Biochemistry

pubs.acs.org/biochemistry Article

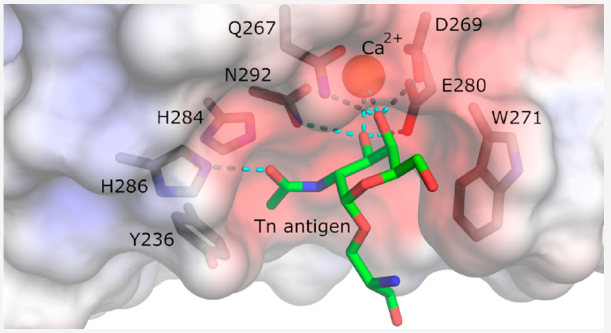
Crystal Structure of the Carbohydrate Recognition Domain of the Human Macrophage Galactose C-Type Lectin Bound to GalNAc and the Tumor-Associated Tn Antigen

Adele Gabba, Agnieszka Bogucka, John G. Luz, Ana Diniz, Helena Coelho, Francisco Corzana, Francisco Javier Cañada, Filipa Marcelo, Paul V. Murphy, and Gabriel Birrane*

Cite This: *Biochemistry* 2021, 60, 1327−1336 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The human macrophage galactose lectin (MGL) is an endocytic type II transmembrane receptor expressed on immature monocyte-derived dendritic cells and activated macrophages and plays a role in modulating the immune system in response to infections and cancer. MGL contains an extracellular calcium-dependent (C-type) carbohydrate recognition domain (CRD) that specifically binds terminal *N*-acetylgalactosamine glycan residues such as the Tn and sialyl-Tn antigens found on tumor cells, as well as other *N*- and *O*-glycans displayed on certain viruses and parasites. Even though the glycan specificity of MGL is known and several binding glycoproteins have been identified, the molecular basis for substrate recognition has remained elusive due to the lack of high-resolution structures. Here we present crystal structures of the MGL CRD at near endosomal pH and in several complexes, which reveal details of the interactions with the natural ligand, GalNAc, the cancer-associated Tn-Ser antigen, and a synthetic GalNAc mimetic ligand. Like the asialoglycoprotein receptor, additional calcium atoms are present and contribute to stabilization of the MGL CRD fold. The structure provides the molecular basis for preferential binding of *N*-acetylgalactosamine over galactose and prompted the re-evaluation of the binding modes previously proposed in solution. Saturation transfer difference nuclear magnetic resonance data acquired using the MGL CRD and interpreted using the crystal structure indicate a single binding mode for GalNAc in solution. Models of MGL1 and MGL2, the mouse homologues of MGL, explain how these proteins might recognize Lewis^X and GalNAc, respectively.



The human macrophage galactose Ca^{2+}-dependent (C-type) lectin (MGL, also known as CLEC10A, CD301, or DC-ASGPR) is the only known receptor on antigen-presenting cells that shows a strong binding preference for O-linked *N*-acetylgalactosamine (*α*- and *β*-GalNAc),[1,2] an important component of the tumor cell antigens, Tn (*α*GalNAc-Ser/Thr) and the sialylated Tn antigen [Neu5NAc*α*(2→6)-GalNAc*α*1→O-Ser/Thr], as well as the LacdiNAc (GalNAc*β*1−4GlcNAc) epitope present on the surface of many pathogens. MGL can therefore discriminate tumor tissue from healthy tissue and detect infections. Interactions between MGL and tumor-associated glycans orchestrate distinct immune responses. Recognition of Tn structures present on the tumor-associated MUC1 glycoprotein and the self-glycoproteins CD43 and CD45 by MGL induces suppressive antitumor immune responses characterized by increased IL-10 production and induction of regulatory and effector T cells.[3−5] However, MGL-expressing cells are also able to take up Tn-derived peptide structures for antigen presentation and induction of T cell responses, thus eliciting antitumor immunity.[1,6,7] The MAG-Tn3 vaccine that contains three Tn moieties linked to a $CD4^+$ T cell epitope exploits this capability to induce robust anti-Tn antibody responses.[8,9]

MGL also interacts with the LacdiNAc epitope[3] present in bacteria such as *Schistosoma mansoni*,[10] *Neisseria gonorrheae*,[11] *Campylobacter jejuni*, and *Trichuris suis*,[11] and binding of MGL to the outer core of *Escherichia coli* strain R1 lipooligosaccharide has been demonstrated by saturation transfer difference nuclear magnetic resonance (STD-NMR) spectroscopy.[12] It has also been proposed that MGL and other C-type lectins, including hepatic asialoglycoprotein receptor (ASGPR), dendritic cell-specific intercellular adhesion molecule-3 grabbing non-integrin (DC-SIGN), and liver/lymph node-specific ICAM-3 grabbing non-integrin (L-SIGN), act as attachment

Received: January 4, 2021
Revised: February 23, 2021
Published: March 16, 2021

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

points to promote filovirus (e.g., Ebola and Zaire) and possibly other infections.[13] Indeed, it was recently demonstrated that MGL is also able to recognize the glycan structures present on the receptor binding domain (RBD) of the SARS-CoV-2 spike protein.[14,15] MGL also appears to function as a negative regulator of autoimmune-driven neuroinflammation.[16]

Structurally, MGL is a type II transmembrane protein composed of a short N-terminal cytoplasmic domain containing a YxxΦ endocytosis motif (where Y is a tyrosine, x represents any amino acid, and Φ is a hydrophobic amino acid), a transmembrane domain, and an extracellular domain (ECD). The ECD contains a coiled-coil neck stabilizing a homotrimeric organization and a C-terminal calcium-dependent (C-type) carbohydrate recognition domain (CRD). The MGL CRD contains the characteristic glutamine-proline-aspartic acid (QPD) motif in a long loop region and the tryptophan-asparagine-aspartic acid (WND) motif in strand β4 that are responsible for Ca^{2+} coordination and substrate recognition.[17] The availability and orientation of the equatorial/axial OH-3 and OH-4 groups of Gal/GalNAc residues are crucial for Ca^{2+} binding, and additional contacts established with the 2-acetamido group (NHAc) determine the MGL binding preference for GalNAc over Gal.[18]

Despite the importance of MGL in infection and cancer immunology, and its relevance as a potential target for the development of therapeutics, vaccines, and biomarkers, progress in this regard has been hindered due to the lack of high-resolution crystal structures to guide drug design efforts. Some of us previously employed STD-NMR spectroscopy to investigate the recognition of GalNAc by MGL. STD-NMR experiments recorded using MGL ECD in combination with molecular dynamics and CORCEMA-st, using a homology model derived from the crystal structure of ASGPR CRD, proposed the coexistence of two binding modes for GalNAc (A and B), with mode B seemingly preferred (Figure S1).[18] Mode A is characterized by the orientation of galactose or galactosamine binding in the crystal structures of C-type lectin CRDs that contain the signature WND and QPD motifs such as CEL-I from the sea cucumber *Cucumaria echinata*[19] and the mouse scavenger receptor.[20] In these structures, CH−π interactions, recurrently observed in many carbohydrate receptors,[21] are also observed between the side chain of a Trp residue following the QPD motif and H6 and H5 of the bound Gal or GalNAc residues. In binding mode B, the carbohydrate ring is flipped around an axis that swaps the positions of the 3- and 4-hydroxy groups coordinating the bound Ca^{2+} atom (Figure S1B). Mode B is similar to that observed in C-type CRDs that lack the signature QPD motif, for example, in the crystal structures of tunicate *Polyandrocarpa misakiensis* lectin bound to galactose,[22] and in human langerin in complex with galactose 6-sulfate.[23] In the tunicate lectin, a Trp residue is present on the opposite side of the carbohydrate binding pocket and makes CH−π interactions with CH6 when the pyranose ring is flipped. Binding mode B is in direct contrast to the mode of carbohydrate recognition by ASGPR, the closest relative of MGL for which structural information is available and which binds carbohydrates in binding mode A. More recently, the $^{1}H-^{15}N$ HSQC-based titration of the MGL CRD revealed that GalNAc binds in a slow exchange regime (small k_{off}) on the NMR chemical shift time scale, indicating long residence times for GalNAc in the MGL CRD binding site.[24] Nevertheless, the chemical shift perturbation analysis of the $^{1}H-^{15}N$ HSQC of MGL CRD upon addition of GalNAc was not sufficient to clarify the major binding mode. To obtain a more accurate picture of the carbohydrate binding mode in solution, we set out to obtain the relevant crystal structures of the MGL CRD and re-examine the previous binding modes proposed by STD-NMR analysis using the MGL ECD. Here we describe the development of an expression, refolding, and purification protocol to produce MGL CRD in *E. coli* cells that led to the recovery of the protein in high yield. This enabled the determination of appropriate crystallization conditions for the MGL CRD and several complexes with physiologic ligands, including GalNAc and the Tn-Ser antigen, and a synthetic glycomimetic compound. The crystal structures identified key protein−carbohydrate interactions that explain the MGL CRD selectivity for GalNAc over Gal. New STD-NMR data for GalNAc were recorded, this time in the presence of the MGL CRD. Using the MGL crystal structure as a model and taking into consideration the slow exchange binding mechanism deduced from $^{1}H-^{15}N$ HSQC titrations, a contact epitope map deduced from the theoretically calculated saturation transfer showed that the binding orientation in solution corresponds to mode A, in agreement with that observed in the crystal structures.

