

# UC San Diego

## UC San Diego Previously Published Works

### Title

Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses

### Permalink

<https://escholarship.org/uc/item/2s2628jr>

### Journal

Chem, 7(12)

### ISSN

1925-6981

### Authors

Lucas, Taryn M  
Gupta, Chitrak  
Altman, Meghan O  
et al.

### Publication Date

2021-12-01

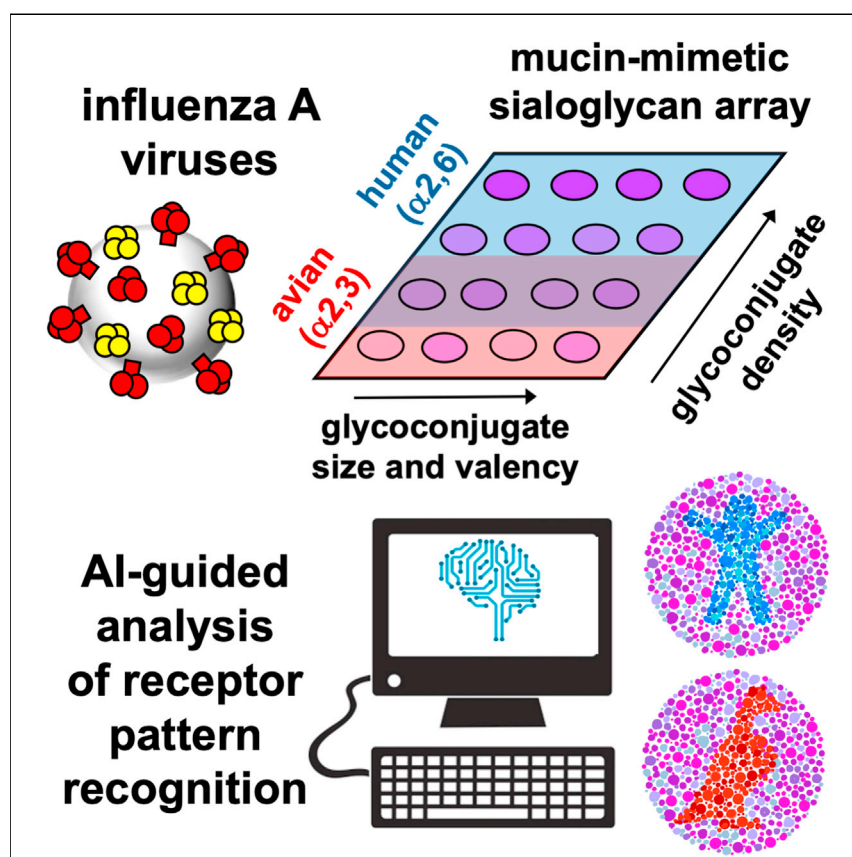
### DOI

10.1016/j.chempr.2021.09.015

Peer reviewed

## Article

## Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses



Glycan array comprising synthetic mucin-like displays of sialoglycans enables profiling of influenza A virus interactions based on receptor structure and presentation. Machine learning can detect and predict changes in receptor pattern binding behavior of the viruses.

Taryn M. Lucas, Chitrak Gupta, Meghan O. Altman, ..., Pascal Gagneux, Abhishek Singharoy, Kamil Godula

kgodula@ucsd.edu

#### Highlights

Mucin-mimetic array profiles viral binding based on glycan receptor presentation

SVM learning can detect and predict viral binding patterns

Organization of human sialoglycans in mucin-type displays disfavors viral binding

H1N1 better utilizes mucin-like receptor displays when produced in mammalian cells

Article

# Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses

Taryn M. Lucas,<sup>1</sup> Chittrak Gupta,<sup>2,3</sup> Meghan O. Altman,<sup>1,6</sup> Emi Sanchez,<sup>1</sup> Matthew R. Naticchia,<sup>1</sup> Pascal Gagneux,<sup>4,5</sup> Abhishek Singharoy,<sup>2,3</sup> and Kamil Godula<sup>1,5,7,\*</sup>

## SUMMARY

**Influenza A viruses (IAVs) exploit host glycans in airway mucosa for entry and infection. Detection of changes in IAV glycan-binding phenotype can provide early indication of transmissibility and infection potential. While zoonotic viruses are monitored for mutations, the influence of host glycan presentation on viral specificity remains obscured. Here, we describe an array platform that uses synthetic mimetics of mucin glycoproteins to model how receptor presentation and density in the mucinous glycocalyx may impact IAV recognition. H1N1 and H3N2 binding in arrays of  $\alpha$ 2,3- and  $\alpha$ 2,6-sialyllactose receptors confirmed their known sialic-acid-binding specificities and revealed their different sensitivities to receptor presentation. Further, the transition of H1N1 from avian to mammalian cell culture improved the ability of the virus to recognize mucin-like displays of  $\alpha$ 2,6-sialic acid receptors. Support vector machine (SVM) learning efficiently characterized this shift in binding preference and may prove useful to study viral evolution to a new host.**

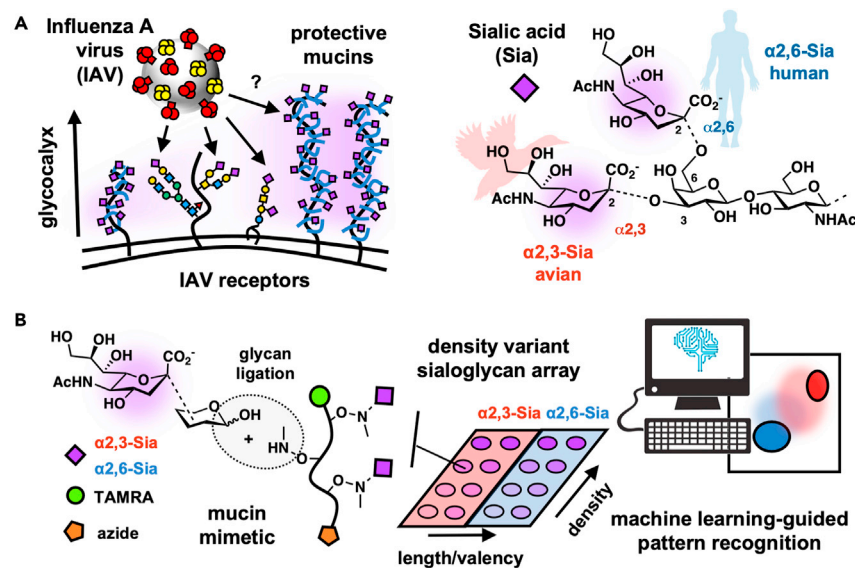
## INTRODUCTION

The periodic emergence of new respiratory viruses capable of spreading across the human population continues to exact a significant toll on human life and the global economy. The novel coronavirus, SARS-CoV-2, which is responsible for the ongoing COVID-19 pandemic,<sup>1</sup> provides a stark example of the risks of zoonotic virus adaptation to our society. Other animal pathogens, such as avian influenza A viruses (IAVs), continuously pose a threat of crossing to human hosts and require close monitoring.<sup>2</sup> Many respiratory viruses, including IAVs, utilize specific glycan receptors on airway epithelial cells to initiate entry and replication.<sup>3</sup> Characterization of the glycan-binding phenotype of IAVs may provide an early indicator of increased infection potential.<sup>4–6</sup>

IAVs carry two types of glycoproteins in their viral coat with specificity for terminal sialic acid modifications on cell surface glycans—the receptor-binding hemagglutinins (HAs) and the receptor-destroying neuraminidases (NAs).<sup>7–9</sup> The configuration of the sialic acid glycosidic bond linkage to the underlying glycans in glycoproteins and glycolipids plays a prominent role in defining IAV host specificity (Figure 1A). According to the prevailing paradigm,<sup>4,10</sup> avian viruses preferentially recognize  $\alpha$ 2,3-linked sialic acids abundant in the gastrointestinal tract of birds, while human IAVs have affinity for  $\alpha$ 2,6-sialosides expressed on lung epithelial cells in our upper airways. A switch in HA specificity from  $\alpha$ 2,3- to  $\alpha$ 2,6-linked sialic acids is associated with increased infection and transmission in humans.<sup>4,11,12</sup>

## The bigger picture

Influenza A exploits sialic acid glycans on host cells to gain entry during infection. Mucosal tissues are protected against viral assault by a physical barrier of soluble and cell surface mucin glycoproteins, which are, paradoxically, also heavily decorated with sialic acids. To explore protective functions of mucins, we examined binding of H1N1 and H3N2 in mucin-mimetic arrays modeling the structure and molecular crowding of the mucosal glycocalyx. Machine-learning analysis showed that displaying human-like sialic acid receptors on mucin-mimetic scaffolds disfavored viral binding. Adhesion was further decreased by crowding on the array, providing insight into the possible role of membrane mucins in limiting viral association. H1N1 propagated in mammalian cells showed greater binding to denser receptor displays than when grown in eggs. This platform may thus find utility as a surveillance tool for assessing changes in receptor-binding phenotype of circulating viruses.



**Figure 1. Machine-learning-enabled glycomimetic array platform for assessing receptor pattern recognition by influenza A viruses (IAVs)**

(A) IAVs begin their infection cycle by binding to sialylated host glycans, but these receptors are also present on mucins that have a proposed protective function. Avian and human IAVs show distinct preferences for the binding of  $\alpha 2,3$ - and  $\alpha 2,6$ -sialoglycan receptors.

(B) Glycopolymers, which mimic the architecture and composition of mucins, were used to build models of the mucinous glycoconjugate on microarrays. A SVM learning algorithm enabled analysis of viral binding response to changing receptor patterns in the synthetic glycoconjugate arrays.

Glycomics screens<sup>13,14</sup> and cell-based studies using glycosylation mutants<sup>15,16</sup> have revealed that, in addition to a particular sialic acid linkage configuration, IAVs can also discriminate between distinct glycan classes and glycoconjugate types (i.e., N- and O-glycosylated proteins and glycolipids). Spatial combinations of these sialylated glycoconjugates give rise to three-dimensional, hierarchically organized receptor patterns in the host cell glycoconjugate that determine the specificity and avidity of IAV binding. Glycan arrays, which present ensembles of chemically defined glycans printed and immobilized on glass substrates, are routinely used to analyze viral HA-receptor specificity.<sup>17–21</sup> However, a recent cross-comparison between glycan composition of *ex vivo* human lung and bronchus tissues with glycan array binding data pointed to a limited ability of the arrays to predict infection events.<sup>14</sup> This indicates that the current platforms may not accurately recapitulate the receptor presentation in the glycoconjugate environment as encountered by viruses at the mucosal epithelium.

