BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

Insight into Glycan Diversity and Evolutionary Lineage Based on Comparative Avio-*N*-glycomics and Sialic Acid Analysis of 88 Egg Whites of Galloanserae

Kazuko Hirose,[†] Maho Amano,[*,†,‡] Ryo Hashimoto,[†] Yuan Chuan Lee,[§] and Shin-Ichiro Nishimura[*,†,‡,∥,⊥]

[†]Graduate School of Life Science, Hokkaido University, Sapporo 001-0021, Japan

[‡]Division of Quantification of Health State (Feel Fine Co.), Graduate School of Life Science, Hokkaido University, Sapporo 001-0021, Japan

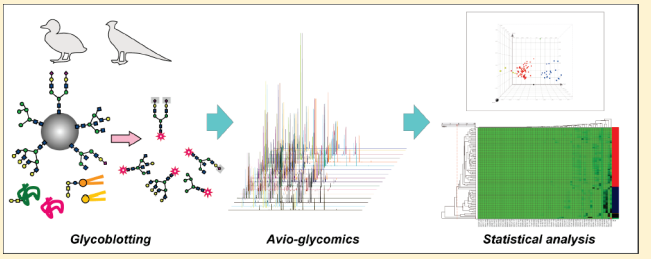
[§]Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, United States

[∥]Ezose Sciences, Inc., 25 Riverside Drive, Pine Brook, New Jersey 07058, United States

[⊥]Medicinal Chemistry Pharmaceuticals, LLC, 1-715, N7, W4, Kita-ku, Sapporo 060-0807, Japan

Supporting Information

ABSTRACT: A large set of glycome information was obtained from egg white proteins of 88 samples from Galloanserae (63 Anseriformes and 25 Galliformes). The data were obtained on whole *N*-glycan structures and types of sialic acids of these egg whites by glycoblotting-based high-throughput and quantitative glycomics. The results revealed clear trends and complexity patterns as well as diversity among taxonomic groups. It is well-known that chicken, a representative domesticated poultry involved in Galliformes, can become an influenza host. However, our data demonstrate that duck, wild goose, and swan of Anseriformes are representative migratory birds that are known as natural hosts of the influenza virus. Hierarchical clustering analysis of the expression pattern of *N*-glycome (total of 61 *N*-glycan peaks) revealed that the members of Galloanserae can be classified into two major groups and five submajor clusters (clusters 1−5) on the basis of simple *m*/*z* values obtained by MALDI-TOF MS. It is clear that expression patterns of *N*-glycomes in the five clusters are influenced significantly by the features such as the body size of the birds, rather than by the difference of the family. On the other hand, quantitative analysis showed that the total amounts of sialic acids in egg whites of Galliformes were distinctly larger than those of Anseriformes. However, it was also revealed in Anseriformes that Neu5Gc and KDN, in addition to common Neu5Ac, were expressed significantly in both *N*- and *O*-glycans of glycoproteins and glycosphingolipids, suggesting the influence of their lifestyles and diet. This is the first report that KDN exists in egg white. These results and the environmental factors are discussed preliminarily with respect to their evolutionary lineage.



Glycans existing on cell surfaces, extracellular proteins, and lipids are recognized by a group of proteins called lectins. Glycans are known to play many important roles in fundamental, vital processes such as differentiation, development, and immunity.[1] They are also known as targets of infectious pathogens, such as bacteria, viruses, and parasites.[2] It is believed that the diversity of the glycan structure in biological species developed so that they were equipped for various environmental conditions, for example, to defend against infection. The diversity and complexity of glycans are thought to have resulted from endogenous factors (intraspecies) and exogenous factors (interspecies). Endogenous factors are thought to be the diversification products through the different biosynthetic routes of glycans, exemplified by the human ABO-blood group system,[3] while the exogenous factor is the diversification of defense for glycan-mediated pathogen infection, as mentioned above.[2] The species-specific glycan structure reveals the evolutionary lineage. For example, the terminal Galα1→3Galβ1→4GlcNAc structure ("α-Gal epitope") is found in all mammals except for the Old World primates, including humans, apes, and Old World monkeys.[4−6] The major phylogenetic shift in glycosylation, the absence of α-Gal, would lead to a compensatory increase in the amounts of other terminal structures like sialic acids. The Galα1→4Gal structure in glycoproteins is also species-specific but had been rarely found in vertebrates except in a few species of birds and amphibians until recently. However, it has been reported that egg white and serum glycoproteins from some species of birds such as pigeon, gull, and perching bird have abundant *N*-glycan with the Galα1→4Gal structure at the nonreducing terminus. Interestingly, some avian species belonging to Ratitae and Galloanserae, such as chicken, quail, duck, and ostrich, did not express Galα1→4Gal on the glycoprotein of egg white.[7−9]

Received: December 6, 2010
Revised: March 11, 2011
Published: April 15, 2011

Sialic acids are often known to be involved in pathogen infection. Sialic acids show remarkable structural diversity, currently with more than 50 naturally occurring members. Natural structural variations of sialic acids frequently occurred at carbon 5 (C-5), which can be substituted with either an acetamido, hydroxyacetamido, or hydroxyl moiety to form *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc), or 2-keto-3-deoxynonulosonic acid (KDN), respectively.[10] Their abundance ratio is highly diverse and dependent on species and tissues.[2,11] Because of their propensity to occupy the outermost terminus of glycans, carry negative charge, and have structural variations, sialoglycans (sialic acid-containing glycans) are known as key determinants in various molecular recognition events.[12−14] Sialic acids are recognized by many receptors (lectins) and often become the target of many viral and bacterial infections via specific lectins.[10] For example, the lectin on the surface of influenza virus, called hemagglutinin (HA), binds to sialic acids of the host cell surfaces as the initial stage of virus infection. The strength of binding depends on the molecular species and the linkage of sialic acids.

Birds must be regarded as one of the most important animals in relation to infectious diseases. It is generally considered that the natural host of the influenza virus is waterfowl. Because some birds move their habitats ("migration"), once pathogen infections occur, they could easily spread to remote areas. Ducks, wild geese, and swans of Anseriformes are representative migratory birds that are also natural hosts of the influenza virus.[15,16] Regardless of the species, Anseriformes inhabiting different environments show different recognition patterns of the influenza virus HA. Their binding affinity for sialic acids differs considerably depending on the different molecular species and linkages.[17] Since 1996, H5N1, a highly pathogenic avian influenza virus, has repeatedly infected waterfowl such as ducks, and when such infection spreads to domestic poultry, including chickens, very high pathogenicity often emerges.[18,19]

Egg white is an ideal material for examining the diversity of glycans for their roles in protecting the embryo in the yolk from various infections. In other words, the glycans in egg white might have been used as a decoy of the pathogen recognition through the specific interaction between pathogenic lectins and designated glycan clusters of egg white glycoconjugates and protect the embryo from infection in the yolk. Therefore, it can be surmised that the history of antipathogenic offense and defense in the course of evolution brought about the diversity of glycans in egg whites.[7−9,20]

If the species-specific structural diversity of egg white glycans is the result of biological evolution as an adaptation to the environmental changes via variation of the glycan antigen structures to protect embryos against various pathogenic invasions, the studies of such glycan structure promise to improve our understanding of the course of biological evolution. However, there is an enormous extent of glycan diversity in nature, yet its actual biological functions are almost unknown. There is also no clear explanation for the extreme complexity and diversity of glycans that can be found on a given glycoconjugate or cell type. The scope and distribution of the diversity among taxonomic groups are quite limited, and it is difficult to envisage clear trends or patterns consistent with different evolutionary lineages. This is why it is necessary to increase the amount of basic information to unravel them.

Until now, technical limitations have restricted acquisition of the total glycan structures of biological samples.[21] Recently, the "glycoblotting method", a practical and highly sensitive method for glycan enrichment, was developed[22] and can be used for high-throughput analysis of biological samples such as sera, cells, and egg whites.[23−27] We have already demonstrated the much higher sensitivity of glycoblotting-based human serum *N*-glycan analysis compared to that of a conventional HPLC-based approach on the basis of PA-derivatized *N*-glycans.[23] Via a comparison with a recent result[28] for human serum glycomics that showed 47 *N*-glycans by using the permethylation method,[28−33] it is obvious that glycoblotting-based glycomics is a highly efficient and facile alternative. Our protocol allowed for rapid profiling of 44 kinds of *N*-glycans when 83 serum samples (10 μL) of patients suffering from hepatocyte cellular carcinoma were tested in comparison with those of 20 healthy human controls (the races are different).[24] It seems likely that both methods have approximately equal potential in terms of the high sensitivity needed for diagnostic biomarkers. Furthermore, high-throughput instrumentation suitable for the large-scale and automated glycoblotting, namely "SweetBlot", was developed, which allowed the truly comprehensive analysis of a series of glycan samples.[24] Using this technology, we can now acquire basic information about glycan profiles at a dramatically fast rate, which can be used to elucidate the function of glycans in the different evolutionary lineage.

EXPERIMENTAL PROCEDURES

Materials. Peptide *N*-glycosidase F (PNGase F) and Proteinase K were purchased from Roche (Mannheim, Germany). 3-Methyl-1-*p*-tolyltriazene (MTT) and *O*-benzylhydroxylamine hydrochloride (BOA) were obtained from Tokyo Chemical Industry Co. (Tokyo, Japan). Dithiothreitol (DTT), α-cyano-4-hydroxycinnamic acid (CHCA) diethylamine salt, 5-*N*-glycolylneuraminic acid (Neu5Gc), and 2-keto-3-deoxynonulosonic acid (KDN) were from Sigma-Aldrich, Inc. (St. Louis, MO). 5-*N*-Acetylneuraminic acid (Neu5Ac) was purchased from Japan Food & Liquor Alliance Inc. Food & Bio Research Center (Kyoto, Japan). BlotGlycoH was purchased from Sumitomo Bakelite, Co. (Tokyo, Japan). 1,2-Diamino-4,5-methylenedioxybenzene (DMB) was purchased from Dojindo Laboratories (Kumamoto, Japan). Other reagents and solvents were obtained from Wako Pure Chemicals Co. (Tokyo, Japan), unless otherwise stated; 88 egg white samples from various species of Galloanserae [2 orders, 5 families, 35 genera, and 72 species (Table 2)] collected by M. Laskowski, Jr.,[34−36] were maintained at −20 °C until they were used.

Nomenclature. The Latin names of the birds were those given by Gruson and Forster[37] and Sibley and Monroe.[38] The common English names were as described previously.[34−36] Orders and families are listed according to Sibley and Monroe.[38,39] The phylogeny of birds based on DNA−DNA hybridization by Sibley et al.[38−40] was used for our basic reference because their classification of birds in the world is complementary to the phylogenetic analysis. The zoogeographical region is one of the geographical divisions of the world devised for the study of the distribution of animals. It consists of the Palearctic (Eurasia), the Nearctic (North America), the Neotropic (Central America and South America), the Afrotropic (Africa south of the Sahara), the Indomalaya (Asia south of the Himalayas), and the Australasia (Australia, New Zealand, and nearby islands) (Table S1 and Figure S1 of the Supporting Information).

Glycoblotting-Based Quantitative *N*-Glycomics of Egg Whites. *Release of N-Glycans from Egg Whites.* Lyophilized

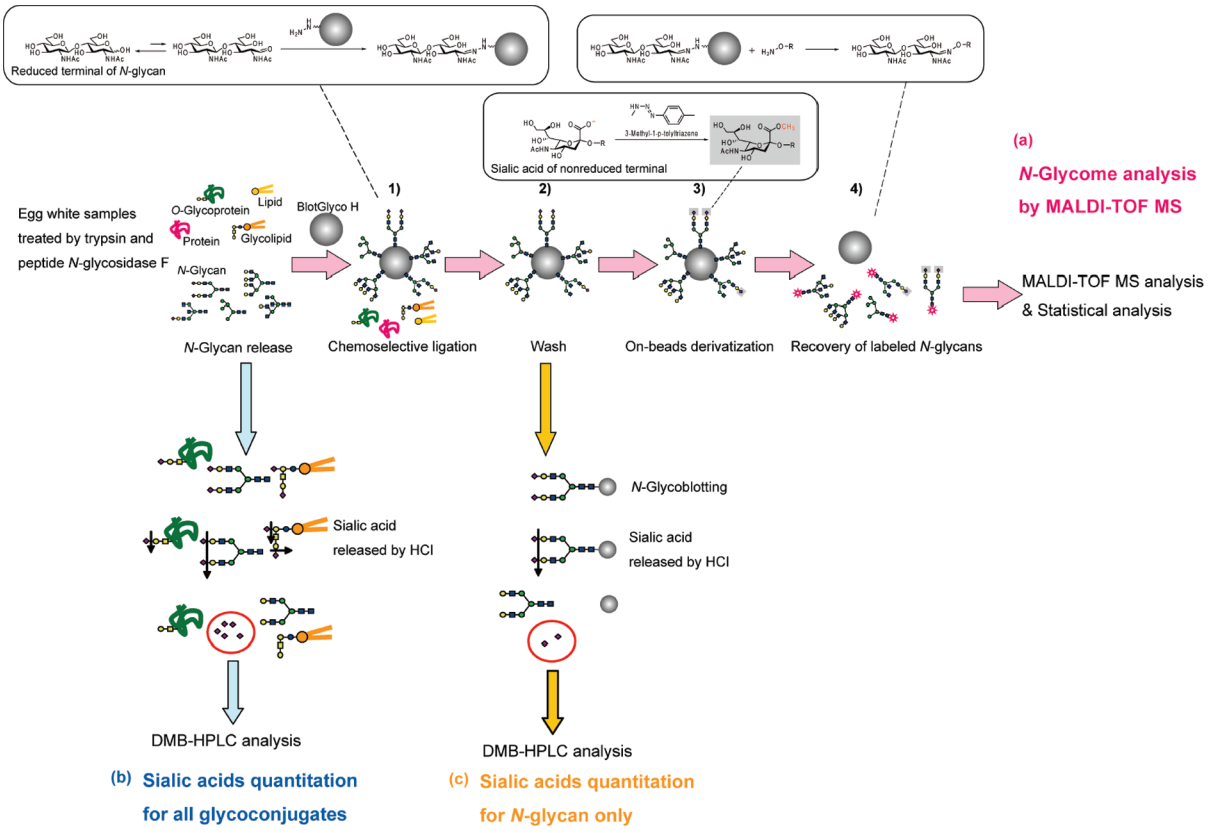


Figure 1. Workflow of glycoblotting-based large-scale avioglycomics for egg whites. (a) *N*-Glycome analysis by MALDI-TOF MS. The steps include (1) chemoselective capturing of reducing sugars onto a hydrazide-functionalized bead, BlotGlycoH, (2) washing to remove any impurities, (3) on-bead methyl esterification of sialic acid residues followed by *trans*-iminization, and (4) recovery of glycans in the form of oxime derivatives of the aminooxy compound that had been added. (b) Sialic acids quantitation for all glycoconjugates from egg whites. The same sample mixtures for *N*-glycoblotting from egg whites of Anseriformes and Galliformes were acid-hydrolyzed to release sialic acids. The hydrolysate was reacted with DMB to develop the fluorescence for the assessment of sialic acids. The reaction mixtures were applied to a reversed-phase column. (c) Sialic acid quantitation for only *N*-glycan from egg whites. The sample mixtures containing released *N*-glycans from egg whites were applied onto BlotGlycoH to capture glycans specifically onto beads via stable hydrazone bonds. Unreacted hydrazide functional groups on beads were acetyl-capped. *N*-Glycans on beads were acid-hydrolyzed to release sialic acids. The filtered hydrolysate containing sialic acids was reacted with DMB to develop the fluorescence for the assessment of sialic acids. The reaction mixtures were applied to a reversed-phase column.

