Fabijan Božajić
Drug Innovation
Utrecht University
Research report - Master internship

Role of the virus-glycan interaction in the infection mechanism of coronaviruses

Abstract

This study explores how human coronaviruses (CoVs), namely MERS-CoV, HCoV-HKU1, SARS-CoV-1, and SARS-CoV-2, infect cells, focusing on the role of their spike (S) proteins and interactions with cell surface receptors. It aims to understand how CoVs use both proteinaceous and sialylated glycan receptors to attach to cells. To safely study these interactions, we used pseudoviruses which are non-infectious and non-replicative particles that express S proteins.
Initial luciferase assay showed no infection from pseudoviruses, including SARS-CoV-2. Due to issues like variability in results and potential mismatches in S protein domains, the assay was repeated. After improving the preparation methods and ensuring the S1 and S2 protein regions were homogenized and compatible, we observed infection by SARS-CoV-2 pseudoviruses in HEK ACE2 TMPRSS2 cells, which was expectedly lost in RD cells. However, pseudoviruses for MERS-CoV and HCoV-HKU1 still failed to infect either cell line.
Further testing of S protein domains-cell line interactions using FACS showed that while some S protein regions could bind certain cell lines, it was never observed for both domains of the same CoV simultaneously which might explain the absence of infection in the luciferase assay. A key factor may also be the incompatibility of the S protein's transmembrane domain (TMD), which helps stabilize the S protein. Mismatched TMDs could disrupt the S protein's function. The next step is to improve pseudovirus design by matching TMDs with other S protein regions. Achieving successful infection will allow us to study how different receptors contribute to viral attachment and entry, advancing strategies to fight CoV infections.

Layman's summary

Coronaviruses (CoVs) are a group of viruses that can infect humans and cause diseases like the flu, pneumonia, and more severe respiratory illnesses. The most well-known coronaviruses are SARS-CoV-1, MERS-CoV, and the recent SARS-CoV-2, which was responsible for the COVID-19 pandemic. SARS-CoV-2 has proven especially challenging because of its high spread rate and impact on global health.
The spike (S) protein on the surface of these viruses is highly important for infection. The viruses use it to attach to cells in the body and enter them. The S protein is made up of two parts: one helps the virus stick to cells (called the receptor-binding domain, or RBD), and the other helps the virus fuse with the cell membrane to release its genetic material.
These viruses can use different "entryways" on the surface of the cells to enter. Some coronaviruses recognize specific proteins on the cell surface, while others, like certain human coronaviruses (HCoV-HKU1, HCoV-OC43), prefer to attach to sugar molecules (called glycans) on the cell surface. There are different types of glycans, which can affect the type of tissue the virus can infect.

For this reason, the interactions between the S protein of the virus and the proteins and sugars on the cell surface are interesting to researchers. Studies are conducted, including this one, to better understand how the virus infects cells, which could help in developing treatments and vaccines or in early detection of future pandemics.
For this study, we used pseudo viruses to safely study how the virus interacts with cells in the lab. Pseudo viruses are capsules that imitate the virus by containing the spike proteins, but they cannot multiply and cause disease. The goal of this project was to create these imitations of coronaviruses and confirm that they can attach and enter human cells so that in the future different cell types with various surface proteins and sugars can be tested. Identifying these surface proteins and sugars and their effect on infection is the first step in prevention and treatment of diseases caused by coronaviruses.

Introduction

There are seven coronaviruses (CoVs) that infect humans, with Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-1) and Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV) among the more recent outbreaks in 2002 and 2012, respectively (1). The most recent outbreak of SARS-CoV-2, starting in 2019, has reached pandemic proportions by infecting more than 737 million people around the world. Similarly to other coronaviruses that circulate in the human population, it causes upper and lower respiratory tract infections but can exhibit a wide range of clinical manifestations including fever, dry cough, and fatigue (2,3). While some CoVs have a higher burden of mortality, such as SARS-CoV-1 (approximately 10%) and MERS (approximately 40%), the SARS-CoV-2 pandemic was more severe and continues to pose a significant challenge for governments and population as a whole (2).

Coronaviruses are single-stranded, enveloped positive-sense RNA viruses from the *Coronaviridae* family. They are typically round in shape and include crown-like surface projections caused by their surface spike (S) proteins (Figure 1), which is a common feature of coronaviruses (4).

