Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22

Di Carluccio, Cristina; Forgione, Rosa Ester; Montefiori, Marco; Civera, Monica; Sattin, Sara; Smaldone, Giovanni

Published in:
iScience

DOI:
10.1016/j.isci.2020.101998

Publication date:
2021

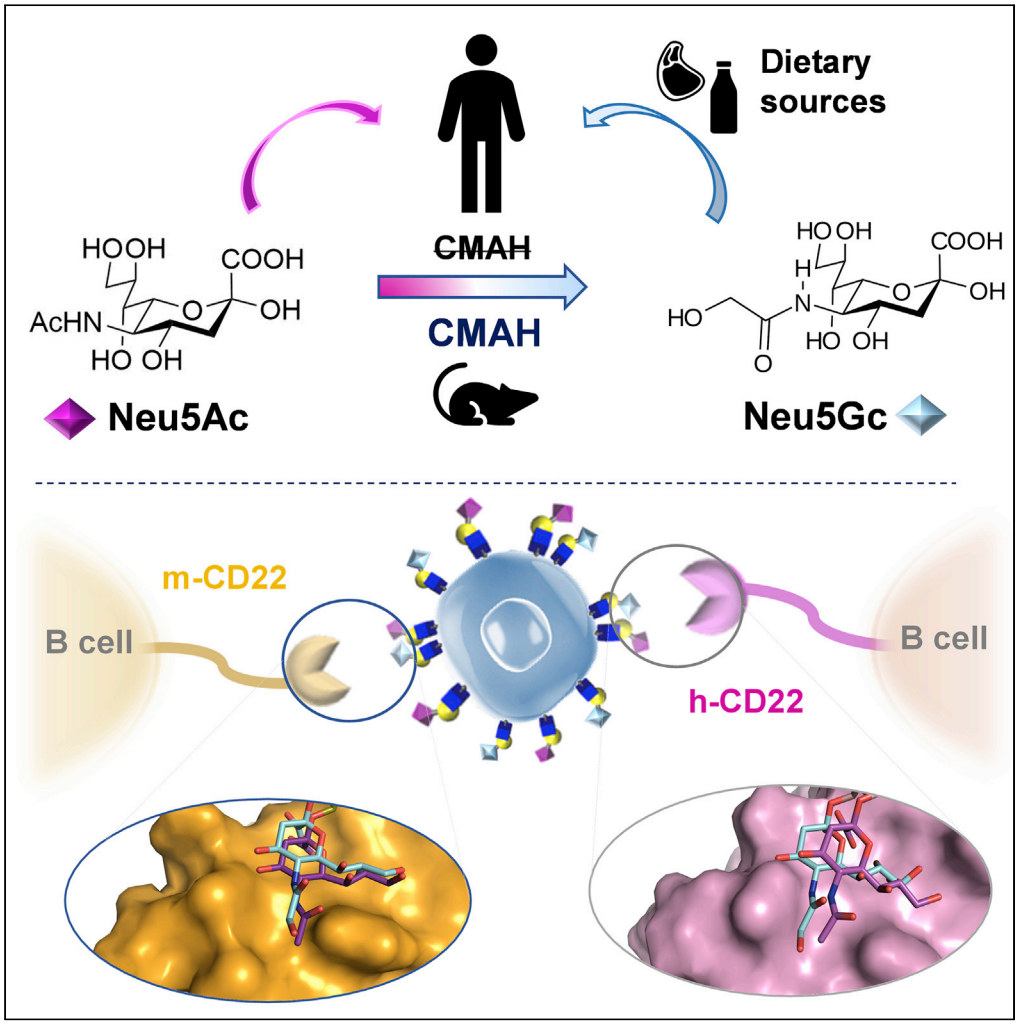
Document Version
Publisher's PDF, also known as Version of record

iScience

CellPress
OPEN ACCESS

Article

Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22



Cristina Di Carluccio, Rosa Ester Forgione, Marco Montefiori, ..., Antonio Molinaro, Roberta Marchetti, Alba Silipo

roberta.marchetti@unina.it (R.M.)
silipo@unina.it (A.S.)

HIGHLIGHTS

The structural basis of sialoglycans recognition by h/m CD22 has been investigated

The binding modes of Neu5Gc-/Neu5Ac-containing ligands to m/h-CD22 were compared

The bioactive conformation of sialoglycans has been derived

Our findings may help in the regulation of immune response and cancer prevention

Di Carluccio et al., iScience 24, 101998
January 22, 2021
https://doi.org/10.1016/j.isci.2020.101998

iScience

CellPress
OPEN ACCESS

Article

Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22

Cristina Di Carluccio,[1,6] Rosa Ester Forgione,[1,6] Marco Montefiori,[2] Monica Civera,[2] Sara Sattin,[2] Giovanni Smaldone,[3] K. Fukase,[4] Y. Manabe,[4] Paul R. Crocker,[5] Antonio Molinaro,[1] Roberta Marchetti,[1,7,*] and Alba Silipo[1,*]

SUMMARY

Siglecs (sialic acid binding immunoglobulin (Ig)-like lectins) constitute a group of 15 human and 9 murine cell-surface transmembrane receptors belonging to the I-type lectin family, mostly expressed on innate immune cells and characterized by broadly similar structural features. Here, the prominent inhibitory CD22 (Siglec-2), well known in maintaining tolerance and preventing autoimmune responses on B cells, is studied in its human and murine forms in complex with sialoglycans. In detail, the role of the *N*-glycolyl neuraminic acid (Neu5Gc) moiety in the interaction with both orthologues was explored. The analysis of the binding mode was carried out by the combination of NMR spectroscopy, computational approaches, and CORCEMA-ST calculations. Our findings provide a first model of Neu5Gc recognition by h-CD22 and show a comparable molecular recognition profile by h- and m-CD22. These data open the way to innovative diagnostic and/or therapeutic methodologies to be used in the modulation of the immune responses.

INTRODUCTION

Sialic acids (Sias) comprise a family of nearly 50 members of acidic monosaccharides, expressed by species belonging to the vertebrates and characterized by a particular nine-carbon sugar backbone. The two predominant forms of sialic acid in mammals are the *N*-acetyl neuraminic acid (Neu5Ac) and the *N*-glycolyl neuraminic acid (Neu5Gc). The latter is biosynthesized by the enzymatic addition of a hydroxyl group to *N*-acetyl moiety at 5′-position of Neu5Ac, catalyzed by a hydroxylase/monooxygenase enzyme, the cytidine monophospho-*N*-acetyl neuraminic acid hydroxylase (CMAH). However, in contrast to mouse or great apes such as chimpanzee, in humans the specific loss of Neu5Gc expression is ascribed to a fixed genomic mutation in CMAH that leads to the gene loss in the hominin lineage (Okerblom and Varki, 2017). Despite the inability to produce Neu5Gc, it can be exogenously introduced from specific dietary sources, such as red meat and cow's milk, and metabolized via the Neu5Ac biochemical pathway (Okerblom and Varki, 2017; Varki, 2017). Different studies reported the presence of Neu5Gc on fetal tissues and on tumor cells, such as melanoma, retinoblastoma, colon cancer, and breast cancer (Samraj et al., 2014). Low levels of Neu5Gc were also detected on the surfaces of human secretory epithelia and small- and large-blood vessels endothelia (Varki, 2017). Therefore, Neu5Gc can be considered a pioneering example of "xeno-autoantigen." The anti-Neu5Gc antibodies detected in humans contribute to establish "xenosialitis," a chronic inflammation state in Neu5Gc-enriched tissues that can significantly impact cancer progression, increasing tumor-associated inflammation (Altman and Gagneux, 2019; Zhou et al., 2020).

Sias are considered self-associated molecular patterns (SAMPs), as they function as determinant of "self" through their intrinsic recognition by specific inhibitory receptors belonging to the Siglec family (Häubli and Varki, 2020; Duan and Paulson, 2020). These receptors assist immune cells in the discrimination between "self" and "non-self" and constitute important regulators of the immune system (Macauley et al., 2014). Their functions depend on their unique and precise sialoglycans' binding specificity as well as on their expression pattern within different cellular compartments (Di Carluccio et al., 2021). Siglecs represent a family of immune proteins that can be classified in two different and evolutionary conserved sub-groups; one comprises Sialoadhesin (Siglec-1), CD22 (Siglec-2), MAG (Siglec-4) and Siglec-15, whereas the other

[1]Dipartimento di Scienze Chimiche, Complesso Universitario Monte Sant'Angelo, Università di Napoli Federico II, Via Cintia 4, 80126 Napoli, Italy

[2]Dipartimento di Chimica, Università Degli Studi di Milano, Via Golgi, 19, 20133 Milano, Italy

[3]IRCCS SDN, Via E. Gianturco 113, Napoli, Italy

[4]Department of Chemistry, Graduate School of Science, Osaka University, Suita, Japan

[5]Division of Cell Signalling and Immunology, School of Life Sciences, University of Dundee, Dundee, United Kingdom

[6]These authors contributed equally

[7]Lead contact

*Correspondence: roberta.marchetti@unina.it (R.M.), silipo@unina.it (A.S.)

https://doi.org/10.1016/j.isci.2020.101998

includes the so-called CD33 (Siglec-3)-related family, rapidly evolving and displaying high homology with the precursor, namely Siglec-3. Despite this differentiation, Siglecs share the ability to recognize and bind to sialic acid epitopes through the N-terminal V (variable)-set domain. Siglecs also possess a variable number (from 1 to 16) of "C2-set" Ig-like domains followed by a single pass trans-membrane region; most of Siglecs feature cytoplasmic tyrosine motifs, such as ITIM (immunoreceptor tyrosine-based inhibitory motif) and ITIM-like regions, which confer inhibitory signaling properties. A few activatory-type Siglecs, such as Siglecs-14–16, contain instead a positively charged amino acid in their transmembrane region and associate with immunoreceptor tyrosine-based activatory motif (ITAM)-containing adaptor proteins (Ajit and Angata, 2006).

The sialic acids are diversely recognized among the Siglec family, for example MAG only binds to Neu5Ac, human and murine sialoadhesin show a strong preference for Neu5Ac over Neu5Gc, whereas murine and human CD22 bind both, with m-CD22 preferring Neu5Gc over Neu5Ac (Angata, 2018). Interestingly, major human pathogens have evolved the ability to mimic Neu5Ac-containing structures, either synthesized *de novo* or scavenged from the host, to escape immune surveillance. In contrast, only a few pathogens express Neu5Gc or Neu5Gc-like structures (Varki, 2017).

Within this context, we here considered an interesting Siglec member, the inhibitory CD22 (h-Siglec-2 and its murine ortholog m- Siglec-2), expressed on the surface of B cells and to a lesser extent of T cells, able to modulate B cell tolerance by counteracting B cell receptor (BCR) signaling. On resting B-cells, CD22 is "masked" by *cis* interactions with sialoglycans exposed on the same plasma membrane, resulting in the formation of CD22 homo-oligomers (Han et al., 2005). Conversely, upon BCR activation, CD22 clusters are recruited to the BCR and, together with Siglec-10, negatively regulate the B cell signaling. CD22 malfunctioning and the resulting lack of appropriate BCR inhibition has been linked to several B-cell-related pathologies, such as hairy cell leukemia, marginal zone lymphoma, chronic lymphocytic leukemia, and non-Hodgkin lymphoma (Müller and Nitschke, 2014; Mahajan and Pillai, 2016; Dörner et al., 2012).

Because Siglecs, and indeed CD22, are considered effective glyco-immuno checkpoints within cancer immunotherapy (Duan and Paulson, 2020), and because changes in the Neu5Gc/Neu5Ac ratio can potentially modulate Siglecs' binding and signaling properties, understanding the basis of Neu5Gc - Siglec interaction may have therapeutic implications. Thus, also considering that the molecular details of Neu5Gc recognition by Siglecs are still far from being explored in-depth, we here aimed to understand the differences in the ability of human and murine CD22 to recognize and bind to various sialoglycans. To achieve this aim, we elucidated the molecular mechanisms of Neu5Gc recognition by, and binding to, human CD22/Siglec-2 (h-CD22) and compared it with the murine ortholog m-Siglec-2 (m-CD22). We combined NMR spectroscopy to computational approaches such as Molecular Dynamics, docking, and CORCEMA-ST methods, thus collecting pivotal information concerning the binding epitope and the conformational behavior of Neu5Gc containing glycans.

