



OPEN

# Rational identification and characterisation of peptide ligands for targeting polysialic acid

Divya G. Shastry<sup>1,2</sup>✉, Flaviyan Jerome Irudayanathan<sup>3</sup>, Asher Williams<sup>4</sup>, Mattheos Koffas<sup>2,4</sup>, Robert J. Linhardt<sup>2,4,5,6</sup>, Shikha Nangia<sup>3</sup> & Pankaj Karande<sup>2,4</sup>✉

The alpha-2,8-linked form of the polysaccharide polysialic acid (PSA) has widespread implications in physiological and pathological processes, ranging from neurological development to disease progression. Though the high electronegativity and excluded volume of PSA often promotes interference of biomolecular interactions, PSA-binding ligands have important implications for both biological processes and biotechnological applications. As such, the design, identification, and characterisation of novel ligands towards PSA is critical for expanding knowledge of PSA interactions and achieving selective glycan targeting. Here, we report on a rational approach for the identification of alpha-2,8-PSA-binding peptides, involving design from the endogenous ligand Siglec-11 and multi-platform characterisation of peptide binding. Microarray-based examination of peptides revealed charge and sequence characteristics influencing peptide affinity to PSA, and carbohydrate–peptide binding was further quantified with a novel fluorescence anisotropy assay. PSA-binding peptides exhibited specific binding to polymeric SA, as well as different degrees of selective binding in various conditions, including competition with PSA of alternating 2,8/9-linkages and screening with PSA-expressing cells. A computational study of Siglec-11 and Siglec-11-derived peptides offered synergistic insight into ligand binding. These results demonstrate the potential of PSA-binding peptides for selective targeting and highlight the importance of the approaches described herein for the study of carbohydrate interactions.

Polysialic acid (PSA) is a unique polysaccharide playing critical roles in a number of physiological and pathological processes. PSA occurs naturally in a variety of forms, comprising different sialic acid monomers and glycosidic linkages<sup>1,2</sup>. Of the three homopolymeric forms of N-acetylneuraminic acid (the most common sialic acid in humans), only  $\alpha$ -2,8-linked PSA is found in our species<sup>1–3</sup>. The roles of PSAs across biology are diverse; humans display highest PSA expression on neural stem and progenitor cells, where expression primarily guides neurological development, but in other systems and organisms, this biopolymer is implicated in bypassing immune surveillance, promoting tissue development, enabling cell migration, and even possibly serving as a molecular reservoir<sup>1–14</sup>. Changes in polysialylation on proteins and cell surfaces are often tied to functional states. For example, change in PSA expression during human embryonic development influences neuronal growth and migration, while in adults, long-chain PSA is largely restricted to regions of the nervous system undergoing remodelling<sup>1–5</sup>. In pathological context, PSA is of great interest due to its occurrence on pathogenic bacteria (*e.g.*, *Neisseria meningitidis* groups B and C), where expression in the polysaccharide capsule enables evasion of the host immune system<sup>1,2,9–12</sup>. Additionally, PSA has been found on tumour cells, and its expression has been correlated with poorer prognosis of certain cancers, possibly due to an increase in metastatic potential<sup>1,2,13</sup>. Despite the broad diagnostic and therapeutic application space available with the ability to understand, target, and detect PSA, *in vitro* or *in vivo*, few studies have sought to develop and characterise robust ligands against PSA.

<sup>1</sup>Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA. <sup>2</sup>Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA. <sup>3</sup>Department of Biomedical and Chemical Engineering, Syracuse University, Syracuse, NY, 13244, USA. <sup>4</sup>Howard P. Isermann Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA. <sup>5</sup>Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA. <sup>6</sup>Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA. ✉e-mail: [shastd3@gmail.com](mailto:shastd3@gmail.com); [karanp@rpi.edu](mailto:karanp@rpi.edu)

Several naturally occurring proteins have been identified to bind to PSA, including neurotrophins and poly-sialyltransferases<sup>1,2</sup>. However, information (e.g., residue-level and structural) about direct interaction of PSA with endogenous ligands is limited<sup>14–23</sup>. Additionally, challenges in protein isolation and stability deter characterisation and applicability of such ligands for PSA targeting. The generation of anti-PSA antibodies, together with their biochemical characterisation (as of mAb735), has proved useful<sup>2,21,22</sup>. Yet, as with native ligands, factors including production, size, and stability may limit the versatility of antibodies in applications; these factors comparatively promote the versatility of low molecular weight ligands. Besides such advantages for the latter ligands (described elsewhere<sup>24</sup>), the identification of small targeting agents, like peptides, for PSA can add to basic knowledge of PSA molecular interactions, especially when rational design strategies are incorporated in ligand development. Furthermore, ligand design with a focus on elucidating binding characteristics can provide critical information for iterative design and future targeting efforts.

In this study, we report on the design and identification of  $\alpha$ -2,8-PSA-binding peptides, with biochemical and thermodynamic characterisation of peptide binding and insight into PSA interaction with the endogenous ligand Siglec-11<sup>23</sup>. We have previously identified peptide ligands binding to  $\alpha$ -2,8-PSA, demonstrating the success of a high-throughput peptide microarray platform for evaluating design strategies and supporting preliminary peptide characterisation<sup>25</sup>. Here, we examined amino acid sequences derived from a naturally occurring lectin to PSA in tandem with previously identified peptides, while simultaneously expanding on peptide characterisation using a synergistic experimental and computational approach.

## Results

**Library-level examination of peptide–PSA interaction.** Peptides (762 unique sequences) were screened for binding to  $\alpha$ -2,8-PSA using a high-throughput microarray screening platform<sup>25</sup>. A library of peptides was designed from linear epitope mapping of the protein Siglec-11 (all domains) with peptides 15 residues in length overlapping by 13 residues (329 peptides; 1 previously reported<sup>25</sup>). Siglec-11 is a member of the highly homologous Siglec family of sialic acid binding-proteins with 2,8-linkage specificity<sup>23,26–28</sup>. An additional 172 peptides (1 previously reported<sup>25</sup>) were designed based on Siglec-derived sequences. Modifications to these peptides included select rational mutations and sequence scrambling. Siglec-modified sequences designed with preliminary binding hypotheses were synthesized alongside the parent library; other sequences based on screening of the original library were synthesized iteratively, where binding and non-binding peptides were modified to study variation in target affinity and selectivity. Random sequences from *de novo* design and from prior reports in literature<sup>29</sup> (38 peptides), as well as PSA-binding and non-binding peptides previously designed from mAb735 and phage display screening (223 peptides)<sup>25</sup>, were concurrently screened for intra-assay comparison to binding of Siglec-derived peptides.

Sequences of peptides exhibiting the highest (approximately top 5%) binding intensities from the complete peptide library are provided in Table 1. As expected, all sequences display a prevalence of positively charged residues. The charge dependence on binding at the library-level is apparent from Fig. 1, which displays an increase in microarray binding with higher peptide basicity and charge. However, a few neutral and negatively charged peptides display measurable affinity towards PSA, and not all positively charged peptides interact with PSA; this suggests that observed binding cannot be attributed to non-specific electrostatic interactions alone. Differentiating peptides based on binding, charge, and origin does not indicate that of the various peptide development strategies chosen, one provides a distinct advantage in increasing charge-based peptide affinity (Supplementary Fig. S1).

The relationship between basic residues and PSA binding is supported by compositional and positional analyses of sequences of high affinity peptides. Figure 2 and Supplementary Fig. S2 display statistically significant increases in basic residues in the top 5% of binders. Several residues show agreement with our prior work on mAb and phage display-derived peptides (specifically, significant increases in the prevalence of arginine, lysine, and phenylalanine and decrease in that of serine)<sup>25</sup>. However, changes in the occurrence of asparagine and glycine were reversed; here, asparagine showed significant decrease and glycine showed significant increase. These differences are likely due to examination of a larger peptide library in this study, as well as inclusion of a larger number of non-phage peptides (lacking the inherent biases in residue propensity observed in phage-derived lead candidates<sup>30</sup>) and restriction of analyses to the top 5% of binders (as compared to the top 10% reported previously).

**Fluorescence anisotropy assay for assessment of peptide binding affinity.** A fluorescence anisotropy (FA) assay was developed for the determination of peptide–PSA binding affinity. In FA assays, in-solution binding of an analyte to a smaller, fluorescently-labelled ligand is quantified through titrations with increasing concentrations of the analyte<sup>31</sup>. FA assays have been used to study various interactions of proteins, DNA, carbohydrates, and small molecules<sup>31–34</sup>. However, use of these assays to study peptide–carbohydrate interactions has been limited. In the experiments described here, anisotropy should be interchangeable with polarization; anisotropy is employed as it is normalized by total intensity<sup>31</sup>.

Nine peptides of different origins were selected for study with the FA assay, and binding affinities of fluorescently labelled-peptides to PSA were determined (Fig. 3(a) and Table 2). Peptides were selected, in part, to represent a range of PSA-binding abilities in microarray studies. Binding analysis of I-P50, a peptide from the putative binding region of Siglec-11, could not be performed due to high insolubility of the peptide in aqueous media (peptides modified from I-P50 to increase hydrophilicity also displayed poor solubility). Under the conditions used, peptides bound PSA in approximately 1:1 ratio. Given this stoichiometry, together with the high conformational flexibility of peptides and PSA<sup>35</sup> (as compared to proteins or monosaccharides), the mid-to-high micromolar affinity values obtained are not unexpected.