Finally, homology models of the mouse MGL1 and MGL2 CRDs derived from the crystal structure of MGL might explain how these molecules recognize $Lewis^{X}$ and GalNAc, respectively.

MATERIALS AND METHODS

Protein Expression and Purification. A cDNA fragment corresponding to the MGL CRD (amino acids 181−277) was codon-optimized for expression in *E. coli*, synthesized (Integrated DNA Technologies), and cloned into a modified pTRIEX vector (Novagen) at the *Bam*HI and *Eco*RI restriction sites. The expressed protein contained a six-His tag followed by a tobacco etch virus (TEV) protease recognition sequence at the N-terminus. Three vector-derived amino acids (Gly, Thr, and Ser) remain after TEV protease removal of the six-His tag. The protein was produced as inclusion bodies in OverExpress C41(DE3) *E. coli* cells (Lucigen) by induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6 for 4 h at 37 °C. Bacterial pellets were harvested by centrifugation, washed with phosphate-buffered saline (Corning), and purified with alternating washing and sonication steps in buffer A [20 mM tris(hydroxymethyl)aminomethane (Tris) (pH 8.0), 0.5% Triton X-100, 5 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0), and 25 mM β-mercaptoethanol (β-ME)] followed by buffer B [buffer A supplemented with 1% Triton X-100 and 500 mM sodium chloride (NaCl)] followed by buffer C (buffer A supplemented with 1.0 M NaCl). Purified inclusion bodies were harvested by centrifugation at 12000*g* for 30 min at 4 °C, denatured, and solubilized by resuspending and stirring pellets from 2 L of the original cell culture in 25 mL of ammonium hydroxide at 4 °C for 2 h. Differential scanning fluorimetry (DSF) was used to determine the optimal conditions for refolding of MGL, based on the guided protein refolding approach developed by Groves and co-workers.[25] Optimal refolding was found to occur by dialyzing the denatured and solubilized inclusion bodies against a buffer containing 20 mM Tris (pH 7.5), 500 mM NaCl, 400 mM L-arginine, 25 mM calcium chloride ($CaCl_2$), 1 mM reduced glutathione, and 0.1 mM oxidized glutathione. A Tycho NT.6 (Nanotemper Technologies) was used to monitor the refolding experiment in real time. Refolding was judged to

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

be complete after 4−6 h when no further changes were seen in the thermal unfolding profile of the protein. The MGL CRD was further purified by affinity chromatography using cOmplete Ni-NTA resin (Roche), and the six-His tag was removed by overnight incubation at 4 °C with TEV protease (produced in house). The protein was applied to a HiPrep Q1 FF anion ion exchange column (GE Healthcare) equilibrated with 20 mM Tris (pH 7.5), 40 mM galactose, and 2 mM $CaCl_2$. The MGL CRD was eluted with a salt gradient from 0 to 1.0 M NaCl, concentrated to 1 mL in a 3 kDa molecular weight cutoff centrifugal concentrator (Millipore), and applied to a Superdex S75 size-exclusion column (GE Healthcare) pre-equilibrated with a running buffer composed of 20 mM Tris (pH 7.5), 2 mM $CaCl_2$, and 150 mM NaCl and containing 40 mM galactose to prevent interaction of the protein with the carbohydrate matrix of the column. Fractions containing MGL CRD were pooled, concentrated, and applied to a polyacrylamide desalting column (Pierce) equilibrated with 20 mM Tris (pH 7.5), 2 mM $CaCl_2$, and 150 mM NaCl to remove the galactose. This procedure routinely yielded at least 10 mg of >98% pure MGL protein per liter of original cell culture as judged by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (data not shown). The protein was concentrated to 70 mg/mL and flash-frozen in liquid nitrogen prior to crystallization screening.

Crystallization. Crystallization trials were set using a Mosquito nanoliter pipetting robot (SPT Labtech) using JCSG crystallization screening kits (Qiagen). Initial conditions were optimized to produce single crystals that were cryo-protected by being briefly dipped into a crystallization solution supplemented with increasing amounts of ethylene glycol. The MGL CRD at a concentration of 23 mg/mL was crystallized in 20% (w/v) polyethylene glycol (PEG) 3000 and 100 mM sodium citrate tribasic/hydrochloric acid (pH 5.5) and cryo-protected in crystallization mother liquor supplemented with 20% ethylene glycol. Complexes were prepared by mixing Tn-Ser antigen (Biosynth Carbosynth), GalNAc (Fisher Scientific), and methyl 2-(acetylamino)-2-deoxy-1-thio-α-D-galactopyranose (GalNAcα-1-SMe)[26] in a 10-fold molar excess with the MGL CRD and allowed to stand at 4 °C for 2 h. All complexes were crystallized under nearly identical conditions including 100 mM Tris (pH 6−7), 200 mM magnesium chloride, and 2−2.8 M sodium chloride and cryo-protected in crystallization buffer supplemented with 15% ethylene glycol.

Data Collection and Processing. The unbound MGL CRD produced rod cluster crystals in the $P2_12_12_1$ space group, whereas those with ligands produced needlelike crystals in the $P3_12$ space group. All diffraction data were recorded at the European Synchrotron Radiation Facility, Grenoble beamline ID-30B, on a Pilatus3 6M (Dectris) photon counting detector using X-rays at a wavelength of 0.9763 Å. Data processing and scaling were performed with XDS,[27] and the structures were determined by molecular replacement using PHASER[28] from the CCP4 suite.[29] The coordinates of the CRD of the H1 subunit of the human hepatic asialoglycoprotein receptor (ASPGR) lectin[30] [Protein Data Bank[31] (PDB) entry 1DV8] were used as the search model. Ligands were built into electron density maps using COOT[32] and refined in REFMAC5,[33] and the final refined models were validated using MolProbity.[34] Data collection and refinement statistics for the structures reported here are listed in Table S1.

