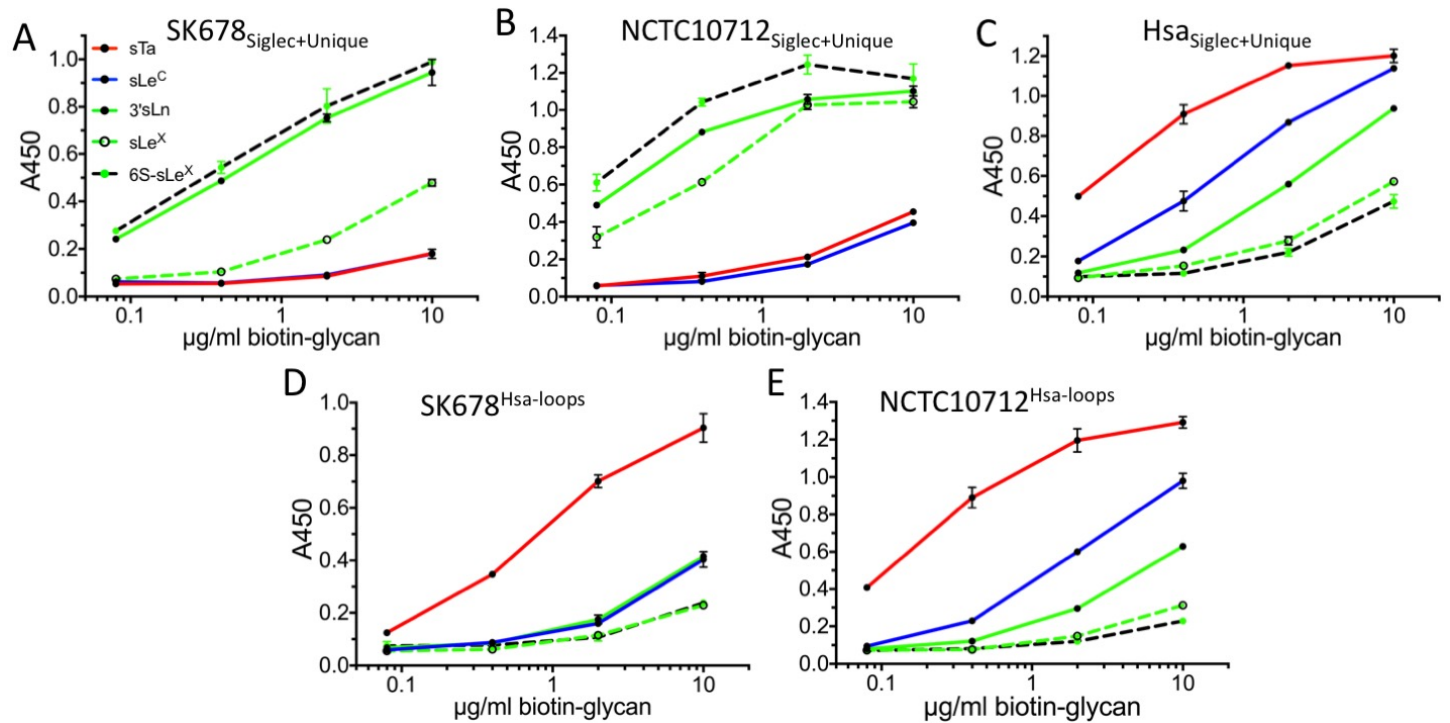


**Figure 3.**



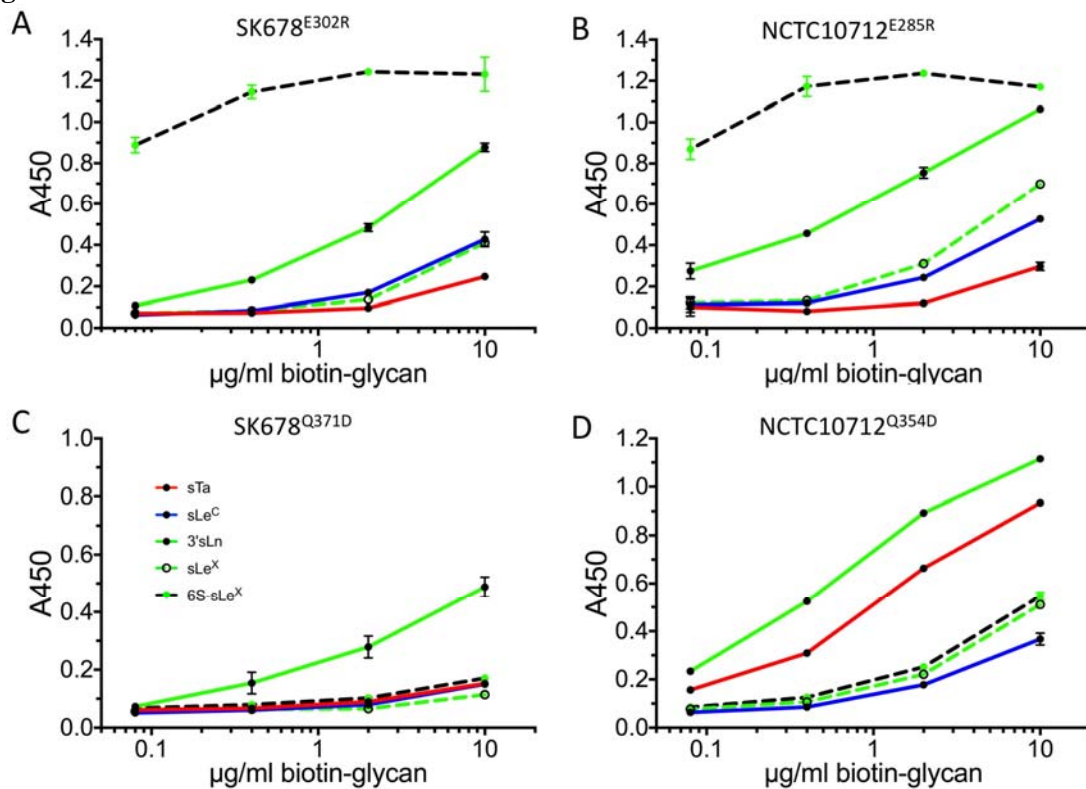
**Figure 3. Conformational changes associated with sTa binding.** A. The FG loop of GspB<sup>Siglec</sup> rotates 10° upon sTa binding. B. The EF loop of Hsa<sup>Siglec+Unique</sup> adjusts to promote formation of hydrogen-bonding interactions between Hsa<sup>K335</sup> and the Neu5Ac of sTa. C. Probability of distance distribution between the Neu5Ac O4 hydroxyl of sTa and the Hsa<sup>K335</sup> backbone carbonyl, as calculated by MD simulations. A bimodal distribution of distances exhibit maxima at 7.5 Å, which reflects the unliganded crystal structure, and at 3.5 Å, which approaches the liganded crystal structure. The formation of the hydrogen-bond between the Hsa<sup>K335</sup> carbonyl and Neu5Ac likely further shifts the conformational equilibrium to a pose that supports the 2.9 Å distance observed in the bound state.

**Figure 4.**



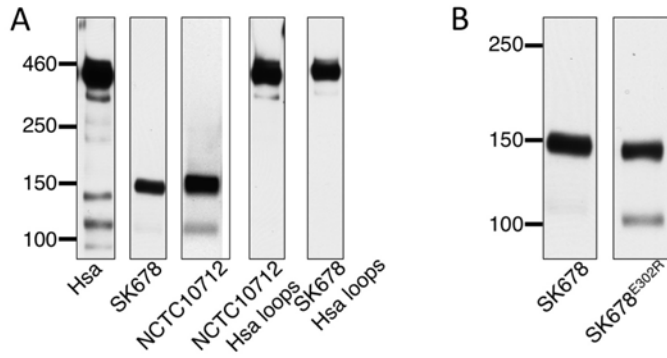
**Figure 4. Chimeragenesis of Hsa-like adhesins.** Dose-response curves of A. wild-type GST-SK678<sup>Siglec+Unique</sup>, B. wild-type GST-NCTC10712<sup>Siglec+Unique</sup>, and C. wild-type GST-Hsa<sup>Siglec+Unique</sup> to five selected ligands. D and E. Dose-response curves of the chimeras D. GST-SK678<sup>Hsa-loops</sup> and E. NCTC10712<sup>Hsa-loops</sup> which contain the CD, EF, and FG loops of Hsa. In each case, sTa binding increases. Measurements were performed using 500 nM of immobilized GST-adhesin and the indicated concentrations of each ligand, and are shown as the mean±SD (n = 2).

**Figure 5.**



**Figure 5. Binding selectivity of select engineered adhesins.** Dose response curves of A. GST-SK678<sup>E302R</sup>, B. GST-NCTC10712<sup>E285R</sup>, C. GST-SK678<sup>Q371D</sup>, and D. GST-NCTC10712<sup>Q345D</sup>. When compared to wild-type (see **Fig. 5A, 5B**), the GST-SK678<sup>E302R</sup> and GST-NCTC10712<sup>E285R</sup> variants exhibit increased binding to 6S-sLe<sup>X</sup>, and reduced binding to 3'sLn and sLe<sup>X</sup>. Conversely, both the GST-SK678<sup>Q371D</sup> and GST-NCTC10712<sup>Q345D</sup> variants have substantially reduced binding to the fucosylated ligands sLe<sup>X</sup> and 6S-sLe<sup>X</sup>. Measurements were performed using 500 nM of immobilized GST-adhesin and the indicated concentrations of each ligand, and are shown as the mean±SD (n = 2).

**Figure 6.**



**Figure 6. Far Western analysis of wild-type, chimeric, and engineered Hsa-like SRR adhesins binding to plasma proteins.** A. Hsa-like chimeras. B. The SK678<sup>E302R</sup> point mutant. As previously identified by affinity capture and mass spectrometry (17), the 460 kD band is proteoglycan 4, the 150 kD band is GP1b $\alpha$ , and the 100 kD band is C1-esterase inhibitor. Each Far Western was performed at least twice, with a representative blot shown.

