Glycoconjugate Journal 21, 35–40, 2004

Lectin-based structural glycomics: Glycoproteomics and glycan profiling

Jun Hirabayashi

Glycostructure Analysis Team, Research Center for Gycoscience, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 6, 1-1-1, Higashi, Tsukuba, Ibaraki 305-8566, Japan

Structural glycomics (SG) plays a fundamental part of concurrent glycobiology aiming at comprehensive elucidation of glycan functions (*i.e.*, functional glycomics) in the context of post-genome sciences. The SG project started in April 2003 and will continue for 3 years in the framework of NEDO (New Energy and Industrial Technology Organization) under the METI (the Ministry of Economy, Trade, and Industry), Japan. The main purpose of the project is the development of high-throughput and robust machines, which should greatly contribute to the structural analysis of complex glycans. In this chapter, 2 major research items, *i.e.*, (1) glycoproteomics, which enables comprehensive analysis of glycoproteins, and (2) "glycan profiling" by means of lectins, are described. For the latter, frontal affinity chromatography has been adopted as a starting tool for comprehensive analysis of the interaction of 100 lectins and 100 oligosaccharides under the concept of "hect-by-hect," which refers to 100 × 100.
Published in 2004.

Keywords: **glycoproteomics, glyco-catch method, lectin, glycan profiling, frontal affinity chromatography**

Introduction

Glycans are the third group of bio-informative macromolecules, whose structures and functions are to be elucidated in various contexts of post-proteome sciences. Apparently, the most important issue facing glycobiology is how rapidly, efficiently, sensitively, and accurately complex glycan structures can be analyzed in a high-throughput manner. In this chapter, I describe 2 research subjects involved in the SG (Structural Glycomics) project, which is proceeding under the control of the New Energy and Industrial Technology Development Organization (NEDO). One is "glycoproteomics," aimed at comprehensive analysis of glycoproteins; and the other is "glycan profiling," for extracting core information about glycan structures by means of lectin affinity technologies. The essence of these subjects was previously stated, in which studies the model organism *Caenorhabditis elegans* was targeted [1]. Each subject described in this chapter is intended to develop high-throughput machines for analysis of the complex carbohydrate structures of much higher organisms including human beings under the concept of "glycoproteomics," which should lead eventually to a complete understanding of the actual world of the glyco-clothed proteome.

To whom correspondence should be addressed: Jun Hirabayashi, Tel.: 81-29-861-6530; Fax: 81-29-861-6083; E-mail: jun-hirabayashi@aist.go.jp

Glycoproteomics

Glycoproteomics plays an "entry" part in the SG project. Actually, the subject is divided into 3 sub-themes achieved by different research groups that collaborate with core laboratories in The Research Center for Glycoscience in Tsukuba (RCG Tsukuba): (1) *proteome strategy group*, (2) *glycome strategy group*, and (3) *glycoform strategy group*. The proteome strategy is being carried out for large-scale analysis at Tokyo Metropolitan University based on the established procedure (described below), whereas the glycome strategy is being implemented at the RCG Tsukuba for the development of a total system enabling both identification of glycoprotein genes and profiling of glycans attached to each glycosylation site. For this realization, automated and versatile machines are being developed in collaboration with GL Sciences Inc. (Iruma, Japan). On the other hand, the glycoform strategy group is developing an automated glycan liberation machine for glycoform analysis applicable to both *N*- and *O*-glycoproteins in collaboration with Kinki University and Shimadzu Corporation.

The proteome strategy occupies a fundamental part in that it collects basic information about mammalian glycoproteins in a comprehensive manner by using a recently established procedure for analysis of glycoproteins (Figure 1A) [2]. The procedure is actually a combination of the recently developed lectin-affinity technology named "glyco-catch" [3] and an advanced proteome strategy, "2D-LC/MS/MS" (2-dimensional