^{1}H−^{15}N HSQC Titrations and Saturation Transfer Difference NMR Spectroscopy. All NMR analysis used α-Me-GalNAc (Biosynth Carbosynth) as the ligand (hereafter termed Me-GalNAc). Experiments were performed on a Bruker Avance 600 MHz spectrometer equipped with a triple-channel cryoprobe. MGL CRD (amino acids C181−H316) and ECD (amino acids Q61−H316) proteins were expressed and purified as described previously[24] and lyophilized before use. STD-NMR experiments (Bruker pulse sequence stddiffesgp) were performed at 310 K with 1.1 mM Me-GalNAc and 30 μM MGL CRD or ECD in 10 mM perdeuterated Tris (D11)-DCl (Sigma-Aldrich) (pD 7.5) in D_2O containing 20 mM $CaCl_2$, 75 mM NaCl, and 0.09% sodium azide. Spectra were recorded with 1728 scans in a matrix with 64K data. An excitation sculpting module with gradients was used to suppress the water proton signals. Selective saturation of the protein resonances (on-resonance spectrum) was performed by irradiating at 7 ppm for aromatic residues and −0.5 ppm for aliphatic residues using a series of 40 Eburp2.1000-shaped 90° pulses of 50 ms for a total saturation time of 2 s and a relaxation delay of 4 s (D1 on the stddiffesgp Bruker sequence). For the reference spectrum (off-resonance), the samples were irradiated at 100 ppm. The same experimental conditions (Bruker pulse sequence stddiff, irradiating at −0.5 ppm for aliphatic protons with a 20-fold molar excess of Me-GalNAc) were used to reproduce the STD-NMR experiment reported using the MGL ECD[18] (Figure S2). A control experiment was performed for Me-GalNAc in the absence of MGL. STD spectra were obtained by subtracting the on-resonance spectrum from the off-resonance spectrum. Residual STD intensities for Me-GalNAc in the absence of the protein were subtracted from those of the complex. Finally, the relative percentages of spin saturation were calculated by setting the STD signal of the proton with the highest intensity to 100% and calculating the other STD signals accordingly.

The ^{15}N-labeled MGL CRD was expressed and purified as described previously.[24] The ^{1}H−^{15}N HSQC titration experiment was performed at 310 K (the same temperature as the STD experiment) with 200 μM MGL CRD in a buffer containing 10 mM Tris (pH 7.5), 75 mM NaCl, 20 mM $CaCl_2$, and 10% D_2O. Me-GalNAc was added to achieve protein:ligand molar ratios of 1:0.1, 1:0.25, 1:0.5, 1:0.75, 1:1, 1:2, and 1:10. Data were acquired with 2048 × 128 points and 32 scans in a spectral window of 9615.4 Hz (centered at 2801 Hz) × 1946 Hz (centered at 7175 Hz), in ^{1}H and ^{15}N sweep's width, respectively. The data were processed in TopSpin 4.0.6 (Bruker) and CcpNMR Analysis version 2.4.[35] Superposition of ^{1}H−^{15}N HSQC spectra for protein:ligand ratios of 1:0, 1:0.5, 1:0.75, and 1:10 is shown in Figure S3.

CORCEMA-st Calculations. Matlab scripts in CORCEMA-st[36,37] were used to calculate proton ligand saturation transfer for the MGL CRD in complex with Me-GalNAc using the crystal structure as a model. Conformer A was used for all amino acid side chains that were modeled in multiple conformations in the crystal structure. Because the experimental values were obtained with Me-GalNAc and the crystal structure contained GalNAc, we replaced the hydroxyl in the model with a methoxy group. To accelerate the computation, only amino acids within a radius of 10 Å around the ligand were included in the calculation. Because the MGL CRD is rich in aromatic residues, changing the irradiation frequency from −0.5 to 7 ppm led to an increase in absolute STD enhancements in the experimentally acquired STD spectra and

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

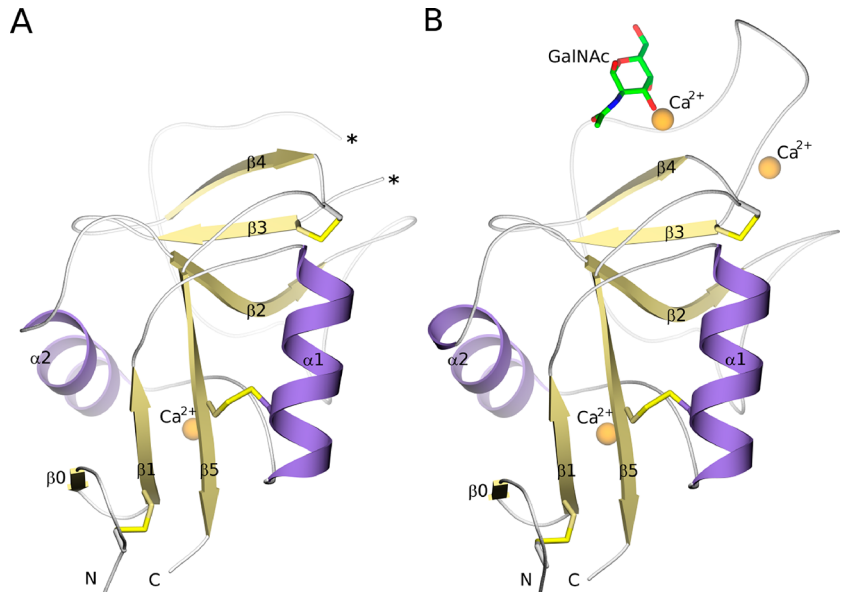


Figure 1. Crystal structure of the MGL CRD. (A) Secondary structure of the apo MGL CRD. α-Helices are depicted as purple coils, and β-strands as khaki arrows. Disulfide bonds are shown as yellow sticks. Calcium atoms are depicted as orange spheres. The positions of the N- and C-termini are indicated. A section of the loop region after the QPD motif from Asp270 to Asp281 is flexible (indicated by asterisks) and could not be modeled. (B) Secondary structure of the MGL CRD with bound GalNAc. Secondary structure elements are colored as in panel A. Carbon atoms of the bound GalNAc molecule are shown as green sticks. The loop region from Asp270 to Asp281 is ordered in the bound form, and two additional calcium atoms not present in the apo structure are seen. This figure was generated using Povscript+ and ray-traced with POVRAY (http://www.povray.com).

CORCEMA-st simulations (Figure S4). For this reason, STD spectra acquired at a saturation frequency at 7 ppm were selected for calculations and comparison with CORCEMA-st simulations. The other parameters used in the calculations were a 2 s saturation time, an estimated K_d of 14.5 μM at 310 K, and experimental concentrations of 1.1 mM ligand and 30 μM protein. Correlation times of 0.28 ns for free Me-GalNAc, 7.58 ns for Me-GalNAc bound to the MGL CRD, and 45.69 ns for Me-GalNAc bound to the MGL ECD were calculated with an online tool (http://nickanthis.com/tools/tau) using experimental values determined at 293 K in H_2O for the MGL CRD in complex with Me-GalNAc (9.6 ns) (unpublished data) and considering a molecular weight of 84 kDa corresponding to the trimeric form of the MGL ECD. In contrast to the previous work in which a diffusion-controlled kinetic model was assumed ($k_{on} = 1 \times 10^8$ M^{-1} s^{-1}),[18] we used lower values for k_{on} (2×10^5 M^{-1} s^{-1} for the MGL CRD and 1×10^6 M^{-1} s^{-1} for the MGL ECD) compatible with the slow exchange regime observed in $^1H-^{15}N$ HSQC titrations.[24] Decreasing the k_{on} led to a significant decrease in the percentage of spin saturation for Me-GalNAc H3 and H4 protons and a slight increase for both H6 protons as well as the acetyl CH_3 protons (Figure S5). Adjusting the k_{on} to a value compatible with a slow exchange regime observed in $^1H-^{15}N$ HSQC-based titrations was required to achieve a good match between the experimental and theoretical STD values (Figure S6).

Molecular Modeling. Molecular dynamics simulations were carried out with Amber16[38] using the ff14SB[39] and GLYCAM06j-1[40] force fields for the MGL CRD complex with GalNAc in binding modes A and B as described previously,[24] and the crystal structure of the MGL CRD in complex with GalNAc. The protocol was that described previously,[24] except for a 500 ns production step instead of 100 ns. A detailed analysis of each MD trajectory was carried out (Figures S7 and S8).

Homology models for the mouse MGL1 and MGL2 were generated by mutating amino acids in MGL to the corresponding MGL1 and MGL2 amino acids in COOT and optimizing the side chain orientations to prevent clashes and maximize hydrogen bonding and hydrophobic interactions. Calcium, water molecules, and GalNAc were added to the MGL2 model by comparison with the crystal structure of MGL in complex with GalNAc. The Lewis^X glycan was manually docked into the MGL1 model by superimposing the crystal structure of the mouse scavenger receptor C-type lectin in complex with Lewis^X and copying the coordinates of the trisaccharide into the new model. The resulting models were energy minimized using Chiron,[41] and the stereochemistry was assessed using MolProbity.[34] All amino acids for both models fell in the allowed or additionally allowed regions of the Ramachandran plot.