$K_D$  determinations confirmed microarray-based classification of peptides as PSA-binding or non-binding. Peptide affinity from microarray studies represents intensity values and is not equivalent to affinity constants; hence, it is expected that microarray affinity does not correspond exactly with  $K_D$  values. Microarray intensities

Peptide index	Peptide sequence	Peptide origin
I-P23	YWF <b>K</b> GRTSPKTGAPV	Siglec-11
I-P24*	FKGRTSPKTGAPVAT	Siglec-11
I-P59	LSNAFFLKVTALTKK	Siglec-11
I-P60	NAFFLKVTALTKKPD	Siglec-11
I-P62	LKVTALTKKPDVYIP	Siglec-11
I-P85	AALSPRRTRPSTSHF	Siglec-11
I-P86	LSPRRTRPSTSHFSV	Siglec-11
I-P102	VDFSRKGVSAQRTVR	Siglec-11
I-P103	FSRKGVSQRTVRLR	Siglec-11
I-P104	RKGVSQRTVRLRVA	Siglec-11
I-P143	WGPRTLGLRLGVRA	Siglec-11
I-P278	FRVKICRKEARKRAA	Siglec-11
I-P279	VKICRKEARKRAAAE	Siglec-11
I-P330	WFKGRTSPKTGAPVA	Siglec-11
I-P333	WFKG <b>K</b> TSPKTGAPVA	Siglec-11: R5K mutation
I-P378	KGKGKGGKGGKGGKGGK	<i>De novo</i>
I-P379	KGGGKGGKGGKGGKGGK	<i>De novo</i>
I-P380	GGGKGGKGGKGGKGGK	<i>De novo</i>
II-P35	GSGSGTDFTLKISR	mAb735
II-P50	VPYTFGGGTRLEIKG	mAb735
II-P77	PGSGNTKYNEKFKGK	mAb735
II-P79	NTKYNEKFKGKATLT	mAb735
II-P213	AISSPLLRNPFRRGGGS	Phage display screening: W8R mutation
II-P214*	AISSPLLRNPFRRGGGS	Phage display screening: W8K mutation
II-P336*	NRTVLRNLNGTSLP	Siglec-11: E6R mutation

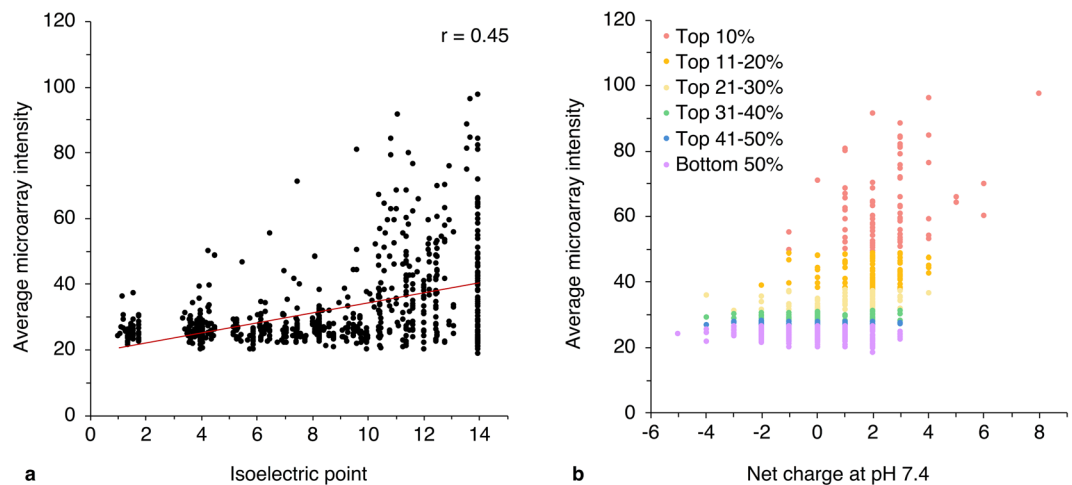
**Table 1.** Sequences and origins of 25 high-binding peptides from microarray screening against  $\alpha$ -2,8-polysialic acid. Peptides shown exhibit binding intensities in the top 5% in three independent screens, with triplicate measurements within each screen and inter-assay coefficients of variation <25% (peptides exhibiting intensities in the top 5% with higher inter-assay CVs excluded). Bolded residues represent mutations from parent peptides. \* Peptides with selectivity >80%.

show strong correlation to anisotropy values (and not  $K_D$  itself) at higher PSA concentrations, with  $r \approx 0.8$  at concentrations  $\approx 100 \mu\text{M}$  (at lower concentrations, higher relative error in anisotropy prevents accurate assessment of correlation). I-P264 was used as a negative control<sup>25</sup> and displays the weakest binding, with  $K_D > 1.5 \text{ mM}$  (binding does not approach saturation even at the solubility limit of PSA). Along with microarray screening, SPR spectroscopy<sup>25</sup> and molecular dynamics (MD) simulations (see below) corroborated the non-binding property of this peptide.

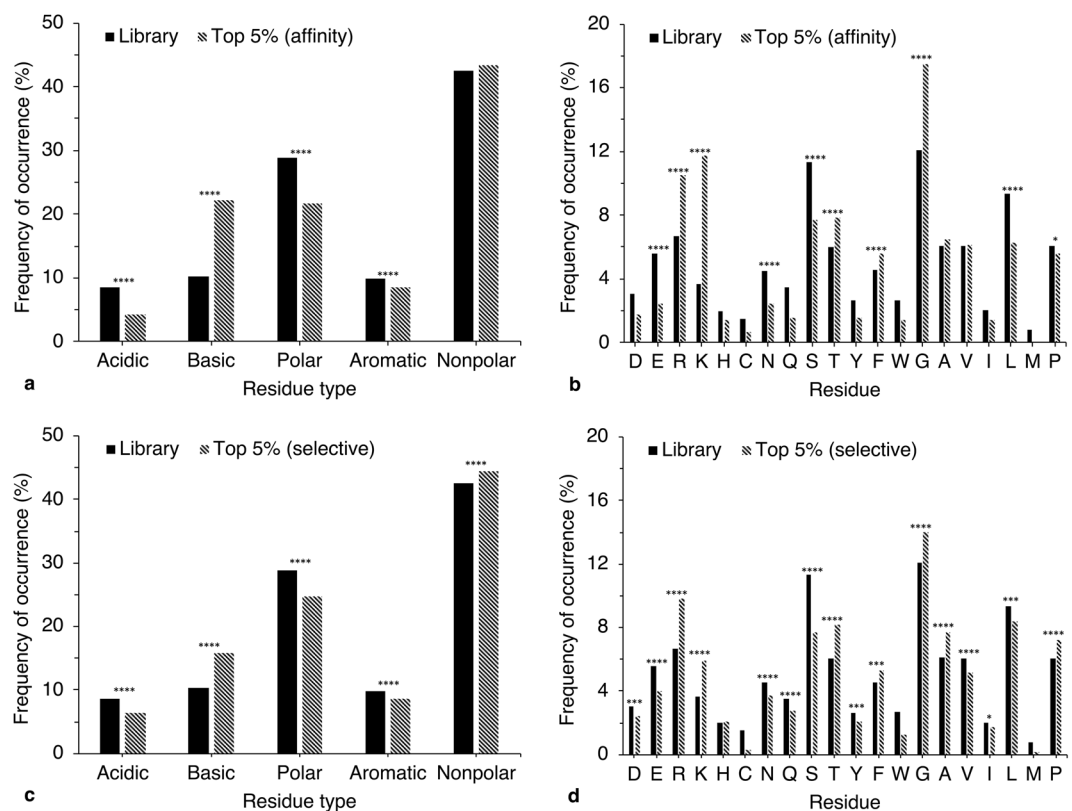
The ratio  $B_{\text{max}}/K_D$  shows strong correlation to microarray intensities ( $r = 0.77$ ; Fig. 3(b)). Though  $n$  is low, this relationship demonstrates that the identification of PSA-binding peptides through selection of high intensity binders on microarrays is likely to isolate peptides of moderate-to-high “binding potential,” *i.e.*, peptides with high  $B_{\text{max}}$  and/or low  $K_D$ . Hence, quantification of binding potential effectively enables selection of a peptide candidate pool which captures two potentially desired properties in peptide ligand applications. (The term binding potential is derived from PET imaging; use of the term here does not indicate any relationship to this technique<sup>36</sup>). Of the peptides characterised with FA, I-P24 and II-P336 notably display the highest  $B_{\text{max}}/K_D$  ratios and reproducibly high microarray affinity intensities (Fig. 3(b)). Since  $B_{\text{max}}/K_D$  is roughly equivalent to the initial slope of the binding curve, a higher ratio generally corresponds to a greater response with a smaller amount of target, which may be useful in assays where peptide ligand sensitivity is critical.

**Assessment of peptide selective binding to  $\alpha$ -2,8-PSA.** Competitive microarray screening with  $\alpha$ -2,8/9-PSA was conducted in order to assess selective binding of peptides to  $\alpha$ -2,8-PSA. The definition of percent selectivity at 10% competing glycan was applied to enable selectivity comparisons amongst a larger set of peptides<sup>25</sup> (nearly all peptides displayed minimal binding to  $\alpha$ -2,8-PSA with equimolar  $\alpha$ -2,8/9-PSA). In contrast to peptide affinity analysis, where higher positive charge and affinity generally correlated, selectivity analysis of high affinity peptides indicated that affinity is not sufficient to enable selective binding. We have previously demonstrated this for selective binding of peptides to PSA over chondroitin sulphate<sup>25</sup>. However, the use of an isomeric polymer is a more stringent assessment of selectivity; here, selective binding is guided by three-dimensional conformation and not additionally by differences in functional groups.

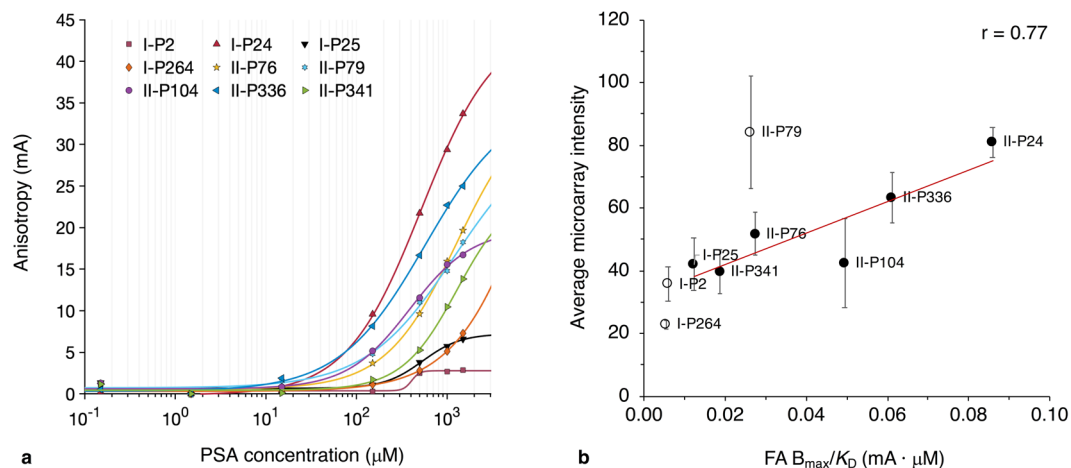
Of the high affinity peptides in Table 1, peptides with selectivity greater than 80% are I-P24, II-P214, and II-P336. (Selectivities of high affinity peptides and of peptides with selectivities >80% are provided in Supplementary Table S3.) These three peptides have different origins. I-P24 is derived from the N-terminal binding domain of Siglec-11. II-P214 was designed with a lysine point mutation to modify binding of a



**Figure 1.** Binding of peptides to  $\alpha$ -2,8-PSA based on (a) isoelectric point (pI) or (b) net charge at pH 7.4. (a) Peptide binding intensity with pI displays a very weak correlation ( $r = 0.45$ ) due to a large number of high pI peptides demonstrating lack of interaction with PSA. All peptides plotted with a pI of 14 displayed the output “>14” with EMBOSS iep. (b) Charge-based binding for 10% intensity divisions of the peptide library are shown. The outlying point with net charge = 8 represents a *de novo* designed peptide with alternating Lys and Gly residues. Binding intensities represent the mean of three independent experiments, with triplicate intra-assay measurements (error bars excluded for clarity).