- Langereis MA, *et al.* (2015) Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep* 11(12):1966-1978.
- Varki A (2006) Nothing in glycobiology makes sense, except in the light of evolution. *Cell* 126(5):841-845.
- Bensing BA, Lopez JA, & Sullam PM (2004) The Streptococcus gordonii surface proteins GspB and Hsa mediate binding to sialylated carbohydrate epitopes on the platelet membrane glycoprotein Ibalpha. *Infect Immun* 72(11):6528-6537.
- Lizcano A, Sanchez CJ, & Orihuela CJ (2012) A role for glycosylated serine-rich repeat proteins in gram-positive bacterial pathogenesis. *Mol Oral Microbiol* 27(4):257-269.
- Obert C, *et al.* (2006) Identification of a Candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect Immun* 74(8):4766-4777.
- Siboo IR, Chambers HF, & Sullam PM (2005) Role of SraP, a Serine-Rich Surface Protein of Staphylococcus aureus, in binding to human platelets. *Infect Immun* 73(4):2273-2280.
- Takahashi Y, Konishi K, Cisar JO, & Yoshikawa M (2002) Identification and characterization of hsa, the gene encoding the sialic acid-binding adhesin of Streptococcus gordonii DL1. *Infect Immun* 70(3):1209-1218.
- Takamatsu D, *et al.* (2005) Binding of the Streptococcus gordonii surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibalpha. *Mol Microbiol* 58(2):380-392.
- Takamatsu D, Bensing BA, Prakobphol A, Fisher SJ, & Sullam PM (2006) Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. *Infect Immun* 74(3):1933-1940.
- Yang YH, *et al.* (2014) Structural insights into SraP-mediated Staphylococcus aureus adhesion to host cells. *PLoS Pathog* 10(6):e1004169.
- Sanchez CJ, *et al.* (2010) The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. *PLoS Pathog* 6(8):e1001044.
- Seo HS, Mu R, Kim BJ, Doran KS, & Sullam PM (2012) Binding of glycoprotein Srr1 of Streptococcus agalactiae to fibrinogen promotes attachment to brain endothelium and the development of meningitis. *PLoS Pathog* 8(10):e1002947.
- Sequeira S, *et al.* (2018) Structural basis for the role of serine-rich repeat proteins from Lactobacillus reuteri in gut microbe-host interactions. *Proc Natl Acad Sci U S A* 115(12):E2706-E2715.
- Singh AK, Woodiga SA, Grau MA, & King SJ (2017) Streptococcus oralis Neuraminidase Modulates Adherence to Multiple Carbohydrates on Platelets. *Infect Immun* 85(3).
- Ramboarina S, *et al.* (2010) Structural insights into serine-rich fimbriae from Gram-positive bacteria. *J Biol Chem* 285(42):32446-32457.
- Bensing BA, *et al.* (2016) Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. *Glycobiology*.
- Bensing BA, Li Q, Park D, Lebrilla CB, & Sullam PM (2018) Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: implications for infective endocarditis. *Glycobiology* 28(8):601-611.
- Bensing BA, *et al.* (2016) Structural Basis for Sialoglycan Binding by the Streptococcus sanguinis SrpA Adhesin. *J Biol Chem* 291(14):7230-7240.
- Deng L, *et al.* (2014) Oral streptococci utilize a Siglec-like domain of serine-rich repeat adhesins to preferentially target platelet sialoglycans in human blood. *PLoS Pathog* 10(12):e1004540.
- Loukachevitch LV, *et al.* (2016) Structures of the Streptococcus sanguinis SrpA Binding Region with Human Sialoglycans Suggest Features of the Physiological Ligand. *Biochemistry* 55(42):5927-5937.
- Pyburn TM, *et al.* (2011) A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. *PLoS Pathog* 7(7):e1002112.
- Takahashi Y, *et al.* (2006) Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by Streptococcus gordonii DL1. *Infect Immun* 74(1):740-743.
- Plummer C, *et al.* (2005) A serine-rich glycoprotein of Streptococcus sanguis mediates adhesion to platelets via GPIb. *Br J Haematol* 129(1):101-109.
- Urano-Tashiro Y, Takahashi Y, Oguchi R, & Konishi K (2016) Two Arginine Residues of Streptococcus gordonii Sialic Acid-Binding Adhesin Hsa Are Essential for Interaction to Host Cell Receptors. *PLoS One* 11(4):e0154098.
- Thamadilok S, Roche-Hakansson H, Hakansson AP, & Ruhl S (2016) Absence of capsule reveals glycan-mediated binding and recognition of salivary mucin MUC7 by Streptococcus pneumoniae. *Mol Oral Microbiol* 31(2):175-188.
- Bashore TM, Cabell C, & Fowler V, Jr. (2006) Update on infective endocarditis. *Curr Probl Cardiol* 31(4):274-352.
- Xiong YQ, Bensing BA, Bayer AS, Chambers HF, & Sullam PM (2008) Role of the serine-rich surface glycoprotein GspB of Streptococcus gordonii in the pathogenesis of infective endocarditis. *Microb Pathog* 45(4):297-301.
- Haliloglu T & Bahar I (2015) Adaptability of protein structures to enable functional interactions and evolutionary implications. *Curr Opin Struct Biol* 35:17-23.
- Birmingham WR, *et al.* (2014) Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat Chem Biol* 10(5):392-399.
- Bhabha G, *et al.* (2013) Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol* 20(11):1243-1249.
- May AP, Robinson RC, Vinson M, Crocker PR, & Jones EY (1998) Crystal structure of the N-terminal domain of sialoadhesin in complex with 3' sialyllactose at 1.85 Å resolution. *Mol Cell* 1(5):719-728.
- Vinson M, *et al.* (1996) Characterization of the sialic acid-binding site in sialoadhesin by site-directed mutagenesis. *J Biol Chem* 271(16):9267-9272.

33. Alphey MS, Attrill H, Crocker PR, & van Aalten DM (2003) High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family. *J Biol Chem* 278(5):3372-3377.
34. Ielasi FS, Verhaeghe T, Desmet T, & Willaert RG (2014) Engineering the carbohydrate-binding site of Epa1p from *Candida glabrata*: generation of adhesin mutants with different carbohydrate specificity. *Glycobiology* 24(12):1312-1322.
35. Yabe R, *et al.* (2007) Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry. *J Biochem* 141(3):389-399.
36. Salomonsson E, *et al.* (2010) Mutational tuning of galectin-3 specificity and biological function. *J Biol Chem* 285(45):35079-35091.
37. Hu D, Tateno H, Kuno A, Yabe R, & Hirabayashi J (2012) Directed evolution of lectins with sugar-binding specificity for 6-sulfo-galactose. *J Biol Chem* 287(24):20313-20320.
38. Hu D, Tateno H, Sato T, Narimatsu H, & Hirabayashi J (2013) Tailoring GalNAc $\alpha$ 1-3Gal $\beta$ -specific lectins from a multi-specific fungal galectin: dramatic change of carbohydrate specificity by a single amino-acid substitution. *Biochem J* 453(2):261-270.
39. Abo H, *et al.* (2015) Mutated Leguminous Lectin Containing a Heparin-Binding like Motif in a Carbohydrate-Binding Loop Specifically Binds to Heparin. *PLoS One* 10(12):e0145834.
40. Imamura K, Takeuchi H, Yabe R, Tateno H, & Hirabayashi J (2011) Engineering of the glycan-binding specificity of *Agrocybe cylindracea* galectin towards  $\alpha$ (2,3)-linked sialic acid by saturation mutagenesis. *J Biochem* 150(5):545-552.
41. Sato T, *et al.* (2017) Engineering of recombinant *Wisteria floribunda* agglutinin specifically binding to GalNAc $\beta$ 1,4GlcNAc (LacdiNAc). *Glycobiology* 27(8):743-754.
42. Hu D, *et al.* (2015) Engineering of a 3'-sulpho-Gal $\beta$ 1-4GlcNAc-specific probe by a single amino acid substitution of a fungal galectin. *J Biochem* 157(4):197-200.
43. Pilobello KT, Slawek DE, & Mahal LK (2007) A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome. *Proc Natl Acad Sci U S A* 104(28):11534-11539.
44. Taga M, *et al.* (2015) A potential role for 6-sulfo sialyl Lewis X in metastasis of bladder urothelial carcinoma. *Urol Oncol* 33(11):496 e491-499.
45. Matsusako T, Muramatsu H, Shirahama T, Muramatsu T, & Ohi Y (1991) Expression of a carbohydrate signal, sialyl dimeric Le(x) antigen, is associated with metastatic potential of transitional cell carcinoma of the human urinary bladder. *Biochem Biophys Res Commun* 181(3):1218-1222.
46. Holst S, Wuhrer M, & Rombouts Y (2015) Glycosylation characteristics of colorectal cancer. *Adv Cancer Res* 126:203-256.
47. Patil SA, *et al.* (2014) Overexpression of  $\alpha$ 2,3sialyl T-antigen in breast cancer determined by miniaturized glycosyltransferase assays and confirmed using tissue microarray immunohistochemical analysis. *Glycoconj J* 31(6-7):509-521.
48. Munkley J (2016) The Role of Sialyl-Tn in Cancer. *Int J Mol Sci* 17(3):275.
49. Cazet A, Julien S, Bobowski M, Burchell J, & Delannoy P (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res* 12(3):204.
50. Loureiro LR, *et al.* (2015) Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules* 5(3):1783-1809.

# Selectivity and engineering of the sialoglycan-binding spectrum in Siglec-like adhesins

Barbara A. Bensing<sup>a,&</sup>, Lioudmila V. Loukachevitch<sup>b,&</sup>, Rupesh Agarwal<sup>c,d,%</sup>, Izumi Yamakawa<sup>b,%</sup>, Kelvin Luong<sup>b</sup>, Azadeh Hadadianpour<sup>e</sup>, Hai Yu<sup>f</sup>, Kevin P. Fialkowski<sup>b,†</sup>, Manuel A Castro<sup>g</sup>, Zdzislaw Wawrzak<sup>h</sup>, Xi Chen<sup>f</sup>, Jerome Baudry<sup>i</sup>, Jeremy C. Smith<sup>d,i</sup>, Paul M Sullam<sup>a</sup>, T M Iverson<sup>b,e,j,k,\*</sup>

From the <sup>a</sup>Division of Infectious Diseases, Veterans Affairs Medical Center, Department of Medicine, University of California, San Francisco, and the Northern California Institute for Research and Education, San Francisco, California 94121, USA; Departments of <sup>b</sup>Pharmacology, <sup>c</sup>Microbiology, Pathology, and Immunology, and <sup>g</sup>Biochemistry, <sup>i</sup>Center for Structural Biology and <sup>k</sup>Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37232, USA; <sup>e</sup>UT/ORNL Center for Molecular Biophysics, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6309, USA; <sup>d</sup>Graduate School of Genome Science and Technology, <sup>h</sup>Department of Biochemistry and Cellular and Molecular Biology, 37996 University of Tennessee, Knoxville, Tennessee, 37996, USA; <sup>f</sup>Department of Chemistry, University of California, Davis, CA 95616, USA, USA; <sup>h</sup>LS-CAT Synchrotron Research Center, Northwestern University, Argonne, Illinois, 60439, USA; <sup>i</sup>University of Alabama at Huntsville, Department of Biological Sciences. SST, Huntsville, Alabama 35899, USA

<sup>&</sup>These authors contributed equally to this manuscript.

