



Universiteit  
Leiden  
The Netherlands

## Glycoproteomics-based signatures of cancer cell lines

Pirro, M.

### Citation

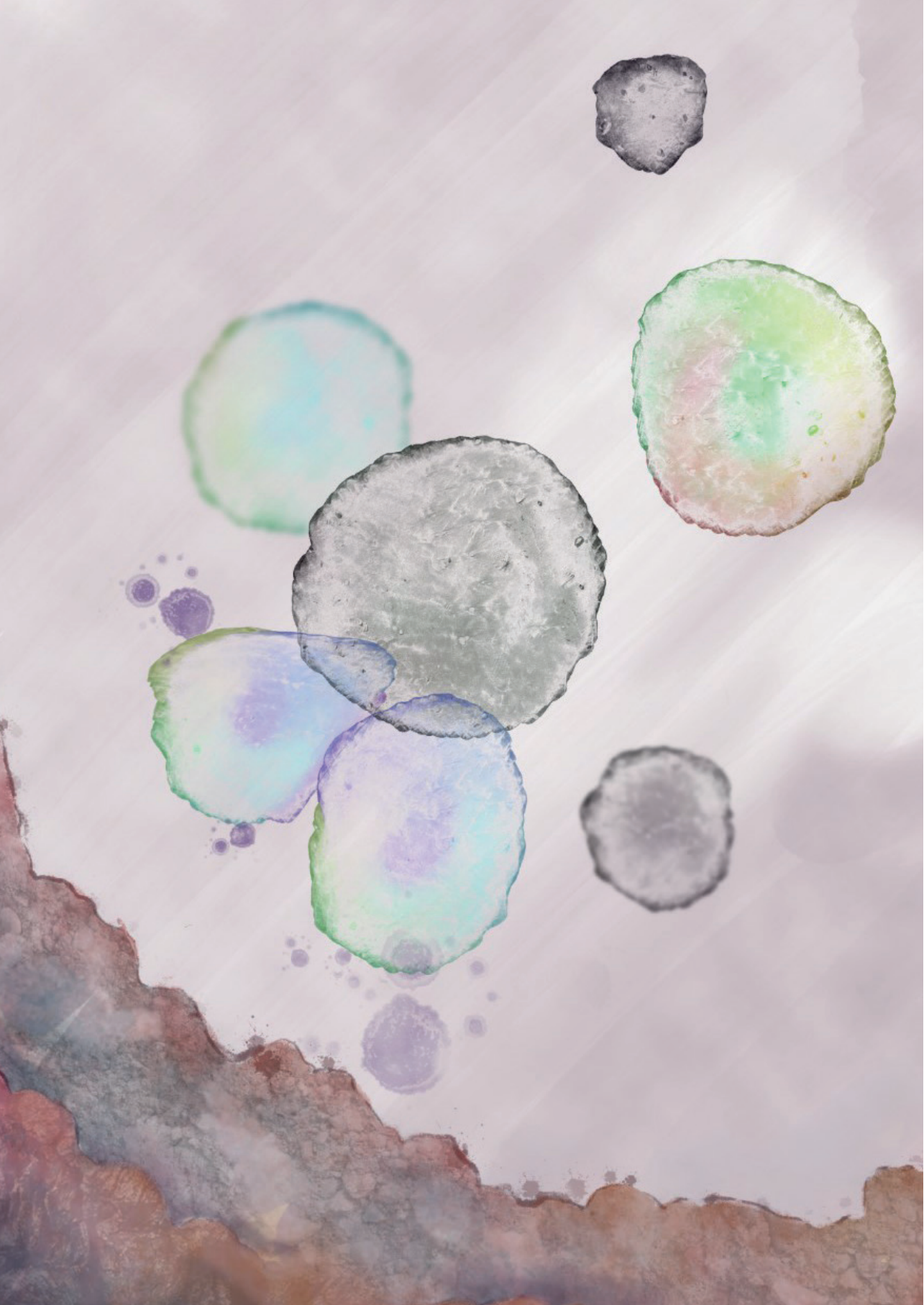
Pirro, M. (2021, September 21). *Glycoproteomics-based signatures of cancer cell lines*. Retrieved from <https://hdl.handle.net/1887/3212951>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3212951>

**Note:** To cite this publication please use the final published version (if applicable).



# CHAPTER 2

## GLYCOPROTEOMIC ANALYSIS OF MGL-BINDING PROTEINS ON ACUTE T-CELL LEUKEMIA CELLS

Martina Pirro<sup>1</sup>, Esmee Schoof<sup>1</sup>, Sandra J. van Vliet<sup>2</sup>, Yoann Rombouts<sup>3</sup>, Alexandre Stella<sup>3</sup>, Arnoud de Ru<sup>1</sup>, Yassene Mohammed<sup>1</sup>, Manfred Wuhrer<sup>1</sup>, Peter A. van Veelen<sup>1</sup>, Paul J. Hensbergen<sup>1</sup>

<sup>1</sup> Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

<sup>2</sup> Amsterdam UMC, Vrije Universiteit Amsterdam, dept. of Molecular Cell Biology and Immunology, Cancer Center Amsterdam, Amsterdam Infection & immunity Institute, Amsterdam, the Netherlands

<sup>3</sup> Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, Toulouse, France.

