BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 5 2014, pages 706–711
doi:10.1093/bioinformatics/btt573

Data and text mining

Advance Access publication October 17, 2013

Frequent glycan structure mining of influenza virus data revealed a sulfated glycan motif that increased viral infection

Tomomi Ichimiya[1], Shoko Nishihara[1], Sayaka Takase-Yoden[1], Hiroshi Kida[2,3] and Kiyoko Aoki-Kinoshita[1,*]

[1]Department of Bioinformatics, Faculty of Engineering, Soka University, 1-236 Tangi-machi, Hachioji, Tokyo, 192-8577, [2]Laboratory of Microbiology, Department of Disease Control, Graduate School of Veterinary Medicine, Hokkaido University, Kita 18, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0818 and [3]Research Center for Zoonosis Control, Hokkaido University, Kita 20, Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0020, Japan

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: It is well known influenza viruses recognize and bind terminal sialic acid (SA) on glycans that are found on the cell surface. In this work, we used a data mining technique to analyze the glycan array data of influenza viruses to find novel glycan structures other than SA that may be involved in viral infection.

Results: In addition to SA structures noted previously, we noted the sulfated structures in the mining results. For verification, we overexpressed the sulfotransferase that is involved in synthesizing these structures, and we performed a viral infection experiment to assess changes in infection in these cells. In our results, we found that there is a 70-fold increase in these cells compared with the control. Thus, we have found a novel pattern in glycan structures that may be involved in viral infection.

Availability and implementation: The Glycan Miner Tool is available from the RINGS resource at http://www.rings.t.soka.ac.jp.

Contact: kkiyoko@soka.ac.jp

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 25, 2013; revised on September 22, 2013; accepted on September 26, 2013

1 INTRODUCTION

Influenza viruses cause highly contagious respiratory disease in humans with potentially fatal outcomes, and millions of people are affected by influenza every year. In addition, many animal species, including birds, swine, horses, dogs, cats, whales and seals, can be infected by influenza viruses. Viruses of all known hemagglutinin (H1–H16) and neuraminidase (N1–N9) subtypes are maintained in aquatic birds, and therefore, they are considered the natural reservoir of influenza viruses. In humans, the known pandemics in 1918, 1977 and 2009; in 1957; and in 1968 were caused by the H1, H2 and H3 viruses, respectively (Wright *et al*., 2006). The known epidemics from 1977 until now were associated with viruses of the H1 and H3 subtypes. Although not yet pandemic, the first human infections with chicken H5N1 viruses were reported in 1997 in Hong Kong (Claas *et al*., 1998; Subbarao *et al*., 1998).

Influenza viruses bind to sialic acids (SAs) on the cell surface to initiate infection and replication. Human influenza viruses preferentially bind to *N*-acetylneuraminic acid (i.e. SA) attached to the penultimate galactose sugar by an α2,6 linkage (SA α 2-6 Gal), whereas avian viruses preferentially bind to SA with an α2,3 linkage (SA α 2-3 Gal) (Connor *et al*., 1994). The receptor specificity is shared by the majority of human and avian viruses, but it has been reported that the receptor-recognition of some avian viruses was affected by the third saccharide and more distant moieties of the (SA α 2-3 Gal)-terminal receptor (Gambaryan *et al*., 2003). In addition, sulfation of the saccharide core has been shown to produce no effect on binding of duck viruses, but chicken and human viruses isolated in 1997 in Hong Kong showed increased affinity to a sulfated sialyl-Lewis structure (Gambaryan *et al*., 2004, 2006). Thus, it has been shown that negatively charged groups, such as SA and sulfate groups, on glycans might function in binding to viruses.

In this work, we used a data mining method called α-closed frequent subtree mining (Hashimoto *et al*., 2008) to analyze glycan binding data as produced by glycan array experiments (Wang *et al*., 2013), which are available on the web by the Consortium for Functional Glycomics (CFG). After data mining of those glycan structures with high binding affinity, we noted that sulfated structures were found in the results in addition to other terminal SA and sialyl-Lewis structures, as illustrated previously in other works (Connor *et al*., 1994; Gambaryan *et al*., 2003, 2004, 2006, 2008).

