Three Amino Acid Changes In Avian Coronavirus Spike Protein Allows Binding To Kidney Tissue

Running title: **Three amino acids in IBV spike alters receptor tropism**

Kim M. Bouwman[1] *, Lisa M. Parsons[2], Alinda J. Berends[1], Robert P. de Vries[3], John F. Cipollo[2], Monique H. Verheije[1] *

[1] Division of Pathology, Department of Pathobiology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands.

[2] Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, United States.

[3] Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

Corresponding authors: K.M. Bouwman k.m.bouwman@uu.nl
Monique H. Verheije m.h.verheije@uu.nl

Abstract:

Infectious bronchitis virus (IBV) infects ciliated epithelial cells in the chicken respiratory tract. While some IBV strains replicate locally, others can disseminate to various organs, including the kidney. Here we elucidate the determinants for kidney tropism by studying interactions between the receptor binding domain (RBD) of the viral attachment protein spike from two IBV strains with different tropisms. Recombinantly produced RBDs from the nephropathogenic IBV strain QX and from the non-nephropathogenic strain M41 bound to the epithelial cells of the trachea. In contrast, only QX-RBD binds more extensively to cells of the digestive tract, urogenital tract, and kidneys. While removal of sialic acids from tissues prevented binding of all proteins to all tissues, binding of QX-RBD to trachea and kidney could not be blocked by pre-incubation with synthetic alpha-2,3-linked sialic acids. The lack of binding of QX-RBD to a previously identified IBV-M41 receptor was confirmed by ELISA, demonstrating that tissue binding of QX-RBD is dependent on a different sialylated glycan receptor. Using chimeric RBD proteins, we discovered that the region encompassing amino acids 99-159 of QX-RBD was required to establish kidney binding. In particular, QX-RBD amino acids 110-112 (KIP) were sufficient to render IBV-M41 with the ability to bind to kidney, while the reciprocal mutations in IBV-QX abolished kidney binding completely. Structural analysis of both RBDs suggests that the receptor binding site for QX is located at a different location on the spike than that of M41.

Importance:

Infectious bronchitis virus is the causative agent of Infectious bronchitis in chickens. Upon infection of chicken flocks, the poultry industry faces substantial economic losses by diminished egg quality and increased morbidity and mortality of infected animals. While all IBV strains infect the chicken respiratory tract via the ciliated epithelial layer of the trachea, some strains can also replicate in the kidneys, dividing IBV in two pathotypes: non-nephropathogenic (example IBV-M41) and nephropathogenic viruses (including IBV-QX). Here we set out to identify the determinants for the extended nephropathogenic tropism of

47 IBV-QX. Our data reveal that each pathotype makes use of a different sialylated glycan
48 ligand, with binding sites on opposite sides of the attachment protein. This knowledge should
49 facilitate the design of antivirals to prevent coronavirus infections in the field.
50
51 **Introduction:**
52 Infectious bronchitis is a disease in chickens caused by infectious bronchitis virus (IBV). In
53 the poultry industry, infection of chicken flocks with IBV causes economic losses by reducing
54 egg quantity and quality. In addition, animals become more susceptible to secondary
55 bacterial infections like *E. coli* (1). The severity of disease and organs affected depends
56 primarily on the IBV strain (2). Phylogenetic classification of IBV strains results in 32
57 phylogenetic lineages (G-I 1-27, G-II-GVI) (3), of which GI-1 includes historically the first IBV
58 genotype identified, Massachusetts (IBV-Mass). IBV-Mass infections are reported worldwide,
59 and in Europe GI-1 is currently the 3rd most prevalent genotype (2). The more prevalent IBV
60 genotype circulating in Europe is IBV-QX (GI-19) (2, 3), which has been reported to cause
61 kidney disease in contrast to IBV-Mass (2).
62
63 IBV primarily infects the respiratory tract, where the virus can bind and infect the ciliated
64 epithelial lining of the trachea (4, 5). Upon infection of IBV clinical symptoms such as
65 snicking, wheezing and/or nasal discharge are reported (6). While infection of IBV-Mass (of
66 which strain M41 is the prototype) is predominantly detected in the upper respiratory tract (7)
67 including the trachea (2), replication of IBV-QX is additionally found in the kidneys (7-9),
68 oviduct and the gastrointestinal tract (10, 11), leading to additional clinical symptoms like
69 swollen proventriculus (12) and reduction of egg production (13, 14). Because of these
70 additional clinical symptoms, IBV-QX is described as a nephropathogenic IBV strain (2).
71
72 Binding to host tissues is the first step in the viral life cycle of IBV and therefore a critical
73 factor in determining tissue tropism. Tissue tropism differs based on the amino acid
74 composition of the spike protein as shown by recombinantly produced proteins (15-17) and

infection assays with recombinant viruses (18). The spike of IBV is post-translationally cleaved into two subunits, S1 and S2, where S2 is anchored in the virus membrane and important for membrane fusion. S1 comprises the head domain of spike and is responsible for host receptor binding (19). Using recombinantly expressed M41-S1 proteins, alpha-2,3-linked sialic acids were identified as the IBV receptor on a glycan array, where specific binding to the ligand Neu5Acα2-3Galβ1-3GlcNAc was observed (19). Recently the cryo-EM structure of the M41 spike has been resolved (20), indicating that the S1 subunit consists of two independent folding domains, NTD (amino acids 21-237) and CTD (amino acids 269-414), with a proposed receptor binding site in both domains. Experimental evidence using recombinantly expressed spike domains have indicated that amino acids 19-272 of the M41 spike are sufficient for binding to trachea as well as binding to alpha-2,3-linked sialic acids (15). This domain thus contains a receptor binding domain (RBD) and can be used to study biological implications of genetic variation in circulating IBV genotypes.

In this study we set out to identify how genetic variations in IBV spike proteins have contributed to different host tropisms. We demonstrate that QX-RBD binding to trachea and kidney is dependent on a different sialylated glycan ligand compared to M41-RBD. In particular, introduction of amino acids 110-112 (KIP) of the QX spike into M41-RBD was sufficient to extend its tropism toward the kidney. Previous docking experiments (17)and structural analysis suggest that the binding pockets for the different glycans are located at opposite sites of each spike protein.

Results:

The N-terminal domain of IBV-QX spike contains a receptor binding domain.

Eighty-five percent of the amino acids between the sequences of the first 257 amino acids of IBV-QX and IBV-M41 are either identical or similar. Here we set out to determine which of the dissimilar amino acids are the determinants for the difference in tissue tropism.

102

103 In previous work we demonstrated the M41-RBD was sufficient to bind chicken trachea (15).
104 To verify that no additional sites are present in M41 that could bind kidney or trachea tissue,
105 we produced recombinant proteins consisting of the full ectodomain (ED), the S1 portion of
106 the ED, the RBD (NTD of S1), and the CTD of S1. Each protein was assessed for binding
107 using trachea and kidney tissue slides. Binding to trachea tissue was observed using M41-
108 ED, S1 and RBD but not CTD to ciliated epithelium of the trachea, specifically located at the
109 base of the cilia (Fig. 1), confirming previous observations (15, 19, 21). None of the proteins
110 bound kidney tissue, which is shown by a representative picture using M41-RBD (Fig 1B).
111 Binding affinity to the known ligand (Neu5Acα2-3Galβ1-3GlcNAc) in ELISA was observed
112 using M41-RBD, -S1 and -ED, not significantly different when compared to each other, but
113 significantly higher compared to M41-CTD and TCoV-S1 (Fig. 1C). These results indicate
114 ligand binding of M41-RBD is not significantly different compared to M41-S1 and -ED,
115 suggesting no additional ligand binding motifs are present in S1 and ED, thus, in the
116 remaining experiments we used M41-RBD as the tissue tropism of the virus is reflected using
117 this recombinant protein.

118 Amino acid alignment of the mature protein sequence of the receptor binding domain (RBD)
119 of M41 and a comparable size fragment of the QX spike displayed a sequence identity of
120 73.6% (Fig. 2A), with highest sequence diversity between amino acids 37-60 and 98-115.
121 These regions include the previously described hyper variable regions (HVRs, highlighted in
122 grey) of M41-S1 (22). Before studying whether sequence diversity between the RBDs of M41
123 and QX contributes to the reported broader tropism of QX *in vivo* we first determined if the
124 potential RBD of QX behaved like that of M41 (Fig. 1) and that it contains a receptor binding
125 domain (15). Both proteins were produced as soluble recombinant protein in mammalian
126 cells and analyzed on western blot after purification. Before loading, a fraction was pre-
127 treated with PNGaseF to remove post-translational glycosylation. QX- and M41-RBD
128 migrated comparable at around 55kDa (including glycosylation) and had a backbone of

around 32kDa as expected after PNGaseF treatment (Fig. 2B). Circular dichroism (CD) spectroscopy was used to assess similarities in secondary structure between M41- and QX-RBD. Spectra at all temperatures followed the same curve, and both proteins had similar broad melting curves, indicating that both proteins are equally stable (data not shown). Subsequent secondary structure calculations using Dichroweb (23) presented that M41- and QX-RBD contain 29 and 25% α-helix, 16 and 17% β-strands, and 55 and 58% random structures, respectively (Fig. 2C). Finally, we confirmed that the QX-RBD was biologically active by applying it to chicken trachea tissue slides in protein histochemistry. We observed clear binding to the ciliated lining of epithelial cells and structures present in the kidney (Fig. 2D) indicating that QX-RBD, like M41-RBD, contains a receptor binding site.