Reprinted and adapted from *J. Proteome Res.* 2019, 18, 3, 1125–1132

DOI: [10.1021/acs.jproteome.8b00796](https://doi.org/10.1021/acs.jproteome.8b00796)

## Abstract

C-type lectins are a diverse group of proteins involved in many human physiological and pathological processes. Most C-type lectins are glycan binding proteins, some of which are pivotal for innate immune responses against pathogens. Other C-type lectins, such as the macrophage galactose-type lectin (MGL), have been shown to induce immunosuppressive responses upon recognition of aberrant glycosylation on cancer cells. MGL is known to recognize terminal *N*-acetylgalactosamine (GalNAc), such as the Tn-antigen, which is commonly found on malignant cells. Even though this glycan specificity of MGL is well described, there is a lack of understanding of the actual glycoproteins that bind MGL. In this paper, we present a glycoproteomic workflow for the identification of MGL binding proteins, which we applied to study MGL ligands on the human Jurkat leukemia cell line. In addition to the known MGL ligands and Tn-antigen carrying proteins CD43 and CD45 on these cells, we have identified a set of novel cell surface ligands for MGL. Importantly, for several of these, *O*-glycosylation has hitherto not been described. Altogether, our data provide new insight into the identification and structure of novel MGL ligands that presumably act as modulatory molecules in cancer immune responses.

## Introduction

C-type lectins belong to a family of glycan-binding proteins (GBPs) which interact with their ligands in a  $\text{Ca}^{2+}$ -dependent manner through a carbohydrate recognition domain (CRD) [30, 57]. They exist as monomers or can oligomerize to enhance the avidity by multivalent glycan interactions. Their activation on immune cells, such as DCs, triggers different immune responses depending on the crosstalk with other receptors, the ligand/carbohydrate specific signaling and the cell subset they are expressed in [58]. Within the CLR family, the macrophage galactose-type lectin (MGL) is expressed on tolerogenic dendritic cells (DC) and macrophages [31] and has the capability to alter DC and macrophage phenotypes [41].

MGL is the only lectin that exclusively binds terminal GalNAc residues, including three well known tumor-associated glycan epitopes [59]: the sialylated and non-sialylated Thomsen/nouvelle (Tn) antigen (GalNAc $\alpha$ 1-Ser/Thr) and LacdiNAc (GalNAc $\beta$ 1-4GlcNAc) [31]. In normal cells, the Tn-antigen is usually elongated with other carbohydrate residues, for example a galactose to form the core 1 T antigen. This step is mediated by the enzyme T-synthase with the assistance of its chaperone *Cosmc*. Mutations in *Cosmc* are responsible for the abortive elongation of *O*-glycans and higher expression of the Tn antigen [60], as observed in Tn syndrome patients [61]. High levels of Tn antigen have also been found in multiple human cancer types such as colon, cervix, stomach, ovary and breast [62]. Moreover, higher levels of this surface truncated glycan were found on cancer cells with high metastatic behaviour and consequently poor prognosis of cancer patients [61]. Interestingly, the interaction of MGL with aberrant glycosylation on cancer cells is associated with the activation of immunosuppressive responses, immune tolerance [63] and poor survival of stage III colorectal cancer patients [45], suggesting a role of MGL in mediating cancer progression.

Notwithstanding our current knowledge on the role of MGL and the glycans it binds, the proteins carrying the MGL ligands are hitherto largely unknown. Therefore, in this paper we sought to identify MGL-binding proteins in a T cell leukemia model cell line, Jurkat, which is known to have high levels of Tn antigen due to a single nucleotide deletion in *Cosmc* and has been used as a model system to study immune modulation mediated by MGL [30, 39].

## Material and Methods

### **Cell lines culture and lysis**

Jurkat (provided by S.J. van Vliet, VUMC, Amsterdam, the Netherlands) cells were cultivated in RPMI-1640 medium containing 10% fetal bovin serum (FBS) (Invitrogen) and streptomycin/penicillin (Sigma-Aldrich) at 5% CO<sub>2</sub> and 37°C. Cells were harvested upon approximately 70% confluency. A total of 2 x 10<sup>7</sup> cells were obtained, harvested, washed and centrifuged at 1500 rpm to obtain cell pellets. Cell pellets were stored at -20°C until use. Then, cells were lysed as described before [39], for 20 min on ice in lysis buffer (10mM triethanolamine pH 8.2, 150mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub> and 1% (volume/volume) Triton X-100, containing EDTA-free protease inhibitor (Roche Diagnostics)). Protein concentration was quantified by BCA assay (BCA Protein Assay Kit, Pierce™).

### **Antibodies and reagent**

Chimeric MGL-Fc was provided by S.J. van Vliet (VUMC, Amsterdam, the Netherlands). The following antibodies were used for immunoblotting: mouse IgG1 anti-human CD43 (eBio84-3C1, eBioscience™) and goat anti-mouse immunoglobulins/HRP (Dako).

### **Pull-down assay and immunoblot analysis**

MGL ligands were pulled down from 1 mg of protein extracts with 2 µg of chimeric MGL-Fc coupled to 50 µL of Dynabeads protein G (Invitrogen). The specific ligands were then eluted by the addition of 100 mM EDTA and concentrated under vacuum. The addition of 100 mM EDTA, prior to the addition of MGL-Fc, was used as a negative control. Captured products were separated by SDS-PAGE (NuPAGE 4–12% Bis-Tris protein gels, Thermo Fisher Scientific) and transferred to a PVDF membrane (Amersham Hyband P 0.45). Five % low-fat milk (Campina) in 0.1% phosphate-buffered saline with Tween 20 (PBS-T) was used to block the blots for 1 h. Immunoblotting was performed with specific antibodies, followed by secondary antibody peroxidase-conjugated goat anti-mouse (Dako). Immunodetection was done by enhanced chemiluminescence (ECL) using Clarity Western ECL substrate (Bio-Rad) and an Amersham Imager 600.