In this work, however, we focused only on sulfation, and to verify the effects of sulfation as shown from these results, we overexpressed the sulfotransferase that is involved in synthesizing these structures. Then we performed a viral infection experiment to assess the change in the level of viral infection using the overexpressed cells. In our results, we found a surprisingly large increase in viral infection using our overexpressed cells. We were particularly surprised because we only overexpressed the sulfation and did not modify any sialyltransferases. From these results, we can claim that we have found a novel pattern in glycan structures that is involved in viral infection using our glycan mining tool.

*To whom correspondence should be addressed.

2 BACKGROUND

2.1 Glycan arrays

The study of glycan binding patterns has been much aided by the development of glycan arrays (Galustian *et al*., 2004) and the availability of such data on the web (Blixt *et al*., 2004; Raman *et al*., 2005). For example, the CFG provides mammalian glycan arrays to analyze the glycan binding specificity of lectins, antibodies, antisera, microorganisms and viruses, to name a few. Samples are fluorescently labeled to detect primary binding to the glycans on the array. To generate the array, a library of natural and synthetic mammalian glycans with amino linkers is printed onto *N*-hydroxysuccinimide-activated glass microscope slides (SCHOTT Nexterion), forming covalent amide linkages. The current CFG glycan arrays contain >600 glycan structures, as listed at http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh8.shtml. Among these structures, ~10% are sulfated structures. Previous studies using glycan arrays have confirmed that SA is an important component in influenza virus binding (Stevens *et al*., 2006). It has also been shown using a solid-phase binding assay that 6-sulfo sialyl Lewis X is recognized by H5, H6, H7 and H9 influenza viruses (Gambaryan *et al*., 2008). Therefore, the data from these technologies have been shown to be effective in analyzing glycan binding affinity.

2.2 Enzymes involved in glycan structure synthesis

The glycan structures expressed on the cell surface are synthesized in the endoplasmic reticulum and Golgi apparatus by a series of glycosyltransferases and sulfotransferases. Overexpression or RNA interference of a glycosyltransferase or a sulfotransferase is generally used to remodel glycan structures on the cell surface (Hirano *et al*., 2012; Sasaki *et al*., 2011). For example, overexpression of a sulfotransferase can increase a specific sulfation on a saccharide core (Hirano *et al*., 2012). Therefore, in this work, to perform our viral infection experiment, in preparing the cell culture, it was important to identify those structures in the extracted data for which glycosyltransferases and/or sulfotransferases were known.

2.3 Data mining for glycan analysis

The Resource for INformatics of Glycomes at Soka (RINGS) is a web resource providing a variety of data mining and algorithmic tools for glycomics data analysis (Akune *et al*., 2010). The data mining method we used in this work is implemented by the Glycan Miner Tool (Aoki-Kinoshita, 2013) in RINGS. This tool uses the α-closed frequent subtree algorithm (Hashimoto *et al*., 2008) for glycan structures. Details of this algorithm are provided in Methods.

For the data, we used the glycan array data of influenza viruses from the CFG (Blixt *et al*., 2004). In particular, we selected the data for H1 and H3 influenza viruses that were associated with the epidemics from 1977 until now. In addition, we selected the data for H5 influenza viruses because the first human infections with chicken H5 viruses were reported in 1997 in Hong Kong. These viruses were readily available, such that viral infection experiments could be performed using them.

2.4 Glycan structure notation

The graphical notation of glycan structures in this work is based on the symbols proposed by the CFG. In particular, each monosaccharide making up glycans is depicted by each given symbol. Thus, a glycan structure is represented as a tree structure, where nodes are monosaccharides (or sugars) and edges are glycosidic linkages (including an α or β anomer and two numbers indicating the linkage positions on each sugar). Table 1 lists the symbols and abbreviated names for the monosaccharides that appear in this manuscript.

Glycan structures may be represented using a variety of text formats. In this work, we used KEGG Chemical Function (KCF) format for the input data to the Glycan Miner Tool, which we obtained by converting from International Union of Pure and Applied Chemistry (IUPAC) format, which is the format used by the CFG. Details of these and other formats can be found in Aoki-Kinoshita, 2010. In brief, KCF format uses a graph notation for glycans by specifying monosaccharides as nodes and glycosidic linkages as edges. It also requires the specification of x- and y-coordinates for each node so that structures can be drawn clearly on a 2D plane. On the other hand, IUPAC format is a linear notation where branches are indicated by parentheses. RINGS provides utilities to convert glycan structures between a number of different formats, and the IUPAC to KCF converter was used in this work.