QX-RBD shows a broader tissue tropism than M41-RBD.

Next, we used M41- and QX-RBDs to study the distribution of host attachment factors across chicken tissues. To this end, we allowed both proteins to bind to tissue microarray slides containing 28 different chicken tissues (24). Binding of M41-RBD was primarily found on the ciliated lining of the epithelium of the proximal and distal trachea (Fig. 1), but additional staining was observed in the epithelial lining of the colon, cecal tonsil, ureter, oviduct, and conjunctiva (Table 1). QX-RBD bound to the same tissues as M41-RBD, but additional binding was observed in gizzard, ileum, and cloaca of the digestive tract, as well as liver and kidneys (Table 1 and Fig. 2D), reflecting that observed *in vivo* for replication of both genotypes. Detailed analysis of staining present in the kidney showed that binding of QX-RBD was restricted to the parietal epithelium of Bowmans capsule in the glomerulus (Fig. 2D). No binding to the glomeruli was observed when using M41-RBD in three independent experiments using different protein batches. Taken together, QX-RBD shows a markedly broader binding profile than M41-RBD, which is in line with the reported broader tissue tropism *in vivo* (2).

QX-RBD binds to sialic acids on chicken tissues.

To investigate whether the expanded tropism of QX-RBD can be explained by binding with similar specificity, but higher affinity, to the previously identified M41 receptor (19), we pre-incubated both RBD proteins with the synthetic Neu5Acα2-3Galβ1-3GlcNAc before applying them to trachea and kidney tissue slides. As expected, binding of M41-RBD to the trachea was completely prevented (Fig. 3A, middle column) in the presence of the synthetic M41 ligand. In contrast, QX-RBD still showed strong binding to the ciliated epithelium of the trachea and glomeruli of the kidney. To confirm the loss of binding of QX-RBD to Neu5Acα2-3Galβ1-3GlcNAc a solid-phase ELISA was performed, in which Neu5Acα2-3Galβ1-3GlcNAc was coated. As expected, no binding of QX-RBD to this particular glycan was observed at any of the protein concentrations, comparable to that of the negative control TCoV-S1 (only binding longer branched galactose terminated glycans (25)), while M41-RBD bound to Neu5Acα2-3Galβ1-3GlcNAc in a concentration-dependent manner (Fig. 3B).

To reveal whether QX-RBD exclusively depends on sialic acids, trachea and kidney tissue slides were pre-treated with *Arthrobacter ureafaciens* neuraminidase (AUNA) before applying M41- and QX-RBD. Removal of sialic acids from trachea and kidney tissue completely prevented binding of both RBD proteins (Fig. 3A, right column), indicating that QX-RBD binding is dependent on the presence of sialic acids on host tissues.

M41-RBD gains kidney binding upon MLQ107-109KIP mutation.

To gain in depth knowledge on the interaction of the IBV-RBD proteins and chicken tissue, we set out to determine the critical amino acids of viral spike proteins involved in binding to these glycan receptors, thereby leading to the ability to bind to kidney tissue. Chimeric RBD proteins were generated by dividing each wildtype RBD into three domains and mixing them to get six different combinations (schematic representations in Fig. 4A). These chimeras were then applied to trachea and kidney tissue slides. Chimeric proteins containing amino acids 98-156 (middle domain) of M41 (M-M-Q, Q-M-M, and Q-M-Q) demonstrated reduced binding to trachea and no detectable binding to kidney (Fig. 4B). In contrast, chimeric

proteins containing this region of QX (Q-Q-M, M-Q-Q, and M-Q-M) had comparable binding to tissues as QX-RBD. In particular, strong binding to the ciliated epithelial lining of trachea and specific staining in Bowmans capsule in the glomerulus was observed (Fig. 4B). Like wild type RBDs, binding of all chimeric proteins was dependent on the presence of sialic acids, as pre-treatment of host tissues with AUNA abrogated binding (data not shown). M-M-Q, Q-M-M and Q-M-Q proteins had reduced affinity for Neu5Acα2-3Galβ1-3GlcNAc (Fig. 4C), potentially explaining the reduced staining of these proteins to trachea tissue (Fig. 4B). None of the RBD proteins containing the middle QX sequence (Q-Q-M, M-Q-Q, and M-Q-M) had affinity for this glycan in the ELISA as expected based on tissue staining (Fig. 4B and C), which is in line with the hypothesis these proteins are dependent on binding to the QX receptor instead of the known M41 receptor. These results indicate that the receptor binding site responsible for recognition of the QX receptor is determined by amino acids 99-159 of the spike.

To ultimately determine the critical residues of the RBD for the interaction with chicken kidney tissue, additional chimeric proteins were produced and used in protein histochemistry. We exchanged two triplets (highlighted in dark green in Fig. 2A) of amino acids in HVR 2 (amino acids 99-115 of M41), either alone or in combination, that had the high diversity in amino acid characteristics (schematic representations Fig. 5A). Introduction of the M41 sequence in the QX-RBD protein, SGS100-102Y (QX-Y) and KIP110-112MLQ (QX-MLQ) and their combination (QX-Y-MLQ), all resulted in a loss of binding to trachea and kidney tissues (Fig. 5B, right panel). In contrast, introduction of MLQ107-109KIP into M41-RBD (M41-KIP) resulted in gain of binding to glomeruli in kidney, both in a wild type background and in the Y99SGS (M41-SGS) mutant (Fig. 5B, left panel). In the ELISA both M41-SGS and M41-KIP demonstrated a decreased affinity for alpha-2,3-linked sialic acids compared to M41-RBD, while introduction of both triplets SGS and KIP (M41-SGS-KIP) completely abolished binding to this glycan (Fig. 5C). Taken together these results suggest that a

receptor binding site critical to establish kidney binding requires amino acids KIP at position 107-109 in M41-RBD.

Receptor binding site of the QX specific receptor differs from that proposed for M41.

Finally, we modelled QX-RBD based on a structural overlay with the recently resolved cryo-EM structure of M41 spike (20), and focused on the amino acids allowing kidney binding. The overall structure of both proteins is comparable (Fig. 6A, green ribbon M41, blue ribbon QX), however the loop consisting of HVR 2 is slightly larger in QX-RBD as expected, as there are two additional amino acids present (Fig. 6A, SGS100-102 for QX-RBD versus Y99 in M41-RBD). Interestingly this loop was predicted to be involved in sugar binding (20), which we showed to be true for QX-RBD, but not for M41-RBD. In detail, the tyrosine (Y99) in the M41 structure (Fig. 6A, beige) occupies more space than serine (S in QX) and can be seen reaching toward a neighboring loop. Furthermore the 110-112KIP sequence identified in QX-RBD (Fig. 6A, dark blue) places a positive charge at the protein surface which is not present in 107-109MLQ in M41-RBD (Fig. 6A light blue). Previous *in silico* docking analysis performed with potential alpha-2,3-linked ligands to the M41-RBD protein, identified amino acids S87, N144 and T162 to potentially be involved in receptor binding (17). When we highlighted these amino acids predicted to be involved in binding to alpha-2,3-linked sialic acids (Fig. 6B red spheres) and the amino acid triplicates involved in binding to the QX specific receptor (100-102 (SGS, yellow spheres) and 110-112 (KIP, dark blue spheres) in the overlayed RBD ribbon structure, we demonstrate that binding of the different ligands recognized by M41 and QX are on different sides of the protein (Fig. 6B). Furthermore, when these amino acids were highlighted in the full cryo-EM resolved structures of M41 (Fig. 6C) and QX (Fig. 6D), it clearly shows the potential ligand binding site of M41 is at a different location compared to the QX ligand binding site (Fig. 6C and D).

231 In conclusion we demonstrate that IBV-QX recognizes a sialylated glycan receptor present
232 on chicken tissues that differs from that recognized by M41, and that this binding is likely
233 required for the extended *in vivo* tissue tropism of the virus.

234 **Discussion**

235 In this study we reveal that nephropathogenic IBV-QX shows expanded binding tropism
236 based on interactions with sialic acid(s) on chicken tissues that differs from the receptor
237 elucidated for IBV-M41. Using chimeric proteins and *in silico* modeling, we conclude that
238 amino acids in hyper variable region 2 are critical for recognizing such a sialylated glycan
239 receptor.

240 The N-terminal domain of IBV-QX spike protein comprises, like previously shown for M41
241 (15), a receptor binding domain. Interestingly, QX-RBD shows no affinity for known ligand of
242 M41 (Neu5Acα2-3Galβ1-3GlcNAc) in glycan ELISA, while it gained binding to a novel
243 unidentified sialylated glycan receptor. Other avian gammacoronaviruses, including guinea
244 fowl and turkey coronavirus are dependent on long glycans (linear or branched) capped with
245 either an alpha-2,6-linked sialic acid (GfCoV only) or galactose ending glycans (both TCoV
246 and GfCoV) (25, 26). Viruses of other coronavirus genera are dependent on sialic acid
247 receptors, like alpha coronaviruses TGEV and PEDV (27, 28) and beta coronaviruses HCoV-
248 OC43 and BCoV (29). Whether other nephropathogenic IBV strains are dependent on the
249 same sialic acid receptors for binding and subsequent infection as QX, remains to be
250 determined.