<sup>%</sup>These authors contributed equally to this manuscript.

<sup>†</sup>Present address: College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

<sup>\*</sup>To whom correspondence may be addressed: ([tina.iverson@vanderbilt.edu](mailto:tina.iverson@vanderbilt.edu)).

## Supporting Information

### Materials and Methods

#### *Sequence analysis*

Sequences of the tandem Siglec and Unique domains were resected from select adhesins and were aligned using the MUSCLE (1) subroutine in Geneious Pro 11.1.4 (2). The JTT-G model of evolution was selected using the ProtTest server (3), and the phylogenetic tree was built using the MrBayes (4) subroutine in Geneious Pro 11.1.4 (2). A distantly-related adhesin from *S. mitis* strain SF100 (5) was used to root the tree.

#### *Cloning, expression, and purification for crystallization*

DNA encoding the adjacent Siglec and Unique domains of GspB, SK150, NCTC10712, or SK678 or the Siglec domain of GspB were cloned into the pBG101 vector (Vanderbilt University), which encodes an N-terminal His<sub>6</sub>-GST tag that is cleavable using 3C protease. Hsa<sup>Siglec+Unique</sup> was cloned into the pSV278 vector (Vanderbilt University), which encodes a His<sub>6</sub>-maltose binding protein (MBP) tag at the N-terminus followed by a thrombin cleavage site. Proteins were expressed in *E. coli* BL21 (DE3) in Terrific Broth medium (for GspB proteins and Hsa<sup>Siglec+Unique</sup>) or LB (for SK150<sup>Siglec+Unique</sup>, NCTC<sup>Siglec+Unique</sup> and SK678<sup>Siglec+Unique</sup>) with 50 µg/ml kanamycin at 37°C. When the OD<sub>600</sub> reached 0.6-1.4, expression was induced with 0.5-1 mM IPTG at 24°C for 3-7 hrs. Cells were harvested by centrifugation at 5,000 × *g* for 15 min, optionally washed with 0.1 M Tris-HCl, pH 7.5, and stored at -20 °C before purification. Frozen cells were resuspended in homogenization buffer (20-50 mM Tris-HCl, pH 7.5, 150-200 mM NaCl, 1mM EDTA, 1 mM PMSF, 2 µg/ml Leupeptin, 2 µg/ml Pepstatin) then disrupted by sonication. Lysate was clarified by centrifugation at 38500 × *g* for 35-60 min and passed through a 0.45 µm filter. Purification was performed at 4 °C. His<sub>6</sub>-GspB-fusion proteins were purified using a Glutathione Sepharose 4B column and were eluted with 30 mM GSH in 50 mM Tris-HCl, pH 8.0. His<sub>6</sub>-SK150<sup>Siglec+Unique</sup>/NCTC10712<sup>Siglec+Unique</sup>/SK678<sup>Siglec+Unique</sup> proteins were purified using Ni<sup>2+</sup> affinity chromatography and eluted with 20 mM Tris-HCl, 150 mM NaCl, 250 mM imidazole, pH 7.6. His<sub>6</sub>-MBP-Hsa<sup>Siglec+Unique</sup> was purified with an MBP-Trap column and eluted in 10 mM maltose. Eluted proteins were concentrated in a 10kD MW cut-off concentrator and exchanged into either PreScission cleavage buffer (GspB<sup>Siglec</sup>, GspB<sup>Siglec+Unique</sup>, SK150<sup>Siglec+Unique</sup>, NCTC10712<sup>Siglec+Unique</sup>, or SK678<sup>Siglec+Unique</sup>; 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1mM DTT) or thrombin cleavage buffer (Hsa<sup>Siglec+Unique</sup>; 20 mM Tris-HCl pH 7.5 and 200 mM NaCl). Affinity tags were cleaved with 1 U of appropriate protease (thrombin or 3C) per mg of protein overnight at 4 °C. For the SK150<sup>Siglec+Unique</sup>, NCTC10712<sup>Siglec+Unique</sup>, and SK678<sup>Siglec+Unique</sup>, the affinity tag has a similar molecular weight as the target protein; in these cases, the cleaved sample was passed through a Ni-column to remove the His<sub>6</sub>-GST tag. For GspB domains, adhesin was separated from the affinity tag by passing the cleavage reaction over the second Glutathione Sepharose 4B column in PreScission Buffer. Protein

aggregates were removed from GspB domains using a Superose-12 column in 50 mM Tris-HCl pH 7.6 and 150 mM NaCl. For the remaining proteins, aggregates were removed using a Superdex 200 increase 10/30 GL column equilibrated in 20 mM Tris-HCl pH 7.6 (NCTC<sub>Siglec+Unique</sub>, SK150<sub>Siglec+Unique</sub>, SK678<sub>Siglec+Unique</sub>) or in 20 mM Tris-HCl pH 7.5 and 200 mM NaCl (Hsa<sub>BR</sub>). After purification, all proteins were >95% pure as assessed by SDS-PAGE and were stored at -80 °C.

#### *Crystallization, data collection, and structure determination*

All crystallization reactions were performed at room temperature (~23 °C). Unless otherwise noted, diffraction data were collected at -180 °C, processed using HKL200 (6), and structures were determined by molecular replacement using the Phaser (7) subroutine of Phenix (8) and the search model indicated. Riding hydrogens were included at resolutions better than 1.4 Å. X-ray sources and data collection statistics are found in **Supporting Tables 1 & 2**.

**GspB**—GspB domains were crystallized by the sitting drop vapor diffusion method by equilibrating 1 µL protein and 1 µL reservoir solution over 50 µL of a reservoir solution. Purified GspB<sub>Siglec+Unique</sub> was concentrated to 9 mg/ml in 20 mM Tris-HCl, pH 7.6 and crystallized using a reservoir containing 0.2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 25% polyethylene glycol (PEG) 3350. Crystals were flash cooled by plunging into liquid nitrogen without the addition of cryo protectant. Purified GspB<sub>Siglec</sub> was concentrated to 22.8 mg/ml in 20 mM Tris-HCl, pH 7.2. Crystals in space group P2<sub>1</sub>2<sub>1</sub>2 were grown with a reservoir solution containing 0.2 M MgCl<sub>2</sub>, 0.1 M Tris-HCl, pH 8.5, 30% w/v PEG 4000; crystals in space group R32 were grown with a reservoir containing 4.0 M HCOONa. GspB<sub>Siglec</sub> was cocrystallized with sTa using reservoir conditions associated with the P2<sub>1</sub>2<sub>1</sub>2 space group and 1 µL of protein-ligand complex (20.5 mg/ml GspB<sub>Siglec</sub>, 10 mM sTa, 18 mM Tris-HCl, pH 7.2). Structures were determined using the appropriate domain(s) of GspB (PDB entry 3QC5 (9)) resected from the three-domain structure.

**SK150**—Purified SK150<sub>Siglec+Unique</sub> was concentrated to 3.5 mg/ml in 20 mM Tris-HCl, pH 7.6. Crystals were grown by the hanging drop vapor diffusion method by mixing 1 µL protein and 1 µL reservoir solution (0.2 M ammonium sulfate, 25% PEG 4000, 15% ethanol, and 0.1M Bis-tris, pH 7.0) and equilibrating over the reservoir solution. Diffraction data were collected at room temperature (~23 °C) and were processed using the PROTEUM suite. The structure was determined using the Siglec and Unique domains of GspB (PDB entry 3QC5 (9)) as the search model.

**Hsa**—Crystals of Hsa<sub>Siglec+Unique</sub> (21.6 mg/ml in 20 mM Tris-HCl, pH 7.2) grew by sitting drop vapor diffusion by equilibrating 1 µL protein and 2 µL reservoir solution over 50 µL of reservoir solution (0.1 M Succinate/Phosphate/Glycine pH 10.0 and 25% PEG 3350). Co-crystals of Hsa<sub>Siglec+Unique</sub> with sTa were prepared by soaking fully formed crystals in reservoir solution supplemented with 5 mM sTa for 20 hr. Crystals did not require cryoprotection beyond the reservoir solution. The structure of unliganded Hsa<sub>BR</sub> was determined using *S. sanguinis* SrpA<sub>Siglec+Unique</sub> (PDB entry 5EQ2 (10)) as the search model. The structure of sTa-bound Hsa<sub>BR</sub> was determined by rigid body refinement of unliganded Hsa<sub>Siglec+Unique</sub> in Phenix (8).

**NCTC10712**—Crystals of NCTC10712<sub>Siglec+Unique</sub> (3.5 mg/ml in 20 mM Tris-HCl pH 7.5) grew via the hanging drop vapor diffusion method using reservoir containing 0.1 M Tris-HCl pH 7.5 and 32% w/v PEG 4000. Crystal quality was improved by microseeding (Hampton Seed Bead kit) using 0.3 µL of seed, 1.2 µL protein (3.5 mg/ml), and 1.5 µL modified reservoir solution (0.1 M Tris-HCl pH 7.5 and 28% w/v PEG 4000). Crystals were cryoprotected in using a solution containing 50% of the reservoir and 50% glycerol, then cryocooled by plunging in liquid nitrogen. Data were processed using XDS (11). The structure was determined Hsa<sub>Siglec+Unique</sub> as the search model.

**SK678**—Crystals of SK678<sub>Siglec+Unique</sub> (7 mg/ml in 20 mM Tris-HCl pH 7.6) were grown via the hanging drop vapor diffusion method by equilibrating 1 µL of SK678<sub>Siglec+Unique</sub> and 1 µL reservoir solution over the reservoir solution (0.1M Bicine pH 7.6 and 25% PEG 6,000, 0.005M hexamine cobalt(II) chloride). Crystals were cryoprotected in artificial reservoir solution containing 15% glycerol, and 15% ethylene glycol, then cryo cooled by plunging into liquid nitrogen. Diffraction data were processed using XDS (11). The structure was determined using NCTC10712<sub>BR</sub> as the search model.