### **Mass spectrometry**

The samples obtained with the MGL pull-down assay were cleaned up by a short SDS-PAGE run (NuPAGE 4–12% Bis- Tris protein gels, Thermo Fisher Scientific). For MS analysis, in-gel trypsin digestion was performed using a Proteineer DP digestion robot (Bruker). Prior to digestion, proteins were first reduced and alkylated using dithiothreitol (10 mM) and iodoacetamide (50 mM), respectively. Tryptic peptides were extracted from the gel slices, lyophilized, dissolved in solvent A (95/3/0.1 water/acetonitrile/formic acid (FA) v/v/v), and subsequently analyzed by online C18 nano-HPLC MS/MS with a system consisting of an Easy nLC 1000 gradient HPLC system (Thermo, Bremen, Germany) and a LUMOS mass spectrometer (Thermo). Fractions were injected onto a homemade precolumn (100  $\mu\text{m}$   $\times$  15 mm; Reprosil-Pur C18-AQ 3  $\mu\text{m}$ , Dr. Maisch, Ammerbuch, Germany) and eluted via a homemade analytical nano-HPLC column (15 cm  $\times$  50  $\mu\text{m}$ ; Reprosil-Pur C18-AQ3  $\mu\text{m}$ ). The gradient was run from 10 to 40% solvent B (20/80/0.1 water/acetonitrile/FA v/v/v) in 20 min. The nano-HPLC column was drawn to a tip of  $\sim$ 5  $\mu\text{m}$  and acted as the electrospray needle of the MS source. The LUMOS mass spectrometer was operated in data-dependent MS/MS (top-10 mode) with collision energy at 32 V and recording of the MS<sup>2</sup> spectrum in the Orbitrap. In the master scan (MS<sup>1</sup>) the resolution was 120 000, and the scan range was  $m/z$  400–1500 at an AGC target of 400 000 with maximum fill time of 50 ms. Dynamic exclusion after  $n = 1$  with exclusion duration of 10 s was applied. Charge states 2–5 were included for MS<sup>2</sup>. For this, precursors were isolated with the quadrupole with an isolation width of 1.2 Da. The MS<sup>2</sup> scan resolution was 30 000 with an AGC target of 50 000 with maximum fill time of 60 ms. During acquisition, a product ion trigger was set on the HexNAc oxonium ion at  $m/z$  204.087. Upon the detection of the oxonium ion, three additional data-dependent MS<sup>2</sup> scans of the same precursor were executed with higher-energy collisional dissociation (HCD) collision energies of 32, 37, and 41 V, respectively, at an AGC target of 500 000 with a maximum fill time of 200 ms. In addition, a collision-induced dissociation (CID) spectrum of the same precursor was recorded. In a post-analysis process, raw data were analyzed with MaxQuant 1.5.1.2 using the Andromeda search engine and the Homo sapiens (2015) database. Modifications taken into account in the search were oxidation (M), acetylation (N-term), and HexNAc (ST) as variable and carbamidomethyl (C) as fixed. Trypsin was selected as the enzyme, allowing up to two miss cleavages. A maximum of 12 modifications per peptide was allowed. The mass tolerance

used for precursor ions was 10 ppm, whereas it was 0.05 Da for fragment ions. Subsequently, only proteins that were either never found in the negative control experiment but were found in at least two of the MGL pull-downs or proteins that were identified in all three MGL pull-downs and identified a maximum of one time in the negative control samples were subsequently considered MGL-binding proteins. Subsequently, these MGL-binding proteins were further filtered based on data on the subcellular location from the UniProtKB database, thereby selecting only cell-surface membrane proteins. The surface location was subsequently manually curated by a literature search. Alternatively, MS<sup>2</sup> spectra containing the specific HexNAc oxonium ions at  $m/z$  204.087 (HexNAc, [C<sub>8</sub>H<sub>14</sub>NO<sub>5</sub>]<sup>+</sup>) and 186.076 (HexNAc-H<sub>2</sub>O, [C<sub>8</sub>H<sub>12</sub>NO<sub>4</sub>]<sup>+</sup>) were selected and written to an MGF file (in-house software), which was subsequently submitted to Byonic version 2.13.2 (proteinmetrics.com) using default settings. The fixed modification was carbamidomethyl (Cys). Variable modifications set were oxidation (Met) and “N-glycan 309 mammalian no sodium” or “O-glycan 78 mammalian” databases as glycosylation parameters. Only glycopeptides with a Byonic score >200 were further selected. Manual interpretation of spectra was done using Xcalibur (Thermo) for the spectrum visualization. The corresponding peptide fragment masses were calculated using Protein Prospector (prospector.ucsf.edu).

### **Data availability**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [64] partner repository with the dataset identifier PXD011307.

## Results

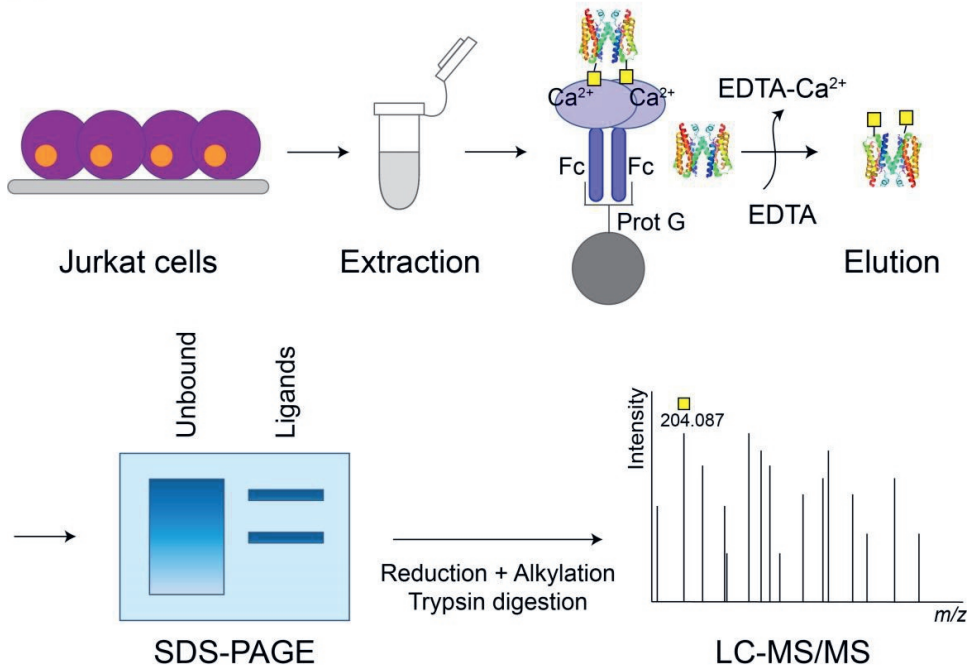
### **Workflow of the MGL pull-down assay**

