Origins of glycan selectivity in streptococcal Siglec-like adhesins suggest mechanisms of receptor adaptation

Barbara A. Bensing1,2,16, Haley E. Stubbs3,16, Rupesh Agarwal4,5, Izumi Yamakawa6,14, Kelvin Luong6, Kemal Solakyildirim7,8, Hai Yu8, Azadeh Hadadianpour9, Manuel A. Castro10, Kevin P. Fialkowski6,15, KeAndreya M. Morrison11, Zdzislaw Wawrzak12, Xi Chen8, Carlito B. Lebrilla8, Jerome Baudry13, Jeremy C. Smith4,5, Paul M. Sullam1,2 & T. M. Iverson6,10

Bacterial binding to host receptors underlies both commensalism and pathogenesis. Many streptococci adhere to protein-attached carbohydrates expressed on cell surfaces using Siglec-like binding regions (SLBRs). The precise glycan repertoire recognized may dictate whether the organism is a strict commensal versus a pathogen. However, it is currently not clear what drives receptor selectivity. Here, we use five representative SLBRs and identify regions of the receptor binding site that are hypervariable in sequence and structure. We show that these regions control the identity of the preferred carbohydrate ligand using chimeragenesis and single amino acid substitutions. We further evaluate how the identity of the preferred ligand affects the interaction with glycoprotein receptors in human saliva and plasma samples. As point mutations can change the preferred human receptor, these studies suggest how streptococci may adapt to changes in the environmental glycan repertoire.

1 Division of Infectious Diseases, Veterans Affairs Medical Center, Department of Medicine, University of California, San Francisco, CA, USA. 2 the Northern California Institute for Research and Education, San Francisco, CA 94121, USA. 3 Graduate Program in Chemical and Physical Biology, Vanderbilt University, Nashville, TN 37232, USA. 4 University of Tennessee/Oak Ridge National Laboratory, Center for Molecular Biophysics, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6309, USA. 5 Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA. 6 Department of Pharmacology, Vanderbilt University, Nashville, TN 37232, USA. 7 Department of Chemistry, Erzincan Binali Yildirim University, Erzincan 24100, Turkey. 8 Department of Chemistry, University of California, Davis, CA 95616, USA. 9 Department of Microbiology, Pathology, and Immunology, Vanderbilt University, Nashville, TN 37232, USA. 10 Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA. 11 Department of Pharmacology, School of Graduate Studies and Research, Meharry Medical College, Nashville, TN 37208, USA. 12 LS-CAT Synchrotron Research Center, Northwestern University, Argonne, IL 60439, USA. 13 Department of Biological Sciences, The University of Alabama in Huntsville, Huntsville, AL 35899, USA. 14 Present address: School of Nursing, Belmont University, Nashville, TN 37212, USA. 15 Present address: College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA. 16 These authors contributed equally: Barbara A. Bensing, Haley E. Stubbs. ✉email: tina.iverson@vanderbilt.edu

https://doi.org/10.1038/s41467-022-30509-y
Selection among many possible host receptors determines whether a bacterium can adhere to a preferred anatomical niche or can infect a particular host.1 Streptococci and staphylococci are among the organisms that use host-associated carbohydrates as receptors; they may specifically bind to sialic acid-containing glycans (sialoglycans; Fig. 1). As an example, human O-linked glycosylated proteins commonly contain a terminal α2-3-linked sialic acid-galactose disaccharide, (Neu5Acα2-3Gal). Additional forms of sialic acid and alternative linkages are found in animal sialoglycans.3,4

Neu5Acα2-3Gal is present on the human salivary mucin MUC7,5-7, on several glycoproteins in blood plasma,8 and on surface platelet proteins.9,10 Bacterial binding to glycoproteins terminating with Neu5Acα2-3Gal may therefore allow colonization of the oral cavity as a commensal. In animal models, sialoglycan binding is also implicated in the persistence of these organisms in the bloodstream as an endovascular pathogen11 although it is not known whether all streptococci can act as pathogens.

Siglec-like binding regions (SLBRs) are among the streptococcal adhesive modules that bind sialoglycans. SLBRs are usually found within the context of serine-rich repeat proteins, which form fibrils extending from the bacterial surface. SLBRs contain two adjacent modules: a Siglec domain, which shares some features with mammalian Siglecs, and a Unique domain13 with no close homologs outside of the family. The Siglec domain contains a TRX sequence motif15 that recognizes Neu5Acα2-3Gal in the context of larger glycans. Reported mutagenesis of the TRX motif demonstrates its importance in sialoglycan binding5,16 and in endovascular disease in animal models.13 This has motivated the development of compounds that bind the TRX motif as a potential therapy for human endovascular infections caused by these organisms.17,18

SLBRs display a range of selectivity. Some SLBRs bind selectively to the α2-3-linked trisaccharide sialyl-T antigen (sTa, Neu5Acα2-3Galβ1-3GalNAc; Fig. 1a)5,19. Others have intermediate selectivity and bind to a small number of closely related glycans.8,19 Still others can bind to a broad range of sialoglycans and do not distinguish between related structures.5,19 The binding profile of these SLBRs is likely adapted to the host display of sialoglycans. In the oral cavity for example, the display of sialylated O-glycans on MUC7 varies between individuals, making it possible that the binding preferences of the SLBRs reflect the specific glycosylation display of an individual.5-7,20,21 The binding profile can also affect virulence; streptococci containing SLBRs that preferentially bind to sTa in vitro exhibit higher pathogenicity in an animal model of endocarditis.22

Despite the importance of the sialoglycan binding profile in the interaction between streptococci and host,22 the sequence determinants that underlie glycans selectivity are not currently clear. Here, we determine the molecular basis for glycan selectivity of a phylogenetically-informed library of SLBRs. We test our predictions for selectivity by engineering the binding spectrum of selected SLBRs and assessing host receptor switching in human saliva and plasma glycoproteins. Collectively, these studies improve our understanding of the glycan selectivity that underlies commensalism and pathogenesis. In addition, they suggest how these bacteria may adapt to host sialoglycan repertoires.

Results

Selection of SLBRs for study. Starting with SLBRs with at least some previously-reported selectivity, we correlated selectivity with phylogeny (Fig. 2)5,8,19,23. Our initial trees contained two major branches. This identified that evolutionary relatedness of SLBRs is a moderate, but not strong, predictor of glycan selectivity. Most SLBRs of the first major branch of the tree (blue in Fig. 2) are broadly-selective and recognize two or more related tri-, tetra-, or hexasaccharides (see examples in Fig. 1). However, sequence similarity does not clearly predict the preferred glycan5,8,19,23. In contrast, characterized SLBRs of the second major branch (green in Fig. 2) are selective for sTa (Fig. 1a)5,8,19,23.