RESULTS

The interaction between different sialoglycans bearing acetylated or glycolylated sialic acid, namely Neu5Ac and Neu5Gc with human and murine CD22, was investigated as follows.

Binding specificity of m- and h- CD22 toward Neu5Ac and Neu5Gc ligands

The binding affinities of both m-CD22 and h-CD22 toward Neu5Gc- and Neu5Ac-containing ligands were evaluated by fluorescence analyses; in detail, fluorescence titrations of increasing amounts of sialoglycans into a fixed concentration of the proteins were performed. The results demonstrated the ability of m- and h-CD22 to similarly recognize acetylated and glycolylated sialoglycans (Figure 1), as supported by the derived values of binding constants (K_b), all in the micromolar range. Thus, human and murine CD22 recognized the examined sialoconjugates comparably.

Glycolylated 6'SLN displays a comparable molecular behavior in the binding pocket of murine and human CD22

The glycolylated 6'-sialylactosamine [Neu5Gc-α-(2,6)-Gal-β-(1,4)-GlcNAc-β-OR, 6'SLN] was investigated upon binding to human and murine CD22. A detailed STD NMR (Meyer and Peters, 2003; Angulo and

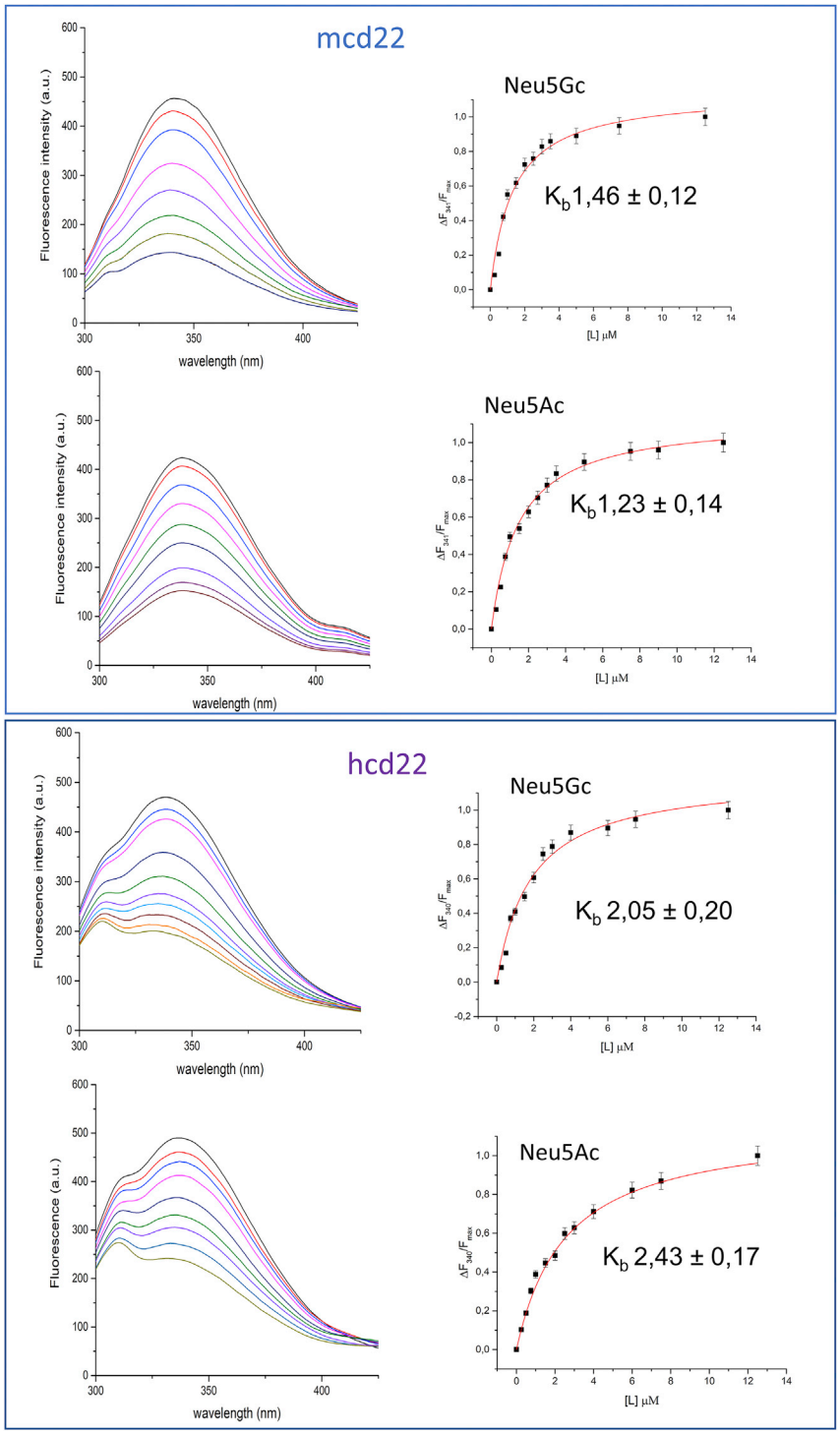


Figure 1. Fluorescence titrations

Fluorescence spectra of m-CD22 (upper panel, black lines) or hCD22 (lower panel, black lines) in the presence of increasing amounts of Neu5Gc-containing ligand (colored lines) or Neu5Ac-containing ligand (colored lines), respectively. The binding isotherm and the values of the binding constants (K_b) are also reported. For each data point, 10% Y error bars are shown.

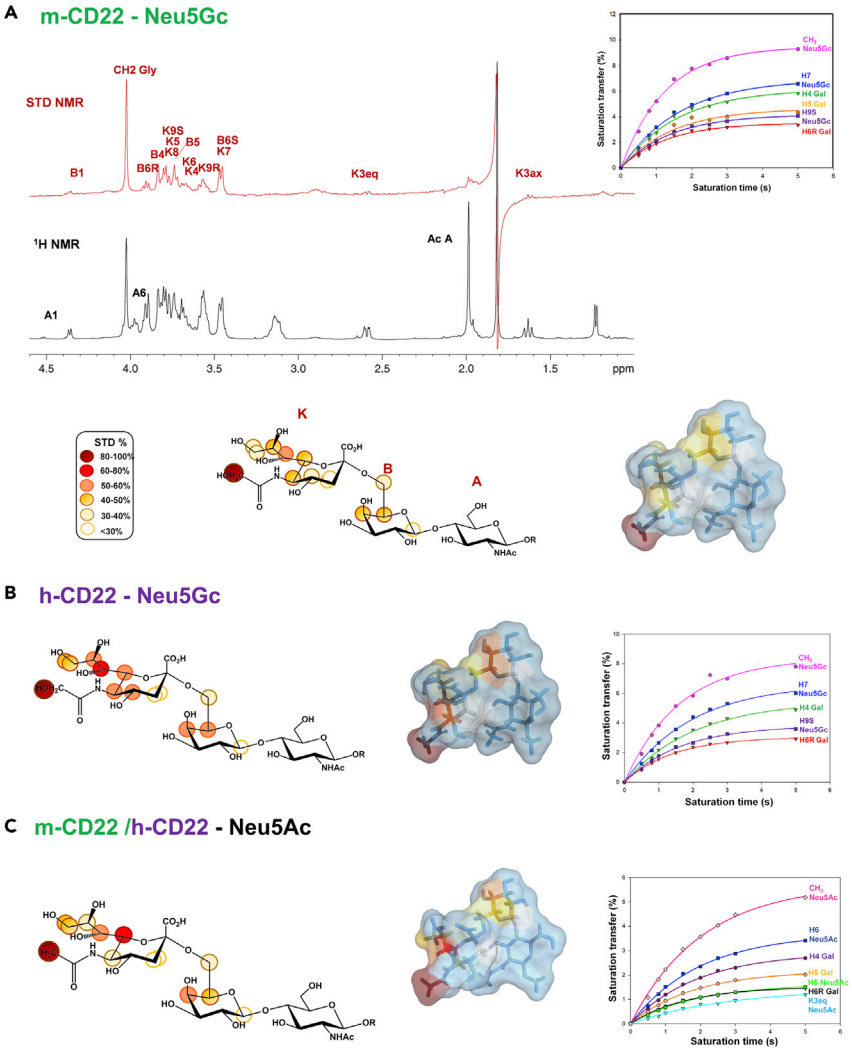


Figure 2. STD NMR analysis of glycolylated and acetylated Sia-α-(2,6)-Gal-β-(1,4)-GlcNAc-β−$OCH_2CH_2NH_2$ interacting with murine and human CD22

(A) Reference ^{1}H NMR spectrum (black) and STD 1D NMR spectrum (red) with a molecular ratio m-CD22/glycolylated ligand of 1:100 at saturation time of 2s. On the right, STD build-up curves are reported. STD build-up curves are calculated using the following monoexponential equation: $STD\ (t_{sat}) = STD^{max}(1 - e^{-k_{sat}t_{sat}})$, where STD ($t_{sat}$) is the STD signal intensity of each proton at t_{sat} saturation time, STD^{max} is the asymptotic maximum of the curve and k_{sat} represents the observed saturation rate constant measuring the speed of STD build-up. The highest STD signal intensity is referred to the glycolyl/acetyl group of sialic acid, set to 100%, whereas the other protons were normalized to this value. The STD-derived epitope mapping on the molecular envelope of Neu5Gc ligand in the bioactive conformation with color code according to the observed STD effects is also shown.

(B) On the left, epitope map of the glycolylated ligand interacting with h-CD22, calculated from the ratio $(I_0-I_{sat})/I_0$, where (I_0-I_{sat}) is the STD signal and I_0 is the peak intensity of the unsaturated reference spectrum. STD effects lower than 10% are not indicated. In the middle, the STD-derived epitope mapping on the molecular envelope of the ligand in its bioactive conformation is shown. On the right, STD build up curves are reported.(C) On the left, epitope map of the acetylated ligand interacting with m/h-CD22, calculated from the ratio $(I_0-I_{sat})/I_0$, where (I_0-I_{sat}) is the STD signal and I_0 is the peak intensity of the unsaturated reference spectrum. STD effects lower than 10% are not indicated. In the middle, the STD-derived epitope mapping on the molecular envelope of the ligand in its bioactive conformation is shown. On the right, STD build up curves are reported.H6R, H6S, H9R, and H9S protons refer to H6-proR, H6-proS, H9-proR, and H9-proS, respectively.

Nieto, 2011) analysis allowed us to map the epitopes of the ligand when interacting to both m-CD22 (Figure 2A) and h-CD22 (Figure 2B). The glycolylated ligand was recognized similarly by both receptors, as highlighted by an almost comparable binding epitope (Figures 2A and 2B). The highest STD signal

belonged to the glycolyl moiety of Neu5Gc, whose signal was set to 100%. The sialic acid—galactose moiety was the main determinant of the binding to both h-CD22 and m-CD22; in detail, H-7 of Neu5Gc (K7, Figure 2A) was saturated more than 50%, whereas H-5, H-6, and H-8 of Neu5Gc (K5, K6 and K8 protons in Figure 2A), as well as H-5 and H-4 of Gal (B5 and B4, Figure 2A) were in the range of 40%–50%. Furthermore, the STD signals of the diastereotopic H-3_{eq} and H-3_{ax} protons (K3) showed the lowest effects with both murine and human CD22. The GlcNAc residue (**A**) was completely excluded from the CD22-binding pocket (Figures 2A and 2B), indicating its solvent exposure. The construction of STD build-up curves (Yan et al., 2003) then allowed to accurately define the glycolylated trisaccharide epitope excluding potential artifacts caused by differences in the longitudinal relaxation time T1 of the ligand protons (Marchetti et al., 2016) (Figures 2A and 2B, Tables S1 and S2).

The topology and conformation adopted by 6′SLN when interacting with h- and m-CD22 *(the bioactive conformation)* was achieved by transferred NOESY (tr-NOESY) analyses (Meyer and Peters, 2003). The stability of φ and ψ dihedral angles of the glycosidic linkages in the free state was monitored during 100 ns Molecular Dynamic simulation in explicit solvent, carried out using the Amber18 package (Case et al., 2018) (See also Methods and Figure S2). Differently from the Neu5Ac trisaccharide (Di Carluccio et al., 2019; Forgione et al., 2020), that in solution explores different populations depending on the values of φ torsion angle (−60°/180°), the Neu5Gc glycan preferentially adopts a conformation with φ around −60° (see Table S3).