To identify novel binding partners of MGL on Jurkat T-cells, we developed a protocol where we used Fc-coupled MGL as a bait in pull-down assays (Figure 1A). MGL-Fc is a chimeric molecule, formed by the extracellular domains of MGL fused to the human immunoglobulin G1 Fc tail [38], allowing binding to magnetic Protein G beads. Because the binding to the CRD of MGL is calcium-dependent, captured proteins were eluted using EDTA and subsequently analyzed by SDS-PAGE and processed for mass-spectrometry-based protein identification.

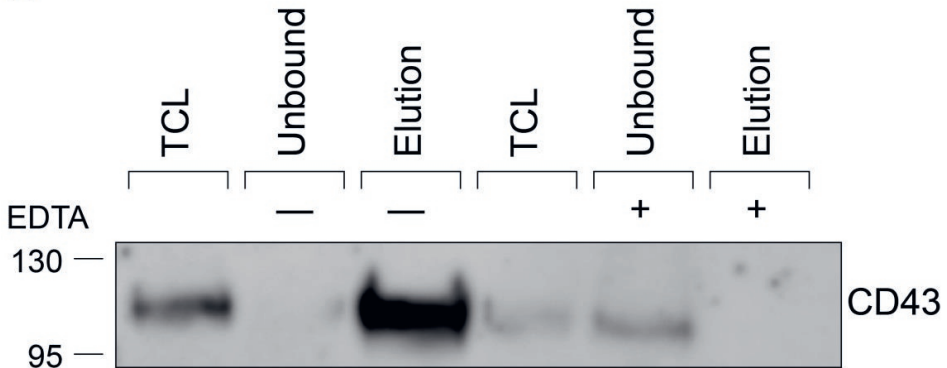


CD43 and CD45 are hitherto the only described MGL-binding proteins on Jurkat cells [39]. Therefore, to test the effectiveness of our workflow, we first determined the capturing of CD43 by western blot. This clearly showed that CD43 was captured by MGL-Fc and could be eluted with EDTA (Figure 1B). Likewise, the addition of EDTA during the pull-down assay prevented the binding of CD43 to MGL (Figure 1B).

**A**



**B**



**Figure 1: Schematic representation and validation of the experimental workflow for the identification of MGL ligands in Jurkat cells.** (A) Pull-down workflow to capture MGL-binding proteins. As negative control, EDTA was added to the sample prior to incubation with MGL-Fc to prevent the binding to MGL. (B) Western blot analysis, using monoclonal anti-CD43 staining of the unbound fraction and captured proteins (elution) from a MGL-Fc pull-down assay in the absence (-) and presence (+, negative control) of EDTA (100 mM). Fc: Fragment crystallisable region; Prot G: protein G; TCL: total cell lysate.

### **MS-based identification of MGL binding proteins from Jurkat cells**

Next, we performed three biologically independent pull-down experiments with MGL-Fc and analyzed the bound proteins by LC-MS/MS following trypsin digestion. As a negative control, pull-down assays in the presence of EDTA were performed. Altogether, these experiments resulted in the identification of 775 proteins (data not shown), of which 540 were identified in at least two experiments. To filter for proteins specifically binding to MGL, we selected proteins that were not observed in the negative controls and at least two times in the pull-downs with MGL or proteins that were found in all three pull-downs with MGL and at most once in a negative control. This resulted in a list of 85 proteins (Table S1). The candidate MGL-binding proteins included intracellular, plasma membrane, and predicted secreted proteins. Because we used total cell lysates for the pull-down assays, some of these proteins may be physiologically less relevant. To filter for cell-surface proteins, which may be expected to be in direct contact with MGL, we next selected only those proteins that were annotated as cell-surface proteins in the UniProtKB database, resulting in a final list of 17 MGL-binding cell-surface proteins (Table 1). Importantly, these proteins were not found in the negative control samples (see also Table S1), making them strong candidate MGL-binding proteins. As expected, CD45 (PTPRC) and CD43 (SPN) were the two main MGL-binding proteins identified based on the total number of peptide-spectrum matches (summed for the three biological replicates) for these two proteins, respectively.

**Table 1: MGL binding cell surface proteins from Jurkat cells.** Using total cell lysates from Jurkat cells, three independent pull-down assays with MGL-Fc were performed and captured proteins were identified by LC-MS/MS. Shown are the MGL-binding cell surface proteins, for a full list see Suppl. Table S1. Each protein is represented with the number of unique peptides identified in the three pull-down assays and the summed spectral count from all three assays. Mol. Weight: Molecular weight; Exp: Experiment number; #pep: number of unique peptides per protein; Sum MS/MS count: summed spectral count from the three assays; Glycopeptide: O-glycopeptide(s) of the respective protein found (yes) or not (no).

