



HOKKAIDO UNIVERSITY

Title	Synthetic Mucin-Like Glycopeptides as Versatile Tools to Measure Effects of Glycan Structure/Density/Position on the Interaction with Adhesion/Growth-Regulatory Galectins in Arrays
Author(s)	Artigas, Gerard; Hinou, Hiroshi; Garcia-Martin, Fayna et al.
Citation	Chemistry-an asian journal, 12(1), 159-167 https://doi.org/10.1002/asia.201601420
Issue Date	2017-01-03
Doc URL	https://hdl.handle.net/2115/68039
Rights	This is the peer reviewed version of the following article: G. Artigas, H. Hinou, F. Garcia-Martin, H.-J. Gabius, S.-I. Nishimura, Synthetic Mucin-Like Glycopeptides as Versatile Tools to Measure Effects of Glycan Structure/Density/Position on the Interaction with Adhesion/Growth-Regulatory Galectins in Arrays, Chem. Asian J. 2017, 12, 159-167, which has been published in final form at 10.1002/asia.201601420. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Type	journal article
File Information	manuscript (final).pdf



Synthetic Mucin-like Glycopeptides as Versatile Tools to Measure Effects of Glycan Structure/Density/Position on Interaction with Adhesion/Growth-regulatory Galectins in Arrays

Dr. Gerard Artigas,^[a] Dr. Hiroshi Hinou,^{*,[a,b]} Dr. Fayna Garcia-Martin,^[a] Prof. Dr. Hans-Joachim Gabius,^{*,[c]} Prof. Dr. Shin-Ichiro Nishimura,^{*,[a,b]}

[a] Faculty of Advanced Life Science, Hokkaido University, N21W11, Kita-ku, Sapporo 001-0021 (Japan); E-mail: hinou@sci.hokudai.ac.jp, shin@sci.hokudai.ac.jp

[b] Medicinal Chemistry Pharmaceuticals, Co., Ltd., N9W15, Chuo-ku, Sapporo 060-0009 (Japan)

[c] Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 München (Germany); E-mail: gabius@tiph.vetmed.uni-muenchen.de

Abstract: Functional pairing of cellular glycoconjugates with tissue lectins is a highly selective process, whose determinative factors have not yet been fully delineated. Glycan structure and modes of presentation, i.e. its position and density, can contribute to binding, as different members of a lectin family can regulate degrees of responsiveness to these factors. Using a peptide repeat sequence motif of the glycoprotein mucin-1, the principle of introducing synthetic (glyco)peptides with distinct variations in these three parameters to array-based screening of tissue lectins is illustrated. Interaction profiles of seven adhesion/growth-regulatory galectins cover the range from intense signals with core 2 pentasaccharides and core 1 binding for galectins-3 and -5 to lack of binding for galectin-1 and also the galectin-related protein, which was included as negative control. Remarkably, the two tandem-repeat-type galectins-4 and -8 were distinguished by core 1 sialylation, as the two separated domains were. These results encourage further synthetic elaboration of the glycopeptide library and testing of the network of natural galectins and rationally engineered variants of the lectins.

Keywords: agglutinin, glycoprotein, glycosylation, lectin, TF antigen

Introduction

The concept of the sugar code assigns functionality to the glycan chains of glycoproteins via a sweet complementarity with endogenous receptors, which ‘read’ the sugar-encoded information and ‘translate’ it into cellular effects.^[1] Dynamic shifts in the glycome, for example associated to diseases, hereby acquire a physiological meaning as an alteration in the display of cellular signals.^[2] With focus on mucin-type O-glycosylation, the general trend for presentation of short determinants in cancer, known as the Tn antigen (CD175; GalNAc α 1-O-Ser/Thr) or the Thomsen-Friedenreich antigen (T(F), CD176; Gal β 1,3GalNAc α 1-O-Ser/Thr),^[3] has attracted considerable attention. That α 2,6-sialylation of Tn or T antigen preclude further O-glycan elaboration with impact on breast cancer cell growth and metastasis formation in breast and lung cancer ^[4] invigorates efforts to understand the ligand properties of these epitopes and of glycans, whose production is blocked by the α 2,6-sialylation. As physiological receptor, a member of the galectin family, i.e. galectin-3, has already been identified. Although its interaction is weak when tested with the free mono (Tn)- or disaccharide (TF), glycan presentation by mucins, especially the mucin-1 glycoprotein (MUC1), led to high avidity.^[5] However, the diversity of mucin-type O-glycosylation in natural glycoproteins including potentially confounding contributions, e.g. the N-glycans and 20 amino acids tandem-repeat structure of MUC1, sets inherent limits to discern structure-activity relationships. In addition, topological parameters of glycan presentation are increasingly becoming of interest to explain the high degree of selectivity of endogenous lectins for their counterreceptors.^[1] In other words, positions and density of glycan epitopes appear to operate as switches for affinity and as source of selectivity of lectin binding. But the underlying rules of how they do so remain to be defined. To address this issue properly synthetic glycopeptides offer a valuable tool, as we herein illustrate.

We have previously reported on the design of a glycopeptide microarray platform and its applicability for epitope mapping of monoclonal antibodies against the MUC1 glycoprotein.^[6] Of note, glycopeptides with a length of between eight and 19 amino acids and one O-linked TF epitope have recently been shown to bind galectin-3.^[7] These two lines of research set the stage for applying this platform to a thorough evaluation of galectins interaction for glycopeptides. Toward this end the sugar(headgroup) structure, the site of glycosylation and its density in the peptide

backbone are systematically varied. Because this galectin can be subject to proteolytic processing in situ with impact on its cross-linking capacity,^[8] the full-length protein and the truncated form (without the N-terminal tail) are comparatively tested. Due to the presence of galectins as a network^[9] and indications for galectin interaction to glycoproteins with mucin-type O-glycosylation beyond galectin-3,^[10] we also have taken the testing from the chimera-type galectin to members of the two other groups of galectins, i.e. tandem-repeat- and proto-type proteins (please see Figure 1 for an overview of the test proteins). Our report thus presents the biology-driven design of an array with synthetic glycopeptides^[6, 11] for analyzing the impact of distinct parameters of ligand presentation on binding process to endogenous effectors comparatively.

Results and Discussion

The natural pairing of a tissue lectin with a cellular glycoconjugate depends on much more than presence of suited glycan (for an overview on the six levels of affinity regulation, please see Figure 3 in ref. 3c^[3c]). The challenge for synthetic chemistry thus is to prepare the scaffold and decorate it site-specifically with glycans, in a manner that lets density of the natural epitopes become programmable. In the cases of Tn, TF, and core 2 O-glycosylation, MUC1 is an accepted role model for a glycoprotein harbouring these determinants.^[12] The N-terminal domain is the major part of the extracellular region. Its 20-amino acid sequence motif with five Ser/Thr residues that occurs in a variable number of tandem repeats endows MUC1 with acceptor properties for O-glycosylation. In secreted MUC1 from T47D breast cancer cell, the five Ser/Thr were occupied with Tn/TF antigen structure and Ser residues are oriented to Tn antigen.^[12c] In urinary MUC1 from healthy donor, core 1/2 O-glycosylation are found as major modifications.^[12d] Tumor-associated epitope property of the MUC1 motif was sensitively affected by modification of adjacent O-glycosylation sites.^[6d] Plant lectins exhibit different binding strength to Tn antigen on the three Thr residues in the MUC1 motif.^[12e] The affinity of TF antigen for galectin-3 is enhanced by the MUC1 peptide scaffold.^[7] Knowing the sequence and feature of O-glycosylation of this unit provides the starting point to prepare a panel of (glyco)peptides to relate their structure to lectin-binding potential in the array system.