The mucosal epithelial cell glycoconjugate is dominated by membrane-tethered mucins (MUCs), which are large, heavily glycosylated proteins projecting tens to hundreds of nanometers above the cell surface (Figure 1A).<sup>22,23</sup> Mucins carry primarily, but not exclusively, O-glycans linked to tracks of serine and threonine residues within the core protein. As much as 80% of mucin mass derives from glycans, giving these glycoproteins an extended semi-flexible bottlebrush form.<sup>24</sup> The O-glycans in mucins are frequently terminated with sialic acids that can serve as IAV receptors<sup>25</sup>; however, epithelial mucins are believed to primarily provide protection against infection. Mucins can serve as decoys, which shed from the cell surface upon virus binding,<sup>22</sup> or assemble into dense extended glycoprotein brushes that restrict virus access to apical membrane receptors and interfere with internalization.<sup>26,27</sup> Due to

<sup>1</sup>Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, USA

<sup>2</sup>School of Molecular Sciences, Arizona State University, Tempe, AZ, USA

<sup>3</sup>Biodesign Institute, Arizona State University, Tempe, AZ 85281, USA

<sup>4</sup>Department of Pathology, Division of Comparative Pathology and Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, USA

<sup>5</sup>Glycobiology Research and Training Center, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

<sup>6</sup>Present address: Abbott Labs, 2251 Faraday Ave, Carlsbad, CA 92008, USA

<sup>7</sup>Lead contact

\*Correspondence: [kgodula@ucsd.edu](mailto:kgodula@ucsd.edu)  
<https://doi.org/10.1016/j.chempr.2021.09.015>

the prominence of these extended glycoproteins within the glycocalyx and their extensive modification with sialic acids, mucins are most likely the first, and likely non-productive, site of virus attachment in its infection cycle.<sup>28</sup> Interestingly, the IAV subtype, H1N1, was found to colocalize with some (i.e., MUC1), but not other (i.e., MUC13 and MUC16), mucins on the surfaces of A549 lung epithelial cells,<sup>29</sup> revealing a preference of the virus for distinct mucin family members within the same cell and produced by a shared glycosylation machinery. The type of mucin and its presentation at the cell surface is likely to influence initial viral interactions at the epithelium and determine the course of infection. A more complete understanding of IAV interactions at the mucosal glycocalyx may, thus, provide a more accurate assessment of the potential of IAVs to infect human hosts.

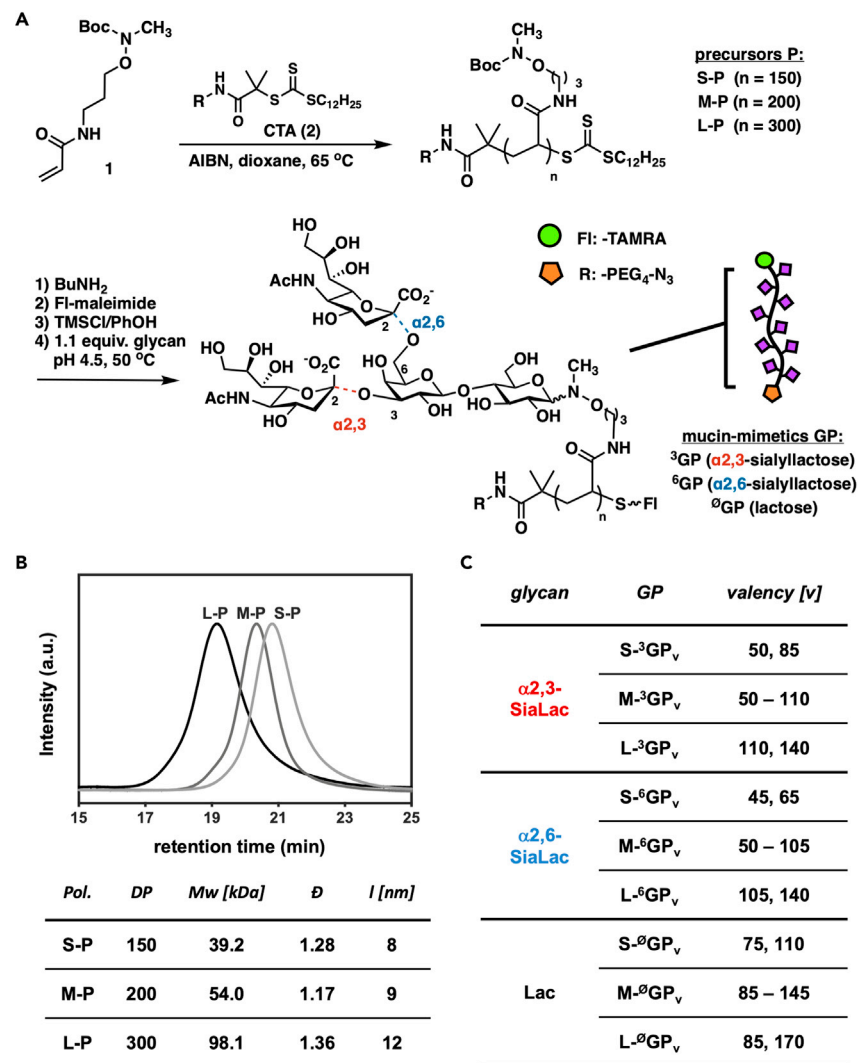
Here, we report the development of an array platform, in which synthetic mucin mimetics are used to model the mucosal epithelial cell glycocalyx, to evaluate receptor pattern recognition by IAVs (Figure 1B). We applied support vector machine (SVM) learning to identify and analyze effects of variations in glycan receptor type, mucin-mimetic valency, nanoscale dimensions, and crowding in the glycocalyx models on shifts in the binding specificity of H1N1 and H3N2 IAV strains. We found that mucin-like polyvalent presentations of  $\alpha$ 2,3- and  $\alpha$ 2,6-sialoglycans and the surface crowding of the glycoconjugates differentially impacted adhesion of the viruses, consistent with the proposed protective functions of mucins in the airway epithelium. The mucin-mimetic arrays also revealed an evolution of receptor pattern recognition by H1N1 produced in avian or mammalian cells, which could be characterized through machine learning.

## RESULTS AND DISCUSSION

### Construction of glycopolymers for mucin-like glycan receptor presentation

To model the mucinous glycocalyx environment in glycan arrays, we have devised a method for generating synthetic glycopolymers (GPs) that replicate key structural features of mucins (i.e., polyvalent glycans displayed along extended linear polypeptide chains) while allowing for tuning of the mimetic size and glycosylation pattern (Figure 2A). Using the reversible addition-fragmentation chain transfer (RAFT) polymerization, we have generated a collection of mucin mimetics of increasing length glycosylated with  $\alpha$ 2,3- and  $\alpha$ 2,6-sialyllactose trisaccharides ( $\alpha$ 2,3-SiaLac and  $\alpha$ 2,6-SiaLac) as model avian and human IAV receptors, respectively. The polymers were terminated with an azide functionality and used either as soluble probes or covalently grafted on cyclooctyne-coated glass via the strain-promoted alkyne-azide cycloaddition (SPAAC) reaction to produce a mucin-like glycocalyx display.<sup>30</sup> A tetramethylrhodamine (TAMRA) fluorophore was appended to the opposing chain end to allow for characterization of mucin-mimetic density on the arrays.

The mucin-mimetic synthesis began with the polymerization of a Boc-protected *N*-methylaminoxypropyl acrylamide monomer (1) in the presence of a chain transfer agent (CTA, 2) and a radical initiator (AIBN) to generate a set of azide-terminated short (S), medium (M), and long (L) polymeric precursors, P, with a degree of polymerization (DP) of 150, 200, and 300, respectively (Figure 2A). Size exclusion chromatography analysis (SEC, Figure 2B; Table S2) confirmed good control over the target molecular weight (MW) and dispersity ( $\mathcal{D}$ ) of the polymers. Next, the trithiocarbonate end groups in polymers P were removed by aminolysis, and the newly exposed thiol groups were capped with TAMRA-maleimide (Figure 2A; Scheme S1). The fluorophore labeling efficiency was determined for each polymer by UV-vis



**Figure 2. Generation of mucin-mimetic probes**

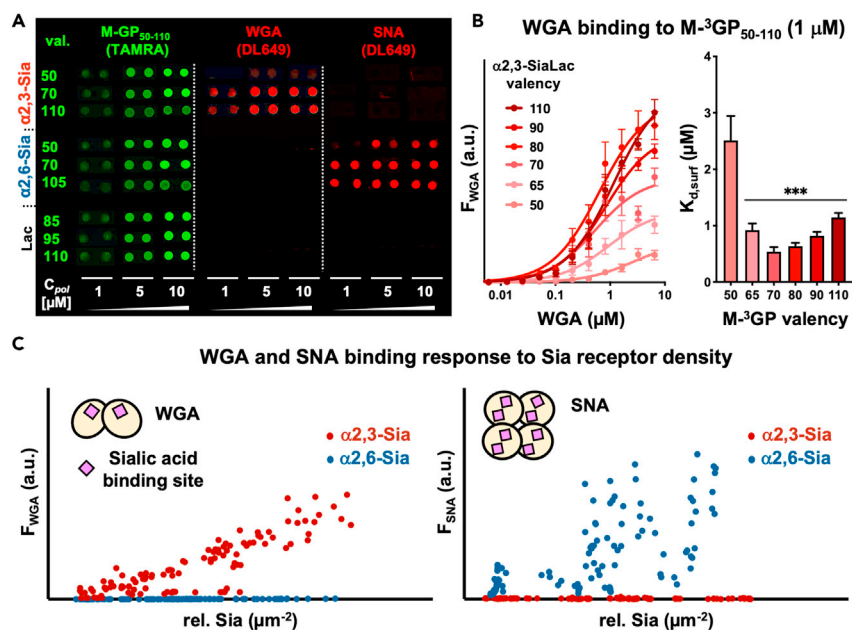
(A) Fluorescently labeled azide-terminated short (S), medium (M), and long (L) mucin-mimetic glycopolymers (GPs) were generated via RAFT polymerization.

(B) Size exclusion chromatography (SEC) analysis of the polymeric precursors P.

(C) The naming scheme for the GPs indicates the polymer backbone length (S-, M-, and L-), sialic acid linkage type (superscripts 3 and 6, and  $\emptyset$  designate  $\alpha$ 2,3-SiaLac,  $\alpha$ 2,6-SiaLac, and Lac, respectively), and glycan valency (final subscript).

spectrometry and ranged between 6%–30% (Tables S3). Side-chain Boc-group deprotection in the presence of phenol and trimethylsilyl chloride (TMSCl)<sup>31</sup> followed by conjugation of the released *N*-methylaminoxy groups with reducing glycans under acidic conditions completed the synthesis of the mucin-mimetic glycopolymers GP (Figure 2A).