Protein Name	Gene Name	Uniprot Entry	Mol. Weight [kDa]	Exp 1: #pep	Exp 2: #pep	Exp 3: #pep	Sum MS/MS Count	Glycopeptide
Receptor-type tyrosine-protein phosphatase C (CD45)	PTPRC	P08575	147	73	60	62	1125	yes
Leukosialin (CD43)	SPN	P16150	40	11	9	6	145	yes
Leucine-rich repeat-containing protein 8D	LRR8D	Q7L1W4	98	21	22	10	73	yes
Podocalyxin	PODXL	O00592	59	7	8	6	66	yes
Leucine-rich repeat-containing protein 8A	LRR8A	Q8IWT6	94	15	16	6	53	no
Protein EVI2B	EVI2B	P34910	49	7	10	4	35	yes
Dyslexia-associated protein KIAA0319-like protein	KIAA0319L	E7EN73	113	12	9	7	30	no
Leucine-rich repeat-containing protein 8C	LRR8C	Q8TDW0	92	8	11	4	28	no
Protein tyrosine phosphatase receptor type C-associated protein	PTPRCAP	Q14761	21	5	3	0	11	no
Semaphorin-4D	SEMA4D	Q92854	96	7	2	2	10	yes
Trem-like transcript 2 protein	TREML2	Q5T2D2	35	3	2	2	6	yes
Protein HEG homolog 1	HEG1	Q9JUL3	147	1	2	1	6	no
Lysosome-associated membrane glycoprotein 3	LAMP3	E7ETP9	42	3	1	1	5	yes
Chloride intracellular channel protein 1	CLIC1	O00299	27	2	3	0	5	no
Low-density lipoprotein receptor	LDLR	HOYM92	34	2	1	1	4	no
Leucine-rich repeat-containing protein 8B	LRR8B	Q6P9F7	92	1	3	0	4	no
Tumor necrosis factor receptor superfamily member 8	TNFRSF8	P28908	64	2	3	0	3	yes

### **Characterization of glycopeptides**

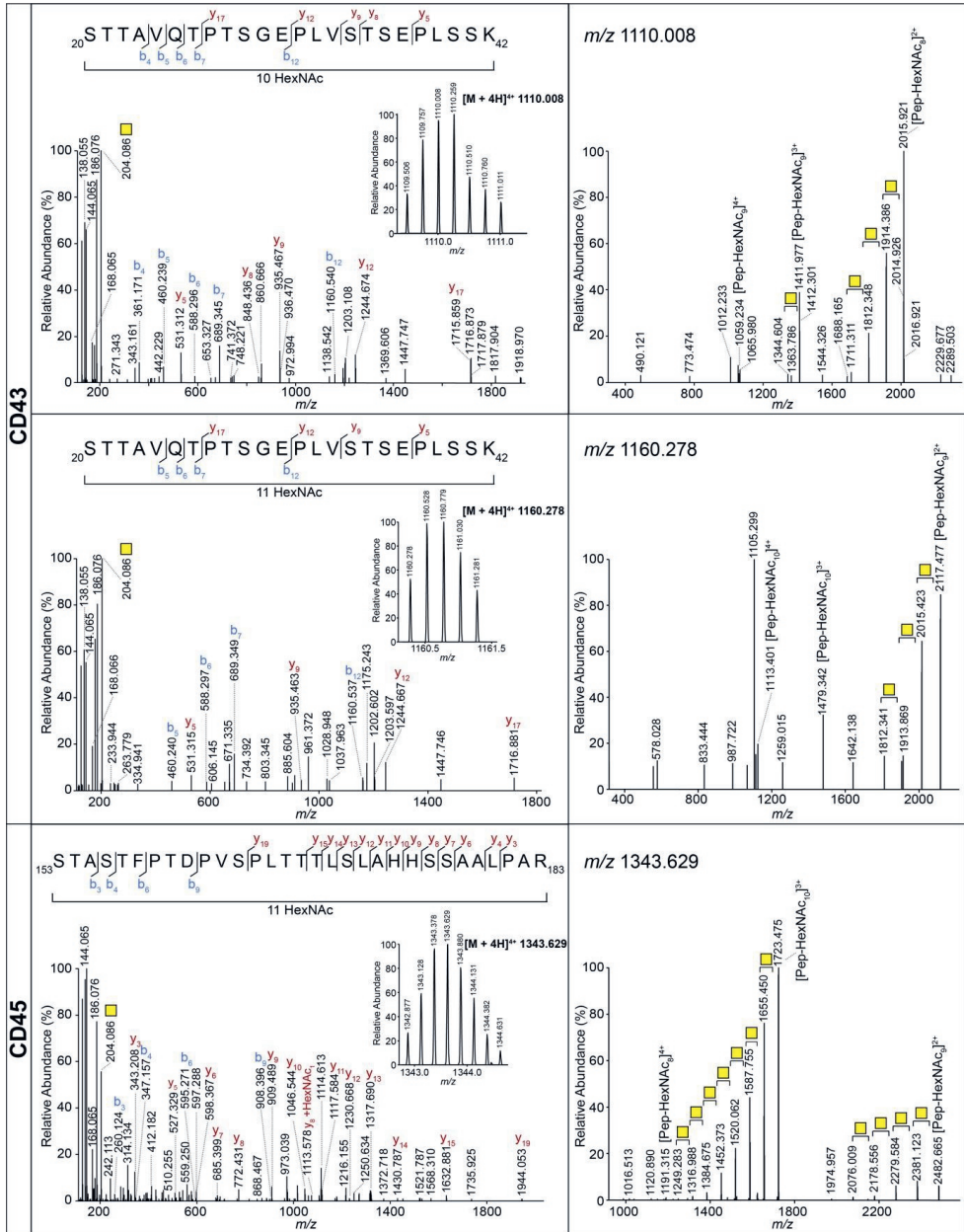
MGL is known to bind terminal residues of N-acetylgalactos-amine. To validate the specificity of the interaction of the proteins listed in Table 1 with MGL, we next evaluated our data for the presence of glycopeptides carrying this motif. During our LC-MS/MS analyses of the tryptic digests, we used a method that triggered additional MS/MS acquisitions once the MS/MS spectrum showed the presence of the characteristic HexNAc oxonium ion at  $m/z$  204.087 ( $[C_8H_{14}NO_5]^+$ ). First of all, for these additional MS/MS events, we used a higher number of ions for fragmentation, resulting in a higher signal-to-noise for the fragment ions. Moreover, to improve the confidence of the glycopeptide identifications, we collected Orbitrap spectra using different HCD collision energies in addition to an ion trap CID spectrum. Next, we searched for glycopeptides carrying the Tn antigen following database searches using MaxQuant and Byonic. With these approaches, we could confirm the presence of the Tn antigen on CD45 and CD43 (Table 1) with peptides containing a maximum of four HexNAcs (Table 2). However, it is known that certain regions in these two proteins contain a high density of *O*-glycans, which may have been missed by the automatic data analysis due to limitations of the particular software used, for example, due to the high number of occupied glycosylation sites in one individual peptide. Indeed, upon manual inspection of our data, we observed peptides containing up to 11 HexNAcs (Figure 2). Of note, one peptide with an extended *O*-glycan on the CD45 tryptic peptide LNPTPGSNAISDVPGER (HexNAc<sub>2</sub>Hex<sub>1</sub>NeuAc<sub>1</sub>) was found using Byonic, which, considering the Cosmc mutation in Jurkat cells, could correspond to a core 6 *O*-glycan, a specific MGL binder (GlcNAc $\beta$ 1-6GalNAc $\alpha$ Ser/Thr) [38]. In addition to CD43 and CD45, our database searches confirmed the presence of the Tn antigen on another 11 proteins (Table 2). We confirmed the Tn antigen on all of these peptides by manual interpretation of the data. As an example, two glycopeptides from EVI2B and TREML2, respectively, are shown in Figure 3. Between the peptides presented in Table 2, one peptide (GLFIPFSVSSVTHK) with three HexNAcs was identified from P-selectin glycoprotein ligand 1 (SELPLG), which was not identified as a specific binder in the proteomics data (Table 1). Inspection of the data showed that, unexpectedly, it was filtered out due to the fact that normal tryptic peptides of SELPLG were observed in only one of the three MGL pull-downs. Terminal *N*-acetylgalactosamines can also be part of *N*-glycans, for example, as part of the LacdiNAc epitope. Hence, we also searched our data for the presence of *N*-

