Development of basic tools for glycoscience and their application to cancer diagnosis

— A 10-year strategy of the Research Center for Medical Glycoscience of AIST —

Hisashi NARIMATSU

[Translation from *Synthesiology*, Vol.5, No.3, p.190-203 (2012)]

We proposed a 10-year strategy for the development of a new scientific field, glycoscience. Initially, we developed basic technological tools to help scientists and engineers entering this field. As the first project, we exhaustively discovered glycogenes and carried out their functional analyses. The fruits of this work led to several follow-on projects: (1) technology for enzyme synthesis of glycans, (2) technology for structural analysis of glycans, and (3) analysis of biological functions of glycans. The basic tools, developed in the first 5 years of our 10-year strategy, were applied to the development of more useful products, e.g., disease biomarkers, particularly for cancer diagnosis. We are also close to achieving the practical use of a liver fibrosis marker and a cholangiocarcinoma marker for diagnosis. Moreover, we are pursuing development of biomarkers for diagnosis of other cancers. The successful research results for these 10 years have now been transferred to the world, in particular, Asian countries, and yielded collaborative research contracts with domestic and overseas research groups.

Keywords : Glycan, *N*-glycan, *O*-glycan, glycosyltransferase, glycogene, lectin, lectin array, mass-spectrometry, IGOT, biomarker, liver fibrosis, liver cancer, cholangiocarcinoma

1 Introduction

Glycans are often regarded as the third biological chains of fundamental biopolymers, following nucleic acids (first chain) and proteins (second chain). Proteins are easily understandable as they are the principal gene products (nucleic acids). The protein synthetic mechanism is similar among species, and thus the principles clarified in lower organisms are also applicable to humans. In contrast, glycans are synthesized sequentially by more than 180 kinds of enzymes called glycosyltransferases.

The substrate specificity of glycans amazingly changes along with the evolution of species, resulting in wide differences in monosaccharides which make up glycans and also in their structural sequences between the lower organisms and humans. This is interesting, as nucleotides and amino acids which make up nucleic acids and proteins have barely evolved. There are a few monosaccharides that are common between bacteria and humans, but the majority varies. Equally, with animals and plants, a large part of the monosaccharides are uncommon. Glycans also differ between anthropoids including humans and other mammals such as pigs, cattle and lower species.

In general, glycans are attached to proteins and phosphates forming glycoproteins and glycolipids, respectively. Most of the membrane proteins and serum proteins are glycoproteins, and they become functionally-matured molecules after receiving proper glycan attachments (glycosylation). A group of glycoproteins consist of uniform protein moiety, but their glycan moieties are quite inhomogeneous. For example, immunoglobulin G (IgG) has double stranded, simple *N*-linked glycans (*N*-glycans), which have a wide variety of structures (36 types). However, it is very difficult to purify glycoproteins consisting of homogenous glycan structures. Moreover, it is almost impossible to synthesize glycoproteins consisting of homogenous glycan structures. For these reasons, the variety of glycan functions attributed to slight alteration of glycan structures is so far hardly analyzed.

Glycans binding to glycoproteins are roughly classified into *N*-glycans and *O*-linked glycans (*O*-glycans). *N*-glycans are attached to the asparagine-X-threonine or -serine (Asn-X-Thr/Ser) sequence, and this preference is relatively conserved among species. In contrast, *O*-glycans can be attached to any part of threonine or serine. In humans, there are about 20 members of ppGalNac-T, a glycotransferase family which initiates synthesis of *O*-glycans, and each of them exhibits various expression patterns along with differentiation or canceration of cells. This means binding sites of *O*-glycans are affected by differentiation or canceration of cells. Unfortunately, no technology to identify the binding site of *O*-glycan has been developed yet.

It is a natural habit of scientists to think that there should

Research Center for Medical Glycoscience (RCMG), AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba 305-8568, Japan
E-mail : h.narimatsu@aist.go.jp

Original manuscript received March 5, 2012, Revisions received April 13, 2012, Accepted April 25, 2012

be some important functions in the existence of glycan molecules. However, functional analysis of glycan-related proteins is often avoided by many scientists, as there is no appropriate basic technology for analysis.

The whole human genome sequencing was completed 11 years ago, and major attention was given to proteomics (comprehensive analysis of proteins) taking the resultant genome sequences as the templates. We were sure that this flow of scientific approach will be directed toward the comprehensive analysis of glycoproteins, the final form of functional molecules. It would require development of the basic technology tools for glycoscience. By means of such tools, analysis of glycan functions would be enhanced under the concept of glycoproteome. We were aiming for the goal of glycoproteomics as the biomedical application, in other words, diagnostic and therapeutic technologies.

2 About glycans

"Glycans are like clothing for cells and proteins" (Fig.1). Characteristics of glycans as summarized are expected to be applied in many fields.

(1) Sequential structures of glycans are dramatically altered along with differentiation, maturation, and activation of cells. If a normal cell becomes a cancerous cell, the cell will dedifferentiate and its glycan structures will be widely modified as well. Such alterations would make glycans the best prospective cancer markers and be applicable to regenerative medicine. Glycan structures change regularly along with the direction of differentiation, which supposed to be useful in the lineage determination of cultured cells. The germ cells mature and differentiate most rapidly among other cells. Glycans of sperm and ova indicate fast and amazing changes, which suggest an important role of glycans in maturation.

Moreover, glycan structures also react and change quickly upon activation and inactivation of cells. Cells related to the immune system show alterations of glycan structures along with their activation and inactivation.

(2) Glycan structures are varied by the derivate tissue of the carrier protein. As an example, transferrin is produced by both the liver cell and choroid plexus, but the glycan structures of transferrin derived from different origins are different but their carrier proteins are the same. Thus the origin of a protein can be detected based on its glycan structure. Moreover, expression levels of certain sialyltransferases and sulfotransferases often increase dramatically in cancerous cells, and as a result, many glycoproteins are sialylated or sulfated. The negative charge of the cancerous cell surface increases due to the increase of sialic acid and the sulfate group. This indicates that controlling only a few glycotransferases would affect functions of abundant glycoproteins. As a result, characteristics of cells are widely affected by glycotransferases.

(3) Glycan structures in a variety of glycoproteins produced by one particular cell vary widely. Despite the fact that the glycoproteins derived from the same cell have the same expression pattern of glycosyltransferases, each kind of glycoproteins has a characteristic glycan structure, and the mechanism for this system has not yet been clarified.

(4) Glycans have individual specificity. A representative example is the blood type. Not only the ABO blood group system, but also Lewis, P, and Ii systems are based on

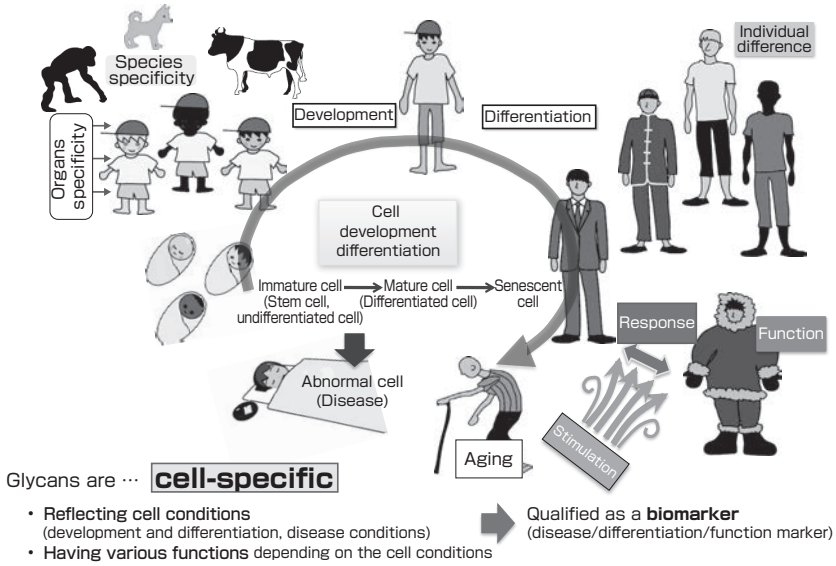

Fig. 1 Glycans are like clothing for cells and proteins

the individual differences of glycan structures. Mutation of the relevant glycotransferase genes induces loss of enzyme activity or alteration of substrate specificity, causing the differences in glycan structures. This mutation is inherited from parent to child. The largest obstacle of organ transplantation in humans is the difference of glycan structures concerned in the ABO blood group system.

(5) Glycans have species specificity. Evolution of glycosyltransferase genes is the fastest among that of all genes. This is probably due to the fact that the glycans on the cell surface are most directly affected by the changes of the outer environment, and the glycan structures have long been selected in response to those changes. Erythropoietin being used as a pharmaceutical antibody and hematopoietics is produced by using hamster cells. Thus its glycan structures are the same as those of hamsters, not of humans. Erythropoietin could be used for doping by athletes. Therefore, testing for hamster-type glycans is applicable for doping inspections. Studies for transplantation of pig organs (xenotransplantation) have been ongoing, but acute rejection occurs due to different glycan structures, as pigs have specific glycan structures that are not found in humans.