3 METHODS

3.1 Glycan array data

The glycan array data from the CFG were used as the data source for glycan binding affinity information of various influenza strains. We first selected all virus-binding array data, which resulted in 191 datasets. Among these, those virus strains for which virus experiments were feasible were selected, resulting in 15 wild-type strains as listed in Table 2. We then selected those glycans that demonstrated high binding affinity in each dataset (provided in Supplementary Information). The smallest number of selected glycans was 8 (human H1N1; A/South Carolina/1/1918) and the largest was 40 [Equine H3N8; A/equine/MA/2003 (H3N8)]. The human H1N1 virus strains bound to 15 glycans on average. Table 2 also lists the number of glycans selected from each dataset. In general, the top binding glycans were selected, except for those whose structures were not given, such as 'Fibrinogen' and 'Transferrin.' The glycan structure data provided by the CFG are provided in IUPAC format, so it was necessary to convert them to KCF format such that they could be input into the Glycan Miner Tool. Once this was done using the IUPAC to KCF utility in RINGS, and the structures were confirmed to be accurately converted, the data were then ready to be analyzed.

3.2 Glycan Miner Tool

The Glycan Miner Tool implements the α-closed frequent subtree algorithm (Hashimoto *et al*., 2008) for glycan structures. In brief, α-closed frequent subtrees are defined as those substructures within the input dataset that frequently appear (more than a specified minimum threshold, called the minimum support, or *minsup*). Those subtrees that appear in at least *minsup* number of input glycan structures are considered to be *frequent subtrees*. Moreover, to avoid the retrieval of overly similar subtrees, the parameter α was defined to specify the amount of overlap allowed among the retrieved subtrees. Thus, the following condition must be met for a subtree T to be included in the results:

$$\text{support}(P) < \max(a * \text{support}(T), \textit{minsup},$$

Table 1. The monosaccharides and their symbols used in this work

Monosaccharide name	Abbreviated name	Symbol
Mannose	Man	
Glucose	Glc	
Galactose	Gal	
N-acetylglucosamine	GlcNAc	
N-acetylgalactosamine	GalNAc	
Sialic acid (N-acetylneuraminic acid)	SA	

Table 2. Glycan Miner Tool parameters used in performing α-frequent subtree analysis of influenza data

Virus strain	Number of glycans	α	*minsup*	Number of subtrees retrieved
A/New York/221/2003 (H1N1)	17	0.8	3	7
A/Kawasaki/173/2001 (H1N1)	26	0.8	5	5
A/Texas/36/1991(H1N1)	14	0.7	3	9
A/South Carolina/1/1918 (H1N1)	8	0.8	3	2
A/New York/1/1918 (H1N1)	11	0.8	3	9
A/Duck/Alberta/35/1976 (H1N1)	12	0.7	3	4
A/Moscow/10/1999 (H3N2)	11	0.8	3	7
A/Duck/Ukraine/1/1963 (H3N8)	14	0.8	3	5
A/canine/Florida/2004 (H3N8)	21	0.8	5	7
A/canine/Florida/2004 (H3N8)	39	0.8	7	10
A/equine/MA/2003 (H3N8)	40	0.7	7	7
A/Indonesia/05/2005 (H5N1)	28	0.75	5	9
A/Vietnam/1203/2004 (H5N1)	27	0.75	5	8
A/Hong Kong/213/2003 (H5N1)	23	0.8	5	6
A/Duck/Singapore/Q-F119-3/19 (H5N3)	10	0.7	3	5

where P is a supertree of T and support(X) refers to the number of times subtree X appears in the input dataset.

The Glycan Miner Tool returns the α-closed frequent subtrees in order of P-value, which is computed based on the input data. This P-value is obtained by computing the support of randomly generated structures with similar topology to the target subtree. Then the probability of obtaining the target subtree among the randomly generated subtrees is computed.