As for the bound state, tr-NOESY analyses confirmed the preference of the glycolylated glycan for the energetic minimum characterized by φ/ψ dihedral angles of −60°/180°. The absence of NOE contacts between the H6-proR of galactose and the diastereotopic (axial and equatorial) H-3 protons of sialic acid and the key NOE established between the acetyl group of GlcNAc and H-5 of sialic acid (Table S3) observed in the tr-NOESY spectra (Figures S1B and S1C) revealed a bent conformation of the ligand, which assumed a shape characterized by an umbrella-like topology when bound either to h-CD22 and m-CD22 (Chandrasekaran et al., 2008).

Molecular modeling showed similar binding features of murine and human CD22 in the interaction with glycolylated ligands

Computational studies including homology modeling, docking, and MD simulations were carried out to describe the binding of CD22 with Neu5Gc ligands. The crystal structure of h-CD22 (PDB: 5VKJ) (Ereño-Orbea et al., 2017) was used as structural template to model m-CD22, whose three-dimensional structure is not available. The sequence encoding for m-CD22 extracellular V-set and C2-set domains was aligned to the template sequence using BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990). According to the sequence alignment displayed in Figure 3A, m-CD22 expectedly showed significant sequence identity (above 58%) relatively to h-CD22, in agreement with the conserved nature of CD22. The target template alignment was submitted to the SWISS-MODEL (Waterhouse et al., 2018) server to obtain a three-dimensional model of m-CD22; the structure quality was assessed through PROCHECK web server (Laskowski et al., 1993), giving a Ramachandran plot in which 80.3% of model torsional angles were in the mostly favored region and 19.4% in the additional allowed region. The m-CD22 structural model obtained showed the typical Siglec Ig-like folding, with the sialic acid binding site located at the summit of the *N*-terminal V-set domain (Figure 3B). Similarly to other Siglecs, m-CD22 binding site architecture features a shallow pocket constituted by the A, F, and G strands and bound by the CC′ and GG′ variable loops. Compared with h-CD22, the composition of the binding site residues entitled to sialylated epitopes binding was overall conserved. The most relevant differences lied in the replacement of $Lys23^{h}$, $Tyr64^{h}$, and $Lys127^{h}$ with $Asp25^{m}$, $Phe68^{m}$, and $Arg131^{m}$, which slightly affected the polarity of the binding region. Prior to docking calculations, the structural model was subjected to MD simulation, monitoring along the trajectory the backbone RMSD and the RMSF of the whole structure, as well as the RMSD of the CC′ and GG′ loops (Figures S3A–S3D). The analysis of the potential energy along the simulation allowed to establish the m-CD22 structure at lowest energy, which was subsequently employed for docking calculations (Figure S3D).

To analyze its binding mode, the Neu5Gc ligand was docked into h-CD22- and m-CD22-binding sites by means of Autodock 4.20 (Morris e al., 2009). The energies and populations of the top clusters were very similar for both receptors, ranging from −3.5 to −2 kcal mol^{-1} (Table S4). Thus, from analysis of the docking results, the h-CD22/and m-CD22/ligand complexes displaying lower relative energy and higher cluster populations were selected to run MD simulations. Notably, in the aforementioned poses Neu5Gc ligand

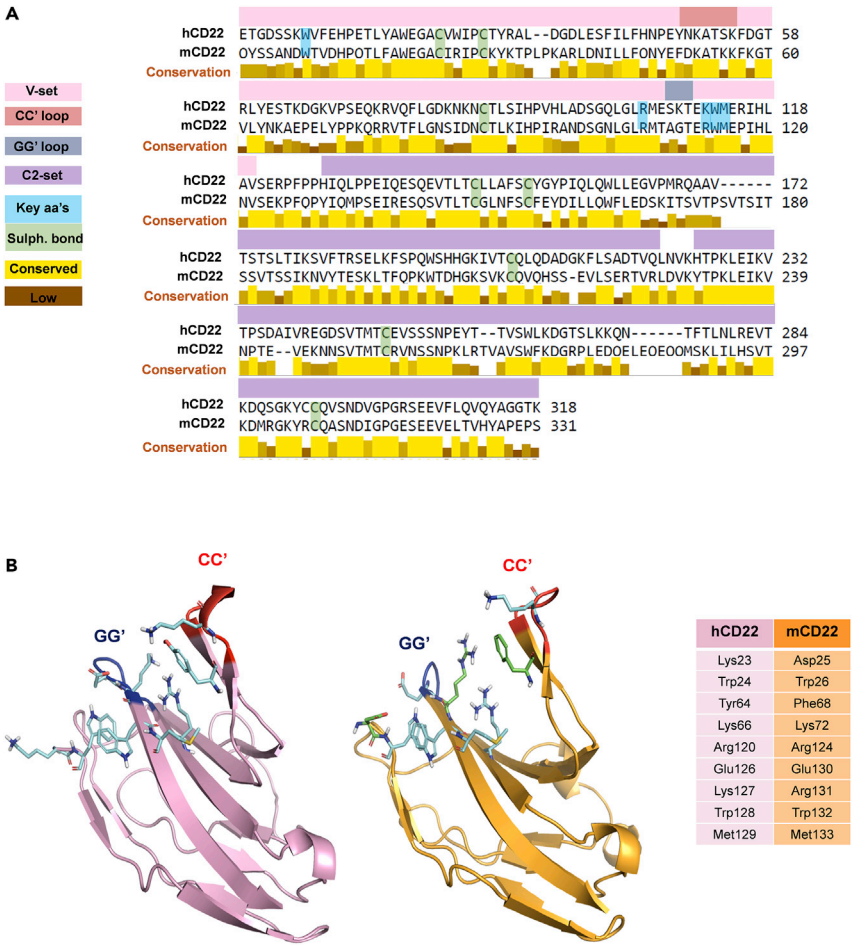


Figure 3. Comparison of h-CD22 and m-CD22 structures

(A) BLAST alignment of the extracellular regions of murine CD22 and human CD22. Key amino acids are highlighted in blue and Cys forming disulfide bridges in green. Sequence corresponding to the V-set domain is evidenced in pink, sequence of C2-set domains in purple. Conservation between the two sequences is evaluated using Jalview (Waterhouse et al., 2009).

(B) Comparison of the *N*-terminal V-set domains of h-CD22 (pink), PDB: 5VKM, and m-CD22 homology model (orange). Common residues constituting the binding sites are highlighted in cyan. Residues of m-CD22 pocket differing from h-CD22 are colored in green. A direct comparison of the binding site residues can be found in the table on the right.

displayed a similar binding mode inside the receptors pocket, in accordance with NMR data, showing the involvement of the following major determinants of sialylated ligands binding, i.e. the conserved arginine ($Arg120^{h}$ and $Arg124^{m}$) and aromatic residues ($Trp24^{h}$, $Trp128^{h}$ and $Trp26^{m}$, $Trp132^{m}$) (Ereño-Orbea et al., 2017). It is worth noting that with both h-CD22 and m-CD22, the ligand assumed an umbrella-like conformation in the chosen clusters.

To finely describe the complexes between Neu5Gc-containing glycans and h- and m-CD22, the aforementioned structures were used as starting point to run MD simulations throughout 100 ns. Along the trajectory, the receptor, and ligand RMSD, the ligand dihedral angles fluctuations, as well as hydrogen bonds and contacts between the ligand and the receptor were monitored (Figures S3–S5). In either complexes, the ligand remained anchored to both h-CD22 and m-CD22 receptors until the end of the simulation, as demonstrated by the ligand RMSD values within ~1.5/2 Å, suggesting the stability of the binding poses (Figures S4A and S4C). Representative complexes selected on the basis of the cluster analysis of the dynamics (see Methods section for details) were then analyzed by means of the CORCEMA-ST program (Jayalakshmi and Krishna, 2002) that allowed the comparison between the theoretical and the experimental STD data and the validation of 3D models of the complexes (Figures 3 and 4). Thus,

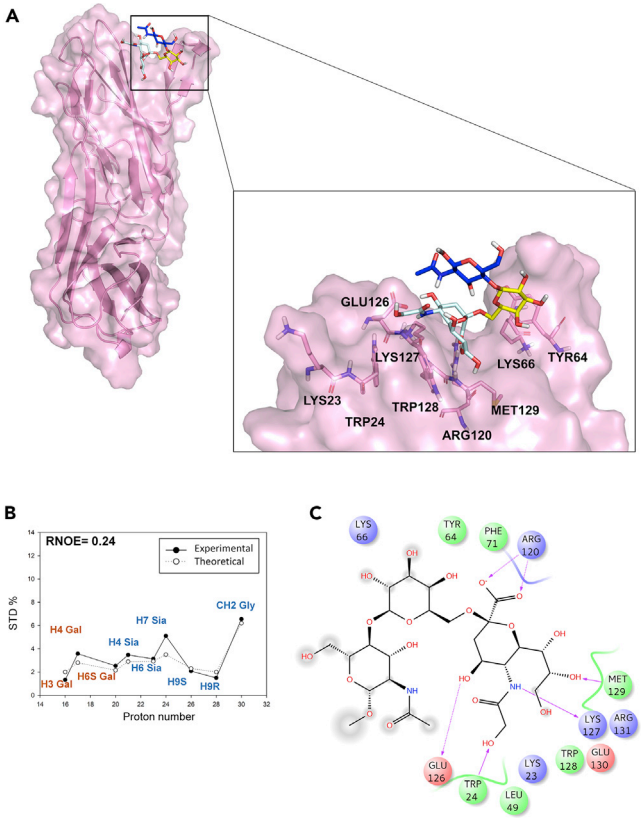


Figure 4. Interaction between h-CD22 and Neu5Gc ligand

(A) Three-dimensional model of Neu5Gc ligand bound to h-CD22 V-set domain as derived by STD, tr-NOESY, and MD.

(B) The three-dimensional h-CD22/Neu5Gc complex showing the best fit between theoretical (solid line) and experimental (dashed line) STD data derived by CORCEMA-ST analysis. (R-NOE values of 0.24).

(C) Two-dimensional plots representing the interactions between the glycolylated trisaccharide and h-CD22-binding site residues, derived from a representative frame of the MD simulation. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone. The interaction diagram was produced by Maestro 10.4.018 program.

as for the h-CD22/Neu5Gc model, several contacts between the ligand and the receptor binding pocket residues were observed; the majority of them were retained for most of the simulation time (Figure S4B). In particular, the polar network between h-CD22 and the glycolylated ligand was similar to that already described for the corresponding Neu5Ac ligand. (Di Carluccio et al., 2019) Indeed, the highly conserved Arg120 established a salt bridge between its guanidine group and Neu5Gc carboxylate. Neu5Gc glycerol moiety was involved in hydrogen bonds with Met129 backbone oxygen and amide, as well as CH-π contacts with Trp128. Also, the Neu5Gc OH-4 formed a polar interaction with Glu126. Most importantly, the *N*-glycolyl group of Neu5Gc engaged a stable hydrogen bond with Lys127 and hydrophobic contacts between its methylene protons and both Trp24 and Trp128 aromatic moieties (present 82% and 100% of the simulation time, respectively, as shown in Figure S4B). The Gal unit contributed to the receptor binding by means of CH–π interaction with Tyr64 aromatic residue (present for 100% of the simulation time). On the contrary, the GlcNAc unit was far from the h-CD22 surface for most part of the simulation, and this sugar moiety displayed higher RMSD with respect to the other sugar units (Figure S5A). Concerning the conformational behavior along the MD simulation, the bound ligand maintained the umbrella topology for 82% of the simulation (Figure S5B). Indeed, the distance between the CH_3 group of the *N*-acetyl glucosamine and the H-5 of sialic acid assumed an average value of 4.9 Å, in accordance with the NOE derived distance (Figure S5C and Table S3).