Synthesis of MUC1-based (glyco)peptides

Our panel consists of 35 compounds. Six spaced glycan derivatives linked to Ser or Thr and twenty nine (glyco)peptides are presented in Figure 2A. Three sugar(head group) structure, Tn, TF, and core2, were arranged systematically on the five potentially glycosylation site (compound **2-16**). Combination of di- and tri-Tn antigen modification (compound **17-24**), and TF antigen with two Tn antigens on adjacent Ser residues (compound **25**), Sialyl-TF antigen (compound **33**) and its heterogeneous modification model (compound **32** and **33**)^[6d] were also added to this panel. They were synthesized by standard solid-phase peptide synthesis (SPPS) methodology using the Fmoc/tBu strategy^[13] assisted by microwave irradiation^[14] on a polyethylene glycol-based resin (NovaPEG) with a Rink linker. O-Acetylated glycans attached to amino acid moieties were processed during peptide elongation by a double activation procedure, employing a low-molar excess of N α -Fmoc-glycosyl-Ser/Thr.^[15] As final synthesis step, according to the coupling procedure followed with Fmoc-amino acids, a polyethylene glycol linker (except for compounds **32-35**)^[6d] and 5-oxohexanoic acid were introduced at the N-terminus of each glycopeptidyl resin.^[6] Cleavage from the resin and the removal of side-chain protecting groups were performed simultaneously by treatment with 90% aqueous TFA. Then, deacetylation of the sugar moiety was carried out to afford compound **1-32**. In order to produce compounds **33-35** with TF sialylation and extension of the core 2 (β 1,6-branch), enzymatic chain elongation by using α 2,3-sialyltransferase and β 1,4-galactosyltransferase was performed.^[6d, 16]

Combining our microarray approach for glycopeptides with the glycoblotting technology,^[6, 17] in which an oxime bond is formed chemoselectively by conjugation to the ketone linker attached to the N-terminus and aminooxy (AO) groups presented by the slides coated with phosphorylcholine (PC)-copolymer,^[6a] printing resulted in the display of the test compounds in quadruplicates, as shown in Figure 2B. To determine extent of binding of biotinylated galectins, commercially available cyanine3-labeled streptavidin was used as probe for detection in a GlycoStation system [<http://www.glycotechnica.com>]. In order to preclude missing low-affinity binding, assay conditions were further modified (please see Materials and Methods for further detail) by using the evanescent-field fluorescence-assisted detection principle in combination with the GlycoStation system. This set-up enabled real-time monitoring without a drying-up the microarray surface.^[18] Thus, comparative analysis of the binding of galectins (please see Figure 1) under identical conditions was possible.

Binding profile of galectin-3

This protein is the chimera-type galectin with a trimodular design composed of an N-terminal peptide with two sites for Ser phosphorylation and non-triple-helical collagen-like repeats (these two regions forming the N-terminal tail) as well as the carbohydrate recognition domain (CRD) (Figure 1a). While monomeric in solution, galectin-3 can aggregate in the presence of ligands by engaging contacts via the tail and the CRD (for a recent review on ligand-induced self-aggregation, please see ref. 18^[19]).

Full-length galectin-3 has a very restricted positivity on this array: signals were obtained for the two glycopeptides presenting the core 2 pentasaccharide, consisting of sialylated TF antigen and the β 1,6-linked N-acetyllactosamine, and the N-acetyllactosamine can contribute the interaction (Figure 3). The two Tn antigen on compound **34** can contribute negatively to the interaction by glycoform-dependent positional effects.^[6d, 16d] The CRD (after proteolytic removal of the N-terminal tail) lost this activity, as it loses binding affinity to ganglioside GM1 on neuroblastoma cells.^[8b] Monovalent TF antigen presentation in a glycopeptide context (or as disaccharide derivative, please also see ref. 5f^[5f]) is not sufficient for generating a signal in this system. Because combined neuraminidase/*O*-glycanase digestion of MUC1's N-terminal ectodomain "almost completely abolished" its binding by galectin-3,^[5m] a dense TF clustering or presence of core 2 glycans appear to be required to implement interaction. Whether an ionic interaction of the basic protein with sialyl Tn/mono- and disialyl TF determinants may occur in situ is an open question. Having herewith documented signal generation in a highly selective manner for the chimera-type family member, a comparison to other galectins was possible, starting with two tandem-repeat-type galectins.

Binding profiles of tandem-repeat-type galectins

This group of galectins presents two different CRDs (N- and C-terminal) connected by a peptide linker (Figure 1b). Obviously, the bivalent proteins can readily bridge counterreceptors. This property is instrumental for a role as matricellular protein or as routing/delivery device for glycoproteins in apical or axonal transport.^[20] With respect to contact points, galectins-4 and -8 share a strong interaction with 3'-sulfated galactosides via their N-terminal domains.^[21] Both tested proteins give broader specificity and tighter binding than with galectin-3 (Figure 4).

Of particular note, core 1 glycans are reactive with these two tandem-repeat-type galectins in this setting. The sialyl TF antigen (**33**) is a good binder for galectin-8 (Figure 4A and E). Adding the core 2 pentasaccharide (**34**, **35**) enhanced signal intensity. A single TF disaccharide can also confer ligand activity to the glycopeptides. Remarkably, as the comparison of glycopeptides with glycan attachment at different sites revealed, positional effects are operative (please compare the results for the four glycopeptides **8-11**), and glycoamino acids **28** and **29** are not active. The same applies to core 2 trisaccharides **15** and **16**, but active to Ser-linked glycoamino acid **31**.

In contrast to the data for galectin-8, the core 2 pentasaccharides and the sialyl TF epitope fail to be a strong ligand for galectin-4. In fact, this lectin distinguishes nearly qualitatively between 3'-O-sulfation and respective sialylation of core 1 O-glycans.^[22] In this case, also with positional effects, a single TF epitope is the main binder, along with core 2 trisaccharide for both glycopeptides and glycoamino acids (Figure 4B and F). When testing its two CRDs separately, Gal-4N, indeed, can target sialylated glycopeptides **34** and **35**, establishing the difference between the profiles of Gal-4N and Gal-4C (Figure 4C and D, G and H). Considered in the network concept, both tandem-repeat-type galectins appear likely to interact with TF antigen-presenting mucins, and galectin-8 has additional affinity to sialylated core 1/core 2 tri- or pentasaccharides. If the linker of galectin-4 is cleaved in situ, both of the free CRD, and the Gal-4N CRD is capable to bind the core 2 pentasaccharide. Proteolytic processing can thus turn a tandem-repeat-type protein into single CRDs and modify the ligand selectivity. Naturally present, besides chimera- and tandem-repeat-type galectins, are proteins of the class of proto-type galectins (Figure 1). These proteins occur either as a monomer (without or with N-terminal extension) or a non-covalently associated homodimer (Figure 1c).