3.3 Cell line and culture

Madin Darby Canine Kidney (MDCK) cells were acquired from the Japanese Collection of Research Bioresources Cell Bank (JCRB9029). MDCK cells were cultured in Eagle's minimal essential medium containing 10% fetal bovine serum and antibiotic at 37°C under 5% CO_2 atmospheric conditions.

3.4 Establishment of stable transformant expressing *hGlcNAc6ST-1*

The vector construction for the expression of *human GlcNAc-6-sulfotransferase-1*(*hGlcNAc6ST-1*) (Uchimura and Rosen, 2006) was prepared according to Invitrogen GATEWAY systems by two-step polymerase chain reaction (PCR). The full-length open reading frame of *hGlcNAc6ST-1* was amplified from EST clone 4562846 (supplied by Open Biosystems), by using the forward primer, 5′-AAAAAGCAGGCTTGGCCGCCA CCATGAAGGTGTTCCGTAGGAAGG-3′ and reverse primer, 5′-AGAAAGCTGGGTTTAGAGACGGGGCTTCCG-3′. For addition of the adaptor sequence necessary to GATEWAY system, the open reading frame fragment was amplified again by using the forward primer, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3′ and reverse primer, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGT-3′. The amplified fragment was recombined with the pCXN2 expression vector modified for GATEWAY system to yield pCXN2-*hGlcNAc6ST-1*.

MDCK cells were seeded at 4×10^5 cells/well in a 12-well plate 24 h before transfection. The 3 μg of pCXN2-*hGlcNAc6ST-1* or pCXN2 (for mock transfectant) was mixed with 9 μl of Lipofectamine2000 (Life Technologies, CA, USA) in 200 μl of Opti-MEM medium (Life Technologies) and transfected to MDCK cells. From 1 day after transfection, cells were cultured in growth medium containing 0.4 μg/ml G418 for selection of the transformants. The transformant cells were passed every 3–4 days for 20 days, and the increase of *hGlcNAc6ST-1* gene expression was checked. Flow cytometry analysis was performed by using MECA-79 monoclonal antibody (sc-19602, Santa Cruz, TX, USA), which is known to recognize 6-sulfo GlcNAc structures (Yeh *et al*., 2001).

3.5 Viral infection to cells and measurement of the amount of viral gene in culture supernatant

A human influenza virus, A/Hokkaido/11/2002 (H1N1), was inoculated into cells at a multiplicity of infection of 0.01 in Eagle's Minimal Essential Medium containing antibiotic (minimal medium). After incubation at 37°C for 1 h, the cells were washed twice in phosphate-buffered saline, and cultured in minimal medium with presence of 1 μg/ml of acetyltrypsin (SIGMA, MO, USA). At 48 and 72 h post-infection, culture supernatants were harvested. Viral RNA was extracted from 140 μl of the supernatant using QIAamp Viral RNA Mini Kit (QIAGEN, Venlo, The Netherlands) according to the manufacture's protocol. These RNA were eluted in a final volume of 70 μl of RNase-free water contains 0.04% sodium azide, and a portion was subjected to real-time reverse transcription-polymerase chain reaction (RT-PCR) using LightCycler 480 Instrument (Roche, Basel, Switzerland). Primers used to quantitate the amount of matrix (M) gene of virus were forward: 5′-CCMAGGTCG AAACGTAYGTTCTCTCTCTATC-3′, reverse: 5′-TGACAGRATYG GTCTTGTCTTTAGCCAYTCCA-3′, probe: 5′-ATYTCGGCTTTGA GGGGGCCTG-3′. The probe was labeled at the 5′-end with the reporter dye, 3FAM, and at the 3′-end with the quencher dye, MGB. Reactions were performed using QuantiTect Probe RT-PCR Kit (QIAGEN) as follows: reverse transcription, 30 min at 50°C; amplification, 50 cycles of 15 s at 94°C and 75 s at 54°C.

4 RESULTS

4.1 Glycan Miner Tool analysis

Table 2 shows the parameters used in analyzing the glycan array data. For each dataset, a value for α and minimum support *minsup* was selected such that the number of subtrees extracted was within the range of 5–10. In general, α ranged from 0.7 to 0.9 and *minsup* ranged from three to seven.

