

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 $< 110^\circ$. The experimental error in the calculated proton-proton distances is estimated to be less than $\pm 10\%$.

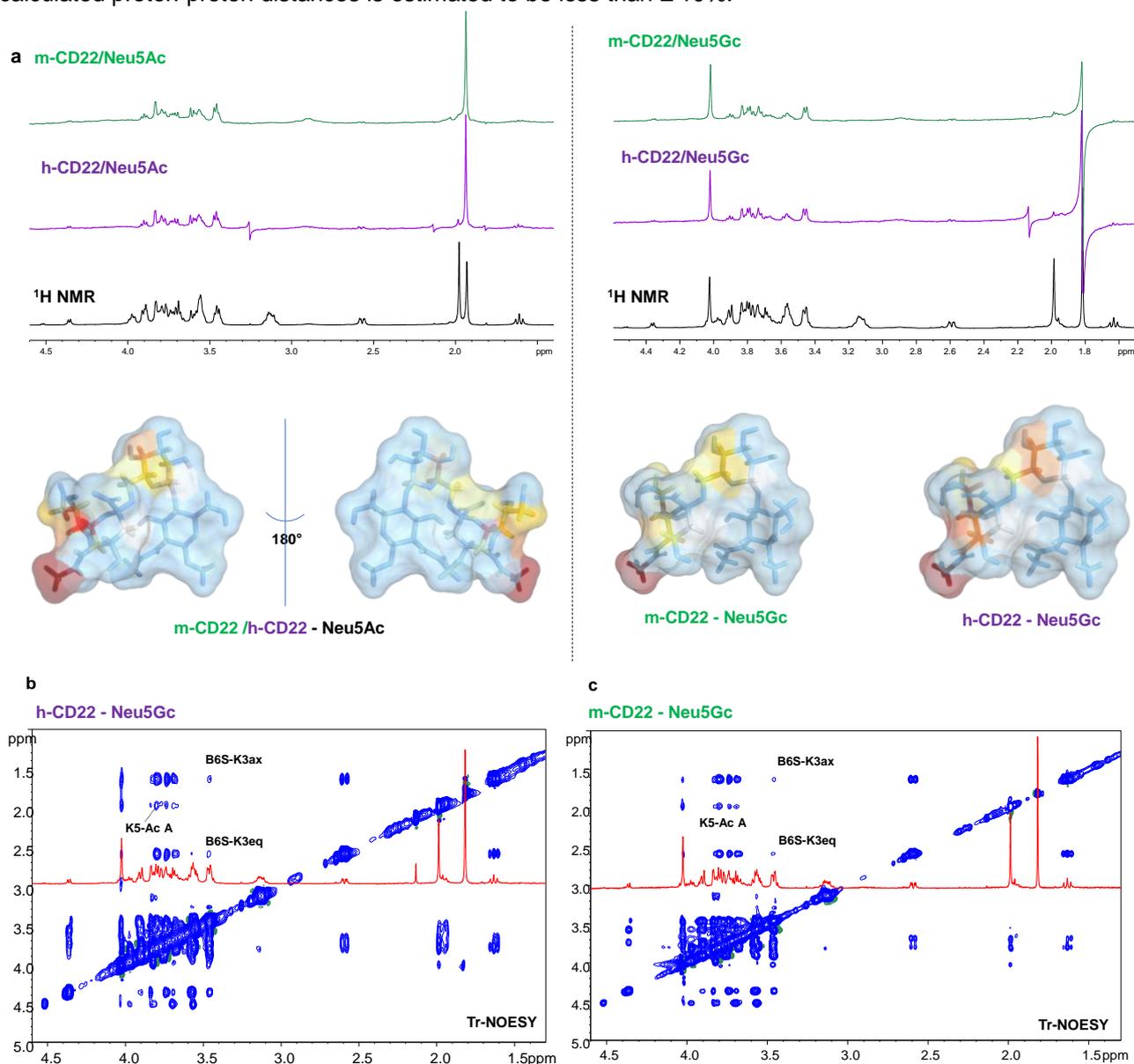


Figure S2. [MD simulation analysis of the glycolylated trisaccharide in its free state], related to Figure 2 and Table S3

- a) $\phi/\psi/\omega$ dihedral angles of Neu5Gc-Gal linkage along the MD trajectory.
- b) $\phi/\psi/$ 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).