RESULTS

The MGL CRD (amino acids 181−277) bearing an N-terminal six-His tag followed by a TEV protease recognition site was expressed as inclusion bodies in *E. coli* cells. Inclusion bodies were washed, denatured with ammonium hydroxide, and refolded by dialysis into a redox buffer containing reduced and oxidized glutathione to promote correct disulfide bond formation, arginine to suppress aggregation, and calcium chloride required for stabilization of the CRD fold. The correctly folded protein was further purified by immobilized metal affinity chromatography (IMAC) on Ni-NTA resin, and the six-His tag was removed with TEV protease. Subsequent Q ion exchange and size-exclusion chromatography steps were carried out in the presence of galactose to further purify the protein and remove aggregates, respectively. The protein was concentrated and passed through a desalting column to

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

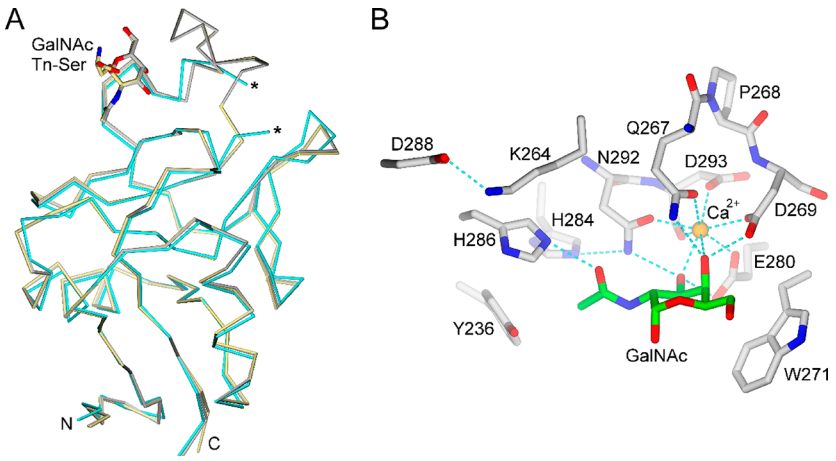


Figure 2. (A) Cα superposition of the structures of the MGL CRD unbound (cyan) and in complex GalNAc (gray) and the Tn-Ser antigen (khaki). Asterisks indicate the location of the disordered loop in the apo MGL CRD structure. The bound GalNAc and Tn-Ser antigen molecules are shown as stick models. (B) Stick representation of the carbohydrate binding site of MGL (gray carbons) showing the bound GalNAc (green carbons). Metal-coordinating and hydrogen bonds are shown as cyan dashed lines. Oxygen and nitrogen atoms are colored red and blue, respectively. The bound calcium is shown as an orange sphere.

remove galactose. The unbound MGL crystallized under a range of conditions, all of which contained citrate or tartrate buffers at acidic pH. X-ray data were collected and processed in space group $P2_12_12_1$ to a high-resolution limit of 1.2 Å. The structure was determined by molecular replacement with the coordinates of the CRD of ASGPR (PDB entry 1DV8)[30] as the search model. The refined model had R_{cryst} and R_{free} values of 15.5% and 18.9%, respectively (Table S1). The purified MGL CRD was mixed with a 10-fold stoichiometric excess of GalNAc, GalNAcα-1-SMe, and Tn-Ser and crystallized in space group $P3_12$. Data sets were collected to high-resolution limits of 1.7, 2.3, and 2.0 Å for the GalNAc, GalNAcα-1-SMe, and Tn-Ser complexes and refined to R_{cryst} and R_{free} values of 16.5% and 18.8%, 16.2% and 20.2%, and 18.3% and 21.1%, respectively (Table S1). For atoms with representative electron density, there are no significant differences observed in the positions of main chain atoms when comparing the apo and halo forms (Figure 1). The root-mean-square deviations for main chain atoms when compared to the apo structure are 0.86, 0.87, and 0.98 Å for the GalNAc-, GalNAcα-1-SMe-, and Tn-Ser-bound forms, respectively (Figure 2A).

The overall structure of the MGL CRD was found to be similar to that observed in previously determined C-type lectin CRD structures[42] with five β-strands (β1−β5) arranged in two β-sheets with a bent β2 strand shared between both sheets. The core sheet structure is flanked on opposite sides by two orthogonally oriented α-helices (Figure 1). As predicted by UniProt (accession number Q8IUN9), there are three intrachain disulfides between cysteine residues 181 and 192, 209 and 304, and 282 and 296 (Figure 1). MGL is a long-form CRD that typically contains an extra β-strand (β0) and a disulfide bond (from C181 to C192) connecting the N-terminus to β1. Three calcium ions are consistently observed in the ligand-bound structures (Figure 1B).[43] A structurally important calcium ion termed calcium site 3 present in the apo and halo forms (Figure 1) is located close to the N- and C-termini of the MGL CRD. Calcium site 3 is coordinated by the side chain oxygen atoms of Asn220, Glu224, and Glu305, and the carbonyl of Val218 and links α2 with β5 and the α1−α2 loop. Two water molecules complete the coordination sphere. Site 3 is also present in the ASGPR CRD[30] and the rattlesnake venom lectin.[44] In other C-type CRD structures that do not contain a calcium atom at this position, a similar structural role is performed by a network of salt bridges and hydrogen bonds in the case of the rat liver mannose binding protein A[45] and the sea cucumber CEL-I lectin,[19] or by hydrophobic interactions in the case of the E-selectin CRD.[46] Under low-pH conditions and in the absence of a ligand, a portion of the glycine rich long loop after the QPD motif from Asp270 to Asp281 is flexible and could not be modeled due to the absence of electron density (Figure 1A). This is consistent with previous NMR studies that showed that amide resonances for distinct amino acids in this same region of the long loop could not be detected in the absence of GalNAc.[24] In the crystal structures reported here of MGL in complex with GalNAc, the Tn antigen, and the synthetic compound GalNAcα-1-SMe, the long loop is ordered and binds the two additional calcium atoms (Figure 1B). One of these is termed calcium site 2 and plays a structural role by coordinating the side chains of Asp243, Asp270, and Asp281 and the main chain carbonyl of Glu280 in the long loop. Three water molecules complete the coordination sphere of site 2 (Figure S9). Finally, and a signature of C-type lectins, a third calcium site termed site 1 is located 8.4 Å from site 2 on the other side of the long loop. Site 1 coordinates the side chains of Gln267 and Asp269 in the QPD motif, Asn292 and Asp293 in the WND motif, Glu280 in the long loop, and the main chain carbonyl of Asp293 as well as the 3- and 4-hydroxyls of the bound GalNAc (Figure 2B). The GalNAc 3-hydroxyl hydrogen bonds to Glu280 Oε2 and Asn292 Nδ2, and the 4-hydroxyl to Nε2 of Gln267 and Oδ2 of Asp269. The only other direct hydrogen bond between the sugar and the protein is between the carbonyl of the acetamide group and Nε2 of the His286 imidazole group, an interaction that is unique to MGL and for which there is no equivalent in ASGPR, the closest cousin of MGL, which contains a serine at this position. This interaction was not predicted in the extensive cassette transfer and mutagenesis experiments performed using rat mannose binding protein and likely results in a higher affinity of MGL for GalNAc compared to ASGPR.[47] The methyl group of the acetamide and the α-face of GalNAc moiety (H5, H6a, or H6b) are engaged in CH−π interactions with the aromatic side chain of Tyr236 and