Screenshots of all of the analysis results are provided in Supplementary information. Here, we describe the results that were used as the basis of selecting sulfotransferase in the viral infection experiment; in particular, we focused on the human H1N1 virus strains A/Texas/36/1991 and A/New York/1/1918. Figures 1 and 2 are snapshots of these results, respectively.

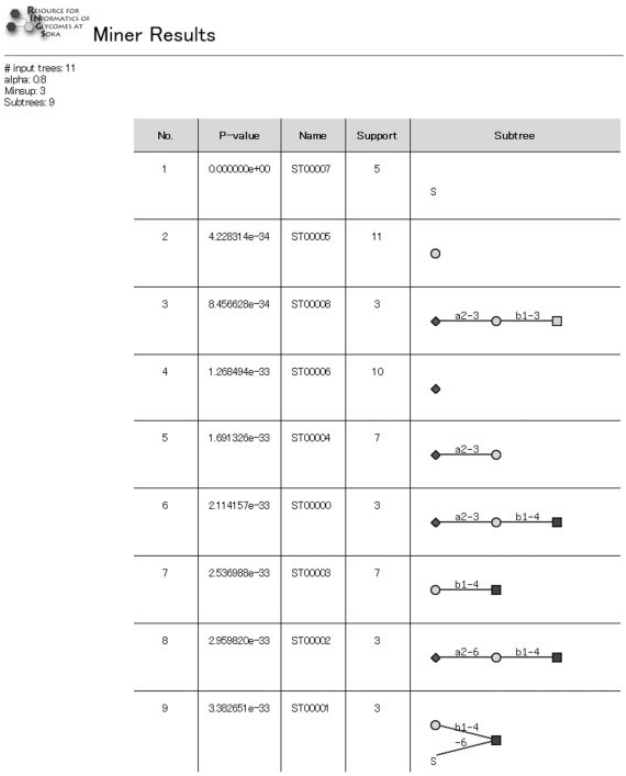

Fig. 1. Screenshot of Glycan Miner analysis results using human H1N1 (New York, 1918)

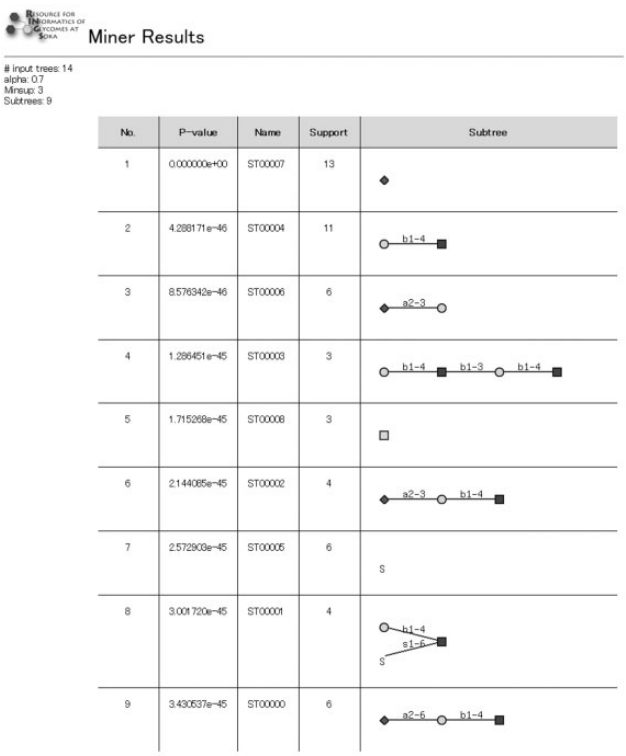

Fig. 2. Screenshot of Glycan Miner analysis results using human H1N1 (Texas)

First, it was interesting to note that in addition to the expected sialylated structures, sulfate itself was retrieved in the results in addition to the same sulfated trimer Gal β1,4 ((S03-) 1,6) GlcNAc in both A/Texas/36/1991 and A/New York/1/1918. Although SA is not included, there is the possibility that the galactose residue is sialylated. However, we were interested in the sulfated structures in the results. Disregarding monomers, Table 3 summarizes the structures that we selected as candidate structures from the mining results.