egg whites (0.5 mg each) were dissolved in 20 μL of 200 mM ammonium bicarbonate, and then aliquots were placed in a 96-well polymerase chain reaction (PCR) plate. As the internal standard, 26 μL of 42 μM disialogalactosylated biantennary *N*-glycan containing amidated sialic acids (A2amide glycan) was also added and mixed in each well. A mixture (54 μL) of 0.06% 1-propanesulfonic acid, 2-hydroxyl-3-myristamido (PHM), and 12 mM dithiothreitol (DTT) in 105 mM ammonium bicarbonate was added to the sample mixture. The solubilized proteinaceous materials were reduced by DTT at 37 °C for 90 min followed by alkylation with 10 μL of 123 mM iodoacetamide (IAA) by incubation in the dark at 25 °C for 60 min. The mixture was then treated with 5 μL of 40 units/μL trypsin (Sigma-Aldrich, Inc.) at 37 °C for 3 h, followed by heat inactivation of the enzyme at 90 °C for 10 min. After being cooled to room temperature, *N*-glycans were enzymatically released from trypsin-digested glycopeptides by incubation with 2 units of PNGase F (Roche Applied Science, Basel, Switzerland) at 37 °C for 16 h. Then the sample mixture was digested with 10 μL of 0.5 unit/μL Proteinase K at 37 °C for 3 h, followed by heat inactivation of the enzyme at 90 °C for 10 min, and stored at −20 °C until the mixture was use. This series of protocols was applied to the automatic operation for multiple samples simultaneously by SweetBlot according to the previous report[24] (Figure 1a).

Enrichment and Labeling of N-Glycans. Aliquots of BlotGlycoH beads (500 μL, Sumitomo Bakelite, Co.) in a 10 mg/mL suspension with water were placed into a well of a MultiScreen Solvinert filter plate (Millipore, Billerica, MA), and the water was removed by vacuum. The sample mixtures (20 μL) containing released *N*-glycans from egg whites were applied to the well followed by the addition of 180 μL of 2% acetic acid (AcOH) in acetonitrile (ACN). The plate was incubated at 80 °C for 45 min to dryness in a thermostat to capture glycans in sample mixtures onto beads via stable hydrazone bonds. The plate was washed with 200 μL of 2 M guanidine-HCl in ammonium bicarbonate followed by washing with the same volume of water and 1% triethylamine in methanol (MeOH). Each washing step was performed twice. Unreacted hydrazide functional groups on beads were capped by incubation with 10% acetic anhydride in MeOH at 25 °C for 30 min. The solution was removed by vacuum, and then the beads were washed twice with 200 μL of 10 mM HCl, MeOH, and dioxane, successively. On-bead methyl esterification of carboxyl groups in sialic acids was conducted via incubation with 100 mM 3-methyl-1-*p*-tolyltriazene (MTT) in

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

dioxane at 60 °C for 60 min to dryness.[41] Then the bead was serially washed with 200 μL of dioxane, water, MeOH, and water. The glycans blotted on beads were subjected to the *trans*-iminization reaction with 20 μL of 50 mM *O*-benzyloxyamine hydrochloride (BOA), followed by the treatment with 180 μL of 2% AcOH in ACN for 45 min at 80 °C. The BOA-labeled *N*-glycans were eluted via treatment with 100 μL of water (Figure 1a). We have already reported that the glycoblotting method achieved approximately 72.5 ± 5.0% of *N*-glycans from whole human serum.[25]

MALDI-TOF MS Aanalysis. The recovered BOA-labeled *N*-glycans were mixed with an equivalent volume of the liquid matrix solution consisting of 100 mM α-cyano-4-hydroxycinnamic acid (CHCA) diethylamine salt, followed by deposition of 2.5 μL of the mixture on an MTP 384 target plate (polished steel TF, Bruker Daltonics), and dried under vacuum to afford crystals of the analytes. Then the analytes were subjected to MALDI-TOF MS analysis using Ultraflex III (Bruker Daltonics) in reflector, positive ion mode, typically totaling 200 × 5 shots.[24−26]

Statistical Analysis. The *N*-glycan peaks detected in MALDI-TOF MS spectra were picked using FlexAnalysis (Bruker Daltonics). The area of the isotopic peaks of each glycan was normalized to 1092 pmol of the internal standard. The compositions and structures of glycans were suggested by GlycoMod Tool (http://br.expasy.org/tools/glycomod/) and GlycoSuiteDB (http://glycosuitedb.expasy.org/glycosuite/glycodb). To identify *N*-glycans to categorize to Anseriformes and Galliformes, we performed classification by hierarchical clustering and principal component analysis (PCA) using Spotfire (Somerville, MA) DecisionSite version 9.0. Hierarchical clustering was performed on an MS data set for a total of 61 *N*-glycans or a data set of 15 glycotypes[26] in 88 samples (Figure 2b and Figure S2 of the Supporting Information, respectively). Hierarchical clustering was initiated by calculation of the similarity between all the possible combinations of two records using a selected similarity measure. The calculated similarity data are then used to derive the similarity between all clusters that are formed from the records during the clustering. The following methods were used in this study: clustering methods, unweighted pair-group method with arithmetic mean (UPGMA); similarity measure, Euclidean distance.[42−45] PCA was performed on an MS data set for a total of 61 *N*-glycans in 88 samples, and five clusters of hierarchical clustering of total *N*-glycomes (Figure 2d) or glycotypes (Figure S2 of the Supporting Information) were separately colored. PCA transforms a set of correlated variables into a smaller set of uncorrelated variables called principal components. The result can be regarded as a new data set with fewer variables. The results of PCA calculation were displayed in a three-dimensional scatter plot (scores plot) mapping the principal component score of each projected record. Each point in the plot represents a record in the MS data set of 61 *N*-glycans. The position along a certain axis represents the score of the record on that principal component.[45]

Fluorometric HPLC Method for the Quantitation of Sialic Acids in Egg Whites. *Sialic Acids Derived from All Glycoconjugates.* Fluorometric analysis to determine the levels of *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc), and 2-keto-3-deoxynonulosonic acid (KDN) was conducted via HPLC using DMB as previously described.[11] The mixtures (2.5 μL each) prepared from egg whites of Anseriformes and Galliformes were diluted with water (1:1) and subjected directly to the hydrolysis by being treated with 200 μL of 25 mM HCl at 80 °C for 1 h. The hydrolysate was reacted with 200 μL of 1,2-diamino-4,5-metylenedioxybenzene (DMB) reagent (7 mM DMB-2HCl, 1 M 2-mercaptoethanol, and 18 mM $Na_2S_2O_4$) and heated at 60 °C for 2.5 h in the dark to develop the fluorescence for the determination of the amounts of sialic acids. The reaction mixtures (30 μL each) were subjected to chromatography on a reversed-phase column (Inertsil ODS-3, 4.6 mm × 250 mm, GL Sciences Inc., Tokyo, Japan). The retention times of individual peaks on the column were normalized with standard mixtures of Neu5Ac, Neu5Gc, and KDN. The DMB derivatives are usually detected within 12 min (ε_{ex} = 373 nm, and ε_{em} = 448 nm). The column was eluted at 1 mL/min with a MeOH/ACN/H_2O mixture (3:1:10, v/v). For DMB-HPLC analysis, a D-7000 HPLC system equipped with an L-7485 fluorescence detector (Hitachi High-Technologies Co., Tokyo, Japan) was used (Figures 1b and 3).

Sialic Acids Derived from N-Glycans. The crude mixtures (20 μL each) containing released *N*-glycans from egg whites were subjected to glycoblotting with BlotGlycoH beads, followed by the addition of 180 μL of 2% AcOH in ACN. The plate was incubated at 80 °C for 45 min to dryness to capture whole *N*-glycans in the sample mixtures specifically onto beads via stable hydrazone bonds. The plate was washed with 200 μL of 2 M guanidine-HCl in ammonium bicarbonate followed by washing with the same volume of water and then with 1% triethylamine in MeOH. Each washing step was performed twice. Unreacted hydrazide groups on beads were capped by incubation with 10% acetic anhydride in MeOH at 25 °C for 30 min. The solution was removed by vacuum, and then the beads were washed twice with 200 μL of 10 mM HCl, MeOH, and water, successively. To each well was added 100 μL of 25 mM HCl, and the wells were sealed; the plate was incubated at 80 °C for 1 h to hydrolyze *N*-glycans on beads. After evacuation and filtration, we collected approximately 100 μL of hydrolysate containing sialic acids. The hydrolysate was treated with 100 μL of DMB reagent (7 mM DMB-2HCl, 1 M 2-mercaptoethanol, and 18 mM $Na_2S_2O_4$) and heated at 60 °C for 2.5 h in the dark to develop the fluorescence for determination of the amounts of sialic acids. The reaction mixtures (10 μL each) were subjected to chromatography on a reversed-phase column. The following steps were performed in a similar manner as described above (Figures 1c and 4).

RESULTS

Large-Scale *N*-Glycomics of Egg Whites from Anseriformes and Galliformes. A glycoblotting method[22] is a powerful tool for large-scale and quantitative glycomics. Glycoblotting-based high-throughput glycomics has proven to be a valuable, comprehensive approach for the discovery of novel clinical biomarkers using human serum, urine, cultured cells, and formalin-fixed paraffin-embedded tissue sections.[23−27] Our recent development of SweetBlot,[24] an automated machine designed for large-scale glycoblotting, is suitable for the collection of sufficient data to be analyzed by such method. Here we examined total *N*-glycomics of egg whites from 88 birds of two orders to demonstrate the feasibility of the glycoblotting-based large-scale glycan analysis. The strategy of avio-*N*-glycomics by MALDI-TOF MS is diagrammed in Figure 1a. First, glycoblotting, chemoselective capturing of reducing sugars onto a stable hydrazide-functionalized polymer, was conducted by using BlotGlycoH beads.[25] The hydrazide groups of BlotGlycoH beads were proven to react with aldehyde or ketone groups that are rather rare in common biological samples except carbohydrates