#### *Crystallographic refinement, and analysis*

All models were improved with iterative rounds of model building in Coot (12) and refinement in Phenix (8). In all structures of GspB subdomains, the unliganded structure of Hsa<sub>BR</sub>, and the structure of NCTC<sub>BR</sub>, electron density for hydrogens was observed in later rounds of refinement and riding hydrogens were included in the final model, which reduced the R<sub>free</sub> by over 1% in each case. Bound cations were assigned as either Na<sup>+</sup>, Mg<sup>2+</sup>, or Ca<sup>2+</sup> depending upon the abundance of these ions in either the purification or the crystallization conditions, and the previous observation that cations bound to this site are readily exchanged with cations in the buffer (9). The final models are associated with the statistics listed in **Supporting Tables 1 and 2**. When Ramachandran outliers are associated with the models, these are unambiguously defined by clear electron density.

For sTa-bound Hsa<sub>Siglec+Unique</sub> and GspB<sub>Siglec</sub>, the crystals were isomorphous with unliganded crystals. Accordingly, R<sub>free</sub> reflections were selected as identical. In both cases, unambiguous electron density for all three sugars of sTa was apparent in the initial maps. Ligand occupancies were held at 1.0 during refinement.



### *Sialoglycan binding assays*

DNA encoding wild-type and variant adhesins were cloned into pGEX-3X. Chimeras were designed using an overlay of the coordinates from each adhesin crystal structure. DNA encoding adhesin chimeras were cloned into pGEX-3X. SK678-Hsa chimeras had the Siglec and Unique domains of SK678 and the loops from Hsa. GspB-SK150 chimeras had the Siglec and Unique domains of GspB with selectivity loops of SK150.

The pGEX vectors encode an N-terminal glutathione *S*-transferase (GST) affinity tag, which was used for purification. Individual GST-Siglec+Unique fusions were expressed and purified using glutathione-sepharose, and the binding of biotinylated glycans to immobilized GST-binding regions was performed as described previously (5).

### *Far Western and lectin blotting of human plasma proteins*

Far-western blotting of human plasma proteins using the indicated GST-binding regions (15 nM) as probes was performed as described (13).

### *Interdomain angle calculations*

The torsion angle between Siglec and Unique domains for each system (GspB, SK150, Hsa, SK678, NCTC10712, SrpA) were defined as the angle between the planes formed between center of mass (COM) of Siglec and Residue 1 (R1) and COM of Unique and Residue 2 (R2). The two residues (R1 & R2) were chosen based on crystal structure alignment and are listed in **Table S4**. Missing residues of SK150<sub>Siglec+Unique</sub> were modeled using GspB<sub>Siglec+Unique</sub> as a template (PDB entry 3QC5, (9)).

### *Molecular dynamics (MD) simulations and analyses*

For MD simulations, each system (GspB or Hsa) was solvated in a 10 Å octahedral box of TIP3P (14) water. The Amber16 ff14SB (15) force field was used for the protein. In the first step of the MD simulation, the backbone and side chains of the protein was restrained using 500 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic potentials while the system was energy minimized for 500 steps of steepest descent (16). This step was followed by 500 steps with the conjugate gradient method (17). In a second minimization step, restraints on the protein were removed and 1000 steps of steepest descent minimization were performed followed by 1500 steps of conjugate gradient. The system was then subjected to MD and heated to 300 K with the backbone and side chains of the protein restrained using 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic potentials for 1000 steps. The restraints were released and 1000 MD steps were performed. The SHAKE(18) algorithm was used to constrain all bonds involving hydrogen in the simulations. MD runs (200 ns) were performed at 300 K in the NPT ensemble and a 2 fs time step. The probability distribution analyses and RMSF calculations were performed on 200 ns of 3 independent runs for each system. All analyses were performed using the cpptraj and pytraj (19) python modules of AMBER16.

**Table S1.**

	<b>Hsa<sup>Siglec+Unique</sup></b>	<b>Hsa<sup>Siglec+Unique</sup> + sTa</b>	<b>NCTC10712<sup>Siglec+Unique</sup></b>	<b>SK678<sup>Siglec+Unique</sup></b>
PDB entry	6EFC	6EFD	6EFF	6EFI
DATAID	328	329	509	510
Resolution	1.4 Å	1.85 Å	1.6 Å	1.7 Å
<i>Data collection</i>				
Beamline	APS 21-ID-F	APS 21-ID-G	APS 21-ID-F	SSRL 9-2
Wavelength	0.978 Å	0.978 Å	0.978 Å	0.979 Å
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P1	P2 <sub>1</sub>
Unit cell	a=46.6 Å b=58.1 Å c=76.0 Å	a=46.7 Å b=58.0 Å c=76.1 Å	a=39.8 Å b=48.9 Å c=99.8 Å α=101.8° β=91.4° γ=89.9°	a=59.6 Å b=59.58 Å c=61.8 Å β=100.7°
R <sub>sym</sub>	0.084 (0.650)	0.107 (0.638)	0.075 (0.730)	0.099 (0.530)
R <sub>pim</sub>	0.024 (0.281)	0.037 (0.218)	0.047 (0.479)	0.040 (0.213)
I/σ	49.7 (2.3)	31.7 (2.9)	22.9 (1.9)	15.0 (4.4)
Completeness (%)	93.3% (60.9%)	98.8 % (89.5%)	92.4% (70.9%)	97.7% (97.3%)
Redundancy	12.6 (5.6)	9.5 (8.7)	3.6 (3.4)	7.0 (7.1)
CC <sub>1/2</sub>	0.837	0.940	0.648	0.998
<i>Refinement</i>				
R <sub>cryst</sub>	0.146 <sup>&amp;</sup>	0.196 <sup>&amp;</sup>	0.180	0.177
R <sub>free</sub>	0.179	0.217	0.207	0.210
No. Mol per ASU	1	1	4	2
RMS deviation				
bond lengths	0.01 Å	0.01 Å	0.01 Å	0.01 Å
bond angles	1.6°	0.9°	0.9°	0.7°
Ramachandran				
favored	97.0%	97.1%	96.8%	99.0%
allowed	2.5%	2.9%	3.1%	1.0%
outliers	0.5%*	0.0%*	0.1%*	0.0%

**Table S1. Crystallographic data collection and refinement statistics for Hsa-like adhesins.** Values in parentheses are for the highest resolution shell. Raw data are deposited with SBGrid and can be accessed at: [data.sbgrid.org/dataset/DATAID](http://data.sbgrid.org/dataset/DATAID).

<sup>&</sup>The main chain of Hsa<sup>Siglec+Unique</sup> displays two alternative conformations between residues 378-384 that is accompanied by a *cis-trans* isomerization of the non-proline peptide bond between Hsa<sup>E381</sup> and Hsa<sup>S382</sup> and results in disallowed bond angles for Hsa<sup>S383</sup>. In other isoforms of Hsa, the equivalent residue is a proline. Current technologies do not allow main chain alternative conformations to be refined within the same model. One position is modeled in the unliganded Hsa<sup>Siglec+Unique</sup> structure and one conformation is modeled in the sTa-bound Hsa<sup>Siglec+Unique</sup> structure; however, electron density for both conformations is clearly visible in both structures.

\*The Ramachandran angles identified as outliers (Hsa<sup>S253</sup>, Hsa<sup>L363</sup>, NCTC10712<sup>S253</sup>, NCTC10712<sup>L361</sup>) are equivalent in the homologs and are associated with clear electron density.

**Table S2.**

	<b>GspB<sub>Siglec</sub></b>	<b>GspB<sub>Siglec</sub> + sTa</b>	<b>GspB<sub>Siglec</sub> Form 2</b>	<b>GspB<sub>Siglec+Unique</sub></b>	<b>SK150<sub>Siglec+Unique</sub></b>
PDB entry	6EF7	5IUC	6EF9	6EFA	6EFB
DATAID		507	601	604	508
Resolution	1.03 Å	1.25 Å	1.3 Å	1.6 Å	1.90 Å
<i>Data collection</i>					
Beamline	21-ID-F	21-ID-F	21-ID-G	21-ID-F	Bruker X8R
Wavelength	0.979 Å	0.979 Å	0.979 Å	0.979 Å	1.542 Å
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	R32	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub>
Unit cell	a= 33.9 Å b= 46.2 Å c= 73.0 Å	a=67.7 Å b=66.6 Å c=55.9 Å	a=b=92.1 Å c=143.9 Å	a=33.0 Å b=47.6 Å c=136.2 Å	a=24.3 Å b=62.6 Å c=62.9 Å β=98.6°
R <sub>sym</sub>	0.049 (0.430)	0.066 (0.406)	0.061 (0.771)	0.057 (0.610)	0.139 (0.538)
R <sub>pim</sub>	0.024 (0.233)	0.018 (0.111)	0.017 (0.302)	0.019 (0.213)	0.044 (0.295)
I/σ	25.5 (4.4)	35.6 (7.9)	59.3 (2.8)	43.8 (3.3)	9.3 (1.9)
Completeness (%)	95.4% (90.6%)	95.7% (90.5%)	99.9% (98.0%)	88.9% (48.6%)	97.3% (91.6%)
Redundancy	9.3 (8.2)	14.9 (14.3)	13.8 (7.1)	9.2 (7.7)	9.0 (3.6)
CC <sub>1/2</sub>	0.911	0.964	0.941	0.975	0.996
<i>Refinement</i>					
R <sub>cryst</sub>	0.125	0.156	0.131	0.166	0.172
R <sub>free</sub>	0.141	0.178	0.144	0.209	0.188
No. Mol per ASU	1	2	2	1	1
RMS deviation					
bond lengths	0.01 Å	0.01 Å	0.01 Å	0.02 Å	0.01 Å
bond angles	1.3°	1.5°	1.1°	1.6°	0.7°
Ramachandran					
favored	100%	99.2%	97.0%	98.0%	99.0%
allowed	0%	0.8%	2.2%	1.5%	1.0%
outliers*	0%	0%	0.8%	0.5%	0.0%

**Table S2. Crystallographic data collection and refinement statistics for GspB-like adhesins.** Values in parentheses are for the highest resolution shell. Raw data are deposited with SBGrid and can be accessed at: [data.sbgrid.org/dataset/DATAID](http://data.sbgrid.org/dataset/DATAID)

