Sulfated glycans have been found to be associated with various diseases and therefore have significant potential in molecular pathology as biomarkers. Although lectins are useful reagents for detecting glycans, there is a paucity of sulfate-recognizing lectins, and those that exist, such as from Maackia amurensis, display mixed specificities. Recombinant lectin engineering offers an emerging tool for creating novel glycan recognition by altering and/or enhancing endogenous specificities. The present study demonstrated the utility of computational approaches in the engineering of a mutated form of E-selectin that displayed highly specific recognition of 6'-sulfo-sialyl Lewis X (6'-sulfo-sLe\(^\alpha\)), with negligible binding to its endogenous nonsulfated ligand, sLe\(^\alpha\). This new specificity mimics that of the unrelated protein Siglec-8, for which 6'-sulfo-sLe\(^\alpha\) is its preferred ligand. Molecular dynamics simulations and energy calculations predicted that two point mutations (E92A/E107A) would be required to stabilize binding to the sulfated oligosaccharide with E-selectin. In addition to eliminating putative repulsions between the negatively charged side chains and the sulfate moiety, the mutations also abolished favorable interactions with the endogenous ligand. Glycan microarray screening of the recombinantly expressed proteins confirmed the predicted specificity change but also identified the introduction of unexpected affinity for the unfucosylated form of 6'-sulfo-sLe\(^\alpha\) (6'-sulfo-sLacNAc). Three key requirements were demonstrated in this case for engineering specificity for sulfated oligosaccharides: 1) removal of unfavorable interactions with the 6'-sulfate, 2) introduction of favorable interactions for the sulfate, and 3) removal of favorable interactions with the endogenous ligand.

Sulfation is a ubiquitous and important posttranslational modification of many biological molecules, including proteins (1), carbohydrates (2, 3), lipids (4), and glycolipids (5), and mediates many biological functions. Sulfated glycans are associated with various diseases, such as cancers (6, 7), cystic fibrosis (8–10), and osteoarthritis (11, 12), and have great potential in molecular pathology as biomarkers. However, the isolation and detection of sulfated glycans is challenging because of their low abundance in cells, their low ionization efficiency for detection by mass spectroscopy, and the fact that the modification is labile under even relatively mild isolation conditions (13, 14). Although lectins are often used to detect glycans (for example, in histology) or to enrich them chromatographically before further analysis, their application to sulfated glycans is challenging due to the paucity of sulfate-recognizing lectins as well as their broad or mixed specificities. For example, lectins from Maackia amurensis recognize both 3'-sulfated and 3'-sialylated oligosaccharides (15, 16), which is perhaps not that unexpected given that sulfate and sialic acid are both anionic. Even more surprising is the observation that the lectin from Langerin cross-reacts with 6'-sulfated glycans (25, 26); however, many of these have broad specificities, with the notable exceptions of Siglec-8 and Siglec-9. Siglec-8 displays a strong preference for ligands that contain sulfation at the O6 position of galactose in sialyl Lewis X (6'-sulfo-sLe\(^\alpha\); Neu5Acα2-3Gal[6S]β1-4[Fucα1-3]GlcNAcβ1) or at that position in sialyl LacNAc (6'-sulfo-sLacNAc; Neu5-Acα2-3Gal[6S]β1-4GlcNAcβ1) (27). In contrast to Siglec-8, Siglec-9 prefers ligands that are sulfated at the O6 position of GlcNAc and shows enhanced glycan array...
binding when fucose is present; that is, 6-sulfo-sLe\(^{\alpha}\) \(\gg\) 6-sulfo-sLacNAc (27). These two Siglecs display orthogonal ligand specificities, and they are also selectively expressed on different leukocytes. Siglec-8 is found on eosinophils, basophils, and mast cells where it regulates their function and survival (28), while Siglec-9 is expressed by neutrophils and monocytes where it modulates the function of neutrophils during infection (29).

Understanding the biological mechanisms by which subtle differences in sulfation patterns govern these specificities is an area of active research. It is worth mentioning here that the specificity of monoclonal antibody S2 parallels that of Siglec-9; although, in contrast to Siglec-9, the presence of fucose in the ligand does not enhance S2 binding (30). Thus, different proteins may display unique binding modes for the same oligosaccharide ligand.

Much of the latest data pertaining to the roles of glycan sulfation (25, 31–33) have come somewhat indirectly, and with considerable effort, from genetic studies in which activity is inferred from the impact of the transfection or deletion of sulfotransferase genes in model cell lines. Given the emerging evidence that sulfation can dramatically enhance (25, 34) or abrogate (35, 36) protein binding, the paucity of reagents that are able to detect specific sulfation patterns in vitro or in vivo creates a barrier to advancing this already challenging field. The narrow specificity of Siglec-8 presents a remarkable example of a highly specific interaction between an endogenous protein and 6'-sulfated oligosaccharides. Because such a degree of specificity is rare among naturally occurring lectins, there is a need for an alternative to serendipitous lectin discovery for the generation of novel carbohydrate detection reagents. One potential approach would be to engineer the desired specificity into an existing lectin scaffold. To be truly specific, however, the reagent should also display reduced or eliminated binding to the endogenous glycan(s). Several examples of lectin specificity engineering have been reported (37–50), with varying degrees of success.