To understand selectivity of these SLBRs for human glycans, we chose comparators from each branch for detailed study. From the first branch of the tree (blue in Fig. 2), we selected the SLBRs of the Hsa adhesin from S. gordonii strain Challis (termed SLBRGspB) and the equivalent SLBRs from Streptococcus sanguinis strain SK678 (SLBRK678) and Streptococcus gordonii strain UB10712 (SLBRUB10712; see footnote). Although these three SLBRs are >80% identical in amino acid sequence, when they were tested with arrays containing 49 sialoglycans, they exhibited distinct binding profiles5,19. SLBRGspB was quite broadly selective and bound to a range of α2-3-linked sialoglycans, but not to the corresponding fucosylated derivatives.5,19 In comparison, SLBRUB10712 and SLBRK678 were more narrowly selective, although both bound to multiple sialoglycan ligands. Specifically, SLBRUB10712 bound strongly to 3'-sialyl-N-acetyllactosamine (3'SLN; Neu5Acα2-3Galβ1-4GlcNAcβ); Fig. 1b) and a small range of related structures5,19, while SLBRK678 bound to only two of the glycans on this array, 3'SLN and 6-O-sulfo-sialyl Lewis X (6'SLeX; Neu5Acα2-3Galβ1-4(Fucα1-3-GlcNAcβ); Fig. 1c)5. In summary, all three of these SLBRs bind multiple ligands with promiscuity following SLBRGspB > SLBRUB10712 > SLBRK678.

The second major branch of the evolutionary tree (green in Fig. 2) includes the well-characterized SLBRGspB from S. gordonii strain M979,13,24,25. SLBRGspB exhibits narrow specificity for the sTa trisaccharide, as have other previously-characterized members of this evolutionary branch5,19,23,24. The binding results for GST-SLBRGspB with sTa, 3'SLN, and sialyl LewisC (sLeC; Neu5Acα2-3Galβ1-4GlcNAcβ) (Fig. 1d) were recapitulated here by ELISA showing concentration-dependent binding (Supplementary Fig. 1a).

In seeking comparators of SLBRGspB, we evaluated close homologs for their binding spectrum. We identified a previously-uncharacterized SLBR from Streptococcus sanguinis strain SK150 (termed SLBRK150) displays 62% identity to SLBRGspB but exhibits a distinct binding profile (Supplementary Fig. 1b). In short, there was modest binding to each of the three trisaccharides, i.e., sTa, 3'SLN, and sLeC; but little detectable binding to any of the tetrasaccharides (i.e., 6'SLeX; Fig. 1c), sialyl Lewis X (6'SLeX; Neu5Acα2-3Galβ1-4(Fucα1-3-GlcNAcβ; Fig. 1d), and sialyl LewisA (sLeA; Neu5Acα2-3Galβ1-3(Fucα1-4-GlcNAcβ; Fig. 1e)) (Supplementary Fig. 1b). The high sequence similarity and distinct binding properties of SLBRGspB and SLBRK150 make these good comparators for understanding selectivity.

Structural basis for recognition of sialoglycan elaborations. To reveal how similar SLBRs could include or exclude different sialoglycans, we determined crystal structures of these five SLBRs at resolutions between 1.0 Å and 1.7 Å (Supplementary Tables 1, 2, Fig. 3, and Supplementary Fig. 2). This included a structure of SLBRGspB with improved resolution as compared to a previous report.19 In each structure, the N-terminal Siglec domain is organized around a V-set variation of the Ig fold (Fig. 3), named for its discovery in antibody variable domains.26 The C-terminal Unique domain of the SLBRs displays a fold that has only been observed in other members of this family (Supplementary Fig. 2).

We next evaluated how these SLBRs interact with preferred versus disfavored ligands. Only the crystallization conditions for SLBRGspB and the isolated Siglec domain of SLBRGspB...
Fig. 1 Sialoglycans used in this study. a sialyl-T antigen, b sialyllactosamine, c 6S-sialyl Lewis X, d sialyl Lewis C, e sialyl Lewis X, f sialyl Lewis A, g 6S-sialyllactosamine. 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.
The sialoglycan-bound structures of SLBR<sub>Hsa</sub> and SLBR<sub>GapB-Siglec</sub> identifies that the sialic acid of all glycans binds above the Φ<sub>TRX</sub> motif in a similar way. This suggests that while the Φ<sub>TRX</sub> motif is important for binding, it does not select between potential ligands. More careful comparison suggests that the distinct selectivity may originate from three loops of the V-set Ig fold that surround the sialoglycan binding site: the CD loop (SLBR<sub>Hsa</sub>284–296 or SLBR<sub>GapB</sub>440–453), the EF loop (SLBR<sub>Hsa</sub>330–336 or SLBR<sub>GapB</sub>475–481), and the FG loop (SLBR<sub>Hsa</sub>352–364 or SLBR<sub>GapB</sub>499–511) (Fig. 4 and Supplementary Fig. 8). Variation of both sequence and structure of SLBRs disproportionately maps to these loops (Supplementary Figs. 8 and 9). Moreover, temperature factor analysis suggests that these loops have high flexibility in the absence of ligand (Supplementary Fig. 10). Finally, molecular dynamics (MD) simulations of unliganded SLBR<sub>Hsa</sub> and SLBR<sub>GapB</sub> predict that these
loops exhibit considerably more flexibility than other parts of the protein (Fig. 5a and Supplementary Fig. 11). The MD further suggests that these loops sample the ligand-bound form even in the absence of sialoglycan. This supports a conformational selection mechanism, where structural change of the protein precedes binding in enzymes affects fidelity and may similarly contribute to ligand selectivity in binding proteins.

Distinct loops in SLBRHsa and SLBRGspB showed the largest conformational differences between the unbound and sialoglycan-bound structures. This provides the first hints into how narrow- versus broad-selectivity is conferred in this family. In the sTa-bound structure of SLBRGspB-Siglec, the helix of the FG loop is rotated 10° as compared to the unliganded conformation. This rotation results in a maximum physical displacement of 1.3 Å (Fig. 5b), which optimizes contacts to the GalNAc of sTa. Mechanistically, this would be consistent with the conserved region of the glycan first interacting with a relatively pre-formed binding pocket comprised of the CD and EF loops prior to interaction with the FG loop.