251

252 To elucidate the specific ligand used by QX-RBD we have performed several binding studies
253 using previously developed glycan arrays (30, 31), containing multiple linear, and branched
254 glycans capped without or with alpha-2,3-linked sialic acids or alpha-2,6-linked sialic acids.
255 Unfortunately, no binding was observed using our RBD proteins. This may be explained by
256 the usage of RBD proteins, instead of the full S1, like used previously (19) or the composition

and fine structure of the glycans present in both arrays. On the arrays used, most glycans contain the linkage found in mammals (Galβ1,4GlcNAc), while the minority contain Galβ1,3GlcNAc linkage. The exact nature of the receptor recognized by IBV-QX could be a more complex glycan containing a Galβ1,3GlcNAc that is scarely populated on glycan arrays.

Comparison of the spikes of various IBV strains with reported nephropathogenicity, including IBV clade GI-14 (including strain B1648 (3)) and clade GI-13 (including strain 793B (3)), shows only nephropathogenic IBV clades contain an amino acid triplicate at position 100-102, whereas in IBV Mass genotypes 99Y/H is expressed, thereby shortening HVR 2 with two amino acids. Sequence alignment of this amino acid triplicate (100-102 in QX-RBD) varies in nephropathogenic IBV genotypes from SGS/SGT for clade GI-19 (IBV-QX), NQQ/SQQ for clade GI-13 (IBV-793B) and SGA for clade GI-14 (IBV-B1648) at that position. Furthermore, the amino acid triplet 110-112 KIP is not conserved across IBV genotypes. In these genotypes amino acid triplets LIQ for B1648 and MIP for 793B are present, which are sequence combinations of amino acids found in Mass (clade GI-1) and QX (clade GI-19). In terms of hydrophobicity and size, amino acid triplet MLQ (M41) is very similar to LIQ (B1648), whereas the proline (P) in KIP (QX) and MIP (793B) reduces the flexibility of the loop.

Structural analysis of the RBD of IBV suggests that the receptor binding sites for M41 and QX are positioned at different sides of the RBD (Fig. 6B, C and D). Previous *in silico* predictions of the interaction with alpha-2,3-linked sialic acid ligands in M41-RBD pointed towards three amino acids S87, N144 and T162, which are in close proximity four essential N-glycosylation sites (N33, N59, N85 and N160) (17). Although the amino acid sequence of S87, N144 and T162 is conserved between M41 and QX, one of the essential N-glycosylation sites is at a different position (N59 in M41, N58 in QX). This may result in a different conformation of the ligand binding site, thereby preventing in the QX-RBD wildtype

binding to the M41 ligand, which was supported by experimental evidence using chimeric M41-RBD protein where this glycosylation site was replaced, resulting in loss of binding to trachea tissue (data not shown). Furthermore, in the publication where the cryo-EM structure of M41 was resolved, the loop consisting of amino acids present in HVR 2 of the spike was proposed to be required for receptor binding (20). Our data points towards involvement of the unglycosylated loop containing HVR 2 for recognition of the QX glycan ligand but not the M41 ligand. Furthermore, the cryo-EM structure of the M41-CTD predicts other putative receptor-binding motif loops in M41 spike (20). In figure 1, we demonstrated no binding to trachea and kidney tissue was observed using our recombinantly expressed M41-CTD, in contrast to their published results. As binding of QX-RBD reflects the tissue tropism of QX infected birds, we speculate whether these loops (in the CTD) are necessary for initial receptor recognition and involved in QX infection.

In conclusion, we demonstrated that IBV-QX binding to chicken trachea and kidney is dependent on a sialylated glycan receptor and that amino acids in HVR 2 of the QX-RBD are critical for this receptor binding profile. This knowledge adds to our understanding of differences in tissue tropism between IBV strains *in vivo* and may contribute in designing new antivirals to prevent coronavirus infections in the field.

Materials and Methods:

Construction of the expression plasmids: The expression plasmids containing the codon optimized M41-ED (amino acids 19-1091 (21)), M41-S1 (amino acids 19-532 (19)), M41-RBD (amino acids 19-272 (15)) and M41-CTD (amino acids 273-532 (15) accession number AY851295) sequence followed by a trimerization domain (GCN4) and strep-tag (ST) was described previously (15). The codon-optimized sequence of QX-RBD (amino acids 19-275, accession number AFJ11176), containing upstream *Nhe*I and downstream *Pac*I restriction sites, was obtained from GenScript and cloned into the pCD5 expression vector by restriction digestion, as previously described (19). Fragments to generate chimeric RBD proteins were

created by splice overlap extension PCR using primers in Table 2 and cloned into pJET vector (Thermo Scientific, USA). The sequences were verified by automated nucleotide sequencing (Macrogen, The Netherlands) before cloning each fragment into the pCD5 expression vector. Mutations up to 9 nucleotides were introduced by site directed mutagenesis using primers listed in Table 2 and the sequences were subsequently verified by automated nucleotide sequencing (Macrogen, The Netherlands).

Production of recombinant proteins: Recombinant RBD proteins were produced in human embryonic kidney (HEK293T) cells. In short, cells were transfected with pCD5 expression vectors using polyethylenimine (PEI) at a 1:12 (wt/wt) ratio. Cell culture supernatants were harvested after 6 days. The recombinant proteins were purified using Strep-Tactin Sepharose beads as previously described (19). Proteins were pretreated (where indicated) with PNGaseF (New England Biolabs, USA) according to manufacturer's protocol before analysis by Western blot using Strep-tactin HRP antibody (IBA, Germany).

Circular Dichroism (CD): Recombinant IBV-RBD proteins were exchanged into buffer containing 10 mM sodium phosphate, pH 7.75 and diluted to 0.06 mg/ml. CD spectra were collected on a JASCO J-810 spectropolarimeter with a Peltier thermostated fluorescence temperature controller module by accumulating 4 scans from 285-190 nm with a scanning speed of 10 nm/min, Digital Integrated Time 1 second, bandwidth 1 nm, and standard sensitivity at 25 ºC. A thermal melt was done from 25 ºC to 95 ºC with a ramp rate of 1 ºC per minute. A full CD scan was collected at 95 ºC. After lowering the temperature to 25 ºC, the protein was allowed to refold for 20 minutes at 25 ºC, and a third CD scan was taken at 25 ºC to measure recovery. Secondary structure calculations for the CD data collected at 25 ºC before the thermal melt were processed by Dichroweb (23) using the CDSSTR (32), Selcon3 (33), and Contill (34) algorithms with protein reference set 7. Results from the 3 algorithms were averaged and plotted in Fig. 2C.

ELISA: Neu5Acα2-3Galβ1-3GlcNAc-PAA (Lectinity Holdings, Russia) was coated in a 96-well Nunc MaxiSorp plate (Sigma-Aldrich, Germany) at 0.5 µg/well overnight at 4 °C, followed by blocking with 3% bovine serum albumin (BSA; Sigma, Germany) in PBS–0.1%

Tween20 overnight. RBD proteins were pre-incubated with Strep-tactin HRP antibody (IBA, Germany) (1:200) for 30 min on ice. Indicated protein amounts were diluted in PBS and applied onto the coated well, followed by incubation for 2 h at room temperature. TMB (3,3=,5,5=-tetramethylbenzidine; Thermo Scientific, Netherlands) substrate was used to visualize binding, after which the reaction was terminated using 1 M H2SO4. The optical density at 450 nm was measured in a FLUOstar Omega instrument (BMG Labtech), and MARS data analysis software was used for data analysis. Statistical analysis was performed using a two-way analysis of variance (ANOVA).

Protein histochemistry: Protein histochemistry was performed as previously described (19). Recombinant proteins pre-complexed with Strep-tactin HRP antibody (IBA, Germany) were applied onto 4 µm sections of formalin-fixed paraffin-embedded healthy chicken tissues at 100 µg/ml (for RBD, S1 and ED in equal molar amount) (24) and binding was visualized using 3-amino-9-ethyl-carbazole (AEC; Sigma-Aldrich, Germany). Where indicated RBD proteins were pre-incubated with 100 µg/ml Neu5Acα2-3Galβ1-3GlcNAc (Lectinity Holdings, Russia) for 30 min on ice, before application onto the tissues. Pre-treatment of tissues was performed using 2 mU of Neuraminidase (sialidase) from *Arthrobacter ureafaciens* (AUNA) (Sigma, Germany) in 10 mM potassium acetate and 2.5 mg/ml Triton X-100, pH 4.2, and incubated at 37 °C overnight, before protein application.

Acknowledgments: The authors would like to thank Geert de Vrieze and Isa Feenstra for technical support and Marius Dwars for help with histopathological analyses of tissue slides. Monique H. Verheije is a recipient of a MEERVOUD grant from the NWO and Robert P. de Vries is a recipient of an ERC starting grant and a Beijerinck Premium of the Royal Dutch Academy of Sciences (KNAW).

365 **References**

366 1. Peighambari SM, Julian RJ, Gyles CL. 2000. Experimental Escherichia coli
367 respiratory infection in broilers. Avian Dis 44:759-69.

368 2. de Wit JJ, Cazaban C, Dijkman R, Ramon G, Gardin Y. 2018. Detection of different
369 genotypes of infectious bronchitis virus and of infectious bursal disease virus in European
370 broilers during an epidemiological study in 2013 and the consequences for the diagnostic
371 approach. Avian Pathol 47:140-151.

372 3. Valastro V, Holmes EC, Britton P, Fusaro A, Jackwood MW, Cattoli G, Monne I. 2016.
373 S1 gene-based phylogeny of infectious bronchitis virus: An attempt to harmonize virus
374 classification. Infect Genet Evol 39:349-364.

