



Glycosylation of the viral attachment protein of avian coronavirus is essential for host cell and receptor binding

Received for publication, January 18, 2019, and in revised form, March 1, 2019. Published, Papers in Press, March 22, 2019, DOI 10.1074/jbc.RA119.007532

Lisa M. Parsons[‡], Kim M. Bouwman[§],  Hugo Azurmendi[‡], Robert P. de Vries[¶], John F. Cipollo^{‡1}, and Monique H. Verheije[§]

From the [‡]Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland 20993, the [§]Division of Pathology, Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands, and the [¶]Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3512 JE Utrecht, The Netherlands

Edited by Gerald W. Hart

Avian coronaviruses, including infectious bronchitis virus (IBV), are important respiratory pathogens of poultry. The heavily glycosylated IBV spike protein is responsible for binding to host tissues. Glycosylation sites in the spike protein are highly conserved across viral genotypes, suggesting an important role for this modification in the virus life cycle. Here, we analyzed the *N*-glycosylation of the receptor-binding domain (RBD) of IBV strain M41 spike protein and assessed the role of this modification in host receptor binding. Ten single Asn-to-Ala substitutions at the predicted *N*-glycosylation sites of the M41-RBD were evaluated along with two control Val-to-Ala substitutions. CD analysis revealed that the secondary structure of all variants was retained compared with the unmodified M41-RBD construct. Six of the 10 glycosylation variants lost binding to chicken trachea tissue and an ELISA-presented α 2,3-linked sialic acid oligosaccharide ligand. LC/MS^E glycomics analysis revealed that glycosylation sites have specific proportions of *N*-glycan subtypes. Overall, the glycosylation patterns of most variant RBDs were highly similar to those of the unmodified M41-RBD construct. *In silico* docking experiments with the recently published cryo-EM structure of the M41 IBV spike protein and our glycosylation results revealed a potential ligand receptor site that is ringed by four glycosylation sites that dramatically impact ligand binding. Combined with the results of previous array studies, the glycosylation and mutational analyses presented here suggest a unique glycosylation-dependent binding modality for the M41 spike protein.

Avian coronaviruses of poultry cause significant disease with subsequent economic losses in several commercially farmed bird species. Avian infectious bronchitis virus (IBV)² is a gam-

This work was supported by The Netherlands Organisation for Scientific Research (NWO) with VENI Grants (to R. P. d.V.) and MEERVOUD (to M. H. V.) and National Institutes of Health, Health and Human Services, Food and Drug Administration Grant Z01 BJ02044-04 LBP (to L. M. P. and J. F. C.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains Figs. S1–S4 and Tables S1 and S2.

¹To whom correspondence should be addressed. E-mail: john.cipollo@fda.hhs.gov.

²The abbreviations used are: IBV, infectious bronchitis virus; RBD, receptor-binding domain; PDB, Protein Data Bank; PNGase, peptide:*N*-glycosidase;

macoronavirus that predominantly affects domestic fowl, primarily chickens (*Gallus gallus*). The virus initially infects upper airway epithelium tissues, and depending on the IBV strain, disease outcomes range from mild respiratory disease to kidney failure and death (1).

The viral envelope of IBV contains the highly-glycosylated spike (S) protein that is post-translationally cleaved into two domains, S1 and S2. This S glycoprotein is the major adhesion molecule of the virus. It is a class I viral fusion protein, in which the variable S1 domain is involved in host cell receptor binding, and the more conserved S2 domain mediates the fusion of the virion with the cellular membrane (2, 3). The role of spike in host cell attachment and the induction of protective immunity has been reviewed (4). The spike protein monomer is a transmembrane glycoprotein with a molecular mass of 128 kDa before glycosylation (3). A cleavable N-terminal signal peptide (5) directs the S protein toward the endoplasmic reticulum (ER), where it is extensively modified with *N*-linked glycosylation (6, 7). After glycosylation in the ER, the monomers oligomerize to form trimers (6–9).

The N-terminal 253 amino acids of S1 were shown to encompass the receptor-binding domain (RBD) of IBV strain M41 (10), which interacts with sialyl- α 2,3-substituted glycans present on the host's cell surface (11, 12). Ten *N*-linked glycosylation sites are predicted to exist on the M41-RBD (5), of which most are highly conserved (Fig. S1). It is interesting that 8 of the 10 sites are 95–100% conserved. Sites Asn-33 and Asn-59 were less conserved at 80 and 25%. However, each had a nearby alternative site that was also highly conserved. Alternative site Asn-36 was conserved 50% of the time, and one or both Asn-33 and Asn-36 was present in 94% of the sequences. Site Asn-57 was conserved at 73%. In 97% of the sequences, either Asn-59 or Asn-57 was present but never together. Therefore, all 10 sites, including the alternatives, likely serve important functions.

The *N*-glycosylation of viral glycoproteins is known to modulate the ability of viruses to infect host cells and to be recognized by the host's immune system (13). Recently, Zheng *et al.* (14) studied extracted spike proteins and mutant viruses with Asn-to-Asp (asparagine to aspartate) and Asn-to-Gln (aspara-

ANOVA, analysis of variance; HILIC, hydrophilic interaction liquid chromatography; ACN, acetonitrile; ER, endoplasmic reticulum; AUNA, *A. ureafaciens* neuraminidase.

Table 1
Techniques used in this paper

Material	Samples	Technique	Outcome
Protein ^a	All ^b	CD	Secondary structure and stability
Protein	All	Tissue histochemistry	Binding affinity to tissues
Protein	M41, Asn-to-Ala variants	ELISA	Binding affinity to sialic acid
Released glycans	All	MALDI-TOF MS	Percent abundance of glycoforms
Sugar-free peptides	M41, N59A, N145A	LC/MS ^E	Site occupancy
Glycopeptides	M41, N59A, N145A	LC/MS ^E	Assignment of site-specific glycoforms
Protein structure	M41	<i>In silico</i> docking	Potential binding sites

^a Recombinant protein consists of the first 253 residues of the RBD of the IBV M41 spike protein, a GCN4 trimerization motif, and a Strep-tag.

^b M41 (unmodified), all the Asn-to-Ala variants, and two nonglycosylation variants, V57A and V58A are shown.

gine to glutamine) mutations at 13 predicted glycosylation sites in the S protein of the Beaudette IBV strain (14). Their results indicate that glycosylation at some sites on the Beaudette S1-RBD was important for viral fusion and infectivity, which may include host recognition. However, the Beaudette strain is a cell culture-adapted strain, is nonvirulent in chickens (15), and does not bind chicken tissues known to be important for infectivity (11), making it difficult to extrapolate these results to clinically relevant IBVs.

To characterize and assess the role that glycosylation plays when interacting with host tissues through the RBD of pathogenic IBV strain M41, we used a combination of molecular and analytical techniques, including histochemistry, ELISA, circular dichroism (CD), MS, and docking analyses as listed in Table 1. Systematic deletion of each glycosylation site and histochemical analysis of each variant revealed which of the 10 glycosylation sites affect the binding of IBV S protein to host epithelial tissue. Site occupancy analysis by LC/MS^E indicated that at least 9 of 10 predicted *N*-glycosylation sites in the M41-RBD domain are glycosylated. Analysis of site occupancy and signature *N*-glycan patterns at each site in combination with single glycosylation site deletions provided insight toward the biological relevance of each of those sites in binding to host tissue receptors. Overall, our data confirm that *N*-glycosylation plays a critical and likely unique role in binding of the IBV spike domain to its host tissue receptors.

Results

Gel electrophoresis and CD analysis indicate that M41-RBD and glycosylation variants are similarly expressed, folded, and stable

To analyze the role of glycosylation of M41-RBD in receptor binding, missense mutants (Asn-to-Ala) were generated on a site-by-site basis at each of the predicted *N*-glycosylation sites. Recombinantly produced glycovariant RBD proteins migrated with the same electrophoretic mobility as unmodified M41-RBD (Fig. 1). The RBD proteins were evaluated by CD spectroscopy to assess similarity to the WT secondary structure. WT M41-RBD, all 10 glycosylation-site variants, and two nonglycosylation variants, V57A and V58A, were analyzed for secondary structure differences at 25 °C. Thermal melts were performed on each construct from 25 to 95 °C followed by full scans collected at 95 °C and again at 25 °C after the melt. Overlays of all the CD spectra can be found in Fig. S2. Visually, all spectra at all temperatures follow the same curve. The N85A spectra were generated at higher protein concentrations but aligned well to CD spectra of all other variants when normal-

ized to the percent of maximum signal. Likewise, all the proteins had analogous broad melting curves suggesting the proteins were similarly stable. Protein folding was reversible for all proteins, with comparable recovery rates (see CD-25 °C-after-melt-normalized in Fig. S2). Dichroweb (16) was used to calculate the percent of α -helix, β -strand, turn, and unordered portions of the protein in the initial 25 °C spectra to estimate secondary structure differences between the proteins (Fig. 2). The percent of α -helix varied with the extremes being unmodified RBD and N145A. N145A exhibited $19.5 \pm 0.3\%$ α -helix character as compared with WT, which has $31.6 \pm 2.4\%$. Interestingly, N145A gave a very strong signal in the histochemical assay (Fig. 3A) and had the most notably different released glycans' signature compared with the other constructs. We conclude that all proteins maintained a very similar structure and therefore suggest that single *N*-glycosylation sites are by themselves not indispensable for protein folding or stability.