**Figure 2.** Compositional analysis of high affinity (a,b) and high selectivity (c,d) peptides. Occurrence of residue types (a,c) or specific residues (b,d) in the peptide library ( $n = 762$ ) is compared to occurrence in approximately the top 5% affinity or selectivity peptides ( $n = 38$  or 41, respectively). Acidic = D and E; basic = R and K; polar = H, C, N, Q, and S; aromatic = Y, F, and W; and nonpolar = G, A, V, I, L, M, and P. (Two-tailed  $z$  test for population proportions; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ , two-tailed;  $p$  values provided in Supplementary Table 1). Statistical significance of frequency changes of the following residues was not determined as low residue occurrence in the sample population precluded the assumption of normal distribution: D, H, C, Q, Y, W, I, and M (b) and C, W, and M (d).



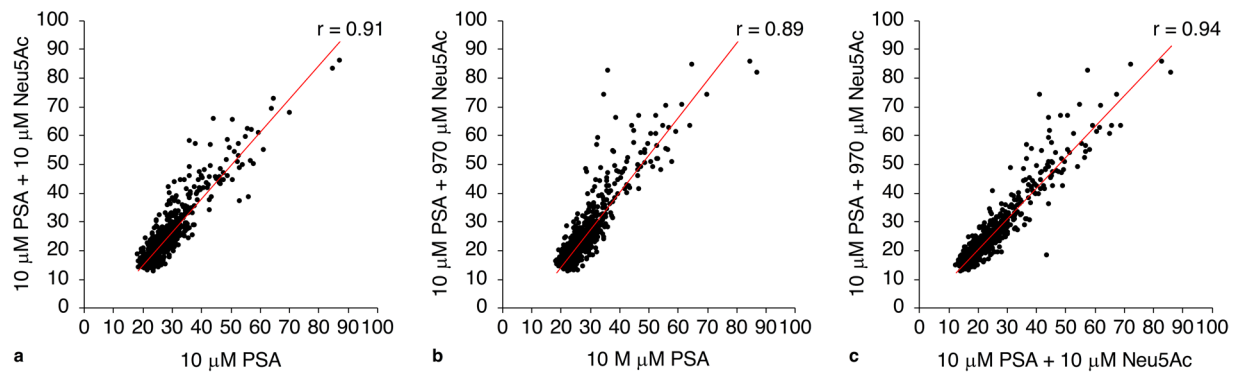
**Figure 3.** (a) Interaction of PSA-binding and non-binding peptides with PSA as assessed by fluorescence anisotropy titrations (solubility limit of PSA  $\sim$ 1.5 mM). Anisotropy values were transformed by subtraction of minimum FA values for each curve (approximately equivalent to anisotropies of free peptides) to enable comparative analysis between peptides. Error bars are excluded for clarity; experimental errors (standard deviations) are provided in Supplementary Table S2 along with non-adjusted anisotropy values. (b) Relationship of microarray “affinity” (intensity values) to peptide  $B_{\max}/K_D$  (from FA assay). Open circles represent peptides excluded from determination of Pearson correlation coefficient due to higher imprecision in  $B_{\max}$  and  $K_D$  from anisotropy curve fits for these peptides (criteria for exclusion: fit standard error of  $B_{\max}$  or  $K_D$   $>$  fit value); values for these peptides are represented for comparison of microarray intensities to other peptides. Microarray intensities represent the mean of three independent experiments, with triplicate measurements within each screen. FA  $K_D$ 's are from at least two independent experiments, with at least triplicate measurements within each assay. Error from the assay with greater inter-assay variability (*i.e.*, microarray screening) is shown (inter-assay standard deviation).

Peptide index	Peptide sequence	Peptide origin	$K_D$ (mM)	SE (mM)	$R^2$
I-P2	KDPSYSLQVQRQVPV	Siglec-11	0.405*	13.896	0.955
I-P24	FKGRTSPKTGAPVAT	Siglec-11	0.521	0.077	1.00
I-P25	GRTSPKTGAPVATNN	Siglec-11	0.536	0.174	0.977
I-P264	EHGGGLGLGAALGAG	Siglec-11	9.85*	53.43	0.994
II-P76	IYPGSGNTKYNEKFK	mAb735	1.29	0.70	0.998
II-P79	NTKYNEKFKGKATLT	mAb735	1.20*	1.31	0.995
II-P104	HLSLKNPLRMDLGGGS	Phage display screening	0.388	0.091	0.997
II-P336	NRTVLRNLGNGTSLP	Siglec-11: E6R mutation	0.577	0.164	0.999
II-P341	NRTVLEKLGNGTSLP	Siglec-11: N7K mutation	1.29	0.886	0.995

**Table 2.** Binding affinity constants ( $K_D$ 's) of peptides as determined by FA assays. SE = standard error of the fit (for  $K_D$ ) with 95% confidence bounds.  $R$  square values of curve fits are also provided. Bolded residues represent mutations from parent peptides. Accuracy of values marked \* is likely poor due to high imprecision ( $SE > K_D$  for II-P79 and  $SE \gg K_D$  for I-P2 and I-P264); corresponding peptides are excluded from correlation in Fig. 6(b).

previously discovered non-binding peptide from phage display screening. Similarly, II-P336 was designed from a non-binding Siglec-derived peptide through substitution of glutamic acid with arginine. In the second two cases (where poor binders were modified), distinctly separated positive charges were incorporated in peptide regions lacking positive charge; charge spacing was based on preliminary hypotheses from principles of lectin-carbohydrate binding<sup>37–42</sup>. Of these three peptides, only I-P24 consistently displayed high affinity binding in the top 2% of microarray intensities. Other than I-P24 and II-P336, peptides characterised with FA showed moderate to poor binding to  $\alpha$ -2,8-PSA in the presence of different concentrations of  $\alpha$ -2,8/9-PSA (Supplementary Fig. S2). Interestingly, I-P24 and II-P79 (from mAb735 and also assessed with FA) demonstrated markedly differing affinity and selectivity responses as compared to their respective overlapping peptide sequences (Supplementary Fig. S2), suggesting sequence-specific PSA binding by these peptides.

Peptides of higher selectivity have highly similar residue composition to peptides of higher affinity, except for a relatively lower propensity of lysine in selective sequences (Fig. 2). This similarity is expected as some degree of affinity is necessary for selective binding to occur, and hence, selective binding was assessed in peptides displaying binding above background level. However, positional occurrences of residues differ in high affinity and high selectivity sequences. In selective sequences, some positions display a reduction in positively charged residues,



**Figure 4.** Selective binding of peptides to polymeric sialic acid over sialic acid monomer. Mean microarray intensities are compared between peptide binding to (a) 10  $\mu\text{M}$  PSA and 10  $\mu\text{M}$  PSA with equimolar (10  $\mu\text{M}$ ) Neu5Ac, (b) 10  $\mu\text{M}$  PSA and 10  $\mu\text{M}$  PSA with equivalent mass (970  $\mu\text{M}$ ) of Neu5Ac, and (c) 10  $\mu\text{M}$  PSA with equimolar Neu5Ac and 10  $\mu\text{M}$  with equivalent mass Neu5Ac. Given that PSA contains an average of 100 sialic acid units, competition of a mass equivalent of Neu5Ac represents a more stringent condition than equimolar competition. Binding intensities represent the mean of two independent experiments, with triplicate intra-assay measurements (error bars excluded for clarity).

while others show an increase in small, hydrophobic residues (Supplementary Fig. S3 and Supplementary Table S4). This aligns with the observation that the highest affinity peptides do not necessarily bind selectively to PSA; such sequences contain numerous positively charged residues that may bind indiscriminately to any negatively charged target. Thus, increased net charge promotes peptide affinity to PSA, but may not enhance selective binding. In fact, *de novo* peptides composed of lysine and glycine residues display high affinities but mediocre selectivity (approximately 55–60%, e.g., I-P378–I-P380 in Table 1). With the additional consideration that selectivity quantification was performed with only 10% competing PSA, these highly basic peptides provide poor examples of applicable selectivity.