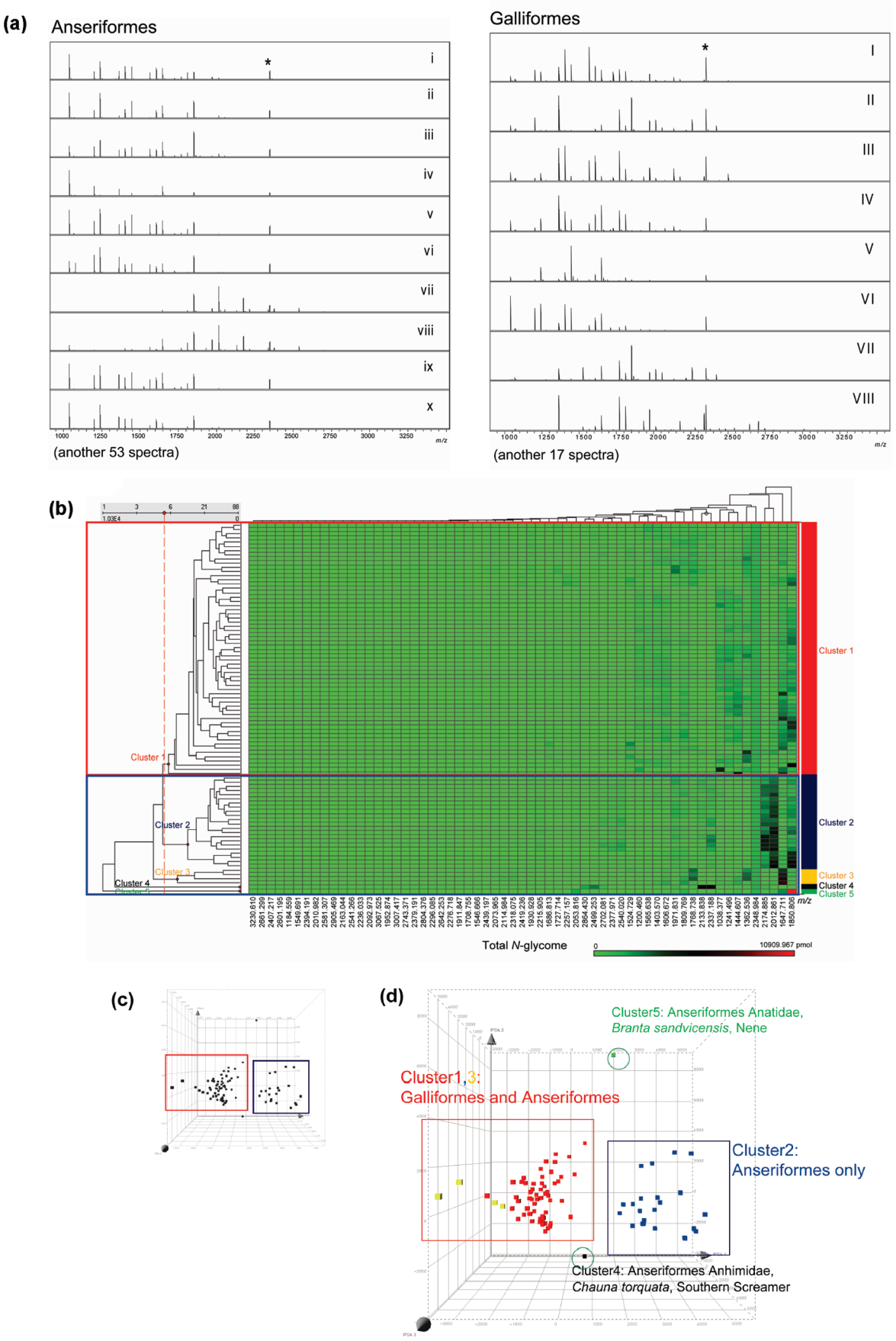


Figure 2. Continued

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

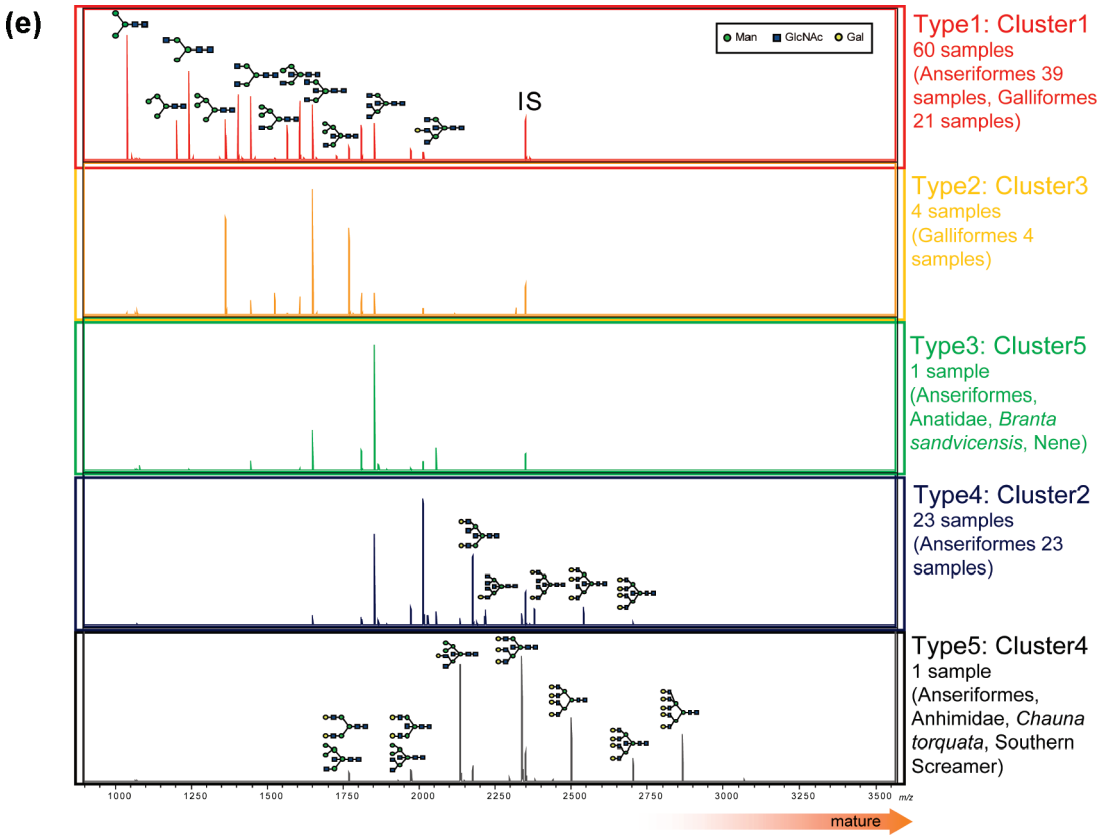


Figure 2. Profiles of *N*-glycomes in egg white from Anseriformes and Galliformes. (a) MALDI-TOF MS spectra of *N*-glycans in egg whites (10 representative spectra of Anseriformes and eight representative spectra of Galliformes): (i) *Aix sponsa*, wood duck; (ii) *Anas clypeata*, northern shoveler; (iii) *Anas laysanensis*, Laysan duck; (iv) *Anas platyrhynchos*, mallard; (v) *Anas strepera*, gadwall; (vi) *Anas rubripes*, American black duck; (vii) *Anser albifrons*, greater white-fronted goose; (viii) *Anser rossii*, Ross's goose; (ix) *Aythya affinis*, lesser scaup; (x) *Aythya fuligula*, tufted duck; (I) *Francolinus adspersus*, red-billed francolin; (II) *Galloperdix spadicea*, red spurfowl; (III) *Syrmaticus reevesii*, Reeves's pheasant; (IV) *Tympanuchus phasianellus*, sharp-tailed grouse; (V) *Callipepla squamata*, scaled quail; (VI) *Colinus virginianus*, bobwhite quail; (VII) *Gallus gallus domesticus*, chicken; (VIII) *Meleagris gallopavo*, wild turkey. We analyzed a total of 88 avian egg whites (63 samples of Anseriformes and 25 samples of Galliformes). Egg whites digested with trypsin and PNGase F were directly subjected to the protocol for *N*-glycan enrichment and derivatization using BlotGlycoH. Note that sialyl *N*-glycans processed with on-bead methyl esterification were stabilized from desialylation so that quantitative glycomic analysis of total *N*-glycans could be conducted. An asterisk indicates a given amount of internal standard, A2amide. (b) Hierarchical clustering of 88 samples of egg whites from Anseriformes and Galliformes in the total *N*-glycome (total of 61 glycans, including an internal standard). This visualization shows records in a dendrogram (a tree graph) based on the similarity between them. Columns include m/z values of 61 glycans (Table 1 and Table S3 of the Supporting Information). (c) Principal component analysis (PCA) of all 88 data sets of the egg white *N*-glycome from Anseriformes and Galliformes. The three-dimensional PCA plot shows the principal component score of each projected record. Columns included m/z values of 61 glycans (Table 1 and Table S3 of the Supporting Information). (d) PCA of all 88 data sets of the egg white *N*-glycome from Anseriformes and Galliformes (panel c) colored by five clusters of hierarchical clustering of total *N*-glycomes (panel b and Table 2). (e) MALDI-TOF MS spectra of *N*-glycans in egg whites (representative spectra of five types): type 1, cluster 1, 60 samples (39 samples from Anseriformes and 21 samples from Galliformes); type 2, cluster 3, four samples (Galliformes); type 3, cluster 5, one sample (Anseriformes, *Anatidae*, *Branta sandvicensis*, nene); type 4, cluster 2, 22 samples (Anseriformes); type 5, cluster 4, one sample (Anseriformes, *Anhimidae*, *Chauna torquata*, southern screamer).

with a reducing terminus. Then, they were washed vigorously to remove any impurities. It is well documented that the glycoside linkages between sialic acid and other sugar residues appear to be acid-labile. For example, a sialic acid residue may be released by hydrolysis under common acidic conditions such as treatment with 25 mM HCl at 80 °C for 1 h or 2 M AcOH at 80 °C for 3 h, whereas a fucose residue cannot be released under such mild acid conditions.[46] Under a general condition for glycoblotting protocols using organic solvent-based (ACN) systems containing 2% AcOH, it was demonstrated that hydrolytic digestion both sialic acid and fucose residues is negligible.[41] Similarly, they were not released by treatment with 10 mM HCl at room temperature. Next, on-bead methyl esterification of enriched glycans facilitates interconversion of sialic acid residues into stable neutral sugar residues via protection of free carboxyl groups. The recovered whole *N*-glycans containing neutral and methyl-protected acidic carbohydrates are subjected to quantitative MS analysis without any loss under high-energy laser irradiation during MALDI-TOF MS, while common intact sialyl glycans are known to decompose under general MS conditions because of the free COOH groups at anomeric positions of sialic acid residues.[41] Because stabilization of sialic acids is essential for the quantitative MS analysis of sialyl oligosaccharides, a convenient *O*-methyl esterification of sialic acid residues by 3-methyl-1-*p*-tolyltriazene (MTT)[41] was included in general glycoblotting protocols.[25−27] Finally, stabilized *N*-glycans were released and recovered in the form of oxime derivatives by *trans*-iminization reaction with benzyloxyamine (BOA) and subjected to MALDI-TOF MS. A series of procedures

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

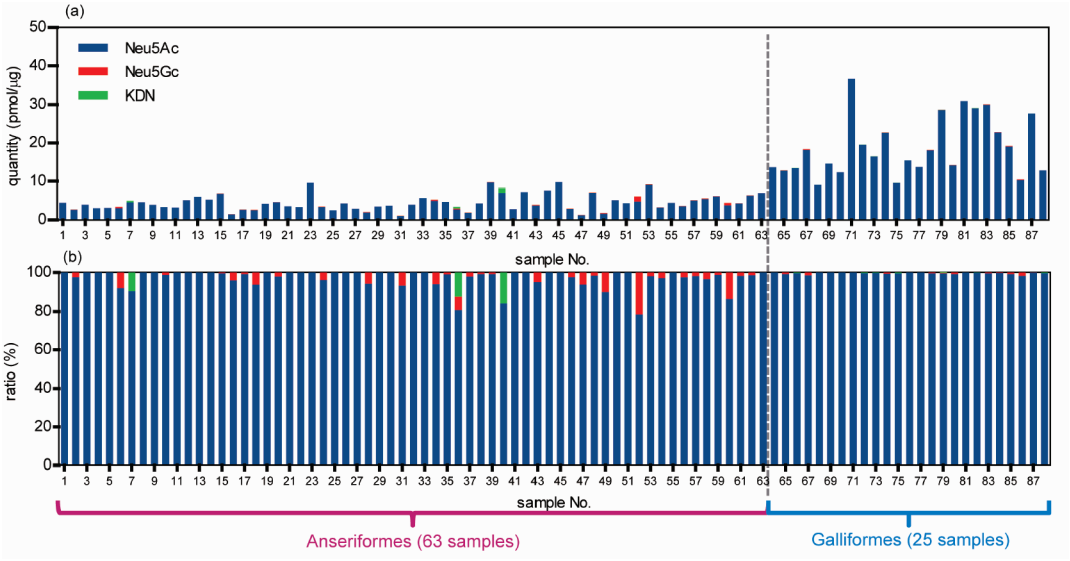


Figure 3. Quantitation of sialic acids for all glycoconjugates in egg whites. Acid-hydrolyzed sialic acids from 88 egg whites (63 samples from Anseriformes and 25 samples from Galliformes) were derivatized using DMB and quantified, and the ratios of Neu5Ac, Neu5Gc, and KDN were measured by reversed-phase HPLC. The bar chart represents the quantities (picomoles per microgram) of Neu5Ac, Neu5Gc, and KDN (a) and the ratios of Neu5Ac, Neu5Gc, and KDN to total sialic acids (Neu5Ac + Neu5Gc + KDN) (b). Blue, red, and green columns represent the quantities or percentages of Neu5Ac, Neu5Gc, and KDN, respectively.

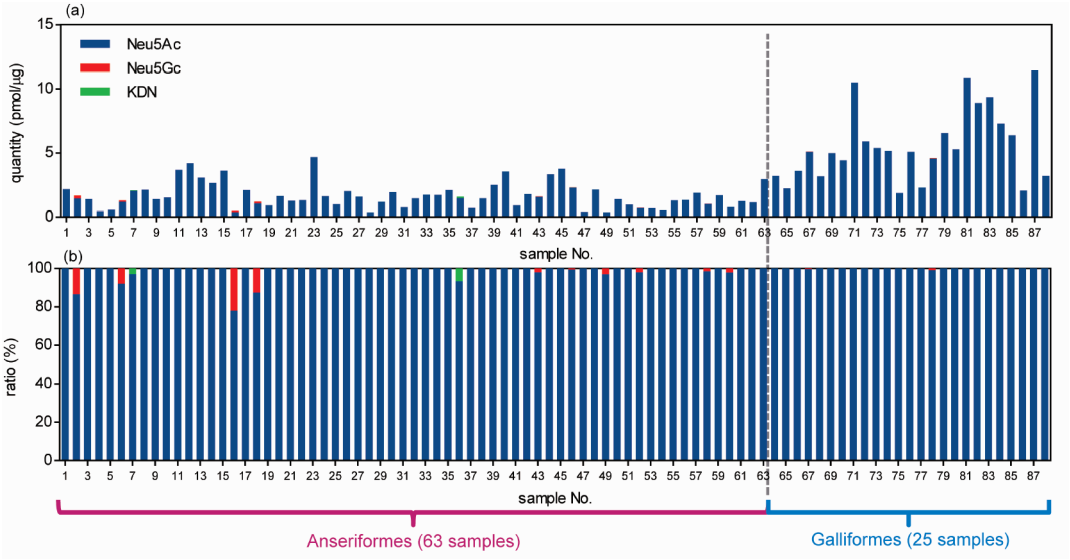


Figure 4. Quantitation of sialic acids for only *N*-glycans in egg whites. Acid-hydrolyzed sialic acids from *N*-glycans of 88 egg whites (63 samples from Anseriformes and 25 samples from Galliformes) were derivatized using DMB and quantified, and the ratios of Neu5Ac, Neu5Gc, and KDN were measured by reversed-phase HPLC. The bar chart represents the quantities (picomoles per microgram) of Neu5Ac, Neu5Gc, and KDN (a) and the ratios of Neu5Ac, Neu5Gc, and KDN to total sialic acids (Neu5Ac + Neu5Gc + KDN) (b). Blue, red, and green columns represent the quantities or percentages of Neu5Ac, Neu5Gc, and KDN, respectively.