375 4. Shen CI, Wang CH, Liao JW, Hsu TW, Kuo SM, Su HL. 2010. The infection of
376 primary avian tracheal epithelial cells with infectious bronchitis virus. Vet Res 41:6.

377 5. Abd El Rahman S, El-Kenawy AA, Neumann U, Herrler G, Winter C. 2009.
378 Comparative analysis of the sialic acid binding activity and the tropism for the respiratory
379 epithelium of four different strains of avian infectious bronchitis virus. Avian Pathol 38:41-5.

380 6. Armesto M, Cavanagh D, Britton P. 2009. The replicase gene of avian coronavirus
381 infectious bronchitis virus is a determinant of pathogenicity. PLoS One 4:e7384.

382 7. Bande F, Arshad SS, Omar AR, Hair-Bejo M, Mahmuda A, Nair V. 2017. Global
383 distributions and strain diversity of avian infectious bronchitis virus: a review. Anim Health
384 Res Rev 18:70-83.

385 8. Gough RE, Cox WJ, Welchman DdB, Worthington KJ, Jones RC. 2008. Chinese QX
386 strain of infectious bronchitis virus isolated in the UK. Veterinary Record 162:99-100.

387 9. Benyeda Z, Szeredi L, Mato T, Suveges T, Balka G, Abonyi-Toth Z, Rusvai M, Palya
388 V. 2010. Comparative histopathology and immunohistochemistry of QX-like, Massachusetts

389 and 793/B serotypes of infectious bronchitis virus infection in chickens. J Comp Pathol
390 143:276-83.

391 10. Raj GD, Jones RC. 1997. Infectious bronchitis virus: Immunopathogenesis of
392 infection in the chicken. Avian Pathol 26:677-706.

393 11. Villarreal LY, Brandao PE, Chacon JL, Saidenberg AB, Assayag MS, Jones RC,
394 Ferreira AJ. 2007. Molecular characterization of infectious bronchitis virus strains isolated
395 from the enteric contents of Brazilian laying hens and broilers. Avian Dis 51:974-8.

396 12. Yudong W, Yonglin W, Zichun Z, Genche F, Yihai J, Xiange L, Jiang D, Shushuang
397 W. 1998. Isolation and identification of glandular stomach type IBV (QX IBV) in chickens, vol

398 13. Yu L, Jiang Y, Low S, Wang Z, Nam SJ, Liu W, Kwangac J. 2001. Characterization of
399 three infectious bronchitis virus isolates from China associated with proventriculus in
400 vaccinated chickens. Avian Dis 45:416-24.

401 14. Liu S, Kong X. 2004. A new genotype of nephropathogenic infectious bronchitis virus
402 circulating in vaccinated and non-vaccinated flocks in China. Avian Pathol 33:321-7.

403 15. Promkuntod N, van Eijndhoven RE, de Vrieze G, Grone A, Verheije MH. 2014.
404 Mapping of the receptor-binding domain and amino acids critical for attachment in the spike
405 protein of avian coronavirus infectious bronchitis virus. Virology 448:26-32.

406 16. Leyson C, Franca M, Jackwood M, Jordan B. 2016. Polymorphisms in the S1 spike
407 glycoprotein of Arkansas-type infectious bronchitis virus (IBV) show differential binding to
408 host tissues and altered antigenicity. Virology 498:218-225.

409 17. Parsons LM, Bouwman KM, Azurmendi H, de Vries RP, Cipollo JF, Verheije MH.
410 2019. Glycosylation of the viral attachment protein of avian coronavirus is essential for host
411 cell and receptor binding. J Biol Chem 294:7797-7809.

412 18. Casais R, Dove B, Cavanagh D, Britton P. 2003. Recombinant avian infectious
413 bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is
414 a determinant of cell tropism. J Virol 77:9084-9.

415 19. Wickramasinghe IN, de Vries RP, Grone A, de Haan CA, Verheije MH. 2011. Binding
416 of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. J
417 Virol 85:8903-12.

418 20. Shang J, Zheng Y, Yang Y, Liu C, Geng Q, Luo C, Zhang W, Li F. 2018. Cryo-EM
419 structure of infectious bronchitis coronavirus spike protein reveals structural and functional
420 evolution of coronavirus spike proteins. PLoS Pathog 14:e1007009.

421 21. Promkuntod N, Wickramasinghe IN, de Vrieze G, Grone A, Verheije MH. 2013.
422 Contributions of the S2 spike ectodomain to attachment and host range of infectious
423 bronchitis virus. Virus Res 177:127-37.

424 22. Niesters HG, Lenstra JA, Spaan WJ, Zijderveld AJ, Bleumink-Pluym NM, Hong F, van
425 Scharrenburg GJ, Horzinek MC, van der Zeijst BA. 1986. The peplomer protein sequence of
426 the M41 strain of coronavirus IBV and its comparison with Beaudette strains. Virus Res
427 5:253-63.

428 23. Whitmore L, Wallace BA. 2008. Protein secondary structure analyses from circular
429 dichroism spectroscopy: methods and reference databases. Biopolymers 89:392-400.

430 24. Wickramasinghe IN, de Vries RP, Eggert AM, Wandee N, de Haan CA, Grone A,
431 Verheije MH. 2015. Host tissue and glycan binding specificities of avian viral attachment
432 proteins using novel avian tissue microarrays. PLoS One 10:e0128893.

433 25. Bouwman KM, Delpont M, Broszeit F, Berger R, Weerts E, Lucas MN, Delverdier M,
434 Belkasmi S, Papanikolaou A, Boons GJ, Guerin JL, de Vries RP, Ducatez MF, Verheije MH.
435 2019. Guinea fowl coronavirus diversity has phenotypic consequences for glycan and tissue
436 binding. J Virol 93.

437 26. Ambepitiya Wickramasinghe IN, de Vries RP, Weerts EA, van Beurden SJ, Peng W,
438 McBride R, Ducatez M, Guy J, Brown P, Eterradossi N, Grone A, Paulson JC, Verheije MH.
439 2015. Novel receptor specificity of avian gammacoronaviruses that cause enteritis. J Virol
440 89:8783-92.

441 27. Schultze B, Krempl C, Ballesteros ML, Shaw L, Schauer R, Enjuanes L, Herrler G.
442 1996. Transmissible gastroenteritis coronavirus, but not the related porcine respiratory
443 coronavirus, has a sialic acid (N-glycolylneuraminic acid) binding activity. J Virol 70:5634-7.

444 28. Liu C, Tang J, Ma Y, Liang X, Yang Y, Peng G, Qi Q, Jiang S, Li J, Du L, Li F. 2015.
445 Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. J Virol 89:6121-5.

446 29. Bakkers MJ, Lang Y, Feitsma LJ, Hulswit RJ, de Poot SA, van Vliet AL, Margine I, de
447 Groot-Mijnes JD, van Kuppeveld FJ, Langereis MA, Huizinga EG, de Groot RJ. 2017.
448 Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase
449 lectin activity. Cell Host Microbe 21:356-366.

450 30. Peng W, de Vries RP, Grant OC, Thompson AJ, McBride R, Tsogtbaatar B, Lee PS,
451 Razi N, Wilson IA, Woods RJ, Paulson JC. 2017. Recent H3N2 viruses have evolved
452 specificity for extended, branched human-type receptors, conferring potential for increased
453 avidity. Cell Host Microbe 21:23-34.

454 31. Broszeit F, Tzarum N, Zhu X, Nemanichvili N, Eggink D, Leenders T, Li Z, Liu L,
455 Wolfert MA, Papanikolaou A, Martinez-Romero C, Gagarinov IA, Yu W, Garcia-Sastre A,
456 Wennekes T, Okamatsu M, Verheije MH, Wilson IA, Boons GJ, de Vries RP. 2019. N-
457 Glycolylneuraminic acid as a receptor for Influenza A viruses. Cell Rep 27:3284-3294.e6.

458 32. Manavalan P, Johnson WC, Jr. 1987. Variable selection method improves the
459 prediction of protein secondary structure from circular dichroism spectra. Anal Biochem
460 167:76-85.

461 33. Sreerama N, Woody RW. 1993. A self-consistent method for the analysis of protein
462 secondary structure from circular dichroism. Anal Biochem 209:32-44.

463 34. van Stokkum IH, Spoelder HJ, Bloemendal M, van Grondelle R, Groen FC. 1990.
464 Estimation of protein secondary structure and error analysis from circular dichroism spectra.
465 Anal Biochem 191:110-8.

466

467 **Figure legends:**

468 **FIG 1** Binding of M41-spike proteins to paraffin-embedded healthy chicken trachea and
469 kidney tissue. A) Schematic representation of spike proteins, M41-ectodomain (ED, amino
470 acids 1-1072), M41-S1 (amino acids 1-513), M41-RBD (amino acids 1-254) and M41-CTD
471 (amino acids 254-513) followed by a trimerization domain (GCN4) and strep-tag (ST).
472 Numbering starts at 1 of the mature protein sequence. B) Protein binding of M41 spike
473 proteins observed in the trachea and kidney is visualized by red staining. C) Affinity of M41
474 spike proteins for the known ligand (Neu5Acα2-3Galβ1-3GlcNAc) in solid phase ELISA.
475 Where at all protein amounts a significant difference of at least $p < 0.01$ was observed
476 between M41-RBD, -S1, -ED and M41-CTD and TCoV-S1 which served as a negative
477 control tested in two-way ANOVA. Elisa was performed in triplicate where average and
478 standard deviations are shown.