The CORCEMA-ST highlighted a good agreement between the theoretical and experimental values (Figure 4B). Indeed, the strongest STD effects in the h-CD22 complex were predicted for protons

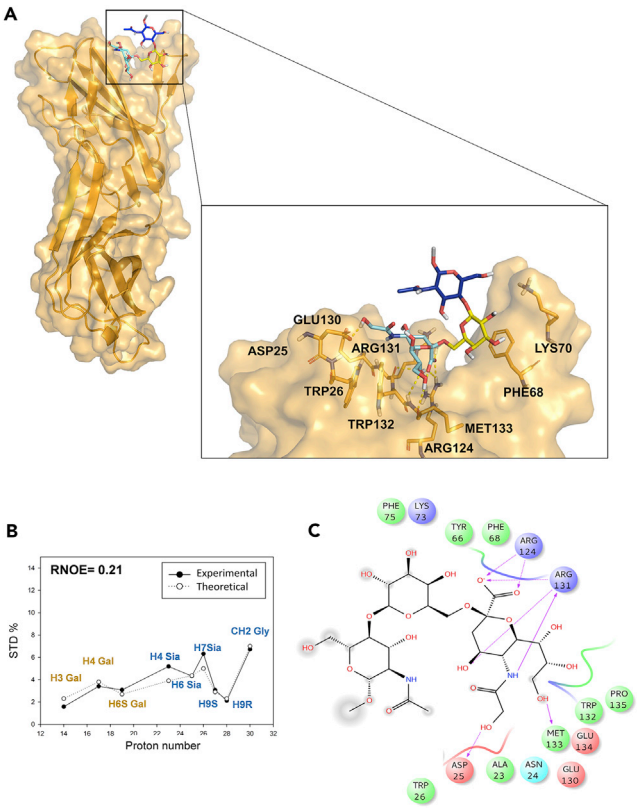


Figure 5. Interaction between m-CD22 and Neu5Gc ligand

(A) Three-dimensional model of Neu5Gc ligand bound to m-CD22 V-set domain as derived by STD, tr-NOESY, and MD.
(B) The three-dimensional m- CD22/Neu5Gc complex showing the best fit between theoretical (solid line) and experimental (dashed line) STD data derived by CORCEMA-ST analysis. (R-NOE values of 0.21).
(C) Two-dimensional plots representing the interactions between the glycolylated trisaccharide and m-CD22 binding site residues, derived from a representative frame of the MD simulation. Dotted arrows represent hydrogen bonds with functional groups from side chains and solid arrows those with functional groups of the backbone. The interaction diagram was produced by Maestro 10.4.018 program.

belonging to Neu5Gc unit; significant saturation was also estimated for some protons of the Gal unit; conversely no saturation was predicted for protons of GlcNAc units, in full agreement with the experimental STD data (Figures 2B and S1A, Table S2). The high STD value of the glycolyl group of the ligand was consistent with the close contacts with Trp24, Trp128, and Glu126 side chains. Regarding STD data of the Neu5Gc glycerol moiety, the higher STD effect observed for the H-7 is due to the strong CH-π interaction of this proton with Trp128 indole group, beyond the contacts between the entire moiety and Met129. Concerning the hydroxymethylene group, only the H-9S was oriented toward Trp128, thus exhibiting a higher STD effect with respect to H-9R. Also, H-4 and H-6 of Neu5Gc displayed significant STD effects for their vicinity to the receptor surface. For the Gal unit, considerable saturation was predicted especially for the proton at position 4, due to its close contacts with Tyr64 aromatic ring.

As for m-CD22/Neu5Gc-validated model, the ligand interaction pattern showed many similarities with its human ortholog. Still, Neu5Gc majorly contributed to the binding, interacting with Arg124, Arg131, Trp132, and Met133 receptor residues, whereas no participation of GlcNAc residue was observed (Figures 5 and S4D). Specifically, the carboxylate of Neu5Gc formed the key electrostatic interactions with the Arg124 guanidinium group. Met133 established numerous polar interactions with the hydroxyl groups of the ligand glycerol moiety. In addition, the Arg131 played the same role of Lys127 in h-CD22 receptor, thus forming a stable hydrogen bond between its backbone oxygen and the amide nitrogen of Neu5Gc *N*-glycolyl moiety, which was also in close contact with Trp26 and Trp132 aromatic residues (present for 93% and 100% of simulation time, respectively).

Differently from h-CD22 complex, the hydroxyl group of the *N*-glycolyl moiety forms polar interactions with Asp25 and Asn24 residues, which lie in proximity of the binding region, although these contacts were observed for 60% and 40% of the MD simulation, respectively. Furthermore, similarly to h-CD22 binding mode, the Gal unit is mostly engaged in CH–π interaction with Phe-68 residue (contact present for 99% of the simulation time); also, the GlcNAc residue did not interact with m-CD22, exhibiting a higher degree of fluctuation with respect to the other residues (Figure S5D). In accordance with the experimental data, the ligand retained an umbrella-like conformation for 89% of the simulation time, as supported by the average distance between the CH_3 group of the *N*-acetyl glucosamine and the H-5 of Neu5Gc along the simulation (Figures S5E and S5F).

Thus, a comparison of CORCEMA-ST results of m-CD22 and h-CD22 in complex with Neu5Gc-containing trisaccharide highlights how the orientation of the glycan inside the receptors binding pockets is comparable. The strong STD value observed for *N*-glycolyl moiety of Neu5Gc in the interaction with m-CD22 could be explained by the contacts between the hydroxymethyl group and Asp25 and Asn24 residues, beyond those observed with the Trp26 and Trp132 aromatic residues and the hydrogen bond with Arg131, analogous to that described in the human receptor. The potential involvement of the Asp25 in the recognition of *N*-glycolyl trisaccharide by m-CD22 was supported by the results of CORCEMA-ST analysis performed on several other structures lacking the hydrogen bond with the Asp25, resulting in higher R-NOE values due to the significantly lower STD value attributed to the *N*-glycolyl moiety protons (data not shown). Considering the Gal moiety, similar STD effects were predicted for the protons directed toward the residue aromatic side chain, namely H-4 and H-3, in line with conservative mutation of $Tyr64^{h}$ into $Phe68^{m}$.

STD NMR analysis revealed a comparable recognition profile of the acetylated 6'SLN by murine and human CD22

The STD NMR analysis of the interaction of Neu5Ac-containing trisaccharide [Neu5Ac-α-(2,6)-Gal-β-(1,4)-GlcNAc] with murine CD22 (Figures 2C and S1A, left panel) revealed an epitope map and a binding mode fully comparable to that of human CD22, previously characterized by our group (Di Carluccio et al., 2019). The several changes in the multiplicity and relative intensity of signals observed in the STD NMR spectrum with respect to the corresponding reference (Figure S1A) were diagnostic of the binding specificity. In detail, the sialic acid residue (**K**) mostly participated to the interaction with m-CD22, with the acetyl group giving the highest STD signal. On the contrary, the acetyl group belonging to the glucosamine residue (**A**) disappeared from the STD spectrum, highlighting its distance from the binding site. In addition, H-6 proton of sialic acid gave a good STD signal, close to 70%, followed by the protons belonging to the glycerol chain and the H5 (range of 30%–50%). The contribution of the diastereotopic H-3 protons was less remarkable (<30%). Interaction of the galactose unit (**B**) was also detected, mainly relative to protons H-4, H-5, and H-6. Notably, as further confirmation of the binding specificity, the multiplet around 3.9 ppm in the off-resonance, deriving from the overlapping of H-6 **A** and H-6 **B**, was converted into a triplet corresponding to the only H-6 Gal **B** in the STD spectrum, further evidence that the *N*-acetylglucosamine moiety was excluded from the recognition process. Additional data gathered from the construction of STD build-up curves (Figure 2C) corroborated the results obtained from the qualitative STD NMR analysis (Table S5). The above experimental data highlighted a totally comparable binding mode of Neu5Ac-containing trisaccharide with human and murine CD22.

Comparison of acetylated and glycolylated glycans interaction with murine and human CD22

To directly compare the mode of interaction of Neu5Ac/Neu5Gc with m-CD22, a computational approach was performed to establish a three-dimensional complex of m-CD22 and Neu5Ac glycans, thus defining a reliable model of interaction (Figure 6). According to our results, the Neu5Ac ligand displayed a similar orientation with respect to Neu5Gc ligand, establishing the crucial salt bridge with Arg124 through its carboxylate. The hydroxyl groups of the glycerol lateral chain interacted through hydrogen bonds with the Met133 backbone. The *N*-acetyl group was involved in the binding with Arg131 as well as hydrophobic interactions with Trp26 and Trp132 aromatic residues. The Gal residue, similarly to Neu5Gc ligand, was involved in CH-π interactions with Phe68, and the GlcNAc was far from the binding region. Thus, it can be assessed that m-CD22 interacts with Neu5Ac and Neu5Gc ligands in analogous manner. A comparison of the 3D structures of the models showed slightly different shape and polarity of the receptor cavities, which accommodate the Neu5Ac and Neu5Gc moieties (Figures 6A and 6B). Indeed, although the h-CD22 region responsible for *N*-acetyl and *N*-glycolyl binding is essentially constituted by aromatic

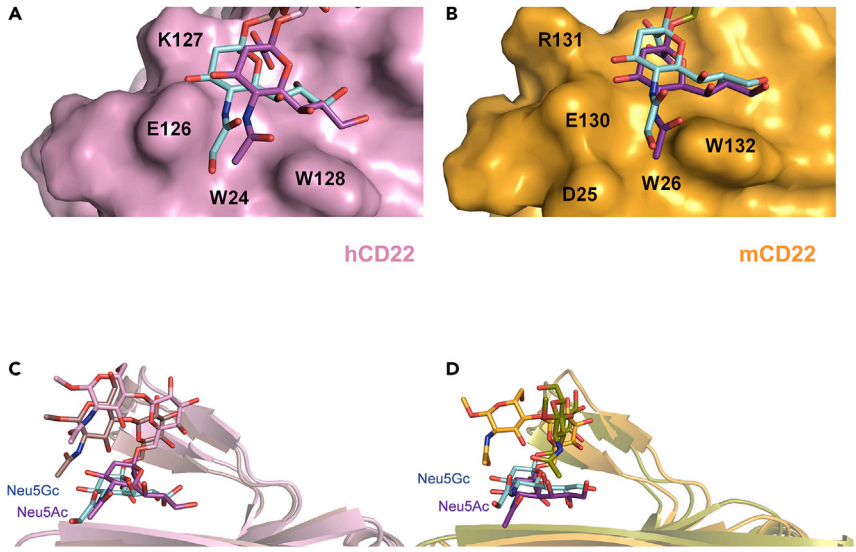


Figure 6. Comparison of the interaction of Neu5Ac/Neu5Gc ligands with m-CD22 and h-CD22

(A) Close up view of *N*-Acetyl (purple)- and *N*-glycolyl (cyan)-binding region of h-CD22, showing the protein surface (pink).

(B) Close up view of *N*-Acetyl (purple)- and *N*-glycolyl (cyan)-binding region of m-CD22, showing the protein surface (orange).

(C) Superimposition of h-CD22/Neu5Gc (pink) and h-CD22/Neu5Ac complexes (dirty violet).

(D) Superimposition of m-CD22/Neu5Gc (bright orange) and m-CD22/Neu5Ac complexes (olive).

residues, m-CD22 also comprises the Asp25 residue in optimal position to interact with the longer *N*-glycolyl chain.

Finally, MM/GBSA and MM/PBSA analysis (Srinivasan et al., 1998) was performed with Amber to have an indication of the relative binding energy of the complexes (Table 1). As result, all four complexes exhibited comparable ΔGb values, in agreement with the similar binding properties of the receptors toward the different forms of sialic acid discussed here (Figures 6C and 6D). Therefore, despite some differences in the binding regions described earlier, it is possible to assess that human and murine CD22 similarly recognize Neu5Ac and Neu5Gc ligands, in accordance with the experimental results.

DISCUSSION

The family of Sias includes nearly 50 structurally diverse members deriving from naturally occurring modifications in different positions of the original nonulosonic acid skeleton (Pearce and Läubli, 2016). Usually, Sias cap the terminal moiety of cell surfaces glycoconjugates and glycolipids as well as secreted glycoproteins are attached to the underlying glycan via α-(2→3), α-(2→6), or α-(2→8) linkages. By virtue of their location, diversity, and ubiquity in vertebrates, Sias serve as ligands of endogenous and exogenous glycan-binding proteins, thus representing important regulators of several biologically relevant recognition processes (Varki, 2008). Sialylated glycans from human cells mainly terminate with the 5′-acetylated isoform of neuraminic acid, Neu5Ac, because of the evolutionary loss of the Neu5Gc, which differs from Neu5Ac by one additional oxygen atom. In contrast with the majority of other mammals, humans do not possess the ability to synthesize Neu5Gc; it can, however, be metabolically incorporated from dietary sources, i.e. red meat, and becomes present on some epithelial and endothelial cell surfaces (Alisson-Silva et al., 2018). Also, the presence of Neu5Gc on nontypeable *Haemophilus influenzae* (NTHi) LOS (lipooligosaccharide) that forages exogenous sialic acids from the host was recently demonstrated (Ng Preston, 2018). Being Neu5Gc a xeno-autoantigen mostly present on malignant cells, the interaction between inhibitory CD22 and Neu5Gc glycans may indeed play a key role in the regulation of the tumor immune response (Samraj et al., 2014), impact on cancer progression, and increase of tumor-associated inflammation. Here, we evaluated the effect of sialic acid glycolylation on the binding with CD22, comparing the behavior of different ligands in complex with murine and human CD22, improving the knowledge of the structural basis of the recognition of sialylated *N*-glycans from CD22 receptor.