glycopeptides using Byonic. Only for CD45 presented in Table 1 we could identify *N*-glycopeptides (on N234, 278, 337, 380, 421, and 470, respectively), but none of these appear to contain a terminal *N*-acetylgalactosamine, as judged on the basis of the glycan compositions as well as the absence of the LacdiNAc (GalNAc $\beta$ 1-4GlcNAc $\beta$ 1) marker ion at *m/z* 407.166 (Table S2).

**Table 2: Identified O-glycopeptides from MGL binding cell surface proteins in Jurkat cells.** O-glycosylated peptides were annotated using three different approaches (Maxquant, Byonic, manual annotation). # HexNAC: number of N-acetylhexoseamine carried by each glycopeptide; N: N-acetylhexoseamine; H: Hexose; S: N-acetylneuraminic acid. Ox: oxidation of methionine.

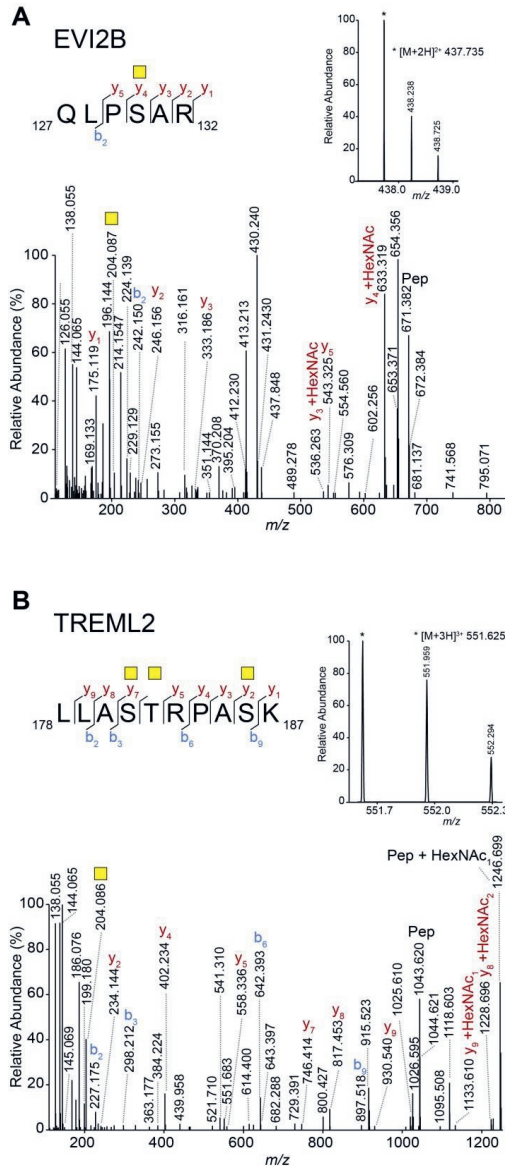
Protein Name	Gene Name	Uniprot Entry	Glycopeptide Sequence	MaxQuant # HexNAC	Byonic # HexNAC	Manual # HexNAC
Lysosome-associated membrane glycoprotein 3	LAMP3	E7ETP9	QAPHQTLAAR	1	1	1
			DYSQPTAAATVQDIK		3	3
Leukosialin (CD43)	SPN	P16150	MYTTSITSDPK	3,4	3	3,4,5
			M <sub>ox</sub> YTTSITSDPK			3,4
			STTAVQTPSTSGEPLVSTSEPLSSK			9,10,11
Podocalyxin	PODXL	O00592	TPSPTVAHESNWAK	2	2,3	2,3
			ANEILASVK		1	1
			CEDLETQTQSEK		1	1
Protein EVI2B	EVI2B	P34910	QLPSAR		1	1
			STPGFILDTSNK	4		4,5
			STPGFILDTSNKNKTPQK			4,5
			QITVHNPSIQPTSTVKNSPR			7
P-selectin glycoprotein ligand 1	SELPLG	Q14242	GLFIPFSVSSVTHK	3	3	3