In SLBRHsa, the conformation of the FG loop is similar in the presence and absence of glycan. Instead, comparing costructure determined with sTa with the costructure determined with sLeC shows that the position of the EF loop differs by 5.9 Å (Fig. 5c). This allows the SLBRHsa backbone carbonyl to form hydrogen-bonding interactions to the invariant portion of the sialoglycans, i.e., the terminal Neu5Acα2-3Gal. In costructures determined with lower-affinity ligands, i.e., 3’sLn or 6S-sLeX, this loop is not associated with clear electron density. This may result from crystal contacts to the EF loop that stabilize its position in the unliganded pose.

Fig. 4 Sialoglycans bound to SLBRHsa and SLBRGspB. SLBRHsa bound to sialoglycans a sTa, b sLeC, c 3’sLn and d 6S-sLeX. e SLBRGspB-Siglec bound to sTa. In each panel, the SLBR is shown as a cartoon with the CD, EF, and FG selectivity loops colored in green, blue, and yellow respectively. The F strand contains the conserved YTRY motif and is shown in cyan. Ions are shown in yellow spheres. Carbon atoms of each sialoglycan are colored in blue, and oxygen in red. |F_r| − |F_s| difference electron density calculated after removing the sialoglycan and performing three rounds of refinement in Phenix are shown in gray mesh and contoured at 3 σ. The standard depiction for each carbohydrate is shown in the upper left, with linkages indicated.

Fig. 5 Conformations associated with SLBRs bound to sialoglycans. a Probability of distance distribution between the position of the Neu5Ac O4-hydroxyl in sTa and the SLBRHsa K335 backbone carbonyl, as calculated by MD simulations. A bimodal distribution of distances exhibit maxima at 7.5 Å (red arrow), which reflects the unliganded crystal structure, and at 3.5 Å (navy arrow), which approaches the liganded crystal structure. The formation of the hydrogen-bond between the SLBRHsa K335 carbonyl and Neu5Ac likely shifts the conformational equilibrium to a pose that supports the 2.9 Å distance (light green arrow) observed in the bound state. b The FG loop of SLBRGspB-Siglec rotates 10° upon sTa binding. The position in the unbound structure is shown in yellow and the position in the bound structure is shown in light green. c The EF loop of SLBRHsa adjusts to promote formation of hydrogen-bonding interactions between SLBRHsa K335 and the Neu5Ac of sTa. The position of this loop in the unbound structure is shown in blue, and the position occupied in the bound structure is shown in light green. The distance between the SLBRHsa K335 backbone carbonyl and the position of the Neu5Ac O4-hydroxyl of the unliganded state are shown in red lines and match the 7.5 Å distance calculated by MD simulations (panel a). The distance between the SLBRHsa K335 backbone carbonyl and the position of the Neu5Ac O4-hydroxyl is shown in light green dots.
resulting in a mixture of open and closed conformations (Supplementary Fig. 12). Comparison of the EF loop positions in the various crystal structures (Figs. 3b, 4, and Supplementary Fig. 13a) with the positions calculated by the MD simulations (Fig. 5c and Supplementary Fig. 11) suggests that closed conformation of the EF loop in the sTa and 3'sLn-bound crystal structures is likely the lowest energy state (Supplementary Fig. 11). Mechanistically, this suggests that for SLBRSK678, the variable, sub-terminal region of a sialoglycan ligand would first interact with the CD and FG loops. The ligand could then adjust in global position to optimize hydrogen-bonding interactions. The flexibility of the EF loop could then adapt to a range of different orientations of bound sialoglycan. This would be expected to promote broad selectivity. Thus, the location of inherent protein flexibility may define whether an SLBR is narrowly- versus broadly-selective.

To further evaluate how the broadly-selective SLBRHsa could select for particular sialoglycans, we compared the binding positions of strong, intermediate, and weak ligands (Supplementary Fig. 13a). In the strong and intermediate ligands, the invariant Neu5Acα2-3 Gal effectively superimposes (Supplementary Fig. 13a, b) and has similar hydrogen bonds. Differences in the SLBR-ligand interactions predominantly map to the variable third sugar of the glycan (Supplementary Fig. 13c–f). Biding strength may therefore be related to these interactions. In contrast, the global binding position of the weak ligand 6S-sLeX is shifted as compared with all other ligands (Fig. 4d and Supplementary Fig. 13b, f). This affects the hydrogen bonds along the entirety of the ligand.

6S-sLeX is both α1,3-fucosylated and O-sulfated at the C6 (6S) of the GlcNAc, modifications that are absent in the strong SLBRHsa ligands (Fig. 6c). The evaluation of the interactions between these groups and SLBRSK678Hsa suggests how related SLBRs include or exclude these elaborations. In considering how the α1,3-fucose in glycans such as sLeX and 6S-sLeX is excluded from SLBRSK678Hsa, our analysis suggests that the β-branching of SLBRSK678HsaD356 on the FG loop favors the binding of a fucosylated glycan (Supplementary Fig. 13c–f). MD simulations also indicate that the FG loop does not sample a position that allows an extra fucose or other large elaboration at this position (Supplementary Fig. 11). This is consistent with the crystal structure, which shows that the loop position does not allow 6S-sLeX to sit optimally in the sialoglycan binding site.

In considering how a 6 S group might be included or excluded, the structure reveals that SLBRSK678HsaE286 of the CD loop contacts the sulfate of 6S-sLeX. This does not exclude a 6 S group per se, but both are negatively charged. The structure suggests that an unknown ligand, possibly a component of the buffer, binds near this site to bridge the interaction (Fig. 4d and Supplementary Fig. 13f). Taken together, these structural and computational analyses show that steric and electrostatic interactions of the broadly selective SLBRHsa exclude specific structural additions to the glycan ligands.

**The CD, EF, and FG loops determine SLBR selectivity.** Because structural studies suggest that the combined action of the CD, EF, and FG loops allow SLBRs to select between ligands, we developed chimeras with the backbone of one SLBR and the loops of a closely-related SLBR. We first replaced the CD, EF, and FG loops of SLBRGspB and SLBRGspB with the equivalent loops from SLBRSK678 and SLBRSK678Hsa to create the SLBRSK678Hsa-CD-loop and SLBRSK678Hsa-FG-loop chimeras. MD simulations would suggest that the loops retain the structure found within the parent SLBRSK678Hsa (Supplementary Fig. 14). Using physiologically-relevant sialoglycans, we measured binding to parent and chimeric SLBRs in ELISAs (Fig. 6a–e). We found that these chimeras bind glycans strongly and had a sialoglycan-binding preference that closely resembled SLBRSK678Hsa rather than the parent SLBR (Fig. 6f and Supplementary Table 4). This change in selectivity occurred via both a gain-of-function that promoted binding to sTa and a loss-of-function that decreased binding to α1,3-fucosylated and O-sulfated sialoglycans. This change of binding spectrum confirms that a major determinant of selectivity in these SLBRs is the loops that surround the ligand-binding pocket.