Six glycosylation variants abrogate binding to host tissue and sialic acid

Because we established that all variant M41-RBD proteins are folded, we investigated their abilities to bind tissue receptors. Recombinant proteins were incubated with chicken trachea tissue sections and examined by histochemical analysis. N145A, N219A, N229A, N246A, V57A, and V58A bound ciliated epithelial cells of the chicken trachea with similar staining intensity as the unmodified RBD with the most intense staining associated with the N145A construct (Fig. 3A). In contrast, binding of constructs N33A, N59A, N85A, N126A, N160A, and N194A to trachea tissue was not detectable. Removal of sialic acids by treatment of the trachea tissues with *Arthrobacter ureafaciens* neuraminidase (AUNA) abrogated binding of all constructs as shown in Fig. S4. These results demonstrate that glycosylation on the RBD affects binding to sialyl ligands on chicken trachea tissue.

The interaction of the variants with Neu5Ac(α 2-3)Gal (β 1-3)GlcNAc, a previously established ligand for M41 (11), was assayed by ELISA. N145A, N219A, N229A, and N246A variants were able to bind the ligand in a concentration-dependent manner (Fig. 3B) like unmodified RBD. Binding affinities of N33A, N59A, N85A, N126A, N160A, and N194A were significantly reduced compared with unmodified RBD and comparable with that of a negative control protein, the S1 of turkey coronavirus, with specificity for nonsialylated diLacNAc glycans (17). Fig. 3C shows the ELISA absorbance at the 75 nmol of ligand concentration for each construct. No significant difference was observed for variants N145A, N219A, N229A, and

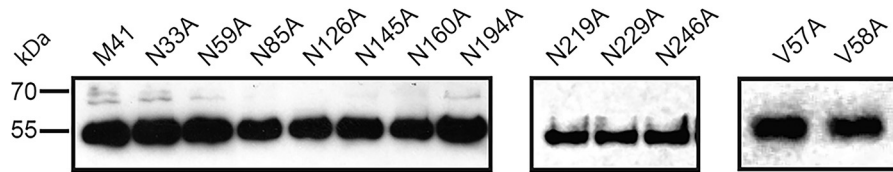


Figure 1. Western blotting verifying production of M41 RBD proteins. Recombinant viral proteins were produced by transfection of HEK293T cells. The soluble proteins were purified from the supernatant using Strep-Tactin beads and analyzed by Western blotting using a Strep-Tactin HRP antibody.

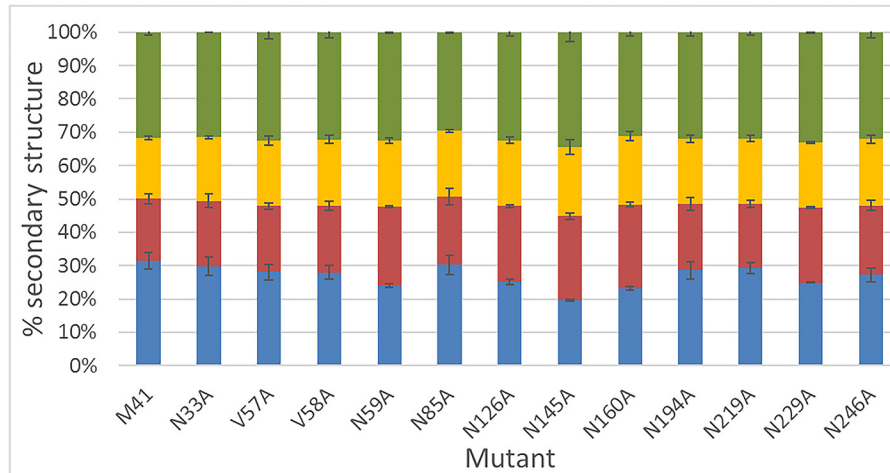


Figure 2. Calculated secondary structure for each variant based on CD data. Each bar represents the average results from three algorithms in Dichroweb. Standard deviations are indicated at the top of each color. From the bottom, the bar segments represent α -helix (blue), β -strand (red), turns (yellow), and unordered (green).

N246A compared with unmodified RBD (shown in *dark gray bars* in Fig. 3C). All other variants (shown in *light gray bars* in Fig. 3C) demonstrated significantly lower affinity for the receptor, consistent with histochemistry and ligand titration plot results.

Overall glycosylation of nonbinding variants is similar to M41-RBD

Six of the 10 single glycosylation site variants lost the ability to bind ligand. To investigate whether global changes in glycosylation may have affected binding, we analyzed release glycans from each protein. Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS) analysis of enzymatically released and permethylated glycans allows for semi-quantitative analysis of glycan compositions. The method is particularly useful for samples containing sialylated glycans because they are stabilized by permethylation. The percent abundances of glycans identified in each sample are shown in Fig. 4.

The majority of the Asn-to-Ala variants, as well as the V57A and V58A control variants, had similar MALDI-TOF-MS permethylation profiles (Fig. 4). Over 100 glycan compositions were identified ranging from high-mannose glycans to large complex ones. Nearly half of the glycans contained at least one and up to three sialic acid molecules in all samples. The most intense glycoforms clustered in five groups with increasing amounts of complexity as reflected by the number of *N*-acetyl glucosamines (HexNAcs). These include high-mannose, complex, and hybrid forms as follows: I, Hex₅₋₉HexNAc₂ (high mannose); II, NeuAc₀₋₁Hex₅₋₆dHex₀₋₁HexNAc₃ (complex and hybrid); III, NeuAc₀₋₂Hex₅dHex₁HexNAc₄(complex); IV,

NeuAc₀₋₁Hex₆dHex₁HexNAc₅ (complex); and V, NeuAc₂Hex₇dHex₁HexNAc₆ (complex). High-mannose glycans were less abundant in unmodified M41 than in variant RBDs. The N194A, N219A, and N229A variants contained diminished amounts of the group V high-mass complex glycans. The N145A variant was the most atypical with less defined clustering in the common clustering regions of the spectrum and higher abundances in spectral regions where compositions had less Hex and more HexNAc overall. For instance, cluster IV was shifted from glycans with 6 hexoses (NeuAc₀₋₁Hex₆dHex₁HexNAc₅) to glycoforms with 3–4 hexoses (NeuAc₀₋₁Hex₃₋₄dHex₁HexNAc₅). More abundance was observed in regions containing 6 HexNAc residues (NeuAc₀₋₂Hex₃₋₆dHex₁HexNAc₆). To better understand the difference between N145A and the other constructs, we calculated the monosaccharide percent mass and average mass for each construct. The average mass percent for glycans across all released glycan pools was Hex (45.8%), HexNAc (42.0%), dHex (5.0%), and NeuAc (7.2%). The N145A construct had the lowest amount of Hex (38.6%) and the highest amounts of HexNAc (46.0%) and NeuAc (9.8%). The former two were 2 S.D. or greater from the mean (see Table S2). This indicates that the N145A construct likely had shorter, more branched, and more highly-charged glycans on average than the other constructs. Two other variants had values more than 2 S.D. from the mean. N229A (normal binding) was most abundant in Hex (53.6%) and least abundant in HexNAc (37.5%) and dHex (3.8%), probably due to its higher high-mannose content. N246A (normal binding) had the lowest amount of NeuAc (3.6%). This is perhaps a reflection of the missing sugars in this variant because site Asn-246 in