(6) In infectious disease, infection is initiated by binding of pathogenic microorganisms with specific glycans of the host cells. Many kinds of viruses including influenza virus bind to glycans. There are also opposite cases. Glycan structures of pathogenic microorganisms are recognized by and bound with lectins on the surface of the host cells, and thus the host cells become infected. The terminal end of glycans is most closely interacted with the outer environment. Infection of many pathogens begins by binding to glycans (or lectins) of cells, suggesting that the individual selection to avoid the infection from pathogens is one of the causes for the fast evolution of the glycan structures. To avoid infection genetically, alteration of glycan structures helps preserving the species. Influenza viruses bind to α2,6 sialic acid, *Helicobacter pylori* to Lewis-type blood group glycans, and noroviruses to ABO and Lewis-type blood group glycans. Individuals which gained tolerance against infection through mutation of glycosyltransferases survive and procreate descendants. Such individual selection can be seen for more than ten-thousand years. The ABO blood group system can be found in species upper than anthropoid, but the Lewis blood group system is specific to humans. The mutation of glycosyltransferase that specifies the Lewis blood group antigens occurred 20 to 30 thousand years ago.

3 Scenario and strategy of glycoscience

As described above, glycan structures well reflect the differentiation and dedifferentiation (canceration) of cells, as well as tissue specificities. These characteristics serve as the principle for development of glycan biomarkers. Glycan structures are determined mainly based on the expression patterns of glycosyltransferases, and thus are estimated to be controlled by the transcriptional regulatory mechanism or epigenetic mechanisms (control of expression patterns by post-translational modification not relying on the gene sequence). However, such basic research had hardly been conducted 10 years ago. Therefore, we had to start from the development of basic technologies required for the glycoscience by ourselves, which greatly contributed to the remarkable progress of the research field. It was the most desired and attractive task to be accomplished as a pioneer of a new science. If a scientist starts research using technologies developed by a foreign country, it cannot be denied that this scientist lags behind the foreign country. This can be applicable to any of the scientific fields.

In the glycoscience, likewise in the genetics and protein science, the firstly-required basic technologies are those for synthesis and structure (or sequence) analysis. These basic technologies must be easy to use for every researcher. Ten years ago, most of the available technologies for synthesis and structure analysis were immature and inappropriate for non-expertized users. Therefore, we set a 10-year perspective to pursue the sequential approach for glycoscience (Fig.2). We focused on the development of basic technologies during the first 5 years, and then application of the technologies during the latter 5 years. The process of research was as follows: (1) Human-derived glycogenes were comprehensively identified and analyzed. (2) Recombinant glycosyltransferases were expressed from the obtained glycogenes, and obtained enzymes were used in combination to synthesize glycans with a variety of structures to make a glycan library. (3) Obtained glycans with known structures were used as the standards for development of glycan analysis technologies. (4) *In vivo* functions of glycans were analyzed.

To elucidate the effects of alteration of glycan structures on the function of glycoproteins and cell phenotypes, the basic technologies for the following purposes are necessary.

(1) Glycogenes:

The human-derived glycogenes were comprehensively identified and analyzed in the project for establishment of a glycogene library, Glycogene (GG) Project. To synthesize a glycoprotein, the protein moiety is regulated by expression of one gene, but the glycan moiety is controlled by the coordinate expression of dozens of glycogenes. Therefore, if all the glycogenes are elucidated, the mechanisms for biological synthesis of glycoproteins and glycolipids should be clarified. To reach the final goal, that is, understanding of the glycan functions, elucidation of all glycogenes was necessary as

the first step.

(2) Synthesis:

Glycosyltransferases were expressed as recombinant enzymes based on the obtained glycogenes, and used in combination to synthesize glycans with a variety of structures to make a glycan library. Biochemical synthesis of glycans requires a huge amount of work and time. Moreover, an innovative organic chemical synthesis method for glycans has not yet been developed, and synthesis with organic solvents is environmentally offensive. Synthesis of complicated glycans with a variety of structures is impossible, and it takes considerable time to create even one kind of structure. The only advantage is that once an organic chemical synthesis method is established, mass production at an industrial level would become possible.

In contrast, the glycogenes that were comprehensively obtained at the previous step were expressed as recombinant enzymes for enzymological synthesis of glycans. Through the combination of a variety of enzymes, quite a lot of desired glycans in a variety of structures could be synthesized freely in a short time. Glycosyltransferases are highly substrate-specific, and thus one kind of enzyme can synthesize only one kind of structure. Therefore, certain glycan structures are synthesized easily and rapidly by the glycosyltransferases with known substrate-specificity. The reaction is hydrolytic and thus is eco-friendly. The disadvantage is that because the enzymes are derived from human origins, they are extremely unstable and the production costs much. However, it is impossible to use glycosyltransferases obtained from lower organisms alternatively for synthesis of human glycan structures. Therefore, the enzyme method is very suitable for synthesis of a small amount of variable glycans, but not for mass production.

(3) Structures:

In the Structural Glycomics (SG) Project, the above-obtained glycans with elucidated structures were used as the standards and contributed to the development of glycan analysis technologies. This enzyme method is suitable to produce a few milligrams of many kinds of glycans required as the standard substances.

We thus established two structure elucidation methods using the standard glycans and glycoproteins: Tandem mass spectrometry (MS^n method) and lectin array method. Each of them has advantages and disadvantages, and is suitable for different purposes. The MS^n method is superior in the following: (1) The analysis is easy for everyone. (2) The glycan structure can be determined. It is inferior in the following: (1) The analysis requires relatively large amounts of glycans (about several micrograms). (2) The glycans must be purified prior to analysis. On the other hand, the lectin array is superior in the following: (1) The sensitivity is very high. (2) The antibody-overlay method does not require complete purification of the target glycoproteins. (3) Comparison of glycan profiling among multiple samples is possible.

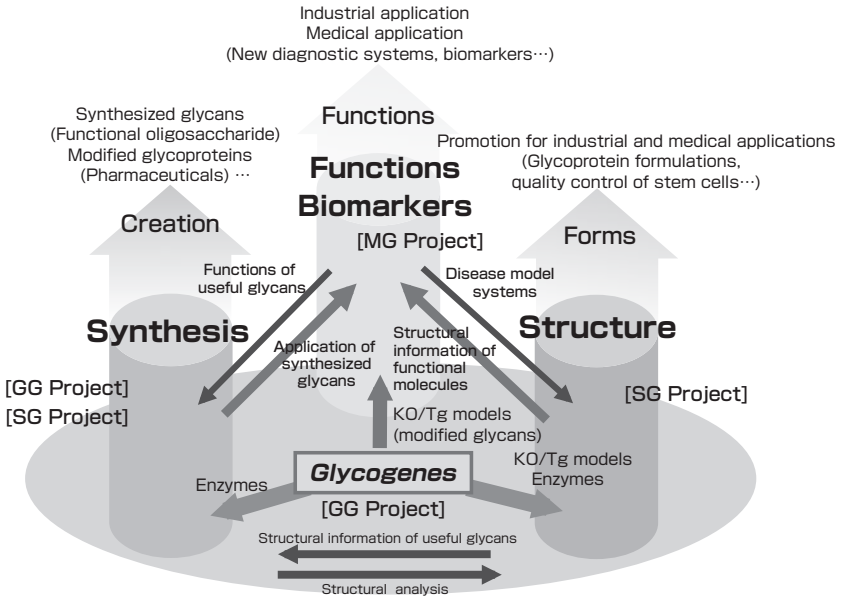

Fig. 2 The three main themes of glycoscience

Glycogene (GG) Project: Establishment of glycogene library, Structural Glycomics (SG) Project: Glycan engineering (structural analysis technologies), Medical Glycomics (MG) Project: Application of determined functions (functions and biomarkers)

It is inferior in the following: (1) Absolute structure determination is impossible. (2) Lectins may not always be readily available.

(4) Functions and biomarkers:

In the Medical Glycomics (MG) Project, effects of the change of glycan structure on the glycoprotein functions and cell phenotypes were determined *in vivo* using the above technologies (1-3). As we went through, "synthesis", "structure", and "functions and biomarkers" have been the three main themes and interactively developed. Their application can be seen as "synthesis" in the production of functional oligosaccharides and glycan-modified glycoproteins, "structure" in the quality management of glycoprotein pharmaceuticals and ES cells, and "functions and biomarkers" in the commercial and medical applications as shown in diagnosis technologies.

4 Development of elemental technologies for basic technology tools

Figure 3 shows a list of elemental technologies developed for glycoscience. The following is the summary of each element.

Elemental technology 1: Identification of glycosyltransferases in human genome databases by bioinformatics technology

We started from the comprehensive identification of candidate genes for glycosyltransferases by full application of the bioinformatics technology. Kikuchi, a member of our project team assigned from Mitsui Knowledge Industry Co., Ltd., developed new software to identify the candidates from the available genome databases. This software is capable of searching not only amino acid homologues but also characteristic glycogenes such as (1) glycogenes possessing a membrane binding site to hydrophobic amino acids at proximal to the *N*-terminus, and their lengths being about 18 to 22 amino acid residues, slightly shorter than cell membrane binding proteins, (2) glycogenes with a subsequent main structure that is rich in proline and has many serine and threonine, (3) glycogenes subsequent to the enzyme activity domain, and have active domains consisting of 300-400 amino acids and containing several cysteine and 3-amino acids of the DXD motif binding to bivalent cations. About 100 glycogenes with such characteristics were identified and their cDNAs were generated mainly based on the RNA of human cultured cells. All the candidate genes coding the entire enzyme were cloned by PCR.