We next assessed the contributions of each loop to selectivity (Supplementary Fig. 15). Substitution of the EF loop of SLBRGspB with the EF loop from SLBRSK678 resulted in decreased binding to sTa, sLeX, sLeX, and 6S-sLeX (Supplementary Fig. 15). This result is consistent with the structural prediction that a SLBR with a flexible EF loop can potentially accommodate a greater range of ligands.

In contrast, substitution of the CD or FG loops altered the identity of the preferred ligands. The altered selectivity of these chimeras involved a combination of reduced binding to some sialoglycans and increased binding to others, i.e., both a loss-of-function and a gain-of-function. For example, both SLBRGspB-CD-loop and SLBRSK678-CD-loop exhibited decreased binding to the fucosylated ligands sLeX and 6S-sLeX while SLBRSK678-CD-loop also increased binding to sTa (Supplementary Fig. 15a, b). This is consistent with the crystallographic interpretation that SLBRSK678Hsa-D356 on the FG loop restricts accommodation of Fucα1-3GlcNAc.

The single-loop chimeras also suggest synergy between these three selectivity loops. For example, the substantial enhancement in binding of SLBRSK678Hsa-CD-loop to 6S-sLeX (Supplementary Fig. 15b) is consistent with a proposal that the binding of 6S-ligands is controlled by the CD loop. However, the SLBRSK678Hsa-FG-loop chimaera retains robust binding to 6S-sLeX (Supplementary Fig. 15a) suggesting that the other loops moderate the effects.

We next turned to SLBRGspB and SLBRGspB, which both bind sTa preferentially (Supplementary Fig. 1). Here, we substituted the loops of SLBRSK150 into SLBRGspB and assessed the binding to sTa and 3'sLn, which are the ligands with the highest affinity for SLBRSK150. In contrast to the results observed with SLBRSK678Hsa and its close homologs, substitution of the EF loop of SLBRSK150 into SLBRGspB had little impact (Supplementary Fig. 15c). In all remaining chimeras, there was little detectable binding to sTa or 3'sLn (Supplementary Fig. 15c). To determine whether protein misfolding may be a contributing factor in variants with loss of binding, we used size exclusion chromatography (Supplementary Fig. 16a–c), which can distinguish between folded and mis-folded SLBRs.23 The chromatogram of the SLBRGspBSK150-loops showed a monodisperse peak with little aggregation, indicating that loss of binding in this case was not due to misfolding. However, the chromatograms of the SLBRGspBSK150-CD-loops and SLBRGspBSK150-FG-loops chimeras showed significant levels of protein aggregates and breakdown products, indicating that misfolding may contribute to loss of binding for these two variants.

The ability to develop functional chimeras for the three SLBRSK678Hsa-like adhesins, but not the two SLBRGspB-like adhesins, might be explained in several ways. First, the broadly-selective scaffolds of SLBRSK678Hsa and SLBRSK150 may have more plasticity, allowing these to better accommodate non-native loops. Conversely, the broadly-selective SLBRs may contain somewhat more flexible loops that more easily adjust to the non-native scaffold. Finally, the sequence identity between SLBRSK678Hsa and SLBRSK150 is higher than that between SLBRGspB and SLBRSK150, allowing a better fit between the scaffold and chimeric loops in the SLBRSK678Hsa-like proteins. To better understand why SLBRSK678Hsa-like proteins were more mutable, we leveraged our crystal structure of SLBRGspB in complex with sTa (Fig. 4e) and identified that SLBRGspB,442 and SLBRGspB,443 closely approach the GalNAc...
We engineered SLBRGspB-SK150 mini-chimeras that swapped single amino acids at these positions with the equivalent residues from SLBRSK150. We then measured binding to sTa, 3'Sln, and sLeC (Supplementary Fig. 17c–f). The SLBRGspB/Y442Y/Y443N mini-chimera had increased binding to 3'Sln and sLeC and was overall more similar in selectivity to SLBRSK150 than to SLBRGspB (compare Supplementary Fig. 17c and Supplementary Fig. 1); however, the converse SLBRSK150/Y300I/N301Y mini-chimera still exhibited reduced binding (Supplementary Fig. 17d) and a size exclusion profile that suggested the presence of breakdown products, indicating that misfolding likely contributes to loss of binding for this variant (Supplementary Fig. 16d). The incomplete success of the mini-chimeras suggests complex origins for the inability to change selectivity in SLBRGspB and SLBRSK150 via mutagenesis.

In summary, the SLBRs from the two branches of the evolutionary tree respond differently to chimeragenesis. The parent SLBRGspB and SLBRSK150 cannot easily undergo alteration of their binding spectrum and tend to exhibit lower stability (Supplementary Fig. 16a–d) and loss of function (Supplementary Fig. 17c–f). In contrast, SLBRHsa, SLBRSK678, and SLBRUB10712 readily tolerate changes in binding spectrum via chimeragenesis to allow strong binding of alternative ligands (Supplementary Table 4).

Identification of selectivity-dictating residues. The identification of the CD, EF, and FG loops as the regions that are of largest natural sequence variation (Supplementary Fig. 4) and as regions that may control glycan selectivity (Fig. 6 and Supplementary Fig. 15) could suggest that these evolved to allow for binding to different host receptors. Natural evolutionary changes in SLBR sequence might involve point mutations rather than substitutions of entire loops. We therefore wanted to test whether point mutations of the loops of SLBRHsa, SLBRSK678, and SLBRUB10712 could change the selectivity. In SLBRHsa, SLBRSK678, and SLBRUB10712, we substituted residues at positions equivalent to SLBRHsaE286 of the CD loop and SLBRHsaD356 of the FG loop (Supplementary Fig. 16a–d) and loss of function (Supplementary Fig. 17c–f). In contrast, SLBRSK678 and SLBRUB10712 bind more strongly to 6'S-leX/3'Sln > sLeC. In contrast, SLBRSK678 and SLBRUB10712 bind sTa > sLeC > 3'Sln > sLeX > 6'S-leX while SLBRHsa binds sTa > sLeC > 3'Sln > sLeX > 6'S-leX. a–f Measurements were performed using 500 nM of immobilized GST-SLBR and the indicated concentrations of each ligand (n = 3 independent experiments performed on protein from a single preparation). Source data are provided as a Source Data file.
because our structures show that these residues closely approach the variable region of the ligands (Supplementary Fig. 13). We then measured relative binding to five physiologically-relevant ligands via ELISA (Fig. 7a–c and Supplementary Table 4).

