HAL open science

Characterization of Macrophage Galactose-type Lectin (MGL) ligands in colorectal cancer cell lines

Martina Pirro, Yoann Rombouts, Alexandre Stella, Olivier Neyrolles, Odile Burlet-Schiltz, Sandra van Vliet, Arnoud de Ru, Yassene Mohammed, Manfred Wuhrer, Peter van Veelen, et al.

To cite this version:

Martina Pirro, Yoann Rombouts, Alexandre Stella, Olivier Neyrolles, Odile Burlet-Schiltz, et al.. Characterization of Macrophage Galactose-type Lectin (MGL) ligands in colorectal cancer cell lines. Biochimica et Biophysica Acta (BBA) - General Subjects, 2020, 1864 (4), pp.129513. 10.1016/j.bbagen.2020.129513 . hal-03000979

HAL Id: hal-03000979

https://hal.science/hal-03000979

Submitted on 16 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

BBA - General Subjects 1864 (2020) 129513

Contents lists available at ScienceDirect

BBA - General Subjects

journal homepage: www.elsevier.com/locate/bbagen

Characterization of Macrophage Galactose-type Lectin (MGL) ligands in colorectal cancer cell lines

Martina Pirro[a], Yoann Rombouts[b], Alexandre Stella[b], Olivier Neyrolles[b], Odile Burlet-Schiltz[b], Sandra J. van Vliet[c], Arnoud H. de Ru[a], Yassene Mohammed[a], Manfred Wuhrer[a], Peter A. van Veelen[a], Paul J. Hensbergen[a,*]

[a] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands
[b] Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, Toulouse, France.
[c] Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Molecular Cell Biology and Immunology, Cancer Center Amsterdam, Amsterdam Infection & Immunity Institute, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Glycoproteomics
C-type lectin
Colorectal cancer
LacdiNAc
Tn antigen

ABSTRACT

Background: The Ca^{2+}-dependent C-type lectin receptor Macrophage Galactose-type Lectin (MGL) is highly expressed by tolerogenic dendritic cells (DC) and macrophages. MGL exhibits a high binding specificity for terminal alpha- and beta-linked GalNAc residues found in Tn, sTn and LacdiNAc antigens. These glycan epitopes are often overexpressed in colorectal cancer (CRC), and, as such, MGL can be used to discriminate tumor from the corresponding healthy tissues. Moreover, the high expression of MGL ligands is associated with poor disease-free survival in stage III of CRC tumors. Nonetheless, the glycoproteins expressed by tumor cells that are recognized by MGL have hitherto remained elusive.
Methods: Using a panel of three CRC cell lines (HCT116, HT29 and LS174T), recapitulating CRC diversity, we performed FACS staining and pull-down assays using a recombinant soluble form of MGL (and a mutant MGL as control) combined with mass spectrometry-based (glyco)proteomics.
Results: HCT116 and HT29, but not LS174T, are high MGL-binding CRC cell lines. On these cells, the major cell surface binding proteins are receptors (e.g. MET, PTK7, SORL1, PTPRF) and integrins (ITGB1, ITGA3). From these proteins, several *N*- and/or *O*-glycopeptides were identified, of which some carried either a LacdiNAc or Tn epitope.
Conclusions: We have identified cell surface MGL-ligands on CRC cell lines.
General significance: Advances in (glyco)proteomics have led to identification of candidate key mediators of immune-evasion and tumor growth in CRC.

1. Introduction

Many studies have highlighted the importance of glycosylation in cancer development and progression [1,2] and altered glycosylation on the cell surface is a prominent hallmark of tumor cells. Among tumor-associated carbohydrate antigens, the most common alterations occur in the mucin-type *O*-glycans, where the abortive elongation results in Tn (GalNAcα1-Ser/Thr) and sialylated sTn (NeuAcα2,6GalNAcα1-Ser/Thr) [3]. These glycan structures are associated with poor prognosis and metastasis in many epithelial cancers [4]. Another glycobiomarker expressed by a large variety of human cancers, such as lung [5], ovary [6] and colon [7,8], is LacdiNAc (GalNAcβ1,4GlcNAcβ1-) on both *N*- and *O*-glycans [9]. Elevated levels of LacdiNAc have been found on glycoproteins during tumor progression [10].

The changes in glycosylation on the cell surface can be recognized by glycan binding proteins, especially lectins. Lectins are classified into families based on their carbohydrate–recognition domains (CRDs), and on their glycan binding specificities that can be determined through the use of glycan arrays [10]. Within the large variety of lectins, the macrophage galactose-type lectin (MGL) is a calcium-dependent C-type lectin expressed by human antigen presenting cells, such as tolerogenic dendritic cells (DCs) and macrophages [11]. MGL has an exclusive specificity for binding to terminal α- and β-linked GalNAc residues [12–14], namely Tn, sTn and LacdiNAc. From a functional point of view, several lines of evidence suggest that the binding of MGL to its ligands dampen the adaptive immune response [11]. First of all, the

* Corresponding author at: Center for Proteomics and Metabolomics, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands.
E-mail address: P.J.Hensbergen@lumc.nl (P.J. Hensbergen).

https://doi.org/10.1016/j.bbagen.2020.129513
Received 23 September 2019; Received in revised form 20 December 2019; Accepted 2 January 2020
Available online 03 January 2020
0304-4165/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

binding of Tn to MGL expressed on DCs, in conjunction with Toll like receptor (TLR) triggering, results in a tolerogenic phenotype, with increased secretion of IL-10 and TNF-α [15]. Moreover, MGL binding to CD45, a known Tn-bearing ligand expressed on effector T cells, combined with T cell receptor activation, suppresses T cell activation, resulting in reduced pro-inflammatory cytokine production and proliferation [11]. Hence, the interaction of MGL with its ligands can have effects both on the cell expressing the lectin as well as on the cell carrying the MGL ligands.

It has been previously demonstrated that MGL binding is a powerful tool to discriminate healthy tissue from that of colorectal cancer (CRC) [16], which is the third most common cancer and the second most common cause of cancer mortality worldwide (http://gco.iarc.fr/today/fact-sheets-cancers). Moreover, high expression of the MGL ligands in the tissues of patients with stage III CRC has been associated with lower disease-free survival [16], thus establishing MGL binding as an independent prognostic marker. Furthermore, MGL ligands expression has a positive correlation with the oncogenic $BRAF^{V600E}$ mutation in CRC [16,17]. More recently, high expression levels of MGL ligands was correlated with worse prognosis not only in CRC but also in cervical squamous cell carcinoma and adenosquamous carcinoma [18].

Notwithstanding the link between MGL and some cancer types, the identity of the MGL ligands expressed by tumor cells have hitherto not been explored in CRC. Therefore, the aim of this study was to characterize the MGL-binding proteins expressed by CRC cell lines (HCT116, HT29 and LS174T) using a combination of flow cytometry, pull-down experiments and (glyco)proteomics.

