Contents

1 Why Glycomics with Lectins?

1.1 Inherent Properties of Glycans

Glycosylation endows a protein with various enhanced properties which the naked protein itself does not possess. Examples of these additional properties in protein glycosylation include polypeptide folding, stability, solubility, destination, functional regulation, and efficiency. In other words, glycosylation makes the role of a protein multidimensional [1, 2]. However, investigating protein glycosylation is not simple experimentally. Glycan structures are extremely diverse, for instance, in comparison with protein phosphorylation, another common post-translational modification. In general, glycans found in natural glycoproteins of higher animals consist of aldohexoses (e.g., Glc, Man, and Gal), their *N*-acetyl derivatives (e.g., GlcNAc and GalNAc), and, in many cases, deoxyhexose (e.g., L-fucose) and sialic acids represented by *N*-acetylneuraminic acid (NeuAc). An aldopentose, xylose, is also a common constituent of proteoglycan, and is involved in a linker region between core protein and glycosaminoglycan chains, such as heparin, heparan sulfate, dermatan sulfate, and chondroitin sulfate. However, the complexity of glycans is not strongly attributed to diversity in such component saccharides but more significantly to that in linkage isomers, because each hexose monosaccharide (e.g., Glc) possesses four potential donor hydroxyl groups (2-, 3-, 4-, and 6-OH) and one acceptor 1-hemiacetal (O–C–OH). The presence of anomerism (α/β) further doubles the number of linkage isomers. Indeed, there are eight possible isomers to

link two identical aldohexoses (e.g., Glc): "Glcα1-2/3/4/6Glc" and "Glcβ1-2/3/4/6Glc." The presence of multiple linkage isomers also allows branching in glycans, which is completely absent in nucleic acids and proteins. In theory, six monosaccharides can make as many as 1.05×10^{12} molecules of structural complexity, according to Laine [3]. Apparently, this figure far exceeds those calculated for hexanucleotides (i.e., $4^6 = 4{,}096$) and hexapeptides (i.e., $20^6 = 64{,}000{,}000$). Nobody knows the actual size of the glycome. However, it is considered that complete separation technology for glycans is not available. It also appears unrealistic currently to develop an automated method for glycan sequencing or synthesis [4, 5].

Because of the indirect association between genome and glycome, it is also difficult to predict glycan structures merely from the gene expression profiles: other information is necessary about the amounts of sugar nucleotide donors and location of a series of glycosyltransferases in the Golgi apparatus, some of which may compete with one another. However, the most important fact and the largest obstacle, which is never overcome by the classic concept workable for nucleic acids and proteins, is "heterogeneity" of glycans, partly because glycan biosynthesis is achieved only as a result of consecutive steps of glycosyltransferase and glycosidase reactions. These biosynthetic reactions can be incomplete or compete with one another. The extent of heterogeneity is further increased at the level of glycoconjugates (e.g., glycoproteins), because glycan structures and occupancies are not always the same between individual glycosylation sites. Of course, glycan structures should change if cell types and states that produce glycoproteins are different, even if the core protein structure is the same. This basic principle of glycoproteomics has been successfully applied to glycoprotein biomarker development for better diagnostic molecules [6, 7].

1.2 Lectin Microarray, an Advanced Technology for Glycan Profiling

Recent advances in separation and analytical technologies are noteworthy. They are represented by high-performance liquid chromatography (HPLC) and mass spectrometry (MS) as well as their combination with other separation principles (e.g., capillary electrophoresis, affinity chromatography). As described, however, it is still difficult for these methods to distinguish completely diverse structural isomers, and all of these conventional methods assume prior liberation of glycans and in most cases their separation and labeling. For the latter purpose, appropriate fluorescent tags (e.g., 2-aminopyridine and 2-aminobenzamide) are used because of increased sensitivity and resolution in HPLC. However, these processes require substantial labor and significant sacrifice of "throughput." Although a commonly-used method for glycan liberation is an enzymatic cleavage of glycans from the core proteins, the procedure has substantial limitation to a class of asparagine-linked

N-glycans. On the other hand, no universal glycosidase has been available for serine/threonine-linked *O*-glycans. As a more fundamental issue, some important biological aspects, e.g., density, depth, and orientation, as well as combination to other biochemical components (e.g., core protein, with which glycan functions are closely associated), would be lost if glycans were to be liberated from the core protein. Mucins are a representative case of such glycoconjugates.