Table 1. Relative binding energies of h-CD22 and m-CD22 with acetylated and glycolylated ligands

Complex	ΔGb (MM/GBSA)	ΔGb (MM/PBSA)
h-CD22/Neu5Gc	−41.24 ± 0.15	−12.66 ± 0.16
h-CD22/Neu5Ac	−38.43 ± 0.18	−10.75 ± 0.18
m-CD22/Neu5Gc	−38.29 ± 0.16	−11.73 ± 0.19
m-CD22/Neu5Ac	−34.69 ± 0.26	−10.72 ± 0.18

All units are expressed in kcal/mol.

NMR analysis revealed a comparable binding epitope of both murine and human CD22 toward glycolylated glycans (Figures 2A–2C and S1). In both complexes a bent umbrella-like conformation of the ligand was adopted, as supported by the results achieved from the NMR data and MD simulations (Figures S1B, S1C, S5C, and S5F). From molecular modeling, it was confirmed that Neu5Gc/Neu5Ac ligands displayed a similar orientation inside the binding site of murine and human CD22, independently from the Sia nature, as supported by a comparison of the molecular surfaces of h-CD22 and m-CD22 in the interaction with *N*-acetyl/*N*-glycolyl Sia chains, in Figure 6. Furthermore, it was evidenced the possibility for m-CD22 of forming additional interactions with Neu5Gc ligand, mainly involving the hydroxymethyl group of *N*-glycolyl moiety. This was further supported by a comparison of the molecular surfaces of h-CD22 and m-CD22 deputed to interact with *N*-acetyl/*N*-glycolyl Sia chains. In addition, the substitution of $Asp25^{m}$ in place of $Lys23^{h}$ in the binding subsite of m-CD22 influences the possibility to establish hydrogen bonds with the hydroxyl group of the glycolyl moiety.

Overall, our studies indicate that, despite the different nature of sialic acid residue, the recognition region of h-CD22 is almost invariant comparing Neu5Ac and Neu5Gc containing glycans. These results agree with the similar affinity of h-CD22 toward Neu5Gc/Neu5Ac structures, as recently reported by Angata (Angata, 2018). In conclusion, the obtained outcomes provide a global vision of how the most diffuse neuraminic acid forms of sialylated *N*-glycans in mammals are arranged in the binding pocket of CD22. Hence, potential high-affinity analogues of ligands naturally recognized from the CD22 could be specifically designed and synthetized for targeting the receptor protein in order to impede the biological interaction, thus modulating immune responses.

Limitations of the study

This study reports the interactions of the biologically relevant CD22, in human and murine variants, with the two predominant forms of sialic acid in vertebrates. A comparison of the binding mode was carried out mainly by NMR spectroscopy, molecular modeling, and CORCEMA-ST calculations. In the absence of high-resolution coordinates of the murine CD22, we employed homology modeling and different computational approaches to validate the generated models (docking, MD simulations, CORCEMA-ST). Despite the high homology with respect to the template, the structural conclusions we derived from homology modeling will require the definition of the three-dimensional structure of the murine CD22 to avoid misinterpretation of the data.

Resource availability

Lead contact

Roberta Marchetti (roberta.marchetti@unina.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This study did not generate datasets.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2020.101998.

ACKNOWLEDGMENTS

This study was supported by PRIN 2017 "Glytunes" (2017XZ2ZBK, 2019-2022) to AS and SS; progetto POR SATIN and Progetto POR Campania Oncoterapia to AM; the European Commission (H2020-MSCA-814102–SWEET CROSSTALK project) to AM, RM, and AS; and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 851356 to RM. PON Ricerca e Innovazione 2014-2020, Azione I.1 "Dottorati Innovativi con caratterizzazione Industriale" is acknowledged for funding the Ph.D. grant to R.E.F.

This paper is dedicated to Prof. Jesús Jiménez-Barbero for his 60th birthday.

AUTHOR CONTRIBUTIONS

A.S. and R.M. conceived and designed the project. R.M., R.E.F., C.D.C., A.S., and A.M. carried out NMR, docking, and CORCEMA experiments and analyzed the results. R.E.F. performed homology modeling. R.E.F., R.M., A.S., S.S., M.M., and M.C. carried out and analyzed MD simulations. P.R.C. and G.S. produced the proteins. K.F. and Y.M. provided the glycans. All the authors wrote, revised, and reviewed the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 5, 2020
Revised: November 27, 2020
Accepted: December 23, 2020
Published: January 22, 2021

REFERENCES

Ajit, V., and Angata, T. (2006). Siglecs—the major subfamily of I-type lectins. Glycobiology *16*, 1–27.

Alisson-Silva, F., Liu, J.Z., Diaz, S.L., Deng, L., Gareau, M.G., Marchelletta, R., Chen, X., Nizet, V., Varki, N., Barrett, K.E., and Varki, A. (2018). Human evolutionary loss of epithelial Neu5Gc expression and species-specific susceptibility to cholera. PLoS Pathog. *14*, e1007133.

Altman, M.O., and Gagneux, P. (2019). Absence of Neu5Gc and presence of anti-Neu5Gc antibodies in humans-an evolutionary perspective. Front. Immunol. *10*, 789.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Angata, T. (2018). Possible influences of endogenous and exogenous ligands on the evolution of human Siglecs. Front. Immunol. *10*, 789.

Angulo, J., and Nieto, P.M. (2011). Std NMR: application to transient interactions between biomolecules—a quantitative approach. Eur. Biophys. J. *40*, 1357–1369.

Di Carluccio, C., Crisman, E., Manabe, Y., Forgione, R.E., Lacetera, A., Amato, J., Pagano, B., Randazzo, A., Zampella, A., Lanzetta, R., et al. (2019). Characterization of the dynamic interactions between complex N-glycans and human CD22. Chembiochem *21*, 129.

Di Carluccio, C., Forgione, R.E., Molinaro, A., Crocker, P., Marchetti, R., and Silipo, A. (2021). Exploring the fascinating world of sialoglycans in the interplay with Siglecs. Carbohydr. Chem. *44*, 31–55.

Case, D.A., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Ghoreishi, D., Gilson, M.K., et al. (2018). AMBER 2018 (University of California).

Chandrasekaran, A., Srinivasan, R., Raman, K., Viswanathan, S., Raguram, T.M., Tumpey, V., and Sasisekharan, R. (2008). Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. Nat. Biotechnol. *26*, 107–113.

Dörner, T., Shock, A., and Smith, K.G. (2012). CD22 and autoimmune disease. Int. Rev. Immunol. *31*, 363–378.

Duan, S., and Paulson, J.C. (2020). Siglecs as immune cell checkpoints in disease. Annu. Rev. Immunol. *38*, 365–395.

Ereño-Orbea, J., Sicard, T., Cui, H., Mazhab-Jafari, M.T., Benlekbir, S., Guarné, A., L Rubinstein, J., and Julien, J.P. (2017). Molecular basis of human CD22 function and therapeutic targeting. Nat. Commun. *8*, 764.

Forgione, R.E., Di Carluccio, C., Guzmán-Caldentey, J., Gaglione, R., Battista, F., Chiodo, F., Manabe, Y., Arciello, A., Del Vecchio, P., Fukase, K., et al. (2020). Unveiling molecular recognition of sialoglycans by human siglec-10. iScience *23*, 101–401.

Han, S., Collins, B.E., Bengtson, P., and Paulson, J.C. (2005). Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking Nat. Chem. Biol. *1*, 93–97.

Häubli, H., and Varki, A. (2020). Sialic acid–binding immunoglobulin-like lectins (Siglecs) detect self-associated molecular patterns to regulate immune responses. Cell. Mol. Life Sci. *77*, 593–605.

Jayalakshmi, V., and Krishna, N.R. (2002). Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. J. Magn. Reson. *155*, 106–118.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck - a program to check the stereochemical quality of protein structures. J. App. Cryst. *26*, 283–291.

Macauley, M.S., Crocker, P.R., and Paulson, J.C. (2014). Siglec-mediated regulation of immune cell function in disease. Nat. Rev. Immunol. *14*, 653–666.

Mahajan, V.S., and Pillai, S. (2016). Sialic acids and autoimmune diseases. Immunol. Rev. *269*, 145–161.

Marchetti, R., Perez, S., Arda, A., Imberty, A., Jimenez-Barbero, J., Silipo, A., and Molinaro, A. (2016). "Rules of engagement" of protein–glycoconjugate interactions: a molecular view achievable by using NMR spectroscopy and molecular modeling. ChemistryOpen *5*, 274–296.

Meyer, B., and Peters, T. (2003). NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew. Chem. Int. Ed. Engl. *21*, 864–890.

Morris, G.M., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (2009). AutoDock treats the ligand as a flexible unit and the protein as a rigid unit. J. Comput. Chem. *30*, 2785–2791.

Müller, J., and Nitschke, L. (2014). The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease. Nat. Rev. Rheumatol. *10*, 422–428.

Ng Preston, S.K. (2018). Nontypeable Haemophilus influenzae has evolved preferential use of N-acetylneuraminic acid as a host adaptation. mBio *10*, 422–519.

Okerblom, J., and Varki, A. (2017). Biochemical, cellular, physiological, and pathological consequences of human loss of N-glycolylneuraminic acid. ChemBioChem *18*, 1–18.

Pearce, O.M.T., and Läubli, H. (2016). Sialic acids in cancer biology and immunity. Glycobiology *26*, 111–128.

Samraj, A.N., Läubli, H., Varki, N., and Varki, A. (2014). Involvement of a non-human sialic Acid in human cancer. Front. Oncol. *4*, 33.

Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A., and Case, D.A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices. J. Am. Chem. Soc. *120*, 9401–9409.

Varki, A. (2008). Sialic acids in human health and disease. Trends Mol. Med. *14*, 351–360.

Varki, A. (2017). Are humans prone to autoimmunity? Implications from evolutionary changes in hominin sialic acid biology. J. Autoimmun. *83*, 134–142.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). Homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, 296–303.

Yan, A.D.K., Mo, H., Shapiro, M.J., and Zartler, E.R. (2003). The effect of relaxation on the epitope mapping by saturation transfer difference NMR. J. Magn. Reson. *163*, 270–276.

Zhou, X., Yang, G., and Guan, F. (2020). Biological functions and analytical strategies of sialic acids in tumor. Cells *9*, 273.

iScience, Volume *24*

Supplemental Information

Behavior of glycolylated sialoglycans in the binding pockets of murine and human CD22

Cristina Di Carluccio, Rosa Ester Forgione, Marco Montefiori, Monica Civera, Sara Sattin, Giovanni Smaldone, K. Fukase, Y. Manabe, Paul R. Crocker, Antonio Molinaro, Roberta Marchetti, and Alba Silipo

Supplemental data

Figure S1. [Comparison of the binding mode of Neu5Ac/Gc containing ligands when interacting with h-/m- CD22], related to Figure 2 and Table S3. a) Left panel: STD-NMR of Neu5Ac ligand interacting with m/h-CD22. Right panel: STD-NMR of Neu5Gc ligand interacting with m/h-CD22. STD NMR analyses were performed using a protein/ligand molecular ratio of a 1:100 and saturation time of 2s. Neu5Ac/Gc ligands 3D epitope maps are also shown. **b)** Tr-NOESY NMR spectrum of the glycolylated trisaccharide in the bound state with h-CD22, using a mixing time of 400ms. **c)** Tr-NOESY NMR of the glycolylated trisaccharide bound to m-CD22, using a mixing time of 400ms. The ligand 6'SLN upon binding with both h- and m-CD22 adopts an umbrella-like topology, depending on the parameter θ, defined as the angle between the carbon C-2 of Sia and C-1 atoms of the Gal and GlcNAc residues, that assumes a value < 110°. The experimental error in the calculated proton-proton distances is estimated to be less than ± 10%.