In the CD loop (i.e., SLBRHsa D286E), our crystallographic analysis suggested that ionic repulsion from the negatively-charged side chain excludes the negative charge of a sulfated ligand. We therefore substituted a positive charge at this location in SLBRUB10712 and SLBRSK678 and conversely substituted Gln for unbranched Gln of SLBR UB10712 and SLBR SK678 can bind fucosylated ligands. We therefore substituted Asp for Gln in the side chain excludes the negative charge of a sulfated ligand. SLBRUB10712Q354D and SLBRSK678Q367D exhibited low binding to all tested ligands. As assessed by size exclusion chromatography, the SLBRSK678Q367D variant was properly folded such that loss of binding did not result from a folding defect (Supplementary Fig. 19c). The observed loss of binding to the fucose-containing sLeX and 6S-sLeXβ by these FG loop variants is consistent with the structural analysis and chimeragenesis showing that the FG loop is particularly important for accommodation of α1,3-fucosylation (Fig. 6 and Supplementary Fig. 15a, b). The converse SLBRHsa D356R, and SLBRHsa D356Q remained broadly-selective but with increased binding to the α1,3-fucosylated sLeX and 6S-sLeXβ as compared to parent SLBRHsa (Fig. 8c, d and Supplementary Fig. 19d).

Taken together, point mutations in the broadly-selective SLBRs can alter the identity of the preferred ligand, and can bind robustly to the newly-preferred ligand. The EC50 values suggest that the binding is strong enough to make physiologically-relevant adhesive interactions to host receptors. A possible evolutionary rationale for facile alteration in sialoglycan binding spectrum is that this allows a bacterium to adapt to changes in host sialoglycan display.

Selectivity variants alter the preferred host receptor. To test whether changes in SLBR binding to synthetic glycans had corresponding effects in the interactions of the SLBRs with human ligands, we examined the binding of parent and variant SLBRs to human salivary and plasma glycoproteins using far western analysis. We focused on the chimeras and variants that had responding effects in the interactions of the SLBRs with human sialoglycan display.

**Figure 7: Binding selectivity of CD loop variants in SLBR_{SK678}, SLBR_{UB10712} and SLBR_{Hsa}**. Dose response curves of biotin-glycan binding to immobilized variant GST-SLBRs (500 nM). a) GST-SLBR_{SK678} E298R, b) GST-SLBR_{UB10712} E285R, c) GST-SLBR_{Hsa} E286E. The respective SLBRs are shown in gray cartoon in the top left corner of each panel with the site of mutation represented as a colored sphere. sTa, shown in red sticks, was placed over the binding site by superimposing sTa bound-SLBRHsa. Measurements were performed using 500 nM of immobilized GST-SLBR and the indicated concentrations of each ligand are shown as the mean ± SD. d-f) Binding of each sugar at 2 µg/mL to each mutant was statistically compared to binding of the same sugar to the SLBRWT with the values presented in a-c using a two-tailed parametric t test. Black circles represent individual data points and bars represent the mean ± SD (n = 3 independent experiments performed on protein from a single preparation). Statistical significance is indicated by: ns, p > 0.05; *, p < 0.05; **, p < 0.01; †††, p < 0.001; ††††, p < 0.0001). In panel d, the p values from left to right are 0.94, 0.15, <0.0001, 0.13, and <0.0001. In panel e, the p values from left to right are 0.0007, 0.0034, 0.046, <0.0001, and 0.24. In panel f, the p values from left to right are 0.047, 0.019, 0.0016, <0.0001, and <0.0001. Source data are provided as a Source Data file.
Both variants had increased binding to 6S-sLeX,3 variant SLBRs (500 nM). Both fucosylated ligands sLeX and 6S-sLeX. In SLBRHsa, charge reversal or neutralization at this same position was assessed in chimeras) or with the presence of 6-...
chimeric SLBRUB10712Hsa-loops and SLBRSK678Hsa-loops chimeras now recognized proteoglycan 4 rather than the preferred receptors for parent SLBRSK678 and SLBRUB10712 (Fig. 9c). We also found that the SLBRSK678E298R variant bound both GPIbα, a receptor associated with infective endocarditis, and the C1-esterase inhibitor (Fig. 9d). Thus, the preferred plasma ligands for the SLBRs appears to be largely determined by the loop residues, as was the case for the recognition of MUC7 glycoforms.

Discussion

Bacterial attachment to host structures is critical for commensalism and is the first committed step in many types of infection. SLBRs can mediate streptococcal binding to a variety of host
glycoproteins \(^5,7–10,14,22,24,25,29,30\), and binding to sTa correlates with pathogenesis in an animal model of endovascular infection \(^22\). But it has not previously been clear how the SLBRs distinguish between the many protein-attached glycans displayed by host. Here, we evaluated how five SLBRs select between sialic glycan receptors. The common element of these glycans, Neu5Ac, interacts with SLBR via the V\(\delta\) TRX motifs \(^13,16,31\), and the EF loop (Fig. 4 and Supplementary Fig. 9) \(^16,23\). The CD and FG loops select for the underlying reducing end (Figs. 6, 8, and Supplementary Figs. 15 and 17), which varies in the identities of its individual sugars, the linkage between the sugars, and the elaborations on the sugars (Fig. 1). This suggests roles for distinct regions of the SLBR structure in glycan selection (Fig. 10) \(^{16,23}\). The substantial sequence and structural variability in the CD, EF, and FG loops as compared to the core fold of the SLBR (Supplementary Figs. 3 and 4) suggests that these regions can tolerate more substitutions while avoiding the liability of misfolding. Indeed, modification of these regions via chimerization or mutation allowed some of the SLBRs to bind different glycoforms of MUC7 or interact with different sialylated glycans (Figs. 6, 7, 8, and Supplementary Figs. 14, 15, 17 and 20) and different host plasma proteins (Fig. 9). Although not previously noted for bacterial SLBRs, the use of loops to control selectivity has been observed in other sialic glycan-binding systems. For example, mammalian Siglec proteins are organized around a V-set Ig-fold but are not detectably related in sequence to the SLBRs \(^{13,23,33}\). The GG’ and CC’ loops are adjacent to the sialic glycan binding site and are variable in structure. In Siglec-7, the CC’ loop \(^{34}\) controls sialic glycan selectivity. In Siglec-8, alteration of this same loop allows the binding of 6’S sialylglycans \(^{35}\). Thus, changes in loop structure may therefore be a common way to evolve changes in ligand binding selectivity.