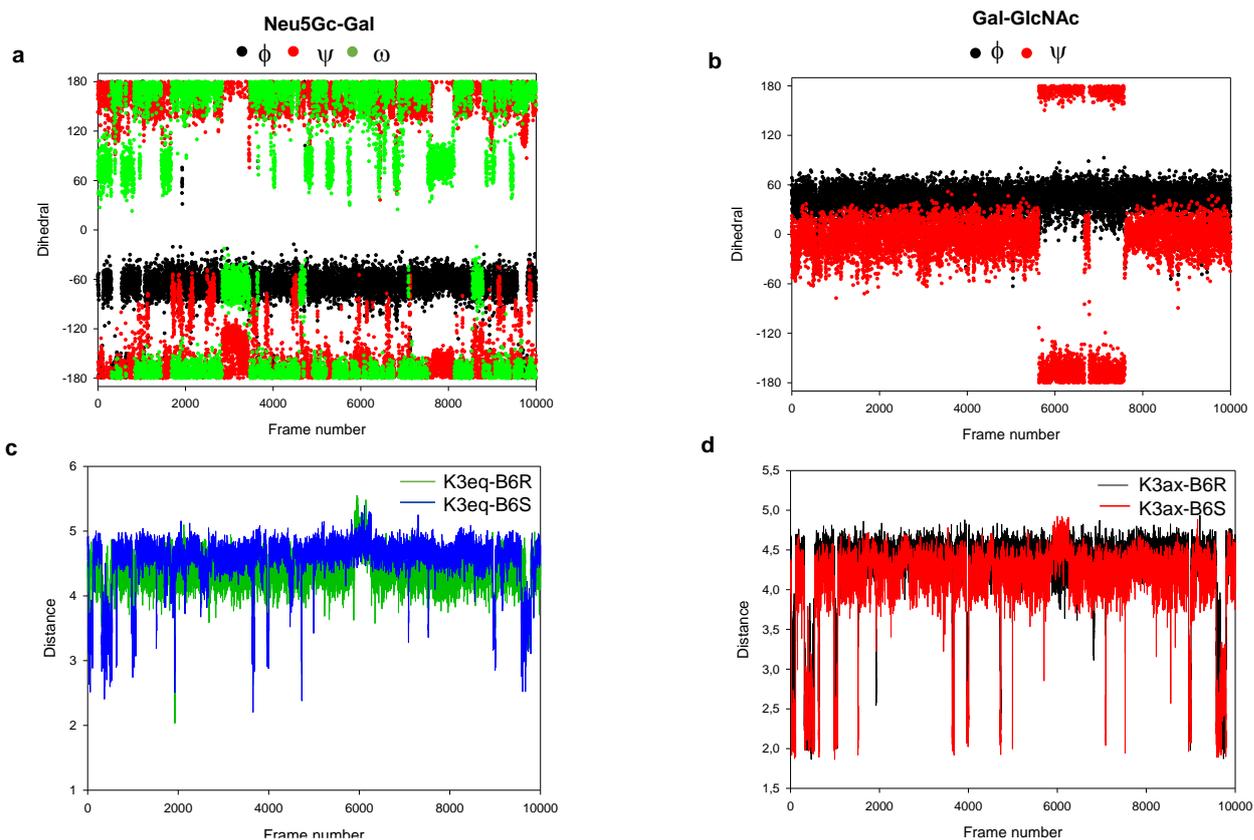


Figure S3. [MD simulation analysis of m-CD22 homology modelling], related to Figure 3.

- Superimposition of the m-CD22 structures each 10 ns of the MD simulations. Along the MD simulation, no relevant conformational changes emerged.
- 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.
- Atomic fluctuation by residue of m-CD22 structure, calculated using the protein C α atoms. The peaks in the RMSF plot corresponded to the mobile loops connecting the β -strands, in both V-set and C2-set Ig-like domains.
- 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.