Elemental technology 2: Recombination of glycosyltransferases into expression vectors and substrate specificity analysis of recombinant enzymes

For the development of this technology, many original members of the Research Center for Glycobiotechnology (predecessor of Research Center for Medical Glycoscience [RCMG]) of AIST (Togayachi, Sato, Goto, Kudo, Tachibana, Cho, Kubota, Sawaki, and more) greatly contributed. Glycosyltransferases are membrane proteins binding to the Golgi membranes and endoplasmic reticulum membranes. To activate the glycosyltransferases as recombinant enzymes for *in vitro* synthesis of glycans, they have to be in the form of

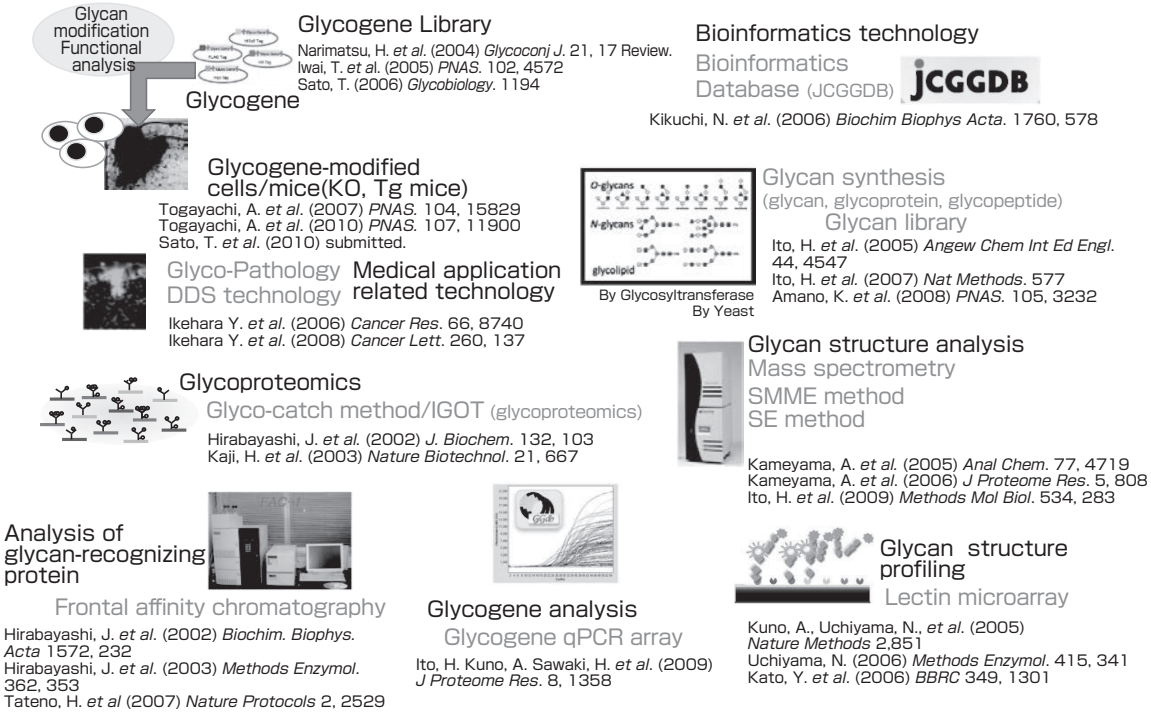

Fig. 3 Representative basic technology tools developed or utilized in RCMG

soluble enzymes. Therefore, the membrane binding moiety of the candidate genes was omitted and the area supposed to be the enzyme activity domain was incorporated into the gateway vector tagged with FLAG, and transfected into the human embryonic kidney blast cells (HEK293T cell). The obtained recombinant proteins secreted into the medium were partially purified from the supernatant. We concentrated on comprehensive, easy, and fast detection of activities in recombinant enzymes derived from many candidate glycogenes. We purchased nine human-derived substrates labeled with radioisotope, and added into the culture media. Moreover, we added monosaccharides and oligosaccharides, as well as a mixture of glycolipid and a mixture of glycoproteins obtained from cultured cells as the acceptor substrates. The HEK293T cell was used for recombinant expression of glycosyltransferases, because it was already known that human derived glycosyltransferases are quite unstable and fragile proteins, and impossible to be expressed in active forms by Escherichia coli or yeast. The HEK293T cell is derived from humans and its glycoproteins are highly glycosylated. It is estimated that many glycosyltransferases are expressed endogenously in the HEK293T cell, which leads to a prediction that there are machineries required for expression of ectogenic human recombinant glycosyltransferases with their own activity. Even under the current knowledge, the human derived HEK293T cell is the most suitable for expression of human recombinant glycosyltransferases. The newly developed glycosyltransferase genes in this project were applied for the substance patents and most of them were published in major journals.[1]-[29] However, in the future, the mass production of glycans will be required. For the mass production, an inexpensive bulk production method is necessary. Production by vertebrate-derived cultured cells costs much and is not suitable for mass production. From this point, Chiba and his group of RCMG are establishing a mass production system of human-derived glycosyltransferases in yeast.[30]

These accomplishments are accumulated in the glycogene databases for open access at Japan Consortium for Glycobiology and Glycotechnology Database (http://jcggdb.jp/). This database is being expanded to contain not only the information about glycogenes, but the wider range of contents to form an advanced database by Shikanai and other members of the RCMG.

Elemental technology 3: Quantitative assay method for 186 kinds of glycogene expressions

Glycosyltransferases, which synthesize the main structure of *N*-glycans, are expressed in every cell. The expression level of glycosyltransferases is large and not affected by the conditions of cells. The expression level of other glycogenes is low; especially those synthesizing the terminal moiety of glycans are very slight compared with other genes. They are impossible to be detected by ordinary DNA chips, and even if it could be detected, their modifications cannot be measured correctly. We developed a technology to measure the expression levels of all 186 glycogenes accurately in a comprehensive, high throughput manner. The quantitative real-time PCR (qPCR) for comprehensive glycogene expression analysis is a matured experimental technique, which is the most reliable biological analysis method in terms of detection sensitivity and measurement accuracy. Glycogenes are experimentally known as low in the expression levels in most of the cases, thus qPCR was considered to be suitable for the development of the expression analysis system.[31] Specifically, customized qPCR arrays for the 186 glycogenes encoding the glycosyltransferases and modification enzymes were established by the members of RCMG with Sawaki as the core member. The plasmid DNA pool of the glycogene clone library used as the calibrator enabled the system to indicate the amount of transcription products of all 186 glycogenes based on their copy number at a one-time measurement. Classification of cells in terms of the expression profiling of glycogenes is well correlated with the cell differentiation or canceration, and the expression of glycogenes is known to be proportional to the expression of glycans.

Elemental technology 4: Establishment of *in vitro* synthesis method of glycans and glycopeptides by recombinant glycosyltransferase

As the recombinant glycosyltransferases are derived from humans, most of them can be purified as soluble recombinant enzymes without losing activity by means of human-derived HEK293T cells. Based on this principle, we established an *in vitro* synthesis method of glycans and glycopeptides. As an exception, the glycosyltransferases attributed to the main structure of *N*-glycans cannot be recombined as they are synthesized on the rough endoplasmic reticulum (ER) and they pass through the lipid membrane several times. As for the main structure of *N*-glycans, commercially available purified natural forms were used as the starting materials. As for the *O*-glycans, representative peptides possessing *O*-glycans, such as mucin were used as the starting materials and glycans were elongated by adding glycosyltransferases sequentially.[30][32][33]

Synthesis methods were established for two objectives by Ito *et al*. of RCMG. The first objective is the mass production of one kind of glycans in the largest quantity. The conditions for enzyme reaction were set at a certain level, and the largest amount of enzymes was reacted for the longest time possible. The obtained products were separated and refined by liquid chromatography. The second objective was the simultaneous synthesis of multiple glycan structures at small quantities in one tube. Reaction of each enzyme was terminated by heat when the production reached about 50 % saturation. Then the next enzyme was added to the tube and similarly the reaction was terminated at the 50 % reaction point. Thearetically, 2^n

kinds of glycans can be yielded in one tube by this method. As the molecular mass of produced glycans is preliminarily known, the desired mass can be measured in an aliquot of the final mixture by mass spectrometer. We named this method Mass-tagged synthesis.[34][35]

Elemental technology 5: Large-scale identification technology of glycoproteins by LC/MS

Based on the newly developed LC/MS analysis method that can simultaneously identify more than 1,000 proteins in a peptide mixture sample at the same time, we established a large scale identification method for elucidation of binding sites of glycoproteins and glycan structures of glycopeptides isolated from protein digestion samples by affinity chromatography. As the peptide moiety of glycopeptides was not fragmented by MS/MS analysis with collision-induced dissociation due to the presence of large glycans, glycopeptides were not suitable for direct identification. Therefore, glycans were released from glycopeptides by glycopeptidases, and the obtained deglycosylated peptides were subjected to the large-scale identification. In this reaction, Asn in the glycosylation site was replaced by Asp, and the mass was increased by 1 Da, which indicated the glycosylation site. During this reaction, as non-glycopeptides receiving de-Asn reaction were also present in the mixture and thus the de-glycosylated peptides could not be distinguished, stable-isotope labeled water ($H_2{}^{18}O$) was added to the solvent for enzymes to incorporate the labeled oxygen. As a result, the glycosylation site was labeled, and thus the highly accurate glycoprotein identification method was actualized (IGOT method). Based on the integration of LC/MS and IGOT method, Kaji *et al*. of RCMG are actively pursuing the high-throughput mass identification of glycoproteins. The current high speed mass analysis system enables a series of identification processes for 500-1,000 kinds of glycoproteins in 1 mg of a tissue-originated protein sample in about 10 days.[36][37]

Elementary technology 6: MS^n-based identification of glycan structures

In the tandem MS^n method, mass of the target glycan (mass of MS^1) is measured first. Then the glycan is dissociated by weak collision of a rare gas such as argon or helium (Collision Induced Dissociation: CID), and the mass of each derived fragment is measured (MS^2). Each of these fragments is isolated and again fragmented by CID to obtain MS^3. Theoretically, MS^n can be measured as long as a sufficient amount of the sample is available. In the actual measurements, even a slight difference of a glycan structure can be distinguished by MS^4 based on the dissociation pattern obtained by CID (Fig.4).