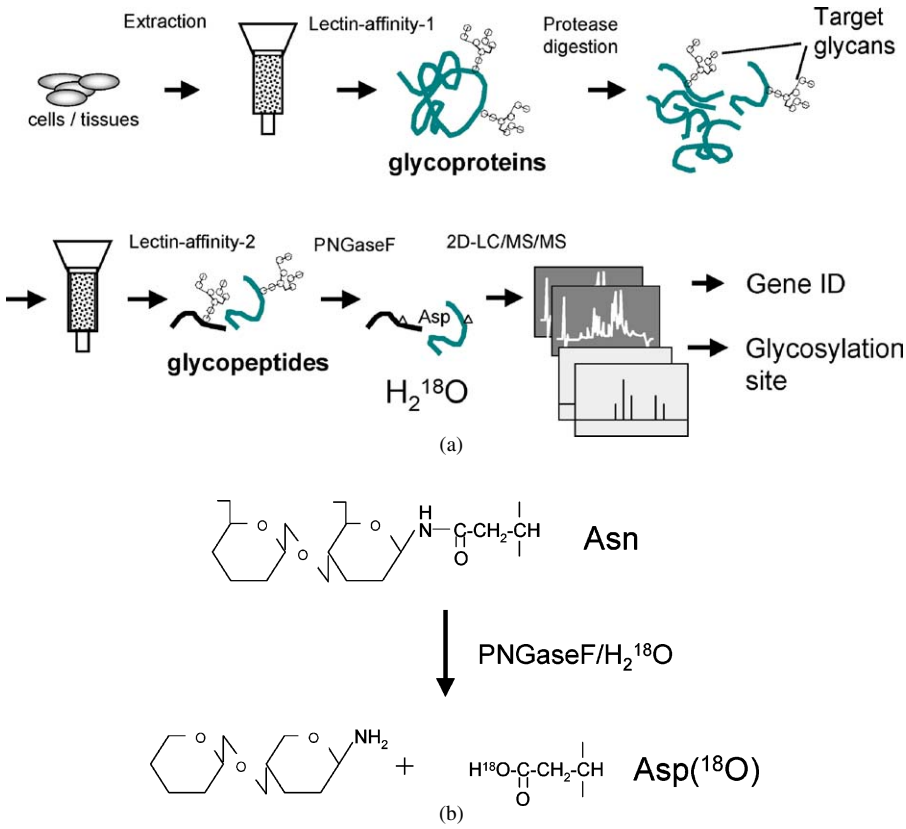

Figure 1. (A) Schema showing glyco-catch procedure for comprehensive identification of glycoproteins. The essence of the strategy is selective capturing of glycopeptides, which are generated by proteolysis and then subjected to conventional proteomic procedures to identify core protein genes in genome databases as well as precise determination of N-glycosylation sites by the special use of peptide-N-glycosidase F (PNGaseF) in the presence of $H_2\,^{18}O$. (B) Reaction scheme of PNGaseF. Since the enzyme is actually an amidase rather than a glycosidase, the ^{18}O atom is introduced into an aspartic acid residue, which makes a useful tag to identify actual N-glycosylated sites.

liquid chromatography/mass spectrometry/mass spectrometry) [4]. In this method, various lectins are used to capture glycopeptides after proteolytic cleavage of isolated glycoproteins, and then the captured glycopeptides are analyzed in the context of focused proteomics. Here, the method must clarify the nature of the glycan moieties, because their structures are not known in advance; and thus, conventional proteome strategy cannot be easily applied. For this solution, a conventional glycotechnological method, *i.e.*, the use of peptide *N*-glycanase F (PNGaseF) in the presence of $H_2\,^{18}O$ [5], was adopted to incorporate stable isotope ^{18}O into each *N*-glycosylation site (designated isotope-coded glycosylation-site-specific tagging: IGOT). Thus, high-throughput, sensitive (by means of LC-MS/MS), accurate (by means of PNGaseF and ^{18}O), and focused (by means of glyco-catch method) analysis of glycoproteins has become quite possible (Figure 1A). Kaji *et al.* applied this strategy for the first time to a simple model organism, *Caenorhabditis elegans* [2]. In the SG project, we focus on targets from much higher organisms including human beings, whose genome sequences are known, by using various lectins with various sugar specificities, *e.g.*, ConA for high-mannose type N-glycans, RCA-I for galactosylated complex type N-glycans, PSA for core-fucosylated N-glycans, SSA for α2,6-sialylated N-glycans, etc.