**Table S3.**

	sTa	sLe <sup>C</sup>	3'sLn	sLe <sup>X</sup>	6S-sLe <sup>X</sup>
<b>Hsa<sub>Siglec+Unique</sub></b>	+++	++	++	+	+
<b>S253G</b>	+++	nd	++	nd	nd
<b>L363G</b>	+++	nd	++	nd	nd
<b>N333P</b>	++	nd	-	nd	nd
<b>G287A/G288P</b>	+++	nd	++	nd	nd
<b>E286R</b>	+++	+++	+++	+++	+++
<b>D356Q</b>	++	++	+++	++	+++
<b>D356R</b>	+	++	+++	+	+++
<b>NCTC10712<sub>Siglec+Unique</sub></b>	+	+	+++	++	+++
<b>E285R</b>	-	-	+	-	+++
<b>Q354D</b>	+	-	+	-	-
<b>+ Hsa CD loop</b>	+	+	+	+	+++
<b>+ Hsa EF loop</b>	-	+	+++	+++	+++
<b>+ Hsa FG loop</b>	++	+	++	+	+
<b>+ all Hsa loops</b>	+++	++	+	-	-
<b>SK678<sub>Siglec+Unique</sub></b>	-	-	++	+	++
<b>E302R</b>	-	-	+	-	+++
<b>Q371D</b>	-	-	+	-	-
<b>+ Hsa CD loop</b>	-	-	-	-	-
<b>+ Hsa EF loop</b>	-	-	++	+	++
<b>+ Hsa FG loop</b>	-	-	+	-	-
<b>+ all Hsa loops</b>	++	+	+	-	-
<b>GspB<sub>Siglec+Unique</sub></b>	+++	-	-	-	nd
<b>L442Y/Y443N</b>	++	+	++	nd	nd
<b>+ SK150 CD loop</b>	-	nd	-	nd	nd
<b>+ SK150 EF loop</b>	+++	nd	-	nd	nd
<b>+ SK150 FG loop</b>	-	nd	-	nd	nd
<b>+ all SK150 loops</b>	-	nd	-	nd	nd
<b>SK150<sub>Siglec+Unique</sub></b>	++	+	+	-	-
<b>Y300L/N301Y</b>	-	nd	-	nd	nd

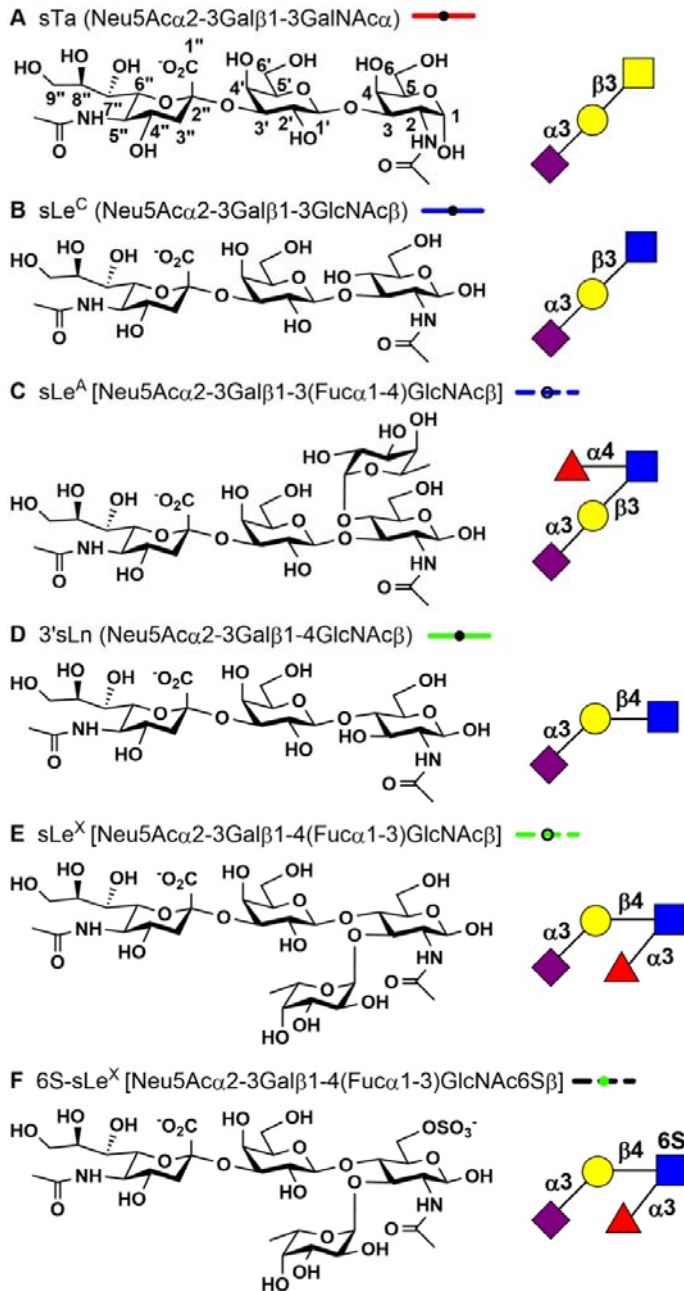
**Table S3. Summary of binding preferences of wild-type and variant SRR adhesins.** sTa, sialyl-T antigen, 3'sLn, 3'-sialyl-N-acetyllactosamine, sLe<sup>C</sup>, sialyl-Lewis<sup>C</sup>, sLe<sup>X</sup>, sialyl-Lewis<sup>X</sup>, 6S-sLe<sup>X</sup>, 6-O-sulfo-sialyl Lewis<sup>X</sup>, nd= not determined

**Table S4.**

<b>Protein</b>	<b>Residues (R1 &amp; R2)</b>
<b>GspB</b>	<b>Glu 417, Asp 557</b>
<b>SK150</b>	<b>Glu 275, Asn 416</b>
<b>Hsa</b>	<b>Glu 262, Asn 411</b>
<b>SK678</b>	<b>Glu 274, Gln 422</b>
<b>NCTC10712</b>	<b>Glu 261, Ser 409</b>
<b>SrpA</b>	<b>Glu 271, Gln 409</b>

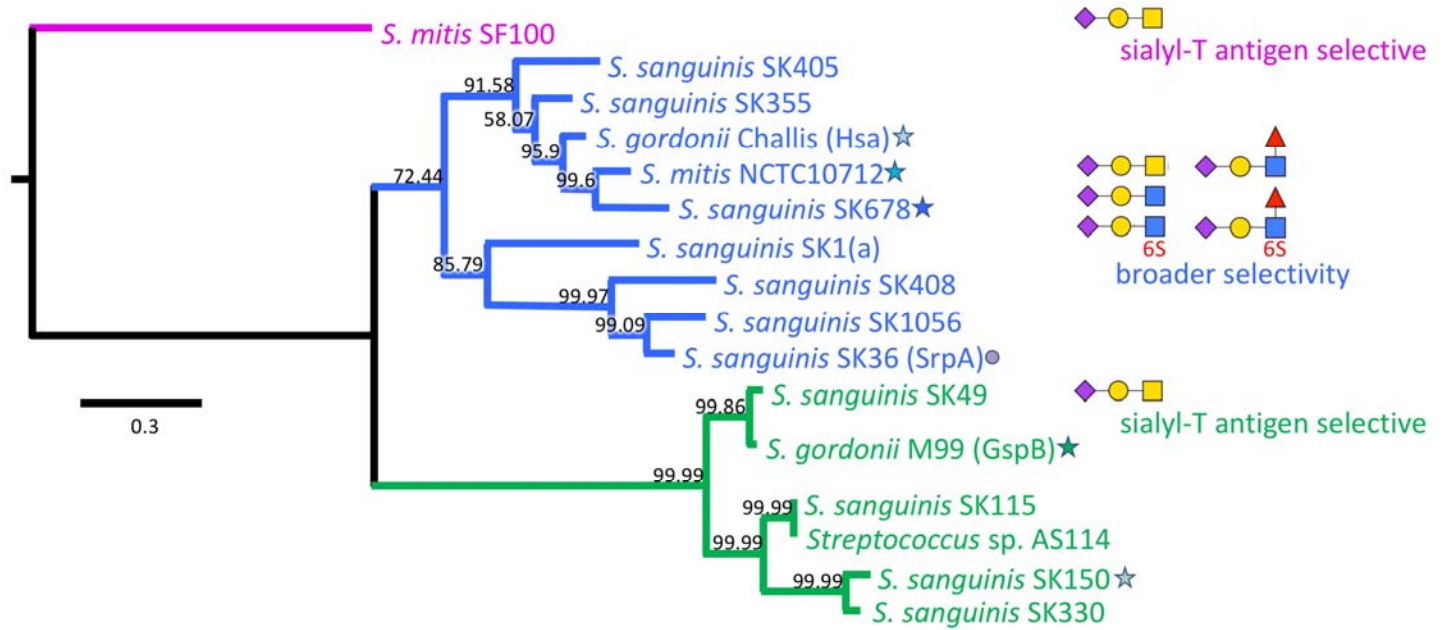
**Table S4. Residues used for interdomain angle calculations.**

Figure S1.



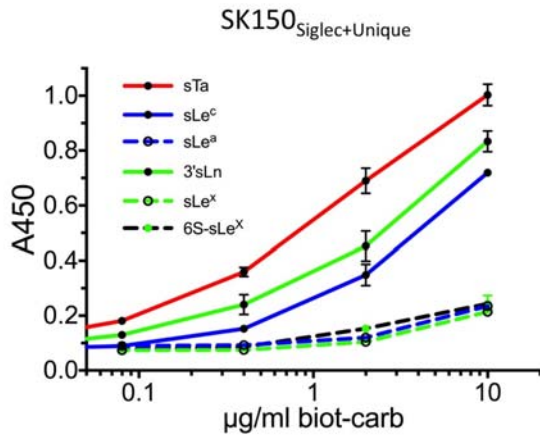
**Figure S1. Sialoglycans used in this study.** The chemical structure of each indicated sialoglycan is shown on the left with the symbolic representation shown on the right. The line style used for all dose response curves is shown to the right of each name.

Figure S2.