The use of loops to control selectivity appears to be a robust way to accommodate a broad range of complex glycans. Indeed, the glycans recognized by SLBRs differ in both the identity of the individual glycans as well as in the linkages between the individual carbohydrates. When bound to these SLBRs (see Supplementary Fig. 13b), glycans with different linkages differ in the overall shape as well as in the pattern of hydrogen-bonding donors and acceptors. However, the glycodic linkage itself does not differ in position with respect to the SLBR binding pocket. Thus, these SLBRs distinguish between glycans with different linkages by changing the steric and electrostatic properties of the region of the pocket that follows the linkage, namely the CD and FG loops. While we focused on SLBRs that recognize tri- and tetrasaccharides, SLBRs can recognize sialylglycans with as few as three and possibly more than six monosaccharide units \(^5,8,19,23\). For example, SLBR\(_{\text{MUC7a}}\) may biologically recognize a hexasaccharide \(^8\) but can bind to partial ligands with lower affinity \(^5,16,23\). SLBRs that recognize larger sialylglycans appear to contain a modular binding site similar to those studied here, albeit with larger binding pockets and with more independent recognition regions. In the oral cavity, this may assist in colonization through interaction with salivary MUC7, which exhibits heterogeneity of its sialylglycan modifications both within and between human hosts (Fig. 9a, b and Supplementary Figs. 20 and 21). Here, sialylglycans are attached to MUC7 and the SLBR binding pocket can bind glycan receptors that are linked to host proteins. The linkage to the receptor protein could affect binding and could involve additional contacts to the SLBR. \(^8\)

In this context, mutation of these loops may be advantageous to the bacterium because it allows facile switching of host receptors. While we do not know how the sequences of the SLBRs actually change during evolution, streptococci compete with numerous other species in the oral cavity \(^36\). As many of these strains contain SLBRs, genetic recombination is likely, which can allow a bacterium to incorporate or modify a SLBR. The ready toleration of mutations in the loops may allow these regions to disproportionately change their sequences. Some of these changes may enable the bacterium to bind a different sialic glycan structure (Figs. 7–9 and Supplementary Figs. 15, 18 and 19). Within a single human host, this could allow colonization of a region of the oral cavity that displays different glycans, could promote binding to different salivary components, or could allow binding to other oral bacteria that are sialylated. This mutability could also permit improved binding to different individuals in the population or allow the colonization of a preferred host, as animals and humans may differ in their glycosylation \(^37\). This mechanism mirrors that of polyomavirus and rotavirus, where single amino acid substitution or a very small number of point mutations can change the identity of preferred host sialic glycan receptors \(^38,39\).

In some of our point mutants, the improvement in affinity and selectivity to alternative ligands exceeds those reported for dedicated engineering studies of glycan-binding lectins \(^40–48\). In those past reports, the maximum enhancement in binding to a non-native glycan is \(>20\)-fold \(^40–45\) and selectivity was often achieved via a decrease in affinity to non-desired ligands in a promiscuous starting lectin \(^46–48\). Development to increase the affinity and narrow the selectivity even further could allow the SLBRs to be used as probes to assess glycan identity and abundance. Key aspects of a probe include the ability to detect glycans in cells and in patient samples. The cellular interaction was shown in recent studies that evaluated the binding SLBRs to engineered HEK293 cell lines with altered glycosylation \(^49\), while the ability to recognize glycans in saliva and plasma suggests that these will be useful in other samples (Fig. 9).

Collectively, our findings give a description for how SLBRs recognize glycans. The conserved sialic acid-recognition motif governs general specificity while sequence diversity in surrounding loop regions allows the SLBR to select between related sialylglycans (Fig. 10). This binding site architecture may be optimized for facile selectivity changes in related SLBRs. This may further explain how bacterial adhesive proteins have evolved to adapt to host receptors. Finally, this work suggests a route for engineering these SLBRs to use as probes to detect specific
glycosylation, which is a focus of ongoing work. A library of SLBR-based binding proteins could be used for glycome mapping or as diagnostic or therapeutic tools for disease states with aberrant glycosylation.

Methods
Sequence analysis. SLBR sequences were aligned using the MUSCLE subroutine in Geneious Pro 11.1.4.50. The JTT-G evolution model was selected using the ProtTest server52, and the phylogenetic tree was built using the MrBayes53 subroutine of MrBayes 13.

Cloning, expression, and purification for crystallization. DNA encoding all SLBRs except SLBRh21 were cloned into the pGB101 vector (Vanderbilt University), which encodes an N-terminal Hist-GST tag cleavable with 3c protease. SLBRh21 was cloned into the pSV278 vector (Vanderbilt University), which encodes a thrombin-cleavable His-maltose binding protein (MBP) tag. Proteins were expressed in E. coli B21 (DE3) with 50 µg/ml kanamycin at 37 °C. Expression was induced with 0.5–1.0 mM IPTG at 24 °C for 3–7 h. Cells were harvested by centrifugation at 5000 × g for 15 min and stored at −20 °C before purification.

Cells were resuspended in 20–30 mM Tris-HCl, pH 7.5, 150–200 mM NaCl, 1 mM EDTA, 1 mM PMFS, 2 µg/ml Leupeptin, 2 µg/ml Pepstatin then disrupted by centrifugation at 5000 × g for 6 min, and washed in 200 mM NaCl.

We note that the S. gordoni strain UB10712 was recently re-typed. Previous strain UB10712 was recently re-typed. Previous S. gordonii strain NCTC10712.

Structure determination. Crystallizations were performed at room temperature (−23 °C) using the conditions in Supplementary Table 5. The SLBRh21-TTa structure used crystals where the ligand was introduced by cotransformation, and the SLBRh21-TTa-ligand structures used crystals where the ligand was introduced by soaking. Data collection and refinement statistics are listed in Supplementary Tables 1, 2, and 3. Structures were determined by molecular replacement using the Phaser54 subroutine of phenix 1.18.25 using the starting models listed in Supplementary Table 5.