479

480 **FIG 2** IBV M41- and QX-RBD protein analysis. A) Amino acid alignment of M41-RBD (amino
481 acids 19-272, accession number AY851295) and amino acids 19-275 (accession number
482 AFJ11176) of the QX spike. Numbering starts at 1 of the mature protein sequence (signal
483 sequence not shown). Dots indicate identical amino acids. Grey highlights, surrounded by
484 black box indicate previously identified hyper variable regions of IBV-Mass (22). Green
485 highlights indicate very different residues. B) M41- and QX-RBD with and without pre-
486 treatment of PNGaseF analyzed by Western blot using Strep-Tactin HRP antibody. C)

Percentage of secondary protein structures calculated based on CD analysis of M41- and QX-RBD. D) Binding of QX-RBD to paraffin-embedded healthy chicken trachea and kidney visualized by red staining in protein histochemistry.

FIG 3 Avidity and affinity of M41- and QX-RBD for host factors. A) Protein histochemistry of M41- and QX-RBD onto paraffin-embedded chicken trachea and kidney tissue (left column), upon pre-incubation of proteins with Neu5Acα2-3Galβ1-3GlcNAc (middle column) or pre-treatment of tissues with *Arthrobacter ureafaciens* neuraminidase (right column). B) Affinity of RBD proteins for Neu5Acα2-3Galβ1-3GlcNAc in ELISA where **= p< 0.01, ****= p<0.001 tested in two-way ANOVA. TCoV-S1 was used as a negative control in equal molar amounts. ELISA was performed in triplicate with all proteins; average is shown with standard deviations.

FIG 4 Chimeric RBD protein binding to chicken tissues. A) Schematic representation of chimeric RBD proteins, grey box indicates M41- and white box QX wildtype sequence. Amino acids surrounding the transitions between the different domains of the chimeric proteins are indicated including the amino acid number of the wildtype sequence. B) Binding of chimeric RBDs to trachea and kidney tissue in protein histochemistry. C) Affinity of 37.5 nmol chimeric RBD proteins in ELISA for Neu5Acα2-3Galβ1-3GlcNAc. Significant differences are *= p<0.05, **= p<0.01, ****= p<0.001, tested in two-way ANOVA. ELISA was performed with all chimeric RBD proteins in triplicates.

FIG 5 Identification of amino acids involved in IBV kidney binding using chimeric proteins. A) Schematic representation of chimeric RBD proteins, grey box indicates M41- and white box QX wildtype sequence. Numbers above indicate the positions of the amino acid triplicates swapped between M41 and QX. B) Binding of chimeric RBDs to trachea and kidney tissue in protein histochemistry. C) Binding of chimeric RBD proteins (37.5nmol) in ELISA to Neu5Acα2-3Galβ1-3GlcNAc. Significant differences are *= p<0.05, **= p<0.01, tested in two-way ANOVA. ELISA was performed with all chimeric RBD proteins in triplicates.

513 **FIG 6** Model of IBV spike with predicted receptor binding sites. A) Structural alignment
514 overlay of QX-RBD (blue ribbon) onto M41-RBD (green ribbon), based on PDB entry 6cv0
515 (20) using Swiss-model. Detailed representation of the receptor binding site identified for QX-
516 RBD; indicated in yellow sticks are amino acids 100-102 SGS in QX-RBD and in beige 99 Y
517 in M41-RBD. 110-112 KIP of QX-RBD is indicated with dark blue sticks, whereas light blue
518 represents 107-109 MLQ in M41-RBD. B) Amino acids involved in receptor binding of IBV.
519 Blue ribbon represents the modeled QX-RBD structure with amino acids 100-102 (SGS) as
520 yellow spheres and 110-112 (KIP) as blue spheres. The green ribbon represents the M41-
521 RBD with 99 (Y) as beige spheres and 107-109 (MLQ) as light blue spheres. Amino acids in
522 red spheres (S87, N144 and T162) are previously predicted to be involved in alpha-2,3-
523 linked sialic acid binding of M41-RBD (17). C) Surface representation of the trimeric M41
524 spike cryo-EM structure (20). S2 is in dark gray for all monomers. S1 is in light gray with one
525 S1 monomer colored bright green for the RBD domain and pale green for the CTD. Amino
526 acids involved in ligand binding are highlighted: yellow is 99Y (100-102 SGS in QX), dark
527 blue is 107-109MLQ (110-112 KIP in QX) and red is S87, N144 and T162. D) Modeled QX
528 spike based on PDB entry 6cv0, colors as indicated in C, except the S1 of QX is blue, and
529 the RBD in bright blue. Representations on right of C and D are structures turned 90 degrees
530 towards the viewer. All representations were made using PyMol viewer.

531 **TABLE 1** Relative binding of IBV-RBD proteins to paraffin-embedded healthy chicken tissues

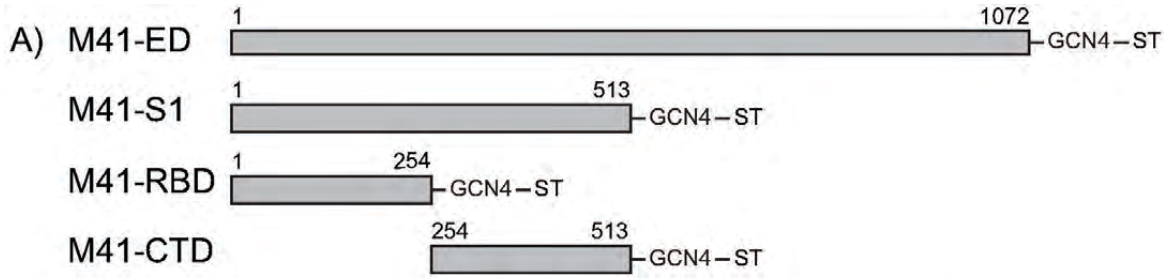
532 white, no visible staining; +- = staining of few cells; + = staining of most epithelial cells

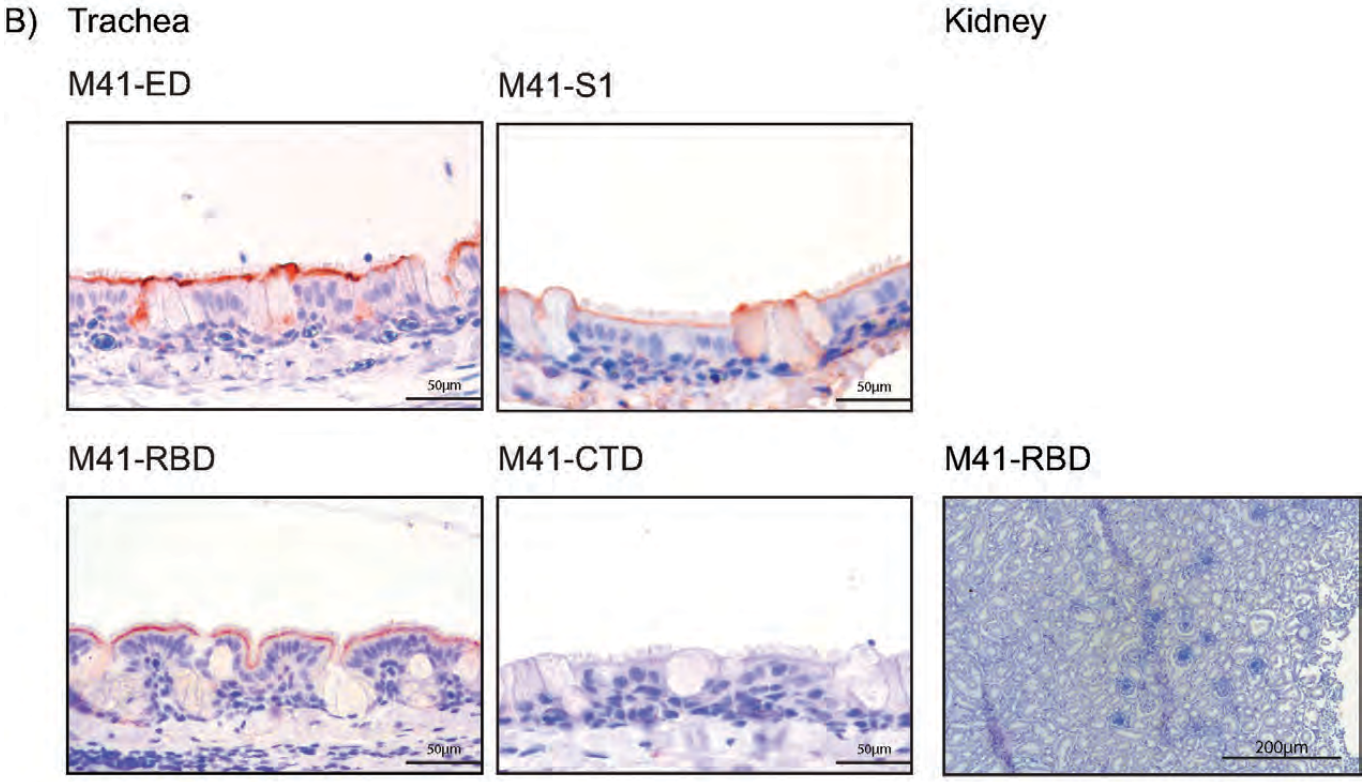
533 **TABLE 2** Primer sequences to create chimeric RBD plasmids