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

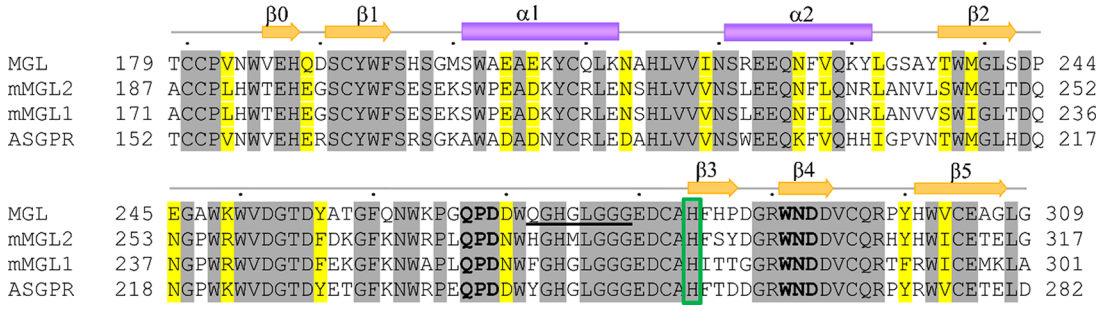


Figure 3. Comparison of the sequences of human macrophage C-type lectin (MGL; Uniprot Q8IUN9), the mouse homologues MGL1 (Uniprot P49300) and MGL2 (Uniprot Q8JZN1), and the human asialoglycoprotein receptor (ASGPR; Uniprot P07306) in the CRD regions. Identical and similar amino acids are highlighted in gray and yellow, respectively. The secondary structure elements of MGL are indicated on the top of the sequence and depicted as beige arrows for β-strands and purple rectangles for α-helices. The region of the glycine rich loop that could not be modeled in the structure of unbound MGL is underlined. The QPD and WND motifs are shown in bold. A histidine that is conserved in several endocytic receptors is highlighted by a green box.

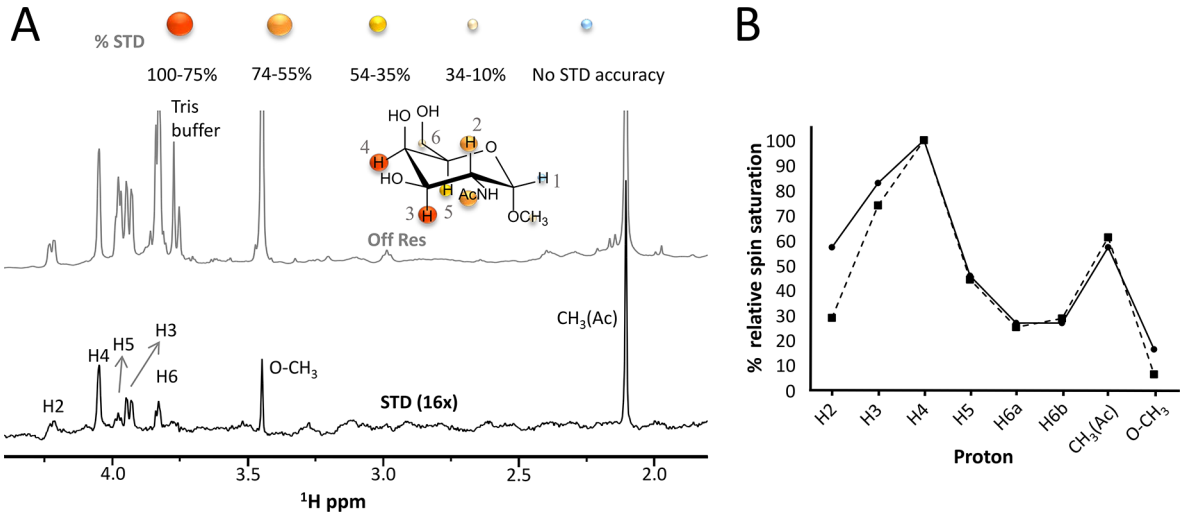


Figure 4. STD-NMR experiment for Me-GalNAc in the presence of the MGL CRD. (A) Reference (off-resonance) and STD spectra are colored gray and black, respectively. Resonances for Me-GalNAc protons are marked in the STD spectrum. The epitope map for Me-GalNAc derived from the STD spectrum is shown on top. The irradiation frequency was 7 ppm (aromatic region), and the saturation time was 2 s. (B) Percentage relative spin saturation for each proton of Me-GalNAc normalized to 100% for the H4 proton that had the highest spin saturation. The solid and dashed lines represent the values obtained from the STD experiment and those calculated values from CORCEMA-st, respectively.

Trp271, respectively (Figure 2B). As in other C-type lectins, the calcium ion coordination is believed to enhance the glycan CH polarization and the CH$-\pi$ interaction between the α-face of the sugar and an aromatic residue, in this case Trp271. The interactions involving the acetamide group justify the fact that MGL has >70-fold greater affinity for the methyl glycoside of GalNAc compared to the corresponding galactopyranoside.[18] Finally, a hydrogen bond between the side chains of Lys264 in the long loop and Asp288 in the $\beta3-\beta4$ loop closes out the carbohydrate binding pocket (Figure 2B). The structure of MGL CRD in complex with Tn-Ser shows that no additional contacts are made with the serine residue of the Tn antigen.

Analysis of the carbohydrate binding site reveals a titratable histidine (His284) that is conserved in MGL, MGL1, MGL2, and ASGPR (Figure 3) and many other endocytic receptors. His284 is a hydrogen bond acceptor for the amide side chain of Asn292, which in turn coordinates Ca^{2+} in the carbohydrate binding site and the GalNAc 3-hydroxyl group (Figure 2B). Protonation of this histidine at endosomal pH would destabilize this hydrogen bonding network and promote ligand release. Indeed, previous studies identified this residue as one that caused a decrease in the pH that supports half-maximal ligand binding when mutated (H284Q) in ASGPR,[48] facilitating the maintenance of ligand binding at physiological pH and release at endosomal pH.

In the crystal structures, the MGL CRD is shown to interact with GalNAc in binding mode A, in contrast with previous STD-NMR work using the MGL ECD that proposed the coexistence of two binding modes with mode B seemingly preferred (Figure S1). To determine whether the interpretation of STD-NMR data was based on STD effects using the MGL ECD protein or on the use of homology models of MGL derived from the crystal structure of the ASGPR CRD, we revisited the analysis using the current crystal structure as a model to interpret the data.

Structural Insights into the MGL CRD/GalNAc Complex Revealed by STD-NMR. The discrepancy between the crystal structures and the previous STD-NMR-proposed binding model prompted us to use molecular dynamics (MD) to examine the stability of the crystal structure of the MGL CRD in complex with GalNAc and the homology models derived from the crystal structure of ASGPR in binding modes A and B (Figure S1). Complexes with GalNAc in binding mode A were stable throughout the 500 ns simulation, whereas

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

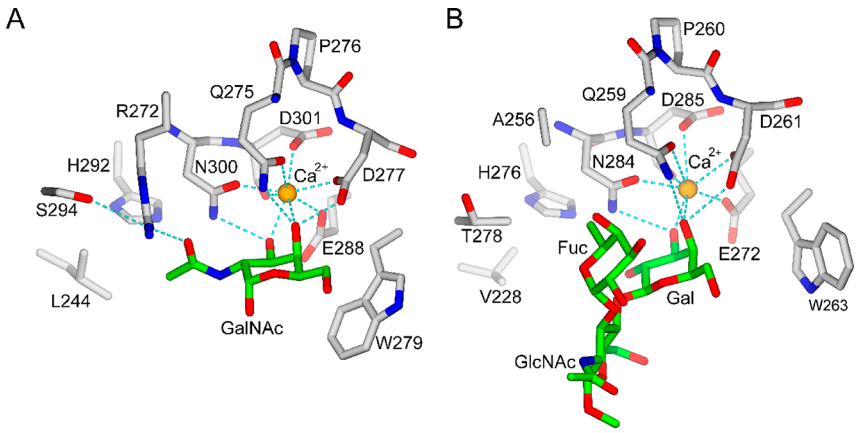


Figure 5. (A) Homology model of MGL2 CRD (gray sticks) in complex with GalNAc (green sticks). The calcium atom is colored orange, and metal-coordinating and hydrogen bonds are shown as dashed cyan lines. Oxygen and nitrogen atoms are colored red and blue, respectively. L244 (Y236 in MGL) makes a hydrophobic interaction with the GalNAc methyl. R282 forms a bridging hydrogen bond between S294 and the acetamide carbonyl of GalNAc. (B) Homology model of the MGL1 CRD in complex with Lewis^X. The orientation of the protein and color scheme is the same as in panel A. A hydrophobic environment composed of the side chains of A256 (R272 in MGL2), T278 (S294 in MGL2), and H276 (H292 in MGL2) accommodates the methyl group of the fucose residue. The side chain of W263 changes its orientation to maintain a CH-π interaction with the galactose residue that is translated slightly relative to its position in MGL and MGL2.