Therefore, we obtained the data up to MS^4 of as many standard glycans as possible, and compiled them into a

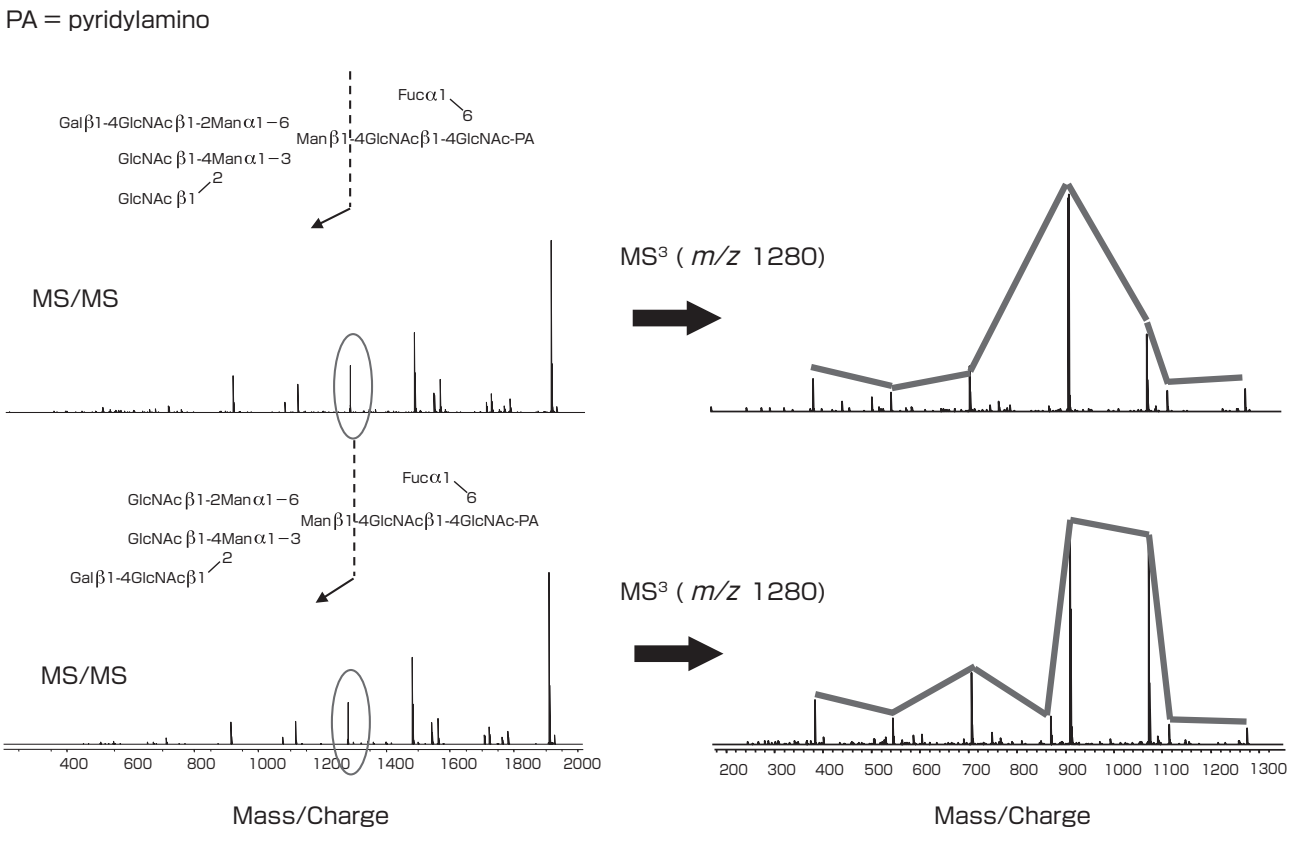

Fig. 4 Application of the tandem MS^n-based structure identification on isomeric glycans

database (DB). A researcher to identify an unknown glycan structure should analyze the target glycan up to MS^2 and refer to the DB. The DB will suggest to the researcher which fragment should be subjected to MS^3. The researcher will refer to the DB with the obtained MS^3. In most of the cases, the structure would be determined at this point, but sometimes the DB would direct the researcher to MS^4.

This MS^n method for the glycan structure identification system was developed collaboratively by Kameyama, Narimatsu, and other members of AIST, Shimadzu Corporation, and Mitsui Knowledge Industry Co., Ltd., and marketed by Shimadzu Corporation.

Elementary technology 7: Antibody-overlay lectin microarray

To determine the alteration of glycan structures on glycoproteins along with disease development in biological samples, "high throughput, highly sensitive, highly reproducible, and rapid" comparative analytical technology is necessary. The most suitable system is the antibody-overlay lectin microarray developed by Kuno and Hirabayashi *et al*. of RCMG.[38] The lectin microarray consists of 43 lectins with a variety of specificity solidified on a glass base plate, which can analyze several samples simultaneously by one plate. In the antibody overlay method, the glycoproteins as the objective samples are applied on the lectin micro array without labeling or any other preprocessing, and the reacted glycoproteins binding to the lectins on the plate are detected by the fluorescence labeled antibody recognizing core proteins. By the excitation light radiated from the glass fragment and its total reflection, about 200-μm thick of evanescent wave is generated around the glass surface. We designated the system so that only the labeled substances within this layer are signaling. This system is highly sensitive and useful even for detection of only a slight amount of glycans (Fig.5). The preceding glycan analysis methods by liquid chromatography or mass chromatography require a considerable amount of preprocessing and time to release glycans from proteins and to label them. In comparison to these methods, this is a progressive method enabling easy detection. Although the sensitivity relies on the quality of antibodies, the amount of the target glycoprotein samples required for western blotting (a few nanograms) is sufficient for this method. Moreover, as the binding signal of the target glycoprotein is detected specifically by the antibody, the only sample preparation required is simple purification by immunoprecipitation or similar methods. In fact, we were successful in comparative analyses of more than 50 glycoproteins in about a 10-ng level of samples efficiently enriched from the serum or tissue samples and cell culture supernatants by the antibody overlay lectin micro array. This technology is applied for verification of candidate glycoproteins for biomarkers and contributed to establishing the development pipeline of useful glycobiomarkers.[31][33][38][39] The detail of this developmental scheme was published as literature.[40] The lectin array glycan profiling system was developed by Kuno and Hirabayashi *et al*. of AIST and GP Bioscience, and made publically available by GP Bioscience.

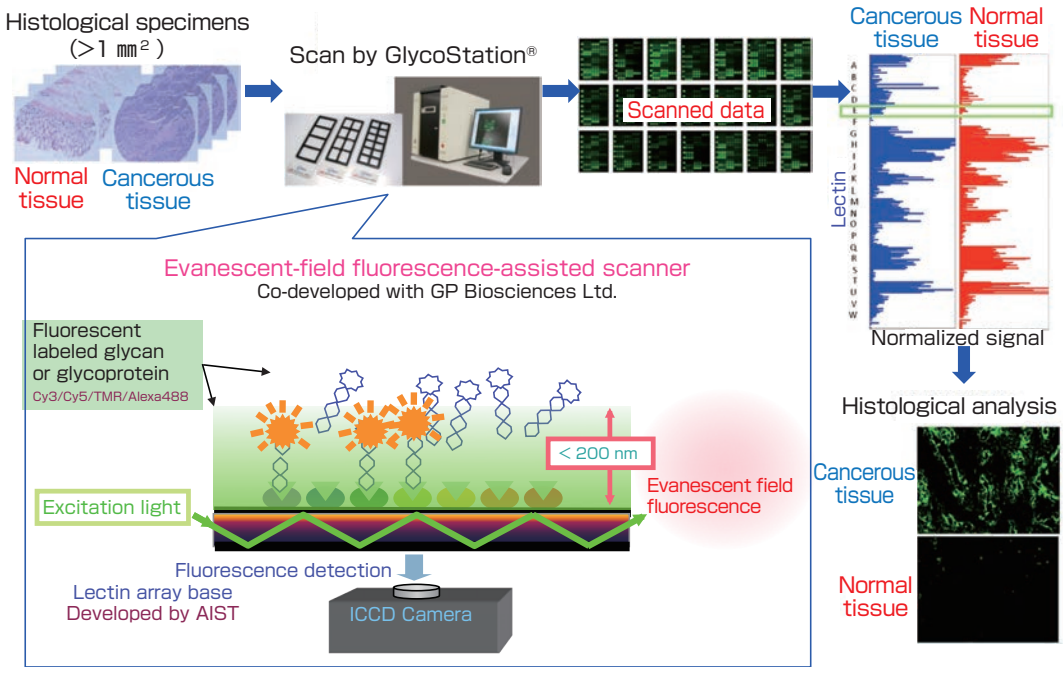

Fig. 5 Application example of lectin microarray
Selection of useful lectins from enriched glycoproteins in slight pieces of histological samples by comparative analysis

5 Search and practical application of disease biomarkers by accumulated base technologies for glycoscience

5.1 Strategy of the disease glycobiomarker search

The search for disease biomarkers by application of proteomics technology is largely pursued. In proteomics, the concept of biomarkers is based on the quantitative difference of proteins. However, our principle is totally different. In our glycoproteomics approach for development of biomarkers, we are targeting on finding qualitatively changed glycoproteins based on the fact that the glycan structure of the disease-derived glycoproteins is altered from that derived from normal cells. Such glycoproteins can be called post-translationally modified isomers.