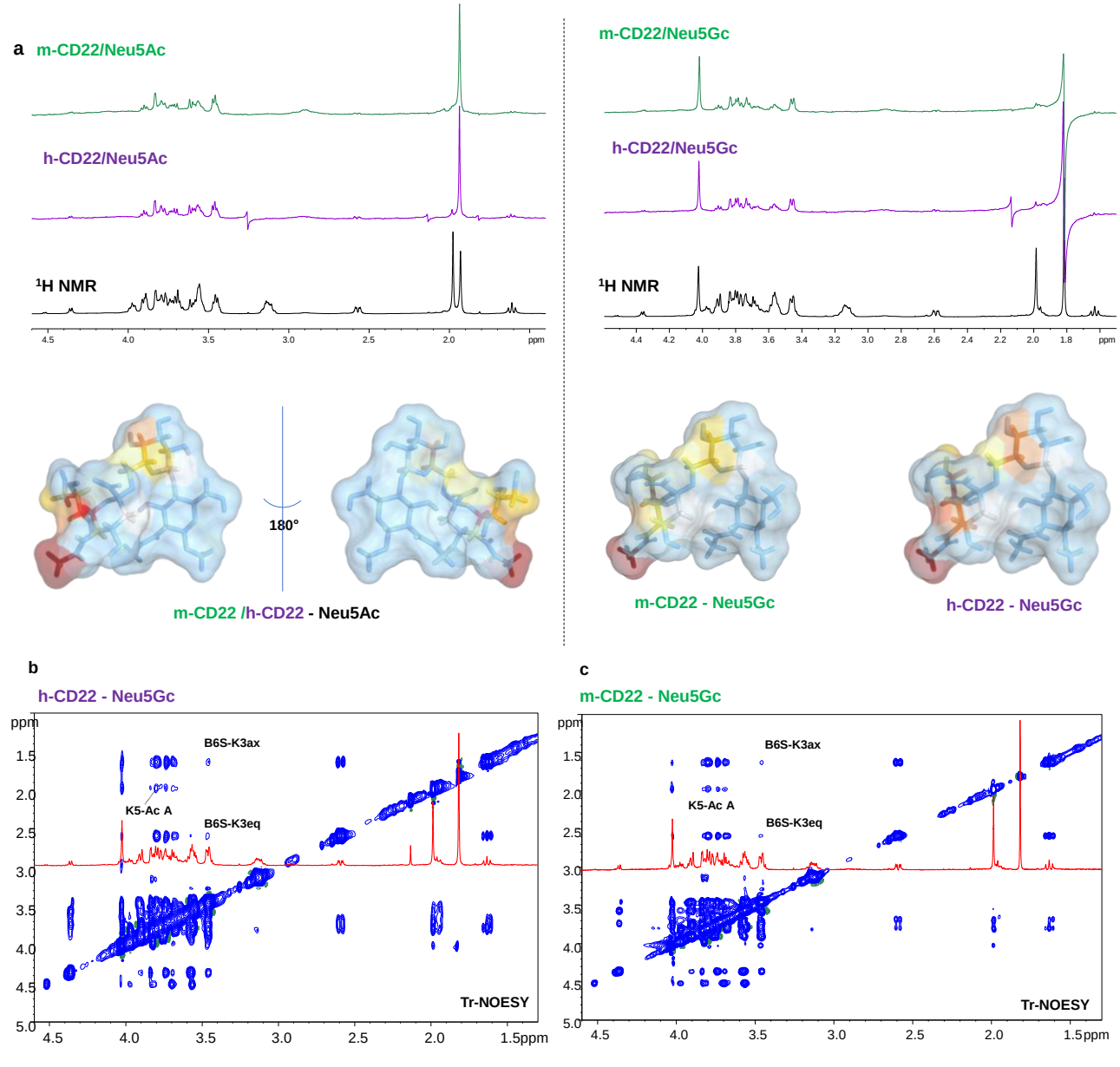


Figure S2. [MD simulation analysis of the glycolylated trisaccharide in its free state], related to Figure 2 and Table S3

a) φ/ψ/ω dihedral angles of Neu5Gc-Gal linkage along the MD trajectory.
b) φ/ψ/ dihedral angles of Gal-GlcNAc linkage along the MD trajectory.
c) $H3_{eq}$ Neu5Gc – H6S/H6R Gal inter-ligand distances.
d) $H3_{ax}$ Neu5Gc – H6S/H6R Gal inter-ligand distances.

The torsion angles were defined as follows: ϕ (C1-C2-O-C'6), ψ (C2-O-C'6- C'5), ω (O−C'6−C'5−O'5).

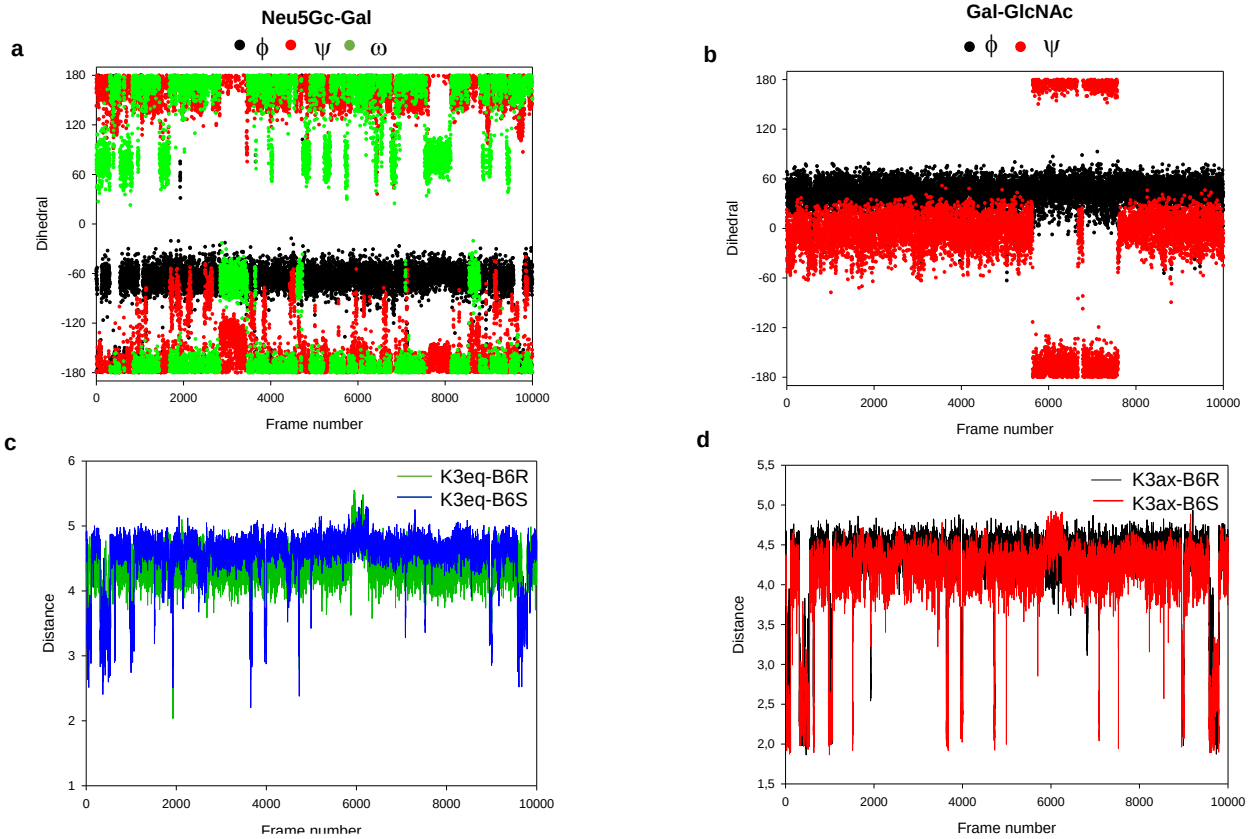


Figure S3. [MD simulation analysis of m-CD22 homology modelling], related to Figure 3.

a) Superimposition of the m-CD22 structures each 10 ns of the MD simulations. Along the MD simulation, no relevant conformational changes emerged.

b) Backbone RMSD of the protein, CC' loop (res 69-74), GG' loop (res 127-130), depicted in black, red and green respectively. The fluctuations of the backbone RMSD of the CC' loop, can be attributed to a dynamic equilibrium between a disordered (high RMSD) and partially ordered (low RMSD) forms of the region.

c) Atomic fluctuation by residue of m-CD22 structure, calculated using the protein Cα atoms. The peaks in the RMSF plot corresponded to the mobile loops connecting the ß-strands, in both V-set and C2-set Ig-like domains.

d) Plot of the potential energy variation of m-CD22 structure along the MD. The structure with the lowest potential energy was considered for the docking calculations.

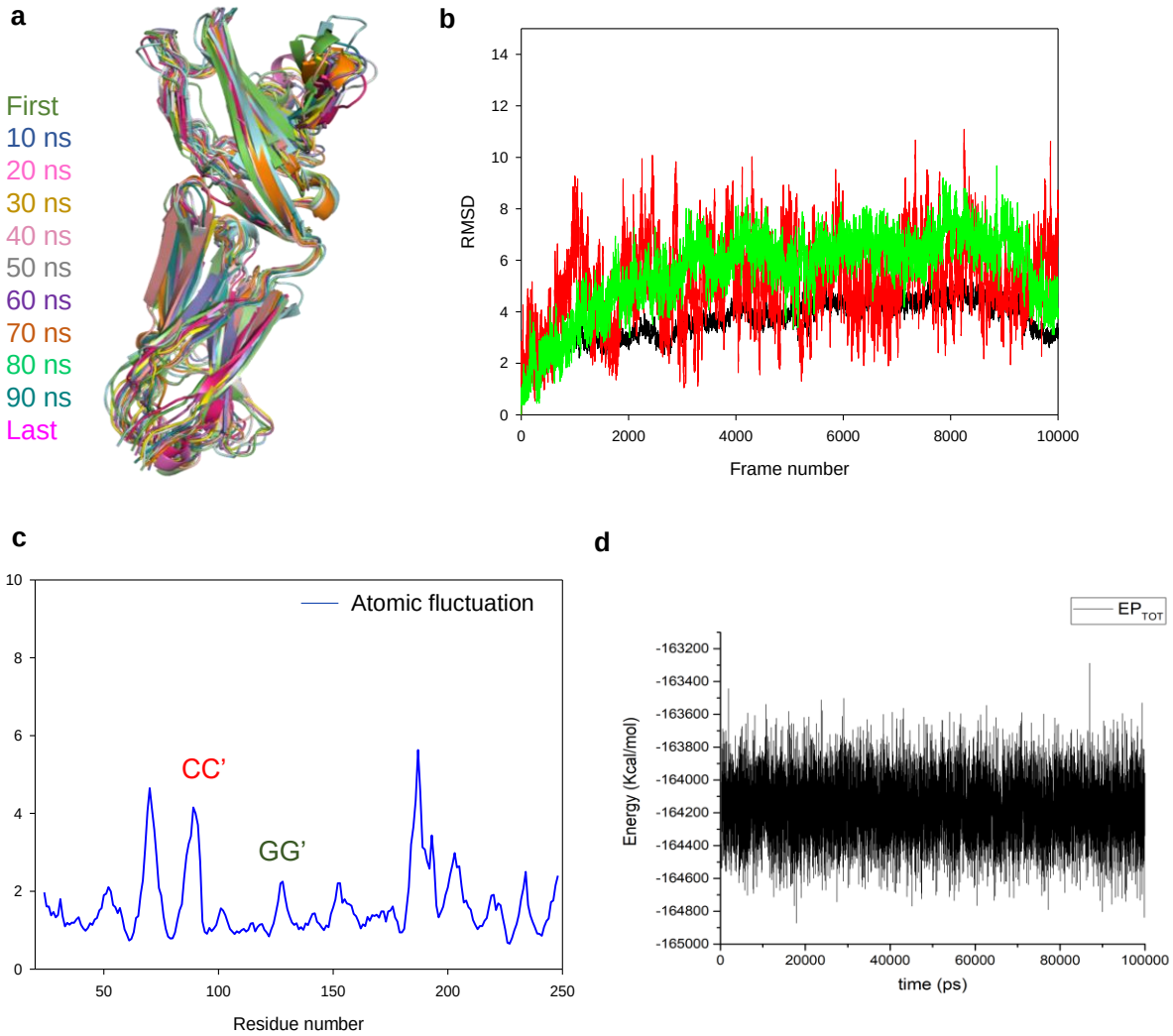


Figure S4. [MD simulation analysis of h-CD22/m-CD22 complexes with the glycolylated ligand], related to Figures 4 and 5.

a) h-CD22 and Neu5Gc ligand RMSD variation along the MD. The ligand RMSD was measured with respect to the protein.
b) Frequency of most representative h-CD22/Neu5Gc inter-molecular distances. A distance cut-off of 5Å was considered for the calculation.
c) m-CD22 and Neu5Gc RMSD variation along the MD. A distance cutoff of 5Å was considered for the calculation.
d) Frequency of most representative m-CD22/Neu5Gc inter-molecular distances.