All models were improved with iterative rounds of model building in Coot 0.956 and refinement in Phenix 1.18.257. Riding hydrogens were included at resolutions better than 1.4 Å. For sialylglycan-bound SLBRh21, the crystals were isomorphous with unliganded crystals and Rp values were similar to identical. Panel representing and reference molecules were docked at 1.5 Å resolution. The agreement between the glycans and SLRBSK678 or SLBRSK678Hsa loops. The CD loop controls specificity, the EF loop controls promiscuity with larger glycans23. The CD loop may control whether the SLBR prefers trisaccharides versus sialosylglycans, and the EF loop may affect selectivity. The CD loop controls specificity, the EF loop controls promiscuity with larger glycans23. The CD loop may control whether the SLBR prefers trisaccharides versus sialosylglycans, and the EF loop may affect selectivity.

Sialoylglycan binding. DNA encoding wild-type and variant SLBRs were cloned into pGEX-3X. Individual GST-SLBR fusions were expressed and purified using glutathione-agarose, and the binding of biotinylated glycans to immobilized GST-SLBRs was performed as described previously. Anti-GST antibody was used at a 1:500 dilution and was from Invitrogen (A8900). Peroxidase-conjugated goat anti-rabbit IgG was from Sigma (A0545). Unclipped gels source data are provided as Source Data file.

Far western and lectin blotting of human proteins. Far-western blotting of human saliva and plasma proteins using the indicated GST-SLBRs (15 nM) as probes was performed as described.55 Plasma was purchased from Innovative Research (Novi, MI). De-identified samples of SMSL saliva were provided by S. Fisher (UCSF), and were collected through a protocol approved by the UCSC Institutional Review Board. Donors confirmed that their samples may be used for other research purposes. Because these specimens were de-identified prior to gifting, our use of this material was exempt from approval by the UCSC Institutional Review Board and was not considered as human subject research. Anti-GST antibody was from Invitrogen (A8900) and peroxidase-conjugated goat anti-rabbit IgG was from Sigma (A0545). Uncropped gels source data are provided as Source Data file.

MUC7 affinity capture and O-glycan profiling. A combination of GST-SLBRh21 and GST-SLBRHsa,loops immobilized on magnetic glutathione beads was used to capture the total sialylated MUC7 from 300 µl of SMSL saliva. The resin-bound GST-SLBRs and affinity-captured MUC7 were co-eluted into LDS sample buffer (Invitrogen) supplemented with dithiothreitol (100 mM final concentration), separated by electrophoresis in 4–12% polyacrylamide gradient gels, and then stained with SimplyBlue SafeStain (Invitrogen). The captured proteins, which ranged from 120–160 kDa, were excised from the gel. A portion of the sample was submitted for protein identification by nanoflow LC-MS/MS of tryptic digests (MSBworks), which confirmed MUC7 as the major component. A second portion of the excised gel slices was mixed, treated by four cycles of rinsing with 100 mM ammonium bicarbonate and dehydration in 100% acetonitrile, and then dried to completion in a vacuum evaporator. The gel pieces were immersed in a mixture of 100 mM NaOH and 1 M NaBH4 and incubated at 45 °C for 18 h to release the O-glycans. The supernatant was collected and placed on ice, and the remaining gel pieces were washed with water and sonicated for 30 min to extract the remaining O-glycans. The initial and secondary extracts were combined and acidified to pH 4.6–5.0 by drop-wise addition of 10% acetic acid. The O-glycan samples were then enriched using porous graphitized carbon cartridges (Agilent, Santa Clara, CA) and dried prior to analysis by mass spectrometry. Glycan samples were analyzed on an Agilent 6520 Accurate Mass QTOF LC/MS equipped with a porous graphitic carbon mass spectrometry chip. A binary gradient consisting of (A) 0.1% formic acid in 80% acetonitrile, and (B) 1% formic acid in 90% acetonitrile was used to separate the glycans at a flow rate of 0.3 µl/min. Data were processed with Agilent MassHunter B.07 software, using the Find by Molecular Feature algorithm with an in-house library of O-glycan masses and chemical formulae to identify and quantitate the O-glycan signals.

In silico structure predictions and MD analyses. The model of SLBRh21-TTa-loops was calculated using MOE. For MD of SLBRh21-TTa-loops, SLBRHsa,loops, and SLBRHsa,loops each set of PDB coordinates was solvated in a 10 Å octahedral box of TIP3P water. The Amber16 IT4SB58 force field was used for the protein. In the first step of the MD simulation, the backbone and side chains of the protein were restrained using 500 kcal mol−1 A−2 harmonic potentials while the system was energy minimized for 200 steps of steepest descent59 and the conjugate gradient method60. Restraints 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 at 300 K with the backbone and side chains restrained using 10 kcal mol−1 A−2 harmonic potentials for 1000 steps. Bonds were constrained using SHAKE68. MD (200 ns) was performed at 300 K in the NPT ensemble and a 2-fs time step. Probability distribution analyses and RMSF calculations were performed on 200 ns of 3 independent runs. Analyses were performed using the cpptraj and pytraj modules of AMBER16. The last snapshot from 20-ns trajectory was used for mapping the interaction between the glycans and SLBRHsa,loops or SLBRHsa,loops.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Source data are provided as a Source Data file. Atomic coordinates and structure factors have been deposited into the RCSB Protein Data Bank at www.rcsb.org under the accession codes

Fig. 10 Model for how SLBRs control sialoylglycan selectivity. The glycans-binding pocket of SLBRs is organized above a ΦTRX sequence motif on the F-strand of the V-set Ig fold that interacts with sialic acid. Three variable loops surround this sialoylglycan binding pocket and affect selectivity. In the broadly-selective SLBRs, flexibility of the EF loop correlates with breadth of selectivity. The CD loop controls specificity for 6-O-sulfated glycans, and the FG loop may control whether the SLBR prefers trisaccharides versus larger glycans. Tetrasaccharides containing α,3-fucosylation were tested here, but past studies of SLBRg21α identify that a small FG loop correlates with the ability to accommodate larger glycans23.
MassIVE (https://massive.ucsd.edu/) with the data identified as available via the accession codes 5IJ3, and 6VT2. Previously 5EQ2, and 3QC5. Published structures used for molecular replacement are available via the accession codes 4. Varki, A. Biological roles of glycans. NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-30509-y ARTICLE