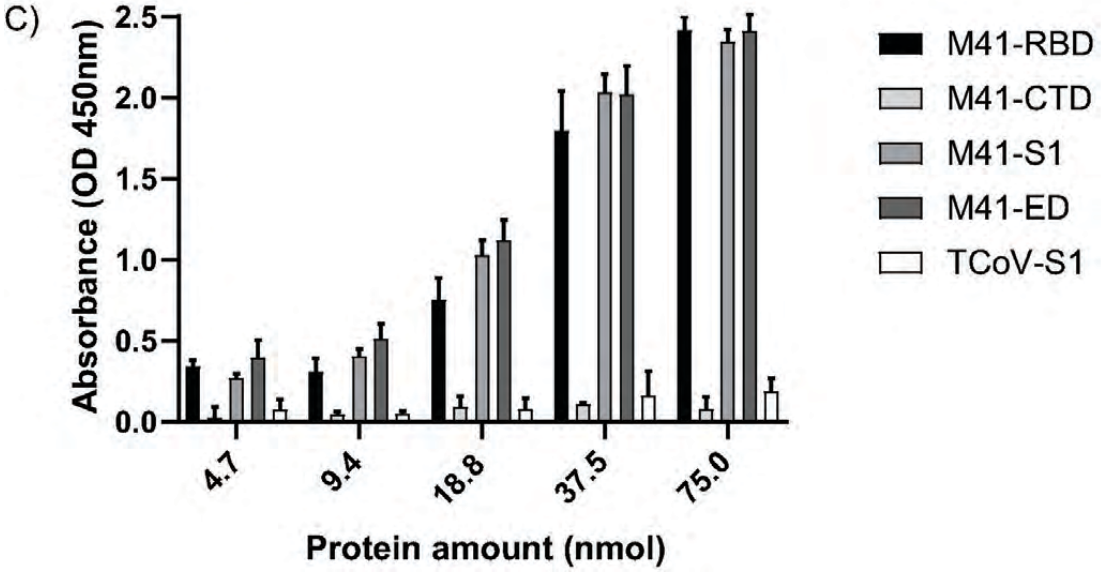
534 R= nt A/G K= nt G/T M=nt A/C Y=nt C/T W= nt A/T S= nt C/G

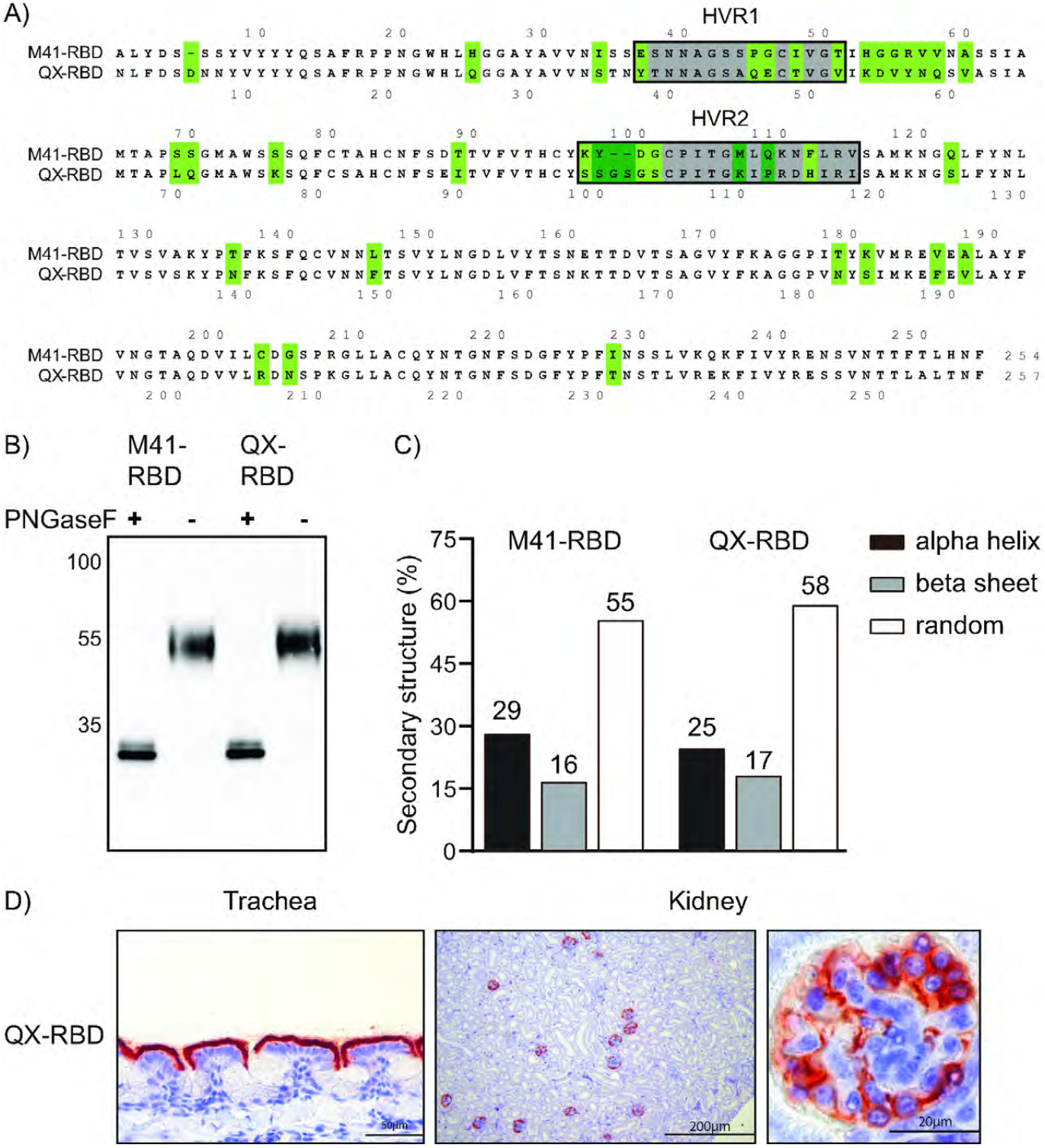
535 Underlined sequences indicate nucleotides changed to introduce the mutation