In an early example of carbohydrate specificity engineering, based on domain swapping, Drickamer (37) introduced galactose-binding activity into a C-type lectin by substituting two amino acids that are conserved in the carbohydrate recognition domain in the mannose-specific lectin (E185 and N187) with two that are conserved in related lectins that prefer galactose. The double mutant (E185Q/N187D) indeed preferred galactose over mannose by 3.5-fold compared to the wild type, which preferred mannoside over galactose by almost 14-fold. Nevertheless, the double mutant retained significant affinity for the endogenous ligands of the parent lectin and unexpectedly introduced high affinity for N-acetylgalactosamine. Subsequently, in the quest to develop a lectin with improved detection capability for the Thomsen–Friedenreich (TF) tumor antigen (Galβ1-3GalNAc), Adhikari et al. (38) introduced point mutations into a recombinant version of peanut agglutinin at a position (N41) that was known to stabilize a water-mediated hydrogen bond with the ligand. Replacing N41 with a glutamine enhanced affinity for the TF antigen by approximately fourfold, rationalized as arising from the replacement of the mediating water by a direct interaction with the Q41 side chain. Although the N41Q point mutation improved affinity for the target ligand, it did not narrow the endogenous specificity of the lectin. In the design of a probe for the disease marker 6-sulfo-galactose, Hu et al. (40) applied error-prone PCR to a recombinant form of the ricin B-domain lectin with endogenous affinity for galactose, work that built off their earlier success applying this approach to introduce specificity for 6'-sialyl-galactose into the same system (45). One mutation (E20K) in particular was identified as being critical for introducing sulfate binding; however, no clones were reported that significantly reduced binding to the endogenous nonsulfated ligands.

The three cases introduced above represent common approaches to lectin specificity engineering, namely, sequence-based domain swapping, structure-based point mutagenesis, and random mutation (directed evolution). Each successfully achieved the goal of introducing either novel or enhanced affinity; however, none were able to simultaneously reduce or remove affinity for the endogenous ligands. This latter property is essential to fully exploit the engineered lectin as a reagent in diagnostic or therapeutic applications.

In the present work, we sought to use computational methods to guide the design of a protein that could recognize 6'-sulfo-sLe\(^{\alpha}\) (a ligand for Siglec-8) based on introducing sulfate specificity into a lectin (E-selectin) known to bind the nonsulfated congener. Selectins recognize the unsulfated core tetrascarhide sLe\(^{\alpha}\), which is found in glycoproteins, such as P-selectin glycoprotein ligand-1 (PSGL-1) (51), E-selectin-ligand-1 (52), and some CD44 isoforms (53). Sulfation of sLe\(^{\alpha}\) can enhance, attenuate, or switch selectin specificity (21, 35, 54, 55). However, in the case of E-selectin, neither direct sulfation of sLe\(^{\alpha}\) nor sulfation of its associated peptide enhanced its affinity (21, 56). No members of the selectin family recognize 6'-sulfo-sLe\(^{\alpha}\) (35).

A computational approach to protein engineering has several benefits over purely experimental techniques, including the ability to predict the effect of hypothetical point mutations on ligand binding. Further, to achieve high specificity, we wanted to test the hypothesis that in addition to introducing affinity for the sulfate group, mutations could be introduced that would reduce or eliminate affinity for the endogenous nonsulfated oligosaccharide. E-selectin was chosen to demonstrate this approach as its specificity and three-dimensional (3D) structure have been reported previously, and it has been shown to have no measurable affinity for 6'-sulfo-sLe\(^{\alpha}\) (35). The results from this study may offer insight into the rules governing oligosaccharide specificity and provide a rational and generalizable approach to developing carbohydrate-specific reagents.

**Results**

**Molecular Models for E-Selectin and Siglec-8 Complexes.** To confirm the validity of the molecular modeling protocol, molecular dynamics (MD) simulations (200 ns each) were first performed on complexes of E-selectin and Siglec-8 with their cognate ligands sLe\(^{\alpha}\) and 6'-sulfo-sLe\(^{\alpha}\), respectively, as reported from experimental structural studies (57, 58). Additionally, to permit a statistical assessment of the variability in the data, three independent MD simulations were performed for each complex. The MD simulation data reproduced the experimentally observed ligand binding poses and glycosidic linkage values (SI Appendix, Fig. S1) and the interatomic interactions (57, 58) (Table 1 and SI Appendix, Table S1). In the case of E-selectin, the endogenous sLe\(^{\alpha}\) ligand maintained stable hydrogen-bond interactions with the protein via the sialic acid, galactose, and fucose residues, as observed in the crystal structure (57). In the case of Siglec-8, the endogenous 6'-sulfo-sLe\(^{\alpha}\) ligand also maintained the key experimentally observed interactions during the simulations (Table 1 and SI Appendix, Table S1), particularly with R56, R109, and K116, residues that are known to be critical for affinity based on experimental alamine scanning (58). Notably, MD simulations of the R56A, R109A, and K116A mutants (computational alamine scanning) in the complex with 6'-sulfo-sLe\(^{\alpha}\) showed that, consistent with the experimental affinity data (58), the loss of R109 completely
abolished affinity for the 6'-sulfo-sLe^c ligand, as evidenced by the ligand diffusing out of the binding site (SI Appendix, Fig. S2). Data from molecular mechanical generalized Born surface area (MM-GBSA) binding energy analyses (59) also supported the experimental observations that the R56A and K116A mutations weaken affinity but do not abolish it (SI Appendix, Fig. S2 and Table S2).

Having thus confirmed the ability of the molecular modeling protocols to reproduce the experimentally observed conformations and interactions for the known E-selectin–sLe^c and Siglec-8–6'-sulfo-sLe^c complexes, we applied the computational mutagenesis method to design 6'-sulfation recognition into E-selectin.