**Figure S2. Bacterial Siglec-like SRR adhesins.** Phylogenetic analysis of the tandem Siglec and Unique domains of select bacterial SRR adhesins reveals three distinct subgroups. Characterized Hsa-like adhesins (*blue*) bind to two or more of the indicated sialoglycans; the four characterized GspB-like adhesins (*green*) have narrow selectivity for sialyl-T antigen. The tree is rooted using the distantly-related *S. mitis* SF100 adhesin (*magenta*). Adhesins investigated here are highlighted with a star, and figure panels comparing properties of these adhesins follow this coloring. The structure and ligand binding properties of SrpA, highlighted with a circle, have previously been reported(10, 20), and SrpA is used as a comparator in this report.

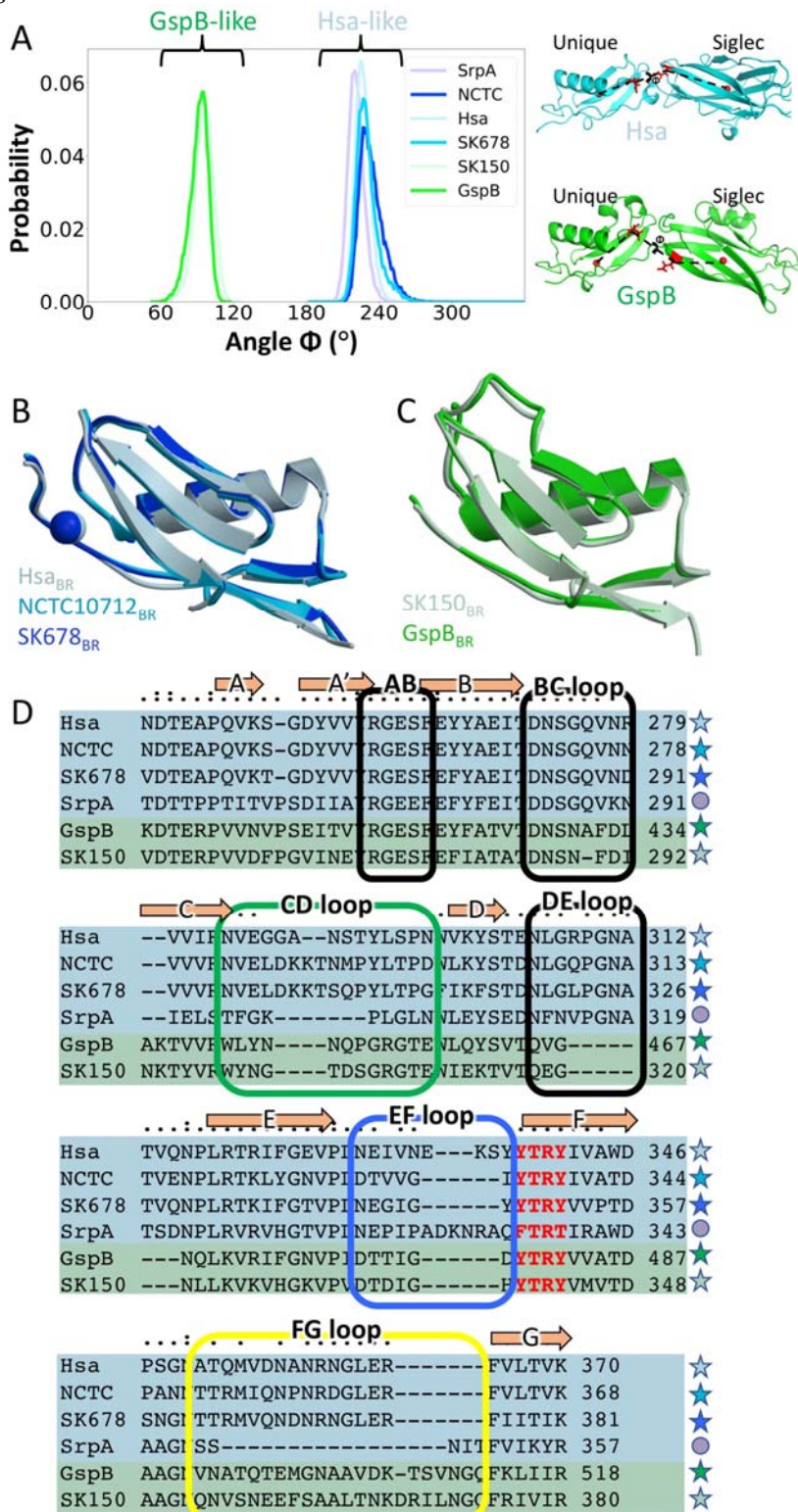
Figure S3.



**Figure S3. Binding of select trisaccharides to GST-SK150<sup>Siglec+Unique</sup>.** GST-SK150<sup>Siglec+Unique</sup> (500 nM) was immobilized in 96-well plates, and biotinylated sialoglycan ligands were added at the indicated concentrations. Binding is reported as the mean  $\pm$  standard deviation, with  $n = 3$ .

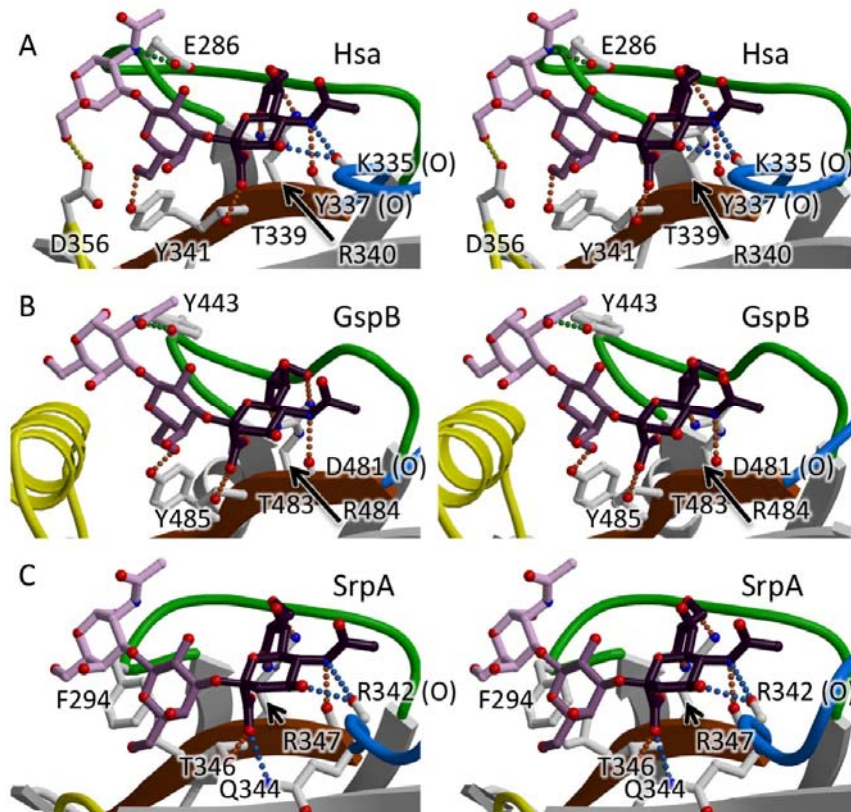


Figure S4.



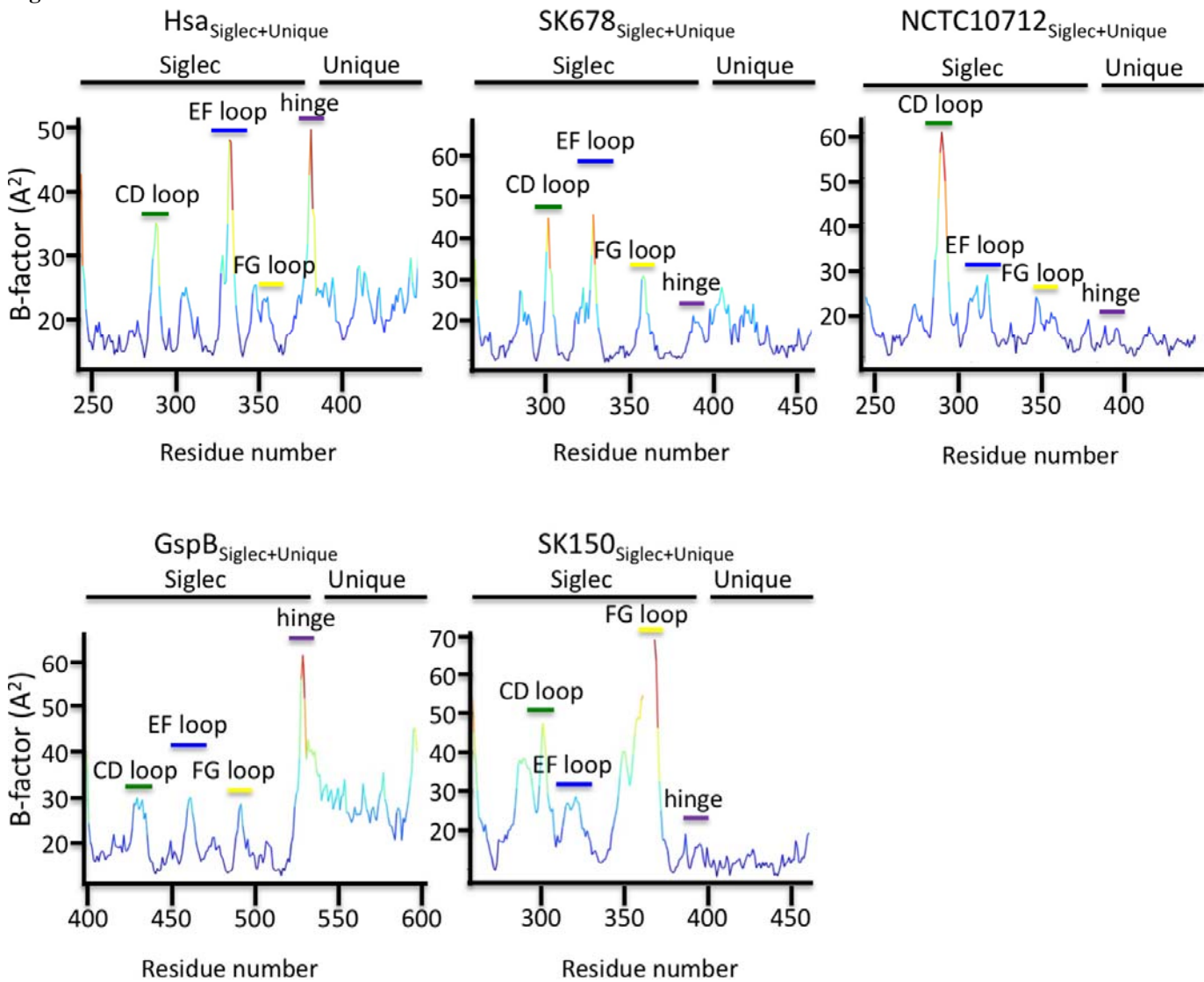
**Figure S4. Comparisons of bacterial SRR adhesins.** A. Probability distributions of the torsion angle ( $\Phi$ ) between the Siglec and Unique domains of GspB, SK150, Hsa, NCTC10712, SK678, SrpA (left) as calculated from MD simulations. Crystal structures showing the  $\Phi$  angle for both GspB-like and Hsa-like proteins (right). The interdomain torsion angles in the corresponding crystal structures are as follows: GspB:  $\sim 100^\circ$ ; SK150:  $\sim 100^\circ$ ; NCTC10712:  $\sim 228^\circ$ ; Hsa:  $\sim 230^\circ$ ; SrpA:  $\sim 216^\circ$ ; SK678:  $\sim 240^\circ$ . B, C. Overlay of the Unique domains of: B. Hsa-like adhesins and C. GspB-like adhesins. The view is rotated as compared to **Fig. 2** in order to highlight the structural similarity between the branches of the phylogenetic tree. D. Sequence alignment of the Siglec domain of SRR adhesins. Hsa-like adhesins are highlighted with a blue background and GspB-like adhesins are highlighted with a green background. Strands conserved in the V-set Ig fold are indicated, and residues of the interstrand loops are highlighted with boxes.