536

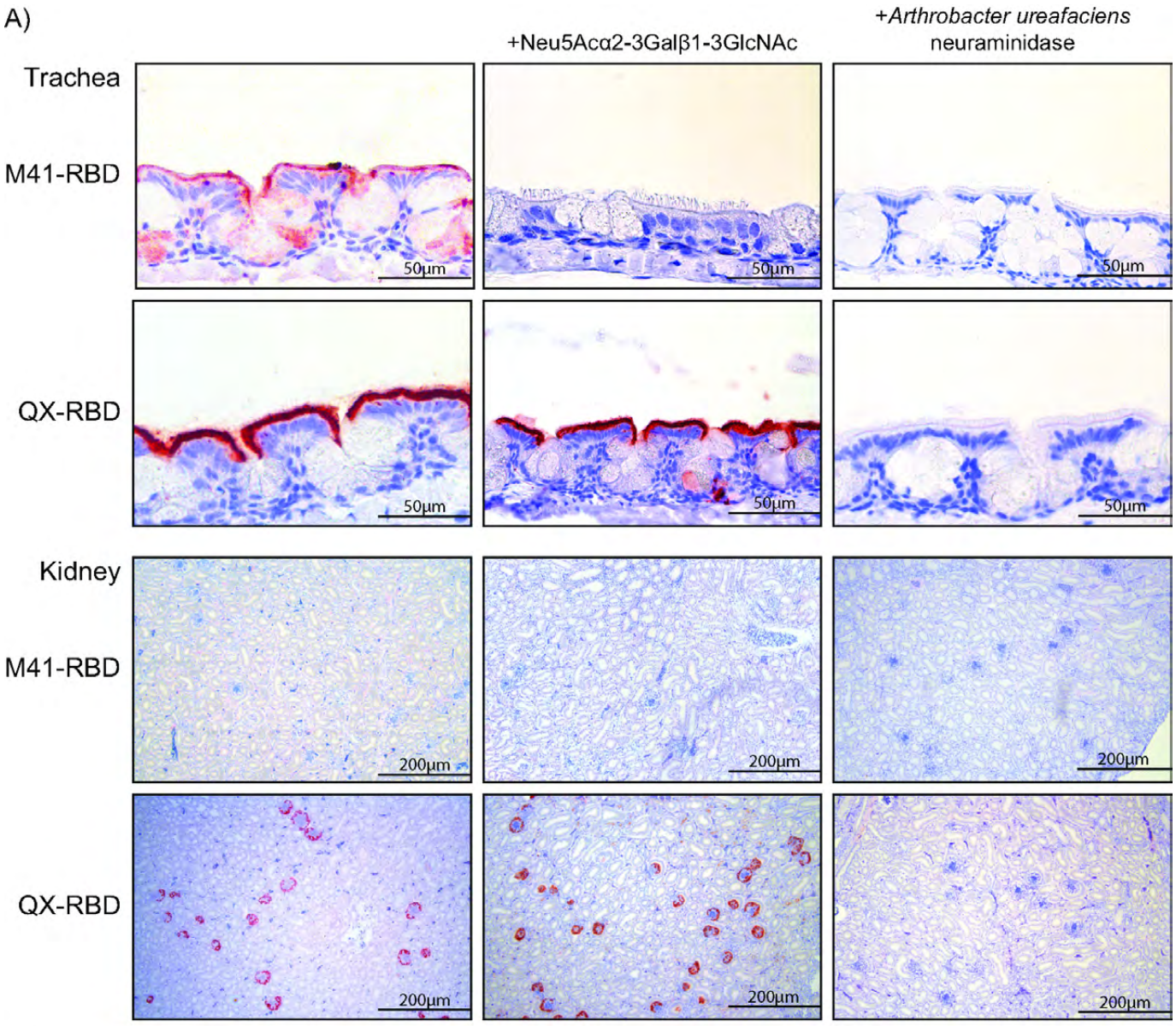


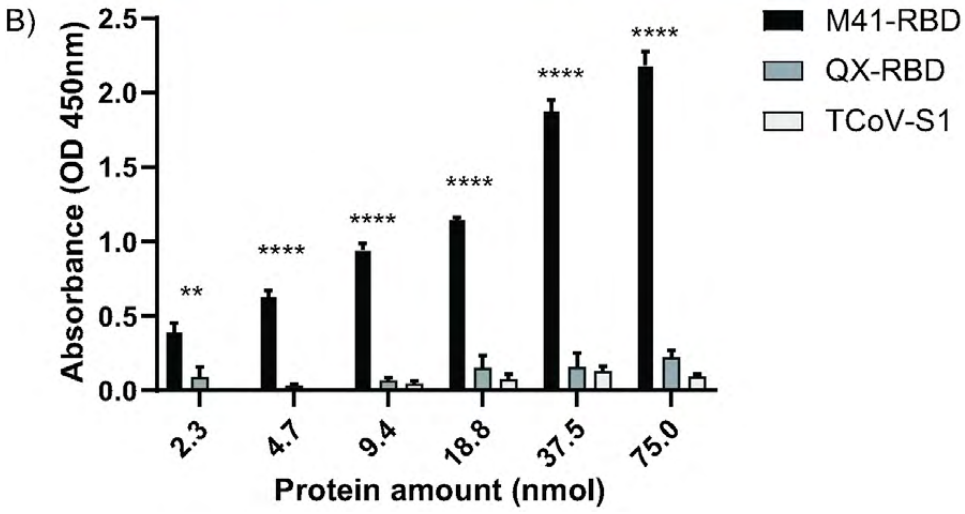


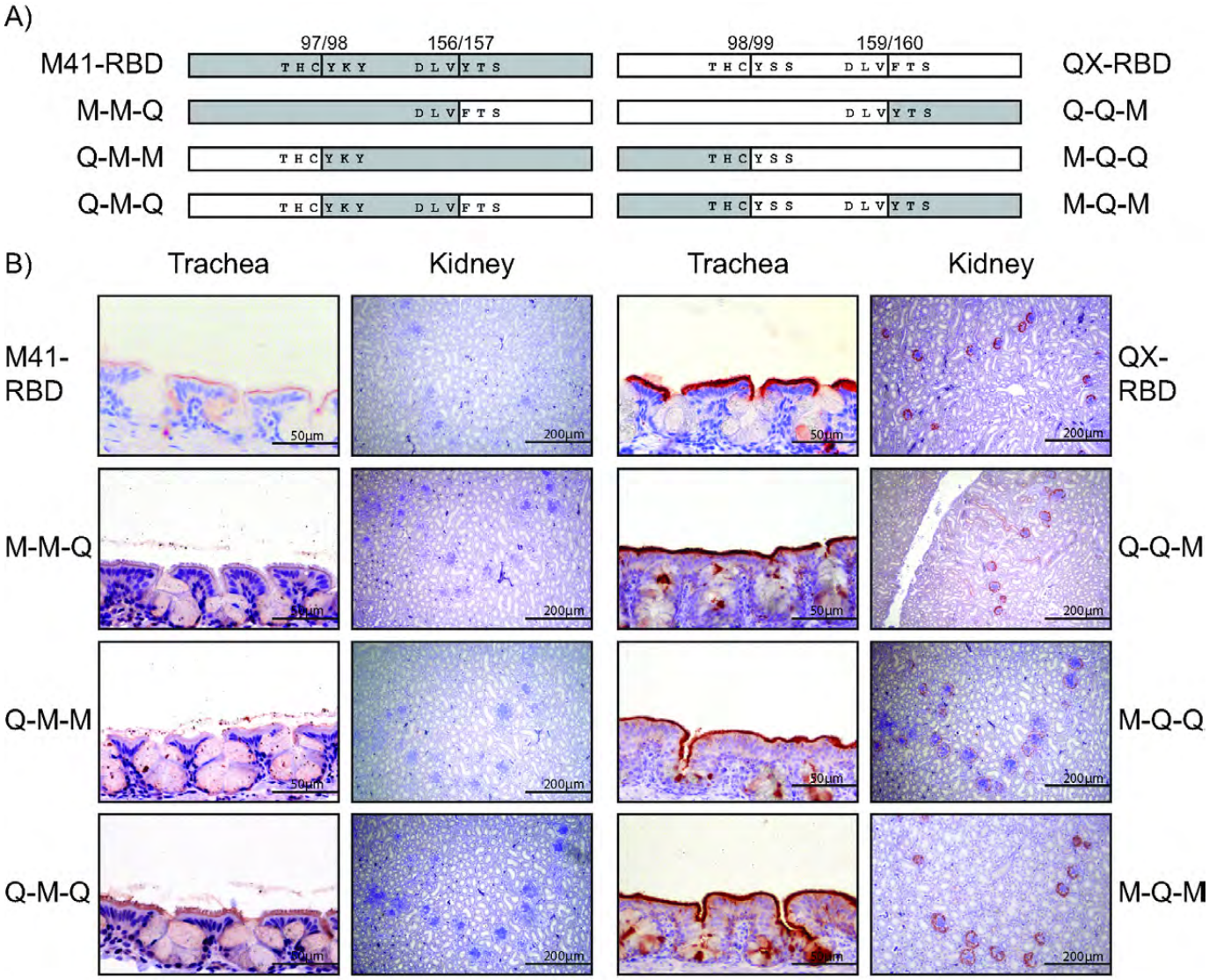


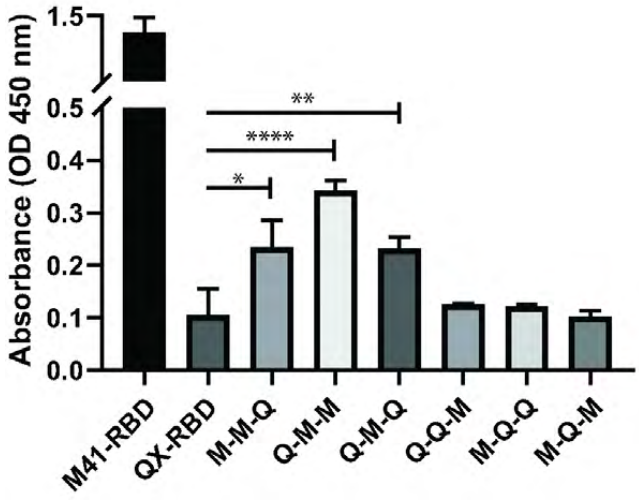


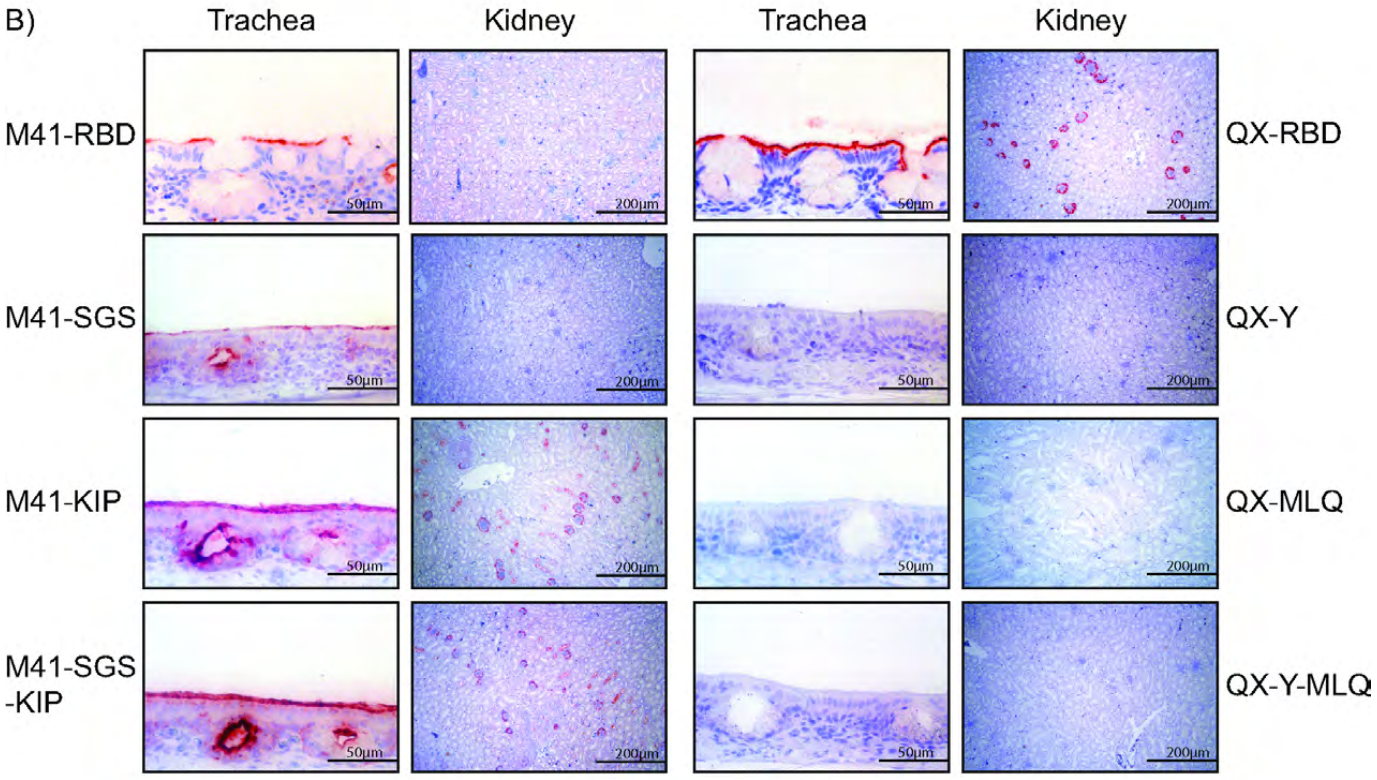
Journal of Virology

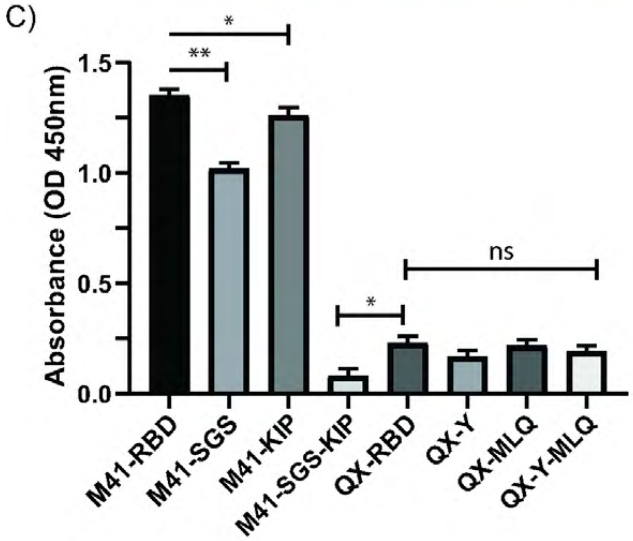












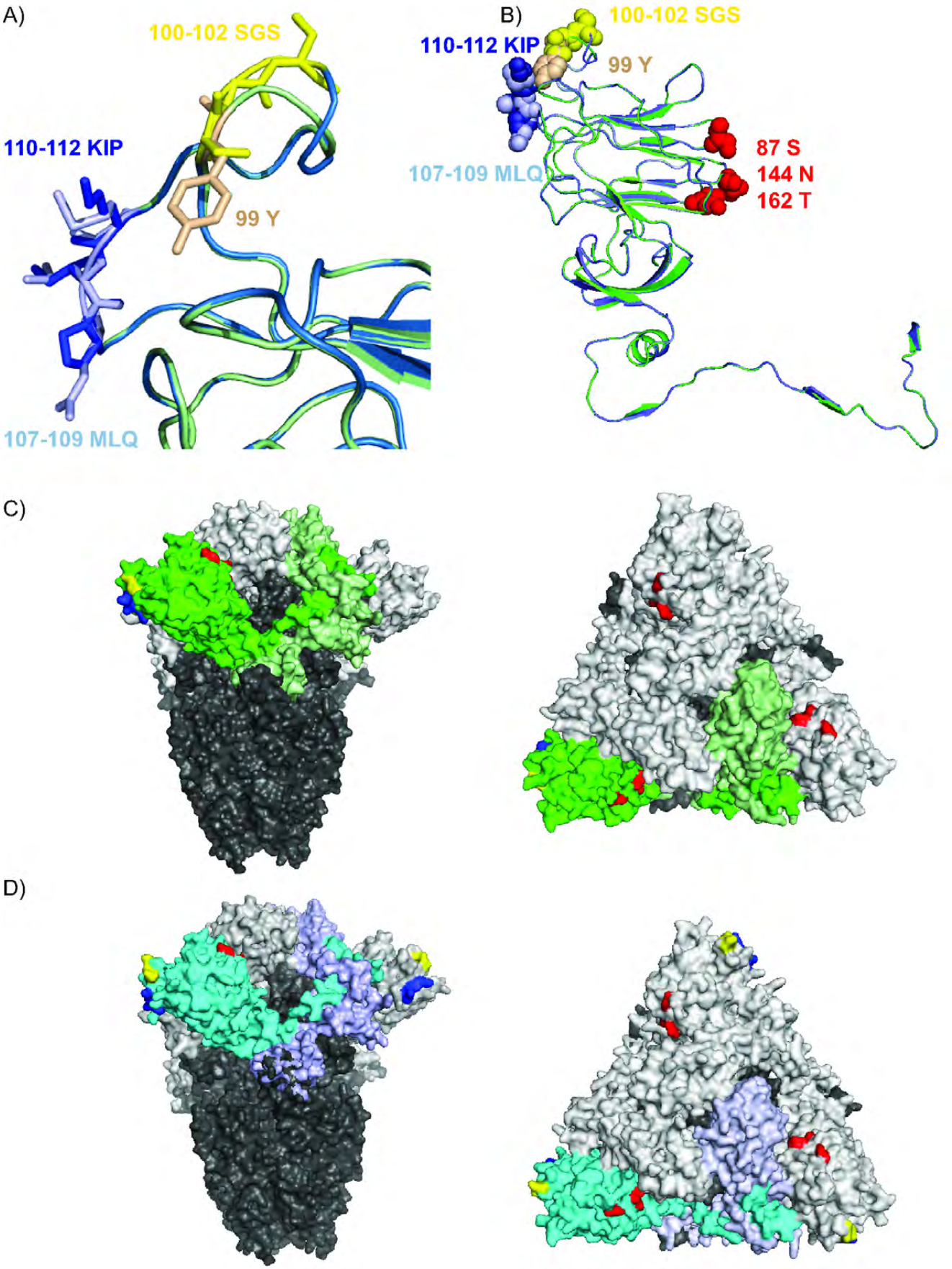


TABLE 1 Relative binding of IBV-RBD proteins to paraffin-embedded healthy chicken tissues

	QX-RBD	M41-RBD
Nostril		
Proximal trachea	+	+
Distal trachea	+	+
Lung		
Esophagus		
Gizzard	+	
Proventiculus		
Duodenum		
Ileum	+	
Colon	+	+
Cecal tonsil	+	+
Spleen		
Liver	+-	
Adrenal glands		
Pancreas		
Kidney	+	
Ureter	+	+
Heart		
Skin		
Conjunctiva	+	+
Muscles		
Ovary		
Oviduct	+	+
Cloaca	+-	
Cerebrum		
Cerebellum		
Brain stem		
Sciatic nerve		

white, no visible staining; +- = staining of few cells; + = staining of most epithelial cells

TABLE 2 Primer sequences to create chimeric RBD plasmids

	Original plasmid	Forward primer	Reverse primer
MMQ	M41-RBD	gtcgcttccgtgctagca	acaccagrtckccgttcag
	QX-RBD	ctgaacggmgayctggtgt	ctgcttcatgcgcttaattaa
		gtcgcttccgtgctagca	ctgcttcatgcgcttaattaa
MQM	MQQ	gtcgcttccgtgctagca	acaccagrtckccgttcag