2. Material and methods

2.1. Cell lines culture and lysis

To recapitulate tumor diversity, we selected three colorectal cancer cell lines with different characteristics: HT29, which contains the $BRAF^{V600E}$ mutation, is classified as colon-like subtype and grade I adenocarcinoma [19] and is well known for the high expression of MGL ligands [16]; LS174T, with same classification as HT29 but lacking the *BRAF* mutation; HCT116, belonging to an undifferentiated subtype and originating from a grade IV carcinoma [20] without the *BRAF* mutation.

HCT116 and HT29 were provided by the Department of Surgery of the Leiden University Medical Center (Leiden, The Netherland), whereas LS174T was obtained from S.J. van Vliet, Amsterdam UMC (Amsterdam, the Netherlands). Cell line authentication was performed using short-tandem repeat (STR) profiling at the forensic laboratory for DNA-research (ISO 17025) and all cell lines matched for 100% with the known profile [21]. Cells were cultivated in RPMI-1640 (for HCT116 and HT29) or in DMEM (for LS174T) medium containing L-glutamine, 10% fetal bovine serum (FBS) (Invitrogen) and streptomycin/penicillin (Sigma-Aldrich) at 5% CO_2 and 37 °C. Cells were maintained till approximately 80% confluence under sterile conditions. For harvesting, cells were washed twice with 1× PBS and incubated for approximately 5 min in 1× trypsin/EDTA solution in 1× PBS, whose activity was inhibited by the addition of serum containing medium following visual cell detachment. Cells were subsequently harvested and counted using the CountessTM Automated Cell Counter (Invitrogen, Paisley, UK) based on trypan blue staining. Aliquots of 2×10^7 viable cells were washed with 1× PBS and centrifuged at 1500 rpm to obtain cell pellets. Cell pellets were stored at −20 °C until use. Before pull-down experiments, cells were lysed as described before [22], for 20 min on ice in lysis buffer (10 mM triethanolamine pH 8.2, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$ and 1% (volume/volume) Triton X-100, containing EDTA-free protease inhibitor (Roche Diagnostics)). Protein quantification was performed using BCA assay (BCA Protein Assay Kit, Pierce™), following the manufacturer's instructions.

2.2. Lectins and reagents

Chimeric MGL-Fc and mutant-MGL-Fc (MGL^{short} H259T-Fc, obtained by site-directed mutagenesis, as described previously [23]) were provided by S. J. van Vliet (Amsterdam UMC, Amsterdam, the Netherlands).

2.3. Flow cytometry

To determine the expression of MGL ligands, 1×10^5 cells per condition were first washed with TSM buffer (20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$) supplemented with 0.5% BSA. Subsequently, cells were incubated with 10 μg/ml MGL-Fc (with or without pre-incubation of MGL-Fc with 100 mM EDTA for 15 min), or mutant-MGL-Fc, for 30 min at 37° C, followed by staining with FITC-labelled rabbit anti-human IgG (Agilent) for 30 min at 4° C. Cells were then fixed with 1% para-formaldehyde and stored at 4° C until analysis on a BD LSRFortessa™ (BD Bioscience) cell analyser with 6000 events per condition. Data analysis was performed with the FlowJo V10 software.

2.4. Pull-down assay

MGL ligands were pulled down from 1 mg of protein extracts using either 2 μg of chimeric MGL-Fc or mutant-MGL-Fc, coupled to 50 μl Dynabeads protein G (Invitrogen) as previously described [22]. Following washing, the elution of specific ligands was performed using 100 mM EDTA.

2.5. SDS-PAGE and NanoLC-MS/MS analysis

For sample clean-up, a short SDS-PAGE run (NuPAGE™ 4–12% Bis-Tris Protein Gels, Thermo Fisher Scientific) of the samples obtained from the MGL pull-down was performed. Gels were stained with SimplyBlue™ SafeStain (Invitrogen) for 3 h and washed with distilled water. Bands corresponding to the whole lane were cut from the gel, and the proteins were then subjected to reduction with dithiothreitol (10 mM), alkylation with iodoacetamide (50 mM) and in-gel trypsin digestion with Promega™ Sequencing Grade Modified Trypsin (Thermo Fisher Scientific), manually or using a Proteineer DP digestion robot (Bruker) [22].

Peptides were analysed by nanoLC-MS/MS using an UltiMate 3000 RSLCnano system (Dionex, Amsterdam, The Netherlands) coupled to a Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA). Peptides were resuspended in 17 μl of 5% acetonitrile, 0.05% TFA. 5 μl of each sample were loaded onto a C-18 precolumn (300 μm inner diameter x 5 mm, Dionex) at 20 μl/min in 5% acetonitrile, 0.05% TFA. After 5 min of desalting, the precolumn was switched online with the analytical C-18 column (75 μm inner diameter × 50 cm; in-house packed with Reprosil C18) equilibrated in 95% solvent A (5% acetonitrile, 0.2% formic acid) and 5% solvent B (80% acetonitrile, 0.2% formic acid). The peptides were eluted using a 5% to 50% gradient of solvent B at 300 nl/min flow rate for 160 min. The mass spectrometer was operated in data-dependent acquisition mode with the XCalibur software. Survey MS scans were acquired in the Orbitrap on the 400–1500 *m/z* range with a resolution of 120,000 (*m/z* 400). The most intense ions per survey scan were selected for higher-energy collisional dissociation fragmentation (time between Master scans: 3 s), the resulting fragments were analysed at a resolution of 30,000 in the Orbitrap. Dynamic exclusion was employed within 60 s to prevent repetitive selection of the same peptide.

For the characterization of glycopeptides, tryptic peptides were analysed on an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific) as described before [22]. MS/MS spectra were acquired in data-dependent mode (top-10) with collision energy at 32 V and recording the MS/MS spectrum in the Orbitrap. During