An apparent disadvantage of the glyco-catch method, followed by IGOT, is its inapplicability to O-glycopeptides, because no universal glycosidase to liberate all kinds of O-glycans is available. Even if any, the above IGOT strategy cannot be applied to identify glycosylation sites, because the stable isotope ^{18}O are not incorporated into the peptide portion like in *N*-glycopeptides (Figure 1B). For systematic identification of mucin-type O-glycoproteins, a novel method using a synthetic donor analogue of GalNAc to be incorporated into polypeptides in appropriate cell culture systems was recently reported [6], though the strategy is not directly applied for natural tissues. Another possible disadvantage of the glyco-catch method is the rather biased collection of glycopeptides, even though various kinds of lectins are used. In this regard, a recently developed procedure can be an alternative to our glyco-catch method [7]. In this context, Aebersold and his co-workers reported a distinct procedure for broader collection of glycoproteins by use of a chemical capture method consisting of (1) periodate oxidation of *cis*-diol moieties of glycans, which occur mostly at

non-reducing terminal ends in glycoproteins, and (2) covalent coupling of the generated aldehyde groups to hydrazide beads. Their following strategy, *i.e.*, identification of peptides by 2D-LC/MS/MS after removal of glycans, is almost the same as in our procedure, though they do not take advantage of the stable isotope ^{18}O to assure determination of N-glycosylation sites. Aebersold's strategy is in principle more comprehensive than the glyco-catch. On the other hand, the latter procedure was developed originally to identify glycoprotein receptors for endogenous lectins, such as galectins [8–10], selectins [11], and siglecs [12]. From a biological viewpoint, "focused glycoproteomics" can be more meaningful over the chemical-catch method. Moreover, the latter also has practical difficulty in identifying poorly expressed glycoproteins whose functions are important. From a global viewpoint, the two procedures can be compensative to each other, and thus, should better be used in a more appropriate context depending on each research objective.

Glycan profiling

Another subject of the SG project described in this chapter is glycan profiling using lectins. MS is expected to play a dominant role in profiling complex glycan structures in a sensitive and accurate manner by means of MS^n. However, an apparent difficulty of MS is discrimination of a number of structural isomers, such as enantiomers (*e.g.*, D/L-glucose), diastereomers (galactose/mannose), anomer (α/β), linkage (1–3/1–4), and position isomers (1–3 branch/1–6 branch). Of course, the SG project challenges the theme from a fundamental aspect. On the other, hand, the use of lectins in glycan profiling is of significant advantage to discriminate the above isomers, because these lectins serve as "glycan decipherers" *in vivo*.

The subject of glycan profiling consists of 2 research projects, *i.e.*, (1) *frontal affinity chromatography* (*FAC*) and (2) *lectin arrays*. FAC takes a fundamental part of glycan profiling by providing basic data of lectin-oligosaccharide interaction under the "hect-by-hect" project, which intends to determine the dissociation constants (K_d's) between 100 lectins and 100 oligosaccharides. For this purpose, a proto-type automated FAC machine (FAC-1) has already been developed in collaboration with Shimadzu Corporation (Kyoto, Japan). The other project, "lectin arrays," aims at coincidental detection of multiple (*e.g.*, > 50) lectin-oligosaccharide interactions, which cannot be done by FAC. For this purpose, various principles should be applied; but the project focuses, as the first choice, on near-field optic technology in the form of an evanescent wave-excited fluorescence detection system [13–15] in collaboration with Nippon Laser and Electronics Lab. (Nagoya, Japan). In this chapter, the hext-by-hect project by FAC is described.