[Figure 1a (1−4)] were needed to achieve reliable and reproducible mass measurements for both neutral and acidic glycoforms in the positive reflector mode in MALDI-TOF (Supporting Information). To illustrate the method, egg white (0.5 mg) was subjected to the optimized protocol using BlotGlycoH beads (Figure 1). An aliquot of the "ready-to-analyze" sample solution equivalent to 9.6 ng of egg white was directly deposited on a target plate and subjected to the MALDI-TOF analysis (positive, reflector mode). We quantified 61 different *N*-glycan peaks that could be detected in the total of 88 egg whites (63 from Anseriformes and 25 from Galliformes) and normalized their abundance to peak 40 (spiked A2amide) as an internal standard (Table 1, Tables S1 and S3 of the Supporting Information, Figure 2a, and Figure S4 of the Supporting Information). Herein, "61" means the number of detected MS peaks due to *N*-glycans, but this does not correspond to the number of *N*-glycan structures, because these MS peaks often involve some isomeric *N*-glycan structures. Thus, the real number of *N*-glycan structures could be much higher

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

than the number of observed MS peaks, and it should be noted that putative glycan structures listed in Table 1 were achieved on the basis of the known structures in refs 47−55 and GlycoSuiteDB online, the ExPASy proteomics server (http://glycosuitedb.expasy.org/glycosuite/glycodb).

Identification of Differential Profiles of *N*-Glycomes by Cluster Analysis. Cluster analysis is useful for classifying observed data sets, and the hierarchical clustering tool groups records and arranges them in a dendrogram on the basis of the similarity between them. The hierarchical clustering procedure for a data set of 61 *N*-glycan peaks of 88 samples from Anseriformes and Galliformes (total 88 records) is as follows: (a) selection of a definition of similarity by Euclidean distance, (b) calculation of the similarity between all the possible combinations of two records using Euclidean distance and formation of a new cluster from the two nearest records, (c) selection of a definition of clustering method, unweighted pair-group method with arithmetic mean (UPGMA), (d) calculation of the similarity between all clusters formed from the records by using UPGMA and formation of a new cluster from the two nearest clusters, and (e) repeating the above operation until 88 records become one cluster.[42−45] The result showed two major classifications (red and blue frames, Figure 2b) and five submajor classifications, clusters 1−5 (Figure 2b,d and Table 2). Cluster 1 consisted of 60 samples: 39 from Anseriformes and 21 from Galliformes. Cluster 2 consisted of 22 samples from Anseriformes. Cluster 3 consisted of four samples from Galliformes. Cluster 4 contained only one sample from Anseriformes (*Anhimidae*, *Chauna torquata*, southern screamer). Cluster 5 also contained only one sample from Anseriformes (*Anatidae*, *Branta sandvicensis*, nene).

Hierarchical clustering of 88 samples of egg whites from Anseriformes and Galliformes in glycotype[26] also showed two major classifications (red and blue frames in Figure S2 of the Supporting Information) and five submajor classifications, clusters A−E (Figure S2 and S3 and Table S1 of the Supporting Information). Cluster A consisted of 61 samples (39 from Anseriformes and 22 from Galliformes), almost corresponding to cluster 1. Cluster B consisted of three Galliformes samples, almost corresponding to cluster 3. Cluster C consisted of one Anseriformes sample (*Anatidae*, *Branta sandvicensis*, nene), corresponding to cluster 5. Cluster D consisted of 22 Anseriformes samples corresponding to cluster 2. Cluster E consisted of one Anseriformes sample (*Anhimidae*, *Chauna torquata*, southern screamer), corresponding to cluster 4.

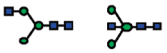
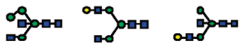
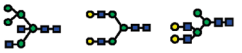
Principal component analysis (PCA) of data set for a total 61 *N*-glycan peaks of 88 samples showed two major classifications (Figure 2c). The larger cluster corresponds to cluster 1 and cluster 3, cluster A and cluster B, respectively, in hierarchical clustering (Figure 2d and Figure S3 of the Supporting Information). The other cluster corresponds to cluster 2 or cluster D in hierarchical clustering (Figure 2d and Figure S3 of the Supporting Information).

MALDI-TOF MS spectra of *N*-glycome in egg whites, representative spectra of five types, are shown in Figure 2e. A type number was assigned when it became a large, mature structure. Type 1 corresponds to cluster 1, type 2 to cluster 3, type 3 to cluster 5, type 4 to cluster 2, and type 5 to cluster 4, as shown in the right column of Figure 2e and Table 2. Type 1 was the major MS pattern in this study, in which both Anseriformes and Galliformes were classified. These profiles had a core structure of *N*-glycan (Man3) and Man4, Man5 corresponding to high-mannose, hybrid, and complex types. Hybrid- and complex-type *N*-glycans detected in this type did not have many galactoses (Gals) attached to outer GlcNAc residues. Type 2 belongs to Galliformes only. Major *N*-glycans in this group had "bisected" structure in both complex and hybrid types. Type 2 also did not have many Gals attached to GlcNAc. Type 3 was classified by one species (Anseriformes, Anatidae, *Branta sandvicensis*, nene). Major structures of *N*-glycans in this group were complex types that have bisected structures, with few Gals. Type 4 was the second major pattern of *N*-glycans, belonging to Anseriformes only. Major structures of *N*-glycans were of the complex type and carried highly branched, bisected structures, and GlcNAc-linked Gals. Type 5 was classified by one sample (Anseriformes, *Anhimidae*, *Chauna torquata*, southern screamer). Major *N*-glycans can be profiled as hybrid and complex types that have highly branched and bisected structures. Almost all *N*-glycans carried GlcNAc-linked Gals. Therefore, type 1, type 2, and type 3 can be called "trans-Gal(−)", but type 4 and type 5 can be called "trans-Gal(+)".

Quantitative Analysis of Sialic Acids of Anseriformes and Galliformes Egg White. *Sialic Acids Derived from All Glycoconjugates.* The sialic acid species were assessed via HPLC using 1,2-diamino-4,5-methylenedioxybenzene (DMB) as a precolumn fluorogenic labeling reagent.[11] Sialic acid analysis of all glycoconjugates, namely *N*- and *O*-glycoproteins and glycosphingolipids, was performed using 88 samples of egg white from Anseriformes and Galliformes (Figure 1b). As shown in Figure 3, the major molecular species of sialic acid in both Anseriformes and Galliformes egg white was *N*-acetylneuraminic acid (Neu5Ac), although small amounts of *N*-glycolylneuraminic acid (Neu5Gc) and 2-keto-3-deoxynonulosonic acid (KDN) were also detected. The total amounts of sialic acids were 1.1−9.8 pmol/μg (average of 4.6 pmol/μg) for egg whites from Anseriformes and 9.1−36.6 pmol/μg (average of 18.9 pmol/μg) for egg whites from Galliformes. Thus, the total amount of sialic acids in egg whites from Galliformes was 4 times greater than that from Anseriformes. However, the amounts of the minor molecular species of sialic acids such as Neu5Gc or KDN in egg whites from Anseriformes were greater than that from Galliformes. The Neu5Gc or KDN ratio (percent) of sialic acids in Anseriformes was higher than those in Galliformes. In particular, samples 51, 60, 6, 34, 28, 37, 36, and 42 showed relatively high levels of Neu5Gc (>10%, in descending order), while samples 36, 7, and 39 exhibited high levels of KDN (>10%, in descending order).

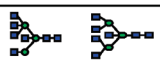
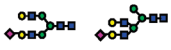
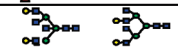
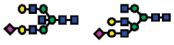
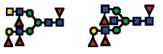
Sialic Acids Derived from N-Glycans. Sialic acid analysis of *N*-glycans captured by glycoblotting of 88 egg whites from Anseriformes and Galliformes (Figure 1c) was also conducted in a manner similar to that described above based on the DMB-HPLC method.[11] As shown in Figure 4, the major molecular species of sialic acids in both Anseriformes and Galliformes egg whites was Neu5Ac, although Neu5Gc and KDN could be detected in small quantities. The total amounts of sialic acids were 0.4−4.7 pmol/μg (average of 1.7 pmol/μg) for egg whites from Anseriformes and 1.9−11.4 pmol/μg (average of 5.6 pmol/μg) for egg whites from Galliformes. The sialic acids contents in egg whites from Galliformes were 3 times greater than those from Anseriformes. However, the contents of the minor variants of sialic acids, Neu5Gc and KDN, in egg whites from Anseriformes were greater than those from Galliformes. The Neu5Gc or KDN ratio (percent) of sialic acids in Anseriformes was higher than those in Galliformes. In particular, samples 16, 2, 18, and 6 exhibited higher levels of Neu5Gc (>5%, in descending order) than others, and samples 36 and 7 carried KDN.

Table 1. Total of 61 *N*-Glycan Peaks, Including an Internal Standard, Derived from Anseriformes (49 *N*-glycan peaks) and Galliformes (52 *N*-glycan peaks) Targeted in This Study[a]

Peak No.	*m/z*	Composition	Putative *N*-glycan structure	Reference No.	Link to DB	
1	1038.38	(Hex)3 (HexNAc)2		49,50,51,52,55	GlycoSuiteDB	
2	1184.56	(Hex)3 (HexNAc)2 (dHex)1		Low abundance	GlycoSuiteDB	
3	1200.46	(Hex)4 (HexNAc)2		47,48	GlycoSuiteDB	
4	1241.5	(Hex)3 (HexNAc)3		48,49,50,53	GlycoSuiteDB	
5	1362.54	(Hex)5 (HexNAc)2		48,49,50	GlycoSuiteDB	
6	1403.57	(Hex)4 (HexNAc)3		48,49,50,53	GlycoSuiteDB	
7	1444.61	(Hex)3 (HexNAc)4		47,48,49,50,53	GlycoSuiteDB	
8	1524.73	(Hex)6 (HexNAc)2		48,49,50	GlycoSuiteDB	
9	1546.67	(Hex)3 (HexNAc)3 (Neu5Ac)1		MS/MS in Fig.S4	Carbbank	
10	1549.69	(Hex)4 (HexNAc)3 (dHex)1		Low abundance	GlycoSuiteDB	
11	1565.64	(Hex)5 (HexNAc)3		48,49,50	GlycoSuiteDB	
12	1606.67	(Hex)4 (HexNAc)4		48,50,49	GlycoSuiteDB	
13	1647.71	(Hex)3 (HexNAc)5		48,49,50,53	GlycoSuiteDB	
14	1686.81	(Hex)7 (HexNAc)2		48,49,50	GlycoSuiteDB	
15	1708.76	(Hex)4 (HexNAc)3 (Neu5Ac)1		MS/MS in Fig.S4	GlycoSuiteDB	
16	1727.71	(Hex)6 (HexNAc)3		48,50	GlycoSuiteDB	
17	1768.74	(Hex)5 (HexNAc)4		48,49,50,52	GlycoSuiteDB	
18	1809.77	(Hex)4 (HexNAc)5		48,50	GlycoSuiteDB	
19	1850.81	(Hex)3 (HexNAc)6		48,49,50,53	GlycoSuiteDB	
20	1911.85	(Hex)4 (HexNAc)4 (Neu5Ac)1		50	GlycoSuiteDB	
21	1930.93	(Hex)6 (HexNAc)4		52	GlycoSuiteDB	
22	1952.87	(Hex)3 (HexNAc)5 (Neu5Ac)1		50	No DB match	
23	1971.83	(Hex)5 (HexNAc)5		48,49,50,52	GlycoSuiteDB	
24	2010.98	(Hex)9 (HexNAc)2		50	GlycoSuiteDB	