Avian coronavirus glycosylation

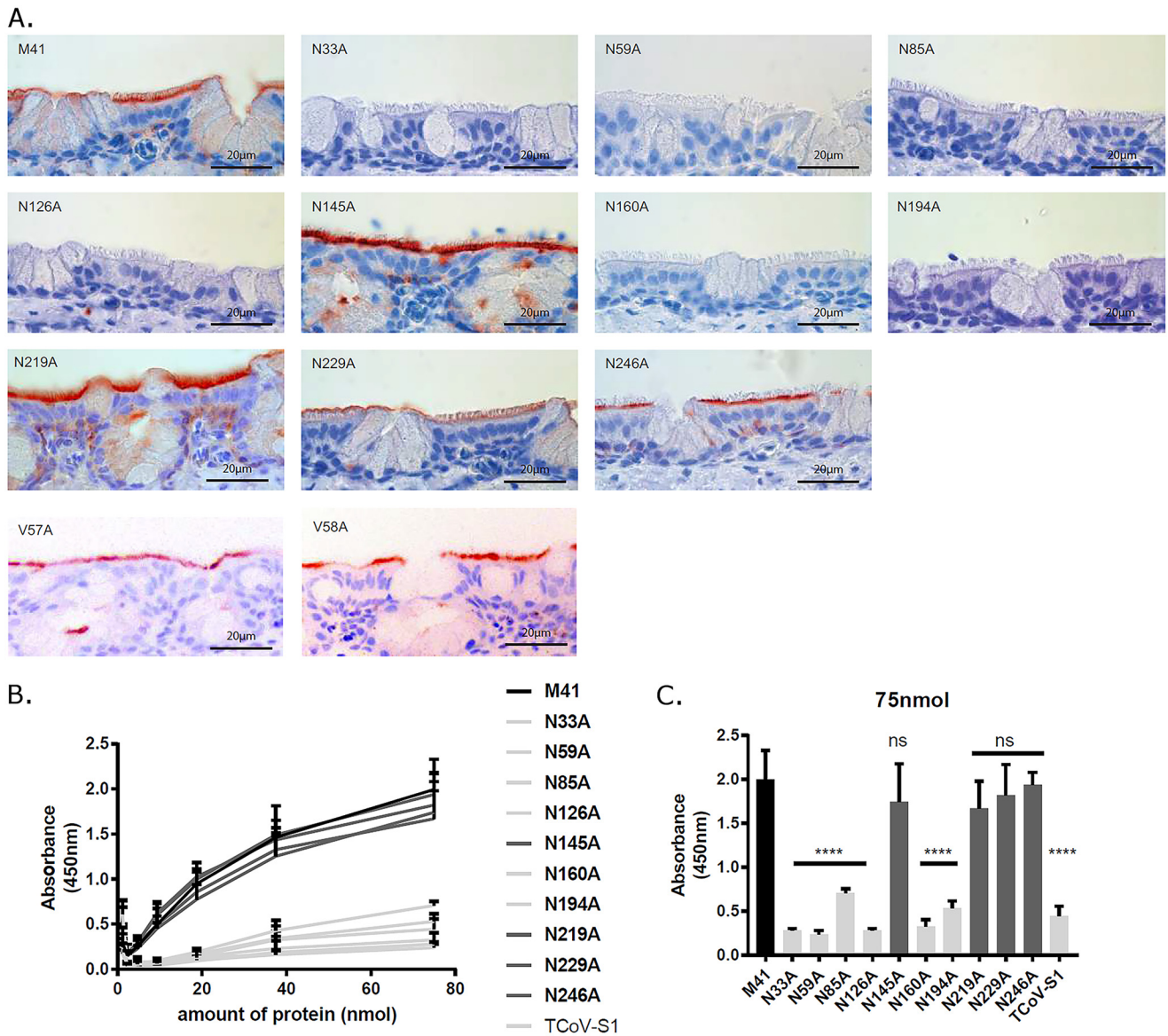


Figure 3. Tissue-binding assay and ELISAs. Histochemical assays of recombinant unmodified M41-RBD and single Asn-to-Ala and Val-to-Ala glycosylation variants to trachea tissue (A) and ELISA-presented Neu5Ac α 2-3Gal β 1-3GlcNAc (B and C). B, concentration dependence of binding. C, absorbance for each protein at the 75-nmol concentration. Two-way ANOVA showed significantly less binding by variant N33A, N59A, N85A, N126A, N160A, and N194A RBD proteins compared with unmodified RBD (compare *light gray bars* (variant) to unmodified (*black bar*)). No significant (*n.s.*) difference was observed for variants with *dark gray bars*. Data points are averaged from three separate assays. ****, $p < 0.0001$.

other variants was populated with many sialylated glycoforms based on site-specific analysis (Table S1).

Glycosylation and site occupancy were similar between M41-RBD, N59A, and N145A

To assess the differences in glycosylation on a site-to-site basis, glycopeptide LC/MS analysis was carried out on unmodified M41 and two single glycosylation site variants, N59A and N145A, that represented a nonbinder and a binder of trachea tissue, respectively. M41-RBD had 10 predicted glycosylation sites, whereas the variant RBDs had nine each. N145A was also of specific interest due to the unique glycosylation pattern observed in its free glycan profile. As cleavage with trypsin alone resulted in glycopeptides with more than one glycosylation site, we also analyzed glycopeptides after an additional treatment with chymotrypsin, which resulted in one glycosite

per peptide, the identification of more glycopeptides, and decreased ambiguity concerning glycosylation site assignment.

Although a protein may contain the sequence (NX(S/T)), where *N*-glycosylation is known to occur, it may not actually be glycosylated, or it may be glycosylated only part of the time. Potential glycosylation sites, their predicted glycosylation state, and their measured site occupancy are shown in Table 2. Of the 10 glycosites, all but Asn-246 were predicted to be glycosylated (occupied) based on NetNGlyc analysis (<http://www.cbs.dtu.dk/services/NetNGlyc-1.0/>).³ Percent occupancy was analyzed by LC/MS; however, a poor signal was obtained for the Asn-219 site in M41 and N59A, and therefore, occupancies were not calculated. All other sites were estimated to be occu-

³ Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

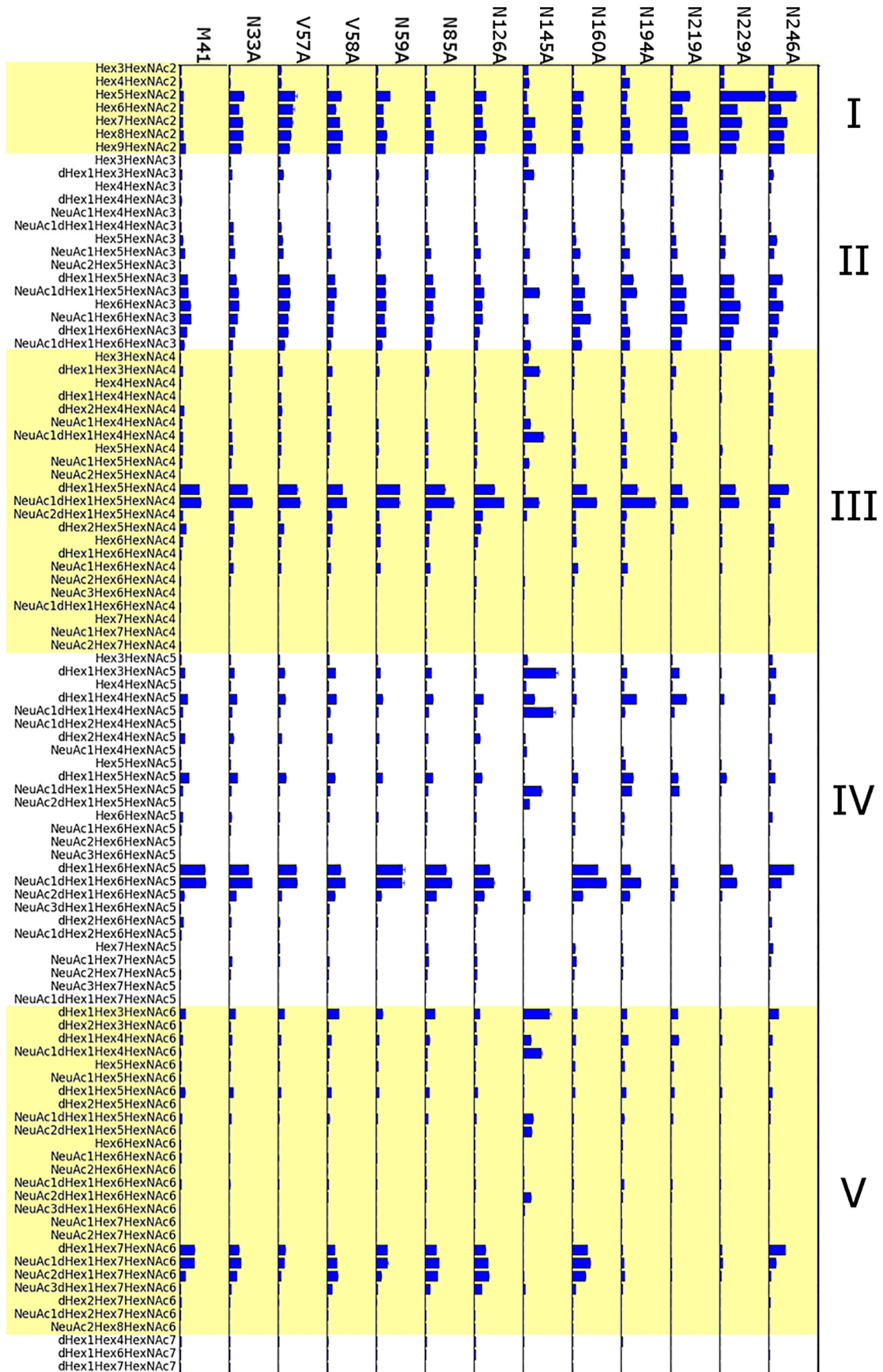


Table 2

Potential glycosylation sites based on sequence, predicted glycosylation sites by NetNGlyc (www.cbs.dtu.dk/services/NetNGlyc),³ and site occupancy measured by LC/MS^E

Potential glycosylation sites		NetNGlyc predictions		Site occupancy ^c		
Position ^a	Sequence	Potential ^b	Result	M41	N59A	N145A
33	NISS	0.7343	++	100 ^d	100 ^d	100 ^d
59	NASS	0.6391	+	99.1 ± 0.2	NA ^e	ND ^{e,f}
85	NFSD	0.6962	+	100 ^d	93.8 ± 0.5	100 ^d
126	NLTV	0.7729	+++	97.3 ± 0.4	98.6 ± 0.4	61.4 ± 2.2
145	NLTS	0.6099	++	97.3 ± 0.2	98.3 ± 0.2	NA ^e
160	NETT	0.5049	+	94.3 ± 0.2	90.6 ± 0.3	96.82 ± 0.04
194	NGTA	0.6832	++	89.2 ± 0.4	91.8 ± 0.5	92.3 ± 0.1
219	NFSD	0.5281	+	ND ^g	ND ^g	100 ^d
229	NSSL	0.5189	+	99.4 ± 0.2	100 ^d	100 ^d
246	NTTF	0.4726	- ^h	94.0 ± 3.4	96.6 ± 0.1	79.4 ± 2.5

^a Sequence position is based on the mature protein.