4.2 Overexpression of sulfotransferase in MDCK cells

From the Glycan Miner Tool results, Gal β 1-4 [(S03-) 1-6] GlcNAc was found in two different H1N1 strains among six H1N1 strains. Five kinds of GlcNAc-6-sulfotransferases (GlcNAc6STs), namely GlcNAc6ST-1 ~ GlcNAc6ST-5, have been reported to have the ability to synthesize Gal β 1-4 [(S03-) 1-6] GlcNAc (Uchimura and Rosen, 2006), and GlcNAc6ST-1 shows ubiquitous tissue expression pattern including lung. Therefore, we performed overexpression of GlcNAc6ST-1 in MDCK cells and analysis of viral infection using a H1N1 strain. We confirmed the upregulation of sulfated GlcNAc structures using flow cytometry and found that they were increased by 257%. The mean value of *hGlcNAcT6ST-1*-overexpression cells was 32.72 against empty-vector-transfected cells having a mean value of 12.71.

Normal MDCK cells, empty-vector-transfected cells, and *hGlcNAc6ST-1*-overexpression cells were infected with H1N1 virus (Fig. 3 and Table 4). At 48 and 72 h post-infection, the amount of viral *M* gene that was contained in culture supernatant was measured by the method of real-time RT-PCR. At 48 h post-infection, the amount of viral M gene in *hGlcNAc6ST-1*-over-expression cells was approximately 70-fold higher than that in normal MDCK cells. At 72 h post-infection, the amount of viral M gene in *hGlcNAc6ST-1*-overexpression cells was ~13-fold higher than that in normal MDCK cells. The amount of viral *M* gene in empty-vector-transfected cells was the same as in normal MDCK cells. These results indicate that Gal β 1-4 [(S03-) 1-6] GlcNAc structure in cells greatly enhances the infection of H1N1 virus.

5 DISCUSSION

In this work, we show that our tool implementing the α-closed frequent subtree method was able to extract a significant pattern from glycan array data, which we proved using a viral infection experiment on cells with modified glycans on the cell surface. This is the first work to our knowledge where experts in bioinformatics, glycobiology and virology collaborated and produced significant and promising biologically-relevant results. The data mining technique involved subtle adjustments of the parameters to obtain results that could be used in wet-lab experiments. The glycobiologists provided expertise in modifying the appropriate glycan structures on the cell surface, and the virologists provided the appropriate cells and viruses by which the overexpressed cells could be experimented. As a result, we were able to illustrate the validity of data mining results by confirming them using wet-lab experiments. More importantly, we show that we have indeed found a novel pattern other than SA in glycan structures that