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

Table 1. Continued

Peak No.	*m/z*	Composition	Putative *N*-glycan structure	Reference No.	Link to DB	
25	2012.86	(Hex)4 (HexNAc)6		48,49,50,53	GlycoSuiteDB	
26	2053.82	(Hex)3 (HexNAc)7		48,49,50,53	GlycoSuiteDB	
27	2073.97	(Hex)5 (HexNAc)4 (Neu5Ac)1 * (Hex)4 (HexNAc)4 (dHex)1 (Neu5Gc)1		50,52	GlycoSuiteDB	
28	2092.97	(Hex)7 (HexNAc)4		52	GlycoSuiteDB	
29	2114.98	(Hex)4 (HexNAc)5 (Neu5Ac)1		50	GlycoSuiteDB	
30	2133.84	(Hex)6 (HexNAc)5		48,51,52	GlycoSuiteDB	
31	2163.04	(Hex)5 (HexNAc)3 (dHex)2 (Neu5Ac)1		Low abundance	Glycosciences.de DB	
32	2174.89	(Hex)5 (HexNAc)6		48,51	GlycoSuiteDB	
33	2215.91	(Hex)4 (HexNAc)7		48,49,53	GlycoSuiteDB	
34	2236.03	(Hex)6 (HexNAc)4 (Neu5Ac)1 * (Hex)5 (HexNAc)4 (dHex)1 (Neu5Gc)1		52	GlycoSuiteDB	GlycoSuiteDB
35	2257.16	(Hex)3 (HexNAc)8		48,49,51,53	GlycoSuiteDB	
36	2276.72	(Hex)5 (HexNAc)5 (Neu5Ac)1		52	GlycoSuiteDB	
37	2296.09	(Hex)7 (HexNAc)5		52	No DB match	
38	2318.08	(Hex)4 (HexNAc)6 (Neu5Ac)1		50	GlycoSuiteDB	
39	2337.19	(Hex)6 (HexNAc)6		48,51	GlycoSuiteDB	
40	2348.98	Internal Standard				
41	2377.97	(Hex)5 (HexNAc)7		48	No DB match	
42	2379.19	(Hex)5 (HexNAc)4 (Neu5Ac)2		50,52	GlycoSuiteDB	
43	2394.19	(Hex)4 (HexNAc)5 (dHex)4		Low abundance	CFG glycan structure DB	
44	2407.22	(Hex)4 (HexNAc)5 (dHex)2 (Neu5Ac)1		Low abundance	GlycoSuiteDB	
45	2419.24	(Hex)4 (HexNAc)8		48,49,50	No DB match	
46	2439.2	(Hex)6 (HexNAc)5 (Neu5Ac)1 * (Hex)5 (HexNAc)5 (dHex)1 (Neu5Gc)1		51,52	GlycoSuiteDB	GlycoSuiteDB
47	2499.25	(Hex)7 (HexNAc)6		51,52	GlycoSuiteDB	
48	2540.02	(Hex)6 (HexNAc)7		48	GlycoSuiteDB	
49	2541.27	(Hex)6 (HexNAc)4 (Neu5Ac)2 * (Hex)5 (HexNAc)4 (dHex)1 (Neu5Ac)1 (Neu5Gc)1		52	GlycoSuiteDB	GlycoSuiteDB
50	2581.31	(Hex)5 (HexNAc)8		48,49	GlycoSuiteDB	_

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

Table 1. Continued

Peak No.	*m/z*	Composition	Putative *N*-glycan structure	Reference No.	Link to DB	
51	2601.2	(Hex)7 (HexNAc)5 (Neu5Ac)1 * (Hex)6 (HexNAc)5 (dHex)1 (Neu5Gc)1		52	GlycoSuiteDB	_
52	2642.25	(Hex)6 (HexNAc)6 (Neu5Ac)1 * (Hex)5 (HexNAc)6 (dHex)1 (Neu5Gc)1		51	GlycoSuiteDB	_
53	2661.3	(Hex)8 (HexNAc)6		51,52	No DB match	
54	2702.08	(Hex)7 (HexNAc)7		50,51	No DB match	
55	2743.37	(Hex)6 (HexNAc)8		48	No DB match	
56	2804.38	(Hex)7 (HexNAc)6 (Neu5Ac)1		51,52	GlycoSuiteDB	
57	2864.43	(Hex)8 (HexNAc)7		52	GlycoSuiteDB	
58	2905.47	(Hex)7 (HexNAc)8		50,55	Carbbank	
59	3007.42	(Hex)7 (HexNAc)7 (Neu5Ac)1 * (Hex)6 (HexNAc)7 (dHex)1 (Neu5Gc)1		51	No DB match	
60	3067.53	(Hex)8 (HexNAc)8		54	Carbbank	
61	3230.61	(Hex)9 (HexNAc)5 (Neu5Ac)2 * (Hex)8 (HexNAc)5 (dHex)1 (Neu5Ac)1 (Neu5Gc)1 * (Hex)7 (HexNAc)5 (dHex)2 (Neu5Gc)2		Low abundance	Carbbank	

* presumptive composition containing Neu5Gc

Detected in both (Black), Anseriformes only (Red), Galliformes only (Blue)

[a] Peak 40 is an internal standard spiked for quantification. Compositional annotation was achieved with GlycoMod, on the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) (http://www.expasy.ch/tools/glycomod/); "61" is the number of *N*-glycan peaks, but not the number of *N*-glycan structures. The real number of *N*-glycan structures could be much higher because of the presence of diverse isomeric *N*-glycans, and putative structures were achieved by references of *N*-glycan structures from egg white[47−55] and GlycoSuiteDB on-line, the ExPASy proteomics server (http://glycosuitedb.expasy.org/glycosuite/glycodb). The linked database pages are shown for all *N*-glycan structures and those references. Asterisks denote the presumptive composition containing Neu5Gc detected in both (black), Anseriformes only (red), or Galliformes only (blue).

DISCUSSION

Although comprehensive genomics and proteomics have been deployed in large-scale studies, glycomics have not been developed to the same extent. Therefore, the argument about the evolution of glycan has been limited only to "partial" structure, such as Galα1→4Gal, existing in the nonreducing terminus, clarified by a conventional glycomics technique. Here we describe the results that allow us to classify birds by the profile of "total" *N*-glycans, namely, *N*-glycome. That was a new challenge tackled with high-throughput glycomics called glycoblotting. Here we demonstrate that the glycoblotting method is a great tool for glycomics, especially for a large-scale analysis.

We quantified 61 different *N*-glycan peaks in MALDI-TOF MS spectra that could be detected in a total of 88 egg whites (63 from Anseriformes and 25 from Galliformes) (Table 1 and Table S2 of the Supporting Information). These *N*-glycan structures share fundamental aspects of the structures elucidated in previous reports.[47−55] Not surprisingly, our new method uncovered a few new *N*-glycans (pink row in Table 1) presumably because of the enhanced sensitivity, but it can also be due to different individual specimens. As we identified the structural characteristics from compositions in Table 1, by known structures in previous reports,[47−55] or databases, or by MS/MS sequencing performed in this study, the real number of proposed glycan structures was much higher than the number of detected MS peaks due to the presence of diverse isomers. From the viewpoints of high-throughput glycan analysis, it seems that even MS pattern information from primary analysis is meaningful because this MS pattern matching technique is quite similar to the DNA−DNA hybridization technique. Although DNA−DNA hybridization is a high-throughput technique that measures the degree of genetic similarity between DNA sequences, it does not require precise analyses of the DNA sequences themselves. Therefore, whether secondary analysis (e.g., MS/MS sequencing) is crucial

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

Table 2. List of 88 Birds (63 from Anseriformes and 25 from Galliformes) in Rank Order of Sample Number[a]

				species		Cluster Analysis		MS pattern
sample No	order	family	genus	Latin name	English name	Fig2(b)	Fig2(d)	Fig2(e)
1	Anseriformes	Anatidae	Aix	*Aix sponsa*	Wood Duck	34	1	type1
2	Anseriformes	Anatidae	Anas	*Anas clypeata*	Northern Shoveler	40	1	type1
3	Anseriformes	Anatidae	Anas	*Anas laysanensis*	Laysan Duck	49	1	type1
4	Anseriformes	Anatidae	Anas	*Anas platyrhynchos*	Mallard	59	1	type1
5	Anseriformes	Anatidae	Anas	*Anas strepera*	Gadwall	38	1	type1
6	Anseriformes	Anatidae	Anas	*Anas rubripes*	American Black Duck	26	1	type1
7	Anseriformes	Anatidae	Anser	*Anser albifrons*	Greater white-fronted goose	74	2	type4
8	Anseriformes	Anatidae	Anser	*Anser rossii*	Ross's Goose,	69	2	type4
9	Anseriformes	Anatidae	Aythya	*Aythya affinis*	Lesser Scaup	37	1	type1
10	Anseriformes	Anatidae	Aythya	*Aythya fuligula*	Tufted Duck	33	1	type1
11	Anseriformes	Anatidae	Branta	*Branta canadensis*	Canada Goose,	61	2	type4
12	Anseriformes	Anatidae	Chloephaga	*Chloephaga picta*	Upland Goose	46	1	type1
13	Anseriformes	Anatidae	Cygnus	*Cygnus atratus*	Black Swan	65	2	type4
14	Anseriformes	Dendrocygnidae	Dendrocygna	*Dendrocygna viduata*	White-faced Whistling Duck	77	2	type4
15	Anseriformes	Anatidae	Mergus	*Mergus serrator*	Red-breasted Merganser	51	1	type1
16	Anseriformes	Anatidae	Oxyura	*Oxyura australis*	Blue-billed Duck	66	2	type4
17	Anseriformes	Anatidae	Somateria	*Somateria mollissima*	Common Eider	42	1	type1
18	Anseriformes	Anatidae	Tadorna	*Tadorna tadornoides*	Australian Shelduck	39	1	type1
19	Anseriformes	Anatidae	Anser	*Anser caerulescens*	Snow Goose	62	2	type4
20	Anseriformes	Anatidae	Aix	*Aix galericulata*	Mandarin Duck	35	1	type1
21	Anseriformes	Anatidae	Anser	*Anser anser*	Greylag Goose	30	1	type1
22	Anseriformes	Anatidae	Anas	*Anas discors*	Blue-winged Teal	31	1	type1
23	Anseriformes	Anatidae	Anas	*Anas erythrorhyncha*	Red-billed Duck	17	1	type1
24	Anseriformes	Anatidae	Anas	*Anas georgica*	Yellow-billed Pintail	32	1	type1
25	Anseriformes	Anatidae	Anas	*Anas gibberifrons*	Sunda Teal	16	1	type1
26	Anseriformes	Anatidae	Anas	*Anas luzonica*	Philippine Duck	29	1	type1
27	Anseriformes	Anatidae	Anas	*Anas platyrhynchos*	Mallard	28	1	type1
28	Anseriformes	Anatidae	Anas	*Anas rubripes*	American Black Duck	23	1	type1
29	Anseriformes	Anatidae	Anas	*Anas superciliosa*	Pacific Black Duck,	36	1	type1
30	Anseriformes	Anatidae	Anas	*Anas versicolor*	Silver Teal	20	1	type1
31	Anseriformes	Anatidae	Anser	*Anser anser*	Greylag Goose	47	1	type1
32	Anseriformes	Anatidae	Anser	*Anser brachyrhynchus*	Pink-footed Goose	73	2	type4
33	Anseriformes	Anatidae	Anser	*Anser cygnoides*	Swan Goose	63	2	type4
34	Anseriformes	Anatidae	Anser	*Anser indicus*	Bar-headed Goose	68	2	type4
35	Anseriformes	Anatidae	Anser	*Anser indicus*	Bar-headed Goose	67	2	type4
36	Anseriformes	Anatidae	Aythya	*Aythya australis*	Hardhead	27	1	type1
37	Anseriformes	Anatidae	Aythya	*Aythya ferina*	Common Pochard	21	1	type1
38	Anseriformes	Anatidae	Branta	*Branta leucopsis*	Barnacle Goose	71	2	type4
39	Anseriformes	Anatidae	Branta	*Branta sandvicensis*	Nene	88	5	type3
40	Anseriformes	Anhimidae	Chauna	*Chauna torquata*	Southern Screamer	87	4	type5
41	Anseriformes	Anatidae	Chenonetta	*Chenonetta jubata*	Maned Duck	22	1	type1
42	Anseriformes	Anatidae	Chloephaga	*Chloephaga poliocephala*	Ashy-headed Goose	41	1	type1
43	Anseriformes	Anatidae	Chloephaga	*Chloephaga rubidiceps*	Ruddy-headed Goose,	43	1	type1
44	Anseriformes	Anatidae	Cygnus	*Cygnus olor*	Mute Swan	64	2	type4
45	Anseriformes	Dendrocygnidae	Dendrocygna	*Dendrocygna arborea*	West Indian Whistling Duck	78	2	type4
46	Anseriformes	Anatidae	Lophodytes	*Lophodytes cucullatus*	Hooded Merganser	24	1	type1
47	Anseriformes	Anatidae	Lophodytes	*Lophodytes cucullatus*	Hooded Merganser	25	1	type1
48	Anseriformes	Anatidae	Mergus	*Mergus serrator*	Red-breasted Merganser	50	1	type1
49	Anseriformes	Anatidae	Netta	*Netta rufina*	Red-crested Pochard	58	1	type1
50	Anseriformes	Anatidae	Oxyura	*Oxyura jamaicensis*	Ruddy Duck,	79	2	type4
51	Anseriformes	Anatidae	Oxyura	*Oxyura vittata*	Lake Duck	82	2	type4
52	Anseriformes	Anatidae	Sarkidiornis	*Sarkidiornis melanotos*	Comb Duck	48	1	type1
53	Anseriformes	Anatidae	Tadorna	*Tadorna tadorna*	Common Shelduck	44	1	type1
54	Anseriformes	Anatidae	Anas	*Anas formosa*	Baikal Teal	57	1	type1
55	Anseriformes	Anatidae	Aythya	*Aythya ferina*	Common Pochard,	45	1	type1
56	Anseriformes	Anatidae	Branta	*Branta canadensis*	Canada Goose,	70	2	type4
57	Anseriformes	Anatidae	Branta	*Branta leucopsis*	Barnacle Goose	72	2	type4
58	Anseriformes	Anatidae	Malacorhynchus	*Malacorhynchus membranaceus*	Pink-eared Duck	55	1	type1
59	Anseriformes	Anatidae	Oxyura	*Oxyura vittata*	Lake Duck	81	2	type4
60	Anseriformes	Anatidae	Oxyura	*Oxyura australis*	Blue-billed Duck	75	2	type4
61	Anseriformes	Anatidae	Oxyura	*Oxyura jamaicensis*	Ruddy Duck,	80	2	type4
62	Anseriformes	Dendrocygnidae	Thalassornis	*Thalassornis leuconotus*	White-backed Duck,	76	2	type4
63	Anseriformes	Anatidae	Anas	*Anas platyrhynchos*	Domestic Duck	52	1	type1
64	Galliformes	Phasianidae	Francolinus	*Francolinus adspersus*	Red-billed Francolin	1	1	type1
65	Galliformes	Phasianidae	Galloperdix	*Galloperdix spadicea*	Red Spurfowl	13	1	type1
66	Galliformes	Phasianidae	Syrmaticus	*Syrmaticus reevesii*	Reeves's Pheasant,	4	1	type1
67	Galliformes	Phasianidae	Tympanuchus	*Tympanuchus phasianellus*	Sharp-tailed Grouse	10	1	type1
68	Galliformes	Odontophoridae	Callipepla	*Callipepla squamata*	Scaled Quail	60	1	type1
69	Galliformes	Odontophoridae	Colinus	*Colinus virginianus*	Northern Bobwhite, Bobwhite Quail	18	1	type1
70	Galliformes	Odontophoridae	Colinus	*Colinus virginianus*	Northern Bobwhite, Bobwhite Quail	19	1	type1
71	Galliformes	Phasianidae	Meleagris	*Meleagris gallopavo*	Wild Turkey	6	1	type1
72	Galliformes	Phasianidae	Pavo	*Pavo cristatus*	Indian Peafowl,	83	3	type2
73	Galliformes	Phasianidae	Alectoris	*Alectoris rufa*	Red-legged Partridge,	54	1	type1
74	Galliformes	Phasianidae	Chrysolophus	*Chrysolophus pictus*	Golden Pheasant,	9	1	type1
75	Galliformes	Phasianidae	Francolinus	*Francolinus coqui*	Coqui Francolin	56	1	type1
76	Galliformes	Phasianidae	Francolinus	*Francolinus sephaena*	Crested Francolin,	2	1	type1
77	Galliformes	Phasianidae	Gallus	*Gallus sonneratii*	Grey Junglefowl	14	1	type1
78	Galliformes	Phasianidae	Pavo	*Pavo cristatus*	Indian Peafowl,	84	3	type2
79	Galliformes	Phasianidae	Pavo	*Pavo muticus*	Green Peafowl,	85	3	type2
80	Galliformes	Phasianidae	Perdix	*Perdix Perdix*	Grey Partridge	53	1	type1
81	Galliformes	Phasianidae	Phasianus	*Phasianus colchicus*	Common Pheasant	8	1	type1
82	Galliformes	Phasianidae	Syrmaticus	*Syrmaticus reevesii*	Reeves's Pheasant,	3	1	type1
83	Galliformes	Phasianidae	Centrocercus	*Centrocercus urophasianus*	Sage Grouse	11	1	type1
84	Galliformes	Phasianidae	Lagopus	*Lagopus lagopus*	Willow Ptarmigan,	12	1	type1
85	Galliformes	Phasianidae	Lagopus	*Lagopus scoticus*	Willow Ptarmigan,	5	1	type1
86	Galliformes	Odontophoridae	Callipepla	*Callipepla gambelii*	Gambel's Quail	86	3	type2
87	Galliformes	Phasianidae	Phasianus	*Phasianus colchicus*	Common Pheasant	7	1	type1
88	Galliformes	Phasianidae	Gallus	*Gallus gallus domesticus*	Chicken	15	1	type1