^b The higher the NetNGlyc potential, the more likely it is to be glycosylated.

^c Average percentages and standard deviations are calculated from three separate LC/MS^E injections.

^d Where percent occupancy = 100, the intensity of the never-glycosylated peptide was too low to detect.

^e Sites missing in the glycosylation variants are noted with NA. Not determined is noted as ND.

^f Both glycosylated and nonglycosylated forms were detected, but incomplete cleavage and low signal intensity precluded accurate approximation of occupancy.

^g Masses matching the spontaneously deaminated and de-glycosylated peptide were not found.

^h This site is not likely to be glycosylated.

pied at 89% or greater in M41 and N59A. The N145A variant exhibited site occupancy at all expected sites, including Asn-219, although signal intensity at that site was low. Two sites had much lower occupancy in N145A as compared with the other samples. Site Asn-126 dropped to 61% occupancy and site Asn-246 to 79% occupancy compared with nearly complete occupancy in the N59A and M41 proteins. Overall site occupancy was high for all sites. The difficulty in detecting some of the peptides, particularly Asn-219, may be due to hydrophobicity. Ionization is partially driven by hydrophobicity, and Asn-219 only had 20% hydrophobic character after the two digestions, which may, in part, explain its low detectability. By comparison, glycopeptides containing Asn-85, Asn-145, and Asn-160 were short and between 21 and 33% hydrophobicity, whereas glycopeptides containing other sites had predicted hydrophobicity ranging from 37 to 61% and tended to produce higher intensity spectra.

Glycoform relative abundances at each site are listed in Table S1. Fig. 5 shows the location of each glycosylation site on the RBD of M41. Overall compositions at each site were similar in charge and size across the three constructs. A representative glycan is shown at each site based on peak intensity. The N145A construct had glycoforms like those identified by MALDI-TOF MS with more HexNAc and fewer Hex compared with M41 and N59A.

Fewer overall glycan compositions were detected on glycopeptides by LC/MS compared with the free glycans observed by MALDI-TOF MS (63 *versus* 100 compositions). This can be expected because the technology of instrumentation used and the physiochemical characteristics of permethylated glycans and glycopeptides differ significantly. The forms detected overlapped between the two analyses.

Docking results are dependent on glycosylation status of the M41-RBD protein

During our investigation, the first structure of the M41 spike protein was solved using electron microscopy (EM) (18). Mapping the glycosylation sites onto the structure did not lead to a clear understanding of how the mutations affect binding. Although EM structural resolution is limited, and the precise coordinates for the attached glycans are not known, an attempt was made to dock a series of potentially sialylated ligands to a glycan-stripped structure of the RBD and a structure that was populated with glycans based on our data. The glycan chosen for each site on the RBD was based on the predominant glycans identified at each site by LC/MS (see Fig. 5).

Seventeen oligosaccharide ligands were chosen based on a previous glycan array study of M41 (11) and ELISA data (this work). Both strong and weak binders were selected (Fig. 6). Each ligand was docked 20 times against both the sugar-stripped and *in silico* glycosylated M41-RBD coordinates. There was no statistically significant difference between the docked binding energies of ligands that did and did not bind on the array. All oligosaccharide ligands, except for 1, 3, 9, 13, 15, and 17, docked seven or more times to one or more of the four sites on the M41 sugar-stripped structure with no clear pattern differentiating between them (Fig. 6). In the sugar-stripped structure, all binding occurred at sites A and B. Site A is under the galectin fold near site Asn-194, and site B encompasses Asn-85 and Asn-59. All three glycosylation sites are required for binding to trachea tissue. The docking pattern changed dramatically when glycans were modeled onto the structure. The most dramatic change was seen at site D where eight ligands bound seven or more times, whereas interactions at all other sites decreased. There were no binders at site A, only two at site

Figure 4. Free glycans identified by MALDI-TOF analysis. Data for M41 and variants are arranged in columns. Assigned glycans are on the y axis. Blue bars represent the average percent abundance across three measurements. Standard deviation is indicated with black lines on top of the bars. Glycan compositions are arranged by increasing complexity, starting with high mannose (*i.e.* Hex₅HexNAc₂) at the top and ending with the larger complex forms at the bottom. Yellow and white shading groups indicate glycan compositions with increasing numbers of HexNAcs moving from top to bottom. Abbreviations are hexose (Hex), GlcNAc (HexNAc), deoxyhexose (dHex), and sialic acid (NeuAc).

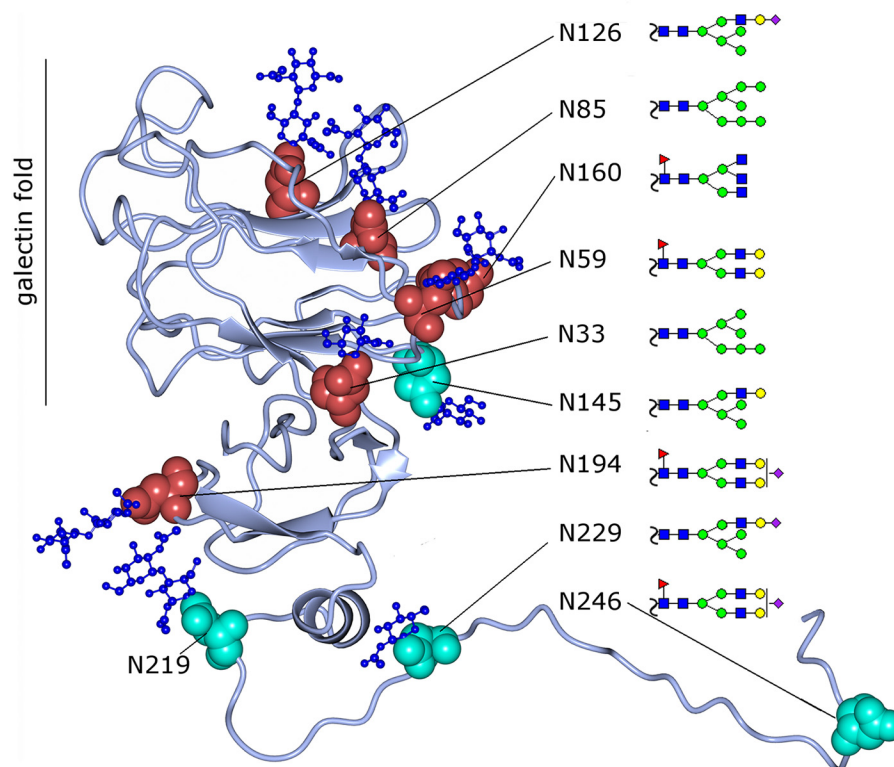


Figure 5. Site-specific glycosylation of M41, N59A, and N145A. The S1–N-terminal receptor binding domain residues 21–268 from PDB entry 6cv0 is represented as *gray ribbons*. The asparagines of glycosylation sites that could still bind trachea tissue after mutation to alanine are in *cyan*, and those that could not are in *dark red*. GlcNAc residues from the structure are *dark blue balls and sticks*. The most predominant glycan for each site across all three constructs is shown to the *right*. Glycoforms shown on the *right* are based on our data, and inferred structural detail is based on accepted knowledge of the cell type used in protein production. Monosaccharides are represented as follows: mannose (*green circles*); galactose (*yellow circles*); GlcNAc (*blue squares*); fucose (*red triangles*); and sialic acid (*purple diamonds*). Numbering of the sites is based on the mature sequence. The figure was made with CCP4MG (38) and GIMP.

C (3 and 16) and three at site B (6, 9, and 17). All of the ligand oligosaccharides that docked at site D were sialylated, consistent with ligands identified by array and ELISA. No control ligand (1 and 2 uncharged; 3 and 4 KDN-charged) bound at site D. The interaction at site D involved both sugar–protein and sugar–sugar contacts, and in some docking runs, the interaction was completely sugar–sugar. Site D is in the center of a circle of glycosylation sites that showed altered binding profiles when mutated; N59A, N85A, and N160A lost the ability to bind, whereas N145A gave a very strong signal in the histochemical assay.

Of note, no ligands docked in the site at the top of the galectin fold where many structural homologs of M41 are thought to bind sugars, such as the bovine coronavirus RBD (19). For comparison, we docked Neu5Ac(α 2–6)Gal(β 1–3)GlcNAc (β -OMe) against the crystal structure of the bovine RBD. Twenty five of 25 times the glycan docked in the proposed binding site at the top of the galectin fold in the negatively-charged area of the bovine RBD control near Asn-198 (Fig. 7B).