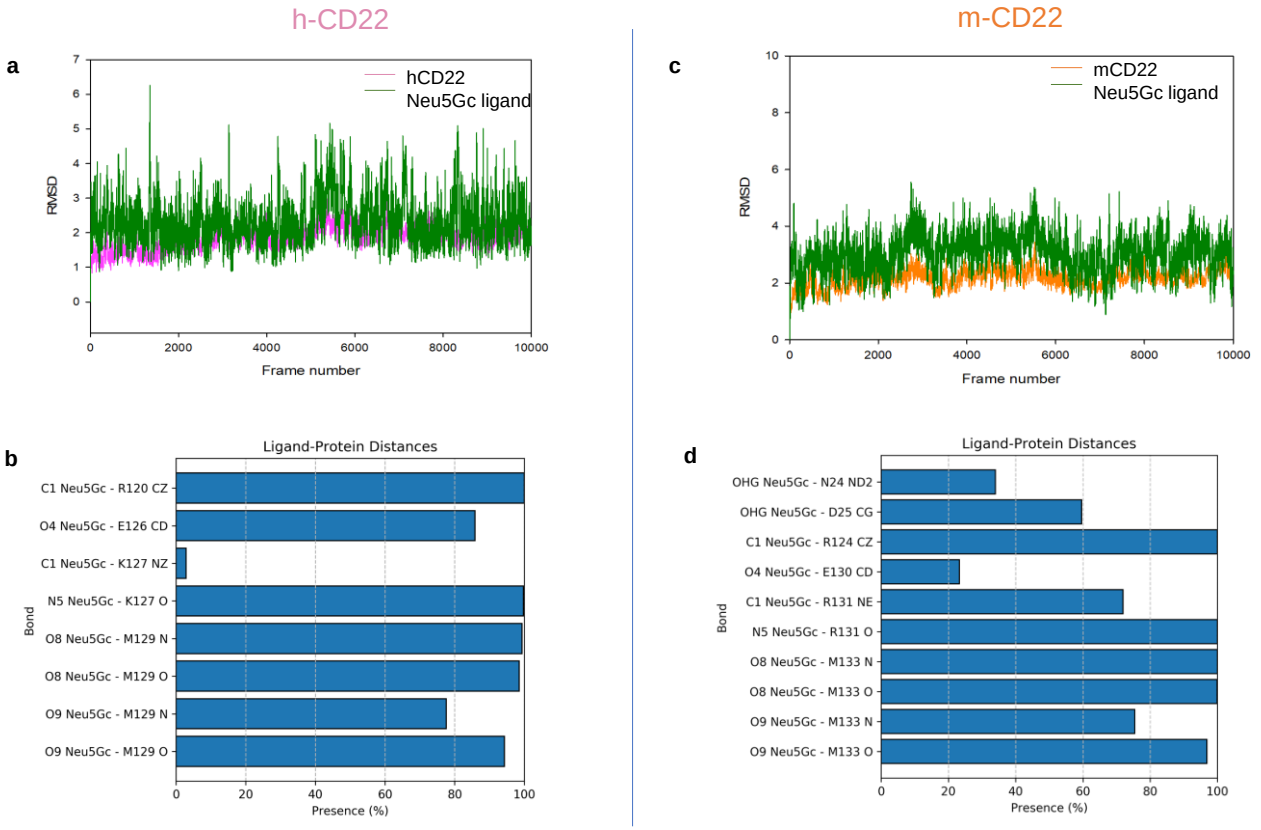


Figure S5. [Analyses of the glycolylated ligand conformation when bound to h-CD22 and m-CD22], related to Figures 4 and 5.

a) RMSD of Neu5Gc ligand residues with respect to h-CD22 protein.
b) Distance between H5 of Neu5Gc and GlcNAc acetyl group C (CME) (average value 5.1 Å).
c) Variation of Neu5Gc θ angle value along h-CD22/Neu5Gc complex simulation. The parameter θ is defined as the angle between the carbon C-2 of Sia and C-1 atoms of the Gal and GlcNAc residues.
d) RMSD of Neu5Gc ligand residues with respect to the m-CD22 protein.
e) Distance between H5 of Neu5Gc and GlcNAc acetyl group C (CME) (average value 4.9 Å).
f) Variation of Neu5Gc θ angle value along m-CD22/Neu5Gc complex simulation.

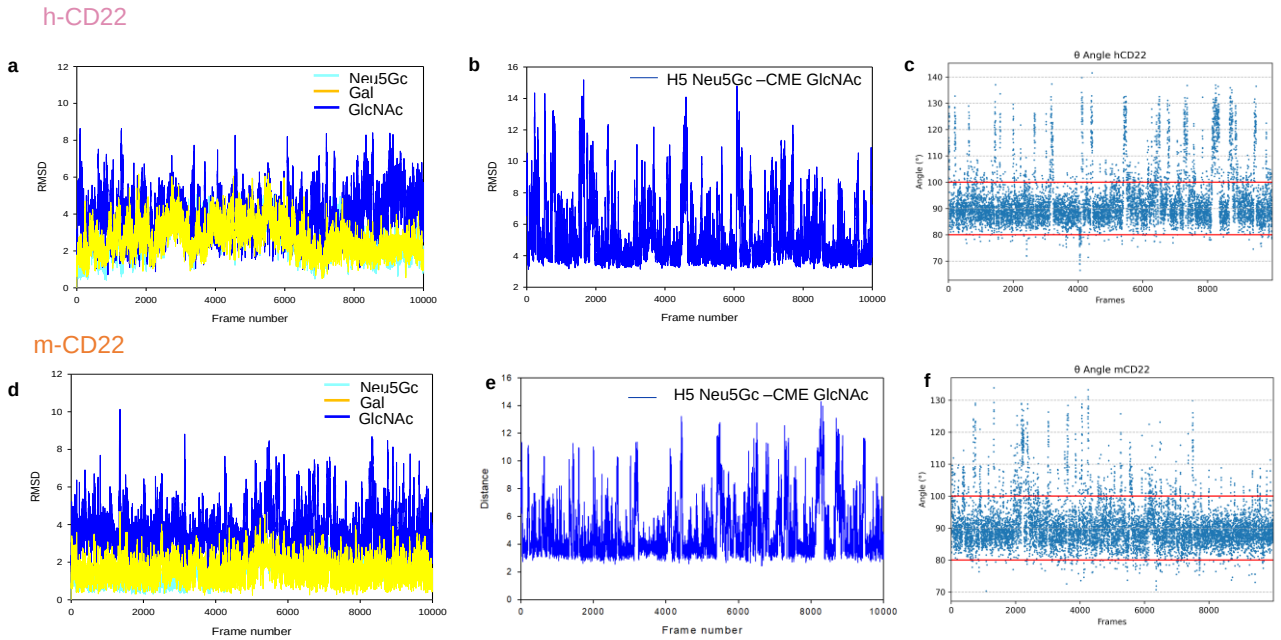
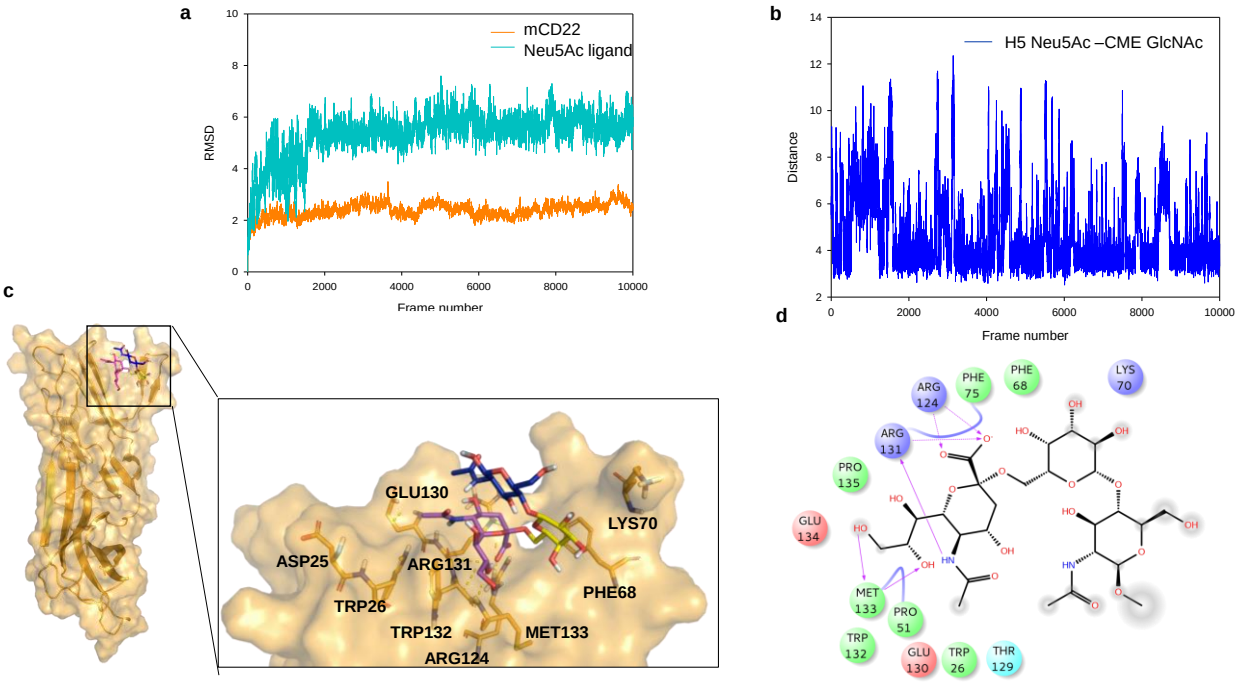


Figure S6. [Interaction between m-CD22 and Neu5Ac ligand], related to Figure 6

a) Protein and ligand RMSD variation along the MD. The ligand RMSD was measured with respect to the protein.
b) Distance between H5 of Neu5Gc and GlcNAc acetyl group C (CME) (average value 4.6 Å)
c) Three-dimensional model derived by STD, tr-NOESY and MD for the Neu5Ac ligand bound form (*gt* conformer) to m-CD22 homology model. The representative frame from the most populated MD cluster was considered to depict the complex.
d) Two-dimensional plots representing the interactions between the glycolylated trisaccharide and the binding site residues of m-CD22. The representative frame of the most populated MD cluster was considered to depict the complex.



SUPPORTING TABLES

Table S1. [**Experimental STD intensities of glycolylated 6'SLN bound to m-CD22], related to Figure 2a.** STD^{max} values were evaluated by fitting the data to a monoexponential equation: $STD = STD^{max}(1 - e^{-k_{sat}t_{sat}})$ (see the Methods for more information).

^1H	STD^{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH_2 Neu5Gc	9.4963	0.8017	7.6132	**100**
H7 Neu5Gc	6.9388	0.6038	4.1897	**55.0**
H4 Gal	6.1757	0.6239	3.8530	**50.6**
H5 Gal	5.7199	0.6108	3.4937	**45.9**
H9S Neu5Gc	4.1764	0.7117	2.9723	**39.0**
H6R Gal	3.5266	0.7722	2.7232	**35.8**

Table S2. [**Experimental STD intensities of glycolylated 6'SLN bound to h-CD22], related to Figure 2b.**

^1H	STD^{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH_2 Neu5Gc	8.3838	0.6205	5.2021	**100**
H7 Neu5Gc	6.6503	0.5120	3.4049	**65.4**
H4 Gal	5.4644	0.4935	2.6967	**51.8**
H9S Neu5Gc	3.8238	0.6113	2.3375	**44.9**
H6R Gal	3.0547	0.7042	2.1511	**41.3**

Table S3. [Theoretical and experimental ^1H-^1H inter-proton distances of the glycolylated trisaccharide in the free and bound states with human and murine CD22], related to Figures 21b,c. Estimated error 5–10%.