The mucin-mimetic library comprised 27 short (S), medium (M), and long (L) sialylated glycopolymers, <sup>3</sup>GP and <sup>6</sup>GP, decorated with increasing amounts of the trisaccharides,  $\alpha$ 2,3-SiaLac and  $\alpha$ 2,6-SiaLac, respectively (Figure 2C; Table S3). In addition, we generated 11 control polymers lacking sialic acid modifications (<sup>0</sup>GP) displaying only the lactose disaccharide (Lac, Figure 2C; Table S3). The extent of



**Figure 3. Construction and validation of mucin-mimetic arrays**

(A) Representative composite images of density-variant arrays of fluorescent mucin-mimetic glycopolymers GP (TAMRA, green) probed with Daylight649-labeled SNA and WGA lectins. Each condition is represented as a duplicate. Full array scans are provided in Figure S5.

(B) Binding isotherms and associated apparent surface dissociation constants ( $K_{d,surf}$ ) for binding of WGA to medium-sized mucin mimetics M-<sup>3</sup>GP<sub>50-110</sub> with increasing α<sub>2,3</sub>-SiaLac valency printed at low surface density ( $c_{GP} = 1 \mu\text{M}$ , \*\*\*p < 0.0005 or greater).

(C) Binding responses of WGA and SNA to increasing glycan receptor density on the array. The dimeric WGA lectin binding is directly proportional to glycan density, whereas the tetrameric SNA lectin exhibits a more complex binding pattern. Insets represent graphical representation of lectin oligomeric state and orientation of sialic-acid-binding sites based on crystallographic data analysis (Figures S8 and S9).

glycosylation for all polymers was determined by <sup>1</sup>H NMR spectroscopy (Data S1) and varied according to glycan structure. Treatment with 1.1 equiv of glycan per polymer side chain was sufficient to achieve maximum polymer glycosylation of ~70% for Lac and ~45% for the negatively charged α<sub>2,3</sub>-SiaLac and α<sub>2,6</sub>-SiaLac (Table S3). The use of sub-stoichiometric amounts of glycans enabled tuning of glycan valency in the mucin mimetics (Figures 2C and S1). The mucin-mimetic lengths (*l*) were estimated to range from ~8 to 12 nm according to their DP using a method by Miura et al. for calculating theoretical end-to-end distances in sialylated glycopolymers (Figure 2B; Equation S1).<sup>32</sup>

### The oligomeric plant lectins, WGA and SNA, show distinct binding behavior in mucin-like receptor displays

In the airways, cell-surface-associated mucins are organized into a dense, brush-like glycocalyx, which projects tens to hundreds of nanometers above the epithelial cell surface.<sup>22</sup> To gain insights into glycan receptor recognition by proteins and pathogens at the mucosal interface, we modeled the mucinous glycocalyx in arrays by printing mucin-mimetic glycopolymers GP on cyclooctyne-functionalized glass (Figure 3A). In addition to varying the structure and glycosylation of the glycopolymer probes, we also modulated their surface crowding by increasing their concentration ( $c_{GP}$ ) from 1 to 10 μM in the printing buffer (PBS supplemented with 0.05% Tween 20, pH = 7.4). The fluorescent TAMRA labels introduced synthetically into the probes

were used to establish their surface grafting efficiency (Figure S2A) and the overall glycan receptor density (Figure S2C) for each polymer condition. The printing conditions yielded spots of uniform morphology (Figure 3A) with linear increase in polymer density across the employed concentrations regardless of polymer size or glycosylation (Figures S3 and S4).

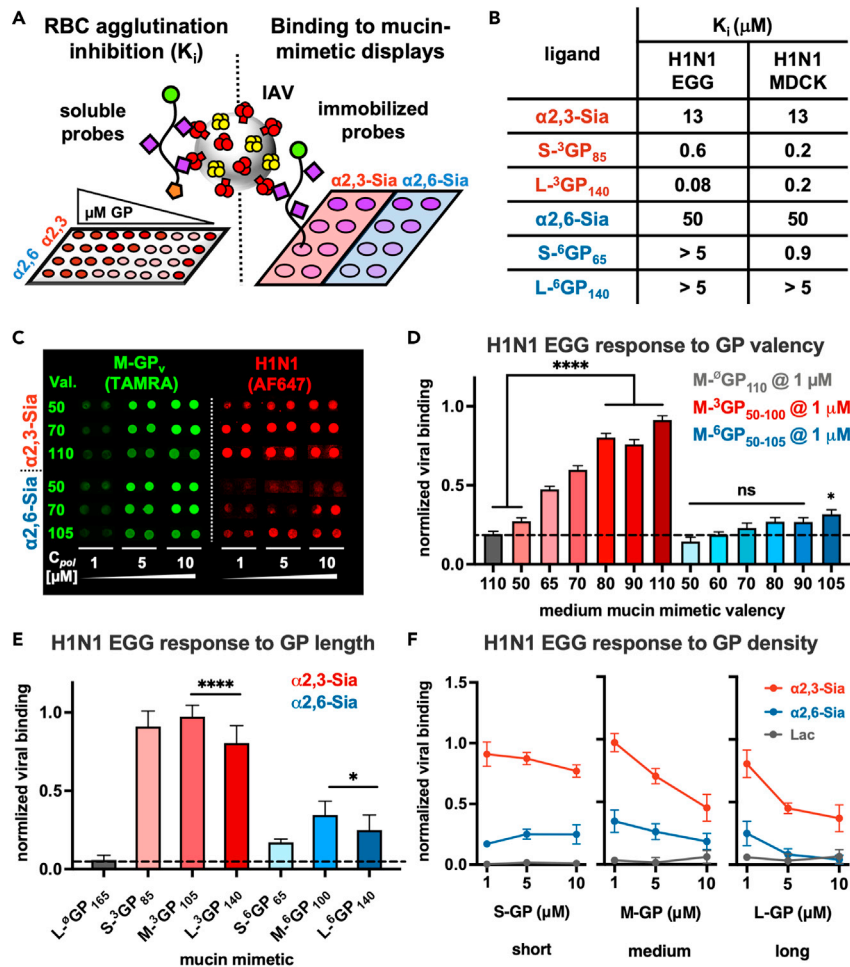
To confirm selective recognition of the mucin mimetics based on to the structure of their pendant glycans, the arrays were probed with Dylight649-labeled lectins wheat germ agglutinin (WGA) and *Sambucus nigra* agglutinin (SNA) (Figures 3A and S5). WGA, which primarily recognizes GlcNAc but has often been used to indicate the presence of  $\alpha$ 2,3-linked sialic acids, is specific for  $\alpha$ 2,3-SiaLac polymers on our arrays,<sup>33,34</sup> while SNA binds exclusively to the polymers containing the  $\alpha$ 2,6-linked isomer.<sup>35</sup> To obtain quantitative assessment of lectin binding to the mucin mimetics, we probed the arrays with increasing concentrations of the lectins to establish binding isotherms and extract apparent surface dissociation constants ( $K_{D,surf}$ ) (Figures 3B and S6). WGA binding to the medium-sized  $\alpha$ 2,3-SiaLac polymers,  $M$ -<sup>3</sup>GP<sub>50-110</sub>, printed at low surface density ( $c_{GP} = 1 \mu\text{M}$ ) showed valency-dependent binding with autoinhibition at the highest valencies caused by glycan crowding on the polymer backbone. This behavior is frequently observed for lectin binding to glycopolymer probes in solution.<sup>36</sup> The low polymer printing concentration produced probe spacing on the array surface that allowed for the measurement of lectin binding responses to the underlying glycoconjugate architecture. Increasing the concentration of the polymers resulted in denser mucin-mimetic arrays, attenuated WGA responsiveness to  $\alpha$ 2,3-SiaLac valency of the individual probes, and increased overall avidity of the dimeric lectin toward the receptor display (Figure S6; Table S4). Our attempts to establish similar binding profiles for SNA were not successful due to protein aggregation at concentrations needed to reach saturation binding (Figure S7).

Collecting thermodynamic binding data for each lectin-probe combination in the array can be time consuming and may not be possible for some lectins, as was the case for SNA. Simplified plots of lectin binding in response to changing relative sialoglycan density in the arrayed glycopolymer spots provide a convenient way to discern different binding modes of the proteins. Using this analysis, we observed that the binding response of WGA to changing glycan density was generally linear, while SNA showed a less correlated binding pattern indicative of contributions from higher-order binding interactions, such crosslinking of neighboring glycopolymers on the array (Figure 3C). Analysis of crystallographic data for WGA (PDB: 2CWG) and SNA (PDB: 3CA4) provides a structural basis for their differences in crosslinking capacity (Figures S8 and S9). WGA exists as a dimer with two sialic-acid-binding domains separated by 3.9 nm and positioned on the same face of the protein.<sup>34</sup> This arrangement reasonably favors WGA binding to glycans presented on the same mucin mimetic and may be responsible for the largely linear relationship between receptor density and lectin binding response. By contrast, SNA can exist as either a monomer, dimer, or tetramer.<sup>37</sup> Each monomer contains two glycan-binding sites that are directed outward on opposite the edges of the protein.<sup>38</sup> The various oligomeric states and the orientation of the binding sites make SNA more likely to engage and crosslink multiple glycoconjugates on the surface, producing the more complex binding behavior observed on the mucin-mimetic arrays.

### H1N1 PR8 virus shows linkage-specific differences in binding to mucin-like sialoglycan presentations

Pathogens, which utilize oligomeric lectins and adhesins for binding to cell surface glycan receptors, may be sensitive to the presentation of glycan receptors at the





**Figure 4. H1N1 EGG binding to mucin-mimetic displays of sialoglycan receptors**

(A) Red blood cell agglutination assays and array screens were used to probe the interactions of H1N1 produced in embryonated chicken eggs (H1N1 EGG) with soluble and surface bound mucin mimetics.

(B) Inhibitory activity,  $K_i$ , of soluble glycan receptors  $\alpha 2,3\text{-SiaLac}$  and  $\alpha 2,6\text{-SiaLac}$  and mucin-mimetic glycopolymers GP in RBC agglutination assays expressed as the minimal ligand concentration needed to prevent cell aggregation. The experimental images are included in [Figure S10](#).