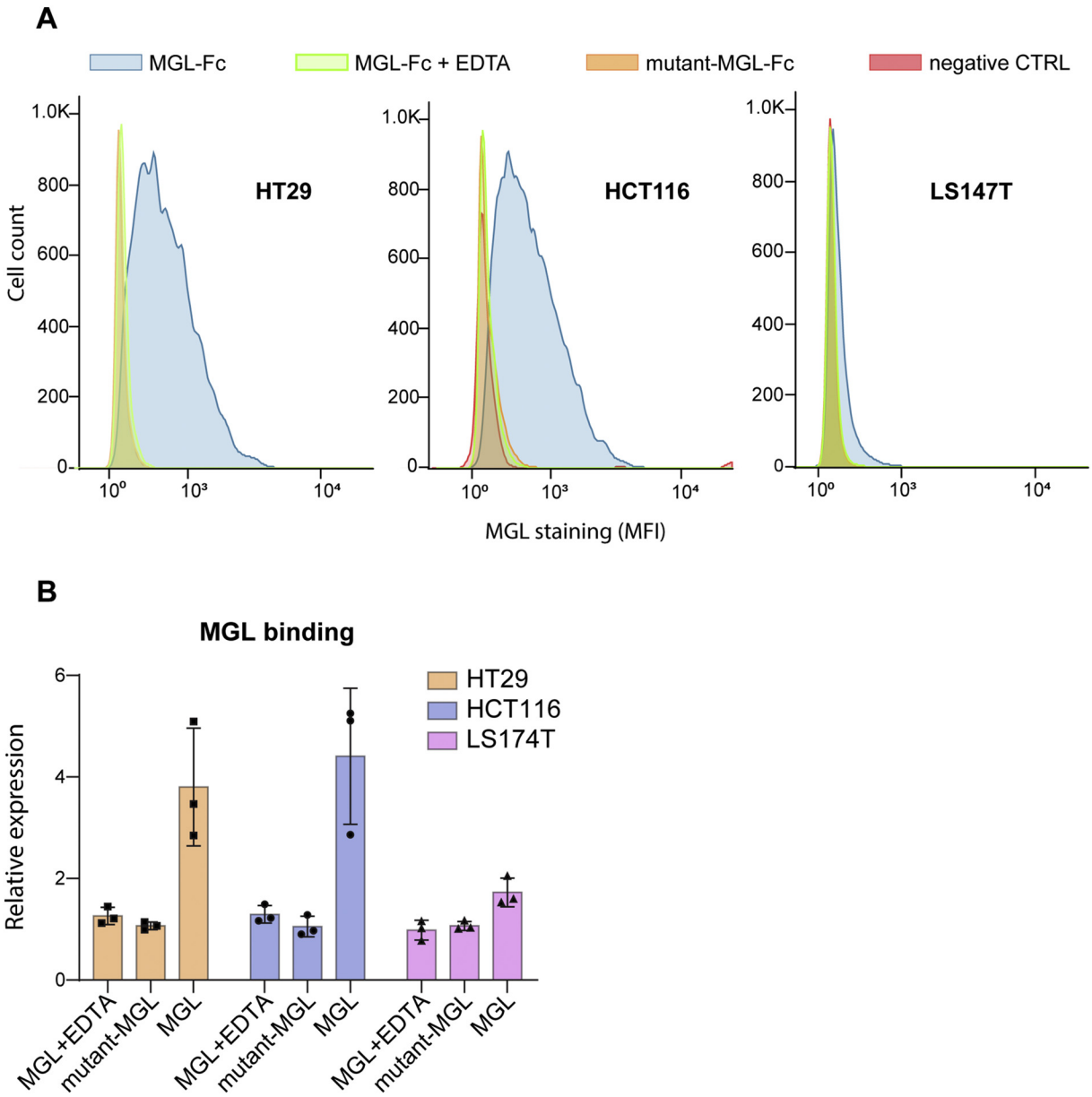

Fig. 1. Expression of MGL ligands on a panel of CRC cell lines. A) Binding of the recombinant soluble MGL-Fc to HT29, HCT116 and LS174T CRC cell lines (blue histograms) was analysed by flow cytometry. MGL-Fc in the presence of EDTA (green), mutant-MGL-Fc (orange) or no MGL (red, negative CTRL) were used as controls. The histograms are representative of three independent experiments. Fc: Fragment crystallisable region; MFI: mean fluorescence intensity; CTRL: control. B) Relative expression of the MGL ligands on HT29, HCT116 and LS174T CRC cells. Shown are mean fluorescent intensities (n = 3, ± S.D.) normalized to staining with the secondary antibody only.

acquisition of MS/MS spectra, the HexNAc oxonium ion at *m/z* 204.087 was set as Product Ion Trigger in order to execute three additional MS/MS scans of the same precursor with HCD collision energies 32, 37 and 41 V, respectively, and with CID at 35 V [22].

2.6. Data analysis

For protein identification, raw data was converted to mzXML using Proteowizard. Peptide and protein identification as well as the after statistical validation were performed in Trans Proteomics Pipeline version 5.1.0 using included software pipeline: X! Tandem Jackhammer TPP (2013.06.15.1-LabKey, Insilicos, ISB) search engine, PeptideProphet and ProteinProphet. The parameters were set as follows: parent mass error of 10 ppm, fragment mass error of 0.04 Da, carbamidomethyl (Cys) and oxidation (Met) as fixed and variable modifications respectively. All results were filtered for FDR threshold of 1% as well as a minimum of two peptides per protein. Data extraction and table generation was done using R version 3.4.4.

In order to select for specific MGL-binding proteins, we further filtered for proteins that were identified with peptide at least in one of the technical replicate LC-MS/MS runs from one MGL pull-down and at least in two out of the three biological replicate MGL pull-downs. Moreover, the selection required that the protein was never found in the corresponding pull-downs with the mutant-MGL (negative controls). From the final list of MGL-binding proteins, we subsequently selected the cell surface proteins based on Gene Ontology (GO) subcellular location, selecting for “cell surface”, “plasma membrane”, “integral component” and/or “external side”.

For the identification of glycopeptides, MS/MS spectra containing the specific HexNAc oxonium ions at *m/z* 204.087 (HexNAc,

[C8H14NO5]$^+$) and 186.076 (HexNAc-H2O, [C8H12NO4]$^+$) were selected from the raw data and written to a .mgf file (in house software). The selected spectra were searched for *N*- and *O*-glycopeptides using Byonic version 2.13.2 (proteinmetrics.com), with default settings. Carbamidomethyl (Cys) was set as fixed modification, whereas oxidation (Met) as variable one. "*N*-glycan 309 mammalian no sodium" or "*O*-glycans 78 mammalian" databases were used for glycosylation annotation. A Byonic score higher than 200 was considered stringent for further glycopeptides selection. All glycopeptide MS/MS spectra carrying potential MGL-binding glycan epitopes were manually validated using Xcalibur (Thermo) for the spectrum visualization and Protein Prospector (prospector.ucsf.edu) for the calculation of theoretical peptide fragment masses.

2.7. Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [24] partner repository with the dataset identifier PXD016726.

3. Results

3.1. MGL binding to CRC cell lines

First, we examined the binding of MGL-Fc to our panel of cell lines using flow cytometry. As expected [23], HT29 exhibited a high expression of cell surface ligands recognized by MGL. A comparable intensity was observed for HCT116 cells. In contrast, very low staining was found for LS174T cells. For all three cell lines, the specific binding to MGL-Fc could be blocked by pre-incubation of MGL-Fc with EDTA, confirming the calcium-dependent interaction. In the Carbohydrate Recognition Domain (CRD) of MGL, there is a secondary binding site, involving histidine-259, that contributes to the interaction of the lectin with the glycan as well as the peptide backbone [23]. Importantly, incubation of cells with MGL in which this histidine was replaced by a threonine (mutant-MGL-Fc) showed that the binding activity of mutant-MGL was completely abrogated in the cell lines used (Fig. 1).