Most conventional methods for glycomics are based on physicochemical principles. However, as emphasized previously, they necessitate separation of "liberated" glycans. This procedure depends largely on the type of glycosylation (i.e., *N*-glycomics, *O*-glycomics, glycosaminoglycomics, glycolipidomics, etc.). Obviously, this is a disadvantage for glycomics if the intention is the comprehensive analysis of all types of glycans/glycoconjugates in a comparative manner. In contrast to these conventional technologies, an alternative is to use biochemical "probes" to glycans, which have a substantial affinity to glycans/glycoconjugates, such as lectins and carbohydrate-binding antibodies. Indeed, various techniques using these glycan probes have already been shown to work in biochemical and cytochemical studies, e.g., lectin-probed western blot analysis, histochemistry, and flow cytometry. While these methods are widely applicable to crude biological samples (e.g., cells and sera), they have a great disadvantage in throughput, sensitivity, and speed, all of which are critical requisites for glycomics. An advanced method, lectin microarray, was developed in 2005 to overcome these drawbacks [8–13]. Lectin microarray has a platform similar to that of antibody microarray, consisting of multiple probes (lectins and carbohydrate-binding antibodies) on an appropriate array substrate (typically a glass slide), which enables simultaneous interaction analyses in an ultrasensitive and high-throughput manner. However, a special attribute of the lectin microarray is differential glycan profiling rather than identification of particular target molecules. It is essentially difficult for the lectin microarray to identify individual glycan structures [4, 14, 15]. Differential glycan profiling can also target a mixture of glycoproteins and clinical samples [16, 17]. Readers can find the essential aspects of the lectin microarray in recent comprehensive reviews (e.g., [18–21]). Therefore, the following are descriptions with a focus on original aspects of the lectin microarray developed in the authors' laboratory.

1.3 Lectins or Carbohydrate-Binding Antibodies?

According to the classic definition, lectins are carbohydrate-binding proteins, with the exception of antibodies and enzymes [22]. However, this criterion should be reconsidered in the light of current understanding of lectins: indeed, siglecs, a group of animal lectins specific for sialic acid, are members of the immunoglobulin superfamily, and an increasing number of enzymes have been found to contain carbohydrate-recognition domains (CRD) other than catalytic domains. On the other hand, it is now evident that even major lectin families, e.g., C-type lectins

and galectins, contain significant numbers of "non-lectin" members. In this context, it seems more reasonable to regard lectins simply as "a group of proteins that have significant ability to bind carbohydrates regardless of the protein families to which they belong." The number of lectins identified is rapidly increasing as glycotechnologies to prove lectin activity have improved [23–25]. Currently, the number of established lectin families exceeds 40, and this number will probably increase significantly in the future. Moreover, it is possible to create artificial lectins not only from natural lectins that have sugar-binding activity but also from those that have no such activity. In other words, all the existing proteins are able to become "novel" lectins if they have potential binding pockets and if they are modified appropriately to make up a binding network to target carbohydrates, e.g., through hydrogen bonds and van der Waals contacts. Another requisite to evolve lectins might be multivalency in various terms; e.g., tandem repeat of CRDs or binding modules, subunit oligomerization, and acquisition of subsite specificity. Siglecs are an example of novel lectins, which have evolved from the immunoglobulin superfamily [26]. The opposite probably applies to galectins-10, 11, 13, and 15, because they have no evident β-galactose-binding activity despite their original definition [27].