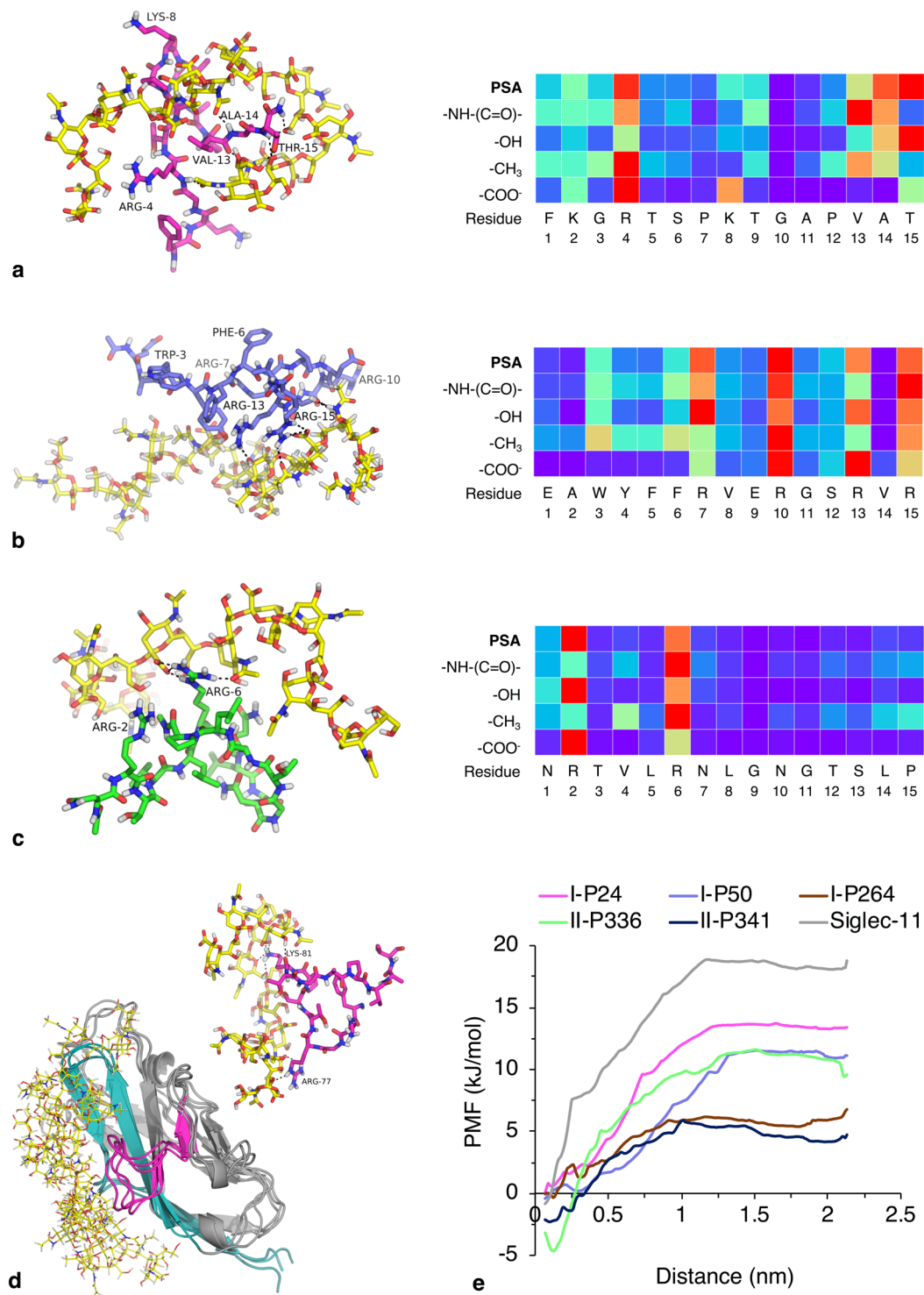
In contrast to the narrow definition of selectivity above, specificity of peptides to polymeric  $\alpha$ -2,8-PSA was demonstrated through competitive microarray screening with N-acetylneuraminic acid (Neu5Ac or sialic acid monomer). Peptide library binding to PSA in the presence of either molar or mass equivalents of Neu5Ac showed strong or very strong correlations (Fig. 4). As proof-of-principle, the FA assay developed for peptide–PSA affinity assessment was applied for I-P24 and I-P264 with Neu5Ac, and the specific interaction of I-P24 with the polymer was confirmed (Supplementary Fig. S4).

Variation in binding between Neu5Ac conditions largely fell within assay error. However, a few peptides demonstrated higher binding to PSA with a mass equivalent of Neu5Ac, possibly due to effects of solution charge and ionic strength on binding. The microarray technique likely lacks the sensitivity needed for systematic evaluation of such effects (*i.e.*, compared to techniques such as FA or isothermal calorimetry [ITC]<sup>43</sup>); however, preliminary assessment of selective binding in differing buffer conditions is possible. Affinity screening of a sub-set of peptides with 100 mM NaCl (the standard buffer concentration) and with 200 mM NaCl did not reveal a clear trend for altered binding with higher ionic strength ( $r = 0.84$ ). I-P2, I-P24, I-P25, and I-P264 (tested with FA) and I-P50 (discussed further below) were included in this sub-set, and these peptides did not display any change in binding (within error). In contrast, nearly all peptides showed minimal binding to PSA when the affinity screen was conducted without NaCl; I-P24 was one of the few peptides displaying detectable binding. It is possible that this large decrease in binding arises from elimination of a positive entropic contribution, similar to the energetic contribution previously described for binding of heparin to peptides, where sodium ion displacement entropically drives the carbohydrate–peptide interaction<sup>37,44</sup>.

Peptide selective binding to PSA at different pH values was also evaluated through microarray screening. Most peptides, including all FA-tested peptides, were largely insensitive to changes in pH in the range 6.0–8.0 (Supplementary Fig. S5). Peptides exhibiting the most prominent differences between conditions contain histidine residues, for which change in protonation at lower pH expectedly enhances PSA interaction (Supplementary Fig. S5).

**Molecular dynamics simulations of PSA–peptide and PSA–Siglec-11 interactions.** MD simulations of PSA–peptide and PSA–Siglec-11 interactions were performed to investigate thermodynamic and structural aspects of peptide–PSA binding, as well as the relationship of this binding to protein–PSA binding (Fig. 5). Simulations of the five peptides I-P24, I-P50, I-P264, II-P336, and II-P341 (all derived from Siglec-11; Table 2) with  $\alpha$ -2,8-PSA largely corroborated microarray and FA observations on PSA–peptide binding, the latter for all but I-P50 (for which poor solubility hindered experimental binding studies). For PSA-binding peptides, interactions were primarily guided by positively charged residues, which align with sialic acid carboxyl groups. Additionally, proximal residues appeared capable of forming hydrogen bonds and Van der Waals contacts with PSA. I-P264, the negative control peptide, showed minimal interaction with PSA.

Simulations of the N-terminal binding domain of Siglec-11 (homology model) with PSA demonstrated target interaction with CC' and GG' loops of the immunoglobulin domain, as well as with the F, G, and G' strands; the arginine conserved amongst Siglecs projects into the binding pocket from the hydrophobic F strand (part of the



**Figure 5.** Interactions of peptides and Siglec-11 with  $\alpha$ -2,8-PSA as assessed by MD simulations. Simulation snapshots from the time point of greatest interaction (with a cut-off distance for interaction set at 3.0 Å) for (a) I-P24, (b) I-P50, and (c) II-P336 with  $\alpha$ -2,8-PSA (decaisialic acid) (1:1) are shown on the left. Polar contacts at the time point of the snapshots are shown as black dashes, and residues for which highest interaction occurs over the course of the complete simulation trajectory are labelled. The carbohydrate is shown in yellow, and peptides I-P24, I-P50, and II-P336 are shown in magenta, blue, and green, respectively. Corresponding residue interaction maps (right) show the normalized number of contacts (3.0 Å cut-off distance) from 0 (purple) to 1 (red) observed over the complete time course of simulation between each peptide and decaisialic acid. II-P341, a moderate to weak binder in microarray and FA assays, did not show PSA binding in 1:1 simulation, but demonstrated binding in other simulations (see Supplementary Information). (d) Siglec-11 N-terminal domain (homology model) with  $\alpha$ -2,8-PSA (decaisialic acid) (1:1; overlay of three frames with highest interaction). The carbohydrate is shown in yellow in rod model, and the protein is shown in grey in ribbon model with the CC'

loop in magenta and F/G strands in teal. *Upper right*, Snapshot of the CC' loop alone in rod model in contact with PSA, with polar contacts shown as black dashes. General loop structure and contact residues are highly similar to those of I-P24. Simulation snapshots for (a–d) are provided as Supplementary PDB files. (e) Potential of mean force (PMF) curves of the five peptides and Siglec-11 binding to  $\alpha$ -2,8-PSA (distance along reaction coordinate).

domain core) to interact with sialic acid carboxyl groups. I-P24 and I-P50 comprise the CC' loop and F strand/FG loop, respectively. Attempts at experimental quantification of PSA–Siglec-11 binding through SPR spectroscopy and ELISAs failed due to low affinity; previous studies have successfully identified but not quantified binding<sup>23,26</sup>. Also, high protein concentrations required for low affinity quantification limited analysis through other techniques, such as FA, that were optimized for the study of peptide binding. However, potential of mean force (PMF) curves of Siglec-11 Domain 1 with PSA suggest a micromolar binding affinity, based on comparison to PMF curves of peptides (Fig. 5(c)) and experimental peptide characterisation. Results from MD simulations of the five peptides and Siglec-11 with  $\alpha$ -2,8/9-PSA are provided in Supplementary Information (Supplementary Fig. S6).

**Targeting of human neural progenitor cells with Siglec-11-derived peptides.** The relative binding of Siglec-11-derived peptides to a PSA-expressing cell line was determined through microarray screening to assess cell-targeting capabilities of PSA-binding peptides. The sub-set of peptides I-P1–I-P59 (from Siglec-11 Domain 1 linear mapping) was selected for cell binding analysis to supplement and elucidate experimental and computational binding data for I-P24. Relative binding of these peptides to human neural progenitor cells (NPCs) is shown in Fig. 6, along with peptide microarray binding and a corresponding residue interaction map or “epitope map” from simulations. There is a significant association between binding ranks for two cell numbers (Spearman correlation coefficient ( $\rho$ ) = 0.83;  $p$  = 0.257,  $\alpha$  = 0.05, two-tailed), indicating selective binding of cells to PSA-binding peptides. Importantly, peptides derived from or near i) the CC' loop and ii) the FG loop and G/G' strand have prominently lower rank values in comparison to other peptides. The corresponding regions of Siglec-11 contact the PSA ligand in simulations, and some peptides from the FG/G region demonstrated moderate-to-high microarray affinity and moderate selectivity in microarray 2,8/9-PSA competition studies.