The amount of *in vivo* glycobiomarkers (isomers) is considered to be very slight. Especially in the search for the early stage cancer markers, the amount is sparse as earlier the stage is. Therefore, it is impossible to find such a molecule in a serum. We fully utilized the preliminarily developed technologies and established a developmental scheme for cancer markers as shown in Figure 6.

1. The RNA is extracted from both cancerous and normal tissues, and expression patterns of their glcyogenes are comprehensively analyzed by the real-time PCR. As a result, the glycan structures receiving modification by canceration are determined.

2. Glycan profiles in the total glycoproteins obtained from cancerous tissue and cultured cancer cells are comparatively analyzed by lectin microarray. Characteristic lectin is selected as the probe.

3. The candidate marker glycoproteins are comprehensively identified with the selected lectin by the LC/MS/IGOT method. The number of candidate glycoprotein is narrowed to about a few hundred at this point.

4. To detect a marker in serum, glycoproteins that are abundantly present in serum even in a normal state are advantageous. We utilized bioinformatics to do the following: (i) Estimation of the serum concentration of candidate glycoproteins, and select candidates with sufficient amounts. (ii) Confirmation of the origin of the candidate is the target tissue. If the candidate has multiple origins, it is avoided due to dilution of the target molecule. (iii) Selection of the molecules with many *N*- or *O*-glycan binding sites, as more binding sites induce higher affinity to the probes. We prioritized the candidate molecules based on these parameters.

5. The candidate molecules are analyzed by western blotting with purchased antibodies according to the priority, and the serum concentrations are estimated.

6. We further narrow down the promising candidates, and immunoprecipitate for rough purification. These concentrated molecules are again analyzed by lectin microarray, and the lectin indicated the most different profiles between cancerous patients and healthy controls (lectin A) is selected.

7. In general, most of commercial antibodies are weak in affinity. In such cases, suitable antibody is also originally developed for the protein moiety of the candidate glycoprotein.

8. With the developed antibody possessing high affinity

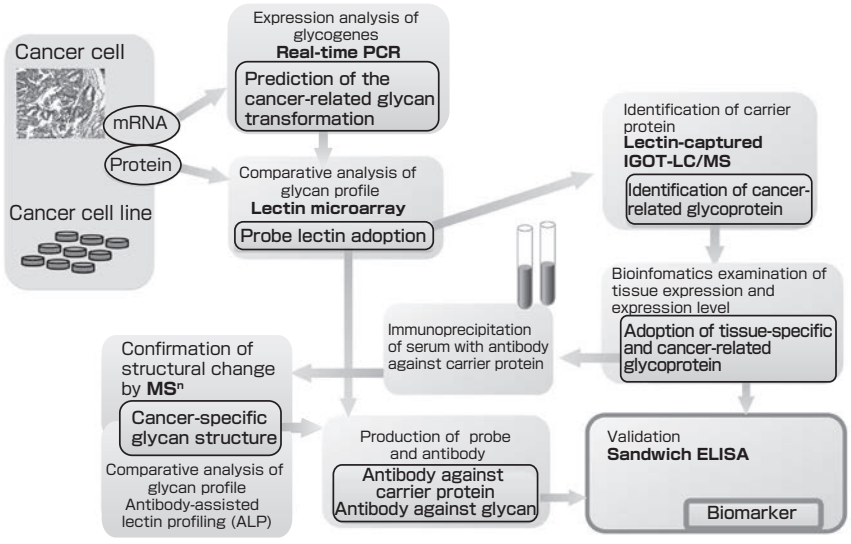

Fig. 6 Strategy for development of disease glycobiomarkers

and specificity and the suitable probe (e.g., lectin A), an antibody-probe sandwich detection kit is established and verified with more than 100 samples.

9. If the advantage over current markers is statistically confirmed, verification of the kit is further conducted with more than 1,000 samples.

10. Furthermore, the modification of glycan structure is determined by MS^n. As the available samples obtained from patients are generally limited, it is difficult to identify the glycan structure of the samples by MS^n. In such cases, we first confirm that cultured cancerous cells indicate the same lectin reactivity, and then the target glycoproteins are purified in large amounts from supernatant of the cultured cells, and are determined for the glycan structure by MS^n.

11. At this point, we ask for cooperation of domestic clinical doctors. With the objective data presented under the cooperation of many doctors, the kit is evaluated comparatively with the current biomarkers, and applied for production and health insurance approval by the collaborative public company for the final production.

We have targeted many diseases. The followings are successful examples in development of the liver fibrosis marker and cholangiocarcinoma marker.

5.2 Development of liver fibrosis marker

We have developed a liver fibrosis marker, which is almost ready to be applied for production approval by the collaborative company.

Infection with hepatitis B virus (HBV) and C virus (HCV) progresses to acute hepatitis, chronic hepatitis, cirrhosis, and finally hepatocarcinoma in 20-30 years after infection. The patients infected with HBV and HCV are estimated to be several hundred million worldwide, 7 % of the Japanese population (about 8 million) and 10 % of the Chinese population (about 150 million). In the course of fibrosis, hepatocytes are disrupted after infection and replaced with fibrin molecules, and the stiffness of the liver increases. Currently the definitive diagnosis is made by needle biopsy; however, this is an invasive and burdening method for patients, requiring 2-3 days of hospitalization. The level of fibrosis (liver stiffness) is classified into F0 (fibrosis 0) to F4. The incidence of hepatocarcinoma increases along with the progression of fibrosis. At F3, the fibrosis level is considerably progressed as a result of chronic hepatitis, and F4 is the state of cirrhosis. Medication with interferon and ribavirin is mostly effective at F3, but cannot be expected at F4. Development of the biomarker for simple blood tests to evaluate the fibrosis level is keenly awaited for the evaluation of therapeutic effect and prediction of the risk of hepatocarcinoma.

According to the above-described developmental scheme, alpha-1 acid glycoprotein (AGP) was selected as the first candidate for the fibrosis marker. AGP is a glycoprotein abundantly present in blood, and mainly secreted from the liver. We premised that AGP would well reflect the status of liver fibrosis, and its lectin avidity would be strong as it has five *N*-glycans. Moreover, it has been long known that the glycan structure of AGP is altered along with liver fibrosis. We could use the serum samples provided from clinical doctors, available with the data of fibrosis levels diagnosed by biopsy. The most suitable lectin for distinguishing F4 and F3 was selected via immunoprecipitation of AGP followed by lectin array. Three lectins, AOL, MAL, and DSA were able to evaluate the fibrosis level quite accurately.[41][42] We established the sandwich detection system consisting of the anti-AGP antibody and three lectins in collaboration with Sysmex Corporation, and optimized the system for an automatic clinical chemistry analyzer, HISCL produced by Sysmex. This system can measure one sample in 17 minutes. However, AGP was not suitable enough for clinical diagnosis, as AGP must be immunoprecipitated from serum before analysis, which requires 2 hours of processing. We further searched for an appropriate glycoprotein for direct measurement of serum by HISCL, and found “molecule X” (the name cannot be disclosed as it is before publication). Glycans on molecule X detected by lectin Y well reflect the status of fibrosis. We prepared a monoclonal antibody against molecule X and established a sandwich assay system with this antibody and lectin Y. This system enabled evaluation of fibrosis within 17 minutes by HISCL without preprocessing of serum. In the future, a patient visiting a clinical doctor would receive the serum test before seeing the doctor, and the doctor would be able to have the fibrosis level of the patient on the same day before consultation.

5.3 Development of cholangiocarcinoma marker

When a hepatic mass is found by an imaging diagnosis method, intrahepatic cholangiocarcinoma derived from bile duct epithelial cells and hepatocarcinoma derived from hepatic cells must be clearly distinguished. Prognosis of intrahepatic cholangiocarcinoma is bad, and its therapeutic strategy is totally different from that of hepatocarcinoma.