	M41-RBD	ctgaacggmgayctggtgt	ctgcttcatgcgcttaattaa
		gtcgcttccgtgctagca	ctgcttcatgcgcttaattaa
MQQ	M41-RBD	gtcgcttccgtgctagca	tagcaatgwgtsacgaacactg
	QX-RBD	cagtgttcgtsacwcattgcta	ctgcttcatgcgcttaattaa
		gtcgcttccgtgctagca	ctgcttcatgcgcttaattaa
QQM	QX-RBD	gtcgcttccgtgctagca	acaccagrtckccgttcag
	M41-RBD	ctgaacggmgayctggtgt	ctgcttcatgcgcttaattaa
		gtcgcttccgtgctagca	ctgcttcatgcgcttaattaa
QMQ	QMM	gtcgcttccgtgctagca	acaccagrtckccgttcag
	QX-RBD	ctgaacggmgayctggtgt	ctgcttcatgcgcttaattaa
		gtcgcttccgtgctagca	ctgcttcatgcgcttaattaa
QMM	QX-RBD	gtcgcttccgtgctagca	tagcaatgwgtsacgaacactg
	M41-RBD	cagtgttcgtsacwcattgcta	ctgcttcatgcgcttaattaa
		gtcgcttccgtgctagca	ctgcttcatgcgcttaattaa
M41-SGS	M41-RBD	atctgatggatgtcccatcacc	ccggacttatagcaatgtgtcacg
M41-KIP	M41-RBD	tcctaagaactttctgcgggtgtc	atcttgccggtgatgggacatcc
M41-SGS-KIP	M-SGS	tcctaagaactttctgcgggtgtc	atcttgccggtgatgggacatcc
QX-Y	QX-RBD	ctgtcccatcaccggcaag	tacccggagctgtagcaatg
QX-MLQ	QX-RBD	gcagcgggaccacatcagaatttc	agcatgccggtgatgggacaaga
QX-Y-MLQ	Q-Y	gcagcgggaccacatcagaatttc	agcatgccggtgatgggacaaga

R= nt A/G K= nt G/T M=nt A/C Y=nt C/T W= nt A/T S= nt C/G

Underlined sequences indicate nucleotides changed to introduce the mutation