**Engineering 6'-Sulfation Recognition into E-Selectin.** To initiate the engineering of E-selectin to recognize 6'-sulfo-sLe^c, MD simulations of E-selectin in complex with the target ligand (6'-sulfo-sLe^c) were performed, in which the sulfated ligand was generated by replacing the 6-OH with a sulfate moiety (Materials and Methods). Consistent with the experimental observation that E-selectin does not show measurable affinity for this sulfated ligand (35), the complex was unstable during each of the three independent simulations (SI Appendix, Fig. S3). A closer examination of the trajectories showed that ligand instability was accompanied by distortions of the glycosidic linkages into high-energy conformations (SI Appendix, Fig. S4). Instability of the complex precluded calculation of the interaction energy for this system; however, examination of the E-selectin–6'-sulfo-sLe^c complex suggested that the ligand instability likely arose from the presence of unfavorable van der Waals and electrostatic interactions between the sulfate moiety and the side chains of glutamate residues E92 and E107. In the E-selectin–sLe^c crystal structure, the 6-OH group in the galactose residue is in close proximity to the side chains of E92 and E107, with a hydrogen bond present involving the sulfate moiety (Fig. 1). Consequently, sulfation of the O6-group would result in unfavorable van der Waals and electrostatic interactions with the side chains of at least one of the glutamate residues. To fully investigate the impact of each of the negatively charged side chains in E92 and E107, ligand complexes with two single mutations (E92A and E107A) and a double mutation (E92A/E107A) were computationally analyzed, with the expectation that the smaller uncharged side chain of alanine would remove the unfavorable interactions with the sulfate moiety.

**E92A in E-selectin–6'-sulfo-sLe^c.** The E92A mutation abolishes the hydrogen bond with the 6-OH moiety in galactose observed in the wild-type E-selectin–sLe^c complex and concurrently eliminates

---

**Table 1. Stable intermolecular hydrogen-bond pairs observed in the MD simulations for E-selectin and Siglec-8 complexes with their endogenous ligands and the E92A/E107A–6'-sulfo-sLe^c complex**

|                  | E-selectin–sLe^c | E92A/E107A–6'-sulfo-sLe^c | Siglec-8–6'-sulfo-sLe^c |
|------------------|------------------|-----------------------------|-------------------------|
| Neu5Ac           | CO_3^-           | Y48*, R97*                  | R109*                   |
| O4               | E98              | E98                         | —                       |
| N5               | —                | —                           | K116*                   |
| O7               | —                | —                           | Y7*                     |
| O8               | —                | —                           | Y58, R109               |
| Core-2 Gal       | O3               | R97                         | —                       |
| O4               | Y94, R97         | Y94, R97                    | —                       |
| O6               | E92*             | —                           | —                       |
| Fuc              | O2               | E88*                        | —                       |
| O3               | E88*, Ca^{2+}    | E88, Ca^{2+}                | —                       |
| O4               | E80*, N82*, Ca^{2+} | E80, N82, Ca^{2+}       | —                       |
| SO_3 (6') O      | —                | N105, K111, K113           | R56*, Q59*              |

*Observed in experimental structures.
†No stable interactions observed.
putative repulsions between the sulfate moiety in 6'-sulfo-sLe^x. This mutation is therefore potentially important for both decreasing the affinity of the endogenous ligand and enhancing that of the target ligand. Indeed, when complexed with E92A, the 6'-sulfo-sLe^x ligand remained bound, although disordered, throughout each independent MD simulation, in contrast to the high degree of instability observed in the complex with wild-type E-selectin. Nevertheless, the ligand populated three distinct poses (pose1, pose2, and pose3 in Fig. 2), indicative of a high degree of disorder and instability. In the complex with E92A, pose1 and pose3 adopted a similar ligand shape (Fig. 2 and SI Appendix, Fig. S3), which is equivalent to that seen in the experimental co-complex with its cognate receptor Siglec-8 (58), but each pose adopted different orientations relative to the mutant protein surface (Fig. 2 and SI Appendix, Table S3). By contrast, the unique shape of the ligand in pose2 resulted from an unexpected flip of the GlcNAc ring from C_1 to C_4, likely induced by the initial placement of the ligand in the hypothetical E92A binding site. Thus, this single point mutation was predicted to be insufficient to lead to a complete conversion in ligand specificities.

To quantify these predictions, the interaction energies for E92A with the 6'-sulfo-sLe^x and with the nonsulfated ligand were computed. Pose1 and pose2 both displayed unfavorable binding interaction energies, 0.4 and 5.0 kcal/mol, respectively (SI Appendix, Table S4). Although pose3 showed a strong favorable binding interaction energy (~7.6 kcal/mol; SI Appendix, Table S4), the average binding interaction energy from the three independent MD trajectories (~0.7 kcal/mol) indicated a significantly weaker affinity than the wild-type E-selectin–sLe^x complex (~3.9 ± 1.0 kcal/mol; Table 2). Without the binding contribution from the 6'-sulfo moiety, the complex of E92A–sLe^x displayed a negligible binding interaction (~0.2 ± 2.1 kcal/mol; SI Appendix, Table S4), despite the observation that the ligand