## Discussion

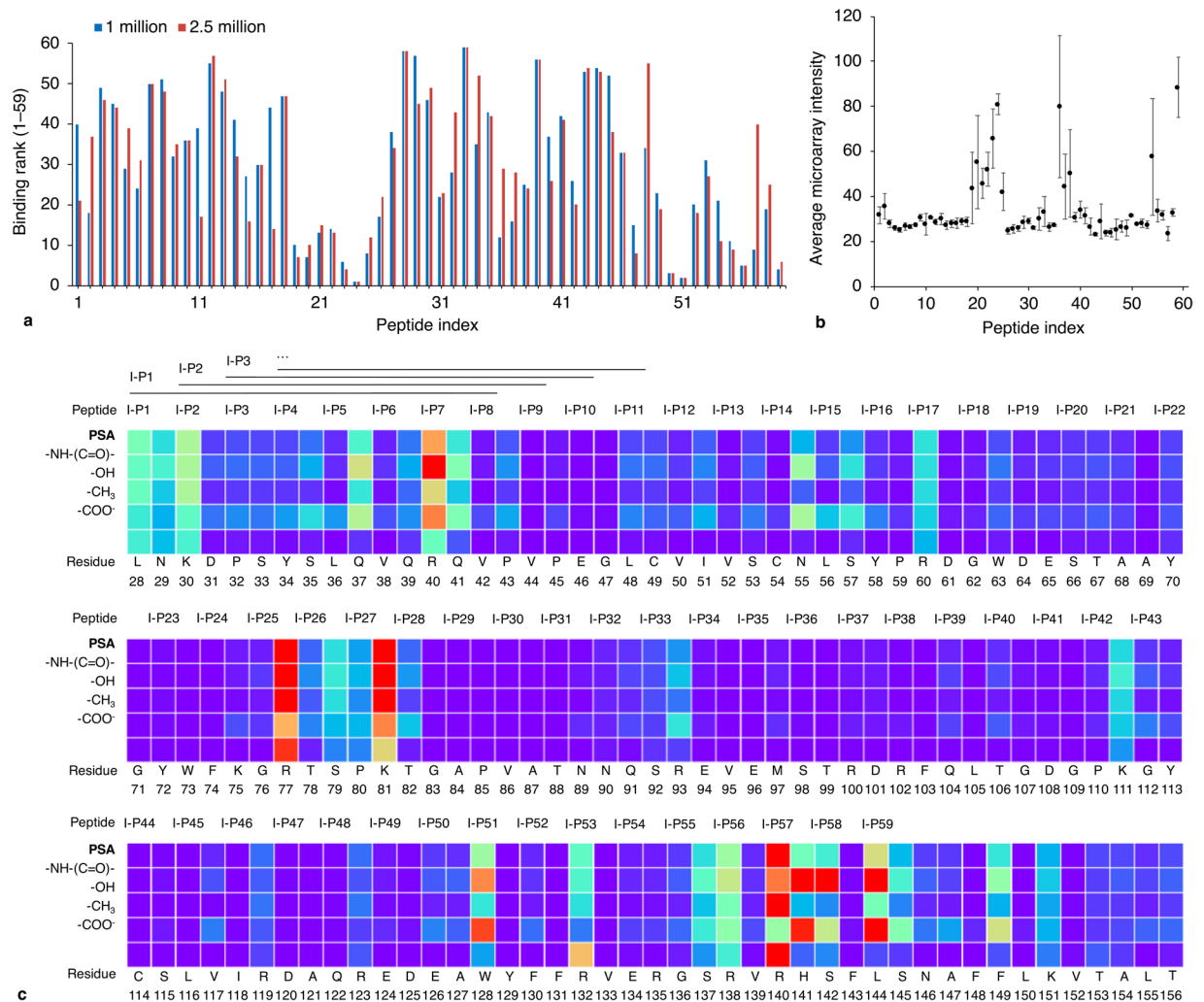
The identification of PSA-binding peptides may be pursued using a number of design strategies, random or rational, and residue-level information related to PSA-interactions may be gained through any strategy. However, design from and comparison to the PSA-binding protein Siglec-11 enabled a complementary study of peptide and protein structure and function, wherein i) peptide design and characterisation provided biochemical understanding on PSA-binding, ii) peptide study led to molecular insight into Siglec-11 interaction, and iii) protein and peptide simulations validated experimental peptide characterisation. Simultaneously, the concurrent analysis of peptides of different origin enabled insight into comparative binding characteristics of Siglec-11-derived peptides.

Assessment of peptide affinity with microarray screening enabled characterisation of residue-based binding properties alongside identification of PSA-binding and non-binding peptides. In comparison, subsequent FA titrations enhanced characterisation through direct quantification of peptide–PSA interaction;  $K_D$  and  $B_{max}$  determinations allowed for property-based discrimination of peptides, and assessment of binding potential supported peptide selection strategy from microarrays. Such subsequent characterisation validates conclusions from high-throughput screening and, furthermore, promotes informed selection of peptide candidates for desired applications.

Despite limited literature on the use of FA assays for the study of carbohydrate interactions<sup>34</sup>, these assays are an attractive platform for in-solution binding analysis of carbohydrate–peptide interactions, which generally demonstrate low 1:1 affinity<sup>24,37–40,43,45–47</sup>. In general, there are few reports on thermodynamic and/or kinetic characterisation of polysaccharide–peptide binding, especially for interactions with high micromolar affinities<sup>34,48–50</sup>. In many binding analysis techniques, low affinities complicate or prevent accurate  $K_D$  determination, as micromolar and millimolar interactions pose detection challenges<sup>43,46</sup>. Additionally, in some techniques, such as SPR spectroscopy, the viscosity of carbohydrate solutions may result in artifacts. Though PSA–peptide binding was detected in prior work with SPR<sup>25</sup>, accurate affinity quantification with this technique was limited by a combination of factors, including the high viscosity of PSA, the affinity detection limit of the system, and the small size of peptides relative to PSA. In contrast, despite limitations with PSA solubility (upper limit of approximately 1.5 mM) and relatively low anisotropy changes, PSA–peptide binding analysis with FA successfully identified and characterised two PSA-binding peptides (I-P24 and II-P336) which performed well in all other experimental and computational assessments. The entropic penalty of binding associated with a low molecular weight peptide interacting with a conformationally flexible, high molecular weight polysaccharide is likely high; hence, the design and characterisation of PSA-binding peptides with  $K_D$ 's comparable to those of Siglec–ligand interactions<sup>51</sup> is notable.

Select peptides, including I-P24 and II-P336, were shown to display some degree of selective binding to PSA of 2,8-linkages. However, specificity to PSA of these linkages was absent, with binding outcompeted by  $\alpha$ -2,8/9-PSA to various degrees. The difference in peptide binding between these isomers, suggested by the selectivity–specificity contrast and supported by MD simulations, may be manipulated to alter selectivity. For example, numerous studies have demonstrated the importance of ligand density in affinity of carbohydrate interactions<sup>37–40,43,52–55</sup>, and there is evidence for altered selectivity with alteration in carbohydrate density<sup>56</sup>; a similar approach could be undertaken wherein ligand density of peptide candidates displaying promising selectivity (with 10% competitor) is modulated to influence selectivity. Furthermore, though peptides with selectivity for  $\alpha$ -2,8-PSA were of focus here, peptides binding to  $\alpha$ -2,8/9-PSA—absent in humans—may be refined for targeting of pathogenic bacterial





**Figure 6.** (a) Binding to peptides derived from Domain 1 of Siglec-11 (I-P1–I-P59, including residues from linker regions) to NPCs. Binding rank represents ranking of binding intensities of peptides (with a rank of 1 corresponding to the peptide showing highest binding amongst 59 peptides). Intensity data (means) are derived from triplicate measurements for each peptide in screening against 1 or 2.5 million NPCs per microarray (error bars excluded for clarity).  $\rho = 0.83$ ;  $p = 0.257$ ,  $\alpha = 0.05$ , two-tailed. Screening with 5 million cells could not be accurately quantified due to high microarray background. Peptides derived from the Siglec-11 CC' loop (residues 74–88) and FG/G region (residues 128–152) demonstrate higher relative binding to NPCs as compared to other peptides, with low rank values for NPC binding (Siglec-11 residue numbering is provided in (c)). While a few other peptides display low binding ranks, the specified regions display low ranks for contiguous peptides. (b) Microarray binding of I-P1–I-P59 to PSA. Binding intensities represent the mean of three independent experiments, with triplicate intra-assay measurements (error bars show inter-assay standard deviations). (c) Interaction map showing the normalized number of contacts (3.0 Å cut-off distance) from 0 (purple) to 1 (red) observed in the MD simulation between Siglec-11 Domain 1 and decasialic acid. Residues correspond to Siglec-11 numbering, and peptide identifiers mark the start of each 15-residue peptide sequence. A similar interaction map for octasialic acid is provided in Fig. S7.

species expressing this PSA<sup>1,2</sup>. Similarly, given the ubiquity of sialic acid in mammalian systems<sup>57</sup>, the demonstrated specificity of PSA-binding peptides to polymeric sialic acid over the monomer can provide considerable advantages in *in vitro* or *in vivo* applications.

The pH insensitivity in binding of high affinity and high selectivity PSA-binding peptides, such as I-P24, may be considered another example of selective binding. This selectivity allows for peptide utilization in environments with variable pH values (e.g., physiological systems). In contrast, pH sensitive peptides demonstrate that selectivity of peptides to PSA may be engineered to switch, which may be useful in “bind and elute” contexts. For example, PSA-binding peptides may be used to target or purify therapeutic proteins with sialic acid modifications of different degrees of polymerization (DPs).

The differing molecular bases<sup>25,58,59</sup> of the thus demonstrated PSA-peptide affinity and selectivity was suggested by experimental techniques (employed for peptide identification and characterisation) and supported by computational approaches (used to validate and further characterise binding). For example, compositional

and positional analyses of peptides suggested that PSA selectivity requires binding of positively charged residues (preferably Arg, as in protein–carbohydrate interactions<sup>41,42</sup>) appropriately spaced by small, hydrophobic residues; simulations likewise demonstrated that electrostatic interactions guide peptide–PSA binding. The apparent lack of a strict consensus motif may be due to conformational flexibility of both the peptide ligand and carbohydrate target. It is possible that DP of PSA affects peptide/protein binding through, for example, conformational differences, entropic effects, stability of species, or multivalency<sup>14,22,35,60,61</sup>. While experimental studies used a polydisperse polymer (with an average DP of 100) and MD simulations assessed binding with decasialic acid, results from the two methodologies showed overall good agreement—especially for I-P24.

Of the peptides evaluated with simulations, I-P24 and I-P50 originated from the N-terminal binding domain (Domain 1) of Siglec-11. The other peptides were derived directly or modified from non-binding regions of the protein, and one of these peptides (II-P336) displayed good affinity and selectivity to PSA. The experimental performance of Domain 1 peptides, specifically high affinity and selectivity of I-P24 and weak binding of I-P50, can be explained by their origins. I-P24 forms the CC' loop in Siglec-11<sup>27</sup>. This loop, which is highly variable amongst Siglecs and bordered by conserved beta strands, has been shown to contain specificity determinants in other Siglecs<sup>27,62</sup>. In Siglec-11, the orientation of the CC' loop is such that it points inwards into the ligand binding pocket as for Siglec-2, -4, -5, and -8<sup>62</sup>. Interestingly, simulations demonstrate that I-P24 also adopts a highly similar omega loop structure on binding PSA. Additionally, the residues of the Siglec-11 loop interacting with PSA are the same residues of I-P24 that orient towards the ligand, with positively charged residues and polar groups contacting the ligand and hydrophobic side chains largely stabilizing the loop's interior. A central Pro (as with Siglec-4, -5, and -8) results in turning of the loop. In both the Siglec-11 CC' loop and I-P24, it is possible that this structural feature, combined with smaller residues maintaining loop flexibility, enable proper orientation of Arg and Lys residues for specific ligand binding. Combined with microarray and FA data on I-P24 selective binding to  $\alpha$ -2,8-PSA, the Siglec-11 CC' loop simulations agree with the prevailing view that this loop confers ligand specificity to Siglecs.