Table 3. A summary of the Glycan Miner Tool analysis results

Virus strain	Glycans selected from the results In order of *P*-value (left < right)			
A/New York/221/2003 (H1N1)	(**S03-**) 1-3 Gal	Gal β 1-4 ((**S03-**) 1-6) Glc	(**S03-**) 1-3 Gal β 1-4 Glc	(**S03-**) 1-6 Gal β 1-4 GlcNAc
A/Kawasaki/173/2001 (H1N1)	SA α 2-6 (G2GN2M3GN2: N-link, α 1-6 Man side)	SA α 2-6 (G2GN2M3GN2: N-link, α 1-3 Man side)		
A/Texas/36/1991(H1N1)	Gal β 1-4 GlcNAc β 1-3 Gal β 1-4 GlcNAc	SA α 2-3 Gal β 1-4 GlcNAc	Gal β 1-4 ((**S03-**) 1-6) GlcNAc	SA α 2-6 Gal β 1-4 GlcNAc
A/South Carolina/1/1918 (H1N1)	SA α 2-6 Gal β 1-4 GlcNAc β 1-3 Gal β 1-4 GlcNAc	SA α 2-6 Gal β 1-4 GlcNAc		
A/New York/1/1918 (H1N1)	SA α 2-3 Gal β 1-3 GalNAc	SA α 2-3 Gal β 1-4 GlcNAc	SA α 2-6 Gal β 1-4 GlcNAc	Gal β 1-4 ((**S03-**) 1-6) GlcNAc
A/Duck/Alberta/35/1976 (H1N1)	SA α 2-3 Gal β 1-3 GalNAc	SA α 2-3 Gal β 1-3 GlcNAc	SA α 2-3 Gal β 1-4 GlcNAc	
A/Moscow/10/1999 (H3N2)	Gal β 1-4 (Fuc α 1-3) GlcNAc:Le^x	SA α 2-6 Gal β 1-4 GlcNAc	Gal β 1-4 GlcNAc β 1-3 Gal β 1-4 GlcNAc	SA α 2-3 Gal
A/Duck/Ukraine/1/1963 (H3N8)	SA α 2-3 Gal β 1-4 GlcNAc	SA α 2-3 Gal β 1-3 GalNAc	SA α 2-3 Gal β 1-3 GlcNAc	
A/canine/Florida/2004 (H3N8)	SA α 2-3 Gal β 1-4 (Fuc α 1-3) GlcNAc:sLe^x	SA α 2-3 Gal β 1-4 GlcNAc		
A/canine/Florida/2004 (H3N8)	Gal β 1-4 (Fuc α 1-3) GlcNAc:Le^x	SA α 2-3 Gal β 1-4 (Fuc α 1-3) GlcNAc:sLe^x	SA α 2-3 Gal β 1-4 GlcNAc	
A/equine/MA/2003 (H3N8)	SA α 2-3 Gal β 1-4 GlcNAc	Gal β 1-4 (Fuc α 1-3) GlcNAc:Le^x		
A/Indonesia/05/2005 (H5N1)	SA α 2-3 Gal β 1-4 (Fuc α 1-3) GlcNAc:sLe^x	SA α 2-3 Gal β 1-4 GlcNAc	SA α 2-3 Gal β 1-4 GlcNAc β 1-3 Gal	
A/Vietnam/1203/2004 (H5N1)	SA α 2-3 Gal β 1-4 (Fuc α 1-3) GlcNAc:sLe^x	SA α 2-3 Gal β 1-4 GlcNAc	SA α 2-3 Gal β 1-4 GlcNAc β 1-3 Gal	
A/Hong Kong/213/2003 (H5N1)	SA α 2-3 Gal β 1-4 GlcNAc	**S03**$^-$	SA α 2-3 Gal β 1-3 GalNAc	
A/Duck/Singapore/Q-F119-3/19 (H5N3)	SA α 2-3 Gal β 1-4 ((**S03-**) 1-6) GlcNAc	SA α 2-3 Gal β 1-4 GlcNAc		

Note: Sulfation is indicated in bold. The underlined structures are those that were selected for further verification in this study.

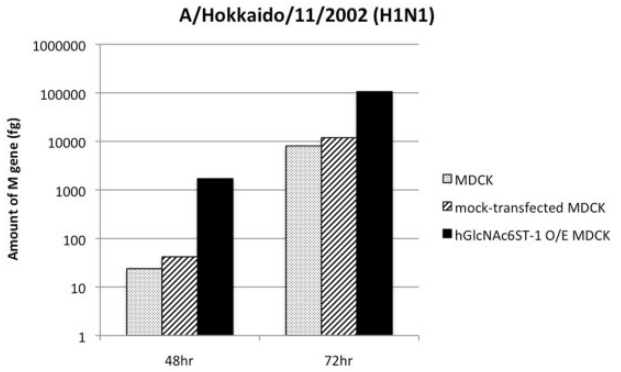

Fig. 3. Plot of viral production in the three types of cells used in this study. The gray bar is the control, the striped bar is the negative control and the black bar is the MDCK cell with overexpressed sulfotransferase. The actual values are listed in Table 4

Table 4. Viral production levels in the three types of cells analyzed in this work

Cells	48 h	72 h
MDCK	1.0	1.0
Mock-transfected MDCK	1.8	1.5
hGlcNAc6ST-1 O/E MDCK	71.5	13.1

may be involved in viral infection, which we reproduced in an independent experimental setting. We make note here that the increased viral infection may still involve SAs, but that the change in infection was solely affected by the increased sulfation. We also note that some interesting results have also been obtained regarding H7N9 influenza A viruses from humans (Watanabe *et al*., 2013), which we plan on analyzing in the near future for comparison with the results of this work.