FAC

Frontal affinity chromatography (FAC) is a quantitative affinity chromatography originally developed by Kasai and Ishii 30 yeas ago [16]. FAC is a qualified technique enabling precise determination of K_d's of biomolecules based on a simple principle analogous to enzyme kinetics (Figure 2A) [17]. Its application to various biomolecules has been reported: *i.e.*, proteases and substrate analogues [18,19], antigens and antibodies [20], and lectins and oligosaccharides [21–24]. However, previous procedures required a considerably long time for each analysis (usually several hours) and relatively large amounts of ligands for immobilization and free analytes for elution. For detection of the eluted analytes, radio-labeling is necessary in some cases. To overcome these problems, major improvement was recently achieved independently in 2 laboratories: Hindsgaul and his coworkers made a small and thin tubular column (*e.g.*, 0. 75 mm diameter $\times$150 mm in length), to which cholera toxin was immobilized. As an innovative point, they connected this column to an electron spray ionization-mass spectrometer for on-line detection of oligosaccharide derivatives. With this system, designated FAC/MS, various chemical compounds synthesized by a series of combinatorial chemistry can be screened for binding to cholera toxin [25]. Hirabayashi *et al.* reinforced the system in a form more specialized for precise analysis of lectin-oligosaccharide interaction by using a fluorescence detection (FD) system [26]. It should be particularly emphasized that their system enables immediately the use of a considerable number of ($>$100) fluorescently labeled standard oligosaccharides, such as the pyridylaminated (PA) ones originally developed by Hase *et al.* [27]. PA is advantageous over other fluorescence labeling methods in its chemical stability, sensitivity, and resolution in multi-dimensional separation systems using HPLC (high-performance liquid chromatography) [28]. Another advantage of the FAC/FD is its accuracy: to determine elution volume (V) of the elution front, a calculation program is available [29]. By using this reinforced FAC system, comprehensive analysis of the interaction between 41 galactose-containing oligosaccharides and 13 animal galectins was made [8]. With this system, functional analysis of galectins was also done [30,31] as well as comprehensive analysis [32].

Hect-by-hect project

In the SG project, a specific machine for FAC/FD has been developed in order to improve the above system for greater high-throughput. For this purpose, further miniaturization of the column system (the above system: 4×10 mm, gourd column type; the present system: 2×10 mm, capsule column type) and incorporation of an auto-sampling system were achieved. Since the proto-type machine is equipped with 2 analysis columns in parallel, it can process more than 100 interaction analyses in 10 h (Figure 2B). Thus, comprehensive lectin-oligosaccharide analysis has become quite possible from both systematic and comparative viewpoints. This recently developed proto-type machine (FAC-1) is the initial step toward the ultimate machine for FAC/FD aiming at versatile profiling of various lectins. Thus, comprehensive analysis, *e.g.*, between 100 lectins and 100 standard oligosaccharides are being undertaken in the context of the