Receptor-type tyrosine-protein phosphatase C (CD45)	PTPRC	P08575	LNPTPGSNAISDVPGER	2	1,2,3	2,3	
			LNPTPGSNAISDVPGER				N2H1S1
			STASTFTDPVSPPLTTLSLAHSSAALPAR				10,11
Semaphorin-4D	SEMA4D	Q92854	NGSAAM <sub>ox</sub> CHFTTK	1	1,2	1	
			VVPKPVVAPTLVVQTEGSR				1,2
			IVINTVPQLHSEK				1,2
Trem-like transcript 2 protein	TREM12	Q5T2D2	LIASTRPASK	3	2,3	3	
			NIPFTHLDNLIK				2
Tumor necrosis factor receptor superfamily member 8	TNFRSF8	P28908	LAQEAASKLTR	2	2	2	
Transferrin receptor protein 1	TFRC	P02786	LAGTESPVREEPGEDFPAAR		1	1	
Voltage-dependent calcium channel subunit alpha-2/delta-4	CACNA2 D4	Q7Z3S7	YKDVESLTK		2	2	
Glycosyltransferase 8 domain-containing protein 1	GLT8D1	Q68CQ7	YTEISNIK		2	2	
Volume-regulated anion channel subunit LRRC8D	LRRC8D	Q7L1W4	TDFALPNQEAK		1	1	



**Figure 2: Highly O-glycosylated peptides of CD43 and CD45.** MS/MS spectra of glycopeptides from CD43 carrying 10 and 11 GalNAcs and CD45 carrying 11 GalNAcs. Left panels: HCD MS/MS spectra. *b* and *y*-ions are representing fragments without the GalNAc, unless indicated otherwise. The insert shows the MS spectrum of the corresponding precursor with charge state and *m/z* value. Right panels: CID MS/MS spectra. Yellow square: GalNAc.





**Figure 3: Tn-bearing O-glycosylated peptides of EVI2B and TREML2.** A: Manually assigned HCD MS/MS spectrum of the tryptic peptide QLPSAR from EVI2B carrying 1 GalNAc. The insert shows the MS spectrum of the precursor ion at  $m/z$  437.735  $[M + 2H]^{2+}$ . B: Manually assigned HCD MS/MS spectrum of the tryptic peptide LAQAASKLTR from TREML2 carrying 3 GalNAcs. The insert shows the MS spectrum of the precursor ion at  $m/z$  551.950  $[M + 3H]^{3+}$ . Yellow square: GalNAc. b and y-ions are representing fragments without the GalNAc, unless indicated otherwise.

## Discussion

We developed a workflow for the identification of MGL ligands, which resulted in the identification of 17 cell-surface proteins of Jurkat cells that bind to MGL. For most of these proteins, we confirmed the specificity of the interaction with MGL through the identification of *O*-glycopeptides carrying the Tn antigen.

The glycopeptide, mediating the interaction with MGL, could not be identified for all MGL ligands. These peptides may have been missed due to insufficient sensitivity or low MS/MS spectrum quality, which would hamper identification by both search algorithms as well as manual inspection. Although in the past years, we and others have made considerable progress in the mass spectrometry-based identification of glycopeptides [65-70], several issues related to their fragmentation behaviour still often impede straightforward (data) analysis, even though the Tn-antigen represents a relatively simple post-translational modification. We combined multiple strategies for data acquisition and analysis to maximize our glycopeptide identification. Other methods, such as electron-transfer dissociation (ETD), have also been valuable tools for the identification of glycopeptides [66], especially in cases where the localisation of the GalNAc(s) was ambiguous. However, for this study we were primarily interested in identifying MGL ligands, and not necessarily the site-specific glycan assignment. Moreover, for several peptides we had full occupancy of all serine and threonine residues present in the identified glycopeptide.

From the 17 potential MGL-binding proteins (Table 1), some could have been obtained through interaction with a genuine MGL ligand. For example, protein tyrosine phosphatase receptor type C-associated protein (PTPRCAP), for which we could not find the corresponding glycopeptide, was captured during the MGL pull-down probably because it is binding partner of CD45 [71]. Similarly, for only one of the four LRRC8 subunits (LRRC8D) that we identified could a specific *O*-glycopeptide be found. Because it is known that the LRRC8 subunits form heterodimers [72, 73], it is conceivable that the whole LRRC8 complex was captured.

As expected, CD43 and CD45 were among the top proteins carrying the Tn-antigen in Jurkat cells. A recent study presented a workflow where a combination of enzymatic and chemical methods was used to selectively tag terminal GalNAc, and not GlcNAc residues [74], allowing enrichment of the corresponding glycopeptides. To study proteins carrying the Tn-antigen, this method was applied to total cell lysates of Jurkat cells. As a result, a total of 97 proteins

harbouring the Tn-antigen were identified, of which 27 were consistently found in all three independent biological replicates. From the total of 97 proteins, 11 were transmembrane signalling molecules, including CD45, Semaphorin-4D and TNFR8, which were also identified in our study. Some other Tn-antigen bearing glycoproteins were not found in our experiments. The observed differences may relate to the different experimental approaches since several Tn-bearing glycoproteins were clearly identified in our experiments, such as CD43, EVI2B and TREML2, but not in the above mentioned study [74]. On the contrary, it could be due to the fact that for MGL binding the presence of Tn is necessary but not sufficient [39, 75]. For example, all CD45 isoforms (CD45ABC/AB/BC/B) except for one (CD45RO) bind MGL. Unlike the others, which are highly glycosylated, CD45RO has only two *O*-linked glycan epitopes.