Binding profiles of proto-type galectins

Homodimeric galectin-1 (galaptin), the first mammalian galectin purified and best-studied family member, exerts activity as cross-linker. In this assay, the lectin did not associate to any compound presented on the array (Figure 5A). On the cellular level, galectin-1 binding was shown to be abrogated by blocking core 2-type O-glycan elongation by overexpression of α 2,3-sialyltransferase-I or α -N-acetylgalactosamine α 2,6-sialyltransferase-I (ST6GalNAc-I).^[23] That a homodimeric galectin can react with glycocompounds on the chip is attested by C-GRIFIN, most closely related to galectin-3

in the phylogenetic tree.^[24] Its binding to the core 1 disaccharide and also to the core 2 trisaccharide on a serine residue resembled aspects of the profile of galectin-4 (Figure 5B and E). Further illustrating the diversity of glycan binding among galectins, monomeric galectin-5, a rat protein implicated in exosomal sorting during reticulocyte maturation,^[25] is a weak binder, its activity confined to the glycopeptides **10** and **11** (Figure 5C and F). This protein has an N-terminal extension, as also present in GRP, which is most homologous to galectin-8C in the phylogenetic tree and deviates from the other galectins due to its lack of canonical interaction with β -galactosides.^[26] This protein from chicken, i.e. C-GRP, did not interact with any test substance. The lack of interactions or weak signal intensities underscore that galectins, which share the β -sandwich folding, distinguish glycosignatures, depending on positional effects. Also, structural properties of the protein, as proteolytic truncation of galectin-3 documents, matter.

Conclusions

The dynamics of shifts in the profile of mucin-type O-glycosylation poses the question on its physiological significance. Assuming a functional pairing with lectins is one route to turn presence and structural changes into cellular effects. Synthetic chemistry had been instrumental to trace presence of endogenous receptors for the TF epitope by glycohistochemistry.^[27] Moving from an inert scaffold used in that approach to the sequence of the natural peptide repeat and implementing O-glycosylation combinatorially is the way to go for comprehensive activity testing of the glycopeptides,^[6-7, 28] here in microarray setting. Because topological features count on both sides of the recognition process,^[29] we here deliberately tested members of a lectin family of different modular architecture. Obviously, protein structure matters, as truncation of galectin-3 reveals, and homologous proteins can markedly differ in binding profiles for distinct glycans (galectins-8 and -4 for the core 2 pentasaccharides) or the panel (galectin-1). (Figure 6) These results encourage to systematically elaborate the glycopeptide library as well as to engineer and test galectin variants with particular structural deviations from the wild-type proteins, e.g. design of tandem-repeat-type proteins from CRDs of galectins-1 and -3.^[30]

Materials and Methods

Materials

Commercially available solvents and reagents were used without purification. NovaPEG Rink Amide resin (loading 0.37 mmol/g) and *N*^α-[9-(fluorenylmethoxy)carbonyl (Fmoc)]-L-amino acids, except for glycosylated compounds, were purchased from Novabiochem Merck KG (Darmstadt, Germany). *N*^α-Fmoc-amino acids, such as Fmoc-(Ac₃GalNAc α)serine/threonine, Fmoc-(Ac₄Gal β 1,3Ac₂GalNAc α)serine/threonine and Fmoc-(Ac₄Gal β 1,3[GlcNAc β 1-6]Ac₂GalNAc α)serine/threonine, were synthesized according to the method reported previously. 1-[Bis(dimethylamino)methylumyl]-1H-benzotriazole-3-oxide hexafluorophosphate (HBTU), 1-hydroxybenzotriazole monohydrate (HOBt), (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP), and *N,N*-diisopropylethylamine (DIEA) were purchased from Kokusan Chemical Co., Ltd (Tokyo, Japan). *N,N*-dimethylformamide (DMF) and 20% piperidine in DMF was purchased from Watanabe Chemical Industries, Ltd. (Hiroshima, Japan). 1-Hydroxy-7-azabenzotriazole (HOAt), acetic anhydride, triisopropylsilane (TIS) and 2,2,2-trifluoroacetic acid (TFA) were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). 5-Oxohexanoic acid was purchased from Tokyo Chemical Industry Co. Ltd. (Tokyo, Japan). All solid-phase reactions were performed manually in a polypropylene tube equipped with a filter (LibraTube, Hipep Laboratories, Kyoto, Japan). Activated N-acetylneuraminic acid (CMP-NANA) and uridine-5'-diphosphogalactose, disodium salt (UDP-galactose) were purchased from Yamasa Corporation (Chiba, Japan). Bovine β 1,4-galactosyltransferase (β 1,4-GalT-1) and rat liver α 2,3-sialyltransferase (α 2,3-SialylT-I) were purchased from Sigma Aldrich and Calbiochem Co. Ltd (Merck Millipore), respectively. Microwave-assisted coupling and deprotection reactions on resins were carried out on an EYELA microwave synthesizer Wave Magic (MWS-1000A, Tokyo Rikakikai Co., LTD., Tokyo, Japan), in which the reaction temperature was kept at 50 °C with a wattage range of 0-100 W. Preparative HPLC purifications were performed on a Prominence Shimadzu HPLC system [(Shimadzu Corporation, Kyoto, Japan) equipped with two LC-6AD pumps, an SPD-20A UV/VIS detector at 220 nm for monitor, and Inertsil ODS-3 reversed-phase C-18 column (250×4.6 mm I.D., GL Sciences Inc., Tokyo, Japan); flow rate, 5.0 mL/min; eluent A, H₂O with 0.1% TFA and eluent B, acetonitrile with 0.1% TFA for non-

glycosylated peptides and glycopeptides having neutral sugar; or buffer A, containing 25 mM ammonium acetate, pH 5.5, and buffer B, acetonitrile containing 10% buffer A, for glycopeptides containing sialic acid residue; composition of the solvent, 0-60 min in a linear gradient flow from (A/B) = (98/2) to (70/30) or (A/B) = (90/10) to (55/45); detection, UV at 220 nm]. Analytical RP-HPLC was conducted using a Waters Acquity Ultra Performance LC system equipped with binary solvent delivery pump, an auto sampler and a UV detector and an Acquity UPLC BEH[®]C18 column (1.7 μ m, 2.1 $\times$ 50 mm, Waters). High-resolution electrospray ionization mass spectra (ESI-HRMS) with JEOL JMS-700TZ, and Amino acid analysis with a JEOL JLC-500/V equipped with ninhydrin detection system, were performed at the Center of Instrumental Analysis at Hokkaido University. MALDI-TOF mass spectra were performed in a Bruker Daltonics Ultraflex MALDI-TOF/TOF mass spectrometer using DHB as a matrix.

Specified buffer used in the present study is a solution of PBS buffer (0.01 M Na₂HPO₄, 0.01 M NaH₂PO₄, 0.138 M NaCl, and 0.0027 M KCl, pH 7.4) containing 0.05% (v/v) Tween-20. Chemicals for buffer preparation and albumin from bovine serum (Cohn fraction V, pH 7.0, min 96%) were purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan). CyTM3-labeled streptavidin was from Life Technologies (Thermo Fisher Scientific Corporation). Microarray slides (75 x 25 x 1 mm) were supplied from Sumitomo Bakelite Co., Ltd. (Tokyo, Japan). Silicon rubber sheets (60 x 24 x 0.1 mm) were purchased from Fuso Rubber Co., Ltd. (Hiroshima, Japan). Micro cover glasses (18 x 18 mm) were purchased from Matsunami Glass Ind., Ltd. (Osaka, Japan). Fluorescence images of microarray slides were measured at on GlycoStationTM Reader 1200 (GlycoTechnica Ltd., Yokohama, Japan). The digital images of fluorescence responses were analyzed using ArrayVisionTM software version 8.0 (GE Healthcare) and GraphPad PrismTM software version 5 (GraphPad Software, Inc.).