(C) Representative composite images of H1N1 EGG virus (red) binding to medium-sized mucin mimetics M-GP (green) printed at low surface density ( $c_{\text{GP}} = 1 \mu\text{M}$ ) according to glycan receptor valency.

(D) Bar graph representation of H1N1 EGG virus binding to medium-sized mucin mimetics M-GP printed at low surface density ( $c_{\text{GP}} = 1 \mu\text{M}$ ) according to glycan receptor valency. Each array condition is represented as a duplicate and full array images are included in [Figures S11, S12, and S14](#). Values and error bars represent averages and standard deviations of experiments from 9 different arrays. Significance is based on viral binding to the Lac polymer control M- $^0\text{GP}_{110}$  (black dashed line).

(E) H1N1 EGG binding to mucin mimetics of increasing length printed at low surface density ( $c_{\text{GP}} = 1 \mu\text{M}$ ). Values and error bars represent averages and standard deviations of experiments from 6 different arrays. Significance was determined against Lac polymer control L- $^0\text{GP}_{165}$  (black dashed line).

(F) H1N1 EGG binding response to increasing crowding of mucin mimetics of all three lengths on the array surface. Values and error bars represent averages and standard deviations of experiments from 6 different arrays. (\* $p < 0.05$ , \*\* $p < 0.005$ , and \*\*\* $p < 0.0001$ ).

mucosal barrier.<sup>39</sup> We examined the binding of the H1N1 (A/Puerto Rico/8/1934 or PR8) virus to different presentations of sialoglycan receptors in our mucin-mimetic platform (Figure 4A). The PR8 strain is a well-characterized, laboratory-adapted human IAV strain, which has the ability to recognize both avian and human sialic acid receptor structures.<sup>40,41</sup> As such, it provides a useful model for assessing how receptor presentation may affect viral binding and selectivity.

H1N1, which was propagated in embryonated chicken eggs and henceforth labeled as H1N1 EGG, bound both receptor types in their soluble monovalent form, with  $\sim 4$ -fold preference for the  $\alpha 2,3$ -SiaLac isomer, as determined in red blood cell (RBC) agglutination inhibition assays ( $K_{i,\alpha 2,3} = 13 \mu\text{M}$  versus  $K_{i,\alpha 2,6} = 50 \mu\text{M}$ , Figures 4B and S10, and Table S5). The array binding data mirrored this preference, while providing additional insights into the effects of receptor presentation on viral interactions (Figures 4C–4F). H1N1 EGG virus binding to the medium size  $\alpha 2,3$ -SiaLac mucin mimetics  $\text{M-}^3\text{GP}_{50-110}$  immobilized at low surface densities ( $c_{\text{GP}} = 1 \mu\text{M}$ ) indicated enhanced viral capture with increasing receptor valency, with a valency threshold for binding above 50  $\alpha 2,3$ -SiaLac residues and a plateau at  $\sim 80$  glycans per polymer (Figures 4D and S11). Shortening the polymer length while maintaining a high receptor valency above 80 ( $\text{S-}^3\text{GP}_{85}$ ) had no negative effect on viral capture (Figures 4E and S12). We observed some decrease in binding to the longest mucin-mimetic  $\text{L-}^3\text{GP}_{140}$  compared with  $\text{M-}^3\text{GP}_{105}$  despite its higher valency, presumably due to its increased chain conformational flexibility. RBC hemagglutination inhibition assays with soluble  $\alpha 2,3$ -SiaLac polymers  $^3\text{GP}$  confirmed the observed valency-dependent binding trend for H1N1 EGG ( $K_{i,^3\text{GP}} = 313 \text{ nM} - 1.25 \mu\text{M}$ , Figure S10; Table S5) and were consistent with prior reports using similar multivalent glycopolymers.<sup>32</sup> In contrast to the behavior of the arrayed mucin mimetics, increasing the polymer length resulted in a more effective inhibition of RBC agglutination by H1N1 EGG in solution ( $K_{i,\text{S-}^3\text{GP}} = 625 \text{ nM}$  versus  $K_{i,\text{L-}^3\text{GP}} = 78 \text{ nM}$ , Figure S10). It should be noted that the increase in inhibitory capacity of the glycopolymers compared with the monovalent receptor can be accounted for based on glycan valency and concentration alone, rather than avidity enhancements due to multivalency. In the case of the short polymer  $\text{S-}^3\text{GP}_{85}$ , when the total amount of glycan on the polymer is taken into account, the per glycan inhibitory activity ( $K_{i,\text{S-}^3\text{GP}} \times \alpha 2,3$ -SiaLac valency =  $53 \mu\text{M}$ ) was effectively reduced compared with the free  $\alpha 2,3$ -SiaLac ( $K_{i,\alpha 2,3\text{-SiaLac}} = 13 \mu\text{M}$ ). Lactose glycopolymers,  $^0\text{GP}$ , lacking sialic acids served as negative controls in both assays (Figures 4 and S10).

Glycopolymers carrying the  $\alpha 2,6$ -SiaLac receptor ( $\text{M-}^6\text{GP}_{50-105}$ ) showed only a limited ability to engage H1N1 and required glycan valency above 90 to reach binding above background (Figures 4D and S11). Extending the length of the mucin mimetic, again, resulted in a decrease in viral capture (Figures 4E and S12). All of the  $\alpha 2,6$ -SiaLac polymers failed to inhibiting RBC hemagglutination by the virus over the range of tested polymer concentrations ( $K_{i,^6\text{GP}} > 5$  or  $325 \mu\text{M}$  with respect to  $\alpha 2,6$ -SiaLac, Figure S10). Considering that monovalent  $\alpha 2,6$ -SiaLac can prevent RBCs agglutination ( $K_{i,\alpha 2,6} = 50 \mu\text{M}$ , Figures 4B and S10), it appears that binding of H1N1 EGG to this glycan receptor is disfavored in the polyvalent glycopolymer presentation.

High levels of mucin expression on the surfaces of epithelial cells produces a dense glycoprotein brush, which has been proposed to restrict IAV access to membrane receptors necessary for infection.<sup>26,27</sup> To examine the effects of polymer size and density on viral adhesion, we modeled glycolyx crowding in our arrays by increasing

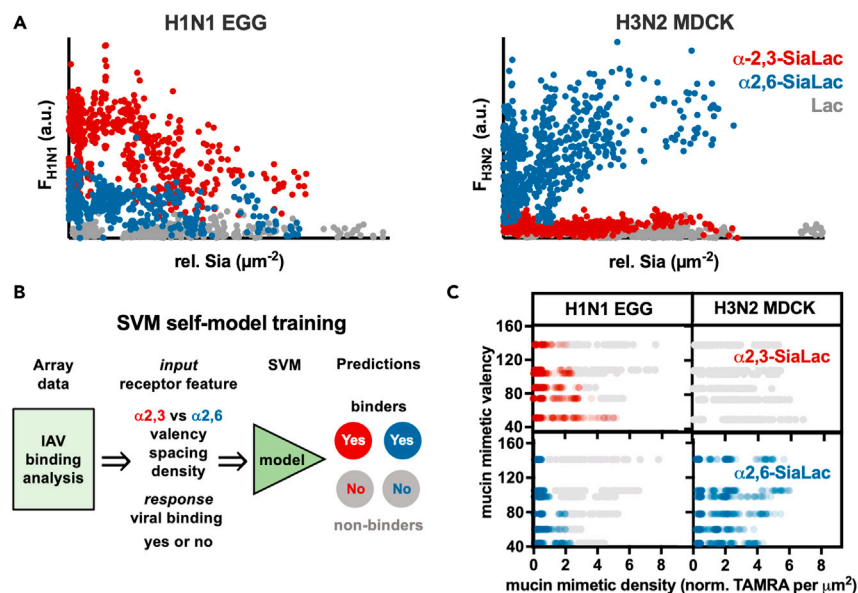
the printing concentration of the mucin mimetics. We assayed H1N1 EGG binding to maximally glycosylated mucin mimetics of all three lengths arrayed at concentrations of 1, 5, and 10  $\mu\text{M}$  (Figures 4F and S12). The virus retained its overall preference for the  $\alpha 2,3\text{-SiaLac}$  probes across all surface densities; however, increased crowding of the polymers led to attenuated viral adhesion, which became more pronounced with increasing mucin-mimetic length. Crowding of the  $\alpha 2,6\text{-SiaLac}$  glycopolymers both enhanced (S- $^6\text{GP}_{65}$ ) and inhibited (L- $^6\text{GP}_{140}$ ) viral adhesion depending on polymer length (Figure 4F). Our data show that, while the H1N1 EGG virus can utilize the less preferred  $\alpha 2,6\text{-SiaLac}$  receptors when presented in surface displays on short mucin-mimetic scaffolds, increasing the length and density of the conjugates generally negatively impacted viral adhesion regardless of receptor type. Such negative impact of increasing receptor density was previously reported for the binding of nanoparticles bearing recombinant HA proteins to sialoglycans in supported lipid bilayers.<sup>42</sup> Thus, crowding of mucins in the glycocalyx may not only shield underlying glycan receptors from the virus,<sup>26</sup> but also limit viral adhesion to the heavily sialylated mucins themselves.

The observed differential H1N1 EGG binding to the mucin-like receptor displays according to sialic acid linkage type supports the distinct functions of secreted and membrane bound mucins comprising the airway mucosal barrier.<sup>26,43</sup> Therein, secreted mucins produced by goblet cells and presenting primarily  $\alpha 2,3\text{-sialic acid}$  modifications serve as decoy receptors for viral capture and clearance. By contrast, the MUCs produced by epithelial cells display  $\alpha 2,6\text{-linked sialic acid}$  receptors and are thought to limit viral adhesion. The binding of H1N1 EGG to polyvalent  $\alpha 2,3\text{-SiaLac}$  mucin mimetics but not the  $\alpha 2,6\text{-SiaLac}$  analogs and the sensitivity of the virus to surface crowding of mucin mimetics carrying both receptor types would provide a rationale for the synergistic but mechanistically distinct functions of secreted and surface bound mucins in limiting viral infection.

### SVM learning-enabled analysis of receptor pattern recognition by influenza A viruses

While the focused analysis of H1N1 EGG binding to some of the key features of glycan presentation in the mucin-mimetic array (e.g., polymer size, valency, or surface density; Figure 4) was informative, the multidimensionality of receptor presentation on the array makes comprehensive assessment challenging and time consuming. For SNA and WGA, plots of lectin capture on the array according to glycan abundance, regardless of polymer structure or brush density, revealed qualitative differences in the lectin binding behavior (Figure 3C). Similar scatterplot representations can highlight major differences in receptor pattern recognition by different IAVs, as shown in Figure 5A for H1N1 EGG and the H3N2 A/Aichi/2/68 strain produced in Madin-Darby canine kidney (MDCK) cells, H3N2 MDCK.