Discussion

Previously, we established that the IBV M41 S1 protein binds sialic acid-substituted glycoconjugate ligands in chicken trachea and lung tissue (11). Intriguingly, the M41 RBD is highly-glycosylated with 10 potential glycosylation sites, and glycosylation appears to be necessary for binding to host tissues because treating the protein with a neuraminidase diminishes binding

(11). This study extends our investigation toward determining the role of glycosylation in the function of the RBD, which encompasses the N-terminal region of the native protein. Each of the potential glycosylation sites was individually ablated, and each construct was examined for its ability to bind tissue and an ELISA-presented ligand. In addition, the global glycosylation profile of every construct was surveyed, and glycosylation of three representative constructs was examined on a site-specific basis.

Six of the 10 glycosylation sites in the RBD domain of IBV M41 were essential for binding to chicken trachea tissue and an ELISA-presented sialylated oligosaccharide ligand. CD analysis demonstrated that both secondary structure and stability were similar across all the RBD constructs indicating the proper fold was likely retained for all. Globally, percent abundances of sialylated glycans differed across mutants, but the differences were not associated with loss of binding. For example, 51 and 20% of the glycans in binding mutants N145A and N246A, respectively, and 46 and 51% of the glycans in the nonbinders N126A and N160A, respectively, were sialylated (summed from Fig. 4). By comparison, 40% of the glycans in the unmodified RBD construct were sialylated. On a site-specific basis, some glycosylation sites had more sialylation than others (Table S1). On average, each of glycosites Asn-126, Asn-194, Asn-229, and Asn-246 were sialylated at least 50% of the time. Sites Asn-229 and Asn-246 were in the less-ordered region of the protein

Avian coronavirus glycosylation

number	A	B	C	D	A	B	C	D	Docked oligos	IUPAC-condensed nomenclature	Array score ¹
1										GlcNAc(b1-3)Gal(b1-4)GlcNAc(b-Ome)	28 ± 35
2	x									GlcNAc(b1-4)Gal(b1-4)GlcNAc(b-Ome)	24 ± 9
3								x		KDN(a2-3)Gal(b1-3)GlcNAc(b-Ome)	NA
4	x									KDN(a2-3)Gal(b1-4)GlcNAc(b-Ome)	0 ± 3
5	x	x								Neu5Ac(a2-3)Gal(b1-3)GlcNAc(a-Ome)	16 ± 17
6		x				x		x		Neu5Ac(a2-3)[Neu5Ac(a2-6)]GalNAc(a-Ome)	33 ± 14
7	x							x		Neu5Ac(a2-3)GalNAc(a-Ome)	12 ± 2
8	x							x		Neu5Ac(a2-3)Gal(b1-3)[Fuc(a1-4)]GlcNAc(b-Ome)	363 ± 20
9						x				Neu5Ac(a2-3)Gal(b1-3)[Neu5Ac(a2-3)Gal(b1-4)]GlcNAc(b-Ome)	3308 ± 257
10	x							x		Neu5Ac(a2-3)Gal(b-Ome)	32 ± 24
11		x						x		Neu5Ac(a2-3)Gal(b1-3)GlcNAc(b-Ome)	285 ± 101
12	x							x		Neu5Ac(a2-3)Gal(b1-4)[Fuc(a1-3)]GlcNAc(b1-3)Gal(b-Ome)	5 ± 7
13										Neu5Ac(a2-3)Gal(b1-4)GlcNAc(b-Ome)	6 ± 5
14	x	x								Neu5Ac(a2-3)Gal(b1-4)Glc(b-Ome)	10 ± 4
15								x		Neu5Ac(a2-6)Gal(b1-4)GlcNAc(b-Ome)	15 ± 14
16	x							x		Neu5Ac(a2-6)Gal(b1-4)Glc(b-Ome)	4 ± 3
17						x				Neu5Ac(a2-8)Neu5Ac(a2-8)Neu5Ac(b-Ome)	11 ± 6

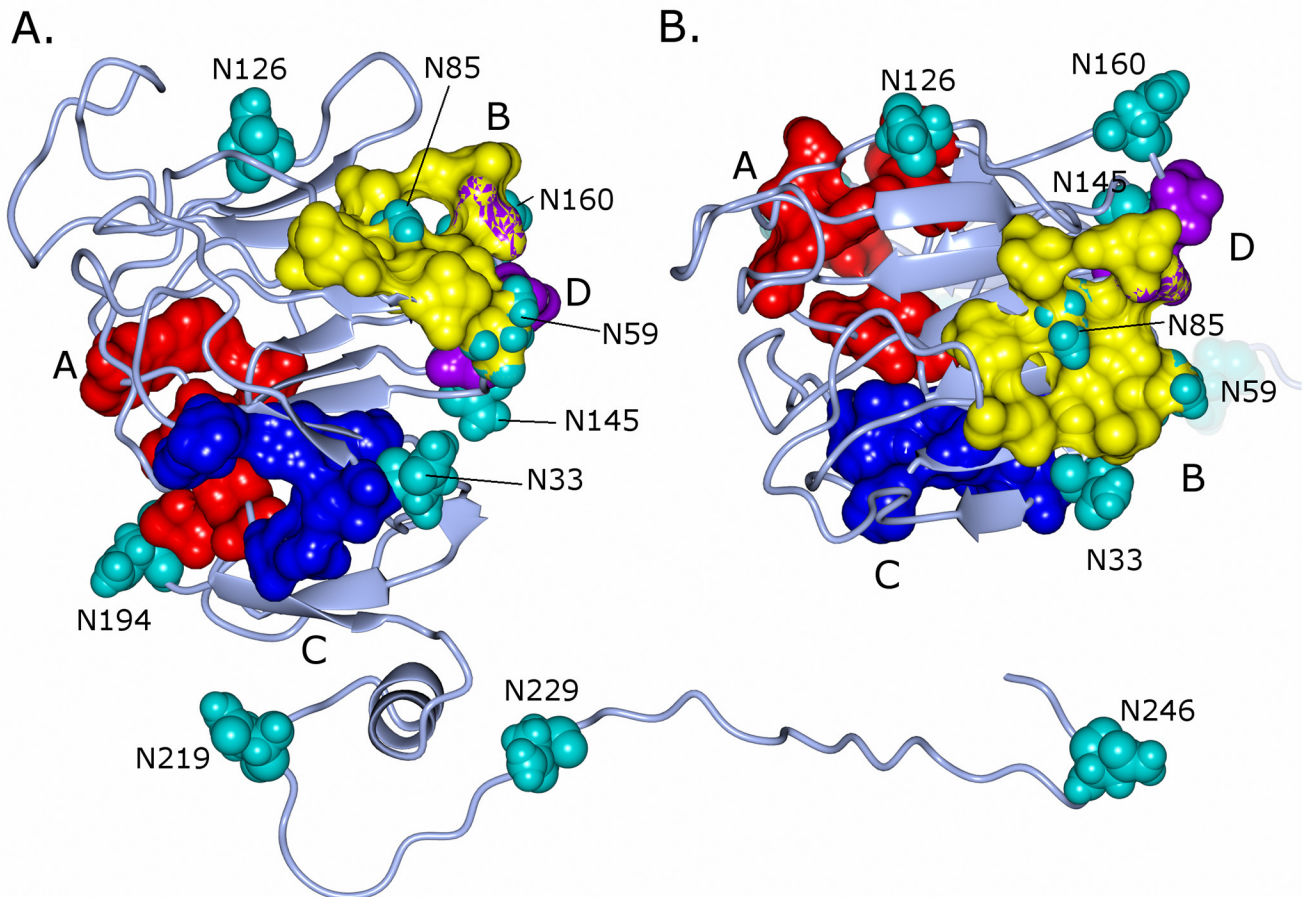


Figure 6. Docking results. Top, list of all oligonucleotides docked to the M41-RBD. Columns with × indicate the sugar in that row was docked seven or more times out of 20 at the indicated site on the protein. Array scores are from Wickramasinghe *et al.* (11) and referenced in the figure as *Array score*¹. White columns were against structure without sugars, and gray columns were LC/MS-identified where the sugars were modeled. Bottom, RBD-binding domain of M41 from PDB structure 6cv0. Glycosylation sites are shown as cyan balls. Sites where two or more oligosaccharides docked seven or more times are indicated as colored space-filled amino acids. Colors and labels match the table above. B is A turned 90° toward the user. Structure representations were made in CCP4-MG (38). Sugar symbols were rendered with DrawGlycan-SNFG (www.virtualglycome.org/DrawGlycan/)³ (39).