Received: 10 April 2021; Accepted: 26 April 2022; Published online: 18 May 2022

References
1. Langereis, M. A. et al. Complexity and diversity of the mammalian sialome. Cell Rep. 11, 1966–1978 (2015).
2. Varki, A. Nothing in glyciobiology makes sense, except in the light of evolution. Cell 126, 841–845 (2006).
3. Varki, N. M., Strober, E., Dick, E. J. Jr., Benirschke, K. & Varki, A. Biomedical differences between human and nonhuman hominids: potential roles for uniquely human aspects of sialic acid biology. Annu. Rev. Pathol. 6, 365–393 (2011).
4. Varki, A. Biological roles of glycans. Glycobiology 27, 3–49 (2017).
5. Bensing, B. A. et al. Novel aspects of sialylglycan recognition by the Siglec-like domains of streptococcal SRR glycoproteins. Glycobiology 26, 1222–1224 (2016).
6. Thamadilok, S., Roche-Hakansson, H., Hakansson, A. P. & Ruhl, S. Absence of capsules reveals glycan-mediated binding and recognition of salmonic mucMUC7 by Streptococcus pneumoniae. Mol. Oral. Microbiol 31, 175–188 (2016).
7. Takamatsu, D., Bensing, B. A., Prakobphol, A., Fisher, S. J. & Sullam, P. M. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. Infect. Immun. 74, 1933–1940 (2006).
8. Bensing, B. A., Li, Q., Park, D., Lebrilla, C. B. & Sullam, P. M. Streptococcal Siglec-like adhesins recognize different subsets of human plasma glycoproteins: implications for infective endocarditis. Glycobiology 28, 601–611 (2018).
9. Takamatsu, D. et al. Binding of the Streptococcus gordonii surface glycoprotein GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibα/αlpha. Mol. Microbiol. 58, 380–392 (2005).
10. Plummer, C. et al. A serine-rich glycoprotein of Streptococcus sanguis mediates adhesion to platelets via GPIb. Br. J. Haematol. 129, 101–109 (2005).
11. Bashore, T. M., Cabell, C. & Fowler, V. Jr. Update on infective endocarditis. Curr. Probl. Cardiol. 31, 274–352 (2006).
12. Lizzano, A., Sanchez, J. I. & Orihuela, C. J. A role for glycosylated serine-rich repeat proteins in gram-positive bacterial pathogenesis. Mol. Oral. Microbiol 27, 257–269 (2012).
13. Pyburn, T. M. et al. A structural model for binding of the serine-rich repeat adhesin GspB to host carbohydrate receptors. PLoS Pathog. 7, e1002112 (2011).
14. Takahashi, Y. et al. Contribution of sialic acid-binding adhesin to pathogenesis of experimental endocarditis caused by Streptococcus gordonii. J. Infect. Immun. 74, 740–743 (2006).
15. Stubbs, H. E. et al. Tandem sialylglycan-binding modules in a Streptococcus sanguinis serine-rich repeat adhesin create target dependent avidity effects. J. Biol. Chem. 295, 14737–14749 (2020).
16. Bensing, B. A. et al. Structural basis for sialylcycan binding by the Streptococcus sanguinis SrpA adhesin. J. Biol. Chem. 291, 7230–7240 (2016).
17. Agarwal, R. et al. Structure based virtual screening identifies novel small molecule effectors for the sialylcycan binding protein Ha. Biochem J. 477, 3685–3707 (2020).
18. Di Carlucchio, C. et al. Molecular recognition of sialylglycans by streptococcal Siglec-like adhesins: toward the shape of specific inhibitors. RSC Chem. Biol. 2, e00406–19 (2021).
19. Deng, L. et al. Oral streptococci utilize a Siglec-like domain of serine-rich repeat adhesins to preferentially target platelet sialylglycans in human blood. PLoS Pathog. 10, e1004540 (2014).
20. Prakobphol, A. et al. Human low-molecular-weight salivary mucin expresses the sialyl Lewis x determinant and has L-selectin ligand activity. Biochemistry 37, 4916–4927 (1998).
21. Karlsson, N. G. & Thomson, K. A. Salivary MUC7 is a major carrier of blood group I type O-linked oligosaccharides serving as the scaffold for sialyl Lewis x. Glycobiology 19, 288–300 (2009).
22. Bensing, B. A. et al. Recognition of specific sialylglycan structures by oral streptococci impacts the severity of endocardial infection. PLoS Pathog. 15, e1007896 (2019).
50. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acid Res. 32, 1792–1797 (2004).
51. Kearse, M. et al. Geneious Basic: an integrated and extensible desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649 (2012).
52. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105 (2005).
53. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542 (2012).
54. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).
55. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D. Biol. Crystallogr. 66, 213–221 (2010).
56. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D. Biol. Crystallogr. 60, 2126–2132 (2004).
57. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935 (1983).
58. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput. 11, 3696–3713 (2015).
59. Arifken G., Weber H. & Harris F. Mathematical methods for physicists—a comprehensive guide, 7th edn. (Academic Press, 2012).
60. Case, D. A. et al. The Amber biomolecular simulation programs. J. Comput. Chem. 26, 1668–1688 (2005).
61. Ryckaert, J., Ciccotti, G. & Berendsen, H. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23, 327–341 (1977).
62. Roe, D. R. & Cheatham, T. E. III. PTRAJ and CIPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theory Comput. 9, 3084–3095 (2013).

Acknowledgements

We thank S Berendzen, I. Loukachevitch, and P Singh for experimental and analytical assistance, 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 (R01AI14513 and U01CA221244 to P.S., R01AI106987 to T.M.I./P.M.S., R03 DE029516 to B.A.B., and GM137458 to T.M.I.), and the American Heart Association (R01AI106987 to T.M.I./P.M.S., R03 DE029516 to B.A.B., and GM137458 to T.M.I.; 14GRNT20390021 to T.M.I.; 17SDG33660424 to B.A.B.). AH was supported by the Department of Veterans Affairs. The remaining authors declare no competing interests.

Competing interests

T.M.I. and B.A.B. hold a provisional patent, PCT/US2021/036983 that covers mutation of SLBRs for use as binding probes. The patent includes mutants described here as well as mutations with different selectivities and the methods to create new mutants via mutation or chimeragenesis. Authors on the patent are affiliated with Vanderbilt University, the Regents of the University of California, and the United States as represented by the Department of Veterans Affairs. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-30509-y.

Correspondence and requests for materials should be addressed to T. M. Iverson.

Peer review information Nature Communications thanks Michael Jarvè, Samantha King, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information are available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.