Peptides from the Siglec-11 region containing the Arg conserved in Siglecs<sup>27</sup> (I-P50 and similar) did not experimentally demonstrate high binding, with binding only evident with increased exposure times in microarray imaging. It is possible that microarray binding is partially affected by high peptide hydrophobicity; due to peptide presentation from a partly hydrophobic surface into aqueous media, hydrophilic peptides may be more accessible for target binding. In fact, a few peptides modified from this region to have higher hydrophilicity demonstrated binding in the top 5% of intensities in affinity screening, but also displayed high inter-assay variability. Furthermore, simulations of I-P50 binding to PSA indicate that the interaction, though weak, occurs through guiding electrostatic interactions of Arg residues (corresponding to the conserved Arg from strand F and other Arg residues from the FG loop/G strand in Siglec-11). It is possible that the hydrophobicity of contiguous residues promotes Arg residue binding to PSA when displayed from deeper within an otherwise hydrophilic protein binding site (though, of note, the conserved Arg is thought to be less critical for Siglec-11 binding than for other Siglecs<sup>23</sup>). However, in isolation (*i.e.*, in peptide form), the residues composing I-P50 likely lack topological context and hence target specificity. This contrasts directly with the structure of the CC' loop, which supported the engineering and comparative study of I-P24.

Unlike in microarray affinity screens with PSA, cell screening exclusively identified binding site-derived peptides. Linearly mapped peptides demonstrating affinity to PSA alone did not show consistently greater binding to NPCs, unlike CC' loop and F/G strand-derived peptides, which is expected as linear epitope mapping is not a surrogate for binding site determination. Thus, as compared to binding and competition assays with few species, cell screening served as a peptide selectivity challenge of greater stringency. Both Siglec-11 simulations and cell binding data suggest that along with the CC' loop, the F/G strand may play a role in ligand specificity. However, the higher performance of I-P24 over F/G strand-derived peptides in various affinity and selectivity assessments makes this peptide exemplary for characterisation and application. The remarkable difference in binding of I-P24 and other select peptides to NPCs demonstrates the potential for these peptides to be used in cell targeting<sup>5,63–65</sup>, especially considering only a sub-population of these cells expressed PSA. For example, PSA-binding peptides may be adapted for selective targeting of cancer cells overexpressing PSA<sup>1,2,13,66</sup>. Hence, the conserved continuous binding pocket of Siglec-11 lent itself to the synergistic study of PSA-binding peptides across various platforms.

Thus, rational peptide design informed on interactions of PSA and the native ligand Siglec-11, supported comparative analysis amongst peptides (including peptides from non-Siglec origins), and provided peptides demonstrating unique applicability. For example, the described design and characterisation strategy saliently revealed the top-performing peptide amongst 762 (I-P24) to originate from the unstructured loop bordering the Siglec binding pocket—thus supporting insight into lectin–PSA interactions while presenting a peptide candidate with desirable affinity and selectivity characteristics across multiple platforms. In this manner, exploiting possible lectin binders to glycans for peptide design can provide lectin-mimics with desirable properties, even if direct experimental study of lectin structure and binding proves challenging. Furthermore, based on the general epitope required for peptide–PSA affinity and selectivity, further investigations may isolate PSA-binding regions of other important proteins thought to interact with PSA (*e.g.*, neurotrophins). In future work, hypothesis-driven approaches can build on knowledge of PSA–protein interactions to investigate structural and thermodynamic/kinetic aspects of PSA interactions, improve upon native ligand-based peptide design for polysaccharides, and use design rules from peptide sequence analysis for controlled modulation of ligand affinity and selectivity.

## Methods

**Peptide microarray screening.** Peptide library synthesis, microarray preparation, and microarray screening were performed as described previously<sup>25</sup> with the following specifications. Siglec-related sequences were N-terminally acetylated to match the charge of the corresponding protein fragment. All peptides were printed in triplicate (60 nL spots), and each initial affinity screening was carried out at least three times. Screening was

conducted with 10  $\mu\text{M}$  colominic acid (CA; 5 mL per microarray), the *Escherichia coli* homolog of  $\alpha$ -2,8-PSA identical in structure to human-derived PSA<sup>67</sup>, and with antibody-based chemiluminescent detection of CA. For selectivity screening, microarrays were challenged with 10  $\mu\text{M}$  CA with 0, 0.1, 1, and 10  $\mu\text{M}$   $\alpha$ -2,8/9-PSA in separate conditions (5 mL total volume each), where molar concentrations were based on molecular weight estimation of synthesized 2,8/9-PSA (details on  $\alpha$ -2,8/9-PSA production, purification, and analysis are provided in Supplementary Information). Selectivity screening with monomeric sialic acid was conducted with mass and molar equivalents of N-acetylneuraminic acid (Rose Scientific, Ltd.) along with 10  $\mu\text{M}$  CA (5 mL total volume each). The mass equivalent amount of sialic acid (specifically, 970  $\mu\text{M}$  sialic acid for a given volume of 10  $\mu\text{M}$  CA) was calculated for CA of an average degree of polymerization of 100. For study of pH-dependent binding, affinity screening with CA was performed using PBS buffers of pH 6.0, 7.0, 7.4, and 8.0 (10 mM phosphate, 100 mM NaCl). The impact of NaCl on peptide binding was assessed at pH 7.4 with 0, 100, and 200 mM NaCl. Control dot blot assays for all alternative microarray conditions (e.g.,  $\alpha$ -8/9-PSA competition or pH 6.0) confirmed that antibody detection of CA was not considerably altered in these conditions and that antibodies did not show cross-reactivity with competitors.

Microarray image and data analysis was performed as described<sup>25</sup>. Image transform values for normalization within the image acquisition software were selected based on relative maximum and minimum intensities across all images and hence differed from values selected in previous work; this results in different absolute intensities reported for peptides II-1–II-223, I-P264, and I-P342, though relative intensities are the same (within expected error)<sup>25</sup>. Of the triplicate experiments used to generate mean microarray affinity intensities in standard conditions (within each of which 762 peptides were represented in triplicate), 225 out of 762 values in one of these three experiments were derived from raw data previously reported<sup>25</sup>, but were alternatively analysed as stated. Peptide isoelectric points were calculated with the EMBOSS iep program<sup>68</sup>.

For compositional and positional analyses, residue occurrence within peptide sequences and at each position within sequences was compared between the compiled peptide library and the top 5% affinity binders, as well as between the library and approximately the top 5% selective binders (consisting of peptides with selectivity >65%). For identification of the top 5% selective binders, an affinity constraint was first applied, where peptides displaying background or non-binding intensities (approximately 75% of the library with intensities in the bottom 20% of the intensity range) were excluded along with an additional 5% of very weakly binding peptides (including peptides with intensities in the bottom 30% of the intensity range). Following application of the affinity criteria, 115 peptides (approximately 15% of the library) with binding above background were considered for selectivity assessment. Though select sequences were chosen for modification in the process of peptide design, each set of modified sequences (from different origins) comprised <10% of all sequences and thus inclusion in affinity and selectivity compositional and positional analyses was not considered to affect statistical outcomes through sampling bias.

**Fluorescence anisotropy assay.** Peptides (Table 2) were synthesized at >95% purity (confirmed with HPLC) by Biomatik (see Supplementary Information). All peptides were N-terminally acetylated and C-terminally labelled with tetramethylrhodamine dye (TAMRA). TAMRA was chosen as a label due to its pH insensitivity compared to FAM (fluorescein) and related dyes commonly used for fluorescence assays. Experiments were conducted in black 384-well plates with non-binding surface (NBS, Corning, Inc.) to minimize background binding and enhance signal-to-noise ratio in the fluorescence-based binding assay. Titrations were carried out through addition of TAMRA-labelled peptide (8  $\mu\text{L}$ , final concentration 20 nM) to CA (72  $\mu\text{L}$ , final concentrations 0–1500  $\mu\text{M}$ ) in PBS, pH 7.4 (10 mM phosphate, 100 mM NaCl), and solutions were incubated with agitation to equilibrium at room temperature (3 h,  $\sim 20^\circ\text{C}$ ). Parallel and perpendicular intensities were subsequently measured at 530 nm excitation (band width 5 nm) and 580 nm emission (band width 20 nm) in fluorescence polarization mode with a Tecan Infinite M1000 Pro plate reader. Total fluorescence intensities were also measured as a control for fluorophore properties with target binding<sup>69</sup>; fluorescence intensities of free peptide were determined to be approximately equivalent to intensities with peptide–CA mixtures. At higher concentrations of free CA (control with no peptide addition), attenuation of detected light differed in parallel and perpendicular planes, so respective intensities were subtracted from peptide–CA intensities prior to anisotropy calculations. Increased fluorescence output at high CA concentrations corresponded to higher solution viscosity and likely resulted in light scattering that considerably affected anisotropies without this correction (rheology studies on CA viscosity are provided in Supplementary Information). Unique samples were screened in at least triplicate within each titration experiment, and each peptide titration was carried out at least twice. In experimental repeats, peptide locations on plates were varied to prevent location-specific effects, if any, and different combinations of peptides were assessed at once. Anisotropies were calculated from corrected parallel and perpendicular intensities<sup>69</sup>, and values were fit to the four-parameter logistic binding model<sup>70</sup> in MATLAB. Prior to curve fitting, anisotropy values for each peptide were transformed through subtraction of minimum values to enhance comparison between peptides. The resulting translations in binding curves were identical, within error, to change in anisotropy curves obtained through subtraction of anisotropy of free labelled peptide.  $B_{\text{max}}$  and  $K_{\text{D}}$  values obtained were used in the calculation of  $B_{\text{max}}/K_{\text{D}}$  ratios for comparison to microarray intensity values, and the Pearson correlation coefficient ( $r$ ) was determined.