As described, both antibodies and lectins are useful tools for glycan profiling. However, one should keep in mind that protein structures are significantly different between species, e.g., human and mouse, whereas many glycan structures are common between these species. For instance, high-mannose structures of *N*-linked glycans are strongly conserved in eukaryotes (e.g., mammals) and, thus, are not antigenic in any animal. On the other hand, representative xenoantigens, the α-Gal epitope and *N*-glycolylneuramic acid (NeuGc) are absent in humans and, thus, are antigenic. However, these xenoantigens are rather exceptional among a large number of common structures. This is a basic reason why only relatively poor anti-carbohydrate antibodies are produced in animals, most of which are attributed to IgM. However, for the sake of differential glycan profiling, rigorously specific antibodies do not work. Rather, probes such as conventional lectins with broader specificity are desirable to ensure the "coverage" of the glycome. Importantly, many of the conventional lectins show significant cross-affinity to structurally-related glycans, but to different extents. On the other hand, in order to detect specific structures (epitopes), carbohydrate-binding lectins or antibodies with rigorously defined specificity should be effective as epitope detectors [28]. The method is also expected to provide an extremely high-throughput means for glycan profiling.

1.4 Preceding Techniques

A previously-developed technique called "serial lectin-affinity chromatography" is available. Although the technique is almost identical to the lectin microarray in its essence, it lacks throughput and speed, because it utilizes open columns and

radio-labeled oligosaccharides [29, 30]. Frontal affinity chromatography (FAC) is a more quantitative method for the analysis of lectin-glycan interactions [31]. The method was recently greatly improved in speed, throughput, and accuracy [32, 33]. Nevertheless, the method is essentially applicable to only purified oligosaccharides. On the other hand, the lectin microarray has a more suitable platform providing simultaneous interaction analyses with a panel of carbohydrate-binding proteins. Moreover, attaining high sensitivity is assured by utilizing an established procedure for labeling proteins with common fluorescent reagents (e.g., Cy3, Cy5). Although special equipment is required for reproducible analysis, especially of monovalent glycans, in many cases lectin-glycoprotein interactions are strong enough to tolerate even repeated washing [8, 9, 11–13]. However, in the case of glycoproteins having only a few glycan chains, and thus relatively weak affinity to a lectin, i.e., in the order of 10^{-3} to 10^{-5} M in dissociation constant (K_d), high-wash procedures after a probing reaction should remove the weak binding. In this case, it is advisable to take advantage of an evanescent-field activated fluorescence detector, which enables in situ observation of immobilized lectin-glycan/glycoprotein interactions while preserving equilibrium [10, 34]. However, substantial affinity enhancement is often observed on an appropriate array platform, particularly when multiple lectin–glycan interactions occur. This phenomenon is generally understood as a glycoside cluster effect, first demonstrated by Y. C. Lee [35], and is closely associated with biological functions of lectins under physiological conditions [36].

2 Basic Fabrication of the Lectin Microarray System

2.1 Equipment

Figure 1 shows a general scheme for glycan profiling using the lectin microarray. Commercial products for this purpose are available [37]: the system consists of the scanner GlycoStation™ Reader 1200 and LecChip™, both of which are produced by GP Biosciences, Co. Ltd. (Yokohama). An advantage is that the system utilizes an evanescent-field activated fluorescence detection principle (described later in greater detail).

2.2 Microarray Plate: LecChip™

LecChip™ Ver.1.0 is a commercial product for the lectin microarray: it contains 45 different lectins (for a list see [5] or [37]), which were carefully selected from a pool of 167 candidate lectins, taking into consideration their binding specificity, stability, and economy. Lectins are briefly classified based on monosaccharide