Fabrication of MUC-1-based (glyco)peptide library

MUC-1 (glyco)peptides and glycoamino acids (**1-35**) were synthesized manually by microwave-assisted SPPS^[15a] by using *N*^α-Fmoc-glycosyl-Ser/Thr amino acid on NovaPEG Rink Amide resin (loading 0.37 mmol/g). All couplings and *N*^α-Fmoc removal reactions were conducted under heating at 50 °C. First, the resin was swollen with CH₂Cl₂ for 1h at room temperature.^[13] Solvents and soluble reagents were

removed by suction in polypropylene syringes fitted with a porous disk. Washings steps between deprotection and couplings were performed at room temperature with DMF (3 x 1 min), CH₂Cl₂ (3 x 1 min) and DMF (3 x 1 min). *N*-fluoren-9-ylmethoxycarbonyl (Fmoc) groups were deprotected with 20% piperidine in DMF for 3 min. Then, protected Fmoc-amino acids (3 equiv) were preactivated with HBTU (3 equiv), HOBT (3 equiv) and DIEA (6 equiv) for 30 sec -1 min, and subsequently incorporated to the resin. Coupling reactions were conducted for 10 min. For the glycosyl amino acid coupling, a double activation protocol^[15b] was carried out, employing *N*^α-Fmoc-glycosyl-Ser/Thr amino acid (1.2 equiv) in combination with PyBOP (1.2 equiv), HOAt (1.2 equiv) and DIEA (2.5 equiv) in DMF. After 10 min under MW irradiation to maintain 50 °C, without filtering, 1.2 equiv of PyBOP-HOAt were added and allowed to react for another 10 min. Unreacted amino groups were capped by acetylation with 13 mM HOBt in Ac₂O/DIEA/DMF solution (4.75:2.25:93.0, v/v/v, 1 mL) for 5 min at room temperature. As final synthesis step, a polyethylene glycol linker (3 equiv) and oxohexanoic acid (3 equiv) were introduced at the *N*-terminus of each glycopeptidyl resin (**1-31**), according to the above coupling procedure for *N*^α-Fmoc-amino acids. Cleavage from the resin and removal of side-chain protecting groups were performed simultaneously by treatment with a solution of 90% aqueous TFA for 2 h at ambient temperature (15 mL cleavage cocktail/g peptide resin). Cleaved *O*-acetylated glycopeptides were precipitated in cold *tert*-butylmethyl ether, and the resulting crudes were dried by a flow of nitrogen gas. Consecutively, the deacetylation reaction was carefully performed at room temperature by a dropwise addition of 0.5 N NaOH (pH 12.4-13.0) over a solution of crude in MeOH (5 mL). Once the reactions was completed (followed by MS spectroscopy.), the reaction mixture was neutralized with 10% aq. AcOH, and the solvent was removed by flowing of nitrogen gas.^[16a-c] Finally, MUC-1 (glyco)peptides and glycoamino acids (**1-35**) were obtained after purification by preparative RP-HPLC. Composition were determined by amino acid analysis. Each product was analyzed by high-resolution ESI-MS and/or MALDI-TOFMS and UPLC chromatography.

For compounds **33-35**, before RP-HPLC purification, the crudes were subjected to an enzymatic sugar elongation by using a combination of β 1,4GalT-I and α 2,3SialT-I.^{[6d,}

Binding assay

Following our optimized protocol,^[6a] microarray AO/PC-copolymer slides were first deprotected by 2N HCl treatment overnight at room temperature and rinsed with MilliQ H₂O. Next, **1-35** compounds and cyanine 3 N-hydroxysuccinimide ester (Cy3-NHS) were robotically printed for quadruplicate at four concentrations (12.5, 25, 50 and 100 μ M) in 25 mM AcOH-Pyr (pH 5.0), 0.0025% (v/v) Triton X-100. To complete the oxime bond formation, printed slides were incubated for 1 h at 80°C. Then, the remaining free aminoxy groups were capped by treating the slides with an aqueous solution of succinic anhydride (10 mg/mL) for 4 h at room temperature. Subsequently, slides were rinsed with MilliQ H₂O and dried by centrifugation before use.

Printed slides were analyzed by combination the evanescent-field fluorescence-assisted detection principle^[31] with the conventional microarray approach of biotinylated probes and fluorescent streptavidin. First, a silicone rubber sheet with 3 chambers was attached to the slide surface. Then, slides were pretreated with PBS buffer containing 0.05% (v/v) Tween-20 for 15 min and dried by centrifugation. Subsequently, a cover glass was set in each well and, a 30 μ L of a premixed solution of biotinylated galectins (90 μ g/mL) and streptavidin-Cy3 conjugate (4.03 μ g/mL) in PBS containing 0.05% (v/v) Tween-20 and 1% (w/v) BSA was added through the interstice of slide and cover. After 1 h of incubation at room temperature in a humidified chamber, the solution was carefully removed and a new addition of the same solution was infused. This process was repeated three times. As a final step, a last incubation with 30 μ L of streptavidin-Cy3 conjugate (4.03 μ g/mL) was performed. In the cases of C-GRIFIN/GPR, this procedure was slightly modified and the incubation with biotinylated galectins (90 μ g/mL) and streptavidin-Cy3 conjugate (4.03 μ g/mL) was performed three times, separately. Afterwards, slides were subjected to a washing step, consisted of an addition of 30 μ L of PBS buffer containing 0.05% (v/v) Tween-20. Slides images were quantified in the presence of PBS buffer containing 0.05% (v/v) Tween-20 and 1% (w/v) BSA.

Fluorescence intensities were measured with a GlycoStation system [<http://www.glycotechnica.com>], and analyzed with ArrayVision software. The net intensity value of each spot was obtained from the backbround value subtraction, and the average relative fluorescence unit (RFU) was plotted as histogram by using

GraphPad Prism software. From high and low RFU value, the galectin specificity was identified, being the error bars are the standard deviation.

Galectin preparation and labelling

The galectin panel was obtained by recombinant production and the optimized protocol with affinity chromatography on lactose-presenting Sepharose 4B, prepared by resin activation with divinyl sulfone, as described for each galectin in the reports of this laboratory, e.g. for galectin-3.^[32] Galectin-related protein was purified by applying fusion-protein engineering as described.^[26a] The proteins were routinely checked for purity by mass spectrometric fingerprinting after tryptic digestion and one- and two-dimensional gel electrophoresis, then biotinylated under activity-preserving conditions and tested for maintained activity by solid-phase and cell assays as described.^[26, 33]

Acknowledgements

We gratefully acknowledge generous financial support by the EC for the ITN GLYCOPHARM (317297) as well as inspiring discussions with Drs. B. Friday and A. Leddoz. We also acknowledge Ms. S. Oka and Mr. T. Hirose of Center for Instrumental Analysis at Hokkaido University for measurements of amino acid analysis. Part of this work was supported by Grant-in-Aid for Scientific Research S (25220206), B (23350074) and “Innovation COE Project for Future Medicine and Medical Research” from MEXT, Japan.