**Molecular dynamics simulations.** The N-terminal binding domain of Siglec-11 (residues 28–156) was modelled from homologous members of the Siglec family of sialic acid-binding lectins<sup>23,27</sup> using homology modelling in YASARA as previously described<sup>71–73</sup>. Models for peptides were obtained from the PEP-FOLD web server<sup>74</sup>. Siglec-11 and peptides were equilibrated under all-atom molecular dynamics (MD) at 300 K for 300 ns using the GROMACS engine<sup>75</sup>. All atoms were described using the CHARMM36m forcefield<sup>76</sup>, and solutes were solvated in TIP3P water<sup>77</sup> and 100 mM NaCl counterions. Equilibrated structures of the peptides were clustered

using gmx cluster utility, and the cluster with the highest population was utilized for binding studies. Decasialic acid structures (with either 2,8 or 2,8/9 linkages) were built using the oligosaccharide builder in YASARA. The binding sites for PSA on Siglec-11 were inferred from co-solvent MD simulations where one PSA chain was equilibrated with one Siglec-11 monomer (1:1). Similarly, one peptide was equilibrated with one PSA chain (1:1). In both cases, simulations were carried out in triplicates. Simulations were sampled for 500 ns for Siglec-11 and 250 ns for the peptides. The longest contacting peptide–PSA conformations were isolated for further analysis. Potential of mean force (PMF) calculations were carried out for peptide interactions with PSA via umbrella sampling simulation using GROMACS and PLUMED<sup>78</sup>. The PMF was calculated using weighted histogram analysis. Python packages MDAnalysis<sup>79,80</sup> and GromacsWrapper were used to build in-house analysis and plotting scripts.

**Cell binding studies.** ReNcell VM human neural progenitor cells (NPCs; EMD Millipore) were cultured in accordance with manufacturer's instructions and passaged at 80–90% confluency with accutase. Immunocytochemistry/immunofluorescence of NPCs with anti-PSA-NCAM antibody (EMD Millipore) was used to confirm PSA expression, as expression varies amongst cell types/populations and decreases with differentiation. NPCs were fixed with methanol prior to blocking with 5% BSA and incubation with anti-PSA-NCAM (EMD Millipore) and Alexa Fluor-labelled goat anti-mouse IgM (Invitrogen).

NPCs were labelled with CellTracker Red CMTPIX dye (Thermo Fisher Scientific; 10  $\mu$ M). Peptide microarray experiments were carried out similarly to carbohydrate screens, except peptide positions were randomized, and microarrays were blocked overnight at 4 °C in 5% BSA (10 mL each) prior to incubation with labelled cells at 37 °C (2 mL on slide surface; 0, 1, 2.5, and 5 million cells). Washes were conducted with PBS, pH 7.4 (3  $\times$  10 minutes, 10 mL each). Microarrays were air-dried overnight before fluorescence imaging (GE Typhoon Trio+ flatbed scanner; 633 nm excitation, 670BP30 nm emission, 450 PMT, 10  $\mu$ m resolution). Fluorescence values for cell binding were adjusted by subtraction of control slide intensities (without cells) to account for autofluorescence and non-specific binding. Since background and dynamic range were different with variation in cell numbers, binding ranks were determined to enable comparison of peptide binding between conditions, and the Spearman rank-order correlation coefficient ( $\rho$ ) was calculated ( $n = 59$ ).

## Data availability

Any data generated or analysed during this study that are not included in the published article and Supplementary Information, including raw microarray images and PDB files generated from MD simulations, are available from the corresponding authors upon reasonable request.

Received: 20 November 2019; Accepted: 8 April 2020;

Published online: 06 May 2020

## References

- Janus, T. & Janus, T. Membrane oligo- and polysialic acids. *Biochim. Biophys. Acta* **1808**, 2923–2932 (2011).
- Sato, C. & Kitajima, K. Disialic, oligosialic and polysialic acids: distribution, functions and related disease. *J. Biochem.* **154**, 115–136 (2013).
- Schnaar, R. L., Gerardy-Schahn, R. & Hildebrandt, H. Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. *Physiol. Rev.* **94**, 416–518 (2014).
- Kleene, R. & Schachner, M. Glycans and neural cell interactions. *Nat. Rev. Neurosci.* **5**, 195–208 (2004).
- Kim, D.-S. *et al.* Highly pure and expandable PSA-NCAM-positive neural precursors from human ESC and iPSC-derived neural rosettes. *PLoS ONE* **7**, e39715, <https://doi.org/10.1371/journal.pone.0039715> (2012).
- Sato, C. A. II *et al.* Structural diversity in the alpha 2 $\rightarrow$ 8-linked polysialic acid chains in salmonid fish egg glycoproteins. Occurrence of poly(Neu5Ac), poly(Neu5Gc), poly(Neu5Ac, Neu5Gc), poly(KDN), and their partially acetylated forms. *J. Biol. Chem.* **268**, 23675–23684 (1993).
- Galuska, S. P. *et al.* Synaptic cell adhesion molecule SynCAM 1 is a target for polysialylation in postnatal mouse brain. *Proc. Natl. Acad. Sci. USA* **107**, 10250–10255 (2010).
- Sato, C. & Kitajima, K. Impact of structural aberrancy of polysialic acid and its synthetic enzyme ST8SIA2 in schizophrenia. *Front. Cell. Neurosci.* **7**, 61, <https://doi.org/10.3389/fncel.2013.00061> (2013).
- Read, R. C. *et al.* The (alpha2 $\rightarrow$ 8)-linked polysialic acid capsule of group B *Neisseria meningitidis* modifies multiple steps during interaction with human macrophages. *Infect. Immun.* **64**, 3210–3217 (1996).
- Kim, K. J., Elliott, S. J., Di Cello, F., Stins, M. F. & Kim, K. S. The K1 capsule modulates trafficking of E. coli-containing vacuoles and enhances intracellular bacterial survival in human brain microvascular endothelial cells. *Cell Microbiol.* **5**, 245–252 (2003).
- Zelmer, A. *et al.* Differential expression of the polysialyl capsule during blood-to-brain transit of neuropathogenic *Escherichia coli* K1. *Microbiology* **154**, 2522–2532 (2008).
- Lehti, T. A., Pajunen, M. I., Skog, M. S. & Finne, J. Internalization of a polysialic acid-binding *Escherichia coli* bacteriophage into eukaryotic neuroblastoma cells. *Nat. Commun.* **8**, 1915, <https://doi.org/10.1038/s41467-017-02057-3> (2017).
- Tanaka, F. *et al.* Prognostic significance of polysialic acid expression in resected non-small cell lung cancer. *Cancer Res.* **61**, 1666–1670 (2001).
- Kanato, Y., Kitajima, K. & Sato, C. Direct binding of polysialic acid to a brain-derived neurotrophic factor depends on the degree of polymerization. *Glycobiology* **18**, 1044–1053 (2008).
- Janus, T., Nowotarski, K. & Janus, T. Polysialic acid can mediate membrane interactions by interacting with phospholipids. *Chem. Phys. Lipids* **163**, 286–291 (2010).
- Schulz, E. C. *et al.* Structural basis for the recognition and cleavage of polysialic acid by the bacteriophage K1F tailspike protein EndoNF. *J. Mol. Biol.* **397**, 341–351 (2010).
- Theis, T. *et al.* Functional role of the interaction between polysialic acid and myristoylated alanine-rich C kinase substrate at the plasma membrane. *J. Biol. Chem.* **288**, 6726–6742 (2013).
- Akobsson, E., Schwarzer, D., Jokilampi, A. & Finne, J. Endosialidases: versatile tools for the study of polysialic acid. *Top. Curr. Chem.* **367**, 29–74 (2015).
- Volkers, G. *et al.* Structure of human ST8SiaIII sialyltransferase provides insight into cell-surface polysialylation. *Nat. Struct. Mol. Biol.* **22**, 627–635 (2015).