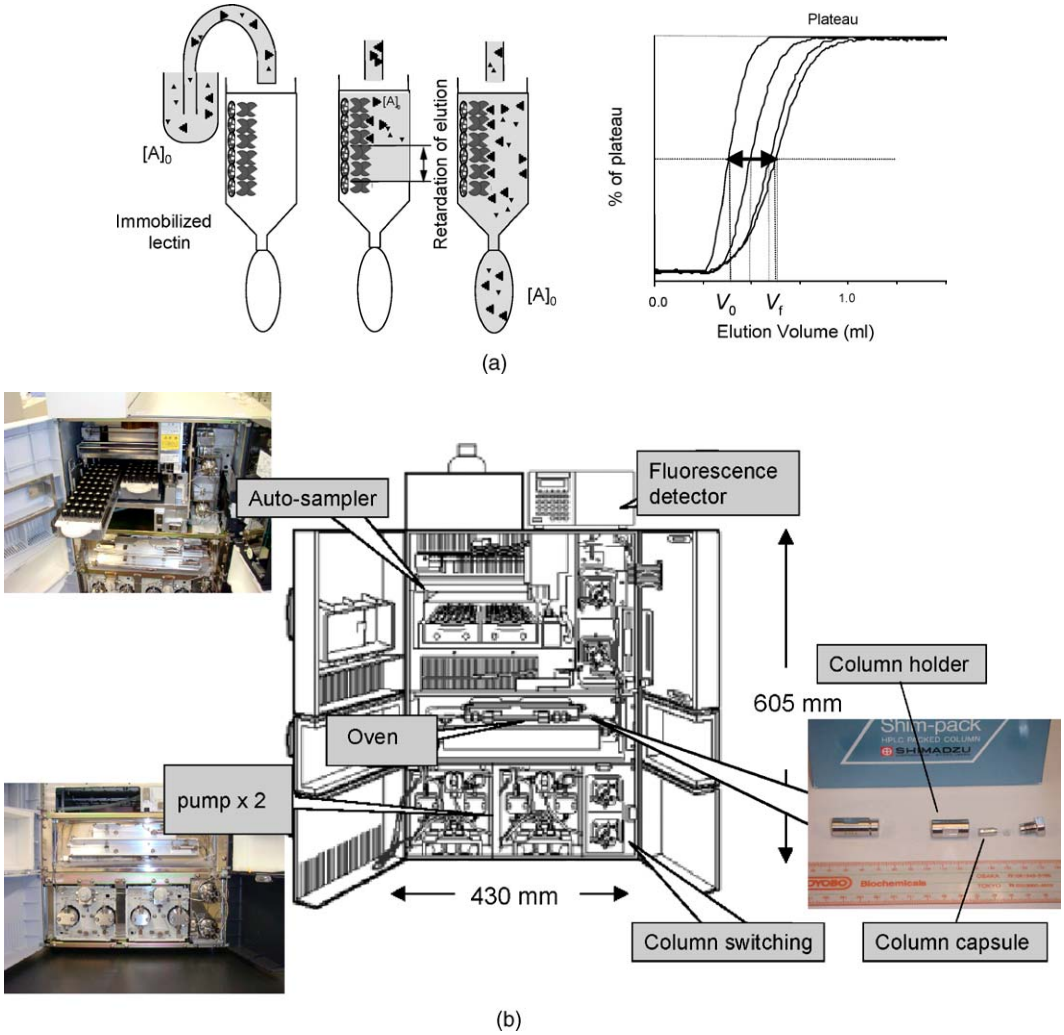

Figure 2. (A) Procedure for frontal affinity chromatography. To a column, which is packed with lectin-immobilized resin (*e.g.*, agarose), an excess amount of diluted saccharide solution is applied continuously. If the lectin has significant affinity to the saccharide, the elution front is retarded relative to that of saccharine, which has no affinity for the lectin. Elution volumes of the former and latter are defined as V_0 and V_f, respectively. (B) Outline of a prototype automated FAC machine. The machine consists of an auto-sampler, dual pumps, oven, column switching system, and fluorescence detector. Capsule-type column size is 2.0 mm in i.d. $\times$10 mm (column volume, 31.4 μl).

"hect-by-hect" project. For the first step to carry out the project, a panel of 49 oligosaccharides was used for brief profiling of individual lectins (Figure 3). The thus selected core oligosaccharides represented glycotope structures such as biantennary, triantennary, tetraantennary, core fucosylated, bisected GlcNAc, α2-3/6 sialylated, αGal, αGalNAc, type 1/2, Le^a, Le^X, etc. The project is also supported by various groups, who provide more complex oligosaccharides and non-commercial lectins, *e.g.*, milk oligosaccharides (Noguchi Institute; Tokyo, Japan), glycosaminoglycan oligosaccharides (Seikagaku Corporation; Higashiyamato, Japan), glycopeptides (RCG Tsukuba and RCG Hokkaido), native lectins (Honen Corporation), and mutated lectins (Tokyo University). Thus, the hect-by-hect project should yield invaluable information regarding lectin-oligosaccharide (glycopeptide) interaction, which is essential for constructing glycan profiler. However, direct application of native lectins, in particular animal lectins, might be limited from various aspects, *e.g.*, stability, availability, specificity. So, in future, development of "reinforced lectins" should be necessary for a more practical purpose.