References

- [1] a) H.-J. Gabius in *The Sugar Code. Fundamentals of glycosciences, Vol.* Wiley-VCH, Weinheim, Germany, **2009**; b) H.-J. Gabius, J. C. Manning, J. Kopitz, S. André and H. Kaltner, *Cellular and molecular life sciences : CMLS* **2016**, *73*, 1989-2016.
- [2] a) S.-i. Hakomori, *Adv. Cancer Res.* **1989**, *52*, 257-331; b) T. Hennet and J. Cabalzar, *Trends in biochemical sciences* **2015**, *40*, 377-384.
- [3] a) T. Ju, V. I. Otto and R. D. Cummings, *Angewandte Chemie* **2011**, *50*, 1770-1791; b) A. P. Corfield, *Biochimica et biophysica acta* **2015**, *1850*, 236-252; c) H.-J. Gabius, H. Kaltner, J. Kopitz and S. André, *Trends in biochemical sciences* **2015**, *40*, 360-376.
- [4] a) N. Murugaesu, M. Irvani, A. van Weverwijk, A. Ivetic, D. A. Johnson, A. Antonopoulos, A. Fearn, M. Jamal-Hanjani, D. Sims, K. Fenwick, C. Mitsopoulos, Q. Gao, N. Orr, M. Zvelebil, S. M. Haslam, A. Dell, H. Yarwood, C. J. Lord, A. Ashworth and C. M. Isacke, *Cancer discovery* **2014**, *4*, 304-317; b) N. E. Reticker-Flynn and S. N. Bhatia, *Cancer discovery* **2015**, *5*, 168-181; c) C. J. Dimitroff, *Cancer research* **2015**, *75*, 3195-3202.
- [5] a) C. Sparrow, H. Leffler and S. H. Barondes, *J. Biol. Chem.* **1987**, *262*, 7383-7390; b) R. S. Bresalier, J. C. Byrd, L. Wang and A. Raz, *Cancer Res.* **1996**, *56*, 4354-4357; c) S. Ramasamy, S. Duraisamy, S. Barbashov, T. Kawano, S. Kharbanda and D. Kufe, *Molecular cell* **2007**, *27*, 992-1004; d) L. G. Yu, N. Andrews, Q. Zhao, D. McKean, J. F. Williams, L. J. Connor, O. V. Gerasimenko, J. Hilken, J. Hirabayashi, K. Kasai and J. M. Rhodes, *The Journal of biological chemistry* **2007**, *282*, 773-781; e) E. M. Rapoport, S. André, O. V. Kurmyshkina, T. V. Pochechueva, V. V. Severov, G. V. Pazygina, H.-J. Gabius and N. V. Bovin, *Glycobiology* **2008**, *18*, 315-324; f) S. R. Stowell, C. M. Arthur, P. Mehta, K. A. Slanina, O. Blixt, H. Leffler, D. F. Smith and R. D. Cummings, *J. Biol. Chem.* **2008**, *283*, 10109-10123; g) P. Argueso, A. Guzman-Aranguez, F. Mantelli, Z. Cao, J. Ricciuto and N. Panjwani, *J. Biol. Chem.* **2009**, *284*, 23037-23045; h) Q. Zhao, M. Barclay, J. Hilken, X. Guo, H. Barrow, J. M. Rhodes and L. G. Yu, *Molecular cancer* **2010**, *9*, 154; i) M. Krzeminski, T. Singh, S. André, M. Lensch, A. M. Wu, A. M. J. J. Bonvin and H.-J. Gabius, *Biochim. Biophys. Acta* **2011**, *1810*, 150-161; j) E. Saeland, A. I. Belo, S. Mongera, I. van Die, G. A. Meijer and Y. van Kooyk, *International journal of cancer* **2012**, *131*, 117-128; k) A. Solatycka, T. Owczarek, F. Piller, V. Piller, B. Pula, L. Wojciech, M. Podhorska-Okolow, P. Dziegiel and M. Ugorski, *Glycobiology* **2012**, *22*, 1042-1054; l) A. B. Yongye, L. Calle, A. Arda, J. Jiménez-Barbero, S. André, H.-J. Gabius, K. Martinez-Mayorga and M. Cudic, *Biochemistry* **2012**, *51*, 7278-7289; m) S. Tanida, Y. Mori, A. Ishida, K. Akita and H. Nakada, *Biochimica et biophysica acta* **2014**, *1840*, 1790-1797.
- [6] a) T. Matsushita, W. Takada, K. Igarashi, K. Naruchi, R. Miyoshi, F. Garcia-Martin, M. Amano, H. Hinou and S.-I. Nishimura, *Biochimica Et Biophysica Acta-General Subjects* **2014**, *1840*, 1105-1116; b) F. Garcia-Martin, T. Matsushita, H. Hinou and S.-I. Nishimura, *Chemistry-a European Journal* **2014**, *20*, 15891-15902; c) H. Coelho, T. Matsushita, G. Artigas, H. Hinou, F. J. Canada, R. Lo-Man, C. Leclerc, E. J. Cabrita, J. Jimenez-Barbero, S.-I. Nishimura, F. Garcia-Martin and F. Marcelo, *Journal of the American Chemical Society* **2015**, *137*, 12438-12441; d) S. Rangappa, G. Artigas, R. Miyoshi, Y. Yokoi, S. Hayakawa, F. Garcia-Martin, H. Hinou and S.-I. Nishimura, *Medchemcomm* **2016**, *7*, 1102-1122.
- [7] M. C. Rodriguez, S. Yegorova, J. P. Pitteloud, A. E. Chavarroche, S. Andre, A. Arda, D. Minond, J. Jimenez-Barbero, H. J. Gabius and M. Cudic, *Biochemistry* **2015**, *54*, 4462-4474.
- [8] a) J. Ochieng, R. Fridman, P. Nangia-Makker, D. E. Kleiner, L. A. Liotta, W. G. Stetler-Stevenson and A. Raz, *Biochemistry* **1994**, *33*, 14109-14114; b) J. Kopitz, C. von Reitzenstein, S. André, H. Kaltner, J. Uhl, V. Ehemann, M. Cantz and H.-J. Gabius, *J. Biol. Chem.* **2001**, *276*, 35917-35923; c) J. Kopitz, S. Vértessy, S. André, S. Fiedler, M. Schnölzer and H.-J. Gabius, *Biochimie* **2014**, *104*, 90-99.
- [9] H. Kaltner and H.-J. Gabius, *Histology and histopathology* **2012**, *27*, 397-416.
- [10] a) K. Wasano and Y. Hirakawa, *J. Histochem. Cytochem.* **1997**, *45*, 275-283; b) A. M. Wu, J. H. Wu, M.-S. Tsai, H. Kaltner and H.-J. Gabius, *Biochem. J.* **2001**, *358*, 529-538; c) A. M. Wu, J. H. Wu, J.-H. Liu, T. Singh, S. André, H. Kaltner and H.-J. Gabius, *Biochimie* **2004**, *86*, 317-326; d) A. M. Wu, T. Singh, J. H. Wu, M. Lensch, S. André and H.-J. Gabius, *Glycobiology* **2006**, *16*, 524-537.
- [11] a) O. Blixt and U. Westerlind, *Current Opinion in Chemical Biology* **2014**, *18*, 62-69; b) N. Laurent, R. Haddoub, J. Voglmeir, S. C. C. Wong, S. J. Gaskell and S. L. Flitsch, *Chembiochem* **2008**, *9*, 2592-2596; c) U. Westerlind, H. Schroder, A. Hobel, N. Gaidzik, A. Kaiser, C. M. Niemeyer, E. Schmitt, H. Waldmann and H. Kunz, *Angewandte Chemie-International Edition* **2009**, *48*, 8263-8267; d) O. Blixt, E. Clo, A. S. Nudelman, K. K. Sorensen, T. Clausen, H. H. Wandall, P. O. Livingston, H. Clausen and K. J. Jensen, *Journal of Proteome Research* **2010**, *9*, 5250-5261.