Noting that molecular recognition is a multi-dimensional problem, we have applied a SVM learning approach (Figure 5B) to resolve the receptor patterns that are best recognized by the viruses. We leveraged the fact that the SVMs are known to predict accurate relationships from semi- or unstructured data.<sup>44</sup> First, we established a binding threshold for the virus based off the background signal from the control Lac polymers (Figure S13). Then, we randomly selected a portion (67%) of the array binding data for each virus to train the SVM using a five-dimensional parameter space of glycan type, glycan valency, glycan spacing on polymer, glycan density on the array, and polymer printing concentration (Figure 5B). In combination, these



**Figure 5. Analysis of receptor pattern recognition by H1N1 and H3N2 strains**

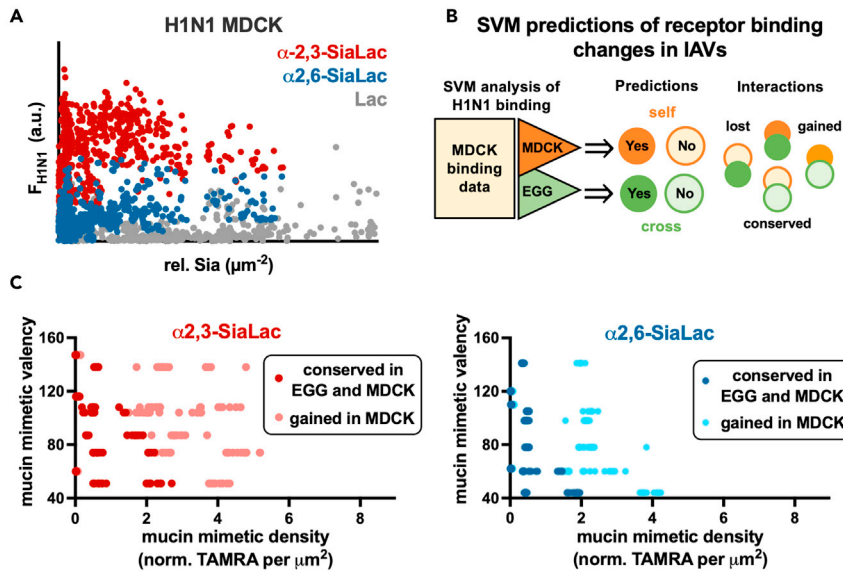
(A) Scatterplots of binding responses to changing glycan receptor density in mucin-mimetic arrays for H1N1 EGG and H3N2 MDCK viruses.

(B) Workflow for creating SVM self-models using viral binding responses and receptor display parameters (i.e., glycan type, valency and spacing on mucin mimetic, mucin-mimetic density, and concentration in printing buffer) in mucin-mimetic arrays.

(C) SVM analysis of H1N1 EGG and H3N2 MDCK binding in mucin-mimetic arrays. In red and blue colors are predicted binding events for each virus to  $\alpha 2,3$ -SiaLac and  $\alpha 2,6$ -SiaLac, respectively. Non-binding events are shown in gray. Color intensity indicates the frequency of the predicted binding events according to the valency and surface density of the mucin mimetics. H1N1 EGG recognizes both  $\alpha 2,3$ - and  $\alpha 2,6$ -SiaLac with preference for low surface densities of mucin mimetics. H3N2 MDCK is specific for  $\alpha 2,6$ -SiaLac glycans and its binding does not diminish with increasing mucin-mimetic density on the array.

parameters defined additional features of our mucin-mimetic receptor displays, such as glycopolymer length (via glycan valency and glycan spacing on polymer) and glycopolymer density on the array surface (via glycan valency, glycan spacing on polymer, and glycan density on array). We used the remaining portion of the binding dataset (33%) to test the accuracy, recall, and precision of the model (Figures S14 and S15). This model predicted binding of H1N1 EGG to both receptor types, with majority of binding events occurring toward lower glycan valencies and polymer surface densities (Figure 5C, left), consistent with our manual analysis of the array data for this virus (Figure 4). The prediction plots in Figure 5C show the parameters of receptor presentation that most influenced viral binding (i.e., glycan receptor type, mucin-mimetic valency and surface density). Additional parameters, such as polymer size and glycan spacing on the polymer, also contribute to viral recognition and can be analyzed. We confirmed that the performance of the model was similar when the training and testing datasets represented either separate array experiments or were selected randomly from combined data across multiple experiments (Figure S16).

The receptor-binding preferences of other IAV strains on the mucin-mimetic array can be rapidly analyzed using the machine learning approach. The application of the SVM to the array binding for H3N2 MDCK correctly predicted the specificity of the virus for  $\alpha 2,6$ -sialoglycans,<sup>10,45</sup> and, newly, identified its better ability to utilize



**Figure 6. Analysis of changes in receptor pattern recognition by IAVs produced in avian and mammalian cells**

(A) Scatterplots of binding responses to changing glycan receptor density in mucin-mimetic arrays for H1N1 MDCK.

(B) The H1N1 MDCK and H1N1 EGG SVM models can be applied to the MDCK array binding data to generate “self” and “cross” predictions, respectively, that can be used to determine lost, conserved, and gained interactions.

(C) The SVM identified better utilization of increasingly crowded mucin mimetic displays by H1N1 upon transition from avian to mammalian cell culture.

increasingly dense glycopolymer displays (Figure 5C). Manual analysis of the array binding data confirmed these predictions (Figures S17 and S18), demonstrating the general applicability of the supervised models for analyzing receptor pattern recognition by IAVs.

### H1N1 PR8 propagation in mammalian cells enhances interactions with $\alpha$ 2,6-sialoglycans in mucin-like displays

Having established the SVMs as an effective method to rapidly identify preferred receptor display parameters for viral binding in the mucin-mimetic arrays, we set to explore how the glycan-binding phenotype of the virus may change depending on the host in which it is propagated. Such information may enhance existing viral surveillance by either eliminating glycan-binding phenotype artifacts, which can be introduced during propagation of field-isolated viruses in the laboratory, or by establishing specific binding phenotype features associated with enhanced human transmission.

We performed a comparative solution and array binding analysis between H1N1 EGG and the same virus produced in MDCK cells (H1N1 MDCK). The MDCK cell line is a commonly used mammalian system for the propagation of IAVs in the laboratory. The binding of H1N1 MDCK to the soluble monovalent  $\alpha$ 2,3-SiaLac and  $\alpha$ 2,6-SiaLac, as measured via RBC hemagglutination inhibition, remained unchanged from that of H1N1 EGG (Figure 4B). However, H1N1 MDCK exhibited improved ability to engage mucin mimetics carrying the  $\alpha$ 2,6-SiaLac receptors both on arrays (Figures 6A, S19, and S20) and in their soluble form ( $K_{i,S-6GP65} = 938$  nM, Figure 4B). The increased ability to recognize  $\alpha$ 2,6-sialoglycans is expected

for H1N1 propagated in MDCK cells, as the cells in the allantoic fluid of embryonated chicken eggs used for viral propagation display mainly  $\alpha$ 2,3-linked sialic acids and the surfaces of MDCK cells are populated by receptors with both linkages.<sup>46</sup> During culture in MDCK cells, viruses can undergo selection for both receptor types, giving rise to virus population with an altered sialoglycan-binding phenotype.<sup>11,46</sup> However, our hemagglutination inhibition assays indicate that the altered glycan-binding phenotype of H1N1 MDCK does not arise from changes in HA affinity for the individual monovalent receptors.

Next, we adapted the SVM analysis to identify viral binding features for H1N1 that are conserved between viruses produced in embryonated eggs and those propagated in MDCK cells. We trained and validated a new SVM model for H1N1 MDCK as described above. The H1N1 EGG and H1N1 MDCK models were able to accurately produce predictions of binding and non-binding events for each respective virus in the mucin-mimetic array (Figures S14C and S15C). These, we termed “self-predictions.” The EGG model was then applied to the array binding data for H1N1 MDCK to produce an EGG-to-MDCK “cross prediction” that was compared with the H1N1 MDCK “self-prediction” results (Figures 6B and S21). The binding events correctly predicted by both models were termed “conserved in MDCK.” Those refer to receptor presentations that are recognized by H1N1 regardless of whether it was produced in avian or mammalian cells (Figure S21). The interactions correctly predicted as binding by the H1N1 MDCK model but were deemed non-binding by the H1N1 EGG model, i.e., interactions that were absent in the EGG-to-MDCK “cross prediction”, were termed “gained in MDCK”. These refer to interactions with glycan receptor patterns in the array that were absent or in primarily non-binding regions for the H1N1 virus produced in eggs but emerged when the virus was propagated in MDCK cells. Finally, the nonbinders that were correctly predicted by the H1N1 MDCK model but were predicted to be binders by the H1N1 EGG model, i.e., interactions that were absent in H1N1 MDCK but were anticipated to occur based on the EGG-to-MDCK prediction, were termed “lost in MDCK”. These refer to glycan patterns recognized by the H1N1 virus produced in eggs but lost in mammalian cell culture.

As shown in Figure 6C, the conserved and gained interactions predicted by the SVM algorithm occur in distinct clusters with most of the interaction gain happening at higher polymer valencies or densities. The predictions are in agreement with our agglutination inhibition data and preliminary manual array analysis (Figures 4 and S19) pointing to improved ability of the H1N1 MDCK virus to engage mucin-mimetic displays, both as individual polymers in solution as well as in ensembles on arrays. For  $\alpha$ 2,6-SiaLac glycopolymers, sensitivity of the virus to receptor crowding continued to persist in H1N1 MDCK, as the majority of binding gains resulted either from higher valency polymers grafted at low densities or from surface crowding of low valency mimetics (Figures 6C and S22). Likewise, the SVM analysis identified stronger interactions of H1N1 MDCK with high densities of  $\alpha$ 2,3-SiaLac receptors in the mucin-mimetic array (Figure 6C). The predictions did not identify loss of any interactions present in the H1N1 EGG upon transition into the mammalian culture system. To validate this method, we compared the results of the cross prediction with experimental data from H1N1-EGG. We confirmed that the interactions identified as conserved and gained in H1N1 MDCK by the SVM also correlated with binding and non-binding events, respectively, observed for H1N1-EGG in the array. We observed a high level of agreement between the cross predictions and experimental data in the low and high receptor valency-density regions on the array, where the conserved and gained interactions are well defined (Figure S22).