![Fig. 2. Analysis of the stability of 6'-sulfo-sLe^x in the binding site of the mutant E92A. (A) Two-dimensional positional RMSD plot for the ring atoms in the ligand. Structures were evenly extracted every 0.1 ns from three MD simulations, and structurally related poses for the ligand are labeled (Bottom). (B) Trajectories of the positional RMSD (black) for the ring atoms and the total CHI energy (61) for the glycosidic linkages (purple) in the ligand in three MD simulations. The RMSD values were computed relative to the position of the endogenous ligand in the crystal structure of the E-selectin–sLe^x complex (57). (C) Glycosidic linkages trajectories (φ: blue; ψ: red) for the ligand in three independent MD simulations; φ = C2-C1-Ox-Cx (for Neu5Ac; φ = C3-C2-Ox-Cx-1) and ψ = C1-Ox-Cx-Cx-1 (for Neu5Ac; ψ = C2-Ox-Cx-Cx-1). (D) Superimposition of the complex for MD frames evenly extracted every 10 ns showing only the pyranoose ring and glycosidic linkage atoms for clarity. The monosaccharides are colored according to SNFG nomenclature (62, 63), and the protein solvent-accessible surface is shown in cyan.](https://doi.org/10.1073/pnas.2117743119)
maintained positional stability in the binding site (SI Appendix, Fig. S5). In the present energy calculations, we applied a value of 3.0 for the internal dielectric constant of the proteins in the MM-GBSA analysis. It should be noted that values greater than 1 for the internal dielectric constant are proposed to serve as an approximate method to account for the effect of charge polarization induced by ligand binding and are therefore somewhat system dependent, with values of 2 to 4 being proposed (64, 65). In the present case, the interaction energy computed with an internal dielectric constant of 3.0 also reproduced the interaction energy for Siglec-8 (3.7 kcal/mol) (66) for the E92A double mutation (Table 2). These MM-GBSA analyses are consistent with the stability of this complex indicating that the E92A/E107A double mutation significantly reduced any unfavorable electrostatic or van der Waals interactions with the 6′-sulfate moiety. Unlike the wild-type E-selectin and its two single mutants, the complex for the double mutant with 6′-sulfo-sLex was stable during each of the three independent MD simulations, with the ligand adopting the same single binding pose as seen in the endogenous sLeX ligand bound to wild-type E-selectin (SI Appendix, Figs. S6 and S8). The stability of this complex indicated that the E92A/E107A double mutation significantly reduced any unfavorable electrostatic or van der Waals interactions with the 6′-sulfate moiety. The stability of the ligand correlated with the presence of strong hydrogen bonds between the ligand and the protein, particularly those involving the sialic acid (carboxylate group and 4-OH), fucose (2-OH, 3-OH, and 4-OH), and galactose (3-OH and 4-OH) residues (Table 1). While the double mutation caused the loss of a hydrogen bond between the 6-OH group in galactose and the side chain of E92, new interactions were formed between the 6′-sulfo group and the side chains of polar residue N105 and positively charged residues K111 and K113 (Table 1).

The binding energies for the E92A/E107A–6′-sulfo-sLeX complex were computed to permit quantitative comparison with the wild-type E-selectin–sLeX complex. The per-residue MM-GBSA energy analysis of the E92A/E107A–6′-sulfo-sLeX complex showed that the fucose residue contributed more to the total binding energy (48%) than did the sialic acid residue (11%; Table 2). This energy distribution was similar to that seen in the E-selectin–sLeX complex (fucose, 52%; sialic acid, 14%; Table 2). These MM-GBSA analyses are consistent with the experimental observation that removing the fucose residue from the sLeX glycan in the PSGL-1 glycopeptide ligand reduced binding below the detection limit (65, 67). The binding energy for the oligosaccharide component (that is, the ligand not including the sulfate group) in the E92A/E107A–6′-sulfo-sLeX complex (−24.8 ± 1.0 kcal/mol) appeared to be

Table 2. Per-residue interaction energies* and entropic penalties† for wild-type and mutated E-selectins and Siglec-8 complexes with their ligands.

|                       | E-selectin–sLeX | E92A/E107A–6′-sulfo-sLeX | Siglec-8–6′-sulfo-sLeX |
|-----------------------|----------------|--------------------------|------------------------|
| **Per-residue interaction energies (MM-GBSA) and percentage of binding** | | | |
| Neu5Ac                | −3.6 ± 0.1 (14%) | −3.2 ± 0.6 (11%) | −12.6 ± 0.2 (64%) |
| Core-2 Gal            | −5.8 ± 0.5 (23%) | −4.7 ± 0.4 (17%) | −3.3 ± 0.0 (17%) |
| GlcNAc                | −2.8 ± 0.1 (11%) | −3.4 ± 0.1 (12%) | −2.5 ± 0.3 (12%) |
| Fuc                   | −13.2 ± 0.4 (52%) | −13.5 ± 0.0 (48%) | 0.0 ± 0.0 (<1%) |
| SO3 (6′)              | −3.3 ± 0.6 (12%) | −1.3 ± 0.2 (7%) | |
| ΔG_{MM-GBSA}          | −25.4 ± 0.7 | −28.1 ± 1.6 | −19.7 ± 0.3 |
| **Entropic penalties** | | | |
| −TΔS_{RTV (all)}      | 19.2 ± 1.1 | 19.7 ± 0.9 | 13.3 ± 0.6 |
| −TΔS_{c}              | 2.3 ± 0.0 | 2.4 ± 0.2 | 1.3 ± 0.1 |
| −TΔS                  | 21.5 ± 1.1 | 22.1 ± 1.1 | 14.6 ± 0.5 |
| **Binding free energies** | | | |
| ΔG_{Binding}          | −3.9 ± 1.0 | −6.0 ± 0.8 | −5.1 ± 0.6 |

*In kcal/mol.
†At 300 K.
slightly weaker than the endogenous sLe^a in the wild-type E-selectin complex (−25.4 ± 0.7 kcal/mol); however, this difference is not statistically significant (P = 0.4425). Thus, the enhanced binding of the sulfated ligand (−2.7 kcal/mol, not including entropy) can be attributed predominantly to new interactions formed with the sulfate moiety (−3.3 ± 0.6 kcal/mol).