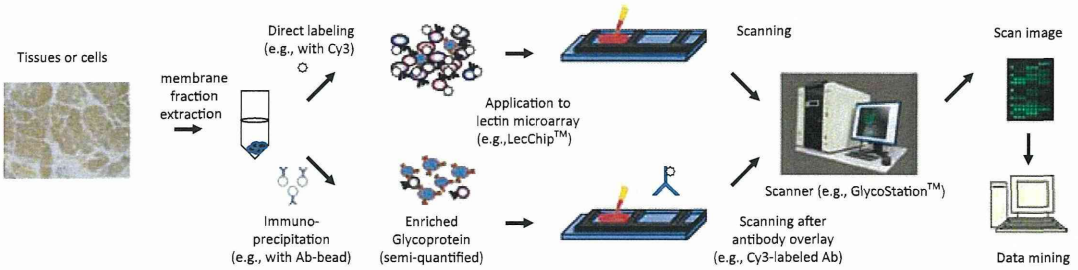

Fig. 1 General procedures of lectin microarray analysis: from sample preparation to data analysis. There are basically two schemes for glycan profiling; i.e., an *upper scheme* for direct profiling and a *lower scheme* for indirect profiling. In the former, all proteins prepared from cells and tissues are fluorescently labeled, e.g., Cy3, and their glycan profiles analyzed directly, while in the latter, target glycoproteins are immuno-precipitated prior to application to the microarray. The target glycoproteins are specifically detected by overlaying fluorescently-labeled detecting antibodies raised against the core protein moiety of the targets (for details, see text or Kuno et al. [28])

specificity; e.g., fucose binders (AAL, etc.), sialic acid binders (SSA, etc.), galactose binders (RCA120, etc.), mannose binders (ConA, etc.), and *O*-glycan binders (Jacalin, etc.), while some of them discriminate branching or more complex features (e.g., PHA-L). LecChip™ (ver.1.0) contains 7 wells on a glass slide, each containing the same set of 45 lectins spotted in triplicate. Basic information about these lectins is available on the website "Lectin frontier DataBase" (L*f*DB; http://riodb.ibase.aist.go.jp/rcmg/glycodb/LectinSearch), in which specificity data obtained by FAC [32, 33] are also available.

2.3 *Detection Principles*

1. *Confocal fluorescence detection method*

   Many researchers using the lectin microarray take advantage of a confocal fluorescence detection principle, because the method has already been adapted for established systems of DNA microarray [18, 19]. The method is simple, and has therefore been widely used. However, it requires washing of the microarray surface after the binding reaction, because a dried sample is required for the confocal-type detection system. One useful application of the system is multi-color detection using, e.g., Cy3 and Cy5, as described later.
2. *Evanescent-field activated fluorescence detection method*

   Kuno et al. developed a unique lectin microarray system, which utilizes an evanescent-field activated fluorescent detection principle. The evanescent wave is generated from the surface of the glass slide when an appropriate light originally introduced from both edges of the slide makes a total reflection [10]. Because the range of this wave is largely restricted to "near optic field (i.e., substantially <200 nm)," fluorescently labeled materials, such as Cy3-labeled glycoproteins, cannot be activated unless they form a significant complex on the array surface. As a result, the method requires no washing

procedures, which is not the case with a conventional confocal detection system. This specific feature enables in situ aqueous-phase detection without perturbing equilibrium conditions, which should contribute to both sensitive and reproducible analysis.

3. *Bimolecular fluorescence quenching and recovery method*

   Hamachi and coworkers manufactured a sophisticated detection principle called "bimolecular fluorescence quenching and recovery" (BFQR) [38]. They used a supramolecular hydrogel matrix, where fluorescent lectins were noncovalently fixed to act as a molecular recognition scaffold. Though the detection mechanism is rather complicated and not straightforward, it enables one to read out fluorescently a series of lectin–saccharide interactions without washing processes, in a manner similar to the evanescent-type detection principle. Moreover, the developed method BFQR does not require prior labeling of the target saccharides. The latter feature resembles that of a surface plasmon resonance detection principle used much more widely [39]. To the present, however, relatively simple, synthetic saccharide quenchers have been utilized, limiting the application of this technology to more complex samples, e.g., body fluids and tissue extracts.