Databases

Lastly, data obtained on all of the above research subjects, *i.e.*, glycoproteomics (described in this chapter), MS, and glycan profiling (described in this chapter) are summarized in the forms of "glycoprotein database," "oligosaccharide database," and "lectin database" (Figure 4), respectively, with the aid of Mitsui Knowledge Industry, Co. Ltd. (Tokyo, Japan). After completion of the SG project, these databases will be opened to

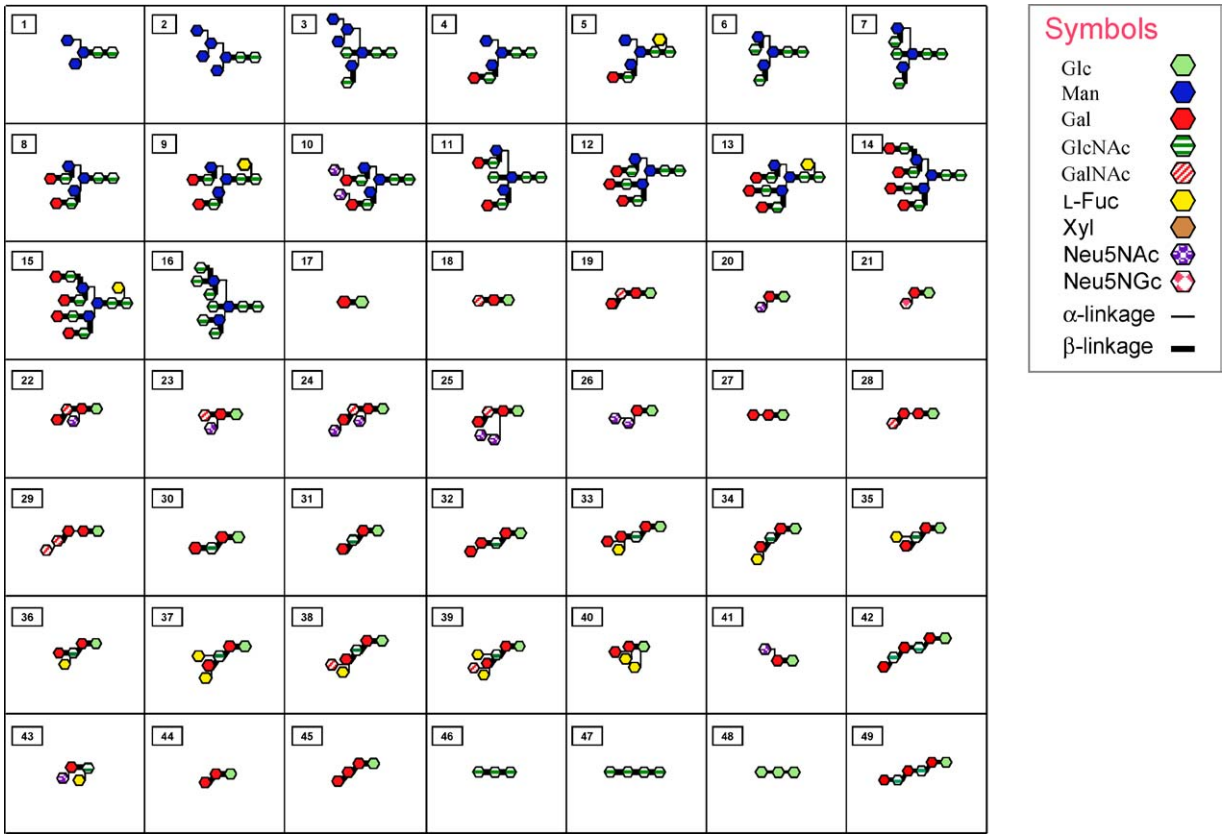

Figure 3. A panel of 49 core oligosaccharides used in the "hect-by-hect" project.

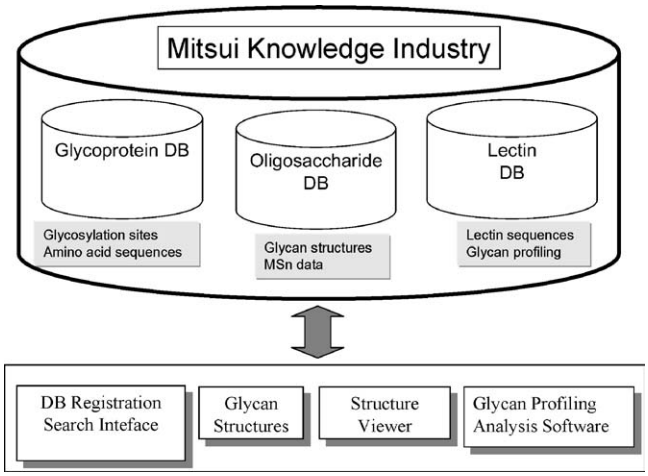

Figure 4. A conceptual image of structural glycome database. The databases are composed of glycoprotein database, oligosaccharide database, and lectin database.