Figure S5.



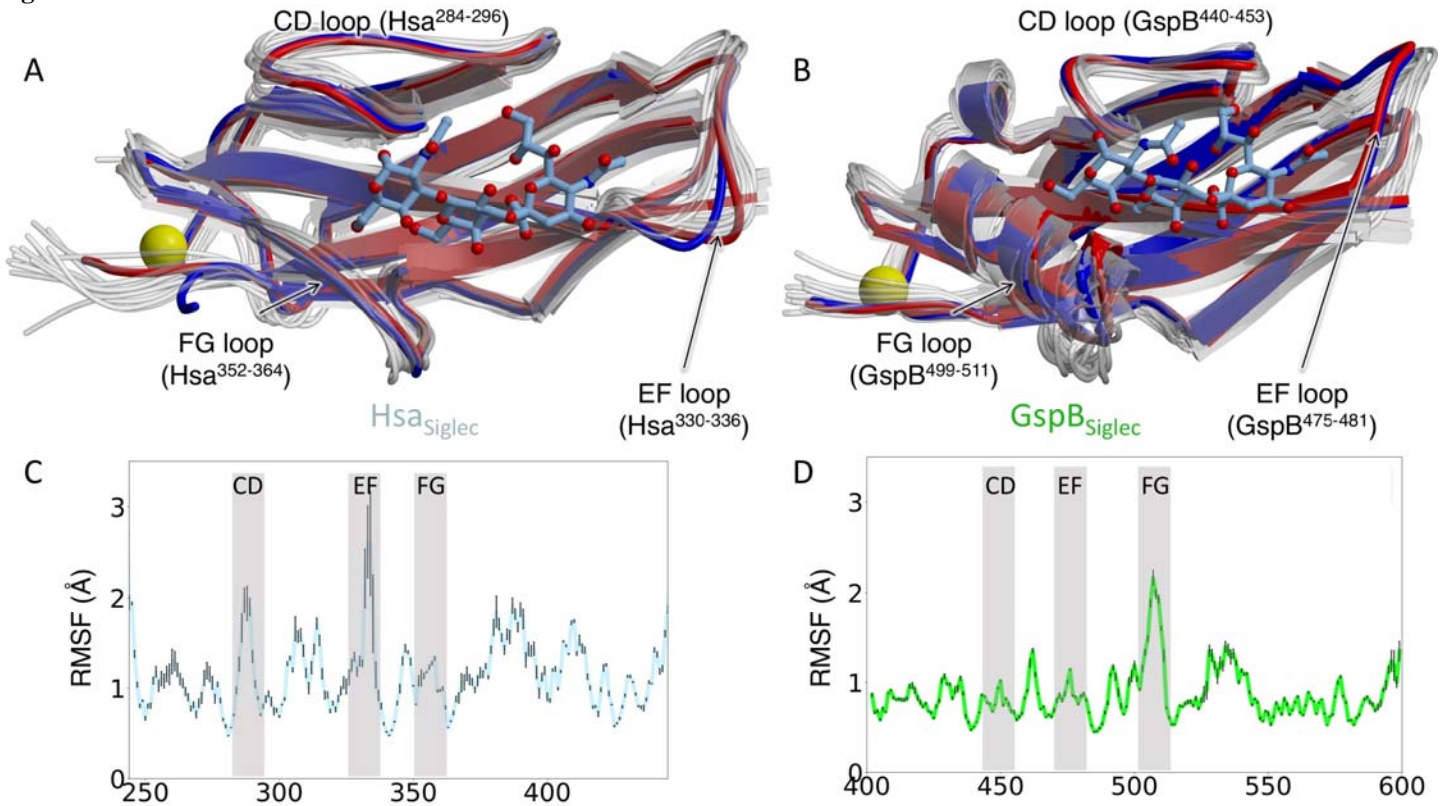
**Figure S5. Stereoviews of the contacts between bacterial SRR adhesins and the sTa ligand.** A. Hsa<sub>Siglec+Unique</sub> bound to sTa. B. GspB<sub>Siglec</sub> bound to sTa. C. SrpA<sub>Siglec+Unique</sub> bound to sTa (PDB entry 5IJ3 (10)). Residues of each respective adhesin within hydrogen-bonding distance of sTa are labeled. Color scheme follows that of Fig. 3 with the CD loop in *green*, the EF loop in *blue*, and the FG loop in *yellow*.

Figure S6.



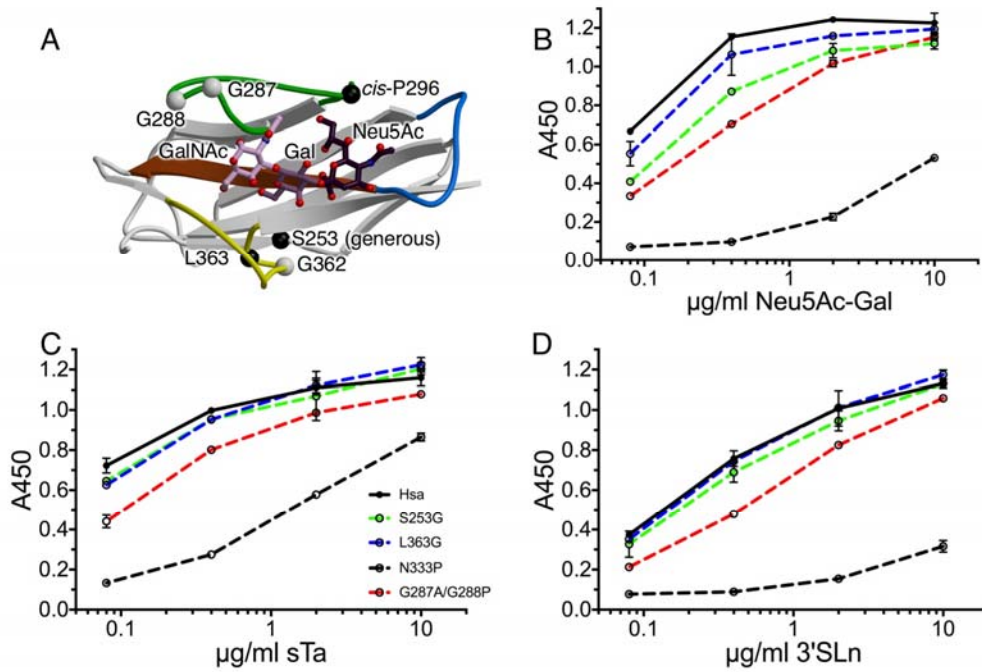
**Figure S6. Temperature factor analysis of adhesins.** For each graph, the residue number is on the x-axis, and the crystallographic temperature factor (B-factor) is on the y-axis. Coloring is by relative B-factor. Regions with the lowest B-factors are predicted to have the lowest mobility (*dark blue*); regions with the highest B-factors are predicted to have the highest mobility (*red*).

Figure S7.