Although it is known that MGL can have immune-modulatory activities, the specific proteins involved in the cellular response elicited by MGL are largely unknown. Tolerogenic antigen presenting cells (APCs) express high levels of MGL on the cell surface, and the binding of MGL to CD45 suppresses TCR-mediated T-cell activation. This results in an anti-inflammatory response characterized by the reduced production of pro-inflammatory cytokines and lower proliferation of T-cells and induction of cell death in Jurkat [39].

For the new MGL ligands identified in this study the outcome of the interaction with MGL remains to be determined. Semaphorin 4D, also known as CD100, is a homodimeric transmembrane protein of 150-kDa that is highly expressed in secondary lymphoid organs and constitutively on naïve T-cells. It binds its ligand CD72, a C-type lectin expressed on the surface of antigen presenting cells, such as B cells and DCs [76], but also macrophages and some subpopulation of T-cells [77]. Our data show that, at least in tumor cells with aberrant glycosylation, Semaphorin 4D also binds to MGL.

For several of the newly identified MGL ligands, literature provides limited information about their function. For example, although KIAA0319L has been demonstrated to be an essential receptor for adeno-associated virus infection, the cellular function, especially in immune responses, is unknown. However, our data support previous findings which suggested high levels of mucin type *O*-glycosylation on KIAA0319L [78]. Our data also provide evidence for *O*-glycosylation on proteins that hitherto were unknown to be *O*-glycosylated, such as EVI2B and TREML2. Although the cellular function of these proteins is largely unknown, it has been demonstrated that EVI2B is the target for the transcription factor CCAAT/enhance binding

protein alpha (C/EBP $\alpha$ ), [79] and TREML2 binding by CD276 leads to enhanced T-cell responses [80]. Altogether, our data warrant further exploration of the functional implications of the interaction of MGL with the newly identified ligands. Notwithstanding the importance of this, some responses elicited by MGL might also be the result of the interaction of MGL with glycolipids [38], which was not the topic of our study.

The high level of Tn antigen is not restricted to leukemias but is frequently observed in other tumors as well. The best characterized ligand carrying Tn or sialyl-Tn (sTn, Neu5Ac $\alpha$ 2,6-GalNAc-O-Ser/Thr) in epithelial cells is the glycoprotein MUC1. MUC1-derived glycopeptides bind to MGL on DCs and induce the activation of the extracellular signal-regulated kinases 1 and 2 (ERK1,2) and the nuclear factor- $\kappa$ B (NF- $\kappa$ B) pathways [81]. In another study, it was demonstrated that these pathways are crucial for IL-10 production, which regulates the DC maturation phenotype [82].

The higher level of Tn in other cancers is not necessarily due to *Cosmc* mutations, as in the Jurkat T-cell used in our study, but may also be the result of other genetic alterations. For example, in colorectal cancer higher levels of several GALNTs have recently been linked to the BRAF<sup>V600E</sup> mutation [83]. A positive correlation between this oncogenic mutation and MGL ligands on tumor cells was previously identified [45]. Hence, it will be interesting to study the proteins carrying MGL in these tumor cells as well.

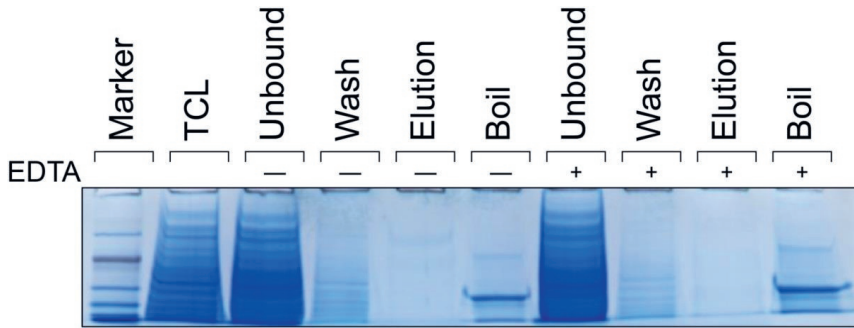
In conclusion, here we provide an optimized method to capture MGL-binding proteins followed by glycoproteomic analysis. The application of this procedure on the Jurkat cell line provides important novel insights into previously unknown MGL ligands. However, further investigations should evaluate the functional immune responses triggered by MGL-specific recognition of those proteins, as has been previously reported for the already known interaction partner CD45 by van Vliet et al. [39]. This will provide a deeper understanding of the MGL involvement in cancer progression and the glycan-specific immune responses mediated by this lectin.

### **Acknowledgements**

This work was supported by the European Commissions Horizon 2020 programme “GlyCoCan” project, grant number 676421, and by the research programme Investment Grant NWO Medium with project number 91116004, which is (partially) financed by ZonMw.

## Supporting Information

Supporting Information is available free of charge via  
<https://pubs.acs.org/doi/suppl/10.1021/acs.jproteome.8b00796>



**Figure S1:** SDS-PAGE of MGL pull-down. TCL: total cell lysate; Unbound: unbound fraction; Wash: washing step with lysis buffer; Elution: eluted proteins with EDTA; Boil: boiling in SDS-PAGE sample buffer after EDTA elution. The picture is a representative of the three replicates performed.

**Table S1:** MGL-binding proteins from Jurkat cells. Mol. Weight: molecular weight; ER: Endoplasmic Reticulum; PM: Plasma Membrane; Exp: experiment number; Peptides: number of unique peptides per protein; Neg CTRL: negative control, MGL-Fc pull-down in presence of EDTA [100mM]; MGL: MGL-Fc pull-down; plasma membrane proteins are highlighted in blue.

**Table S2:** N-glycosylated peptides of CD45. Mox: oxidated methionine; N: N-acetyl-D-hexosamine; H: hexose; F: fucose; S: sialic acid; Exp: number indicates in which of the three replicates it was found.