[a] The results of hierarchical clustering of total *N*-glycomes, consisting of 61 *N*-glycan peaks, with the record of five clusters are listed and colored by each cluster depicted in Figure 2d. The columns corresponding to Galliformes, small waterfowl, and large waterfowl are colored green, black, and blue, respectively.

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

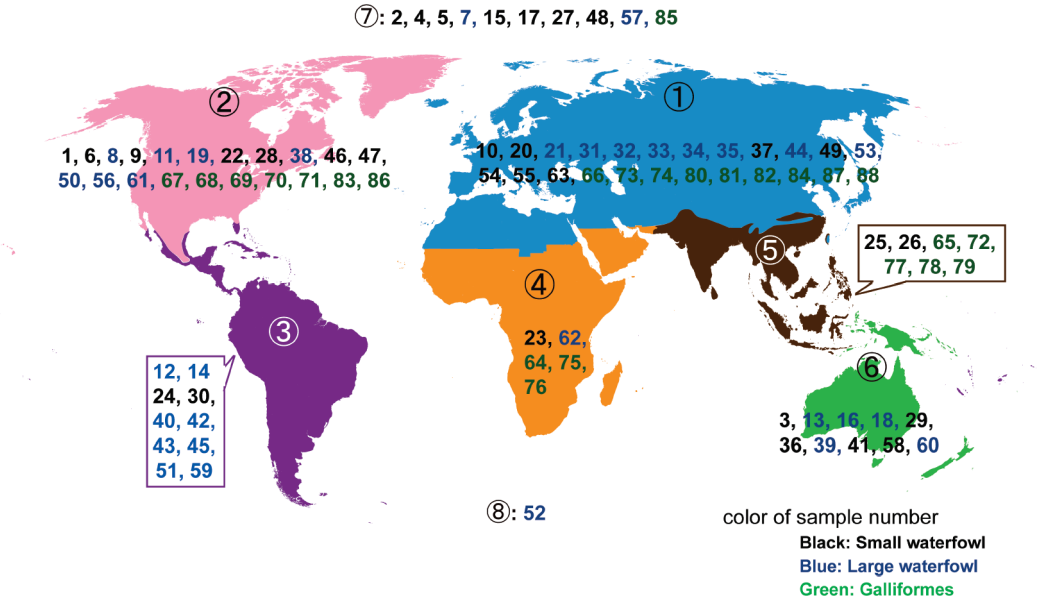


Figure 5. Distribution of egg white samples on the world. Zoogeographical regions 1−6 and their background colors are the same as those in Table S1 of the Supporting Information. The color of the sample number corresponds to the kind of bird. Numbers for Galliformes, small waterfowl, and large waterfowl are colored green, black, and blue, respectively. The character color is the same as that in Table S1.

depends on the needs of the individual study. In particular, statistical analysis for determining the similarity among total glycan profiles by means of large-scale glycome data sets is not always essential for clarifying all isomeric *N*-glycans.

To investigate the relationship among biological reaction, glycosylation, and evolution, we performed clustering analysis of the MS data set of *N*-glycans from 88 egg whites. Interestingly, although we performed hierarchical clustering and PCA analyses of the categorized MS data set prepared by "glycotype"[26] and intact MS data set on the basis of m/z values of glycans (Figure S2 of the Supporting Information and Figure 2), we have obtained approximately similar clustering results. These results mean that many MS data of *N*-glycans can be easily classified without time-consuming steps such as categorization or exact identification of individual glycan structures; hence, we need only intact MS data of whole glycans. It is clear that glycotype analysis is a useful tool, especially when glycans of interest are expressed in minute quantities.[26] However, the information about intact MS data of *N*-glycans is now feasible for large-scale glycomics in classifying biological species and similarities and insight into their evolutional lineage. In other words, we demonstrate that even intact MS data sets of *N*-glycans obtained by large-scale glycomics can be crucial information for the clustering analysis for classifying expression patterns of total *N*-glycans synthesized on the basis individual biosynthetic pathways.

Small waterfowl of Anseriformes were classified as being in cluster 1, the same as Galliformes (Figure 2 and Table 2). It was suggested that the evolution of Galliformes might have proceeded closer to that of small waterfowl than that of large waterfowl such as swans and geese in the same Anatidae family of Anseriformes. While small waterfowl (Figure 2e, cluster 1) had little Gal-modified structure in their nonreducing termini [trans-Gal(−)], large waterfowl (Figure 2e, cluster 2) carried an abundance of highly branched Gal-terminated structures [trans-Gal(+)]. No clear correlation with diet and habitat, namely the zoogeographical region, was observed in the analyses as indicated in Figure 5 (see also Figure S1 of the Supporting Information). Information about plate tectonics tells us that continents changed drastically until the Indian continent crashed into the Eurasian continent 30 million years ago (mya). Via the study of excavated fossils, it was revealed that an ancestor of avian species in the Miocene period (23.5 mya) had complexities almost similar to those of modern birds.[56] Therefore, we can propose two hypotheses. (a) The common ancestor of the small waterfowl [trans-Gal(−)] and the large waterfowl [trans-Gal(+)] had divided after the common ancestor of Galliformes and Anseriformes. It carried little Gal-modified structure and ramified during evolution, which is inferred to have occurred before the topography of the current earth had been completed. The terminal unsubstituted GlcNAc structure might have become disadvantageous with respect to living at the water's edge, because of the possibility of infection by pathogens.[2,57,58] As a result, the structure with terminal Gal was developed to alleviate the problem. This hypothesis seems to disagree with the proposition based on the study of DNA−DNA hybridization by Sibley et al.[38] (Figure 6), which affirms large waterfowl can be placed on the older branch in the avian evolutional dendrogram (Figure 7a). (b) The common ancestor originally was expressed Gal-terminally [trans-Gal(+)]. However, after separation of Anseriforms and Galliforms, it lost the ability to be expressed Gal-terminally, presumably because of the loss of galactosyltransferase activity, which resulted in the present *N*-glycome patterns as classified here (Figure 7b). This hypothesis agreed with results of phylogenomic and morphologic study.[59] Recently, it was clarified that less Gal-terminal structure in chicken egg white [trans-Gal(−)] was caused by a low level of expression of GalT-1 and -2, which are the family members of the α1,4-galactosyltransferase. Although GalT-3 is expressed at a relatively high level, it does not have enzyme activity (personal communication with S. Iijima's group at Nagoya University, Nagoya, Japan). Therefore, GalT-3 might be a key in support of the latter hypothesis.

Our approach is based on the comprehensive analysis of glycome, that is, posttranslational modifications. It is interesting that the information related to avian evolution partially deviates from that from the genomic approach.[59] It may be necessary to

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

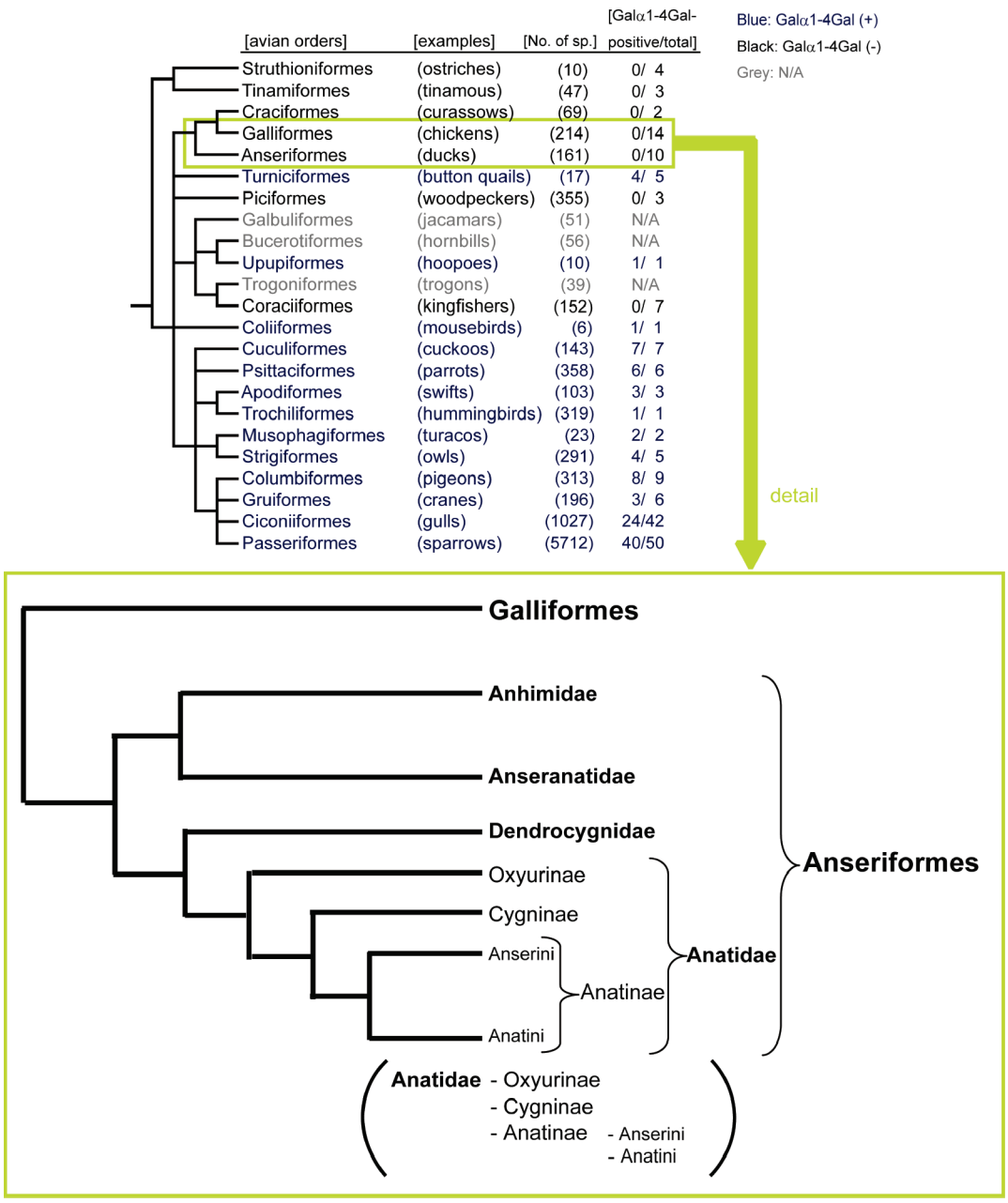


Figure 6. Evolutional dendrogram of modern birds based on DNA−DNA hybridization by Sibley et al.[38]

consider glycomics, in addition to genomics and proteomics, in classifying avian species.