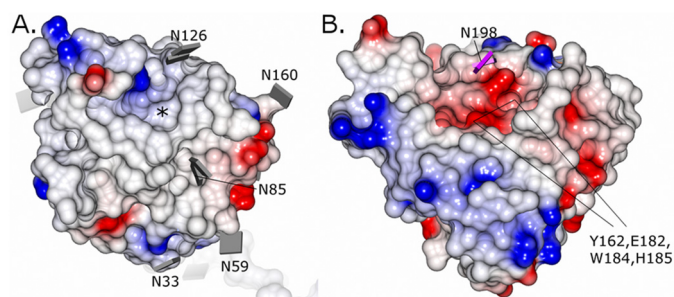


Figure 7. Charge distribution looking down on the potential sialic acid-binding site of M41 (A) and bovine (B) RBDs. Orientation of both proteins matches that of Fig. 6B. Positive electrostatic charge is blue, and negative is red. Sugars are gray boxes on A and pink boxes on B. Y162, E182, W184, and H185 in B are involved in binding to sialic acid. The large asterisk in A indicates possible binding site based on structural comparison between the two proteins. Images were made with CCP4-MG (38). Bovine coordinates are from PDB code 4H14.

away from the galectin fold where binding is associated in the docking study. Site Asn-194 is at the bottom of the galectin fold and is required for ligand binding. Site Asn-126 is at the top of the galectin fold and is also required for binding. Although we cannot conclude that sialylation is required at Asn-194 and Asn-126, it is clear that glycosylation at these sites serves a role in ligand binding.

The publication of the cryo-EM structure of M41 (18), the first structure of a spike protein from a gammacoronavirus, made it possible to visualize the distribution of the glycosylation sites in the tertiary structure of the protein. The study verified the site occupancy we observed on M41–RBD because 9 of 10 of the glycosylation sites in the EM structure were occupied. Site Asn-246, not occupied in the EM structure, is on a β -strand in the EM structure, and it forms close contacts with the S1 C-terminal domain in the native protein. The C-terminal domain was not part of our construct. Therefore, Asn-246 in the recombinant constructs was likely in an environment much different from that found in the full-length protein.

Many human galectins, and also the bovine β -coronavirus spike protein (18), bind sugars at what is the top of the β -sandwich near site Asn-126 in the RBD constructs (see Fig. 5). The bovine RBD site Asn-198 closely aligns with site Asn-126 of M41 (see Fig. 7). In the bovine protein, this demarks the region of proposed ligand binding. Loss of Asn-126 in the M41 RBD abrogates binding to trachea tissue. Although ablation of Asn-126 diminishes ligand binding, our docking study gave no evidence that this is the sialyl ligand-binding site in M41. Evaluation of the charge distribution in the proposed binding sites indicates that the bovine site is negatively charged, whereas the negative charge in the same region in M41 is sparse (Fig. 7). This difference in charge near Asn-126 may explain the lack of ligand docking in this region (*gray* β -strands in Fig. 6B) during docking simulations.

The precise ligand-binding region of proteins with a galectin fold varies. Rotavirus protein VP4, for example, binds sialic acid in a groove between the β -sheets of the sandwich (20). The clustering of five of six required *N*-glycosylation sites suggests the location of the ligand-binding site may be on the right of the galectin fold as shown in Fig. 5. Our docking experiments studying 17 possible oligosaccharide ligands to M41 were not

conclusive in terms of binding energies but did identify four potential saccharide-binding regions (Fig. 6). Docking also demonstrated that glycosylation affects binding *in silico* because one potential site (site A; see Fig. 6) lost favor, whereas another one, site D, dramatically gained favor when the protein was glycosylated. Site D is in the center of three glycosylated asparagines required for binding (Asn-59, Asn-85, and Asn-160), and one whose loss results in a very strong histochemical signal and has a protein-wide effect on glycosylation with increased sialylation (Asn-145). In addition, the site D region is negatively charged (see Fig. 7A) like the proposed sialyl ligand-binding site on the bovine protein (Fig. 7B) (19). All the ligands that interacted with site D were sialylated and included the glycan that bound in our ELISA studies. Interestingly, carbohydrate–carbohydrate contacts were detected in the RBD–ligand interactions at site D. This is an intriguing result because carbohydrate–carbohydrate interactions, although not common, have been reported between nonfucosylated antibodies and their receptor, in cell–cell adhesion interactions, between tumor antigens, and between bacterial receptors and mucin (21–25). A literature search did not uncover any reported carbohydrate–carbohydrate interactions between virus and host. Although our docking study must be evaluated in the context of the higher root mean square deviations typical of EM structures, and the inexactness of modeled oligosaccharides, results suggest that a combination of carbohydrate–carbohydrate and carbohydrate–protein interactions should be considered in the binding mechanism.

In conclusion, we have shown that glycosylation of six sites on the M41 IBV RBD are necessary for the interaction of M41 with both trachea tissue and Neu5Ac(α 2–3)Gal(β 1–3)GlcNAc ligand in ELISA. Based on occupancy data, at least nine sites were glycosylated in the recombinant M41–RBD. Deletion of individual glycosylation sites had little effect on secondary structure, but it did have some effect on overall glycosylation profiles of some variants, especially N145A. Some differences can be expected because one site, with specific glycans, is lost from each variant, thus mildly altering overall profiles. *In silico* docking suggests that glycosylation may guide ligand binding. Especially intriguing is site D, where glycosylation is required for *in silico* docking at that site. The interaction of M41 IBV with sialyl ligand may prove to be a unique interaction involving both carbohydrates and protein. Further investigation is warranted.

Experimental procedures

Ethics statement

The tissues used for this study were obtained from the tissue archive of the Veterinary Pathologic Diagnostic Center (Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, The Netherlands). This archive is composed of paraffin blocks with tissues maintained for diagnostic purposes; no permission from the Committee on the Ethics of Animal Experiment is required.

Plasmid construction

The pCD5 vector containing IBV M41–RBD in-frame with a C-terminal GCN4 trimerization motif and Strep-Tag has been

Table 3**Primers used for site-directed mutagenesis to generate Asn-to-Ala and Val-to-Ala substitutions**

The sequence encoding alanine is in lowercase. FW means forward; RV means reverse.

Mutant	Primer sequence for Asn → Ala and Val → Ala substitutions
N33A	FW: CGCTGTGGTgctATCTCCAGCG RV: TAAGTCCTCCATGCAGG
V57A	FW: GGAGGAAGGgcgGTGAACGCC RV: GTGAATTGTGCCACGATG
V58A	FW: GGAAGGTGgcgAACGCCTCC RV: TCCGTGAATTGTGCCAC
N59A	FW: AAGGGTGGTgcccGCCTCCAGCA RV: CCTCCGTGAATTGTGCC
N85A	FW: AGCCACTGTgctTTTAGCGACACC RV: GTGCAGAACTGGGAGCTG
N126A	FW: GCTGTTCTACgctCTGACAGTGTCCGTGG RV: TGGCCGTCTTCATGGCG
N145A	FW: GTGCGTGAACgctCTGACCTCCG RV: TGAAGCTCTTAAAGGTTG
N160A	FW: GTATACATCCGCTGAGACCACAGATGTGACCAGC RV: ACCAGTTCGCCGTTTCAGG
N194A	FW: CTACTTCGTgctGGCACAGCCCAGGAC RV: GCCAGGCCCTTCACCTCC
N219A	FW: CAACACCGGAgctTTCTCCGATGGC RV: TACTGCAGGCCAGCAGT
N229A	FW: TCCGTTTCATgcccAGCTCCCTGG RV: TAAAAGCCATCGGAGAAATTC
N246A	FW: GAACAGCGTGGCTACCACATTCAC RV: TCGCGGTACACAATAAAC

described previously (10). Site-directed mutagenesis using the Q5 technology (New England Biolabs) was performed to mutate the asparagine-encoding residues of the N-linked glycosylation sequence motif NX(S/T) into alanine or valine using the primers in Table 3. Sequences of the resulting RBDs were confirmed by Sanger sequencing (Macrogen, The Netherlands).

Production of recombinant proteins

HEK293T (ATCC CRL-3216) cells were transfected with pCD5 plasmids using polyethyleneimine at a 1:12 ratio. The recombinant proteins were purified using Strep-Tactin–Sepharose beads, as described previously (11), and their production was confirmed by Western blotting using Strep-Tactin HRP antibody (IBA, Germany).

CD

Recombinant M41 and its variants were prepared for CD spectroscopy by buffer exchange and concentration with four centrifugation cycles through 10-kDa MWCO Amicon Ultra 0.5-ml centrifugal filters (UFC 501024) into 10 mM sodium phosphate, pH 7.75. Final concentrations were measured with a Thermo Fisher Scientific Nanodrop 2000 spectrophotometer. CD spectra were collected on a JASCO J-810 spectropolarimeter with a Peltier thermostated fluorescence temperature controller module. Samples were diluted to 0.06 mg/ml and four scans accumulated from 285 to 190 nm with a scanning speed of 10 nm/min, digital integrated time 1-s, bandwidth 1 nm, and standard sensitivity at 25 °C. A thermal melt was done from 25 to 95 °C with a ramp rate of 1 °C/min. Measurements were taken every 2° at 222, 218, 215, 212, 208, 205, 196, and 194 nm. A full CD scan was collected at 95 °C. The temperature was then lowered to 25 °C. After allowing the protein to refold for 20 min at 25 °C, a third CD scan was taken at 25 °C to measure recov-

ery. A Savitzky-Golay filter was used to smooth CD data at different temperatures for visual comparison (Fig. S2).