- [12] a) S. Gendler, J. Taylor-Papadimitriou, T. Duhig, J. Rothbard and J. Burchell, *The Journal of biological chemistry* **1988**, 263, 12820-12823; b) V. Apostolopoulos, L. Stojanovska and S. E. Gargosky, *Cellular and molecular life sciences : CMLS* **2015**, 72, 4475-4500; c) S. Muller, K. Alving, J. Peter-Katalinic, N. Zachara, A. A. Gooley and F. G. Hanisch, *J Biol Chem* **1999**, 274, 18165-18172; d) Y. Miura, K. Kato, Y. Takegawa, M. Kuroguchi, J. Furukawa, Y. Shinohara, N. Nagahori, M. Amano, H. Hinou and S. Nishimura, *Anal Chem* **2010**, 82, 10021-10029; e) D. Madariaga, N. Martinez-Saez, V. J. Somovilla, H. Coelho, J. Valero-Gonzalez, J. Castro-Lopez, J. L. Asensio, J. Jimenez-Barbero, J. H. Busto, A. Avenoza, F. Marcelo, R. Hurtado-Guerrero, F. Corzana and J. M. Peregrina, *ACS Chem Biol* **2015**, 10, 747-756.
- [13] I. Coin, M. Beyermann and M. Bienert, *Nature protocols* **2007**, 2, 3247-3256.
- [14] S. L. Pedersen, A. P. Tofteng, L. Malik and K. J. Jensen, *Chemical Society reviews* **2012**, 41, 1826-1844.
- [15] a) T. Matsushita, H. Hinou, M. Kuroguchi, H. Shimizu and S.-I. Nishimura, *Organic Letters* **2005**, 7, 877-880; b) F. Garcia-Martin, H. Hinou, T. Matsushita, S. Hayakawa and S.-I. Nishimura, *Organic & Biomolecular Chemistry* **2012**, 10, 1612-1617.
- [16] a) M. Fumoto, H. Hinou, T. Ohta, T. Ito, K. Yamada, A. Takimoto, H. Kondo, H. Shimizu, T. Inazu, Y. Nakahara and S.-I. Nishimura, *Journal of the American Chemical Society* **2005**, 127, 11804-11818; b) T. Matsushita, H. Hinou, M. Fumoto, M. Kuroguchi, N. Fujitani, H. Shimizu and S.-I. Nishimura, *Journal of Organic Chemistry* **2006**, 71, 3051-3063; c) N. Ohyabu, H. Hinou, T. Matsushita, R. Izumi, H. Shimizu, K. Kawamoto, Y. Numata, H. Togame, H. Takemoto, H. Kondo and S.-I. Nishimura, *Journal of the American Chemical Society* **2009**, 131, 17102-17109; d) T. Matsushita, N. Ohyabu, N. Fujitani, K. Naruchi, H. Shimizu, H. Hinou and S.-I. Nishimura, *Biochemistry* **2013**, 52, 402-414.
- [17] S.-I. Nishimura, K. Niikura, M. Kuroguchi, T. Matsushita, M. Fumoto, H. Hinou, R. Kamitani, H. Nakagawa, K. Deguchi, N. Miura, K. Monde and H. Kondo, *Angewandte Chemie-International Edition* **2005**, 44, 91-96.
- [18] A. Kuno, N. Uchiyama, S. Koseki-Kuno, Y. Ebe, S. Takashima, M. Yamada and J. Hirabayashi, *Nature methods* **2005**, 2, 851-856.
- [19] H. Ippel, M. C. Miller, S. Vertesy, Y. Zheng, F. J. Canada, D. Suylen, K. Umemoto, C. Romano, T. Hackeng, G. Tai, H. Leffler, J. Kopitz, S. Andre, D. Kubler, J. Jimenez-Barbero, S. Oscarson, H.-J. Gabius and K. H. Mayo, *Glycobiology* **2016**, 26, 888-903.
- [20] a) E. M. Danielsen and G. H. Hansen, *Mol. Membr. Biol.* **2006**, 23, 71-79; b) Y. Zick, M. Eisenstein, R. A. Goren, Y. R. Hadari, Y. Levy and D. Ronen, *Glycoconjugate J.* **2004**, 19, 517-526.
- [21] a) H. Ideo, A. Seko, I. Ishizuka and K. Yamashita, *Glycobiology* **2003**, 13, 713-723; b) O. A. Vokhmyanina, E. M. Rapoport, S. André, V. V. Severov, I. Ryzhov, G. V. Pazynina, E. Korchagina, H.-J. Gabius and N. V. Bovin, *Glycobiology* **2012**, 22, 1207-1217.
- [22] a) H. Ideo, A. Seko, T. Ohkura, K. L. Matta and K. Yamashita, *Glycobiology* **2002**, 12, 199-208; b) H. Ideo, Y. Hinoda, K. Sakai, I. Hoshi, S. Yamamoto, M. Oka, K. Maeda, N. Maeda, S. Hazama, J. Amano and K. Yamashita, *International journal of cancer. Journal international du cancer* **2015**, 137, 1652-1660.
- [23] a) H. F. Valenzuela, K. E. Pace, P. V. Cabrera, R. White, K. Porvari, H. Kaija, P. Vihko and L. G. Baum, *Cancer research* **2007**, 67, 6155-6162; b) E. M. Yazawa, J. E. Geddes-Sweeney, F. Cedeno-Laurent, K. C. Walley, S. R. Barthel, M. J. Opperman, J. Liang, J. Y. Lin, T. Schatton, A. C. Laga, M. C. Mihm, A. A. Qureshi, H. R. Widlund, G. F. Murphy and C. J. Dimitroff, *The Journal of investigative dermatology* **2015**, 135, 1849-1862.
- [24] G. García Caballero, A. Flores-Ibarra, M. Michalak, N. R. Khasbiullina, N. V. Bovin, S. André, J. C. Manning, S. Vertesy, F. M. Ruiz, H. Kaltner, J. Kopitz, A. Romero and H.-J. Gabius, *Biochimica et Biophysica Acta* **2016**, 128-129, 34-47.
- [25] C. Barrès, L. Blanc, P. Bette-Bobillo, S. André, R. Mamoun, H.-J. Gabius and M. Vidal, *Blood* **2010**, 115, 696-705.
- [26] a) G. Garcia Caballero, A. Flores-Ibarra, M. Michalak, N. Khasbiullina, N. V. Bovin, S. Andre, J. C. Manning, S. Vertesy, F. M. Ruiz, H. Kaltner, J. Kopitz, A. Romero and H. J. Gabius, *Biochimica et biophysica acta* **2016**, 1860, 2285-2297; b) H. Kaltner, G. Garcia Caballero, F. Sinowatz, S. Schmidt, J. C. Manning, S. Andre and H. J. Gabius, *Biochimica et biophysica acta* **2016**, 1860, 2298-2312.
- [27] a) H.-J. Gabius, C. Schröter, S. Gabius, U. Brinck and L.-F. Tietze, *J. Histochem. Cytochem.* **1990**, 38, 1625-1631; b) L. F. Tietze, C. Schroter, S. Gabius, U. Brinck, A. Goerlach-Graw and H. J. Gabius, *Bioconjugate chemistry* **1991**, 2, 148-153.
- [28] D. Solís, N. V. Bovin, A. P. Davis, J. Jiménez-Barbero, A. Romero, R. Roy, K. Smetana Jr. and H.-J. Gabius, *Biochimica et biophysica acta* **2015**, 1850, 186-235.