3.2. Identification of MGL-binding proteins from CRC cell lines

In order to identify the MGL-binding proteins expressed by the three CRC cell lines (HT29, HCT116 and LS174T), we performed pull-down experiments in triplicate with MGL-Fc coupled to magnetic protein G beads (Fig. 2A). Based on our FACS results (Fig. 1), we additionally performed triplicate pull-down experiments with the mutant-MGL-Fc (negative control). Since the binding to the CRD of MGL is calcium-dependent, captured proteins were eluted using EDTA and then subjected to a short SDS-PAGE run (Fig. S1), in-gel digested with trypsin and analysed by LC-MS/MS. Overall, these experiments resulted in the identification of 854 proteins (Fig. 2B, Table S1). In order to select only for specific MGL-binding proteins, we filtered the data of each cell line for proteins that were present in at least two of the pull-downs with MGL-Fc, but never in the negative control (mutant-MGL-Fc). Using this selection criteria, we identified 281 and 246 MGL-binding proteins from HCT116 and HT29, respectively (Fig. 2B, Table S1). As expected, based on the low MGL-Fc staining observed by flow cytometry (Fig. 1), only 12 MGL-binding proteins were found from the LS174T cells (Fig. 2B, Table S1).

Since under physiological conditions MGL on DCs and macrophages binds to cell surface glycoproteins, we further selected only for those proteins that were annotated as "plasma membrane" and/or "cell surface", and localized on the "external side" and/or "integral component" of the membrane, using an automated Gene Ontology (GO) survey. Ultimately, this resulted in the selection of 45, 56 and 2 cell surface MGL-binding protein(s) in HCT116, HT29 and LS174T respectively, (Fig. 2C, Table S2). Next to 18 shared MGL-binding proteins, HCT116 and HT29 showed 27 and 38 unique binding proteins, respectively (Fig. 2C). The top 25 glycoproteins from HCT116 and HT29 (based on the peptide counts) are shown in Table 1. Among these major MGL-binding proteins (Table 1), different classes of proteins were found, such as cell surface signaling receptors (e.g. MET, PTK7, SORL1, PTPRF) and integrins (ITGB1, ITGA3) but also proteins whose function is hitherto less well defined (e.g. IGSF3).

3.3. LacdiNAc and Tn antigens are carried by cell surface MGL ligands in CRC cell lines

To obtain information on the glycan structures that were present on the cell surface proteins captured by MGL (Table S2), and whether these could explain the binding to MGL, we performed additional MGL-Fc pull-down experiments with HCT116 and HT29 cells. To identify intact tryptic glycopeptides, we used a glycoproteomics data acquisition and analysis workflow that we recently optimized for this purpose [22]. Briefly, additional MS/MS experiments were performed once the characteristic HexNAc oxonium ion at *m/z* 204.087 was found in an MS/MS spectrum. For these acquisitions, different HCD collision energies were used. Moreover, an ion trap CID spectrum was recorded. All this helped for the confident identification of glycopeptides. The results revealed that five of the proteins listed in Table 1 (GLG1, SORL1, PTK7, ITGA3 and GOLM1) carried a LacdiNAc epitope on *N*-glycans in HCT116 and/or HT29 cells (Table S3). As examples, Fig. 3 shows the MS/MS spectra of the corresponding *N*-glycopeptides for ITGA3 and PTK7 from HCT116 cells. The glycan on these peptides consists of a $HexNAc_5Hex_4NeuAc_1$ structure. In both the HCD (higher panels) and CID (lower panels) spectra, the presence of the LacdiNAc structure is indicated by the specific oxonium ion at *m/z* 407.166. The fragmentation spectra revealed a different position of the sialic acid on the two glycopeptides. Whereas for ITGA3, sialic acid was located on the galactose, the MS/MS spectra from the PTK7 glycopeptide revealed a sialylated LacdiNAc, characterized by the oxonium ions at *m/z* 495.182 and 698.261, (Fig. 3).

In addition, one plasma membrane MGL-ligand from HCT116, dystroglycan (DAG1, Table 1), showed glycopeptides carrying either two (Fig. 4) or three Tn-antigens. Of note, we observed a glycopeptide belonging to Agrin (AGRN) from HT29 with a Tn-antigen, but Agrin was not selected as a MGL-binding protein in the overall proteomics experiments in this cell line, because it was also identified in the samples using mutant-MGL.

Our glycoproteomic analyses also revealed several other glycopeptides from cell surface MGL-binding proteins as shown in Table S2 but none of these was found to carry a glycan structure that could explain the binding to MGL (Table S3). For example, MET, with 11 potential N-glycosylation sites in its extracellular domain (Table S1), was found with a glycopeptide carrying a sialylated complex *N*-glycan ($HexNAc_4Hex_5NeuAc_1$) at Asn-785 in HCT116 cells. The same *N*-glycan structure was also found on glycopeptides from other MGL-binding proteins such as IGSF3, ITGA3, PTK7 and SORL1. In addition, some other glycan structures were observed on cell surface binding proteins (e.g. high mannose glycans and (di-) sialyl-T antigen, Table S3).

4. Discussion

In this study, we characterized three different CRC cell lines (HCT116, HT29 and LS174T) for their binding to MGL and subsequently identified the cell surface MGL ligands using a (glyco)proteomics approach. From the major cell surface MGL-binding proteins found in our study, around 50% is shared between the two cell lines that showed high staining for MGL ligands on the cell surface (i.e. HT29 and HCT116). With our glycopeptide analyses, we observed both LacdiNAc and Tn epitopes on the cell surface MGL-binding proteins, representing known target ligands of MGL.