Secondary structure calculations for the CD data collected at 25 °C before the thermal melt were processed by Dichroweb (16) using the CDSSTR (26), Selcon3 (27), and Contill (28) algorithms with protein reference set 7. Results from the three algorithms were averaged and plotted in Fig. 2.

Protein histochemistry

Histochemistry was performed as described previously (11). Briefly, chicken trachea tissues from a 7-week-old broiler chicken were sectioned at 4 μm before incubation with RBD proteins at 100 μg/ml. Desialylated tissues were prepared by pre-treatment with 2 milliunits of neuraminidase (sialidase) from *A. ureafaciens* (AUNA, Sigma, Germany) in 10 mM potassium acetate, 2.5 mg/ml Triton X-100, pH 4.2, at 37 °C overnight before protein application. Chicken trachea tissues were from a 7-week-old broiler chicken (*G. gallus*) obtained from the tissue archive of the Veterinary Pathologic Diagnostic Center (Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, The Netherlands).

ELISA

Sialic acids (Neu5Acα2–3Galβ1–3GlcNAc-PAA, 3-SiaLc-PAA, GlycoNZ, Russia) were coated (1 μg/well) in a 96-well Maxisorp plate (NUNC, Sigma) at 4 °C overnight, followed by blocking with 3% BSA (Sigma) in PBS-0, 1% Tween. RBD proteins (100 μg/ml) were preincubated with Strep-Tactin-HRPO (1:200) for 30 min on ice, before applying them to the plates for 2 h at room temperature. 3,3',5,5'-Tetramethylbenzidine substrate was used as a peroxidase substrate to visualize binding, after which the reaction was terminated using 2 N H₂SO₄. Absorbances (A_{450 nm}) were measured in a FLUOstar Omega (BMG Labtech) microplate reader, and MARS data analysis software was used for analysis. Protein samples of each recombinant protein were measured at each concentration in triplicate. Statistical analysis was performed by comparing each protein to the unmodified RBD using two-way ANOVA with Dunnett's multiple comparisons test where α was set to 0.05.

Glycopeptide preparation, enrichment, and N-glycan release

The workflow is shown in Fig. S3. Aliquots between 200 and 400 μg of M41, N59A, and N145A and 50 μg of the remaining proteins were digested with trypsin as per An and Cipollo (29). Approximately 25–100-μg aliquots of protease-digested proteins were processed for deglycosylated glycopeptide and permethylated glycan analyses. Samples were resuspended in 50 mM ammonium bicarbonate, pH 8.0. Glycans were released by digestion with 10 units/μl PNGase F (glycerol-free from New England Biolabs) for 3 h at 37 °C. The samples were adjusted to pH 5.0 with 2–4 μl of 125 mM HCl. To maximize glycan release, samples were further digested with 0.15 milliunits/μl PNGase A overnight at 37 °C. Free glycans and deglycosylated peptides were separated using C18 SPE cartridges (Thermo Fisher Scientific). Intact glycopeptide analyses were performed using 175–300 μg of HILIC-enriched glycopeptides as per An and Cipollo (29). Following data collection on the trypsinized glycopeptides, the remainder of the M41, N59A, and N145A sam-

ples were digested with chymotrypsin at a ratio of 1:20 overnight at 25 °C, and HILIC was enriched a second time (for the M41 and N59A samples only) prior to LC/MS analysis.

Site occupancy

LC/MS^E data were collected on trypsinized peptides deglycosylated with PNGase F as described under *N*-glycan release. Asparagines that are deglycosylated by PNGase F are converted to aspartate with a mass gain of 0.984 Da due to the replacement of $-\text{NH}_2$ with $-\text{OH}$. The percent occupancy for each site is calculated by comparing the intensity of peptides with Asn to those with Asp. However, spontaneous deamidation of unmodified Asn to Asp can also occur. ¹⁸O-Water, which results in mass shift of 2.984 Da, was used to ensure calculated percent occupancy was not skewed due to spontaneous deamidation. This experiment allows for examination of both spontaneous and enzymatically catalyzed deamidation, and therefore, accurate estimations of percent occupancy of glycosites can be determined. Percent occupancy was calculated by comparing the intensities of the deglycosylated (DG) and nonglycosylated (NG) peptides using the equation: $\text{DG}/(\text{DG} + \text{NG}) \cdot 100$.

Purification, permethylation, and semi-quantitation of free glycans

PNGase-released *N*-glycans were applied to C18 SPE and eluted with 0.1% formic acid leaving the deglycosylated peptides bound to the C18 column. The glycan eluate fractions were combined, and butanol was added to a final concentration of 1%. The samples were then loaded onto 100-mg porous graphite columns prepared first by sequential washes of 1 ml of 100% acetonitrile (ACN), 1 ml of 60% ACN in water, 1 ml of 30% ACN in water, and 1 ml of water. All solutions contained 0.1% trifluoroacetic acid (TFA). The loaded columns were washed three times with 1 ml of 0.1% TFA in water, then eluted with 30% ACN, 0.1% TFA, water, followed by 60% ACN, 0.1% TFA, and water. The eluents were pooled and dried in glass vials by rotary evaporation. Permethylation was done following the method of Cincanu and Costello (30) and Cincanu and Kerek (31). MALDI-TOF analysis of permethylated *N*-glycans was performed on a Bruker AutoflexTM speed mass spectrometer in positive polarity reflectron mode. 2,5-Dihydroxybenzoic acid was used as a matrix, and malto-oligosaccharides were used as an external calibrant. Data were processed using FlexAnalysisTM. Each sample was spotted three times, and scans were collected in positive reflectron mode. Peaks were picked and assigned, and intensities were averaged across each set of spots using in-house software. Assignments were based on glycans known to be present in HEK293T cells.

Reverse-phase nanoLC/MS^E analysis of glycopeptides and peptides

Each peptide or glycopeptide sample was analyzed three times. A C18 column (BEH nanocolumn 100 μm inner diameter \times 100 mm, 1.7- μm particle, Waters Corp.) was used for nanoLC/MS^E analyses. A nanoAcquity UPLC system (Waters Corp.) was used for automatic sample loading and flow control. Load buffer was 3% ACN, 97% water. Peptides were eluted via a 60-min gradient from 3 to 50% ACN with a flow of 0.4 $\mu\text{l}/\text{min}$.

All chromatography solutions included 0.1% formic acid. The eluent flowed to an uncoated 20- μm inner diameter PicoTip Emitter (New Objective Inc., Woburn, MA). The mass spectrometer was a SYNAPT G2 HDMS system (Waters Corp.). Applied source voltage was 3000 V. Data were collected in positive polarity mode using data-independent MS^E acquisition, which consists of a starting 4-V scan followed by a scan ramping from 20 to 50 V in 0.9 s. To calibrate internally, every 30 s 400 fmol/ μl Glu-fibrinopeptide B with 1 pmol/ μl leucine enkephalin in 25% acetonitrile, 0.1% formic acid, 74.9% water was injected through the lockmass channel at a flow rate of 500 nl/min. Initial calibration of the mass spectrometer was performed in MS² mode using Glu-fibrinopeptide B and tuned for a minimum resolution of 20,000 full-width at half-maximum.

Data analysis for peptide and glycopeptide identification

NanoLC/MS^E data were processed using BiopharmaLynx 1.3 (Waters Corp.) and GLYMPS (in-house software) (32, 33) to identify specific glycans on each peptide. The search settings included trypsin digest with up to one missed cleavage, fixed cysteine carbamidomethylation, variable methionine oxidation, and variable *N*-glycan modifications based on a building block glycan library. Assignment inclusion criteria were as follows: 1) the presence of a core fragment (peptide, peptide + HexNAc, peptide + HexNAc₂, peptide + dHex₁HexNAc₁, and peptide + Hex₁HexNAc₂); 2) the presence of three or more peptide fragments; 3) the presence of three or more assigned glycopeptide fragments; 4) assignment is made in at least 2 of 3 injections; and 5) the existence of the glycan in GlyConnect (<https://glyconnect.expasy.org>).³

Docking

Residues 21–268 of the M41 spike EM structure were extracted from the published structure (PDB code 6cv0) (18). This corresponds to the M41–RBD used in this paper. Glycamweb's glycoprotein-builder program (34) was used to add the major oligosaccharide found at each glycosylation site onto the protein *in silico*. All glycosites in the M41 EM structure were occupied except Asn-246; however, Asn-246 was occupied in our data and was populated accordingly. All glycosites were glycosylated in the new PDB file based on best evidence from our MS data. The coordinates of M41–RBD without glycans, M41–RBD with modeled glycans, and bovine RBD (PDB code 4H14) were used in docking experiments. A virtual library of 17 oligosaccharides representing a variety of binding epitopes was created based on the CFG array version 4.2 (see Fig. 6 for a list). Raw models of the oligosaccharide ligands were created with the AMBER tool tleap (www.ambermd.org)³ utilizing the GLYCAM06 force field (35), then energy minimized using YASARA (36). Dock screening of the library was performed with the YASARA implementation of Autodock Vina (37) with default parameters. A molecular dynamics simulation with explicit water (TP3) but with fixed coordinates for the backbone atoms was run on the glycosylated M41 RBD model to allow the amino acid side chains to accommodate the added glycans and to find low energy conformations. Two models were extracted from the glycosylated MD RBD run at 5 and 10 ns, which were used for dock screening with the virtual library.