In accordance with the developmental strategy of the cancer marker mentioned above, we incised small tissue specimens of 1-mm diameter from the cholangiocarcinoma tissue and the normal region by microdissection. As a result of the lectin array of the fluorescence labeled glycoprotein extract, the WFA lectin signal was remarkably different between the cancer tissue and the normal tissue. We identified many candidate molecules (230 glycoproteins) of the cholangiocarcinoma marker that bind to WFA by the IGOT method. These molecules were prioritized based on the estimated blood concentrations by bioinformatics. The antibodies were purchased for the 10 glycoproteins of

higher priority and the bile and serum concentrations of the candidate molecules in the patients of cholangiocarcinoma were estimated by western blotting and immunoprecipitation. We confirmed that the candidate molecules were actually produced by the cancer cells based on the immunostaining of the cancer tissues obtained from the patients. Currently, two assay systems, anti-MUC1 antibody/WFA and anti-protein Y/WFA, have been established, and the marker concentration of each system in the bile of patients was quantified. The detection rate of cancer cells in the most widely used diagnosis system is low at about 20-30 %, but our system indicates 85-90 % of high diagnostic accuracy.[43] This methodology is proved to be effective for the assays with patient serum as well as bile.

6 Conclusion

We are applying the same strategy aiming successful development of truly useful clinical diagnosis markers for other cancers, such as lung cancer, ovary cancer, pancreas cancer, prostate cancer.

The most important thing for development of disease markers is the tight relationship with reliable clinical doctors for collaborative research. The research and development should be pursued in careful consideration of the following points: (1) What is the really desired marker? (2) What kind of phenomena or indices should be compared to achieve the goal? Can the collaborative doctors offer the samples for comparison? (3) Are there suitable samples of patients with clear disease history retained at the clinical site? (4) Presence of long-time chronological samples of the same patients is very important if possible. (5) Presence of samples of the same patients before and after treatment is also very important.

There is a bad example of the cancer biomarker search. In the serum of the terminal cancer patients, there are abundant abnormal molecules due to cachexy caused by the state of cancer. Although more than several hundred kinds of abnormal molecules can be identified immediately in comparison of sera from the terminal cancer patients and healthy controls, these are clinically useless. Truly helpful biomarkers for judgment of disease progression cannot be found by the comparison between the terminal cancer patients and healthy controls.

More than 10 years of the scientific activity in the field of glycoscience at AIST is the most substantial scientific experiment for me because of the following: (1) We were able to receive enough grant through NEDO for our research. (2) We were able to invite excellent researchers from outside sources. (3) We were able to have researchers from various scientific backgrounds such as medical science, agriculture, physical science, and engineering. (4) We started the glycoscience with about 30 members, but now we have about 100 members aiming toward the same goal.

In addition, I would like to enhance the collaboration in the Asian region for glycoscience. When I started the glycogene project 11 years ago, I imagined the great development of science in China 10 years later. Therefore, I hired about 10 Chinese post-doctoral researchers and educated them in glycoscience from the beginning. They accomplished their research in 2 to 3 years and returned to China, and now they are holding professorships and important roles for development of Chinese glycoscience. Last year, we established a branch office of RCMG of AIST within the Shanghai Jiao Tong University to facilitate the collaborative research, and the common scientific themes are pursued through frequent interaction of researchers. I am expecting further cooperation with many organizations through acceptance and education at AIST of post-doctoral researchers and doctoral students from domestic and international universities. I am sure that the 21st century will be the era of Asia. To emphasize the cooperation in Asia in the field of glycoscience, we founded Asian Consortium of Glycoscience and Glycotechnology (ACGG) and hosted the first ACGG symposium in Tsukuba 3 years ago. The second symposium was held in Taipei and the third in Shanghai, and its participants are increasing rapidly.

References

[1] L. Cheng, K. Tachibana, H. Iwasaki, A. Kameyama, Y. Zhang, T. Kubota, T. Hiruma, K. Tachibana, T. Kudo, JM. Guo and H. Narimatsu: Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T15, *FEBS Lett.*, 566(1), 17-24 (2004).

[2] L. Cheng, K. Tachibana, Y. Zhang, JM. Guo, K. Tachibana, A. Kameyama, H. Wang, T. Hiruma, H. Iwasaki, A. Togayachi, T. Kudo and H. Narimatsu: Characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-T10, *FEBS Lett.*, 531(2), 115-121 (2002).

[3] K. Fujimura, H. Sawaki, T. Sakai, T. Hiruma, N. Nakanishi, T. Sato, T. Ohkura and H. Narimatsu: LARGE2 facilitates the maturation of α-dystroglycan more effectively than LARGE, *Biochem. Biophys. Res. Commun.*, 329(3), 1162-1171 (2005).

[4] M. Gotoh, T. Sato, T. Akashima, H. Iwasaki, A. Kameyama, H. Mochizuki, T. Yada, N. Inaba, Y. Zhang, N. Kikuchi, YD. Kwon, A. Togayachi, T. Kudo, S. Nishihara, H. Watanabe, K. Kimata and H. Narimatsu: Enzymatic synthesis of chondroitin with a novel chondroitin sulfate *N*-acetylgalactosaminyltransferase that transfers *N*-acetylgalactosamine to glucuronic acid in initiation and elongation of chondroitin sulfate synthesis, *J. Biol. Chem.*, 277(41), 38189-38196 (2002).

[5] M. Gotoh, T. Sato, K. Kiyohara, A. Kameyama, N. Kikuchi, YD. Kwon, Y. Ishizuka, T. Iwai, H. Nakanishi and H. Narimatsu: Molecular cloning and characterization of β1,4-*N*-acetylgalactosaminyltransferases IV synthesizing *N,N′*-diacetyllactosediamine, *FEBS Lett.*, 562(1-3), 134-140 (2004).

[6] M. Gotoh, T. Yada, T. Sato, T. Akashima, H. Iwasaki,

H. Mochizuki, N. Inaba, A. Togayachi, T. Kudo, H. Watanabe, K. Kimata and H. Narimatsu: Molecular cloning and characterization of a novel chondroitin sulfate glucuronyltransferase that transfers glucuronic acid to *N*-acetylgalactosamine, *J. Biol. Chem*., 277(41), 38179-38188 (2002).

[7] J.M. Guo, Y. Zhang, L. Cheng, H. Iwasaki, H. Wang, T. Kubota, K. Tachibana and H. Narimatsu: Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase family, pp-GalNAc-T12, *FEBS Lett*., 524(1), 211-218 (2002).

[8] T. Hiruma, A. Togayachi, K. Okamura, T. Sato, N. Kikuchi, YD. Kwon, A. Nakamura, K. Fujimura, M. Gotoh, K. Tachibana, Y. Ishizuka, T. Noce, H. Nakanishi and H. Narimatsu: A novel human β1,3-*N*-acetylgalactosaminyltransferase that synthesizes a unique carbohydrate structure, GalNAcβ1-3GlcNAc, *J. Biol. Chem*., 279(14), 14087-14095 (2004).

[9] N. Inaba, T. Hiruma, A. Togayachi, H. Iwasaki, XH. Wang, Y. Furukawa, R. Sumi, T. Kudo, K. Fujimura, T. Iwai, M. Gotoh, M. Nakamura and H. Narimatsu: A novel I-branching β-1,6-*N*-acetylglucosaminyltransferase involved in human blood group I antigen expression, *Blood*, 101(7), 2870-2876 (2003).

[10] H. Ishida, A. Togayachi, T. Sakai, T. Iwai, T. Hiruma, T. Sato, R. Okubo, N. Inaba, T. Kudo, M. Gotoh, J. Shoda, N. Tanaka and H. Narimatsu: A novel β1,3-*N*-acetylglucosaminyltransferase (β3Gn-T8), which synthesizes poly-*N*-acetyllactosamine, is dramatically upregulated in colon cancer, *FEBS Lett*., 579(1), 71-78 (2005).

[11] T. Iwai, N. Inaba, A. Naundorf, Y. Zhang, M. Gotoh, H. Iwasaki, T. Kudo, A. Togayachi, Y. Ishizuka, H. Nakanishi and H. Narimatsu: Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide β1,3-*N*-acetylglucosaminyltransferase (β3Gn-T6), an enzyme synthesizing the core 3 structure of *O*-glycans, *J. Biol. Chem*., 277(15), 12802-12809 (2002).

[12] T. Iwai, T. Kudo, R. Kawamoto, T. Kubota, A. Togayachi, T. Hiruma, T. Okada, T. Kawamoto, K. Morozumi and H. Narimatsu: Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells, *Proc. Natl. Acad. Sci*., 102(12), 4572-4577 (2005).

[13] N. Kikuchi, A. Kameyama, S. Nakaya, H. Ito, T. Sato, T. Shikanai, Y. Takahashi and H. Narimatsu: The carbohydrate sequence markup language (CabosML): an XML description of carbohydrate structures, *Bioinformatics*, 21(8), 1717-1718 (2005).

[14] N. Kikuchi, YD. Kwon, M. Gotoh and H. Narimatsu: Comparison of glycosyltransferase families using the profile hidden Markov model, *Biochem. Biophys. Res. Commun*., 310(2), 574-579 (2003).

[15] T. Kudo, T. Iwai, T. Kubota, H. Iwasaki, Y. Takayama, T. Hiruma, N. Inaba, Y. Zhang, M. Gotoh, A. Togayachi and H. Narimatsu: Molecular cloning and characterization of a novel UDP-Gal:GalNAc(α) peptide β1,3-galactosyltransferase (C1Gal-T2), an enzyme synthesizing a core 1 structure of *O*-glycan, *J. Biol. Chem*., 277(49), 47724-47731 (2002).