Our data show that five of the main cell surface MGL-binding

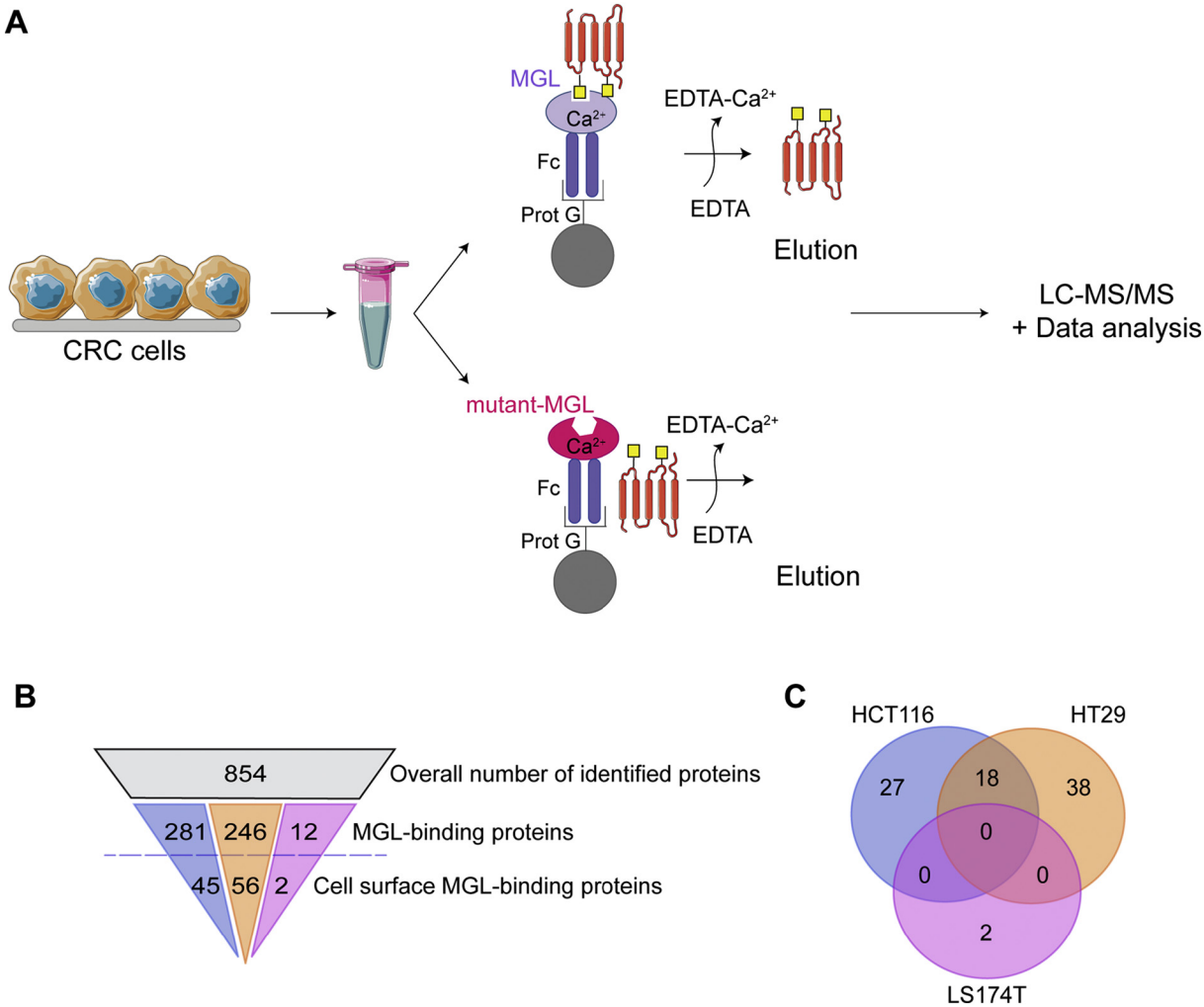

Fig. 2. Pull-down workflow and LC-MS/MS analysis for the identification of cell surface MGL-binding proteins in CRC cell lines. A) Three independent total lysates from the three CRC cell lines (HCT116, HT29 and LS174T) were incubated with either MGL-Fc or mutant-MGL-Fc as a negative control. Captured proteins were analysed by LC-MS/MS. CRC: colorectal cancer; Fc: Fragment crystallisable region; Prot G: protein G. B) Filtering of identified proteins by mass spectrometry (grey) to overall MGL-binding proteins and predicted cell surface MGL-binders in HCT116 (blue), HT29 (orange) and LS174T (pink). C) The Venn diagram shows the number of cell surface MGL-binding proteins unique or shared by the three CRC cell lines, after filtering (see also Table S2).

proteins carry a LacdiNAc motif. LacdiNAc biosynthesis is catalyzed by β4-*N*-acetylgalactosaminyltransferase 3 (β4GalNAc-T3) in stomach and colon [25], which is often overexpressed in primary colon tumors [8]. Moreover, immunohistochemistry of different stages of CRC tissues showed a correlation of β4GalNAc-T3 overexpression with advanced stages (III and IV) and with higher metastatic behaviour [7]. However, the LacdiNAc modification of several MGL-binding proteins found in our study is probably not related to overexpression of β4GalNAc-T3, since its transcriptional levels in HT29 and HCT116 cells are comparable to that in the low MGL-binding, LS174T, cell line [19]. Nevertheless, the protein expression or its activity might still be different in the three cell lines and require further investigations. In our data, we also observed a sialylated LacdiNAc on the PTK7 glycopeptide. From glycan array studies it is known that sialylation of the LacdiNAc epitope is compatible with MGL binding [23]. In fact, since the sialic acid is most likely added in a α2,6-linked manner [26], 3- and 4-hydroxyl groups of the GalNAc are unmodified and exposed for the binding to MGL [23,27].

In this study, we were not able to identify a high number of glycopeptides that carry the Tn antigen on the MGL-binding proteins. In a previous study on Jurkat T cells [22], however, we found the Tn antigen to be the dominant epitope on the cell surface MGL-binding proteins. In Jurkat cells, the expression of the MGL-glycan epitope is mediated by the loss of Cosmc, the chaperone needed for *O*-glycan elongation. The corresponding *Cosmc* mutation and altered T-synthase activity did not appear to be the mechanisms driving Tn expression in colorectal carcinomas [28,29] and colorectal cancer cell lines [17]. These findings suggest that other pathways might be implicated in Tn expression, such as the altered localization in the ER [30–32] or Golgi [33] and/or the up-regulation of initiating enzymes pp-GalNAc-Ts. The latter mechanism has been recently explored by Sahasrabudhe et al. [17], who found a correlation between the MGL ligands synthesis in HT29 cell line and the overexpression of GALNT3.

However, overall, we do not have a complete picture of the relative contribution of Tn antigen compared to LacdiNAc protein modifications because for other major MGL-binding proteins, we did not find glycopeptides carrying the MGL binding specific epitope. In attempts to increase the number of glycopeptide identifications, and get a more comprehensive overview of the relative contribution of Tn compared to LacdiNAc on the MGL-binding proteins, we tested several approaches to specifically enrich for glycopeptides e.g. MGL pull-down experiments using tryptic peptides instead of proteins as input, and hydrophilic interaction liquid chromatography (HILIC) of tryptic peptides from MGL-captured proteins. We also used an alternative protease, i.e. chymotrypsin, for digestion. However, none of these approaches was successful in increasing the number of identified glycopeptides. Moreover,