For future work, we plan on increasing the functionality of the Glycan Miner Tool by adding additional information in the results to indicate the possible classes of glycans (e.g. *N*- or *O*-linked, glycosaminoglycans, etc) in which an extracted substructure may be found. The original glycan(s) in which the substructure was found should also be displayed to assist the user in further analysis of the dataset. It may also be useful to indicate the net negative charge of a resulting structure, as such information may be relevant to binding.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Shota Yamada who performed the parameterization and analysis of the CFG data using the Glycan Miner Tool.

Funding: RINGS was supported by Grant-In-Aid for Young Scientists (A), KAKENHI (20016025), and this work was supported by Challenging Exploratory Research (21650068), the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Conflicts of Interest: none declared.

REFERENCES

Akune,Y. *et al.* (2010) The RINGS resource for glycome informatics analysis and data mining on the Web. *OMICS*, **14**, 475–486.

Aoki-Kinoshita,K.F. (2010) Chapman & Hall/CRC Mathematical and Computational Biology Series. In: *Glycome Informatics: Methods and Applications*. CRC Press, Boca Raton.

Aoki-Kinoshita,K.F. (2013) Mining frequent subtrees in glycan data using the rings glycan miner tool. *Methods Mol. Biol.*, **939**, 87–95.

Blixt,O. *et al.* (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl Acad. Sci. USA*, **101**, 17033–17038.

Claas,E.C. *et al.* (1998) Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*, **351**, 472–477.

Connor,R.J. *et al.* (1994) Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology*, **205**, 17–23.

Galustian,C. *et al.* (2004) High and low affinity carbohydrate ligands revealed for murine SIGN-R1 by carbohydrate array and cell binding approaches, and differing specificities for SIGN-R3 and langerin. *Int. Immunol.*, **16**, 853–866.

Gambaryan,A. *et al.* (2006) Evolution of the receptor binding phenotype of influenza A (H5) viruses. *Virology*, **344**, 432–438.

Gambaryan,A.S. *et al.* (2003) Differences between influenza virus receptors on target cells of duck and chicken and receptor specificity of the 1997 H5N1 chicken and human influenza viruses from Hong Kong. *Avian Dis.*, **47**, 1154–1160.

Gambaryan,A.S. *et al.* (2004) H5N1 chicken influenza viruses display a high binding affinity for Neu5Acalpha2-3Galbeta1-4(6-HSO3)GlcNAc-containing receptors. *Virology*, **326**, 310–316.

Gambaryan,A.S. *et al.* (2008) 6-sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry. *Virol. J.*, **5**, 85.

Hashimoto,K. *et al.* (2008) Mining significant tree patterns in carbohydrate sugar chains. *Bioinformatics*, **24**, i167–i173.

Hirano,K. *et al.* (2012) 3-*O*-sulfated heparan sulfate recognized by the antibody HS4C3 contribute to the differentiation of mouse embryonic stem cells via Fas signaling. *PLoS One*, **7**, e43440.

Raman,R. *et al.* (2005) Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat. Methods*, **2**, 817–824.

Sasaki,N. *et al.* (2011) LacdiNAc (GalNAcβ1-4GlcNAc) contributes to self-renewal of mouse embryonic stem cells by regulating leukemia inhibitory factor/STAT3 signaling. *Stem Cells*, **29**, 641–650.

Stevens,J. *et al.* (2006) Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.*, **355**, 1143–1155.

Subbarao,K. *et al.* (1998) Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science*, **279**, 393–396.

Uchimura,K. and Rosen,S.D. (2006) Sulfated L-selectin ligands as a therapeutic target in chronic inflammation. *Trends Immunol.*, **27**, 559–565.

Wang,Z. *et al.* (2013) A general strategy for the chemoenzymatic synthesis of asymmetrically branched N-glycans. *Science*, **341**, 379–383.

Watanabe,T. *et al.* (2013) Characterization of H7N9 influenza A viruses isolated from humans. *Nature*, **501**, 551–555.

Wright,P.F. *et al.* (2006) In: Knipe,D.M. and Howley,P.M. (eds). *Fields Virology*. Lippincott Williams & Wilkins, Philadelphia, PA, pp. 1691–1740.

Yeh,J.C. *et al.* (2001) Novel sulfated lymphocyte homing receptors and their control by a Core1 extension beta 1,3-N-acetylglucosaminyltransferase. *Cell*, **105**, 957–969.