[16] H. Mochizuki, K. Yoshida, M. Gotoh, S. Sugioka, N. Kikuchi, YD. Kwon, A. Tawada, K. Maeyama, N. Inaba, T. Hiruma, K. Kimata and H. Narimatsu: Characterization of a heparan sulfate 3-*O*-sulfotransferase-5, an enzyme synthesizing a tetrasulfated disaccharide, *J. Biol. Chem*., 278(29), 26780-26787 (2003).

[17] H. Narimatsu: Construction of a human glycogene library and comprehensive functional analysis, *Glycoconj. J*., 21(1-2), 17-24 (2004).

[18] H. Narimatsu: Human glycogene cloning: focus on β3-glycosyltransferase and β4-glycosyltransferase families, *Curr. Opin. Struct. Biol*., 16(5), 567-575 (2006).

[19] C. Peng, A. Togayachi, YD. Kwon, C. Xie, G. Wu, X. Zou, T. Sato, H. Ito, K. Tachibana, T. Kubota, T. Noce, H. Narimatsu and Y. Zhang: Identification of a novel human UDP-GalNAc transferase with unique catalytic activity and expression profile, *Biochem. Biophys. Res. Commun*., 402(4), 680-686 (2010).

[20] T. Sato, M. Gotoh, K. Kiyohara, T. Akashima, H. Iwasaki, A. Kameyama, H. Mochizuki, T. Yada, N. Inaba, A. Togayachi, T. Kudo, M. Asada, H. Watanabe, T. Imamura, K. Kimata and H. Narimatsu: Differential roles of two *N*-acetylgalactosaminyltransferases, CSGalNAcT-1, and a novel enzyme, CSGalNAcT-2. Initiation and elongation in synthesis of chondroitin sulfate, *J. Biol. Chem*., 278(5), 3063-3071 (2003).

[21] T. Sato, M. Gotoh, K. Kiyohara, A. Kameyama, T. Kubota, N. Kikuchi, Y. Ishizuka, H. Iwasaki, A. Togayachi, T. Kudo, T. Ohkura, H. Nakanishi and H. Narimatsu: Molecular cloning and characterization of a novel human β1,4-*N*-acetylgalactosaminyltransferase, β4GalNAc-T3, responsible for the synthesis of *N,N'*-diacetyllactosediamine, galNAc β1-4GlcNAc, *J. Biol. Chem*., 278(48), 47534-47544 (2003).

[22] T. Sato, M. Sato, K. Kiyohara, M. Sogabe, T. Shikanai, N. Kikuchi, A. Togayachi, H. Ishida, H. Ito, A. Kameyama, M. Gotoh and H. Narimatsu: Molecular cloning and characterization of a novel human β1,3-glucosyltransferase, which is localized at the endoplasmic reticulum and glucosylates *O*-linked fucosylglycan on thrombospondin type 1 repeat domain, *Glycobiology*, 16(12), 1194-1206 (2006).

[23] A. Togayachi, T. Akashima, R. Ookubo, T. Kudo, S. Nishihara, H. Iwasaki, A. Natsume, H. Mio, J. Inokuchi, T. Irimura, K. Sasaki and H. Narimatsu: Molecular cloning and characterization of UDP-GlcNAc:lactosylceramide β1,3-*N*-acetylglucosaminyltransferase (β3Gn-T5), an essential enzyme for the expression of HNK-1 and Lewis X epitopes on glycolipids, *J. Biol. Chem*., 276(25), 22032-22040 (2001).

[24] A. Togayachi, N. Kikuchi, T. Kudo and H. Narimatsu: Comprehensive study on glycosyltransferases which determine glycosylation, *Tanpakushitsu Kakusan Koso*, 48(11 Suppl), 1542-1549 (2003).

[25] A. Togayachi, T. Sato and H. Narimatsu: Comprehensive enzymatic characterization of glycosyltransferases with a β3GT or β4GT motif, *Methods Enzymol*., 416, 91-102 (2006).

[26] H. Wang, K. Tachibana, Y. Zhang, H. Iwasaki, A. Kameyama, L. Cheng, JM. Guo, T. Hiruma, A. Togayachi, T. Kudo, N. Kikuchi and H. Narimatsu: Cloning and characterization of a novel UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase, pp-GalNAc-T14, *Biochem. Biophys. Res. Commun*., 300(3), 738-744 (2003).

[27] T. Yada, M. Gotoh, T. Sato, M. Shionyu, M. Go, H. Kaseyama, H. Iwasaki, N. Kikuchi, YD. Kwon, A. Togayachi, T. Kudo, H. Watanabe, H. Narimatsu and K. Kimata: Chondroitin sulfate synthase-2: molecular cloning and characterization of a novel human glycosyltransferase homologous to chondroitin sulfate glucuronyltransferase, which has dual enzymatic activities, *J. Biol. Chem*., 278(32), 30235-30247 (2003).

[28] T. Yada, T. Sato, H. Kaseyama, M. Gotoh, H. Iwasaki, N. Kikuchi, YD. Kwon, A. Togayachi, T. Kudo, H. Watanabe, H.

Narimatsu and K. Kimata: Chondroitin sulfate synthase-3: molecular cloning and characterization, *J. Biol. Chem*., 278(41), 39711-39725 (2003).

[29] Y. Zhang, H. Iwasaki, H. Wang, T. Kudo, T.B. Kalka, T. Hennet, T. Kubota, L. Cheng, N. Inaba, M. Gotoh, A. Togayachi, J. Guo, H. Hisatomi, K. Nakajima, S. Nishihara, M. Nakamura, J.D. Marth and H. Narimatsu: Cloning and characterization of a new human UDP-*N*-acetyl-α-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase, designated pp-GalNAc-T13, that is specifically expressed in neurons and synthesizes GalNAc α-serine/threonine antigen, *J. Biol. Chem*., 278(1), 573-584 (2003).

[30] K. Amano, Y. Chiba, Y. Kasahara, Y. Kato, M.K. Kaneko, A. Kuno, H. Ito, K. Kobayashi, J. Hirabayashi, Y. Jigami and H. Narimatsu: Engineering of mucin-type human glycoproteins in yeast cells, *Proc. Natl. Acad. Sci*., 105(9), 3232-3237 (2008).

[31] H. Ito, A. Kuno, H. Sawaki, M. Sogabe, H. Ozaki, Y. Tanaka, M. Mizokami, J. Shoda, T. Angata, T. Sato, J. Hirabayashi, Y. Ikehara and H. Narimatsu: Strategy for glycoproteomics: identification of glyco-alteration using multiple glycan profiling tools, *J. Proteome Res*., 8(3), 1358-1367 (2009).

[32] M.K. Kaneko, Y. Kato, A. Kameyama, H. Ito, A. Kuno, J. Hirabayashi, T. Kubota, K. Amano, Y. Chiba, Y. Hasegawa, I. Sasagawa, K. Mishima and H. Narimatsu: Functional glycosylation of human podoplanin: glycan structure of platelet aggregation-inducing factor, *FEBS Lett*., 581(2), 331-336 (2007).

[33] Y. Kato, M.K. Kaneko, A. Kuno, N. Uchiyama, K. Amano, Y. Chiba, Y. Hasegawa, J. Hirabayashi, H. Narimatsu, K. Mishima and M. Osawa: Inhibition of tumor cell-induced platelet aggregation using a novel anti-podoplanin antibody reacting with its platelet-aggregation-stimulating domain, *Biochem. Biophys. Res. Commun*., 349(4), 1301-1307 (2006).

[34] H. Ito, A. Kameyama, T. Sato, K. Kiyohara, Y. Nakahara and H. Narimatsu: Molecular-weight-tagged glycopeptide library: efficient construction and applications, *Angew. Chem. Int. Ed*., 44(29), 4547-4549 (2005).

[35] H. Ito, A. Kameyama, T. Sato, M. Sukegawa, H.K. Ishida and H. Narimatsu: Strategy for the fine characterization of glycosyltransferase specificity using isotopomer assembly, *Nat. Methods*, 4(7), 577-582 (2007).

[36] H. Kaji, H. Saito, Y. Yamauchi, T. Shinkawa, M. Taoka, J. Hirabayashi, K. Kasai, N. Takahashi and T. Isobe: Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins, *Nat. Biotechnol*., 21(6), 667-672 (2003).

[37] H. Kaji, Y. Yamauchi, N. Takahashi and T. Isobe: Mass spectrometric identification of *N*-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging, *Nat. Protoc*., 1(6), 3019-3027 (2006).

[38] A. Kuno, Y. Kato, A. Matsuda, M.K. Kaneko, H. Ito, K. Amano, Y. Chiba, H. Narimatsu and J. Hirabayashi: Focused differential glycan analysis with the platform antibody-assisted lectin profiling for glycan-related biomarker verification, *Mol. Cell Proteomics*, 8(1), 99-108 (2009).

[39] A. Kuno, A. Matsuda, Y. Ikehara, H. Narimatsu and J. Hirabayashi: Differential glycan profiling by lectin microarray targeting tissue specimens, *Methods Enzymol*., 478, 165-179 (2010).

[40] H. Narimatsu, H. Sawaki, A. Kuno, H. Kaji, H. Ito and Y. Ikehara: A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics, *FEBS J*., 277(1), 95-105 (2010).

[41] A. Kuno, Y. Ikehara, Y. Tanaka, T. Angata, S. Unno, M. Sogabe, H. Ozaki, K. Ito, J. Hirabayashi, M. Mizokami and H. Narimatsu: Multilectin assay for detecting fibrosis-specific glyco-alteration by means of lectin microarray, *Clin. Chem*., 57(1), 48-56 (2011).