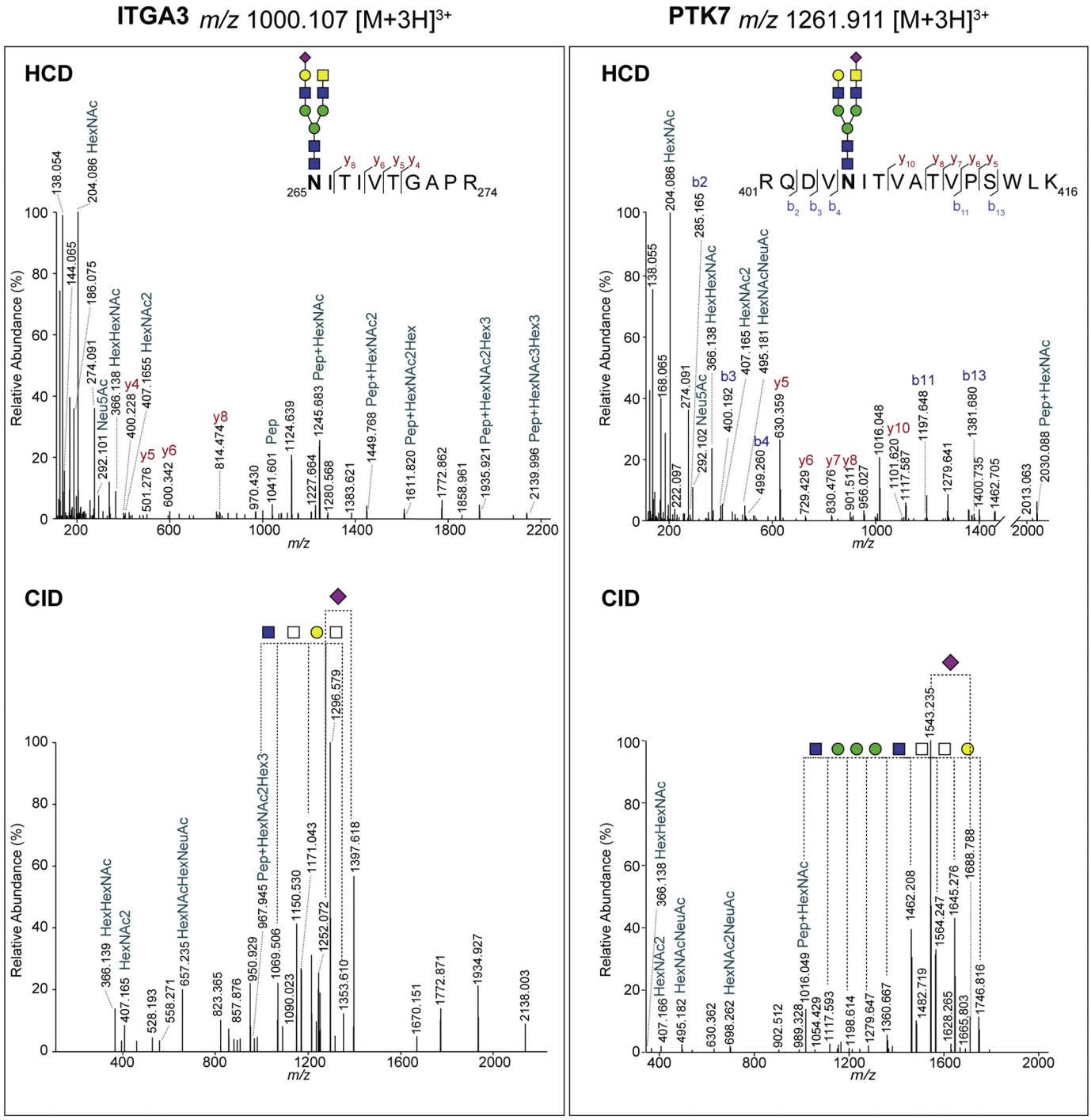

Fig. 3. N-glycosylated peptides of ITA3 and PTK7 carrying a LacdiNAc. Left panels: MS/MS spectra of the tryptic glycopeptide of ITA3 observed at *m/z* 1000.107 $[M + 3H]^{3+}$. Right panels: MS/MS spectra of tryptic glycopeptide of PTK7, m/z 1261.911 $[M + 3H]^{3+}$. b and y ions indicate fragments without glycans, unless otherwise indicated. Upper panels: HCD fragmentation spectra; lower panels: CID fragmentation spectra; Pep: peptide; white square: *N*-acetylhexosamine; blue square: *N*-acetylglucosamine, yellow square: *N*-acetylgalactosamine; green circle: mannose; yellow circle: galactose; purple diamond: *N*-acetylneuraminic acid.

since the Tn antigen epitope can also be found on tyrosines, another known binder of MGL [34], we checked our data for this possibility, but no tyrosines with a HexNAc were observed. We believe that the low intensity of the glycopeptides in combination with their MS/MS fragmentation characteristics complicates their straightforward identification.

Even though we did not find a glycopeptide explaining the binding of several identified proteins to MGL (e.g. MET, ITGB1, PTGFRN and IGSF3), we are confident that these proteins are genuine MGL-interactors as they were strongly enriched in three independent pull-down experiments in both HCT116 and HT29 cells. Moreover, avalaible transcriptomic data [19] show that the expression of the top 25 candidates (Table 1) is similar for all three cell lines. Together, this suggests that the binding of MGL to HCT116 and HT29, but not to LS174T cells, is actually driven by differences in the glycosylation of MGL-binding proteins between the three CRC cell lines. Of note, we have not identified any mucins as MGL-binding proteins in the cell lines even though mucin 1 (MUC1) was shown to bind to MGL [35] and is highly overexpressed in CRC tissues [36]. However, the CRC cell lines used in the current study express low levels of MUC1 [37,38].

Table 1
Major cell surface MGL-binding proteins in HCT116 and HT29 cells. The overall list of cell surface MGL-binding proteins in HCT116 and HT29 (Table S2) was sorted for the total number of peptides that was identified in the twelve LC-MS/MS runs (two cell lines, 3 independent pull-downs per cell line, analysed in duplicate by LC-MS/MS). The resulting top 25 proteins are shown. See Table S2 for full list of combined 83 MGL-binding proteins. M.W.: molecular weight; Sum of peptides: total of peptides identified in all replicates from each cell line. [a]Glycopeptide carrying LacdiNAc or Tn antigen found.

UniProt entry	Protein name	Gene name	M.W.	Sum of peptides	
				HCT116	HT29
Q92896	Golgi apparatus protein 1	GLG1	134	531[a]	200[a]
P08581	Hepatocyte growth factor receptor	MET	155	217	239
Q92673	Sortilin-related receptor	SORL1	248	267[a]	178
P05556	Integrin beta-1	ITGB1	88	197	152
Q13308	Inactive tyrosine-protein kinase 7	PTK7	118	290[a]	30[a]
Q9P2B2	Prostaglandin F2 receptor negative regulator	PTGFRN	98	220	87
O75054	Immunoglobulin superfamily member 3	IGSF3	135	158	142
P26006	Integrin alpha-3	ITGA3	116	168[a]	118
P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF	212		197
Q6ZRP7	Sulfhydryl oxidase 2	QSOX2	77	9	154
Q8NBJ4	Golgi membrane protein 1	GOLM1	45	47	84[a]
P02786	Transferrin receptor protein 1	TFRC	84	40	78
P20023	Complement receptor type 2	CR2	112	105	
P05023	Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	112	77	11
P51572	B-cell receptor-associated protein 31	BCAP31	27		87
O75369	Filamin-B	FLNB	278		72
Q15262	Receptor-type tyrosine-protein phosphatase kappa	PTPRK	162	19	47
P08195	4F2 cell-surface antigen heavy chain	SLC3A2	67	65	
P15529	Membrane cofactor protein	CD46	43		61
P11279	Lysosome-associated membrane glycoprotein 1	LAMP1	44		57
Q14118	Dystroglycan	DAG1	97	57[a]	
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	114	31	26
Q86SJ2	Amphoterin-induced protein 2	AMIGO2	57		56
Q9ULF5	Zinc transporter ZIP10	SLC39A10	94	55	
P23467	Receptor-type tyrosine-protein phosphatase beta	PTPRB	224		51

The MGL-binding proteins identified in this study are involved in many biological processes, and some of them have been identified as drivers of tumor progression. For example, the hepatocyte grow factor receptor (MET), a cell surface receptor involved in downstream signaling, is involved in cell proliferation, survival and morphogenesis in CRC [39], while integrin subunit alpha 3 and beta 1 (ITGA3, ITGB1), together forming the fibronectin receptor, promote invasion. Moreover, Golgi apparatus protein 1 (GLG1), sortilin-related receptor (SORL1) and inactive tyrosine-protein kinase 7 (PTK7) are also implicated in migration, adhesion, morphogenesis and differentiation (Table S1). Although the downstream effect of the binding of MGL to the identified proteins is currently unknown, studies focused on other lectins have demonstrated that intracellular pathways can be activated [40], inducing pro-tumoral effects [41–44].