- [29] R. Roy, P. V. Murphy and H. J. Gabius, *Molecules* **2016**, *21*, 629.
- [30] a) S. Vértessy, M. Michalak, M. C. Miller, M. Schnölzer, S. André, J. Kopitz, K. H. Mayo and H.-J. Gabius, *Protein engineering, design & selection : PEDS* **2015**, *28*, 199-210; b) S. Zhang, R.-O. Moussodia, C. Murzeau, H. J. Sun, M. L. Klein, S. Vértessy, S. André, R. Roy, H.-J. Gabius and V. Percec, *Angewandte Chemie* **2015**, *54*, 4036-4040.
- [31] H. Tateno, A. Mori, N. Uchiyama, R. Yabe, J. Iwaki, T. Shikanai, T. Angata, H. Narimatsu and J. Hirabayashi, *Glycobiology* **2008**, *18*, 789-798.
- [32] N. Agrwal, Q. Sun, S. Y. Wang and J. L. Wang, *The Journal of biological chemistry* **1993**, *268*, 14932-14939.
- [33] a) G. Garcia Caballero, H. Kaltner, M. Michalak, N. Shilova, M. Yegres, S. Andre, A. K. Ludwig, J. C. Manning, S. Schmidt, M. Schnolzer, N. V. Bovin, D. Reusch, J. Kopitz and H.-J. Gabius, *Biochimie* **2016**, *128-129*, 34-47; b) H.-J. Gabius, B. Wosgien, M. Hendrys and A. Bardosi, *Histochemistry* **1991**, *95*, 269-277; c) M. Amano, H. Eriksson, J. C. Manning, K. M. Detjen, S. André, S.-I. Nishimura, J. Lehtiö and H.-J. Gabius, *FEBS J* **2012**, *279*, 4062-4080.

Figure 1. Overview of tested galectins and the chicken galectin-like protein (C-GRP) belong to three types of galectin architecture: a) Chimera-type protein which is composed of three parts, the carbohydrate recognition domain (CRD), proline/glycine-rich collagen-like repeat section, and N-terminal peptide with two sites for Ser phosphorylation, the latter two forming the N-terminal tail. Human galectin-3 can form oligomers via the CRD and N-terminal tail. b) Tandem-repeat-type protein constituted by two non-identical CRDs at the N- and C-terminal sides joined by a linker peptide. c) Proto-type proteins consisting of a single type of CRD. Human galectin-1 and the chicken galectin-related *inter-fiber protein* (C-GRIFIN) form homodimers. Rat galectin-5 and C-GRP have an N-terminal extension, both monomeric.

Figure 2. Glycoblotting-based microarray for galectin profiling. (A) Synthetic MUC-1 (glyco) peptides **1-23** and **31-35**, and glycoamino acids **24-30** used for microarray analysis. The two types of linkers, as well as peptide backbones, are represented as a grey (compounds **1-30**) and a green, (compounds **31-35**) background. Compounds **24-30** correspond to *O*-glycosylated serine or threonine amino acids with α -GalNAc (Tn), core 1 (TF) and core 2 glycans. (B) Compounds **1-35** were robotically printed to give quadruplicates at four concentrations (12.5, 25, 50 and 100 μ M) on an amphiphilic plastic slide.^[6a] Afterwards, a silicone rubber sheet with three chambers was attached. Green spots correspond to spotted cyanine 3 N-hydroxysuccinimide ester (Cy3-NHS), used as fluorescence positive control.

Figure 3. Fluorescence images of microarray chip after treatment with galectin-3 (A) and its CRD (B). Relative fluorescence units (RFU) with its standard deviation for compounds **1-35** after exposure to galectin-3 (C).

Figure 4. Fluorescence images of microarray chip after treatment with galectin-8 (A), galectin-4 (B), Gal-4N (C), and Gal-4C (D). Relative fluorescence units (RFU) with its standard deviation for compounds **1-35** after exposure to galectin-8 (E), galectin-4 (F), Gal-4N (G), and Gal-4C (H).

Figure 5. Fluorescence images of microarray chip after treatment with galectin-1 (A), C-GRIFIN (B), galectin-5 (C), and C-GRP (D). Relative fluorescence units (RFU) with

its standard deviation for compounds **1-35** after exposure to C-GRIFIN (**E**) and galectin-5 (**F**).

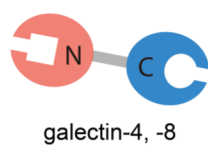
Figure 6. Overview of interaction strength between tested galectins (and galectin-like protein) and glycopeptides (and glycoamino acids). The interaction strength were classified into six grades (see Table S2 in supporting information).

Figure 1.

a) Chimera-type



b) Tandem-repeat-type



c) Proto-type



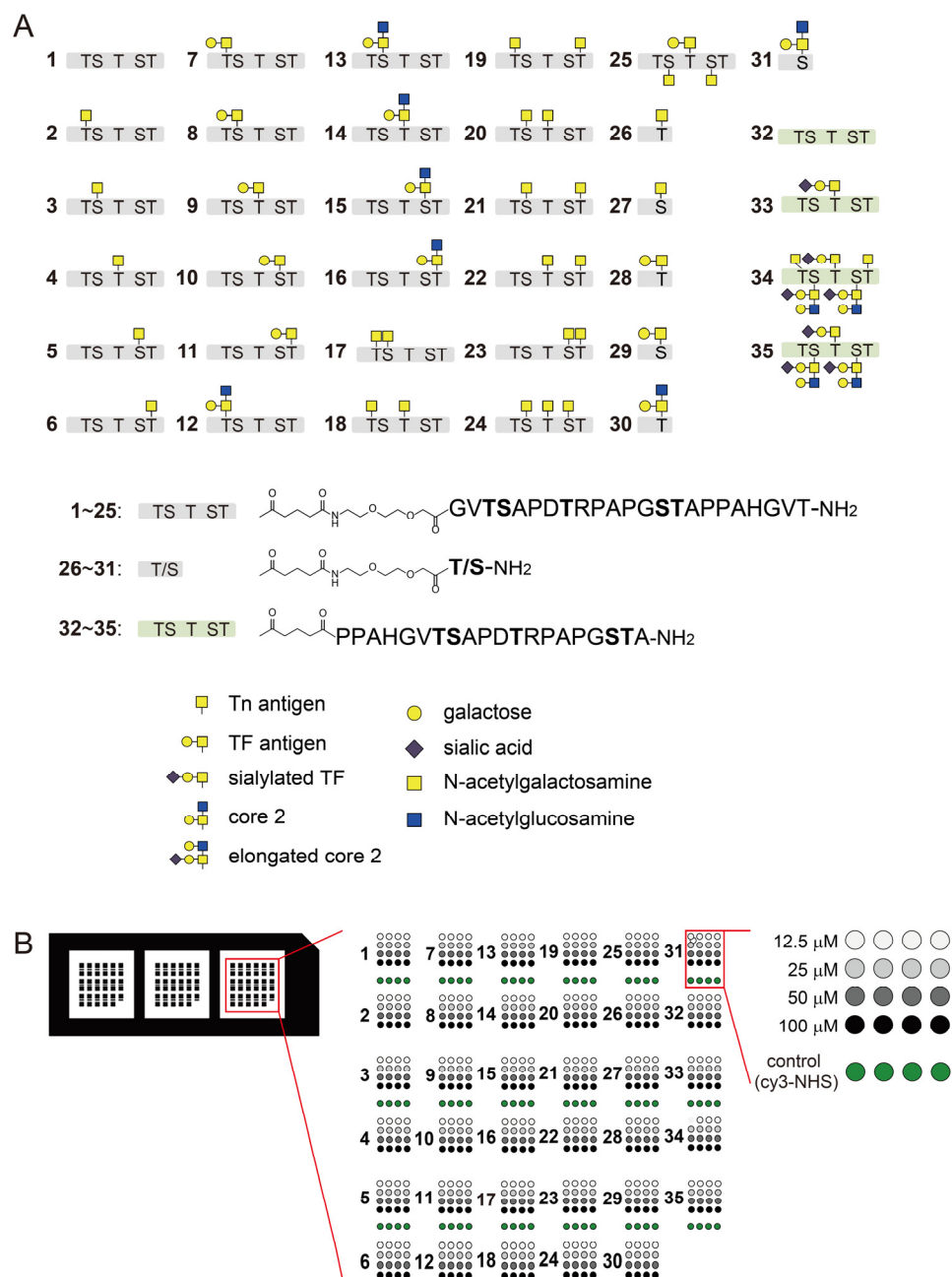
Figure 2.