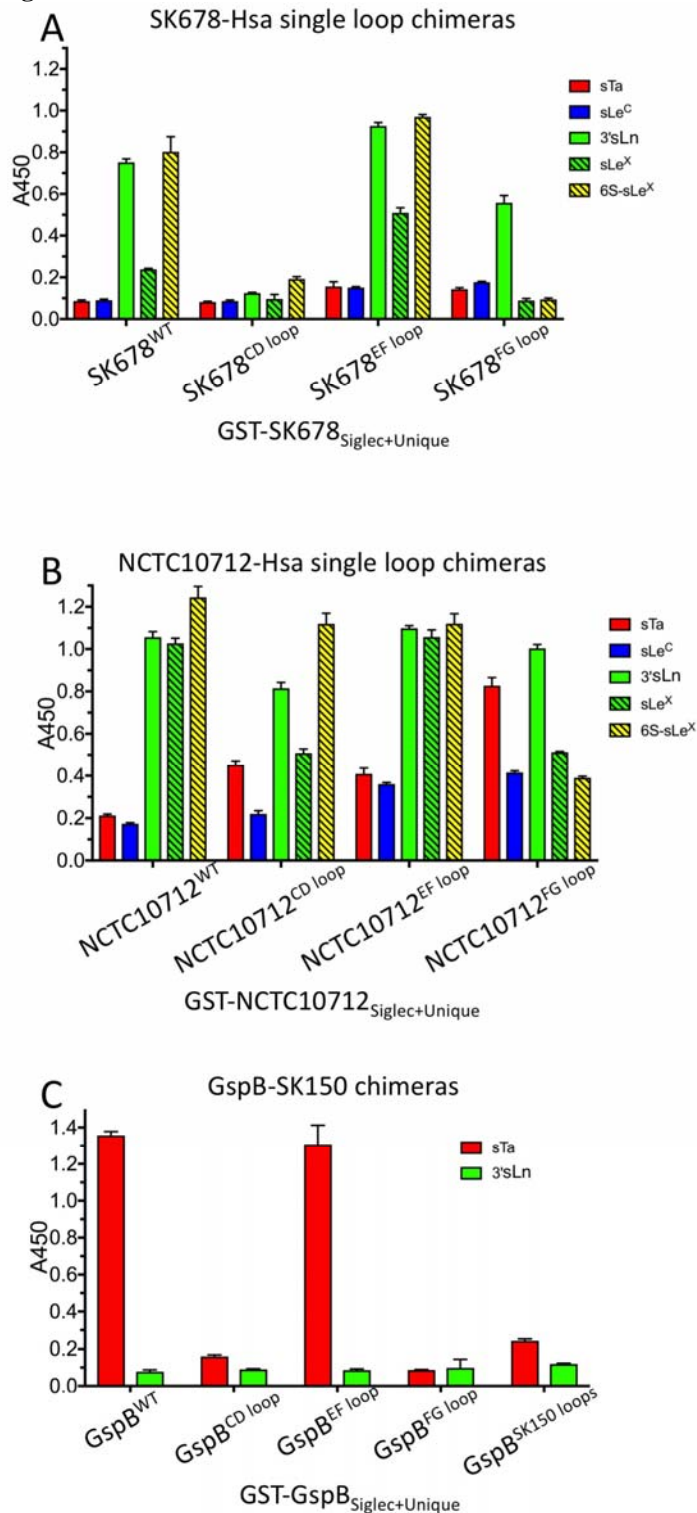
**Figure S7. Conformational selection in SRR adhesins.** A, B. Superposition of a representative subset of MD simulation snapshots (*translucent*) of A. Hsa and B. GspB onto the crystal structures determined in the presence (*blue*) and absence (*red*) of the sTa sialoglycan. MD simulations were performed on the adjacent Siglec and Unique domains; the Siglec domain was resected from the coordinates and is shown in isolation for clarity. MD simulations used structures determined in the absence of ligand as a starting point. C, D. Root mean square fluctuations (RMSF) of the Siglec domain of C. Hsa and D. GspB from the average position of the C $\alpha$  atoms of each residue. Calculations were performed on the adjacent Siglec and Unique domains, with only the resected Siglec domain shown. Error bars correspond to the standard error over 3 independent simulations.

Figure S8.



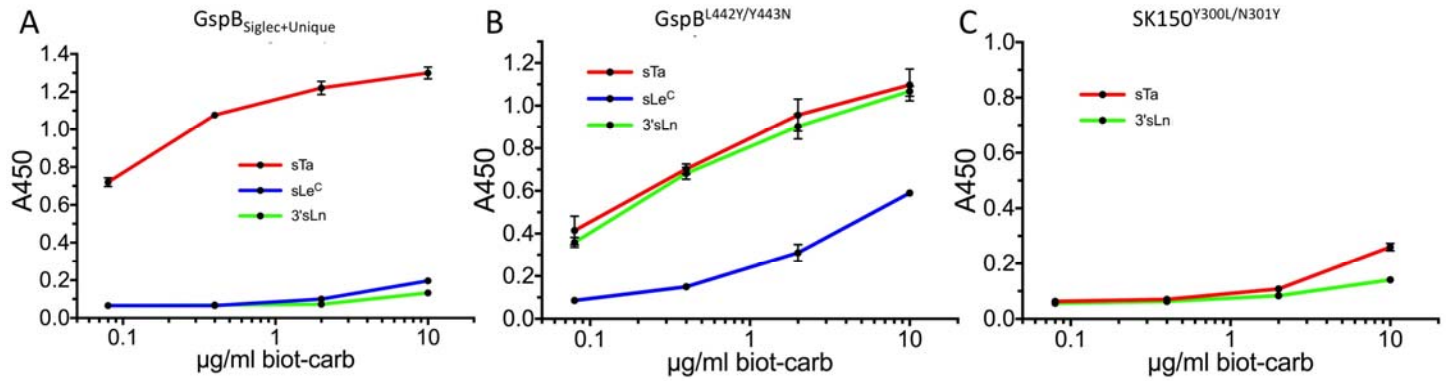
**Figure S8. Impact of flexibility of sialoglycan binding by Hsa<sub>BR</sub>.** A. Locations of variants within the Hsa<sub>Siglec</sub> domain. As a note, Hsa<sup>S253</sup> contains Ramachandran angles in the generously allowed region. B-D. Dose response curves of wild-type and variant GST-Hsa<sub>Siglec+Unique</sub> (500 nM) immobilized in 96-well plates and binding to B. the Neu5Ac-Gal disaccharide, C. sialyl-T antigen, and D. 3'sialyl-*N*-acetylactosamine. Biotinylated sialoglycan ligands were added at the indicated concentrations. Binding is reported as the mean  $\pm$  standard deviation, with  $n = 2$ .

Figure S9.



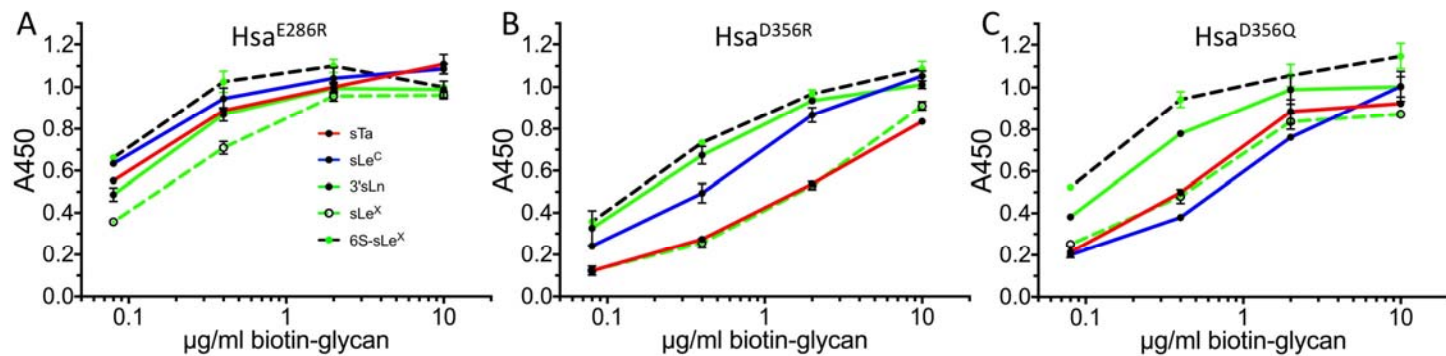
**Figure S9. Chimeras of SRR adhesins.** A. Binding of biotin-glycans (2  $\mu\text{g/ml}$ ) to GST-SK678<sub>Siglec+Unique</sub> containing loops CD, EF, or FG of Hsa, substituted individually. Values correspond to the mean  $\pm$  standard deviation, with  $n = 2$  (wt) or  $n = 3$  (variants). B. Binding of biotin-glycans (2  $\mu\text{g/ml}$ ) to GST-NCTC10712<sub>Siglec+Unique</sub> containing loops CD, EF, or FG of Hsa, substituted individually. Values correspond to the mean  $\pm$  standard deviation, with  $n = 2$  (wt) or  $n = 3$  (variants). C. Binding of biotin-glycans (1  $\mu\text{g/ml}$ ) to GST-GspB<sub>Siglec+Unique</sub> containing loops CD, EF, or FG of SK150 substituted either individually or together. Values correspond to the mean  $\pm$  standard deviation, with  $n = 3$ .

Figure S10.



**Figure S10. Mini-chimeragenesis of the GspB and SK150 adhesins.** Dose-response curves of biotin-glycan binding to immobilized binding regions (500 nM). A. Wild-type GST-GspB<sub>Siglec+Unique</sub> shows a binding preference for sTa. B. Mini-chimeragenesis with the SK150 adhesin was accomplished with the GST-GspB<sup>L442Y/Y443N</sup> double mutant. The mini-chimera becomes more broadly selective by increasing the affinity for 3'sLn and sLe<sup>C</sup>. As a result, it exhibits binding selectivity more similar to wild-type GST-SK150<sub>Siglec+Unique</sub> (see **Fig. S1**). C. The converse mini-chimeragenesis of SK150<sub>Siglec+Unique</sub> exhibited reduced binding for sialoglycan ligands that bind most avidly to both wild-type GspB<sub>Siglec+Unique</sub> and SK150<sub>Siglec+Unique</sub>.

Figure S11.



**Figure S11. Binding characteristics of select GST-fused Hsa<sub>Siglec+Unique</sub> variants.** Dose-response curves of biotin-glycan binding to immobilized binding regions (500 nM). A. Hsa<sup>E286R</sup>, B. Hsa<sup>D356R</sup>, C. Hsa<sup>D356Q</sup>. Values are reported as the mean  $\pm$  standard deviation, with  $n = 2$ .



## Supporting References

1. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
2. Kearse M, *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.
3. Abascal F, Zardoya R, & Posada D (2005) ProfTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104-2105.
4. Ronquist F, *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539-542.
5. Bensing BA, *et al.* (2016) Novel aspects of sialoglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. *Glycobiology*.
6. Otwinowski Z & Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology, Macromolecular Crystallography, part A*, eds Carter C, Jr & Sweet R (Academic Press, NewYork), Vol 276, pp 307-326.
7. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
8. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
9. Pyburn TM, *et al.* (2011) A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. *PLoS Pathog* 7(7):e1002112.
10. Bensing BA, *et al.* (2016) Structural Basis for Sialoglycan Binding by the Streptococcus sanguinis SrpA Adhesin. *J Biol Chem* 291(14):7230-7240.
11. Kabsch W (2010) Xds. *Acta Crystallographica Section D-Biological Crystallography* 66:125-132.
12. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
13. Bensing BA, Li Q, Park D, Lebrilla CB, & Sullam PM (2018) Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: implications for infective endocarditis. *Glycobiology* 28(8):601-611.
14. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* 79:926-935.
15. Maier JA, *et al.* (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.
16. Arfken GB, Weber HJ, & Harris FE (2011) *Mathematical methods for physicists: a comprehensive guide* (Academic press).
17. Case DA, *et al.* (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26(16):1668-1688.
18. Jean-paul Ryckaert GC, Herman J. C. Berendsen (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*
19. Roe DR & Thomas E. Cheatham I (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.
20. Loukachevitch LV, *et al.* (2016) Structures of the Streptococcus sanguinis SrpA Binding Region with Human Sialoglycans Suggest Features of the Physiological Ligand. *Biochemistry* 55(42):5927-5937.