We note that an analysis of conservation, gain or loss of binding is made possible by application of the five-dimensional SVM, wherein the algorithm learns the “rules of binding” from a set of interaction patterns (defined by glycan valency, density, polymer spacing, and printing concentration data), which is then compared with another dataset to bring out the similarities and differences in the interactions. Arriving at these conclusions would be impossible using only chemical intuition or simple visualization of data because the interaction patterns are not apparent when the data are plotted with equal weights on all five dimensions. Using between 200 to 2,000 iterations, the SVM finds linear combinations of the parameters in this five-dimensional space, reweighting the original dimensions, until the most separated classes are created, and their underlying data features are segregated, to distinguish between binding and non-binding events.

It is unlikely that such combinations can be found manually, or using unsupervised clustering schemes, making SVM a natural choice to create an “intelligent” space that correlates binding rules to linear combinations of glycan valency and spacing on the mimetics and their density on the array surface. In contrast to popular simple clustering algorithms, such as K-means, which are limited by the number of dimensions, the strength of SVM is that it does not impose the ground-truth or prior knowledge of the binding patterns to learn the rules of binding.<sup>47</sup> Rather, SVM infers both the rule as well as the outcome directly from higher-dimensional datasets. Although we could have employed more sophisticated algorithms, such as neural networks, the use of a discrete binary classifier in this study (binder = 1 and non-binder = 0) made SVM a more convenient choice. The SVMs are not solved for local optima, thus, they handle well high-dimensional data using one to two classifiers, as is the case here.

The improved range of binding interactions of H1N1 MDCK with  $\alpha$ 2,6-linked sialic acids is expected based on the higher expression of these glycans in the mammalian cells. Even in absence of *Muc* gene expression in MDCK cells,<sup>48</sup> the observed changes in receptor pattern recognition for both glycans do not stem from altered affinity of its HA toward the monovalent glycans in solution (Figure 4B). The HA protein carries several glycosylation sites near the sialic-acid-binding region and the addition of glycans can influence binding<sup>13</sup> and may give rise to the observed changes in receptor pattern recognition.

We performed PAGE analysis of the viral proteins before and after PNGase treatment to remove N-linked glycans. The glycosylated form of HA1 fragment from the egg-grown H1N1 virus had lower MW than that of the MDCK virus (Figure S23). This difference in gel mobility was eliminated by PNGase treatment, which catalyzes the removal of N-linked glycans. This points to a more extensive glycosylation of the MDCK cell-derived virus HA proteins and is consistent with previously observed differences in the glycosylation, but not in the primary amino acid sequence, of HAs from isolated duck H1N1 viruses propagated in MDCK cells and egg cultures.<sup>49</sup> Whether the relationship between the changes in HA glycosylation and the altered receptor pattern recognition of the virus observed in the current array study is causative or correlative is yet to be determined. Nonetheless, it presents one possible explanation for the altered binding behavior of the viruses. Such differences in glycosylation may reflect the influence of species, cell type, or combination of both on receptor pattern recognition by viruses and contribute to their emerging ability to cross between species. Similar analysis of glycan-binding changes upon transition from egg culture to MDCK cells was not possible for the H3N2 virus used in this study, which, due to its restricted receptor specificity, does

not propagate well in embryonated eggs without undergoing mutations in its HA sequence.<sup>50</sup> However, we anticipate that the array can help identify other viruses with glycan-binding phenotype shifts occurring in response to adaptation to a new host.

### Conclusions

The glycocalyx exists as a complex assortment of membrane-tethered and secreted glycoconjugates that serve different roles in multicellular identity, function, and pathogen invasion. In this work, we aimed to model the interactions of IAVs with the mucin glycoprotein components of the mucosal barrier by generating mucin-mimetic glycopolymers with tunable sizes and compositions displaying  $\alpha$ 2,3- and  $\alpha$ 2,6-linked sialyllactose glycans as prototypes for the avian and human receptors for IAVs. RBC hemagglutination inhibition assays with soluble forms of the probes revealed an enhancement in selectivity of H1N1 PR8 viruses produced in embryonated chicken eggs from  $\sim$  4-fold preference for binding of  $\alpha$ 2,3-SiaLac in its soluble monomeric form to more than 60-fold for the polyvalent receptor displays. This differential arose from selectively disfavored binding of the virus to the  $\alpha$ 2,6-SiaLac glycans in the mucin mimetics rather than increased avidity toward the  $\alpha$ 2,3-SiaLac polymers. Systematic evaluation of H1N1 EGG capture on arrays of immobilized mucin mimetics enabled by an SVM learning algorithm showed that the virus bound surface displays of both receptor types but was attenuated at high polymer densities. The receptor pattern recognition changed when the virus was propagated in MDCK cells toward improved utilization of  $\alpha$ 2,6-SiaLac mucin mimetics in both soluble and immobilized forms, and a lower sensitivity to surface crowding of the  $\alpha$ 2,3-SiaLac glycopolymers.

Our findings are consistent with the proposed protective functions of soluble mucins at the airway epithelium which present primarily  $\alpha$ 2,3-linked receptors and prevent infection through viral capture and clearance. Newly, our observations that the presentation of  $\alpha$ 2,6-linked sialoglycans on linear polyvalent scaffolds and the arrangement of such conjugates in increasingly dense surface ensembles disfavors the binding of H1N1 produced in avian cells also provides support for the role of membrane-associated mucins in limiting viral adhesion and infection at the epithelial cell surface. The improved binding of H1N1 viruses produced in mammalian cells to the mucin-mimetic displays of  $\alpha$ 2,6-linked sialoglycans was not accompanied by changes in the affinity or selectivity of their HA proteins toward the individual glycan receptors. While the basis for the differences in receptor pattern recognition needs to be further investigated, our studies show that the mucin probes and their arrays may serve as useful tools to investigate viral interactions at the mucosal barrier and the evolution of their glycan receptor-binding phenotype.

The array and the SVM analysis can be extended to other IAV strains. The platform successfully predicted the  $\alpha$ 2,6-sialoglycan specificity of the H3N2 A/Aichi/2/68 virus and identified its better ability to utilize polyvalent and high-density mucin-like displays of the human receptor prototype,  $\alpha$ 2,6-SiaLac. The modularity of the mucin-mimetic synthesis and the ease with which glycans can be incorporated into these materials<sup>51</sup> should be well suited for the introduction of more complex and biologically relevant glycans into the array, including mucin-type O-glycans. Of particular interest for future array development are glycans with linear or branched core structures variously modified with extended N-acetylglucosamine disaccharide repeats and presenting fucosylation or sulfation motifs, as they continue to emerge from glycomics<sup>13,14,28,52</sup> and genetic screens<sup>16,53</sup> as relevant to IAV recognition and infection. Machine learning, as employed in this study, is well poised to support such



increase in the dimensionality of parameter space on the array and can be expanded to non-binary classification for analysis. This may allow future inclusion of multiple continuous classifiers, such the strength and kinetics of a binding interaction between a virus and a receptor pattern, to construct more sophisticated neural networks that reveal higher-order relationships between pathogens and their hosts.

## EXPERIMENTAL PROCEDURES

### Resource availability

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kamil Godula ([kgodula@ucsd.edu](mailto:kgodula@ucsd.edu)).

#### Materials availability

This study did not generate new unique reagents. Glycopolymers and microarrays can be made available through contact with the lead contact. Influenza virus A/PR/8/34 (H1N1) is available from ATCC using code VR-1469.

#### Data and code availability

All SVM datasets and code generated for our machine-learning analysis of IAV binding to the mucin-mimetic arrays are available on Github and can be accessed through <https://doi.org/10.5281/zenodo.5563187>

A detailed list of chemical and biological reagents including their sources and catalog numbers can be found in [Table S1](#).

### Glycopolymer synthesis and characterization

Fluorescently labeled glycopolymers end-functionalized with an azide (GP) used in this study and their polymeric precursors (P) were prepared using the RAFT polymerization method according to previously published procedures<sup>54</sup> and are summarized in [Figure 2](#) and described in detail in [supplemental information](#).

### Cell culture

MDCK cells: MDCK cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% FBS, 100 U/mL penicillin, and 100 U/mL streptomycin.

### Viral culture

Influenza virus strain A/PR/8/34 (H1N1, ATCC VR-1469) was purchased from ATCC and propagated in MDCK cells that were transferred to DMEM medium supplemented with 0.2% BSA fraction V, 25 mM HEPES buffer, 2  $\mu$ g/mL TPCK-trypsin, and 1% penicillin/streptomycin ("DMEM-TPCK" media). The same strain was used for viral production in embryonated chicken eggs. Briefly, fertilized chicken eggs were obtained and stored at 37°C. When the embryos were 10 days old (assessed by candeling), virus was injected through small holes made in the shell over the air sac. The holes were covered with parafilm, and the eggs were placed back at 37°C. After 2 days, the eggs were chilled to prepare for harvest. The eggshells were cut open above the air sac, and the allantoic fluid was carefully collected into centrifuge tubes without rupturing the yolk. The tubes were centrifuged to pellet any debris and the supernatant-containing virus was aliquoted into 1 mL cryovials and stored at  $-80^{\circ}\text{C}$ .

### Viral titers

Turkey red blood cells were purchased from Lampire and a 1% solution was used to determine viral titers via the hemagglutination test. MDCK cells were used to

determine the 50% tissue culture infective dose (TCID<sub>50</sub>) using the Spearman-Kärber method.

### Hemagglutination inhibition

Glycopolymer solutions in PBS (25  $\mu$ L, ranging from 20  $\mu$ M to 20 nM by 2-fold dilutions) were added to a 96-well plate. The last well in each row was used as a PBS control and did not contain glycopolymer or virus. An equal volume (25  $\mu$ L) of H1N1 diluted to HAU = 4 (turkey RBCs) was added to each well and allowed to incubate at room temperature. After a 0.5 h incubation, 50  $\mu$ L of 1% turkey RBCs in PBS were added to all of the wells. The K<sub>i</sub> value was read after a 0.5 h as the lowest concentration of polymer to inhibit hemagglutination.