The results of quantification of sialic acids (Figures 3 and 4) showed the tendency that Galliformes have much more sialic acid than Anseriformes, which means small and large waterfowl resemble each other in this regard. In addition, these results did not suggest the correlation with diet and zoogeographical region of inhabitation (Figure 5 and Figure S1 of the Supporting Information). Two interpretations of these results are possible. One is that both Galliformes and Anseriformes had the trans-Gal(−) type of *N*-glycome, just after they had evolved from their common ancestor. Then, the trans-Gal(+) type has gradually evolved in Anseriformes. Another is that the trans-Gal(−) type is suitable for terrestrial environments while the trans-Gal(+) type is appropriate for aquatic environments; thus, the environmental factor can primarily be attributed to the evolution of the *N*-glycome. At the water's edge, the pathogens that pose a threat of infection are conceivably different from those on land. Therefore, the trans-Gal(−) type developed in Galliformes that had lived on land, while the trans-Gal(+) type had built up in Anseriformes that were waterfowl. However, after that, in Anseriformes, the new infectious pathogens became conspicuous, the trans-Gal(+) type became unfavorable, and the trans-Gal(−) type emerged.

Next, our attention was directed to the importance of sialic acids in terms of the difference in the modification at the C-5 position, notably by the *N*-acetyl or *N*-glycolyl group, because sialic acid is an important class of sugar for the investigation of the relationship between evolution and infection. Although we planned to characterize sialic acid species on the basis of MS data, we decided to employ the highly sensitive DMB-HPLC method[11] because of the less abundant sialic acid-containing

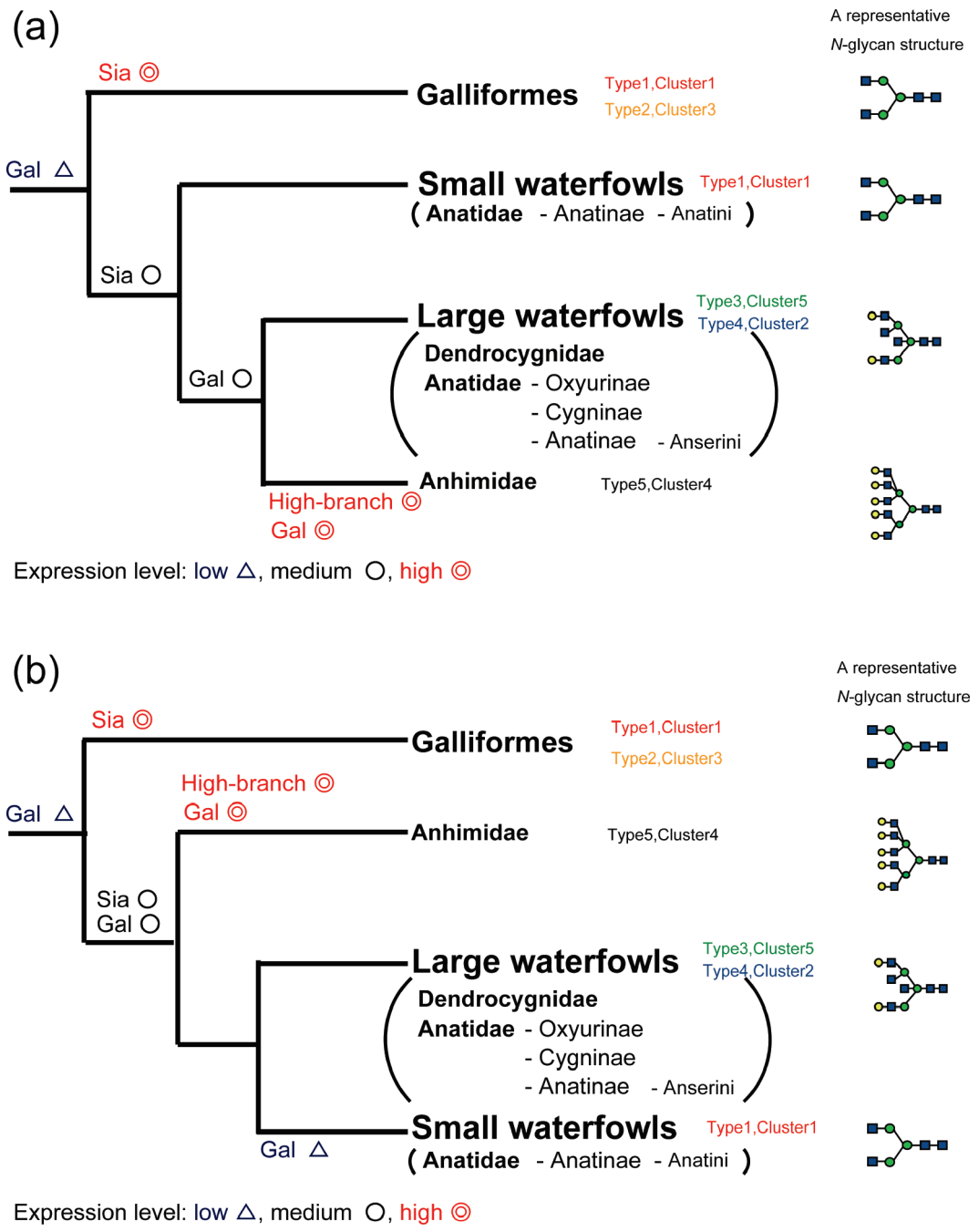


Figure 7. Evolutional dendrogram focused on Galliformes and Anseriformes based on the results of *N*-glycomics. Hypothesis a disagreed with the proposition based on the study of DNA−DNA hybridization by Sibley et al.[38] shown in Figure 6, while hypothesis b agreed.

glycans compared to neutral oligosaccharides. It was demonstrated that the DMB-sialic acids can be used for profiling sialic acid species even in cases of minor and trace amounts of sialylated components, though the information about their linkages was beyond the scope of this study.

In Anseriformes, non-Neu5Ac sialic acids (Neu5Gc and KDN) were particularly noticeable. This result did not correlate well with the species or the habitat but might be a result of a variation in diet. It is well-known that some fish contain KDN-modified glycans,[58] and some waterfowl prey on a small fish. In addition, it seems likely that Neu5Gc from a diet can accumulate for even the species that do not have Neu5Gc such as *Homo sapiens*.[59−64] The efficiency of that incorporation of Neu5Gc into cells under starvation conditions such as hypoxia is greater than that under normal conditions.[64] It was reported that a small amount of Neu5Gc is found in egg whites of some birds incapable of synthesizing Neu5Gc by themselves.[65,66] Anseriformes are waterfowl, while Galliformes are land birds, which tend to become hypoxic. KDN and Neu5Gc may be taken up by the body through predation of small fish having Neu5Gc or KDN, or alternative hypoxia conditions can induce uptake and generation of Neu5Gc or KDN.

We should note that this study is a preliminary trial in an attempt to gain insight into the relationship between the diversity

dx.doi.org/10.1021/bi101940x |*Biochemistry* 2011, 50, 4757–4774

of the avian glycome and its evolution. Therefore, we intended to cover widely avian species of Galloanserae to provide a discussion based on *N*-glycomics or quantitative sialic acid analysis, just between "groups" such as "large waterfowl and small waterfowl" or "Anseriformas and Galliformes", but not between "species" or on detailed structures. In fact, there were many of the same families and the same genera in the 88 samples tested herein, and we did not describe any detailed and critical evolutional dendrogram for each species at this preliminary screening of avio-*N*-glycomics. With a goal of in-depth discussion of the "inter-species" difference of glycomes, focused avian glycomics using multiple individuals will be designed and performed on the basis of a comprehensive approach using a high-throughput glycoblotting technique.

It is clear that large-scale glycomics used herein for avio-glycomics can be applied widely for the investigation of the significance of glycan complexity and diversity in the evolutional lineages of every mammal as well as birds. It should also be noted that sialic acid-focused reverse glycoblotting and glycoproteomics may provide further profound insights into the relationship between such inter-sialic acid alteration and genomic/proteomic or environmental background, and the mechanism of posttranslational modification in both *N*- and *O*-glycosylations.[67,68]

■ ASSOCIATED CONTENT

Supporting Information. Data from MALDI-TOF MS, DMB-HPLC, *N*-glycan structures, and statistical analysis based on glycotype. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Graduate School of Life Science, Hokkaido University, Kita-ku, N21, W11, Sapporo 001-0021, Japan. E-mail: maho@glyco.sci.hokudai.ac.jp or shin@glyco.sci.hokudai.ac.jp. Phone: +81 11 706 9043. Fax: +81 11 706 9042.

Funding Sources

This work was supported partly by a grant for "Development of Systems and Technology for Advanced Measurement and Analysis (SENTAN)" and "The Matching Program for Innovations in Future Drug Discovery and Medical Care" from the Japan Science and Technology Agency (JST) and the Ministry of Education, Culture, Science, and Technology of Japan.

■ ACKNOWLEDGMENT

We express special gratitude to the late Dr. M. Laskowski, Jr., for his generosity in providing invaluable egg white collections. We appreciate all the support and technical assistance for the preparation of egg white samples, the large-scale glycoblotting on SweetBlot, and MALDI-TOFMS analysis of Dr. H. Hinou, Dr. M. Hato, Ms. A. Yamauchi, Ms. M. Yokoyama, Dr. T. Yamashita, Ms. S. Kudo, Ms. M. Yamaguchi, Dr. J. Furukawa, Dr. Y. Shinohara, Dr. Y. Takegawa, Dr. M. Kurogochi, and Dr. N. Miura of Hokkaido University and Dr. Y. Miura and Dr. T. Nakahara of Ezose Sciences, Inc. We also are grateful to Prof. Y. Suzuki of the Department of Biomedical Sciences, Chubu University (Aichi, Japan), for stimulating discussions and suggestions and Dr. N. Fukushima of Science & Technology Systems, Inc., for the use of Spotfire DecisionSite and helpful discussion of the data analysis.

■ ABBREVIATIONS

DTT, dithiothreitol; IAA, iodoacetamide; MTT, 3-methyl-1-*p*-tolyltriazene; CHCA, α-cyano-4-hydroxycinnamic acid; DMB, 1,2-diamino-4,5-methylenedioxybenzene; BOA, *O*-benzylhydroxylamine hydrochloride; PHM, 1-propanesulfonic acid, 2-hydroxyl-3-myristamido; Hex, hexose; Man, mannose; Gal, galactose; HexNAc, *N*-acetylhexosamine; GlcNAc, *N*-acetylglucosamine; Neu5Ac, *N*-acetylneuraminic acid; Neu5Gc, *N*-glycolylneuraminic acid; KDN, 2-keto-3-deoxynonulosonic acid; HPLC, high-performance liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MS, mass spectrometry.

■ REFERENCES

(1) Varki, A. (1993) Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology 3*, 97–130.

(2) Gagneux, P., and Varki, A. (1999) Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology 9*, 747–755.

(3) Martinko, J. M., Vincek, V., Klein, D., and Klein, J. (1993) Primate ABO glycosyltransferases: Evidence for trans-species evolution. *Immunogenetics 37*, 274–278.

(4) Galili, U., Clark, M. R., Shohet, S. B., Buehler, J., and Macher, B. A. (1987) Evolutionary relationship between the natural anti-Gal antibody and the Galα1−3Gal epitope in primates. *Proc. Natl. Acad. Sci. U.S.A. 84*, 1369–1373.

(5) Galili, U., and Swanson, K. (1991) Gene sequences suggest inactivation of α-1,3-galactosyltransferase in catarrhines after the divergence of apes from monkeys. *Proc. Natl. Acad. Sci. U.S.A. 88*, 7401–7404.

(6) Galili, U. (1993) Interaction of the natural anti-Gal antibody with α-galactosyl epitopes: A major obstacle for xenotransplantation in humans. *Immunol. Today 14*, 480–482.

(7) Suzuki, N., Khoo, K. H., Chen, H. C., Johnson, J. R., and Lee, Y. C. (2001) Isolation and characterization of major glycoproteins of pigeon egg white: Ubiquitous presence of unique *N*-glycans containing Galα1−4Gal. *J. Biol. Chem. 276*, 23221–23229.

(8) Suzuki, N., Laskowski, M., Jr., and Lee, Y. C. (2004) Phylogenetic expression of Galα1−4Gal on avian glycoproteins: Glycan differentiation inscribed in the early history of modern birds. *Proc. Natl. Acad. Sci. U.S.A. 101*, 9023–9028.

(9) Suzuki, N., Laskowski, M., Jr., and Lee, Y. C. (2006) Tracing the history of Galα1−4Gal on glycoproteins in modern birds. *Biochim. Biophys. Acta 1760*, 538–546.

(10) Lehmann, F., Tiralongo, E., and Tiralongo, J. (2006) Sialic acid-specific lectins: Occurrence, specificity and function. *Cell. Mol. Life Sci. 63*, 1331–1354.

(11) Hara, S., Takemori, Y., Yamaguchi, M., Nakamura, M., and Ohkura, Y. (1987) Fluorometric high-performance liquid chromatography of *N*-acetyl- and *N*-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids. *Anal. Biochem. 164*, 138–145.

(12) Varki, A. (1997) Sialic acids as ligands in recognition phenomena. *FASEB J. 11*, 248–255.

(13) Angata, T., and Varki, A. (2002) Chemical diversity in the sialic acids and related α-keto acids: An evolutionary perspective. *Chem. Rev. 102*, 439–469.

(14) Schauer, R. (2009) Sialic acids as regulators of molecular and cellular interactions. *Curr. Opin. Struct. Biol. 19*, 507–514.

(15) Hansen, W. (1999) Avian Influenza. In *Field Manual of Wildlife Disease: General Field Procedures and Diseases of Birds*, pp 181−184, U.S. Geological Survey, Reston, VA.

(16) Cardona, C. J., Xing, Z., Sandrock, C. E., and Davis, C. E. (2009) Avian influenza in birds and mammals. *Comp. Immunol., Microbiol. Infect. Dis. 32*, 255–273.

(17) Ito, T., Suzuki, Y., Suzuki, T., Takada, A., Horimoto, T., Wells, K., Kida, H., Otsuki, K., Kiso, M., Ishida, H., and Kawaoka, Y. (2000) Recognition of *N*-glycolylneuraminic acid linked to galactose by the α2,3 linkage is associated with intestinal replication of influenza A virus in ducks. *J. Virol. 74*, 9300–9305.