The MM-GBSA analysis quantifies the magnitude of the energy gained from the formation of hydrogen-bond interactions between the sulfated group and N105, K111, and K113 and that lost with the abolition of the hydrogen bond between the 6-OH group in galactose and the side chain of E92. The impact of the double mutations on the binding affinity of the endogenous sLe^a ligand was determined from MD simulations of the putative complex of E92A/E107A with sLe^a. This complex was stable in each of the three independent MD simulations (SI Appendix, Fig. S5) and gave rise to a binding energy of −24.1 ± 0.5 kcal/mol, not including entropy. Although stable, the interaction energy was reduced by ~4 kcal/mol compared to that of the E92A/E107A-6′-sulfo-sLe^a (SI Appendix, Table S5). With the inclusion of entropic effects, the absolute binding energy of the E92A/E107A-sLe^a complex was −2.4 ± 0.9 kcal/mol compared to −6.0 ± 0.8 kcal/mol for the binding of the 6′-sulfo-sLe^a ligand.

The computational analysis predicted that to engineer recognition of 6′ sulfation into E-selectin would require removing unfavorable electrostatic and van der Waals interactions between the 6′-sulfo group and the negatively charged side chains of E92 and E107. The analysis also predicted that replacing only one of these side chains would only partially stabilize the 6′-sulfo-sLe^a ligand in the binding site. To fully stabilize the complex, a double mutation (E92A and E107A) was required. In the double mutant, the 6′-sulfo-sLe^a oligosaccharide bound in the same low-energy pose observed in the endogenous nonsulfated ligand and formed additional strong interactions involving the sulfate group. Moreover, specificity for the novel sulfated ligand over the endogenous oligosaccharide was predicted to arise from loss of a hydrogen bond to 6-OH in galactose after mutation of E92. Having obtained statistically robust data from MD simulations and MM-GBSA analyses, we then undertook the expression of the relevant mutants of E-selectin in HEK293 cells with the aim of experimentally confirming their specificity by glycan array screening.

**Glycan Microarray Data for Wild-Type and Mutated E-Selectins.** The recombinant E-selectin mutants were submitted to the National Center for Functional Glycomics (NCFG) for glycan microarray screening. Glycan array data for wild-type human E-selectin have been previously reported (68) and, as expected, showed binding to a limited number of sialic acid–containing glycans, including sLe^a and sLe^x (Fig. 3A). By contrast, the E92A/E107A double mutant displayed exclusive specificity for the 6′-sulfo sialylated lactosamine (6′-sulfo-sLacNAc) motif present in 6′-sulfo-sLacNAc (Fig. 3B), a specificity indistinguishable from that of the wild-type Siglec-8. Neither single mutation alone was sufficient to generate a binding signal to 6′-sulfo-sLe^a (Fig. 3 C and D). The loss of detectable affinity for the endogenous sLe^a in the double mutant is consistent with the modeling-based interpretation that the double mutation not only enhanced the binding to 6′-sulfo-sLe^a by introducing new interactions with 6′-sulfo group but also reduced the affinity for the endogenous sLe^a ligand by removing a key hydrogen bond between the Gal-O6 hydroxyl group and E92. Therefore, the specificity of the double mutant demonstrates the importance of combining mutations that enhance binding to the target ligand with ones that attenuate binding to the endogenous ligand. That the computed binding energy for the E92A/E107A-sLe^a complex (ΔGbinding = −2.4 ± 0.9 kcal/mol) was ~3.6 kcal/mol weaker than the binding of the 6′-sulfo-sLe^a ligand suggested that the double mutant retained some affinity for the nonsulfated ligand. However, the fact that this interaction was not observed by glycan array screening indicated that any remaining affinity must be below the detection limit of the experimental assay.

The observation that the double mutant bound to 6′-sulfo-sLacNAc was unexpected given that E-selectin–ligand interactions are characterized by a coordination between the O3 and
O4 hydroxyl groups of the fucose residue and the Ca\(^{2+}\) ion, leading to the classification of this protein as a C-type lectin (70). To confirm the requirement for Ca\(^{2+}\) in wild-type E-selectin binding and to define this dependence for the recombinant mutants, the glycan array screening experiments were repeated for each system in the presence of 10 mM ethylenediaminetetraacetic acid (EDTA). Under these conditions, no binding to any ligands was observed (SI Appendix, Fig. S9). This result would be expected for binding that depends on coordination of the fucose ring to the Ca\(^{2+}\) ion but suggested that some alternative Ca\(^{2+}\)-dependent mode of interaction must be present to explain loss of binding of 6'-sulfo-sLacNAc in the presence of EDTA.

To establish a molecular mechanism for the binding of 6'-sulfo-sLacNAc to the double mutant, an initial model for this complex was generated based on the structure of this mutant bound to 6'-sulfo-sLe\(^{\alpha}\). The fucose was then removed, and the complex was subjected to three independent MD simulations (200 ns). During the MD simulations, the ligand remained bound to the mutant but adopted a modified orientation (ring atom RMSD of 4 Å) compared to the initial ligand position (SI Appendix, Figs. S6 and S10), arising from a change in the conformation of the Neu5AcO2-3Gal linkage angles (SI Appendix, Fig. S11). In the absence of the fucose residue, the reoriented ligand retained its characteristic interactions between the sialic acid and the protein (SI Appendix, Table S6) but was now able to form a stable water-mediated interaction between the sulfate moiety and the Ca\(^{2+}\) ion. This interaction was detected by performing an analysis of high-occupancy water positions in the MD trajectory and was confirmed to be present in all three independent MD simulations (distance for Ca\(^{2+}\)--O (H2O) of 2.0 ± 0.1 Å; distance for O (H2O)--O (SO\(_3\)) of 3.0 ± 0.8 Å; SI Appendix, Fig. S12). For reference, a similar water-mediated interaction between a sulfate group and a Ca\(^{2+}\) ion has been reported in human annexin V [Protein Data Bank ID [PDBID] 1AVR (71)] with a Ca\(^{2+}\)--O (H2O) distance of 2.4 Å and a O (H2O)--O (SO\(_3\)) distance of 2.9 Å. The presence of a persistent interaction between the sulfate moiety and the Ca\(^{2+}\) ion, albeit water mediated, provides a rationalization for the observed Ca\(^{2+}\)-dependent nature of the binding of 6'-sulfo-sLacNAc and the double mutant and simultaneously explains the abolition of a requirement for fucose in the ligand. An equivalent interaction was not observed in the complex of the double mutant with 6'-sulfo-sLe\(^{\alpha}\).