### Array construction

Microarrays were printed on cyclooctyne-coated glass slides as previously described<sup>54</sup> using a sciFLEXARRAYER S3 printer (Scienion) following passivation with a 1% BSA/0.1% Tween-20 solution in PBS for 1 h. Polymer solutions were diluted in printing buffer (0.005% Tween-20 in PBS) to concentrations of 1, 5, and 10  $\mu$ M polymer and printed in replicates of six at a humidity of 70%. Following an overnight reaction at 4°C, excess polymer was removed by vigorous washing in 0.1% Triton X/PBS solution. The slides were then imaged on an Axon GenePix4000B scanner (molecular devices) at the highest photomultiplier setting (PMT) possible without saturation of pixels. The GenePix software was used to calculate the relative polymer density, by dividing the fluorescent intensity at 532 nm by the labeling efficiency for each polymer (obtained through UV-vis measurements) and the spot area (calculated from the spot diameter generated by the software). To obtain relative glycan density, the polymer density was multiplied by the glycan valency for each polymer (attained through NMR integration).

### Array binding assays

Prior to array binding assays, subarrays were blocked with 3% BSA solution in PBS for 1 h. For lectin binding, subarrays were washed three times with lectin binding buffer (0.005% Tween-20 in PBS with 0.1 mM CaCl<sub>2</sub>, MnCl<sub>2</sub>, and MgCl<sub>2</sub>). Daylight-labeled SNA and WGA were diluted in the lectin binding buffer and incubated on the array for 1 h in the dark. After washing with binding buffer, 0.1% Tween-20 solution in PBS, and rinsing with MilliQ water, the slides were imaged at the highest PMT possible without producing saturated pixels. For H1N1 binding, subarrays were washed with 1% BSA/PBST three times following passivation. H1N1 was diluted in 1% BSA/PBST and incubated on the array for 1 h. The slide was washed two times with 1% BSA/PBST and then fixed for 20 min with 4% PFA in PBS. To visualize H1N1, binding a 1:500 dilution of anti-HA in 1% BSA/PBST was incubated on the array for 1 h, followed by 1-h incubation in the dark of a 1:500 dilution of anti-rabbit-AF647 antibody. The subarrays were washed two more times with 1% BSA/PBST, two times with the 0.1% Tween-20 solution in PBS, rinsed with MilliQ and imaged at the highest PMT possible without producing saturated pixels.

### Machine-learning workflow

The machine learning workflow is demonstrated in [Figure S21](#). The H1N1 EGG, H3N2, and H1N1 MDCK datasets contained 8,373, 1,922, and 1,273 data points, respectively. Given the size of our data, the feature space (4–5 variables), and the category of problem (binary classification), we chose to use SVM algorithm for this work. SVM is ideally suited for such binary classification tasks. Likewise, more sophisticated algorithms like random forest shows ideal performance with larger datasets and/or more complex features. Use of random forest in place of SVM did not show any improvements (data not shown). As a preprocessing step, all negative viral

fluorescent intensities (which resulted from background subtraction in the absence of viral binding) were adjusted to zero to indicate the lack of viral binding. The fluorescence intensities were then normalized over the entire dataset in the range of [0,1]. This continuous data were converted to categorical ("binders"/"nonbinders") based on a cutoff fluorescence that was determined from the distribution of fluorescence intensities of lactose samples which served as negative control. Next, the features (glycan type, valency, and polymer density) were scaled to a range of [0,1] in order to avoid bias from higher values. Glycan type was the only categorical feature that was transformed into a numerical value by mapping to a two-dimensional space where  $\alpha$ 2,3-SiaLac was represented by (1,0) and  $\alpha$ 2,6-SiaLac by (0,1). This dataset was then split into a training and test set, where training set contained 67% of the data. Each experiment containing 6 fluorescence measurements was individually split between training and test sets. The SVM algorithm was used for learning from the training data, and predictions were validated against the test data. We confirmed that the performance of the model was similar when the data were split across all experiments for creating training and test sets (Figure S16). Training and testing were performed separately for H1N1 EGG, H3N2, and H1N1 MDCK data. Convergence was reached within 2,000 iterations for H1N1 EGG data and 200 iterations for H3N2 and H1N1 MDCK data (Figure S14). Next, the algorithm trained on H1N1 EGG data was used to predict results for H1N1 MDCK virus data, and the prediction results from this so-called "cross model" was compared with the results obtained from the model trained on H1N1 MDCK virus (the "self-model"). Data were prepared in python using pandas, numpy, and scipy packages. SVM was performed using python's scikit-learn package.<sup>55</sup>

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.chempr.2021.09.015>.

## ACKNOWLEDGMENTS

We thank the UCSD Glycobiology Research and Training Center for access to tissue culture facilities and analytical instrumentation, Dr. Christopher Fisher for his help with virus culture and characterization methods, and the ASU Research Computing and their Agave computing cluster for the computational resources. The authors also wish to thank Prof. Mia Huang (TSRI) for her technical advice and valuable insights over the course of this research. This work was supported in part by the NIH Director's New Innovator Award (NICHD: 1DP2HD087954-01), the NIH Director's Glycoscience Common Fund (1R21AI129894-01), and the Gordon and Betty Moore Foundation via a Scialog grant (#9162.07). K.G. was supported by the Alfred P. Sloan Foundation (FG-2017-9094) and the Research Corporation for Science Advancement via the Cottrell Scholar Award (grant #24119). T.M.L. was supported by the Chemistry and Biology Interface training program (NIGMS: 5T32GM112584-03). A.S. was supported by a CAREER award from the NSF (MCB-1942763) and an RO1 grant from the NIH (GM095583). C.G. was supported by DoD National Defense Education Program Center (Contract #HQ00342110007: Singharoy PI). A.S. and C.G. also acknowledge funds from ASU-Mayo Foundation, AstraZeneca, and start-up grants from Arizona State University School of Molecular Sciences and Biodesign Institute's Center for Applied Structural Discovery. P.G. and M.O.A. are supported by the G. Harold and Leyla Y. Mathers Charitable Foundation. E.S. was supported by the UCSD Paths Program and the Undergraduate Summer Research Award through the UCSD Division of Physical Sciences.

## AUTHOR CONTRIBUTIONS

Conceptualization, K.G. and A.S.; methodology, K.G., A.S., T.M.L., C.G., and M.O.A.; software, A.S. and C.G.; validation, T.M.L. and C.G.; formal analysis, T.M.L., C.G., and M.R.N.; investigation, T.M.L., and C.G.; data curation, T.M.L., C.G., E.S., and M.R.N.; writing – original draft, K.G., A.S., T.M.L., and C.G.; writing – review & editing, K.G., A.S., T.M.L., and C.G.; funding acquisition, K.G. and A.S.; resources, M.O.A. and P.G.; supervision, M.O.A. and P.G.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community.

Received: April 21, 2021

Revised: August 30, 2021

Accepted: September 28, 2021

Published: October 22, 2021

## REFERENCES

1. Keni, R., Alexander, A., Nayak, P.G., Mudgal, J., and Nandakumar, K. (2020). COVID-19: emergence, spread, possible treatments, and global burden. *Front. Public Health* 8, 216. <https://doi.org/10.3389/fpubh.2020.00216>.
2. Long, J.S., Mistry, B., Haslam, S.M., and Barclay, W.S. (2019). Host and viral determinants of influenza A virus species specificity. *Nat. Rev. Microbiol.* 17, 67–81.
3. Lipsitch, M., Barclay, W., Raman, R., Russell, C.J., Belser, J.A., Cobey, S., Kassin, P.M., Lloyd-Smith, J.O., Maurer-Stroh, S., Riley, S., et al. (2016). Viral factors in influenza pandemic risk assessment. *Elife* 5, e18491. <https://doi.org/10.7554/eLife.18491>.
4. Suzuki, Y. (2005). Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. *Biol. Pharm. Bull.* 28, 399–408. <https://doi.org/10.1248/bpb.28.399>.
5. Raman, R., Tharakaraman, K., Shriver, Z., Jayaraman, A., Sasisekharan, V., and Sasisekharan, R. (2014). Glycan receptor specificity as a useful tool for characterization and surveillance of influenza A virus. *Trends Microbiol.* 22, 632–641. <https://doi.org/10.1016/j.tim.2014.07.002>.
6. Air, G.M. (2014). Influenza virus-glycan interactions. *Curr. Opin. Virol.* 7, 128–133. <https://doi.org/10.1016/j.coviro.2014.06.004>.
7. Skehel, J.J., and Wiley, D.C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* 69, 531–569. <https://doi.org/10.1146/annurev.biochem.69.1.531>.
8. McAuley, J.L., Gilbertson, B.P., Trifkovic, S., Brown, L.E., and McKimm-Breschkin, J.L. (2019). Influenza virus neuraminidase structure and functions. *Front. Microbiol.* 10, 39. <https://doi.org/10.3389/fmicb.2019.00039>.
9. Gaymard, A., Le Briand, N., Frobert, E., Lina, B., and Escurat, V. (2016). Functional balance between neuraminidase and haemagglutinin in influenza viruses. *Clin. Microbiol. Infect.* 22, 975–983. <https://doi.org/10.1016/j.cmi.2016.07.007>.
10. Rogers, G.N., and Paulson, J.C. (1983). Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* 127, 361–373.
11. Couceiro, J.N., Paulson, J.C., and Baum, L.G. (1993). Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Res* 29, 155–165.
12. Rogers, G.N., Paulson, J.C., Daniels, R.S., Skehel, J.J., Wilson, I.A., and Wiley, D.C. (1983). Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* 304, 76–78. <https://doi.org/10.1038/304076a0>.
13. Peng, W., de Vries, R.P., Grant, O.C., Thompson, A.J., McBride, R., Tsogtbaatar, B., Lee, P.S., Razi, N., Wilson, I.A., Woods, R.J., and Paulson, J.C. (2017). Recent H3N2 viruses have evolved specificity for extended, branched human-type receptors, conferring potential for increased avidity. *Cell Host Microbe* 21, 23–34. <https://doi.org/10.1016/j.chom.2016.11.004>.
14. Walther, T., Karamanska, R., Chan, R.W.Y., Chan, M.C.W., Jia, N., Air, G., Hopton, C., Wong, M.P., Dell, A., Malik Peiris, J.S., et al. (2013). Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathog* 9, e1003223. <https://doi.org/10.1371/journal.ppat.1003223>.
15. Chu, V.C., and Whittaker, G.R. (2004). Influenza virus entry and infection require host cell N-linked glycoprotein. *Proc. Natl. Acad. Sci. USA* 101, 18153–18158. <https://doi.org/10.1073/pnas.0405172102>.
16. Narimatsu, Y., Joshi, H.J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y.-H., Schjoldager, K.T., Steentoft, C., et al. (2019). An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Mol. Cell* 75, 394–407.e5. <https://doi.org/10.1016/j.molcel.2019.05.017>.
17. Stevens, J., Blixt, O., Glaser, L., Taubenberger, J.K., Palese, P., Paulson, J.C., and Wilson, I.A. (2006). Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* 355, 1143–1155. <https://doi.org/10.1016/j.jmb.2005.11.002>.
18. Stevens, J., Blixt, O., Tumpey, T.M., Taubenberger, J.K., Paulson, J.C., and Wilson, I.A. (2006). Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* 312, 404–410. <https://doi.org/10.1126/science.1124513>.
19. Liao, H.-Y., Hsu, C.-H., Wang, S.-C., Liang, C.-H., Yen, H.-Y., Su, C.-Y., Chen, C.-H., Jan, J.-T., Ren, C.-T., Chen, C.-H., et al. (2010). Differential receptor binding affinities of influenza hemagglutinins on glycan arrays. *J. Am. Chem. Soc.* 132, 14849–14856. <https://doi.org/10.1021/ja104657b>.
20. Bradley, K.C., Jones, C.A., Tompkins, S.M., Tripp, R.A., Russell, R.J., Gramer, M.R., Heimbürg-Molinario, J., Smith, D.F., Cummings, R.D., and Steinhauer, D.A. (2011).