those in binding mode B were stable for only during the first 40 ns (Figure S7). After inspection of the models, the MGL CRD homology model with GalNAc in binding mode A is almost identical to the crystal structure, whereas the model in binding mode B presents considerable differences in the conformation of the long loop (Figure S10). Furthermore, the main interactions (H-bonds and CH$-\pi$) between GalNAc and MGL (Figure S8) are identical in the homology model in binding mode A (Figure S1A) and the crystal structure. Thus, differences in the homology model of MGL derived from the ASGPR structure and the crystal structure of MGL CRD cannot explain the discrepancy in the mixture of binding modes proposed by STD-NMR. For this reason, new STD-NMR data were acquired using both the MGL CRD and the ECD (see Materials and Methods). The experimental spin saturation values derived from the STD-NMR spectra (Figure 4A and Figure S6) acquired using an on-resonance frequency of 7 ppm were compared with the theoretical spin saturation calculated by CORCEMA-st. The best agreement between experimental and theoretical STDs was obtained when low values of k_{on} (2×10^5 for the MGL CRD and 1×10^6 M^{-1} s^{-1} for the MGL ECD) were used in the calculations (Figure 4B and Figure S6B). For decreasing values of k_{on}, the percentage of absolute spin saturation approached the same value for all of the Me-GalNAc protons (Figure S5). This result might be explained by an increase in spin diffusion artifacts due to the longer residence time for Me-GalNAc in the binding site. In the previous analysis, a diffusion-controlled kinetic model with a k_{on} of 1×10^8 M^{-1} s^{-1} was used for CORCEMA-st calculations;[18] however, $^1H-^{15}N$ HSQC-based titrations recorded at 293 K for the ^{15}N-labeled MGL CRD in the presence of Me-GalNAc[24] showed that this assumption is not valid. Indeed, a slow exchange regime between the free and bound states, showing separate signals for both states at 293 K, was observed previously[24] and was now reproduced in experiments recorded at 310 K (the same temperature that was used for the STD-NMR experiments) (Figure S3). A good agreement was obtained between the STD-NMR data and the theoretically calculated values using the crystal structure when k_{on} was adjusted to a value compatible with a slow exchange regime (Figure 4b), indicating that the major binding mode for GalNAc in solution is indeed mode A.

Comparison with MGL1 and MGL2. MGL1 and MGL2 are the mouse homologues of MGL (Figure 3). The degree of sequence similarity in the CRD region between MGL and the mouse proteins is high (87% similar and 66.5% identical between MGL and MGL2 and 83.3% similar and 62.3% identical between MGL and MGL1); therefore, the three-dimensional structure of MGL affords a unique opportunity to use homology modeling to compare MGL with MGL1 and MGL2 and explain their carbohydrate binding preferences. MGL2 has a similar specificity for MGL and is considered its functional homologue, whereas MGL1 recognizes Lewis^X structures.[49]

In a homology model of the MGL2 in complex with GalNAc (Figure 5A), Leu244 (Tyr236 in MGL) makes a hydrophobic interaction with the methyl of the acetamide group. Ser294 (His286 in MGL) does not make a direct hydrogen bond with the GalNAc carbonyl; however, a bridging hydrogen bond could be formed with the side chain Nη1 of Arg272 (Lys264 in MGL). At the equivalent positions in MGL and ASGPR, Lys264 and Arg237 in the long loop make stabilizing hydrogen bonds with the side chains of Asp288 (Figure 2B) and Asp260 in the $\beta3-\beta4$ loop, respectively, but a similar interaction is not possible for Arg272 in MGL2, allowing the side chain guanidinium group to enter the carbohydrate binding pocket (Figure 5A). All interactions involving the galactose ring are mediated by conserved residues in the QPD and WND motifs and are therefore unchanged when compared to MGL (Figure 5A).

In the homology model of an MGL1 in complex with Lewis^X, Val228 and Thr278 replace Tyr236 and His286, respectively, in MGL, effectively removing the binding preference GalNAc and creating a more hydrophobic environment to accommodate the methyl group of the Lewis^X fucose (Figure 5B). Additionally, replacing Lys264 and Asp288 in MGL with Ala256 and Gly280, respectively, in MGL1 removes the hydrogen bond and creates a large hydrophobic pocket to accommodate the fucose residue. Indeed, this model is supported by site-directed mutagenesis experiments that

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

showed that mutation of Ala256 or Thr278 decreased the binding affinity of MGL1 for the Lewis^X trisaccharide.[50] In addition, these amino acids showed weakened signals when the fucose methyl protons of Lewis^X were selectively saturated in saturation transfer NMR experiments.[50] Further structural studies will be essential to validate the predictions made using these models.

CONCLUSION

We report the first crystal structures of the human MGL CRD in the unbound state at near endosomal pH and low calcium concentration, and in complex with GalNAc and the biologically relevant Tn-Ser antigen and the mimetic GalNAc derivative methyl 2-(acetylamino)-2-deoxy-1-thio-α-D-galacto-pyranose. At endosomal pH, a section of the glycine rich long loop from Asp270 to Asp281 is flexible and could not be modeled due to the absence of electron density (Figures 1A and 2A). In the bound state, GalNAc- and GalNAc-derived molecules show the same binding mode (mode A) and all of the interactions are conserved between the different complexes. A unique interaction between the acetamide group of the bound GalNAc and His286 of MGL is present and was not observed in the structure of a GalNAc mimic in complex with the ASGPR CRD that has a threonine at this position.

The carbohydrate binding mode in the crystal structures presented here is in contrast with the results of previously proposed STD-NMR experiments using the MGL ECD that suggested a mixture of binding modes with a preference for mode B. This discrepancy prompted us to revisit the STD-NMR experiments, through the acquisition of new data using the MGL CRD as the construct and the MGL crystal structure as a model to interpret the data. Remarkably, when a k_{on} value compatible with a slow exchange between GalNAc in the free and bound states was used, a very good agreement is achieved between the experimental STD-NMR data and that predicted by CORCEMA-st in binding mode A. This indicates that the prevalent mode of GalNAc recognition by MGL in solution is the same as that in the crystal structures.

Homology models of mouse MGL1 and MGL2 protein generated using the crystal structure of MGL explained their binding preferences for Lewis^X and GalNAc, respectively. The QPD and WND motifs that select for the galactose ring are conserved on both proteins. However, a leucine in MGL2 replaces Tyr236 in MGL in making a hydrophobic interaction with the acetamide methyl group of the GalNAc and the side chain of Arg272 replaces His286 to hydrogen bond with acetamide carbonyl. Three amino acid substitutions in human MGL (Tyr236 to Val, Lys264 to Ala, and His286 to Thr) in the $\alpha2-\beta2$ loop, the $\beta2-\beta3$ loop, and the $\beta3$ strand remove the binding preference for GalNAc in MGL1 and create a more hydrophobic environment to accommodate the methyl group of the Lewis^X fucose residue.

The data reported here provide the three-dimensional information required for the rational design of therapeutics targeting MGL not only to design highly specific ligands but also to investigate their presentation to the immune system.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.1c00009.

Figures S1−S10, Table S1, and additional references (PDF)

Accession Codes

The atomic coordinates and structure factors for molecules presented herein have been deposited in the Protein Data Bank as entries 6PUV, 6PY1, 6XIY, and 6W12.