3 General Procedures

A general procedure for glycan profiling experiments using the GlycoStation™ system has been described [37]. Although there might be a variety of sample forms depending on experimental purposes, e.g., sera, urine, cell lysates, and tissue extracts, the core experimental procedure is common to all of the sample forms: i.e., it comprises (1) sample preparation, (2) protein quantification, (3) labeling with an appropriate fluorescent reagent (e.g., Cy-3), (4) application of the labeled sample onto the lectin microarray plate, and (5) data analysis, essentially as described [40, 41]. General procedures are outlined in Fig. 1.

In the following section, a typical procedure using cell lysates and Cy-3 labeling is described, while variously modified procedures are also available, some of which are described later.

3.1 Sample Preparation: Cell Lysate

The following protocol assumes that a cell pellet is washed extensively with PBS, and is kept frozen at –80°C until use:

1. Melt the cell pellet gradually on ice
2. Add 1 mL of PBS-Tx (PBS containing 1% (w/v) Triton X-100) to the cell pellet and suspend the cells with a pipette

3. Disrupt the cells by sonication (1–2 min on ice)
4. Centrifuge the above sample at 14,000×*g* at 4°C for 5 min
5. Recover the supernatant with a pipette
6. Quantify protein by an appropriate method (e.g., Pierce/Micro BCA™ Protein Assay Reagent Kit, #23235)

3.2 *Fluorescent Labeling and Application to the Lectin Microarray*

1. *Cy3 labeling*
   (a) Dilute samples to 50 μg/mL by adding PBS-Tx based on the result of the above protein assay
   (b) Put 20 μL of each sample (50 μg/mL) into a tube containing 100 μg of Cy3-SE (succinimide ester), mix with a pipette, and spin down by brief centrifugation
   (c) Put the tubes into a shading bag and incubate them for 1 h at room temperature (25°C) in the dark
2. *Gel filtration to remove excess free Cy3-SE*:
   Remove excess Cy3-SE reagent by gel filtration using appropriate equipment (e.g., Pierce/Zeba™ Desalt Spin Columns, 0.5 mL, #89882)
3. *Sample application to LecChip™*
   (a) Measure each volume of Cy3-labeled samples with a pipette
   (b) Prepare a total volume of 500 μL by adding an appropriate probing buffer (e.g., GP Biosciences Probing Solution); the protein concentration becomes 2 μg/mL as 1 μg of protein is eluted
   (c) Dilute these samples to 2 μg/mL, 1 μg/mL, 500 ng/mL, 250 ng/mL, 125 ng/mL, 62.5 ng/mL, and 31.25 ng/mL by adding the same probing buffer
   (d) Put LecChip™ in a frozen state at –20°C into an incubation box to melt the keeping solution (provided by the supplier) in wells of the LecChip™
   (e) Remove the keeping solution
   (f) Apply 100 μL of these samples to each well of the LecChips™ with a pipette
   (g) Incubate the LecChips™ in an incubation box at 20°C for 16 h
4. *Scanning with GlycoStation™ Reader 1200*
   Scan the LecChips with GlycoStation™ Reader 1200. The following conditions are recommended for the first trial: gain (70–125), exposure time (133, 199 ms), cumulative count (4). In order to detect relatively weak signals while avoiding saturation of strong signals, take some other scans while adjusting the gain and the exposure time.

3.3 Data Analysis

Data analysis is a critical part of lectin microarray experiments, because the data obtained from each microarray analysis shows systematic variation in microarray quality, scanner detection stability, sample preparation reproducibility, and labeling efficiency. For this purpose, an initial step to normalize the obtained signal intensities is essential. In this section, the two fundamental procedures required for data analysis are described [42–44].