the scientific community. Hopefully their availability to many other scientists will lead to the development of even more powerful approaches to accumulate further experimental data on glycan structures and to devise other software for the purpose of elucidating glycan functions.

Acknowledgment

This work is supported by NEDO (New Energy and Industrial Technology Organization) under the METI (The Ministry of Economy, Trade and Industry, Japan).

References

1 Hirabayashi J, Arata Y, Kasai K, Glycome project: Concept, strategy and preliminary application to *Caenorhabditis elegans*, *Proteomics* **1**, 295–303 (2001).

2 Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T, Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins, *Nat Biotechno* **21**, 667–72 (2003).

3 Hirabayashi J, Hayama K, Kaji H, Isobe T, Kasai K, Affinity capturing and gene assignment of soluble glycoproteins produced by the nematode *Caenorhabditis elegans*, *J Biochem (Tokyo)* **132**, 103–14 (2002).

4 Mawuenyega KG, Kaji H, Yamuchi Y, Shinkawa T, Saito H, Taoka M, Takahashi N, Isobe T, Large-scale identification of *Caenorhabditis elegans* proteins by multidimensional liquid chromatography-tandem mass spectrometry, *J Proteome Res* **2**, 23–35 (2003).

5 Gonzalez J, Takao T, Hori H, Besada V, Rodriguez R, Padron G, Shimonishi Y, A method for determination of N-glycosylation sites in glycoproteins by collision-induced dissociation analysis in fast atom bombardment mass spectrometry: Identification of the