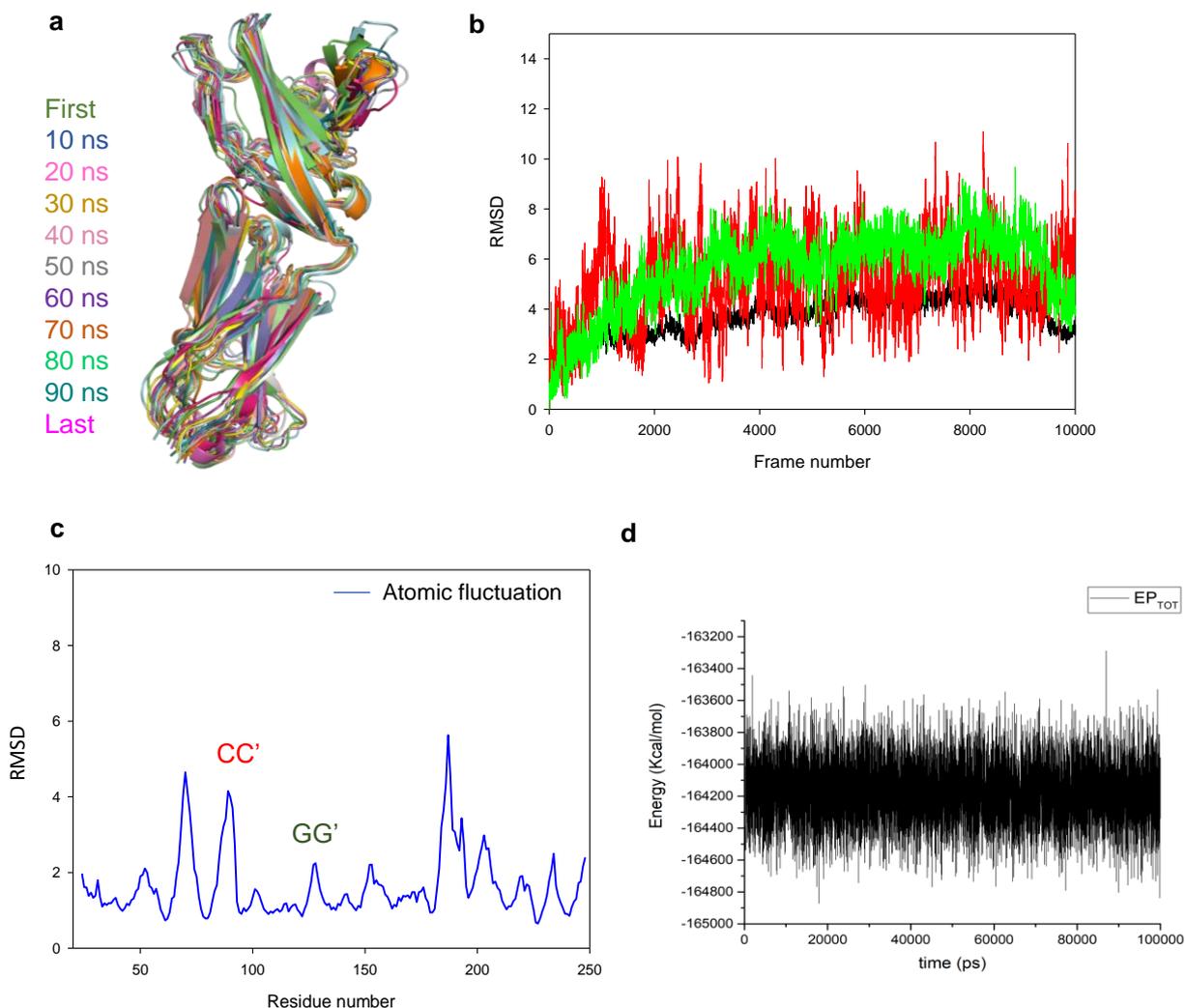


Figure S4. [MD simulation analysis of h-CD22/m-CD22 complexes with the glycolylated ligand], related to Figures 4 and 5.