AUTHOR INFORMATION

Corresponding Author

Gabriel Birrane − *Division of Experimental Medicine, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, United States;* orcid.org/0000-0002-1759-5499; Phone: (617) 667-0025; Email: gbirrane@bidmc.harvard.edu; Fax: (617) 975-5240

Authors

Adele Gabba − *Division of Experimental Medicine, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, United States; School of Chemistry, National University of Ireland Galway, Galway H91 TK33, Ireland;* orcid.org/0000-0001-8240-6482

Agnieszka Bogucka − *Division of Experimental Medicine, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, United States; School of Chemistry, National University of Ireland Galway, Galway H91 TK33, Ireland;* orcid.org/0000-0001-5317-4344

John G. Luz − *Division of Experimental Medicine, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, United States;* orcid.org/0000-0001-7651-2094

Ana Diniz − *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, 2829-516 Caparica, Portugal;* orcid.org/0000-0003-1698-4668

Helena Coelho − *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, 2829-516 Caparica, Portugal;* orcid.org/0000-0003-1992-8557

Francisco Corzana − *Departamento de Química, Centro de Investigación en Síntesis Química Universidad de La Rioja, 26006 Logroño, Spain;* orcid.org/0000-0001-5597-8127

Francisco Javier Cañada − *Centro de Investigaciones Biológicas Margarita Salas, CSIC, 28040 Madrid, Spain; CIBER de Enfermedades Respiratorias (CIBERES), 28029 Madrid, Spain;* orcid.org/0000-0003-4462-1469

Filipa Marcelo − *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade de Nova de Lisboa, 2829-516 Caparica, Portugal;* orcid.org/0000-0001-5049-8511

Paul V. Murphy − *School of Chemistry, National University of Ireland Galway, Galway H91 TK33, Ireland;* orcid.org/0000-0002-1529-6540

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.1c00009

Funding

This work was supported by funding from Gabrielle's Angel Foundation for Cancer Research to G.B., grants 12/IA/1398 and 16/IA/4419 from Science Foundation Ireland to P.V.M.

and GOIPG/2016/858 from the Irish Research Council to A.G. F.M., H.C., and A.D. acknowledge Fundação para a Ciência e a Tecnologia (FCT-Portugal) for funding Projects IF/00780/2015 and PTDC/BIA-MIB/31028/2017 and UCI-BIO Project UIDB/04378/2020, as well as the Ph.D. grant attributed to A.D. (PD/BD/142847/2018). The NMR spectrometers are part of the National NMR Network (PTNMR) and are partially supported by Infrastructure Project 22161 (co-financed by FEDER through COMPETE 2020, POCI, and PORL and FCT through PIDDAC). F.J.C. acknowledges funding from Agencia Estatal de Investigación (Spain) for Grant RTI2018-094751-B-C22 and CIBERES, an initiative from the Spanish Institute of Health Carlos III. F.C. thanks Agencia Estatal de Investigación (Spain) for Grant RTI2018-099592-B-C2.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

X-ray diffraction data were collected at the ID30B beamline at the European Synchrotron Radiation Facility (Grenoble, France). The authors are grateful for the facilities and support provided by Prof. Jerome Groopman at the Beth Israel Deaconess Medical Center.

■ REFERENCES

(1) van Kooyk, Y., Ilarregui, J. M., and van Vliet, S. J. (2015) Novel insights into the immunomodulatory role of the dendritic cell and macrophage-expressed C-type lectin MGL. *Immunobiology 220* (2), 185−192.

(2) van Vliet, S. J., Saeland, E., and van Kooyk, Y. (2008) Sweet preferences of MGL: carbohydrate specificity and function. *Trends Immunol. 29* (2), 83−90.

(3) Saeland, E., van Vliet, S. J., Backstrom, M., van den Berg, V. C., Geijtenbeek, T. B., Meijer, G. A., and van Kooyk, Y. (2007) The C-type lectin MGL expressed by dendritic cells detects glycan changes on MUC1 in colon carcinoma. *Cancer Immunol. Immunother. 56* (8), 1225−1236.

(4) Li, D., Romain, G., Flamar, A. L., Duluc, D., Dullaers, M., Li, X. H., Zurawski, S., Bosquet, N., Palucka, A. K., Le Grand, R., O'Garra, A., Zurawski, G., Banchereau, J., and Oh, S. (2012) Targeting self- and foreign antigens to dendritic cells via DC-ASGPR generates IL-10-producing suppressive CD4+ T cells. *J. Exp. Med. 209* (1), 109−121.

(5) van Vliet, S. J., Bay, S., Vuist, I. M., Kalay, H., Garcia-Vallejo, J. J., Leclerc, C., and van Kooyk, Y. (2013) MGL signaling augments TLR2-mediated responses for enhanced IL-10 and TNF-alpha secretion. *J. Leukocyte Biol. 94* (2), 315−323.

(6) van Vliet, S. J., Aarnoudse, C. A., Broks-Van den Berg, V. C., Boks, M., Geijtenbeek, T. B., and van Kooyk, Y. (2007) MGL-mediated internalization and antigen presentation by dendritic cells: a role for tyrosine-5. *Eur. J. Immunol. 37* (8), 2075−2081.

(7) Freire, T., Zhang, X., Deriaud, E., Ganneau, C., Vichier-Guerre, S., Azria, E., Launay, O., Lo-Man, R., Bay, S., and Leclerc, C. (2010) Glycosidic Tn-based vaccines targeting dermal dendritic cells favor germinal center B-cell development and potent antibody response in the absence of adjuvant. *Blood 116* (18), 3526−3536.

(8) Laubreton, D., Bay, S., Sedlik, C., Artaud, C., Ganneau, C., Deriaud, E., Viel, S., Puaux, A. L., Amigorena, S., Gerard, C., Lo-Man, R., and Leclerc, C. (2016) The fully synthetic MAG-Tn3 therapeutic vaccine containing the tetanus toxoid-derived TT830−844 universal epitope provides anti-tumor immunity. *Cancer Immunol. Immunother. 65* (3), 315−325.

(9) Rosenbaum, P., Artaud, C., Bay, S., Ganneau, C., Campone, M., Delaloge, S., Gourmelon, C., Loirat, D., Medioni, J., Pein, F., Sablin, M. P., Tredan, O., Varga, A., and Leclerc, C. (2020) The fully synthetic glycopeptide MAG-Tn3 therapeutic vaccine induces tumor-specific cytotoxic antibodies in breast cancer patients. *Cancer Immunol. Immunother. 69* (5), 703−716.

(10) van Vliet, S. J., van Liempt, E., Saeland, E., Aarnoudse, C. A., Appelmelk, B., Irimura, T., Geijtenbeek, T. B., Blixt, O., Alvarez, R., van Die, I., and van Kooyk, Y. (2005) Carbohydrate profiling reveals a distinctive role for the C-type lectin MGL in the recognition of helminth parasites and tumor antigens by dendritic cells. *Int. Immunol. 17* (5), 661−669.

(11) van Vliet, S. J., Steeghs, L., Bruijns, S. C., Vaezirad, M. M., Snijders Blok, C., Arenas Busto, J. A., Deken, M., van Putten, J. P., and van Kooyk, Y. (2009) Variation of Neisseria gonorrhoeae lipooligosaccharide directs dendritic cell-induced T helper responses. *PLoS Pathog. 5* (10), No. e1000625.

(12) Maalej, M., Forgione, R. E., Marchetti, R., Bulteau, F., Thepaut, M., Lanzetta, R., Laguri, C., Simorre, J. P., Fieschi, F., Molinaro, A., and Silipo, A. (2019) Human Macrophage Galactose-Type Lectin (MGL) Recognizes the Outer Core of Escherichia coli Lipooligosaccharide. *ChemBioChem 20* (14), 1778−1782.

(13) Takada, A., Fujioka, K., Tsuiji, M., Morikawa, A., Higashi, N., Ebihara, H., Kobasa, D., Feldmann, H., Irimura, T., and Kawaoka, Y. (2004) Human macrophage C-type lectin specific for galactose and N-acetylgalactosamine promotes filovirus entry. *J. Virol. 78* (6), 2943−2947.

(14) Chiodo, F., Bruijns, S. C. M., Rodriguez, E., Li, R. J. E., Molinaro, A., Silipo, A., Di Lorenzo, F., Garcia-Rivera, D., Valdes-Balbin, Y., Verez-Bencomo, V., and van Kooyk, Y. (2020) Novel ACE2-Independent Carbohydrate-Binding of SARS-CoV-2 Spike Protein to Host Lectins and Lung Microbiota. *bioRxiv*, DOI: 10.1101/2020.05.13.092478.