Figure 3.

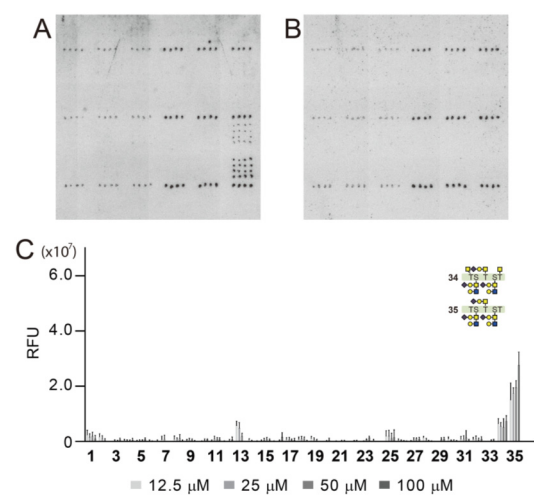


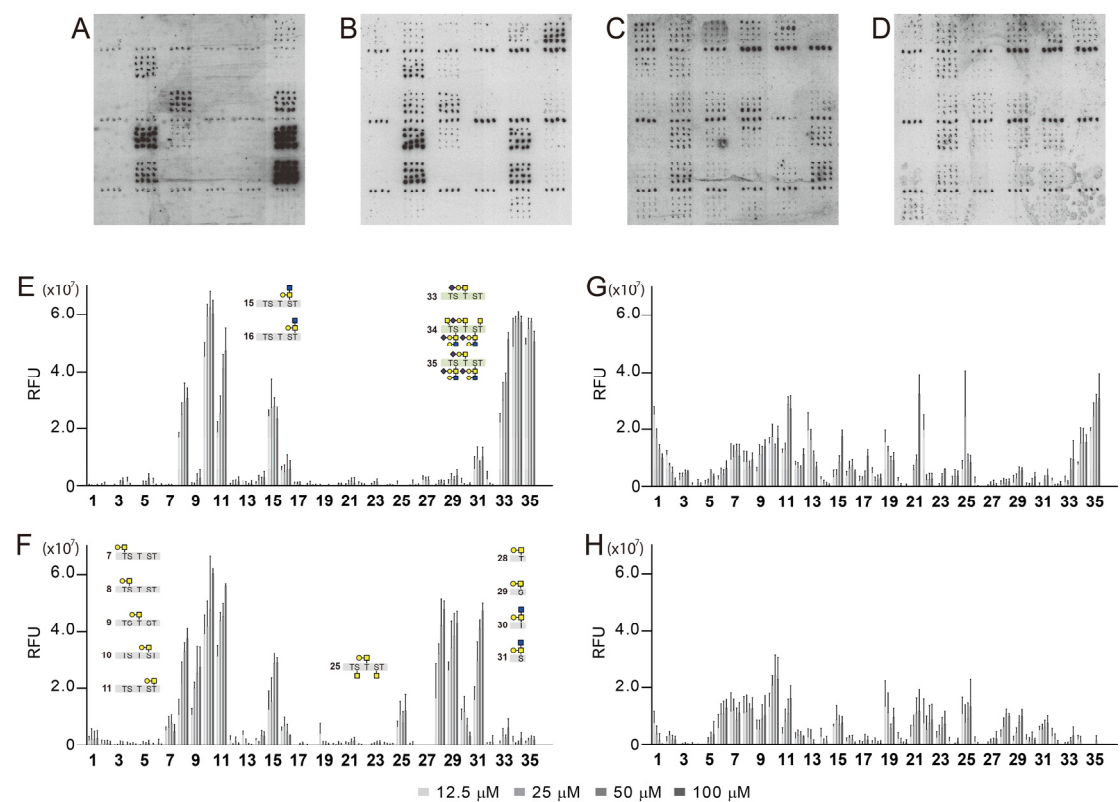
Figure 4.

Figure 5.

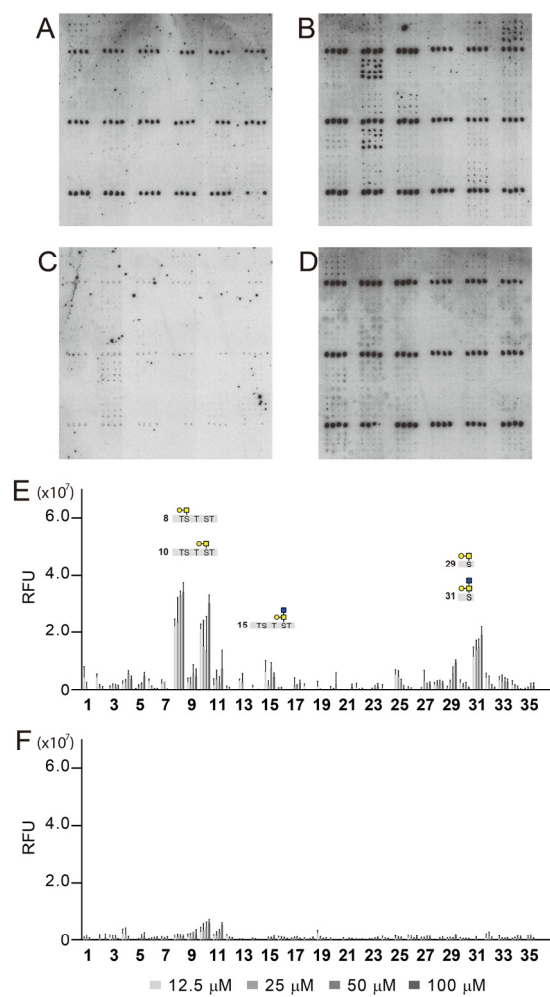


Figure 6.