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

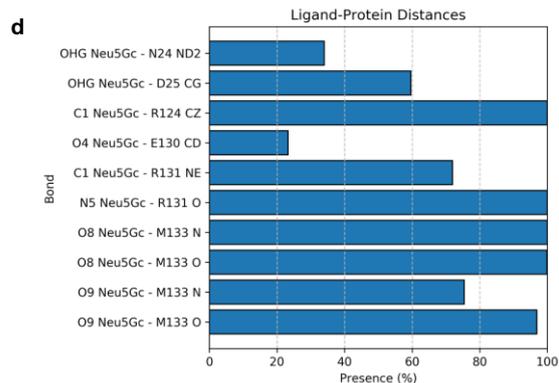
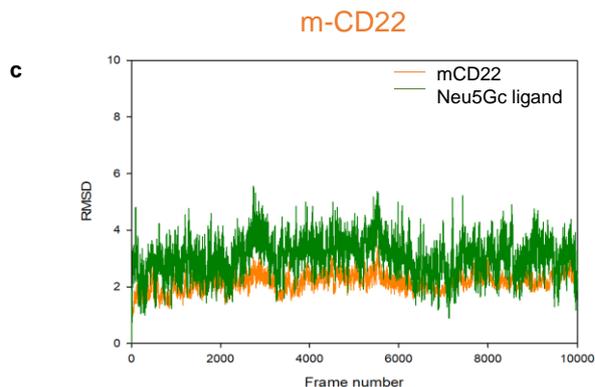
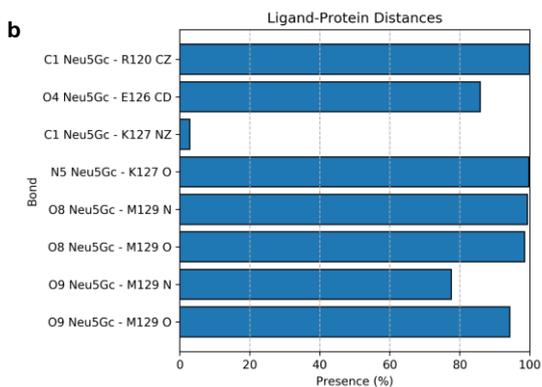
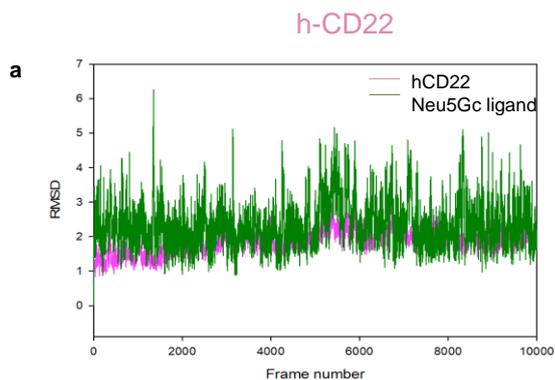
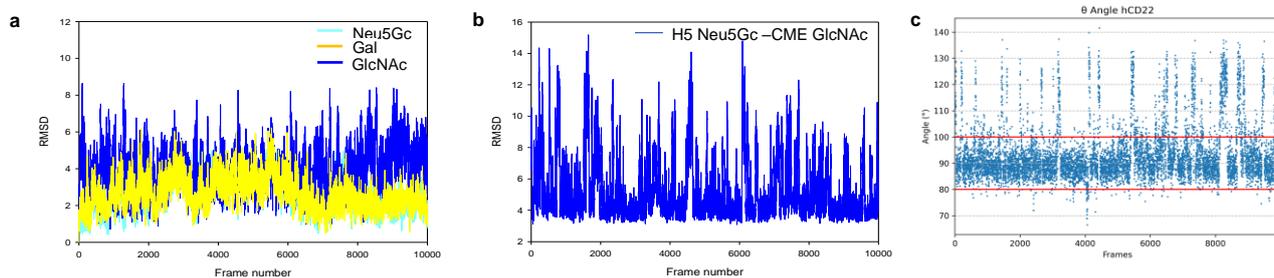


Figure S5. [Analyses of the glycolylated ligand conformation when bound to h-CD22 and m-CD22], related to Figures 4 and 5.

- RMSD of Neu5Gc ligand residues with respect to h-CD22 protein.
- Distance between H5 of Neu5Gc and GlcNAc acetyl group C (CME) (average value 5.1 Å).
- 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.
- RMSD of Neu5Gc ligand residues with respect to the m-CD22 protein.
- Distance between H5 of Neu5Gc and GlcNAc acetyl group C (CME) (average value 4.9 Å).
- Variation of Neu5Gc θ angle value along m-CD22/Neu5Gc complex simulation.