Avian coronavirus glycosylation

Each oligosaccharide ligand was docked against the structures 20 times. Docking results shown in Fig. 6 are for the 10-ns model. Results were similar in the 5-ns models.

Author contributions—L. M. P., K. M. B., R. P. d. V., J. F. C., and M. H. V. conceptualization; L. M. P., K. M. B., H. F. A., and J. F. C. data curation; L. M. P., K. M. B., H. F. A., and J. F. C. formal analysis; L. M. P., K. M. B., H. F. A., and J. F. C. investigation; L. M. P., K. M. B., H. F. A., and J. F. C. methodology; L. M. P. and J. F. C. writing-original draft; L. M. P., K. M. B., H. F. A., R. P. d. V., and M. H. V. writing-review and editing; J. F. C. and M. H. V. supervision; J. F. C. and M. H. V. validation; J. F. C. visualization.

References

1. Cook, J. K., Jackwood, M., and Jones, R. C. (2012) The long view: 40 years of infectious bronchitis research. *Avian Pathol.* **41**, 239–250 [CrossRef Medline](#)
2. Bosch, B. J., van der Zee, R., de Haan, C. A., and Rottier, P. J. (2003) The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**, 8801–8811 [CrossRef Medline](#)
3. Masters, P., and Perlman, S. (2013) Coronaviridae. In *Fields Virology* (Howley, P., and Knipe, D. M., eds), 6th Ed, pp. 825–858, Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, PA
4. Wickramasinghe, I. N., van Beurden, S. J., Weerts, E. A., and Verheije, M. H. (2014) The avian coronavirus spike protein. *Virus Res.* **194**, 37–48 [CrossRef Medline](#)
5. Binns, M. M., Bournsnel, M. E., Cavanagh, D., Pappin, D. J., and Brown, T. D. (1985) Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* **66**, 719–726 [CrossRef Medline](#)
6. Cavanagh, D. (1983) Coronavirus IBV: structural characterization of the spike protein. *J. Gen. Virol.* **64**, 2577–2583 [CrossRef Medline](#)
7. Cavanagh, D. (1983) Coronavirus IBV glycopolypeptides: size of their polypeptide moieties and nature of their oligosaccharides. *J. Gen. Virol.* **64**, 1187–1191 [CrossRef Medline](#)
8. Delmas, B., and Laude, H. (1990) Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J. Virol.* **64**, 5367–5375 [Medline](#)
9. Lewicki, D. N., and Gallagher, T. M. (2002) Quaternary structure of coronavirus spikes in complex with carcinoembryonic antigen-related cell adhesion molecule cellular receptors. *J. Biol. Chem.* **277**, 19727–19734 [CrossRef Medline](#)
10. Promkuntod, N., van Eijndhoven, R. E., de Vriese, G., Gröne, A., and Verheije, M. H. (2014) Mapping of the receptor-binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virology* **448**, 26–32 [CrossRef Medline](#)
11. Wickramasinghe, I. N., de Vries, R. P., Gröne, A., de Haan, C. A., and Verheije, M. H. (2011) Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J. Virol.* **85**, 8903–8912 [CrossRef Medline](#)
12. Winter, C., Schwegmann-Wessels, C., Cavanagh, D., Neumann, U., and Herrler, G. (2006) Sialic acid is a receptor determinant for infection of cells by avian infectious bronchitis virus. *J. Gen. Virol.* **87**, 1209–1216 [CrossRef Medline](#)
13. Raman, R., Tharakaraman, K., Sasisekharan, V., and Sasisekharan, R. (2016) Glycan–protein interactions in viral pathogenesis. *Curr. Opin. Struct. Biol.* **40**, 153–162 [CrossRef Medline](#)
14. Zheng, J., Yamada, Y., Fung, T. S., Huang, M., Chia, R., and Liu, D. X. (2018) Identification of N-linked glycosylation sites in the spike protein and their functional impact on the replication and infectivity of coronavirus infectious bronchitis virus in cell culture. *Virology* **513**, 65–74 [CrossRef Medline](#)
15. Geilhausen, H. E., Ligon, F. B., and Lukert, P. D. (1973) The pathogenesis of virulent and avirulent avian infectious bronchitis virus. *Arch. Gesamte Virusforsch.* **40**, 285–290 [CrossRef Medline](#)
16. Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392–400 [CrossRef Medline](#)
17. Ambepitiya Wickramasinghe, I. N., de Vries, R. P., Weerts, E. A., van Beurden, S. J., Peng, W., McBride, R., Ducatez, M., Guy, J., Brown, P., Eterradosi, N., Gröne, A., Paulson, J. C., and Verheije, M. H. (2015) Novel receptor specificity of avian gammacoronaviruses that cause enteritis. *J. Virol.* **89**, 8783–8792 [CrossRef Medline](#)
18. Shang, J., Zheng, Y., Yang, Y., Liu, C., Geng, Q., Luo, C., Zhang, W., and Li, F. (2018) Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. *PLoS Pathog.* **14**, e1007009 [CrossRef Medline](#)
19. Peng, G., Xu, L., Lin, Y. L., Chen, L., Pasquarella, J. R., Holmes, K. V., and Li, F. (2012) Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **287**, 41931–41938 [CrossRef Medline](#)
20. Dormitzer, P. R., Sun, Z. Y., Wagner, G., and Harrison, S. C. (2002) The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO J.* **21**, 885–897 [CrossRef Medline](#)
21. Ferrara, C., Grau, S., Jäger, C., Sondermann, P., Brünker, P., Waldhauer, I., Hennig, M., Ruf, A., Rufer, A. C., Stihle, M., Umaña, P., and Benz, J. (2011) Unique carbohydrate–carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12669–12674 [CrossRef Medline](#)
22. Lorenz, B., Álvarez de Cienfuegos, L., Oelkers, M., Kriemen, E., Brand, C., Stephan, M., Sunnick, E., Yüksel, D., Kalsani, V., Kumar, K., Werz, D. B., and Janshoff, A. (2012) Model system for cell adhesion mediated by weak carbohydrate–carbohydrate interactions. *J. Am. Chem. Soc.* **134**, 3326–3329 [CrossRef Medline](#)
23. Bucior, I., and Burger, M. M. (2004) Carbohydrate–carbohydrate interaction as a major force initiating cell–cell recognition. *Glycoconj. J.* **21**, 111–123 [CrossRef Medline](#)
24. Sletmoen, M., Gerken, T. A., Stokke, B. T., Burchell, J., and Brewer, C. F. (2018) Tn and STn are members of a family of carbohydrate tumor antigens that possess carbohydrate–carbohydrate interactions. *Glycobiology* **28**, 437–442 [CrossRef Medline](#)
25. Radziejewska, I., Borzym-Kluczyk, M., and Leszczyńska, K. (2013) Are Lewis b and H type 1 on *Helicobacter pylori* involved in binding of bacteria to MUC1 mucin? *Adv. Clin. Exp. Med.* **22**, 347–353 [Medline](#)
26. Manavalan, P., and Johnson, W. C., Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**, 76–85 [CrossRef Medline](#)
27. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44 [CrossRef Medline](#)
28. van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem.* **191**, 110–118 [CrossRef Medline](#)
29. An, Y., and Cipollo, J. F. (2011) An unbiased approach for analysis of protein glycosylation and application to influenza vaccine hemagglutinin. *Anal. Biochem.* **415**, 67–80 [CrossRef Medline](#)
30. Ciucanu, I., and Costello, C. E. (2003) Elimination of oxidative degradation during the per-O-methylation of carbohydrates. *J. Am. Chem. Soc.* **125**, 16213–16219 [CrossRef Medline](#)
31. Ciucanu, I., and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Res.* **131**, 209–217 [CrossRef](#)
32. An, Y., McCullers, J. A., Alymova, I., Parsons, L. M., and Cipollo, J. F. (2015) Glycosylation analysis of engineered H3N2 influenza A virus hemagglutinins with sequentially added historically relevant glycosylation sites. *J. Proteome Res.* **14**, 3957–3969 [CrossRef Medline](#)
33. Parsons, L. M., An, Y., de Vries, R. P., de Haan, C. A., and Cipollo, J. F. (2017) Glycosylation characterization of an influenza H5N7 hemagglutinin series with engineered glycosylation patterns: implications for

- structure-function relationships. *J. Proteome Res.* **16**, 398–412 [CrossRef](#) [Medline](#)
34. Woods Group (2005–2018) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA
35. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655 [CrossRef](#) [Medline](#)
36. Krieger, E., and Vriend, G. (2014) YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics* **30**, 2981–2982 [CrossRef](#) [Medline](#)
37. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 [Medline](#)
38. McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 386–394 [CrossRef](#) [Medline](#)
39. Cheng, K., Zhou, Y., and Neelamegham, S. (2017) DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology* **27**, 200–205 [CrossRef](#) [Medline](#)