20. Shahraz, A. *et al.* Anti-inflammatory activity of low molecular weight polysialic acid on human macrophages. *Sci. Rep.* **5**, 16800, <https://doi.org/10.1038/srep16800> (2015).
21. Park, I. H., Lin, J., Choi, J. E. & Shin, J. S. Characterization of Escherichia coli K1 colominic acid-specific murine antibodies that are cross-protective against Neisseria meningitidis groups B, C, and Y. *Mol. Immunol.* **59**, 142–153 (2014).
22. Nagae, M. *et al.* Crystal structure of anti-polysialic acid antibody single chain Fv fragment complexed with octasialic acid: insight into the binding preference for polysialic acid. *J. Biol. Chem.* **288**, 33784–33796 (2013).
23. Angata, T. *et al.* Cloning and characterization of human Siglec-11. A recently evolved signaling that can interact with SHP-1 and SHP-2 and is expressed by tissue macrophages, including brain microglia. *J. Biol. Chem.* **277**, 24466–24474 (2002).
24. Arnaud, J., Audfrey, A. & Imberty, A. Binding sugars: from natural lectins to synthetic receptors and engineered neolectins. *Chem Soc. Rev.* **42**, 4798–4813 (2013).
25. Shastry, D. G. & Karande, P. Microarrays for the screening and identification of carbohydrate-binding peptides. *Analyst* **144**, 7378–7389 (2019).
26. Wang, Y. & Neumann, H. Alleviation of neurotoxicity by microglial human Siglec-11. *J. Neurosci.* **30**, 3482–3488 (2010).
27. Varki, A. & Angata, T. Siglecs—the major subfamily of I-type lectins. *Glycobiology* **16**, 1R–27R (2006).
28. Gunten, S. & Bochner, B. S. *Basic and clinical immunology of Siglecs*. *Ann. N. Y. Acad. Sci.* **1143**, 61–82 (2008).
29. Röckendorf, N., Bade, S., Hirst, T. R., Gorris, H. H. & Frey, A. Synthesis of a fluorescent ganglioside GM1 derivative and screening of a synthetic peptide library for GM1 binding sequence motifs. *Bioconj. Chem.* **18**, 573–578 (2007).
30. Zade, H. M., Keshavarz, R., Shekarabi, H. S. Z. & Bakhshinejad, B. Biased selection of propagation-related TUPs from phage display peptide libraries. *Amino Acids* **49**, 1293–1308 (2017).
31. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer Science+Business Media, LLC: New York, NY, 2006.
32. Hall, M. D. *et al.* Fluorescence polarization assays in high-throughput screening and drug discovery: a review. *Methods Appl. Fluoresc.* **4**, 022001, <https://doi.org/10.1088/2050-6120/4/2/022001> (2016).
33. Lea, W. A. & Simeonov, A. Fluorescence polarization assays in small molecule screening. *Expert Opin. Drug Discov.* **6**, 17–32 (2011).
34. Kakehi, K., Oda, Y. & Kinoshita, M. Fluorescence polarization: analysis of carbohydrate–protein interaction. *Anal. Biochem.* **297**, 111–116 (2001).
35. Azurmendi, H. F. *et al.* Extracellular structure of polysialic acid explored by on cell solution NMR. *Proc. Natl. Acad. Sci. USA* **104**, 11557–11561 (2007).
36. Innis, R. B. *et al.* Consensus nomenclature for *in vivo* imaging of reversibly binding radioligands. *J. Cereb. Blood Flow Metab.* **27**, 1533–1539 (2007).
37. Ambrosi, M., Cameron, N. R. & Davis, B. G. Lectins: tools for the molecular understanding of the glycode. *Org. Biomol. Chem.* **3**, 1593–1608 (2005).
38. Gabius, H. J., André, S., Jiménez-Barbero, J., Romero, A. & Solís, D. From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **36**, 298–313 (2011).
39. Weis, W. I. & Drickamer, K. Structural basis of lectin–carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441–473 (1996).
40. García-Hernández, E. & Hernández-Arana, A. Structural bases of lectin–carbohydrate affinities: comparison with protein–folding energetics. *Protein Sci.* **8**, 1075–1086 (1999).
41. Malik, A. & Ahmad, S. Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a neural network. *BMC Struct. Biol.* **7**, 1, <https://doi.org/10.1186/1472-6807-7-1> (2007).
42. Banno, M. *et al.* Development of a sugar-binding residue prediction system from protein sequences using support vector machine. *Comput. Biol. Chem.* **66**, 36–43 (2017).
43. Lundquist, J. J. & Toone, E. J. The cluster glycoside effect. *Chem. Rev.* **102**, 555–578 (2002).
44. Rabenstein, D. L. & Wang, J. Interaction of heparin with two synthetic peptides that neutralize the anticoagulant activity of heparin. *Biochemistry* **45**, 15740–15747 (2006).
45. del Carmen Fernández-Alonso, M. *et al.* Protein–carbohydrate interactions studied by NMR: from molecular recognition to drug design. *Curr. Protein Pept. Sci.* **13**, 816–830 (2012).
46. Kang, Y. *et al.* Bacteriophage tailspikes and bacterial O-antigens as a model system to study weak-affinity protein–polysaccharide interactions. *J. Am. Chem. Soc.* **138**, 9109–9118 (2016).
47. Marquardt, A., Bernevic, B. & Przyblyski, M. Identification, affinity characterisation and biological interactions of lectin-like peptide–carbohydrate complexes derived from human TNF- $\alpha$  using high-resolution mass spectrometry. *J. Pept. Sci.* **13**, 803–810 (2007).
48. Morales Betanzos, C. *et al.* A. Bacterial glycoprofiling by using random sequence peptide microarrays. *ChemBioChem* **10**, 877–888 (2009).
49. Jeong, K. J., Butterfield, K. & Panitch, A. A novel assay to probe heparin–peptide interactions using pentapeptide-stabilized gold nanoparticles. *Langmuir* **24**, 8794–8800 (2008).
50. McKeon, J. & Holland, L. A. Determination of dissociation constants for a heparin-binding domain of amyloid precursor protein and heparins or heparan sulfate by affinity capillary electrophoresis. *Electrophoresis* **25**, 1243–1248 (2004).
51. Pröpster, J. M. *et al.* Structural basis for sulfation-dependent self-glycan recognition by the human immune-inhibitory receptor Siglec-8. *Proc. Natl. Acad. Sci. USA* **113**, E4170–E4179, <https://doi.org/10.1073/pnas.1602214113> (2016).
52. Hyun, S., Lee, E. H., Park, J. & Yu, J. Tentacle type peptides as artificial lectins against sulfated Lewis X and A. *Bioorg. Med. Chem. Lett.* **18**, 4011–4014 (2008).
53. Hyun, S., Kim, J., Kwon, M. & Yu, J. Selection and syntheses of tentacle type peptides as ‘artificial’ lectins against various cell-surface carbohydrates. *Bioorg. Med. Chem.* **15**, 511–517 (2007).
54. Kwon, M., Jeong, S., Lee, K. H., Park, Y.-K. & Yu, J. Mimicry of tandem repeat peptides against cell surface carbohydrates. *J. Am. Chem. Soc.* **124**, 13996–13997 (2002).
55. Connaris, H., Crocker, P. R. & Taylor, G. L. Enhancing the receptor affinity of the sialic acid-binding domain of Vibrio cholerae sialidase through multivalency. *J. Biol. Chem.* **284**, 7339–7351 (2009).
56. Horan, N., Yan, L., Isobe, H., Whitesides, G. M. & Kahne, D. Nonstatistical binding of a protein to clustered carbohydrates. *Proc. Natl. Acad. Sci. USA* **96**, 11782–11786 (1999).
57. Varki, A. Sialic acids in human health and disease. *Trends Mol. Med.* **14**, 351–360 (2008).
58. Wang, W. & Woodbury, N. W. Unstructured interactions between peptides and proteins: exploring the role of sequence motifs in affinity and specificity. *Acta Biomater.* **11**, 88–95 (2015).
59. Sood, A. *et al.* Defining the specificity of carbohydrate–protein interactions by quantifying functional group contributions. *Chem. Inf. Model.* **58**, 1889–1901 (2018).
60. Häyrynen, J. *et al.* High affinity binding of long-chain polysialic acid to antibody, and modulation by divalent cations and polyamines. *Mol. Immunol.* **39**, 399–411 (2002).
61. Inoue, S. & Inoue, Y. Developmental profile of neural cell adhesion molecule glycoforms with a varying degree of polymerization of polysialic acid chains. *J. Biol. Chem.* **276**, 31863–31870 (2001).
62. Ereño-Orbea, J. *et al.* Molecular basis of human CD22 function and therapeutic targeting. *Nat. Commun.* **8**, 764, <https://doi.org/10.1038/s41467-017-00836-6> (2017).
63. Mousavizadeh, A., Jabbari, A., Akrami, M. & Bardania, H. Cell targeting peptides as smart ligands for targeting of therapeutic or diagnostic agents: a systematic review. *Colloids Surf., B* **158**, 507–517 (2017).

64. Dodla, M. C. *et al.* Differing lectin binding profiles among human embryonic stem cells and derivatives aid in the isolation of neural progenitor cells. *PLoS One* **6**, e23266, <https://doi.org/10.1371/journal.pone.0023266> (2011).
65. Zhao, W., Yuan, H. & Ma, L. Isolation and initial application of a novel peptide that specifically recognizes the neural stem cells derived from rhesus monkey embryonic stem cells. *J. Biomol. Screen.* **15**, 687–694 (2010).
66. Valentiner, U., Mühlhoff, M., Lehmann, U., Hildebrandt, H. & Schumacher, U. Expression of the neural cell adhesion molecule and polysialic acid in human neuroblastoma cell lines. *Int. J. Oncol.* **39**, 417–424 (2011).
67. Finne, J. Occurrence of unique polysialosyl carbohydrate units in glycoproteins of developing brain. *J. Biol. Chem.* **257**, 11966–11970 (1982).
68. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
69. Rossi, A. M. & Taylor, C. W. Analysis of protein–ligand interactions by fluorescence polarization. *Nat. Prot.* **6**, 365–387 (2011).
70. Findlay, J. W. & Dillard, R. F. Appropriate calibration curve fitting in ligand binding assays. *AAPS J.* **9**, E260–E267, <https://doi.org/10.1208/aapsj0902029> (2007).
71. Irudayanathan, F. J., Trasatti, J. P., Karande, P. & Nangia, S. Molecular architecture of the blood brain barrier tight junction proteins—a synergistic computational and *in vitro* approach. *J. Phys. Chem. B.* **120**, 77–88 (2016).
72. Rajagopal, N., Irudayanathan, F. J. & Nangia, S. Computational nanoscopy of tight junctions at the blood–brain barrier interface. *Int. J. Mol. Sci.* **20**, 5583, <https://doi.org/10.3390/ijms20225583> (2019).
73. Irudayanathan, F. J. *et al.* Self-assembly simulations of classic claudins—insights into the pore structure, selectivity and higher-order complexes. *J. Phys. Chem. B* **122**, 7463–7474 (2018).
74. Lamiable, A. *et al.* PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **44**, W449–W454, <https://doi.org/10.1093/nar/gkw329> (2016).
75. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25, <https://doi.org/10.1016/j.softx.2015.06.001> (2015).
76. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
77. Jorgensen, W. L., Chandrasekhar, J. & Madura, J. D. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
78. Bonomi, M. *et al.* Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).
79. Gowers, R. J. *et al.* MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. In Proceedings of the 15th Python in Science Conference—Austin, Texas (ed. Benthall, S. & Rostrup, S.), SciPy, 98–105; <https://doi.org/10.25080/majora-629e541a-00e> (2016).
80. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

## Acknowledgements

This work was supported by the National Science Foundation Graduate Research Fellowship (to DGS) under Grant No. DGE-1247271. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. DGS thanks Dr. John P. Trasatti, Dr. Steven Cramer, and Dr. Linda McGown for critical input during conception and design of the work. DGS and PK thank Dr. Patrick Maxwell and Dr. Jennifer Hurley (RPI) for use of the ChemiDoc XRS + System for microarray and gel imaging. DGS acknowledges Dr. Catherine Royer (RPI) for advice on fluorescence anisotropy assay development and for use of the Tecan Infinite M1000 Pro plate reader for these assays. DGS and AW thank Dr. Kalib St. Ange for assistance with NMR spectroscopy of PSAs. DGS and PK thank Dr. Jonathan Dordick (RPI) for the culturing and provision of ReNcells for cell binding studies. Parts of this work used facilities in the Cell and Molecular Biology, Microbiology, and NMR Research Cores in the Center for Biotechnology and Interdisciplinary Studies (RPI).

## Author contributions

DGS conceived of and designed the work and acquired and analysed microarray and fluorescence anisotropy data. PK contributed to conception and design of the work. FJI performed molecular dynamics simulations and provided data for Fig. 5, 6, S6, and S7. SN contributed to design and analysis of computational studies. AW produced and purified  $\alpha$ -2,8/9-PSA, contributed to analysis (NMR spectroscopy), and provided material for Fig. S8 and S9, with support and help from MK. RJL contributed to analysis of  $\alpha$ -2,8/9-PSA production/purification data and provided input on design of the work. All other figures were prepared by DGS. DGS, FJI, and AW wrote the manuscript (MD simulations methods by FJI,  $\alpha$ -2,8/9-PSA production, purification, and NMR analysis by AW, and remaining by DGS). All authors reviewed the manuscript.

## Competing interests

Rensselaer Polytechnic Institute has filed a provisional patent application related to peptide sequences described in this paper, with DGS and PK as inventors (U.S. Patent Application No. 62/879,176). FJI, AW, MK, RJL, and SN have no conflicts of interest to declare.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-64088-z>.

**Correspondence** and requests for materials should be addressed to D.G.S. or P.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020