A closer examination of the MD data indicated that the position of this mediating water molecule corresponds to that of one of the Ca\(^{2+}\)-coordinated water molecules in the apo-E-selectin crystal structure [water molecule S15 in PDBID 1ESL (72)] and also to the position that is occupied by fucose hydroxyl group O3 after binding to sLe\(^{\alpha}\) in the reported crystal structure (PDBID 4CSY). It thus appears that the double mutant displays two modes of ligand recognition, one that parallels the canonical C-type lectin binding of E-selectin with additional interactions between the sulfate moiety and residues N105, K111, and K113 and an alternative mode that may occur in the absence of fucose, wherein the sialic acid and sulfate moieties maintain their key interactions with the protein but in which the sulfo group additionally forms a water-mediated electrostatic interaction with the Ca\(^{2+}\) ion (SI Appendix, Table S6). This fascinating possibility will be the subject of further theoretical and experimental investigation.

**Orthogonal Binding Modes Display Equivalent Ligand Specificity.**

The selectivity displayed in the glycan microarray for E92A/E107A appeared to be indistinguishable from that reported for wild-type Siglec-8 (28, 73) (Fig. 3B), suggesting that the recognition of 6'-sulfo-sLe\(^{\alpha}\) has been engineered into E-selectin by the double mutation. Examination of the conformations for 6'-sulfo-sLe\(^{\alpha}\) in the MD simulations of the complexes with E92A/E107A and Siglec-8 showed that the glycocalyx linkages of the ligand in both complexes displayed the same distributions (SI Appendix, Figs. S1 and S6), indicating that the 6'-sulfo-sLe\(^{\alpha}\) ligand adopted the same conformation in each complex (Fig. 4). Remarkably, the engineered 6'-sulfo-sLe\(^{\alpha}\) recognition motif in the E92A/E107A is not equivalent to that in the wild-type Siglec-8 (Fig. 4). Whereas the binding of 6'-sulfo-sLe\(^{\alpha}\) to the E-selectin double mutant is driven predominantly by interactions with the sulfate moiety, the sialic acid, and the fucose (via coordination to a Ca\(^{2+}\) ion conserved across the selectins), in the complex with Siglec-8, the affinity arose primarily from interactions with the sulfate moiety and the sialic acid (Table 2). Similar to the wild-type E-selectin–sLe\(^{\alpha}\) complex, in the double mutant, the fucose contributed nearly half of the total binding energy from 6'-sulfo-sLe\(^{\alpha}\). By contrast, in the wild-type Siglec-8–6'-sulfo-sLe\(^{\alpha}\) complex, the fucose made a negligible contribution to binding, while the sialic acid residue contributed more than half of the total binding energy. The sulfate group in the Siglec-8–6'-sulfo-sLe\(^{\alpha}\) complex was predicted to contribute $-1.3 \pm 0.2$ kcal/mol to the total binding energy, which may be compared to a value of $-3.3 \pm 0.6$ kcal/mol for the interaction with the same moiety in the E92A/E107A mutant. The higher magnitude of the binding energy attributable to the sulfate in the double mutant complex is consistent with the observation that in the double mutant, the sulfate interacts with two charged residues (K111 and K113), while in the complex with Siglec-8, it interacts only with one charged residue (R56). Overall, the energy decomposition pattern correlated well with the presence of hydrogen bonds observed in the MD simulations (Table 1 and SI Appendix, Table S1).

**Discussion**

The present modeling study demonstrated the use of MD simulations and per-residue energy analyses to rationally guide the engineering of sulfate-binding specificity into a nonsulfate-binding protein. Specifically, the modeling was used to convert the specificity of E-selectin from a preference for its endogenous ligand (sLe\(^{\alpha}\)) to the 6'-sulfated form of the same oligosaccharide (6'-sulfo-sLe\(^{\alpha}\)). MD simulations predicted that E-selectin does not recognize 6'-sulfo-sLe\(^{\alpha}\) due to repulsions between the negatively charged side chains of E92 and E107 and the 6'-sulfo group. Additionally, simulations suggested that removing only one of the negatively charged side chains was not sufficient to stabilize binding of the sulfated ligand. However, a simultaneous double mutation, which eliminated the unfavorable repulsions by removing negatively charged side chains, reduced affinity to the endogenous sLe\(^{\alpha}\) ligand with a loss of a key hydrogen bond and introduced new favorable interactions to the 6'-sulfo group. These predictions were confirmed experimentally by screening the double mutant (expressed in HEK293 cells) against a glycan array containing more than 600 glycans, which confirmed not only that the double mutant exclusively bound to glycans terminating in the 6'-sulfo-sialyl motif but also that neither the wild-type E-selectin nor either of the single mutants showed this specificity.

The present study demonstrated a successful example of the rational design of a protein-based probe for a sulfated glycan through modeling the effects of site-directed mutagenesis. The work also led to the serendipitous discovery of a putative interaction mode for nonfucosylated 6'-sulfo-sLe\(^{\alpha}\). As predicted
computationally and confirmed experimentally, a double mutation was required to introduce the desired specificity. Notably, the computational analyses indicated that engineering the desired specificity required three components:

1) Removal of destabilizing steric and electrostatic interactions between the 6'-sulfate and E92 and E107.
2) Creation of favorable electrostatic interactions between the 6'-sulfo group and K111, K113, and N105, enabled by the E92A/E107A mutations.
3) Loss of a favorable hydrogen bond from Gal-O6 in the endogenous ligand to E92.