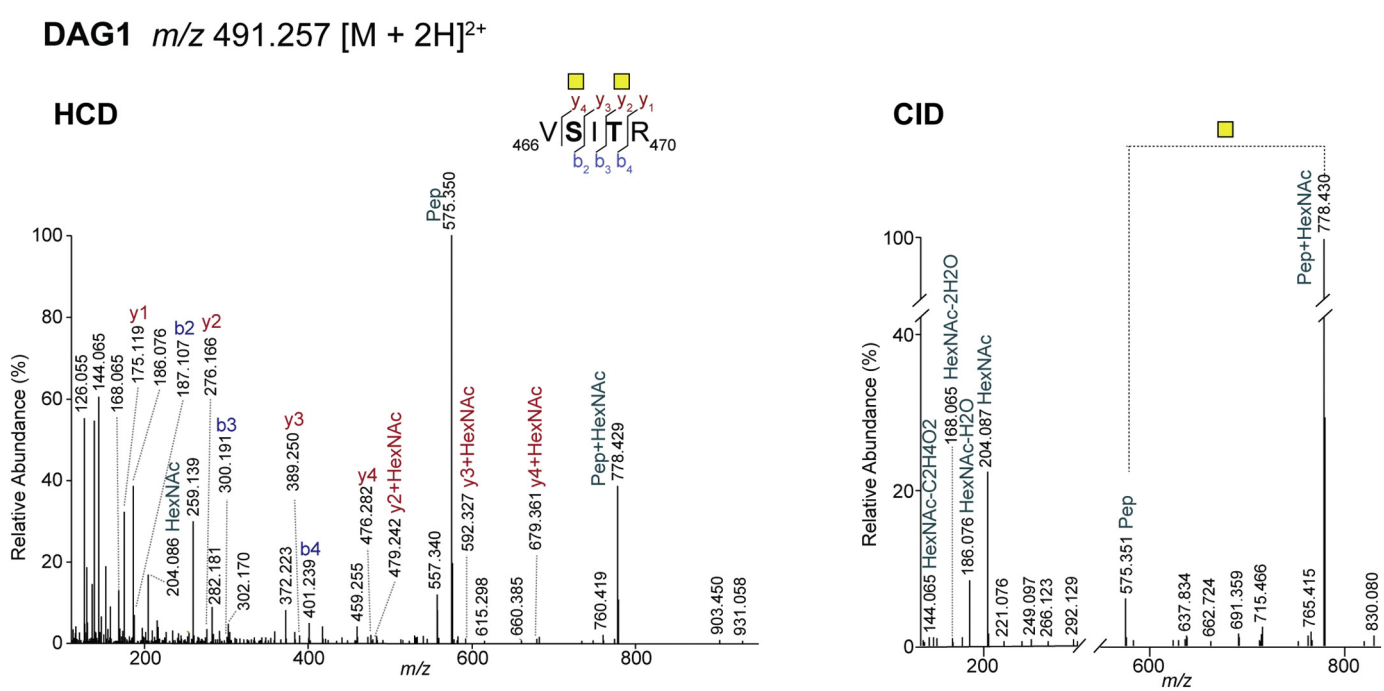

Fig. 4. O-glycosylated peptide of DAG1 carrying Tn antigens. MS/MS spectra of the tryptic glycopeptide of DAG1 at m/z 491.257 $[M + 2H]^{2+}$. Left panel: HCD MS/MS spectrum. Right panel: CID MS/MS spectrum. b and y ions indicate fragments without glycans, unless otherwise indicated. Yellow square: *N*-acetylgalactosamine.

In conclusion, whereas the signaling and effects on the immune system in response to MGL binding to its ligands has been widely explored [11,14,15], the downstream effects of MGL binding to CRC cells and the role of individual ligands remain still unexplored. We believe that the identification of glycoproteins binding to MGL elucidated in this paper represents the first step towards a better understanding of mechanisms driving MGL-mediated tumor progression.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2020.129513.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Gabi W. van Pelt for providing the human CRC cells and the availability to use the cell culture facility from the Department of Surgery, LUMC. We acknowledge the Flow Cytometry Facility (FCF) of the LUMC for providing courses, equipment and support for flow cytometry experiments. This project 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. This work was also supported in part by the Région Occitanie and the French Ministry of Research with the program "Investissement d'Avenir, Infrastructures Nationales en Biologie et Santé" (Proteomics French Infrastructure, ProFI, project ANR 10-INBS-08).