(18) Olsen, B., Munster, V. J., Wallensten, A., Waldenström, J., Osterhaus, A. D., and Fouchier, R. A. (2006) Global patterns of influenza a virus in wild birds. *Science 312*, 384–388.

(19) Li, K. S., Guan, Y., Wang, J., Smith, G. J., Xu, K. M., Duan, L., Rahardjo, A. P., Puthavathana, P., Buranathai, C., Nguyen, T. D., Estoepangestie, A. T., Chaisingh, A., Auewarakul, P., Long, H. T., Hanh, N. T., Webby, R. J., Poon, L. L., Chen, H., Shortridge, K. F., Yuen, K. Y., Webster, R. G., and Peiris, J. S. (2004) Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature 430*, 209–213.

(20) Johnson, J. R., and Berggren, T. (1994) Pigeon and dove egg white protect mice against renal infection due to P fimbriated *Escherichia coli*. *Am. J. Med. Sci. 307*, 335–339.

(21) Pilobello, K. T., and Mahal, L. K. (2007) Deciphering the glycocode: The complexity and analytical challenge of glycomics. *Curr. Opin. Chem. Biol. 11*, 300–305.

(22) Nishimura, S.-I., Niikura, K., Kurogochi, M., Matsushita, T., Fumoto, M., Hinou, H., Kamitani, R., Nakagawa, H., Deguchi, K., Miura, N., Monde, K., and Kondo, H. (2004) High-throughput protein glycomics: Combined use of chemoselective glycoblotting and MALDI-TOF/TOF mass spectrometry. *Angew. Chem., Int. Ed. 44*, 91–96.

(23) Kita, Y., Miura, Y., Furukawa, J.-i., Nakano, M., Shinohara, Y., Ohno, M., Takimoto, A., and Nishimura, S.-I. (2007) Quantitative glycomics of human whole serum glycoproteins based on the standardized protocol for liberating N-glycans. *Mol. Cell. Proteomics 6*, 1437–1445.

(24) Miura, Y., Hato, M., Shinohara, Y., Kuramoto, H., Furukawa, J.-i., Kurogochi, M., Shimaoka, H., Tada, M., Nakanishi, K., Ozaki, M., Todo, S., and Nishimura, S.-I. (2008) BlotGlycoABC^{TM}, an integrated glycoblotting technique for rapid and large scale clinical glycomics. *Mol. Cell. Proteomics 7*, 370–377.

(25) Furukawa, J.-i., Shinohara, Y., Kuramoto, H., Miura, Y., Shimaoka, H., Kurogochi, M., Nakano, M., and Nishimura, S.-I. (2008) Comprehensive approach to structural and functional glycomics based on chemoselective glycoblotting and sequential tag conversion. *Anal. Chem. 80*, 1094–1101.

(26) Amano, M., Yamaguchi, M., Takegawa, Y., Yamashita, T., Terashima, M., Furukawa, J.-i., Miura, Y., Shinohara, Y., Iwasaki, N., Minami, A., and Nishimura, S.-I. (2010) Threshold in stage-specific embryonic glycotypes uncovered by a full portrait of dynamic *N*-glycan expression during cell differentiation. *Mol. Cell. Proteomics 9*, 523–537.

(27) Miura, Y., Kato, K., Takegawa, Y., Kurogochi, M., Furukawa, J.-i., Shinohara, Y., Nagahori, N., Amano, M., Hinou, H., and Nishimura, S.-I. (2010) Glycoblotting-assisted *O*-glycomics: Ammonium carbamate allows for highly efficient *O*-glycan release from glycoproteins. *Anal. Chem. 82*, 10021–10029.

(28) Kyselova, Z., Mechref, Y., Al Bataineh, M. M., Dobrolecki, L. E., Hickey, R. J., Vinson, J., Sweeney, C. J., and Novotny, M. V. (2007) Alterations in the serum glycome due to metastatic prostate cancer. *J. Proteome Res. 6*, 1822–1832.

(29) Hakomori, S.-I. (1964) A rapid permethylation of glycolipid, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide. *J. Biochem. 55*, 205–208.

(30) Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res. 131*, 209–217.

(31) Kang, P., Mechref, Y., Klouckova, I., and Novotny, M. V. (2005) Solid-phase permethylation of glycans for mass spectrometric analysis. *Rapid Commun. Mass Spectrom. 19*, 3421–3428.

(32) Dell, A., Lee, J. J., Pang, P.-C., Parry, S., Smith, M. S., Tissot, B., Morris, H. R., Panico, M. M., and Haslam, S. M. (2008) Glycomics and mass spectrometry. In *Glycoscience, Chemistry and Chemical Biology* (Fraser-Reid, B. O., Tatsuta, K., and Thiem, J., Eds.) Vol. 3, pp 2191−2217, Springer-Verlag, Berlin.

(33) Bones, J., Mittermayer, S., O'Donoghue, N., Guttman, A., and Rudd, P. M. (2010) Ultra performance liquid chromatographic profiling of serum *N*-glycans for fast and efficient identification of cancer associated alterations in glycosylation. *Anal. Chem. 82*, 10208–10215.

(34) Laskowski, M., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, W. J., Park, S. J., Parks, K., and Schatzley, B. L. (1987) Ovomucoid third domains from 100 avian species: Isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochemistry 26*, 202–221.

(35) Laskowski, M., Apostol, I., Ardelt, W., Cook, J., Giletto, A., Kelly, C. A., Lu, W. Y., Park, S. J., Qasim, M. A., and Whatley, H. E. (1990) Amino acid sequences of ovomucoid third domain from 25 additional species of birds. *J. Protein Chem. 9*, 715–725.

(36) Apostol, I., Giletto, A., Komiyama, T., Zhang, W., and Laskowski, M. (1993) Amino acid sequences of ovomucoid third domains from 27 additional species of birds. *J. Protein Chem. 12*, 419–433.

(37) Gruson, E. S., and Forster, R. A. (1976) *Checklist of the World's Birds: A Complete List of the Species, With Names, Authorities, and Areas of Distribution*, Quadrangle New York Times Book Co., New York.

(38) Sibley, C. G., and Monroe, B. L. (1990) *Distribution and Taxonomy of Birds of the World*, Yale University Press, New Haven, CT.

(39) Sibley, C. G., and Monroe, B. L. (1993) *A Supplement to Distribution and Taxonomy of Birds of the World*, Yale University Press, New Haven, CT.

(40) Sibley, C. G., and Ahlquist, J. E. (1990) *Phylogeny and Classification of Birds: A Study in Molecular Evolution*, Yale University Press, New Haven, CT.

(41) Miura, Y., Shinohara, Y., Furukawa, J., Nagahori, N., and Nishimura, S.-I. (2007) Rapid and simple solid-phase esterification of sialic acid residues for quantitative glycomics by mass spectrometry. *Chem.—Eur. J. 13*, 4797–4804.

(42) Hair, J. F. J., Anderson, R. E., Tatham, R. L., and Black, W. C. (1995) *Multivariate Data Analysis*, 4th ed., Prentice Hall, Englewood Cliffs, NJ.

(43) Mirkin, B. (1996) Mathematical Classification and Clustering. In *Nonconvex Optimization and Its Applications* (Pardalos, P., and Horst, R., Eds.) Vol. 11, Kluwer Academic Publishers, Dordrecht, The Netherlands.

(44) Sneath, P., and Sokal, R. R. (1973) *Numerical taxonomy*, 2nd ed., W. H. Freeman, San Francisco.

(45) Jolliffe, I. T. (1986) *Principal Component Analysis*, Springer Series in Statistics, Springer-Verlag, New York.

(46) Yasuda, Y., Takahashi, N., and Murachi, T. (1970) The Composition and Structure of Carbohydrate Moiety of Stem Bromelain. *Biochemistry 9*, 25–32.

(47) Nomoto, H., Endo, T., and Inoue, Y. (1982) Preparation and characterization of fragment glycoasparagines from ovalbumin glycopeptides: Reference compounds for structural and biochemical studies of the oligo-mannose and hybrid types of carbohydrate chains of glycoproteins. *Carbohydr. Res. 107*, 91–101.

(48) Harvey, D. J., Wing, D. R., Küster, B., and Wilson, I. B. (2000) Composition of *N*-linked carbohydrates from ovalbumin and co-purified glycoproteins. *J. Am. Soc. Mass Spectrom. 11*, 564–571.

(49) Lattova, E., Perreault, H., and Krokhin, O. (2004) Matrix-assisted laser desorption/ionization tandem mass spectrometry and post-source decay fragmentation study of phenylhydrazones of *N*-linked oligosaccharides from ovalbumin. *J. Am. Soc. Mass Spectrom. 15*, 725–735.

(50) Sumiyoshi, W., Nakakita, S., Hasehira, K., Miyanishi, N., Kubo, Y., Kita, T., and Hirabayashi, J. (2010) Comprehensive analysis of *N*-linked oligosaccharides from eggs of the family Phasianidae. *Biosci., Biotechnol., Biochem. 74*, 606–613.

(51) Takahashi, N., Khoo, K. H., Suzuki, N., Johnson, J. R., and Lee, Y. C. (2001) *N*-Glycan structures from the major glycoproteins of pigeon egg white: Predominance of terminal Galα(1−4)Gal. *J. Biol. Chem. 276*, 23230–23239.

(52) Suzuki, N., Su, T. H., Wu, S. W., Yamamoto, K., Khoo, K. H., and Lee, Y. C. (2009) Structural analysis of *N*-glycans from gull egg white glycoproteins and egg yolk IgG. *Glycobiology 19*, 693–706.

(53) Yamashita, K., Kamerling, J. P., and Kobata, A. (1983) Structural studies of the sugar chains of hen ovomucoid. Evidence indicating that they are formed mainly by the alternate biosynthetic pathway of asparagine-linked sugar chains. *J. Biol. Chem. 258*, 3099–3106.

(54) Tomiya, N., Awaya, J., Kurono, M., Endo, S., Arata, Y., and Takahashi, N. (1988) Analyses of N-linked oligosaccharides using a two-dimensional mapping technique. *Anal. Biochem. 171*, 73–90.

(55) Risley, J. M., and Van Etten, R. L. (1986) Structures of the carbohydrate moiety attached to one site in the first domain of turkey ovomucoid: Elucidation by ^{1}H NMR spectroscopy. *Carbohydr. Res. 147*, 21–29.

(56) Storrs, L. O. (1985) The Fossil Record of Birds. In *Avian Biology* (Donald, S. F., James, R. K., and Kenneth, C. P., Eds.) Vol. 8, pp 79−252, Academic Press, New York.

(57) Nakamura-Tsuruta, S., Kominami, J., Kamei, M., Koyama, Y., Suzuki, T., Isemura, M., and Hirabayashi, J. (2006) Comparative Analysis by Frontal Affinity Chromatography of Oligosaccharide Specificity of GlcNAc-Binding Lectins, *Griffonia simplicifolia* Lectin-II (GSL-II) and *Boletopsis leucomelas* Lectin (BLL). *J. Biochem. 140*, 285–291.

(58) Annuk, H., Nynes, S. O., Hirmo, S., Mikelsaar, M., and Wadström, T. (2001) Characterisation and differentiation of lactobacilli by lectin typing. *J. Med. Microbiol. 50*, 1069–1074.

(59) Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K. L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., and Yuri, T. (2008) A Phylogenomic Study of Birds Reveals Their Evolutionary History. *Science 320*, 1763–1768.

(60) Inoue, S., and Kitajima, K. (2006) KDN (deaminated neuraminic acid): Dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconjugate J. 23*, 277–290.

(61) Tangvoranuntakul, P., Gagneux, P., Diaz, S., Bardor, M., Varki, N., Varki, A., and Muchmore, E. (2003) Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc. Natl. Acad. Sci. U.S.A. 100*, 12045–12050.

(62) Martin, M. J., Muotri, A., and Varki, A. (2005) Human embryonic stem cells express an immunogenic nonhuman sialic acid. *Nat. Med. 11*, 228–232.

(63) Byres, E., Paton, A. W., Paton, J. C., Löfling, J. C., Smith, D. F., Wilce, M. C., Talbot, U. M., Chong, D. C., Yu, H., Huang, S., Chen, X., Varki, N. M., Varki, A., Rossjohn, J., and Beddoe, T. (2008) Incorporation of a non-human glycan mediates human susceptibility to a bacterial toxin. *Nature 456*, 648–652.

(64) Varki, A. (2009) Multiple changes in sialic acid biology during human evolution. *Glycoconjugate J. 26*, 231–245.

(65) Yin, J., Hashimoto, A., Izawa, M., Miyazaki, K., Chen, G. Y., Takematsu, H., Kozutsumi, Y., Suzuki, A., Furuhata, K., Cheng, F. L., Lin, C. H., Sato, C., Kitajima, K., and Kannagi, R. (2006) Hypoxic culture induces expression of sialin, a sialic acid transporter, and cancer-associated gangliosides containing non-human sialic acid on human cancer cells. *Cancer Res. 66*, 2937–2945.

(66) Schauer, R., Srinivasan, G. V., Coddeville, B., Zanetta, J. P., and Guérardel, Y. (2009) Low incidence of *N*-glycolylneuraminic acid in birds and reptiles and its absence in the platypus. *Carbohydr. Res. 344*, 1494–1500.

(67) Kurogochi, M., Amano, M., Fumoto, M., Takimoto, A., Kondo, H., and Nishimura, S.-I. (2007) Reverse glycoblotting allows rapid enrichment glycoproteomics of biopharmaceuticals and disease-related biomarkers. *Angew. Chem., Int. Ed. 46*, 8808–8813.

(68) Kurogochi, M., Matsushista, T., Amano, M., Furukawa, J., Shinohara, Y., Aoshima, M., and Nishimura, S.-I. (2010) Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay. *Mol. Cell. Proteomics 9*, 2354–2368.