Distances	Family I Φ = -60° Ψ = 180° ω = 60°	Family II Φ = 180° Ψ = 180° ω = 60°	Free state Exp. distances	Exp. h-CD22 bound state	Exp. m-CD22 bound state
$H3_{eq}$ Neu5Gc - H6S Gal	4.93	3.84	4.50	4.61	4.72
$H3_{eq}$ Neu5Gc - H6R Gal	4.58	3.37	nd	nd	nd
$H3_{ax}$ Neu5Gc - H6S Gal	4.43	2.53	4.11	4.80	4.90
$H3_{ax}$ Neu5Gc - H6R Gal	4.25	2.35	nd	nd	nd
H5 Neu5Gc - CH_3 GlcNAc	4.30	9.60	nd	4.96	4.81

Table S4. [Cluster rank, Cluster population, computed binding energy and RMSD (Root Mean Square Deviation) for the molecular docking (AutoDock) of m-CD22/ligand and h-CD22/ligand complexes], related to Figures 4 and 5

Complex	Cluster Rank	No cluster conformations	Estimated free energy of binding (kcal/mol)	RMSD from reference structure (Å)
h-CD22	1	154	-2.57	2.37
m-CD22	3	85	-1.94	3.23

Table S5. [Experimental STD intensities of the acetylated ligand bound to m-CD22], related to Figure 2c.

^1H	STD^{max}	K_{sat}	STD (fit)	% STD epitopes (fit)
CH_3 Neu5Ac	5.7320	0.4884	2.7995	**100**
H6 Neu5Ac	3.7105	0.5080	1.8849	**67.3**
H4 Gal	2.9244	0.5187	1.5169	**54.2**
H5 Gal	2.1719	0.5580	1.2119	**43.3**
H6R Gal	1.5247	0.6170	0.9407	**33.6**
H5 Neu5Ac	1.6331	0.5256	0.8584	**30.6**
$H3_{eq}$ Neu5Ac	1.4283	0.3632	0.5188	**18.5**

Transparent Methods

Protein expression and purification. The plasmids encoding for the three N-terminal Ig-like domains of human CD22 and murine CD22, respectively fused to the Fc region of mouse IgG2b and human IgG1, were expressed in CHO cell lines and purified as described elsewhere (Di Carluccio, et al., 2019).

Fluorescence titration. Steady-state fluorescence spectra have been collected on a Fluoromax-4 spectrofluorometer (Horiba, Edison,NJ, USA) at the fixed temperature of 10°C. Emission spectra were recorded in the emission range of 300–500 nm upon excitation at 285 nm. The slit widths were fixed at 4 nm for the excitation and 10 nm for the emission wavelength. A quartz cuvette with a path length of 1 cm and 0,2 mL volume was used. m-CD22 and h-CD22 solutions at fixed concentration of 0.25 μM in PBS buffer (pH 7.4) were titrated by adding small aliquots of a ligand stock solution of 100 μM) of Neu5Ac and Neu5Gc ligands. The fluorescence of both proteins found to quench in the presence of the ligands. The binding curve was obtained by plotting $\Delta F/\Delta F_{max}$ values versus ligand concentration and fitting the data through non-linear regression using the function described by Ribeiro et al (Ribeiro et al., 2008):

$\frac{\Delta I_f}{I_0} = \frac{\Delta I_{max}}{I_0} X_{FY}$ *where* $X_{FY} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

$a=[F]_tK_b$, $b=1+[Y]_tK_b$, $c=[Y]_tK_b$ where $[F]_t$ and $[Y]_t$ represent the total concentration of protein and ligand, respectively.

NMR analysis. Samples were prepared using 50mM phosphate deuterated buffer, pH 7.4. All NMR experiments were recorded on a Bruker AVANCE NEO 600-MHz equipped with a cryo probe and the analyses were performed with TOPSPIN 3.2 software.

Tr-NOESY analysis. Homonuclear 2D ^{1}H-^{1}H ROESY and ^{1}H-^{1}H NOESY experiments were carried out at 298°K by using data sets of 4096x256 points and mixing times of 600 ms for the free state and 400 ms for the bound states. Proton – proton cross relaxation rates (σ_{ij}) were measured integrating the ROE/NOE cross peaks of interest normalizing against the corresponding cross peak on the diagonal in F1. The experimental distances (r_{ij}) were calculated by employing the isolated spin pair approximation using as reference the intra-residue distance H1-H5 of the N-acetylglucosamine residue as 2.6Å.

STD NMR analysis. STD NMR experiments were acquired with 32 k data points and zero-filled up to 64 k data points prior to processing. 40 Gauss pulses with a length of 50 ms were used to selectively irradiate the protein resonances, setting the on-resonance pulse at 7.5 ppm and the off-resonance pulse frequency at 40 ppm. To suppress the water signals, an excitation sculpting with gradient pulses (esgp) was applied. A protein/ligand molar ratio of 1:100 was used for all systems. The fractional STD effects were calculated by use of $(I_0 – I_{sat})/I_0$, with I_{sat} the intensity of the signal in the STD NMR spectrum and I_0 the peak intensity of an unsaturated reference spectrum (off-resonance). The STD curves were acquired at different saturation times, from 0.8 to 5s. The STD build up curves were performed by fitting the saturation time data to a monoexponential equation of the form: $STD = STD^{max}(1 - e^{-k_{sat}t_{sat}})$, where STD stands for the STD signal intensity corresponding to the saturation transfer of a given proton at a saturation time t_{sat}, STD^{max} represents the asymptotic maximum of the curve, and k_{sat} is the observed saturation rate constant that measures the speed of STD build-up. The value of STD_{fit} was derived by the slope of the STD build-up curve at a saturation time of 0. Once calculated both STD_{fit} and K_{sat} values, all the intensities of different protons ligand were normalized to the largest STD_{fit}, giving $STD_{epitopes\ fit}$.

Homology modeling. The sequence encoding for m-CD22 (Uniprot: NP_033975.3) was obtained from NCBI (http://www.ncbi.nlm.nih.gov). For computational 3D structure calculation by homology modeling, the extracellular V-set, and C2 set domains of murine CD22 were considered. The sequence interval corresponding to the extracellular portion was aligned to hCD22 template (PDB: 5VKJ), using BLAST (Altschul et al., 1990) homology model was generated by means of SWISS-MODEL (Waterhouse et al., 2018). Then, the obtained structure was subjected to 100ns molecular dynamics (MD) simulations for geometry optimization and to evaluate the stability of the model.

Molecular dynamics simulations. To run the MD simulation of h-CD22 and m-CD22, only the corresponding V-set domain and adjacent C2-set domain were considered (19-355). Missing residues in h-CD22 CC' loop were added with the help of ModLoop (Fiser et al., 2000), prior to MD simulation, the structure was then refined; for each system missing hydrogen atoms were added, and protonation state of ionisable groups was computed using Maestro Protein Preparation Wizard (Schrodinger, 2012). MD simulations were carried out using AMBER 18 suite of programs (Case et al., 2018) to investigate the ligands behavior in solution, to assess the stability of the homology models, the mobility of the loops and the stability of the docking poses. Atom types and charges were assigned according to AMBER ff14SB force field for the proteins and GLYCAM-06j-1 force field to represent the ligands. By using the Leap module, the proteins and ligands were hydrated with octahedral boxes containing explicit TIP3P water molecules buffered at 10 Å, also, Na+ counter ions were added to neutralize the system by using the Leap module. The systems minimization was performed using Sander and MD simulations were carried out using the CUDA, which are distributed within the AMBER 18 package.
The smooth particle mesh Ewald method was used to represent the long-range electrostatic interactions in the system while each simulation was under periodic boundary conditions, and the grid spacing was set to 1 Å. In the equilibration procedure, the system was minimized by applying a restriction to the protein which was gradually released in the following steps. Then slow system thermalization from 0°K to 300 °K was carried out applying a solute restraint. Temperature was increased from 0°K to 100°K at constant volume. Then, from 100°K to 300°K in an isobaric ensemble. Thereafter, temperature was kept constant at 300 °K during 50 ps with progressive energy minimizations and solute restraint. Once completed the restraints were removed and the systems then advanced in an isothermal-isobaric ensemble along the production.
Concerning the complex MD simulation, an harmonic restraint to the ligand ω dihedral angle between Neu5Gc and Gal unit was applied to keep its value to keep its value around 60 degrees. considering the *gt* bioactive conformation derived from NOE experimental data.
Coordinates were archived in order to acquire 10000 structures of the progression of the dynamics. Trajectories were analyzed with the ptraj module included in the AMBER18 and visualized with VMD molecular visualization program. Each trajectory was submitted to cluster analysis with respect to the ligand RMSD using K-mean algorithm implemented in ptraj module. The representative structure of the most populated cluster was considered to depict the complexes interactions.

Ligand-protein docking calculations. Preparation of the macromolecules. The crystal structure of h-CD22 and m-CD22 refined 3D coordinates were used for docking purposes. Each structure was then submitted to 100000 steps of steepest descent minimization with OPLS3 force field using MacroModel (Schrödinger Release 2020-2, 2020) before being used for docking calculations.
Building of ligands. The 3D coordinates of Neu5Gc-α-(2-6)-Gal-β-(1-4)-GlcNAc and Neu5Ac-α-(2-6)-Gal -β-(1-4)-GlcNAc were built by means of Glycam (Woods Group, 2005-2020).The ligands geometries were optimized by 100000 step of steepest descent minimization with OPLS3 force field using Macro Model. Ligands were prepared for docking calculations using AutoDockTools, setting all rotatable bonds free to move during the docking calculations. An MD simulation, to investigate the conformational behavior of Neu5Gc-α-(2-6)-Gal-β-(1-4)-GlcNAc was also performed.
Docking calculations. Docking calculations of all compounds were performed using AutoDock 4.2.2 (Morris et al., 2009). Analysis of the docking poses was performed with AutoDockTools. The docking protocol was validated by carrying out the docking of CD22 crystallographic structure in complex with Neu5Ac-α-(2-6)-Gal ligand (PDB: 5VKM). The 3D structure of Sia-α-(2-6)-Gal was extracted from the crystallographic structure of CD22.The grid point spacing was set at 0.375 Å, and a hexahedral box was built with x, y, z dimensions: 64 Å, 46 Å, 56 Å centered in the centroid position among the binding site residues. A total of 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100, and 250000 energy evaluations. Based on energy and cluster populations, promising h-CD22/ and m-CD22/ligand complexes were identified and further subjected to MD simulations.

CORCEMA-ST. CORCEMA-ST protocol was used as previously described (Jayalakshmi and Krishna, 2000). The pdb coordinates of complexes were selected from the MD trajectory analyses. The conformation of the ligand was assumed to be invariant in free and bound state. The input variables, as the concentration of the ligand and the protein, were experimentally derived. The saturation time was set to 2s and the dissociation constants (K_D) were set on the basis on the experimentally derived for h-CD22/ Neu5Ac-α-(2-6)-Gal-β-(1-4)-GlcNAc complex[23] and further adjusted to get the best fit. A binding site cutoff of 8 Å was employed. By

computing the R matrix and the calculation of spectral densities, the fractional intensity changes were calculated for each ligand protons and compared to the experimental STD effects by means of a NOE R factor, a normalized root-means square deviation value. For the calculations, only the STD values of the ligands isolated signals were considered. Figures of the selected complexes were done using Pymol 2.4.0 (Schrödinger, LLC, 2000).

Supplemental references

Ribeiro, M. M. B.; Franquelim, H. G.; Castanho, M. A. R. B.; Veiga, A. S. (2008) Molecular Interaction Studies of Peptides Using Steady-State Fluorescence Intensity. Static (De)Quenching Revisited *J. Pept. Sci*. 14 (4), 401– 406.

Fiser, A., Do R.K., Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci*. 9,1753-73.

Schrodinger. 2012. Epik version 2.3, Impact version5.8, Prime version 3.1, Schrodinger, LLC, New York, NY.Schrödinger Release 2020-2**:** MacroModel, Schrödinger, LLC, New York, NY, 2020.

Woods Group. (2005-2020) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia,Athens, GA. (http://glycam.org).

The PyMOL Molecular Graphics System, Version 2.4.0 Schrödinger, LLC.