(15) Lenza, M. P., Oyenarte, I., Diercks, T., Quintana, J. I., Gimeno, A., Coelho, H., Diniz, A., Peccati, F., Delgado, S., Bosch, A., Valle, M., Millet, O., Abrescia, N. G. A., Palazon, A., Marcelo, F., Jimenez-Oses, G., Jimenez-Barbero, J., Arda, A., and Ereño-Orbea, J. (2020) Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and their Interactions with Human Lectins. *Angew. Chem., Int. Ed. 59* (52), 23763−23771.

(16) Ilarregui, J. M., Kooij, G., Rodriguez, E., van der Pol, S. M. A., Koning, N., Kalay, H., van der Horst, J. C., van Vliet, S. J., Garcia-Vallejo, J. J., de Vries, H. E., and van Kooyk, Y. (2019) Macrophage galactose-type lectin (MGL) is induced on M2 microglia and participates in the resolution phase of autoimmune neuroinflammation. *J. Neuroinflammation 16* (1), 130.

(17) Kolatkar, A. R., and Weis, W. I. (1996) Structural basis of galactose recognition by C-type animal lectins. *J. Biol. Chem. 271* (12), 6679−6685.

(18) Marcelo, F., Garcia-Martin, F., Matsushita, T., Sardinha, J., Coelho, H., Oude-Vrielink, A., Koller, C., Andre, S., Cabrita, E. J., Gabius, H. J., Nishimura, S., Jimenez-Barbero, J., and Cañada, F. J. (2014) Delineating binding modes of Gal/GalNAc and structural elements of the molecular recognition of tumor-associated mucin glycopeptides by the human macrophage galactose-type lectin. *Chem. - Eur. J. 20* (49), 16147−16155.

(19) Sugawara, H., Kusunoki, M., Kurisu, G., Fujimoto, T., Aoyagi, H., and Hatakeyama, T. (2004) Characteristic recognition of N-acetylgalactosamine by an invertebrate C-type Lectin, CEL-I, revealed by X-ray crystallographic analysis. *J. Biol. Chem. 279* (43), 45219−45225.

(20) Feinberg, H., Taylor, M. E., and Weis, W. I. (2007) Scavenger receptor C-type lectin binds to the leukocyte cell surface glycan Lewis(x) by a novel mechanism. *J. Biol. Chem. 282* (23), 17250−17258.

(21) Asensio, J. L., Arda, A., Cañada, F. J., and Jimenez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc. Chem. Res. 46* (4), 946−954.

(22) Poget, S. F., Legge, G. B., Proctor, M. R., Butler, P. J., Bycroft, M., and Williams, R. L. (1999) The structure of a tunicate C-type lectin from Polyandrocarpa misakiensis complexed with D -galactose. *J. Mol. Biol. 290* (4), 867−879.

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336

(23) Feinberg, H., Taylor, M. E., Razi, N., McBride, R., Knirel, Y. A., Graham, S. A., Drickamer, K., and Weis, W. I. (2011) Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J. Mol. Biol. 405* (4), 1027−1039.
(24) Diniz, A., Coelho, H., Dias, J. S., van Vliet, S. J., Jimenez-Barbero, J., Corzana, F., Cabrita, E. J., and Marcelo, F. (2019) The Plasticity of the Carbohydrate Recognition Domain Dictates the Exquisite Mechanism of Binding of Human Macrophage Galactose-Type Lectin. *Chem. - Eur. J. 25* (61), 13945−13955.
(25) Wang, Y., van Oosterwijk, N., Ali, A. M., Adawy, A., Anindya, A. L., Domling, A. S. S., and Groves, M. R. (2017) A Systematic Protein Refolding Screen Method using the DGR Approach Reveals that Time and Secondary TSA are Essential Variables. *Sci. Rep. 7* (1), 9355.
(26) Kaltner, H., Manning, J. C., García Caballero, G., Di Salvo, C., Gabba, A., Romero-Hernández, L. L., Knospe, C., Wu, D., Daly, H. C., O'Shea, D. F., Gabius, H.-J., and Murphy, P. V. (2018) Revealing biomedically relevant cell and lectin type-dependent structure−activity profiles for glycoclusters by using tissue sections as an assay platform. *RSC Adv. 8* (50), 28716−28735.
(27) Kabsch, W. (2010) Xds. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66* (Part 2), 125−132.
(28) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40* (4), 658−674.
(29) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67* (4), 235−242.
(30) Meier, M., Bider, M. D., Malashkevich, V. N., Spiess, M., and Burkhard, P. (2000) Crystal structure of the carbohydrate recognition domain of the H1 subunit of the asialoglycoprotein receptor. *J. Mol. Biol. 300* (4), 857−865.
(31) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28* (1), 235−242.
(32) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60* (12), 2126−2132.
(33) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53* (3), 240−255.
(34) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66* (1), 12−21.
(35) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct., Funct., Genet. 59* (4), 687−696.
(36) Jayalakshmi, V., and Rama Krishna, N. (2004) CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J. Magn. Reson. 168* (1), 36−45.
(37) Krishna, N. R., and Jayalakshmi, V. (2007) Quantitative Analysis of STD-NMR Spectra of Reversibly Forming Ligand-Receptor Complexes. *Top. Curr. Chem. 273*, 15−54.
(38) Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, I. T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Botello-Smith, W. M., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., and Kollman, P. A. *Amber 2016*; University of California, San Francisco: San Francisco, 2016.
(39) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet. 65* (3), 712−725.
(40) Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem. 29* (4), 622−655.
(41) Ramachandran, S., Kota, P., Ding, F., and Dokholyan, N. V. (2011) Automated minimization of steric clashes in protein structures. *Proteins: Struct., Funct., Genet. 79* (1), 261−270.
(42) Valverde, P., Martinez, J. D., Cañada, F. J., Arda, A., and Jimenez-Barbero, J. (2020) Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem 21* (21), 2999−3025.
(43) Jegouzo, S. A., Quintero-Martinez, A., Ouyang, X., dos Santos, A., Taylor, M. E., and Drickamer, K. (2013) Organization of the extracellular portion of the macrophage galactose receptor: a trimeric cluster of simple binding sites for N-acetylgalactosamine. *Glycobiology 23* (7), 853−864.
(44) Walker, J. R., Nagar, B., Young, N. M., Hirama, T., and Rini, J. M. (2004) X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. *Biochemistry 43* (13), 3783−3792.
(45) Ng, K. K., Kolatkar, A. R., Park-Snyder, S., Feinberg, H., Clark, D. A., Drickamer, K., and Weis, W. I. (2002) Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J. Biol. Chem. 277* (18), 16088−16095.
(46) Somers, W. S., Tang, J., Shaw, G. D., and Camphausen, R. T. (2000) Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell 103* (3), 467−479.
(47) Kolatkar, A. R., Leung, A. K., Isecke, R., Brossmer, R., Drickamer, K., and Weis, W. I. (1998) Mechanism of N-acetylgalactosamine binding to a C-type animal lectin carbohydrate-recognition domain. *J. Biol. Chem. 273* (31), 19502−19508.
(48) Wragg, S., and Drickamer, K. (1999) Identification of amino acid residues that determine pH dependence of ligand binding to the asialoglycoprotein receptor during endocytosis. *J. Biol. Chem. 274* (50), 35400−35406.
(49) Singh, S. K., Streng-Ouwehand, I., Litjens, M., Weelij, D. R., Garcia-Vallejo, J. J., van Vliet, S. J., Saeland, E., and van Kooyk, Y. (2009) Characterization of murine MGL1 and MGL2 C-type lectins: distinct glycan specificities and tumor binding properties. *Mol. Immunol. 46* (6), 1240−1249.
(50) Sakakura, M., Oo-Puthinan, S., Moriyama, C., Kimura, T., Moriya, J., Irimura, T., and Shimada, I. (2008) Carbohydrate binding mechanism of the macrophage galactose-type C-type lectin 1 revealed by saturation transfer experiments. *J. Biol. Chem. 283* (48), 33665−33673.

https://dx.doi.org/10.1021/acs.biochem.1c00009
Biochemistry 2021, 60, 1327−1336