h-CD22



m-CD22

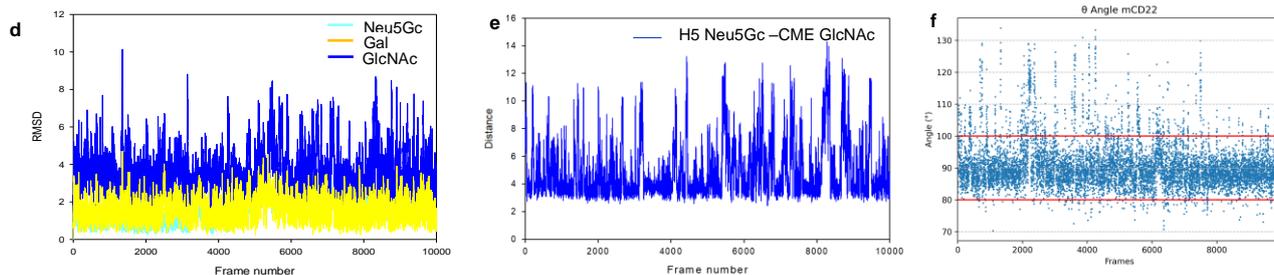
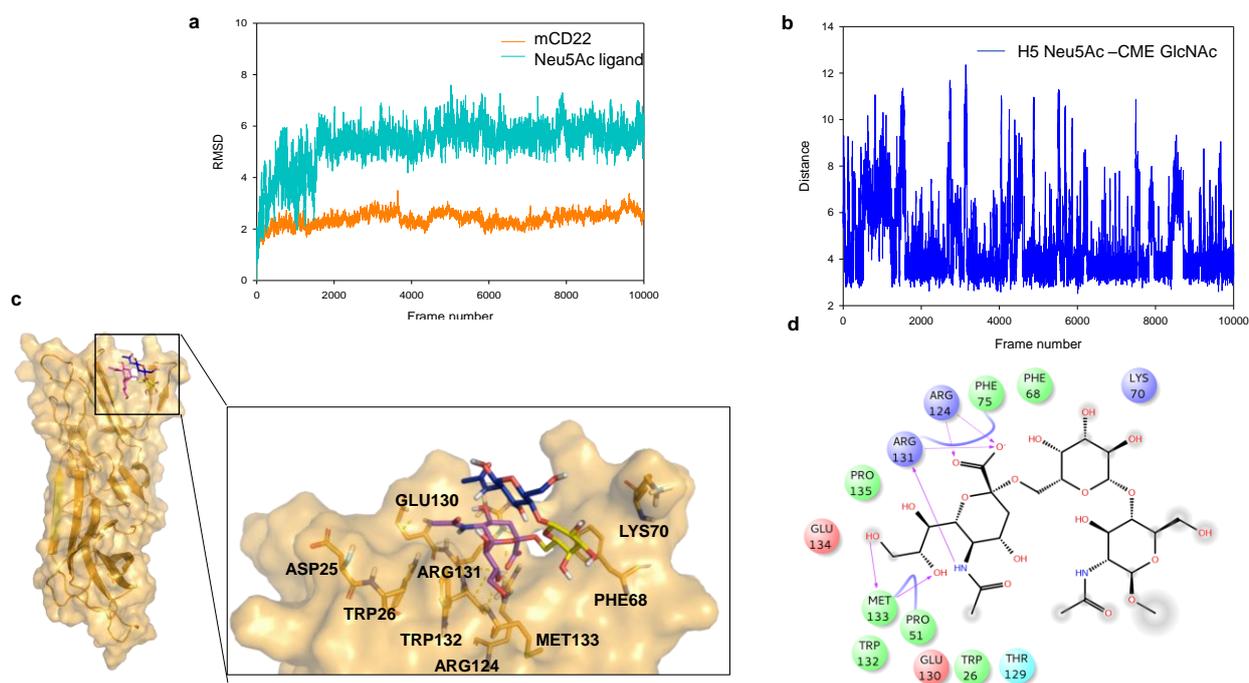


Figure S6. [Interaction between m-CD22 and Neu5Ac ligand], related to Figure 6

- Protein and ligand RMSD variation along the MD. The ligand RMSD was measured with respect to the protein.
- Distance between H5 of Neu5Gc and GlcNAc acetyl group C (CME) (average value 4.6 Å)
- 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.
- Two-dimensional plots representing the interactions between the glycosylated 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 $\Phi = -60^\circ$ $\Psi = 180^\circ$ $\omega = 60^\circ$	Family II $\Phi = 180^\circ$ $\Psi = 180^\circ$ $\omega = 60^\circ$	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)
CH₃ 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):

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

$a = [F]_t K_b$, $b = 1 + [Y]_t K_b$, $c = [Y]_t K_b$ 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 ^1H - ^1H ROESY and ^1H - ^1H 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 $(I_0 - I_{\text{sat}})/I_0$, 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: $STD = STD^{\max}(1 - e^{-k_{\text{sat}}t_{\text{sat}}})$, 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 $STD_{\text{epitopes fit}}$.

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 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 complex²³ 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.