The first two components enhanced affinity for the novel ligand, while the third was required to eliminate affinity for the endogenous glycan. Indeed, it seems reasonable that introducing specificity required not only the creation of favorable interactions with the new ligand but also the introduction of interactions that disfavored binding to the endogenous ligand.

**Materials and Methods**

**Structure Preparation.** The initial coordinates for E-selectin-sLeX and Siglec-8-6'-sulfo-sLeX complexes were obtained from the PDB (entry codes 4CSY [57] and 2N7B [58], respectively). Chain A was extracted from the E-selectin-sLeX complex crystal structure, with the water molecules removed. Mutants of Siglec-8 (R109A, R56A, and K116A) and E-selectin (E92A, E107A, and E92A/E107A) were created by removing the extra atoms in the side chain of the corresponding residues. Both the addition of sulfate group to the O6 position of the galactose residue in sLeX and the removal of the fucose residue from sLeX in the E-selectin complexes were performed by using the tLEaP module in AMBER15 (75). Force
using the mirror image of the reported energy curves (61). Conformation of the ligand that was most similar to its average shape in the protein complex acquired from all three independent MD simulations was extracted and presented as its representative structure. The analyses of high-occupancy water positions in the MD simulation trajectories were performed with the visual molecular dynamics (VMD) volmap plugin (74), which computed the average densities of water molecules over all matrices of cubic voxels (a cell size of 0.5 Å).

Cloning and Protein Expression. The gene for human E-selectin (including residues 22 to 558) was designed to include the transferrin secretion signal and C-terminal human Fc and 8xHis tag and purchased from Genewiz. The resulting gene was cloned into pcDNA3.1 for expression in mammalian cells. The wild-type construct was used as a template to create the mutants E92A and E107A and the double mutant E92AE107A using Quickchange mutagenesis. HEK293 Freestyle cells were transiently transfected with the expression construct using polyethyleneimine, and culture supernatants were harvested 5 to 7 days after transfection. E-selectin were purified from the culture supernatants by nickel affinity chromatography and diazoylated against storage buffer (20 mM Tris-HCl, 250 mM NaCl, and 5 mM CaCl₂, pH 7.4) and flash frozen and stored at −80 °C until use.

Microarray Experiments. The E selectin proteins were run on consortium for functional glycomics (CFG) version 5.2 microarrays (85, 86). Microarray slides were rehydrated for 5 min in TSM buffer (20 mM Tris-HCl, 150 mM NaCl, 2 mM CaCl₂, and 2 mM MgCl₂) before adding 50 μg/mL of Fc-tagged human E-selectin in TSM binding buffer (TSM buffer with 1% bovine serum albumin). Microarrays were washed with TSM + 0.05% Tween-20, and bound selectins were detected with anti-human IgG-Alexa488 (Invitrogen) at 5 μg/mL. Because selectin binding is Ca²⁺ dependent, control experiments for all variants were performed, which included 10 mM EDTA in the binding buffer instead of CaCl₂. Microarray slides were scanned with a Genepix 4300A (Molecular Devices) and quantified with Genepix Pro 7.0 software. The results are shown as relative fluorescent units by averaging the background-subtracted signals of the four replicate spots (after throwing out the highest and lowest value of the six printed spots), with error bars representing the SD of the four averaged values.

Data, Materials, and Software Availability. All study data are included in the article and/or SI Appendix.

ACKNOWLEDGMENTS. We would like to thank Dr. Oliver C. Grant at the University of Georgia for insightful discussions. We would also like to thank Dr. Jamie Heimberg Molinaro and Alyssa McQuillan for glycan microarray experiments and analyses. R.J.W. gratefully acknowledges support from the NIH (1R24GM136984) and the NSF (CHE2002628). This work was also supported by NIH grants P41GM103694 and R24GM137763 to R.D.C. This work was supported in part by GlycoMIP, a National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-193325.