1. *Gain merging*

   Kuno et al. developed a gain-merging procedure to expand the dynamic range of signal intensities obtained by lectin microarray analysis [42]. This procedure is necessary because of the optical properties of the scanning system, but is practically useful when analyzing a series of clinical samples, which often give a wide range of binding signals. When performing an analysis, a lectin microarray slide is scanned under two different gain conditions; one is a higher gain to "rescue" weak signals below 1,000 and the other is a lower gain to "suppress" strong signals over 40,000. Here, appropriate lectins for "merging" are selected, to give moderate signal intensities (i.e., 1,000–40,000) under both higher and lower gain conditions. Then, a merging factor (F) is determined as the average of higher (Int^H)/lower (Int^L) ratios calculated for individual merging lectins by (1):

$$F = \text{Averaged}\left(\text{Int}^{\text{H}}/\text{Int}^{\text{L}}\right) \tag{1}$$

   The over-range intensities (>40,000) obtained under the higher gain condition (e.g., $Int^H_{(\text{lectin A})}$) are replaced with theoretical intensities (Int^T_d lectin cT) by (2):

$$\text{Int}^{\text{T}}{}_{(\text{lectin A})} = \text{Int}^{\text{L}}{}_{(\text{lectin A})} \times F \tag{2}$$

2. *Normalization*

   Four different normalization methods are available to process lectin microarray data: "max," "mean," "particular lectin," and "median." For these normalization procedures, the signal intensity is multiplied by a normalization factor N for each array, which is calculated by $N = 1/\mu$, where μ is either the highest signal intensity of all of the lectins on the array (max), the mean of all of the lectins on the array (mean), the signal intensity of one selected lectin on the array (particular lectin), or the median of all of the lectins on the array (median). Required procedures depend significantly on experimental procedures and research purposes. For example, the author usually uses a mean normalization method for comparative purposes when dealing with a series of stem cells [45–47], because this gave the best result for glycan analysis of CHO and its mutant LEC cells [44]. On the other hand, a max-normalization procedure is widely used for differential glycan profiling targeting clinical samples for glycan-related bio-marker investigation [16, 17, 28, 48, 49].

4 Targets of the Lectin Microarray

Since its first publication in 2005, most reports of the lectin microarray have concerned its basic aspects; e.g., development of the array substrates and their fabrications. Recently, however, an increasing number of applications of lectin microarray technology have been reported, e.g., in glycan-related biomarker investigations, stem cell profiling toward regeneration medicine, microbial infections, and glycoprotein profiling in the light of functional glycomics (for representative applications, see Table 1). In the following section, relevant technologies required in various orders of hierarchy (i.e., molecular, cellular, tissue and body fluid orders) are briefly described.

4.1 Samples of Homogeneous Glycome

1. *Oligosaccharides*

   There are few reports describing the application of the lectin microarray to free oligosaccharides [10, 34], probably because a substantial merit of this microarray is its direct applicability to glycoproteins without liberation of oligosaccharides. Alternatively, most monoamine-coupling fluorescent reagents (e.g., 2-aminopyridine and 2-aminobenzamide) are of the UV-excited type, which are not compatible with the present glass substrate (UV-fluorescence positive). Uchiyama et al. used TAMRA (tetramethylrhodamine)-labeled three representative *N*-linked oligosaccharides (i.e., M6 high-mannose type, α-2-6-disialobiantennary, and asialobiantennary glycans) for analysis by an evanescent-field activated fluorescence-type scanner [34]. Notably, without a washing procedure, relatively weak bindings of these monovalent oligosaccharides could be seen toward a restricted set of lectins, while they were lost immediately after a buffer replacement procedure. Signal patterns observed for these oligosaccharides are relatively clear, as has been evident when sequential glycosidase digestion was performed for a complex-type sialobiantennary *N*-glycan [10].
2. *Purified glycoproteins*

   As described, glycoproteins are major targets for direct analysis by the lectin microarray, either in their purified (e.g., glycoprotein drugs) or crude (e.g., cell supernatant and body fluids) forms. The analysis provides glycan profiles regarding both *N*- and *O*-glycans [8–13]. Because the profiles obtained are unique to individual glycoproteins and states of the cells which produce them, the method contributes to the validation of glycoprotein drugs [12]. In combination with a specific antibody against core protein, highly sensitive monitoring of glycan profiles of target glycoproteins in the course of their production is possible (described later).