positions of carbohydrate-linked asparagine in recombinant alpha-amylase by treatment with peptide-N-glycosidase F in ^{18}O-labeled water, *Anal Biochem* **205**, 151–8 (1992).
6 Hang HC, Yu C, Kato DL, Bertozzi CR, A metabolic labeling approach toward proteomic analysis of mucin-type *O*-linked glycosylation, *Proc Natl Acad Sci USA* **100**, 14846–51 (2003).
7 Zhang H, Li XJ, Martin DB, Aebersold R, Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry, *Nat Biotechnol* **21**, 660–6 (2003).
8 Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Muller WE, Yagi F, Kasai K, Oligosaccharide specificity of galectins: A search by frontal affinity chromatography, *Biochim Biophys Acta* **1572**, 232–54 (2002).
9 Kuwabara I, Sano H, Liu FT, Functions of galectins in cell adhesion and chemotaxis, *Methods Enzymol* **363**, 532–52 (2003).
10 Houzelstein D, Goncalves IR, Fadden AJ, Sidhu SS, Cooper DN, Drickamer K, Leffler H, Poirier F, Phylogenetic analysis of the vertebrate galectin family, *Mol Biol Evol* (in press)
11 McEver RP, Interactions of selectins with PSGL-1 and other ligands. *Ernst Schering Res Found Workshop* **44**, 137–47 (2004).
12 Crocker PR, Zhang J, New I-type lectins of the CD 33-related siglec subgroup identified through genomics, *Biochem Soc Symp* **69**, 83–94 (2002)
13 Oheim M, Loerke D, Chow RH, Stuhmer W, Evanescent-wave microscopy: A new tool to gain insight into the control of transmitter release. *Philos Trans R Soc Lond B Biol Sci* **354**, 307–18 (1999).
14 Garland PB, Optical evanescent wave methods for the study of biomolecular interactions, *Q Rev Biophys* **29**, 91–117 (1996).
15 Hutchinson AM, Evanescent wave biosensors. Real-time analysis of biomolecular interactions. *Mol Biotechnol* **3**, 47–54 (1995).
16 Kasai K, Ishii S, Quantitative analysis of affinity chromatography of trypsin. A new technique for investigation of protein-ligand interaction, *J Biochem (Tokyo)* **77**, 261–4 (1975).
17 Hirabayashi J, Arata Y, Kasai K, Frontal affinity chromatography as a tool for elucidation of sugar recognition properties of lectins, *Methods Enzymol* **362**, 353–68 (2003).
18 Kazama M, Tahara C, Abe T, Kasai K, Quantitative analysis of fibrin-binding affinity of fibrinolytic components by frontal affinity chromatography, *Thromb Res Suppl* **8**, 81–90 (1988).
19 Kasai K, Trypsin and affinity chromatography, *J Chromatogr* **597**, 3–18 (1992)
20 Hearn MT, Davies JR, Evaluation of factors which affect column performance with immobilized monoclonal antibodies. Model studies with a lysozyme-anti lysozyme system, *J Chromatogr* **512**, 23–39 (1990).
21 Oda Y, Kasai K, Ishii S, Studies on the specific interaction of concanavalin A and saccharides by affinity chromatography, Application of quantitative affinity chromatography to a multivalent system, *J Biochem (Tokyo)* **89**, 285–96 (1981).
22 Ohyama Y, Kasai K, Nomoto H, Inoue Y, Frontal affinity chromatography of ovalbumin glycoasparagines on a concanavalin A-sepharose column. A quantitative study of the binding specificity of the lectin, *J Biol Chem* **260**, 6882–7 (1985).
23 Sueyoshi S, Tsuji T, Osawa T, Carbohydrate-binding specificities of five lectins that bind to *O*-glycosyl-linked carbohydrate chains. Quantitative analysis by frontal-affinity chromatography, *Carbohydr Res* **178**, 213–24 (1988).
24 Sueyoshi S, Yamamoto K, Osawa T, Carbohydrate binding specificity of a beetle (*Allomyrina dichotoma*) lectin, *J Biochem (Tokyo)* **103**, 894–99 (1988).
25 Zhang B, Palcic MM, Schriemer DC, Alvarez-Manilla G, Pierce M, Hindsgaul O, Frontal affinity chromatography coupled to mass spectrometry for screening mixtures of enzyme inhibitors, *Anal Biochem* **299**, 173–82 (2001).
26 Hirabayashi J, Arata Y, Kasai K, Reinforcement of frontal affinity chromatography for effective analysis of lectin-oligosaccharide interactions, *J Chromatogr A* **890**, 261–71 (2000).
27 Natsuka S, Hase S, Analysis of *N*- and *O*-glycans by pyridylamination. *Methods Mol Biol* **76**, 101–13 (1998).
28 Takahashi N, Nakagawa H, Fujikawa K, Kawamura Y, Tomiya N, Three-dimensional elution mapping of pyridylaminated *N*-linked neutral and sialyl oligosaccharides, *Anal Biochem* **226**, 139–46 (1995).
29 Arata Y, Hirabayashi J, Kasai K, Application of reinforced frontal affinity chromatography and advanced processing procedure to the study of the binding property of a *Caenorhabditis elegans* galectin, *J Chromatogr A* **905**, 337–43 (2001).
30 Arata Y, Hirabayashi J, Kasai K, Sugar binding properties of the two lectin domains of the tandem repeat-type galectin LEC-1 (N32) of *Caenorhabditis elegans*. Detailed analysis by an improved frontal affinity chromatography method, *J Biol Chem* **276**, 3068–77 (2001).
31 Pelletier I, Hashidate T, Urashima T, Nishi N, Nakamura T, Futai M, Arata Y, Kasai K, Hirashima M, Hirabayashi J, Sato S, Specific recognition of *Leishmania major* poly-beta-galactosyl epitopes by galectin-9: Possible implication of galectin-9 in interaction between L. major and host cells, *J Biol Chem* **278**, 22223–30 (2003).
32 Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Muller WE, Yagi F, Kasai K, Oligosaccharide specificity of galectins: A search by frontal affinity chromatography, *Biochim Biophys Acta* **1572**, 232–54 (2002).

Received 12 March 2004; accepted 21 May 2004