References

[1] S.S. Pinho, C.A. Reis, Glycosylation in cancer: mechanisms and clinical implications, Nat. Rev. Cancer 15 (9) (2015) 540–555.
[2] S. Holst, M. Wuhrer, Y. Rombouts, Glycosylation characteristics of colorectal cancer, Adv. Cancer Res. 126 (2015) 203–256.
[3] T. Ju, et al., Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers, Proteome. Clin. Appl. 7 (9–10) (2013) 618–631.
[4] C. Fu, et al., Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen, HLA 88 (6) (2016) 275–286.
[5] K. Soga, et al., Terminal N-acetylgalactosamine-specific leguminous lectin from Wisteria japonica as a probe for human lung squamous cell carcinoma, PLoS One 8 (12) (2013) e83886.
[6] E. Machado, et al., N-glycosylation of total cellular glycoproteins from the human ovarian carcinoma SKOV3 cell line and of recombinantly expressed human erythropoietin, Glycobiology 21 (3) (2011) 376–386.
[7] M.I. Che, et al., Beta1, 4-N-acetylgalactosaminyltransferase III modulates cancer stemness through EGFR signaling pathway in colon cancer cells, Oncotarget 5 (11) (2014) 3673–3684.
[8] J. Huang, et al., Beta1,4-N-acetylgalactosaminyltransferase III enhances malignant phenotypes of colon cancer cells, Mol. Cancer Res. 5 (6) (2007) 543–552.
[9] K. Hirano, et al., Expression of LacdiNAc groups on N-glycans among human tumors is complex, Biomed. Res. Int. (2014) 981627.
[10] O. Haji-Ghassemi, et al., Molecular basis for recognition of the Cancer Glycobiomarker, LacdiNAc (GalNAc[beta1– > 4]GlcNAc), by Wisteria floribunda agglutinin, J. Biol. Chem. 291 (46) (2016) 24085–24095.
[11] S.J. van Vliet, et al., Regulation of effector T cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45, Nat. Immunol. 7 (11) (2006) 1200–1208.
[12] R. Beatson, et al., The breast Cancer-associated Glycoforms of MUC1, MUC1-Tn and sialyl-Tn, are expressed in COSMC wild-type cells and bind the C-type lectin MGL, PLoS One 10 (5) (2015) e0125994.
[13] N. Mortezai, et al., Tumor-associated Neu5Ac-Tn and Neu5Gc-Tn antigens bind to C-type lectin CLEC10A (CD301, MGL), Glycobiology 23 (7) (2013) 844–852.
[14] S.J. van Vliet, et al., 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) (2005) 661–669.
[15] S.J. van Vliet, et al., MGL signaling augments TLR2-mediated responses for enhanced IL-10 and TNF-alpha secretion, J. Leukoc. Biol. 94 (2) (2013) 315–323.
[16] K. Lenos, et al., MGL ligand expression is correlated to BRAF mutation and associated with poor survival of stage III colon cancer patients, Oncotarget 6 (28) (2015) 26278–26290.
[17] N.M. Sahasrabudhe, et al., Oncogenic BRAFV600E drives expression of MGL ligands in the colorectal cancer cell line HT29 through N-acetylgalactosamine-transferase 3, Biol. Chem. 399 (7) (2018) 649–659.
[18] N.M. Sahasrabudhe, et al., MGL ligand expression is correlated to lower survival and distant metastasis in cervical squamous cell and adenosquamous carcinoma, Front. Oncol. 9 (2019) 29.
[19] K.C.G. Berg, et al., Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies, Mol. Cancer 16 (1) (2017) 116.
[20] S. Holst, et al., N-glycosylation profiling of colorectal Cancer cell lines reveals association of fucosylation with differentiation and caudal type Homebox 1 (CDX1)/villin mRNA expression, Mol. Cell. Proteomics 15 (1) (2016) 124–140.
[21] A.A. Westen, et al., Comparing six commercial autosomal STR kits in a large Dutch population sample, Forensic Sci. Int. Genet. 10 (2014) 55–63.
[22] M. Pirro, et al., Glycoproteomic analysis of MGL-binding proteins on acute T-cell leukemia cells, J. Proteome Res. 18 (3) (2019) 1125–1132.
[23] F. Marcelo, et al., Identification of a secondary binding site in human macrophage galactose-type lectin by microarray studies: implications for the molecular recognition of its ligands, J. Biol. Chem. 294 (4) (2019) 1300–1311.
[24] J.A. Vizcaino, et al., 2016 update of the PRIDE database and its related tools, Nucleic Acids Res. 44 (22) (2016) 11033.
[25] T. Sato, et al., Molecular cloning and characterization of a novel human beta 1,4-N-acetylgalactosaminyltransferase, beta 4GalNAc-T3, responsible for the synthesis of N,N'-diacetyllactosediamine, galNAc beta 1-4GlcNAc, J. Biol. Chem. 278 (48) (2003) 47534–47544.
[26] P. Stanley, R.D. Cummings, Structures common to different glycans, in: A. Varki, R.D. Cummings (Eds.), Essentials of Glycobiology, Cold Spring Harbor (NY), 2015, pp. 161–178.
[27] S.A. Jegouzo, et al., Organization of the extracellular portion of the macrophage galactose receptor: a trimeric cluster of simple binding sites for N-acetylgalactosamine, Glycobiology 23 (7) (2013) 853–864.
[28] X. Sun, T. Ju, R.D. Cummings, Differential expression of Cosmc, T-synthase and mucins in Tn-positive colorectal cancers, BMC Cancer 18 (1) (2018) 827.
[29] N.J. Yoo, M.S. Kim, S.H. Lee, Absence of COSMC gene mutations in breast and colorectal carcinomas, APMIS 116 (2) (2008) 154–155.
[30] F. Bard, J. Chia, Cracking the glycome encoder: signaling, trafficking, and glycosylation, Trends Cell Biol. 26 (5) (2016) 379–388.
[31] D.J. Gill, et al., Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness, Proc. Natl. Acad. Sci. U. S. A. 110 (34) (2013) E3152–E3161.
[32] A.T. Nguyen, et al., Organelle specific O-glycosylation drives MMP14 activation, tumor growth, and metastasis, Cancer Cell 32 (5) (2017) 639–653 e6.
[33] G. Egea, et al., cis-Golgi resident proteins and O-glycans are abnormally compartmentalized in the RER of colon cancer cells, J. Cell Sci. 105 (Pt 3) (1993) 819–830.
[34] R. Gibadullin, et al., GalNAc-tyrosine is a ligand of plant lectins, antibodies, and human and murine macrophage galactose-type lectins, ACS Chem. Biol. 12 (8) (2017) 2172–2182.
[35] I.G. Zizzari, et al., MGL receptor and immunity: when the ligand can make the difference, J Immunol Res 2015 (2015) 450695.
[36] E. Saeland, et al., The C-type lectin MGL expressed by dendritic cells detects glycan changes on MUC1 in colon carcinoma, Cancer Immunol. Immunother. 56 (8) (2007) 1225–1236.
[37] J.H. Chik, et al., Comprehensive glycomics comparison between colon cancer cell cultures and tumours: implications for biomarker studies, J. Proteome 108 (2014) 146–162.
[38] P.L. Devine, et al., Expression of MUC1 and MUC2 mucins by human tumor cell lines, Tumour Biol. 13 (5–6) (1992) 268–277.
[39] J.C. Samame Perez-Vargas, et al., Role of cMET in the development and progression of colorectal cancer, Int. J. Mol. Sci. 14 (9) (2013) 18056–18077.
[40] S. Gout, et al., Death receptor-3, a new E-selectin counter-receptor that confers migration and survival advantages to colon carcinoma cells by triggering p38 and ERK MAPK activation, Cancer Res. 66 (18) (2006) 9117–9124.
[41] T. Aychek, et al., E-selectin regulates gene expression in metastatic colorectal carcinoma cells and enhances HMGB1 release, Int. J. Cancer 123 (8) (2008) 1741–1750.
[42] M.A. Carrascal, et al., A functional glycoproteomics approach identifies CD13 as a novel E-selectin ligand in breast cancer, Biochim. Biophys. Acta Gen. Subj. 1862 (9) (2018) 2069–2080.
[43] M.A. Carrascal, et al., Staining of E-selectin ligands on paraffin-embedded sections of tumor tissue, BMC Cancer 18 (1) (2018) 495.
[44] J. Tomlinson, et al., Human colon cancer cells express multiple glycoprotein ligands for E-selectin, Int. J. Oncol. 16 (2) (2000) 347–353.