[42] A. Kuno, Y. Ikehara, Y. Tanaka, K. Saito, K. Ito, C. Tsuruno, S. Nagai, Y. Takahama, M. Mizokami, J. Hirabayashi and H. Narimatsu: LecT-Hepa: a triplex lectin-antibody sandwich immunoassay for estimating the progression dynamics of liver fibrosis assisted by a bedside clinical chemistry analyzer and an automated pretreatment machine, *Clin. Chim. Acta*, 412(19-20), 1767-1772 (2011).

[43] A. Matsuda, A. Kuno, T. Kawamoto, H. Matsuzaki, T. Irimura, Y. Ikehara, Y. Zen, Y. Nakanuma, M. Yamamoto, N. Ohkohchi, J. Shoda, J. Hirabayashi and H. Narimatsu: Wisteria floribunda agglutinin-positive mucin 1 is a sensitive biliary marker for human cholangiocarcinoma, *Hepatology*, 52(1), 174-182 (2010).

Author

Hisashi NARIMATSU

Graduated the Keio University School of Medicine and received M.D. in 1974. Completed the Keio University Graduate School of Medicine and received Ph.D. in 1979. Became an associate professor at the Keio University School of Medicine and a professor at the Institute of Life Science, Soka University in 1991. Joined the Institute of Molecular and Cell Biology, AIST as a principal research scientist in 2001 and became the director of RCMG since 2006. Currently, also serves as a professor of the Graduate School of Comprehensive Human Sciences, Tsukuba University, a visiting professor of the School of Medicine, Keio University, and the advisory professor for the Shanghai Jiao Tong University. Specialties are glycobiology, biochemistry, immunology, microbiology, and tumor biology. Received the Nikkei BP Technological Award in 2007, Chemistry-Bio Tsukuba Award in 2010, Tsukuba Award in 2011. Affiliations: councilor, The Japanese Society of Carbohydrate Research General; council member, Japan Consortium for Glycobiology and Glycotechnology; council member, Human Proteome Organization (HUPO); council member, Japan Human Proteome Organization (JHUPO); councilor, the Japanese Biochemical Society; councilor, Japan Society for Molecular Marker Research; cooperation member, Science Council of Japan; member of Japan Cancer Association.

Discussions with Reviewers

1 Overall evaluation

Comment (Akira Ono, AIST)

The author clearly describes a developmental strategy (scenario) for the research in a new scientific field as a pioneer of glycoscience. The author has appropriately selected necessary elemental technologies and successfully achieved the development. Moreover, the processes followed by the author in which the elemental technologies are integrated and applied to

the establishment of cancer diagnosis systems are remarkably described.

It is noteworthy that the author regarded the development of basic technology tools, which are accessible for use by many scientists and engineers, as the first target aiming for the future enhancement of a new scientific field, glycoscience. Scientists are apt to gather under a fully developed fruitful tree and wish to become an end-user of fashionable sciences, but often avoid seeding or growing young trees of basic science. In contrast, this approach by the author and his team is worth praise as leading a new scientific field for other researchers.

Moreover, I think that the establishment of the three main themes, "synthesis", "structure", and "functions and biomarkers" at the initial phase as shown in Figure 2 was the key to the steady growth of the following research activities. I am sure that, to successfully lead the big project under a solid scenario, intensive and concerted efforts by a large group such as a research center of AIST for a long period as 10 years were necessary, as the author describes in "6. Conclusion."

Comment (Noboru Yumoto, AIST)

The development of the elemental technologies for analysis of glycan structures by a multi-disciplinary approach and the integration of the elemental technologies in accordance with the scenario of developing the disease biomarker is a great example of "*Type 2 Basic Research*."

2 Domestic and international application of the basic tools

Question (Akira Ono)

I assume that the basic tools developed in this research are broadly utilized in Japan and overseas to enhance glycoscience. Please introduce some of the examples for current application of these developmental achievements by other research groups of other institutes and companies. Are there any collaborative researches with domestic or international organizations? If so, please let us know as much as possible.

Answer (Hisashi Narimatsu)

Here is a list for some of the applications that I know of for our-developed basic tools used by other organizations

Glycogenes

- We hold the patent for about 30 glycogenes, and granted the licenses of 13 genes to GlycoGeneInc.
- We provide unpatented glycogenes for about 20 domestic and oversea research organizations. Moreover, we deposited these glycogenes to the National Institute of Technology and Evaluation (NITE) for wider distribution.
- Our developed glycogene database, GGDB had 172,570 access/year (2011fy).
- We developed 13 strains of glycogene knock-out mice. Collaborative studies with 5 domestic institutes and 3 overseas institutes are ongoing using them as disease models.
- Collaborative studies of glycan synthesis by glycosyltransferases aiming at applications such as to glycan arrays are ongoing with domestic and overseas institutes.
- A collaborative study for glycoprotein synthesis using inexpensively expressed glycosyltransferases by yeast is ongoing with a private company.
- Many research articles have been published as achievements of collaborative studies.

Lectin microarray

- GP Biosciences Ltd. commercialized the lectin microarray
- More than a few tens of related reports are published from external organizations. I picked three important reports that have a strong impact. The first report is the utilization of lectin microarray for evaluation of iPS cells by Dr. Yamanaka of Kyoto University:

  YC. Wang *et al*.: Specific lectin biomarkers for isolation of human pluripotent stem cells identified through array-based glycomic analysis, *Cell Res*., 21(11), 155-1563 (2011).

  EL. Bird-Lieberman *et al*.: Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Berrett's esophagus, *Nature Medicine*, 18(2), 315-321 (2012).

  SA. Fry *et al*.: Lectin microarray profiling of metastatic breast cancers, *Glycobiology*, 21(8), 1060-1070 (2011).

- The Lectin Frontier Database, LfDB, had 23,605 accesses in 2011.
- We have already applied for seven patents concerning glycan biomarkers based on the lectin array.

Glycan structure analysis based on the mass spectrometry

- The system has been marketed from Shimadzu Corporation/ Mitsui Knowledge Industry Co., Ltd. They sold one system each to Qatar, Beijing, Shanghai, and the United States, and three in Japan (one each for National Cancer Center, Gifu University, and Japan Anti-Doping Agency).
- The access counts to the two databases were 18,256 for GMDB and 36,729 for GPDB in 2011.
- We have many accomplishments of collaborative studies in glycan structure analyses. These are some of the representative articles.

  T. Fukuda *et al*.: α1,6-fucosyltransferase-deficient mice exhibit multiple behavioral abnormalities associated with a schizophrenia-like phenotype: importance of the balance between the dopamine and serotonin systems, *J. Biol. Chem*., 286(21), 18434-18443 (2011).

  N. Watanabe *et al*.: Clinicopathological features of 171 cases of primary thyroid lymphoma: a long-term study involving 24 553 patients with Hashimoto's disease, *Br. J. Haematol*., 153(2), 236-243 (2011).

  T. Nakagawa *et al*.: Glycomic analyses of glycoproteins in bile and serum during rat hepatocarcinogenesis, *J. Proteome Res*., 9(10), 4888-4896 (2010).

  N. Sasaki *et al*.: High levels of E4-PHA-reactive oligosaccharides: potential as marker for cells with characteristics of hepatic progenitor cells, *Glycoconj. J*., 26(9), 1213-1223 (2009).

  H. Suzuki *et al*.: Computationally and experimentally derived general rules for fragmentation of various glycosyl bonds in sodium adduct oligosaccharides, *Anal. Chem*., 81(3), 1108-1120 (2009).

  H. Shirato *et al*.: Noroviruses distinguish between type 1 and type 2 histo-blood group antigens for binding, *J. Virol*., 82(21), 10756-10767
  (2008).

  YY. Fan *et al*.: Identification of further elongation and branching of dimeric type 1 chain on lactosylceramides

from colonic adenocarcinoma by tandem mass spectrometry sequencing analyses, *J. Biol. Chem*., 283(24), 16455-16468 (2008).

T. Nishie *et al*.: Development of IgA nephropathy-like disease with high serum IgA levels and increased proportion of polymeric IgA in β-1,4-galactosyltransferase-deficient mice, *Contrib. Nephrol*., 157, 125-128 (2007).

T. Nishie *et al*.: Development of immunoglobulin A nephropathy-like disease in β-1,4-galactosyltransferase-I-deficient mice, *Am. J. Pathol*., 170(2), 447-456 (2007).

J. Iijima *et al*.: Cell-cell interaction-dependent regulation of *N*-acetylglucosaminyltransferase III and the bisected *N*-glycans in GE11 epithelial cells. Involvement of E-cadherin-mediated cell adhesion, *J. Biol. Chem*., 281(19), 13038-13046 (2006).

Quantitative PCR

- Two collaborative studies have been reported.

  K. Moriwaki *et al*.: Combination use of anti-CD133 antibody and SSA lectin can effectively enrich cells with high tumorigenicity, *Cancer Sci*., 102(6), 1164-1170 (2011).

  M. Miyata *et al*.: Membrane sialidase NEU3 is highly expressed in human melanoma cells promoting cell growth with minimal changes in the composition of gangliosides, *Cancer Sci*., 102(12), 2139-2149 (2011).

- One patent was obtained through a collaborative study (in collaboration with Osaka University, Nagoya University, and GP Bioscience).