1. K. L. Moore, The biology and enzymology of protein tyrosine O-sulfation. J. Biol. Chem. 278, 24243–24246 (2003).
2. S. M. Muthana, C. T. Campbell, J. C. Gildersleeve, Modification of glycans: Biological significance and therapeutic opportunities. ACS Chem. Biol. 7, 31–43 (2012).
3. C. I. Gama et al., Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. Nat. Chem. Biol. 2, 467–473 (2006).
4. I. H. K. Dias et al., Sulfate-based lipids: Analysis of healthy human fluids and cell extracts. Chem. Phys. Lipids 215, 52–63 (2018).
5. H. Horike et al., Paramodulation junction formation and spermatogenesis require sulfoglycolipids. Proc. Natl. Acad. Sci. U.S.A. 99, 4227–4322 (2002).
6. I. Fernandez-Vega et al., Specific genes involved in synthesis and editing of heparan sulfate proteoglycans show altered expression patterns in breast cancer. BMC Cancer 13, 24 (2013).
7. C. Ricciardelli et al., Elevated stromal chondromodulin sulfate glycosaminoglycan predicts progression in early-stage prostate cancer. Clin. Cancer Res. 5, 983–992 (1999).
8. B. Xu, J. A. Royall, G. Damera, G. D. Pacheco, R. D. Cummings, Altered O-glycosylation and sulfation of amnion moxes associated with cotic fibrosis. Glycobiology 15, 747–777 (2005).
9. G. Sundblad, S. Kajulla, V. Quanta, H. H. Freire, A. Varki, Sulfated N-linked oligosaccharides in mammalian cells. III. Characterization of a pancreatic carcinoma cell surface glycoprotein with N- and O-sulfate esters on asparagine-linked glycans. J. Biol. Chem. 263, 8977–8993 (1988).
10. T. P. Mawhinney, E. Adelstein, D. A. Morris, A. M. Mawhinney, G. J. Barbero, Structural determination of five sulfated oligosaccharides derived from tracheobronchial mucus glycoproteins. J. Biol. Chem. 262, 2994–3001 (1987).
11. A. H. K. Plass, L. A. West, S. W. Pianigiani, F. T. R. Nelson, Glycosaminoglycan sulfation in human osteoarthritis. Disease-related alterations at the non-reducing terminus of chondroitin and dermatan sulfate. J. Biol. Chem. 273, 12642–12649 (1998).
12. M. T. Bayless, D. Oseborne, S. Woodhouse, C. Davidson, Sulfation of chondroitin sulfate in human articular cartilage. The effect of age, topographical position, and zone of cartilage on tissue composition. J. Biol. Chem. 274, 15982–15990 (1999).
13. M. Leu, M. V. Novotny, Y. Mechref, Sequential enrichment of sulfated glycans by sequential double-permethylation using methyl iodide and deuteromethyl iodide. J. Am. Soc. Mass Spectrom. 21, 1660–1671 (2009).
14. X. Bai, J. R. Brown, A. Varki, J. D. Eko, Enhanced O-sulfation of galactose on Asn-linked glycans and Maackia amurensis lectin binding in a new Chinese hamster ovary cell line. Glyobiology 11, 621–632 (2001).
15. W. C. Wang, R. D. Cummings, The immobilized leukoagglutinin from the seeds of Maackia amurensis binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked alpha 2-3 to penultimate-galactose residues. J. Biol. Chem. 263, 4526–4535 (1988).
16. H. Tateno et al., Dual specificity of Langerin to sulfated and mannosylated glycans via a single C-type carbohydrate recognition domain. J. Biol. Chem. 265, 6390–6400 (2010).
17. W. Gao, T. Xu, J. Liu, M. Ho, Epitope mapping by a Wnt-blocking antibody: Evidence of the Wnt binding domain of hexameric heparin and heparan sulfate. J. Biol. Chem. 278, 41143–41151 (2010).
18. N. C. Smits et al., The heparan sulfate motif (GlcNAc6S-IDa2G2S), common in heparin, has a strict topography and is involved in cell behavior and disease. J. Biol. Chem. 285, 41143–41151 (2010).
33. T. R. McKitrick
21. K. Ohmori et al.
45. R. Yabe
41. S. Thamotharan et al.
44. I. C. Michelow
46. K. Yamamoto, I. N. Maruyama, T. Osawa, Cyborg lectins: Novel leguminous lectins with unique biological activity, Biochim. Biophys. Acta 1521, 1-15 (2000).

25. J. Jung et al.
49. R. Arango et al.
51. W. S. Somers, J. Tang, G. D. Shaw, R. T. Camphausen, Insights into the molecular basis of P-selectin ligand recognition by natural evolution-mimicry, J. Biol. Chem. 280, 137-143 (2005).

29. R. P. Schleimer, R. L. Schnaar, B. S. Bochner, Regulation of airway inflammation by Siglecs, glycans, and other glycan-binding proteins, J. Allergy Clin. Immunol. 135, 535-543 (2015).

16. W. F. Vangunsteren, H. J. C. Berendsen, Algorithms for macromolecular dynamics and constraint optimizations, J. Comput. Chem. 15, 676-689 (1994).

13. A. Sood et al., Method for estimating the conformational entropy of a protein, J. Phys. Chem. B 106, 7392-7400 (2002).

22. R. Aragon et al., Modification by site-directed mutagenesis of the specificity of Erythrina corallodendron lectin for galactose-binding protein by natural evolution-mimicry, J. Biol. Chem. 141, 389-399 (2007).

24. K. Yamamoto, I. N. Maruyama, T. Osawa, Cyborg lectins: Novel leguminous lectins with unique biological activity, Biochim. Biophys. Acta 1521, 1-15 (2000).

23. A. V. Kuznetsov et al., Insight into the binding specifity of Siglec-9 ligands, J. Mol. Biol. 352, 771-788 (2005).

26. B. Ernst, Consortium for Functional Glycomics, GLYCAM06: A generalizable biomolecular force field for carbohydrates, J. Chem. Phys. 121, 28 (2004).

27. A. Gonzalez-Gil et al., Sialylated and sulfated carbohydrate ligands for selectins and siglecs: Structure, cell interaction and homing of diverse glycan binding proteins, Adv. Exp. Med. Biol. 645, 137-142 (2009).

28. M. Karplus, J. N. Kushick, Method for estimating the conformational entropy of a protein, J. Phys. Chem. B 106, 7392-7400 (2002).

29. K. Olomoni et al., P- and E-selectins recognize sialyl 6-sulfo Lewis X, the recently identified L-selectin ligand, Blood 107, 4156-4159 (2006).

30. T. R. McKitrick et al., Synthetic inhibitors of cell adhesion: A glycopeptide from E-selectin ligand 1-1(E3L)-1 with the arabinosyl sialyl Lewis(x) structure, Angew. Chem. Int. Ed. Engl. 40, 3836-3839 (2001).

31. A. Hidalgo, A. J. Perier, M. Wild, D. Vestweber, Complete identification of E-selectin ligands on neutrophils reveals distinct functions of PSGL-1, E-1, and CD44, Immunity 26, 677-689 (2007).