- Comparison of the receptor binding properties of contemporary swine isolates and early human pandemic H1N1 isolates (Novel 2009 H1N1). *Virology* 413, 169–182. <https://doi.org/10.1016/j.virol.2011.01.027>.
21. Gulati, S., Smith, D.F., Cummings, R.D., Couch, R.B., Griesemer, S.B., St George, K., Webster, R.G., and Air, G.M. (2013). Human H3N2 influenza viruses isolated from 1968 to 2012 show varying preference for receptor substructures with no apparent consequences for disease or spread. *PLoS One* 8, e66325. <https://doi.org/10.1371/journal.pone.0066325>.
  22. Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V., and McGuckin, M.A. (2008). Mucins in the mucosal barrier to infection. *Mucosal Immunol* 1, 183–197. <https://doi.org/10.1038/mi.2008.5>.
  23. Möckl, L. (2020). The emerging role of the mammalian glycocalyx in functional membrane organization and immune system regulation. *Front. Cell Dev. Biol.* 8, 253. <https://doi.org/10.3389/fcell.2020.00253>.
  24. Hang, H.C., and Bertozzi, C.R. (2005). The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg. Med. Chem.* 13, 5021–5034. <https://doi.org/10.1016/j.bmc.2005.04.085>.
  25. Mayr, J., Lau, K., Lai, J.C.C., Gagariarov, I.A., Shi, Y., McAtamney, S., Chan, R.W.Y., Nicholls, J., von Itzstein, M., and Haselhorst, T. (2018). Unravelling the role of O-glycans in influenza A virus infection. *Sci. Rep.* 8, 16382. <https://doi.org/10.1038/s41598-018-34175-3>.
  26. Delaveris, C.S., Webster, E.R., Banik, S.M., Boxer, S.G., and Bertozzi, C.R. (2020). Membrane-tethered mucin-like polypeptides sterically inhibit binding and slow fusion kinetics of influenza A virus. *Proc. Natl. Acad. Sci. USA* 117, 12643–12650.
  27. Button, B., Cai, L.-H., Ehre, C., Kesimer, M., Hill, D.B., Sheehan, J.K., Boucher, R.C., and Rubinstein, M. (2012). A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science* 337, 937–941.
  28. Byrd-Leotis, L., Liu, R., Bradley, K.C., Lasanajak, Y., Cummings, S.F., Song, X., Heimbürg-Molinario, J., Galloway, S.E., Culhane, M.R., Smith, D.F., et al. (2014). Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses. *Proc. Natl. Acad. Sci. USA* 111, E2241–E2250.
  29. McAuley, J.L., Corcilus, L., Tan, H.-X., Payne, R.J., McGuckin, M.A., and Brown, L.E. (2017). The cell surface mucin MUC1 limits the severity of influenza A virus infection. *Mucosal Immunol* 10, 1581–1593.
  30. Agard, N.J., Prescher, J.A., and Bertozzi, C.R. (2004). A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* 126, 15046–15047. <https://doi.org/10.1021/ja044996f>.
  31. Kaiser, Sr., E., Picart, F., Kubiak, T., Tam, J.P., and Merrifield, R.B. (1993). Selective deprotection of the N.alpha.-tert-butylloxycarbonyl group in solid phase peptide synthesis with chlorotrimethylsilane and phenol. *J. Org. Chem.* 58, 5167–5175.
  32. Nagao, M., Fujiwara, Y., Matsubara, T., Hoshino, Y., Sato, T., and Miura, Y. (2017). Design of glycopolymers carrying sialyl oligosaccharides for controlling the interaction with the influenza virus. *Biomacromolecules* 18, 4385–4392. <https://doi.org/10.1021/acs.biomac.7b01426>.
  33. Armstrong, G.D., Howard, L.A., and Pepler, M.S. (1988). Use of glycosyltransferases to restore pertussis toxin receptor activity to asialoagalactofetuin. *J. Biol. Chem.* 263, 8677–8684.
  34. Wright, C.S., and Jaeger, J. (1993). Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J. Biol. Chem.* 268, 620–638.
  35. Shibuya, N., Goldstein, I.J., Broekaert, W.F., Nsimba-Lubaki, M., Peeters, B., and Peumans, W.J. (1987). The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *J. Biol. Chem.* 262, 1596–1601.
  36. Sigal, G.B., Mammen, M., Dahmann, G., and Whitesides, G.M. (1996). Polyacrylamides bearing pendant alpha-sialoside groups strongly inhibit agglutination of erythrocytes by influenza virus: the strong inhibition reflects enhanced binding through cooperative polyvalent interactions. *J. Am. Chem. Soc.* 118, 3789–3800.
  37. Van Damme, E.J.M., Peumans, W.J., Barre, A., and Rougé, P. (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit. Rev. Plant Sci.* 17, 575–692.
  38. Maveyraud, L., Niwa, H., Guillet, V., Svergun, D.I., Konarev, P.V., Palmer, R.A., Peumans, W.J., Rougé, P., Van Damme, E.J., Reynolds, C.D., and Mourey, L. (2009). Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*. *Proteins* 75, 89–103.
  39. Raman, R., Tharakaraman, K., Sasisekharan, V., and Sasisekharan, R. (2016). Glycan-protein interactions in viral pathogenesis. *Curr. Opin. Struct. Biol.* 40, 153–162. <https://doi.org/10.1016/j.sbi.2016.10.003>.
  40. Meng, B., Marriott, A.C., and Dimmock, N.J. (2010). The receptor preference of influenza viruses. *Influenza Other Respir. Viruses* 4, 147–153. <https://doi.org/10.1111/j.1750-2659.2010.00130.x>.
  41. Koerner, I., Matrosovich, M.N., Haller, O., Staeheli, P., and Kochs, G. (2012). Altered receptor specificity and fusion activity of the haemagglutinin contribute to high virulence of a mouse-adapted influenza A virus. *J. Gen. Virol.* 93, 970–979. <https://doi.org/10.1099/vir.0.035782-0>.
  42. Di Iorio, D.D., Verheijden, M.L., van der Vries, E., Jonkheijm, P., and Huskens, J. (2019). Weak multivalent binding of influenza hemagglutinin nanoparticles at a sialoglycan-functionalized supported lipid bilayer. *ACS Nano* 13, 3413–3423.
  43. Zanin, M., Baviskar, P., Webster, R., and Webby, R. (2016). The interaction between respiratory pathogens and mucus. *Cell Host Microbe* 19, 159–168.
  44. Uddin, S., Khan, A., Hossain, M.E., and Moni, M.A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Med. Inform. Decis. Mak.* 19, 281.
  45. Suzuki, Y., Nagao, Y., Kato, H., Suzuki, T., Matsumoto, M., and Murayama, J. (1987). The hemagglutinins of the human influenza viruses A and B recognize different receptor microdomains. *Biochim. Biophys. Acta* 903, 417–424.
  46. Ito, T., Suzuki, Y., Takada, A., Kawamoto, A., Otsuki, K., Masuda, H., Yamada, M., Suzuki, T., Kida, H., and Kawaoaka, Y. (1997). Differences in sialic acid-galactose linkages in the chicken egg amnion and allantois influence human influenza virus receptor specificity and variant selection. *J. Virol.* 71, 3357–3362.
  47. Yao, Y., Liu, Y., Yu, Y., Xu, H., Lv, W., Li, Z., and Chen, X. (2013). K-SVM: an effective SVM algorithm based on K-means clustering. *J. Comput.* 8, 2632–2639.
  48. Hudson, M.J., Stamp, G.W., Chaudhary, K.S., Hewitt, R., Stubbs, A.P., Abel, P.D., and Lalani, E.N. (2001). Human MUC1 mucin: a potent glandular morphogen. *J. Pathol.* 194, 373–383.
  49. Inkster, M.D., Hinshaw, V.S., and Schulze, I.T. (1993). The hemagglutinins of duck and human H1 influenza viruses differ in sequence conservation and in glycosylation. *J. Virol.* 67, 7436–7443.
  50. Li, J., Liu, S., Gao, Y., Tian, S., Yang, Y., and Ma, N. (2021). Comparison of N-linked glycosylation on hemagglutinins derived from chicken embryos and MDCK cells: a case of the production of a trivalent seasonal influenza vaccine. *Appl. Microbiol. Biotechnol.* 105, 3559–3572.
  51. Cohen, M., Fisher, C.J., Huang, M.L., Lindsay, L.L., Plancarte, M., Boyce, W.M., Godula, K., and Gagneux, P. (2016). Capture and characterization of influenza A virus from primary samples using glycan bead arrays. *Virology* 493, 128–135.
  52. Byrd-Leotis, L., Jia, N., Dutta, S., Trost, J.F., Gao, C., Cummings, S.F., Brailke, T., Müller-Loennies, S., Heimbürg-Molinario, J., Steinhauer, D.A., and Cummings, R.D. (2019). Influenza binds phosphorylated glycans from human lung. *Sci. Adv.* 5, eaav2554.
  53. Nason, R., Büll, C., Konstantinidi, A., Sun, L., Ye, Z., Halim, A., Du, W., Sørensen, D.M., Durbesson, F., Furukawa, S., et al. (2021). Display of the human mucinome with defined O-glycans by gene engineered cells. *Nat. Commun.* 12, 4070.
  54. Huang, M.L., Smith, R.A.A., Trieger, G.W., and Godula, K. (2014). Glycocalyx remodeling with proteoglycan mimetics promotes neural specification in embryonic stem cells. *J. Am. Chem. Soc.* 136, 10565–10568.
  55. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.