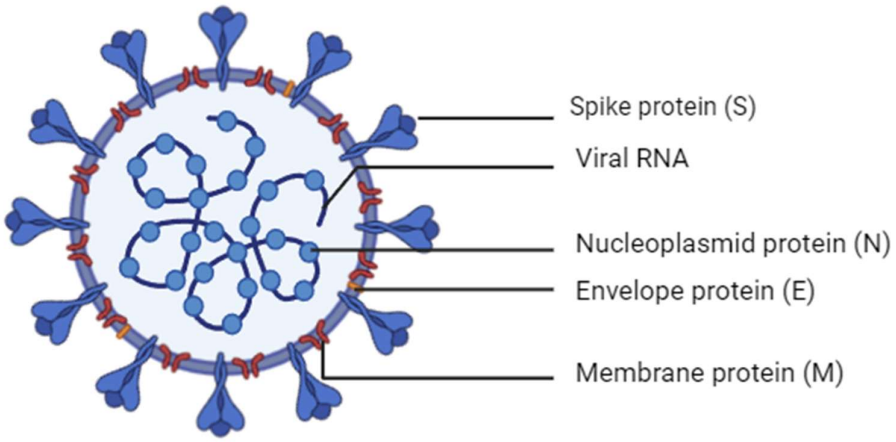


Figure 1: coronavirus structure with structural proteins: spike protein (S), membrane glycoprotein (M), envelope glycoprotein (E), and nucleoplasmid protein (N). (Figure from Biorender)

The role of the S protein is to mediate viral attachment and membrane fusion during infection by binding cell surface receptors.

The importance of the S protein is highlighted by the fact that it can mutate which has negative consequences for the host. This was shown most recently on the example of SARS-CoV-2 S protein mutations which were responsible for the emergence of new viral variants. The new variants had superior infectivity and transmissibility due to the higher affinity of the virus for the ACE2 receptor (5).

The CoV S protein is made up of two subunits and their domains. The S1 subunit contains a C-terminal receptor binding domain (RBD) and the N-terminal domain (NTD), among others, while the S2 subunit facilitates membrane fusion and contains a transmembrane domain (TMD). The TMD plays a vital role in the correct functioning of the S protein by fixating it in the glycoprotein membrane and ensuring the correct orientation and operation of the other domains (6). With the help of the RBD, the CoVs bind the proteinaceous cell-surface receptors, while the NTD is observed to contain a glycan binding cleft used in the previously demonstrated interaction of CoVs with sialogycans (7,8). Together, the S1 and S2 subunits form a clove-like structure of the S protein (Figure 2).

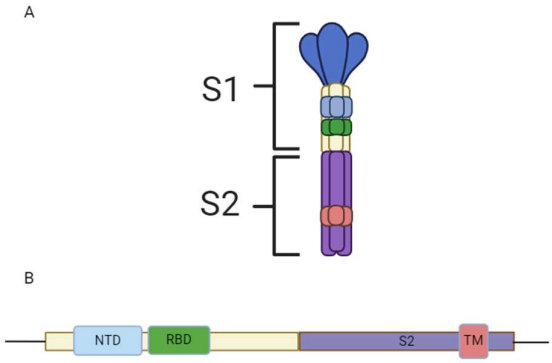


Figure 2: Coronavirus S protein structure with subunits (A) and simplified visualization of the S protein domains (B) (from Biorender).

We differentiate between two kinds of cell surface receptors in this report: proteinaceous and glycan cell surface receptors.
Glycans are sugars found everywhere on cell and protein surfaces. They are either short, branched chains or long chains attached to other cell surface proteins- proteoglycans. It was found that both proteoglycans and the glycans on the surfaces of host and virus proteins play an important role in how viruses attach to cells (9).
The types of cell-surface receptors bound by the spike protein differ among CoVs (Table 1). Among the CoVs that engage non-glycan receptors are human coronavirus NL63 (HCoV-NL63), HCoV- 229E, HCoV-HKU1, MERS-CoV, SARS-CoV-1, and SARS-CoV-2. Additionally, HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV-1, and SARS-CoV-2 use either a sialylated glycan as their primary receptor or have dual receptor binding capabilities. Consequently, the S protein plays a central role in the initial transmission, virus tropism, and ongoing infection of CoVs (9–11)

Cell-surface receptors used by CoVs		
CoV	**Proteinaceous receptor**	**Glycan receptor**
HCoV-NL63	Heparan sulfate proteoglycans (HSPG) (12)	
HCoV-229E	Aminopeptidase N (APN) (13)	
HCoV-OC43		5-N-acetyl-9-O-acetyl neuraminic acid (Neu5,9Ac2) (14)

HCoV-HKU1	Transmembrane protease, serine 2 (TMPRSS2) (15)	9-O-acetylated sialosides (15)
MERS-CoV	Dipeptidyl peptidase 4 (DPP4) (16)	9-O-α2,3-linked sialosides (17)
SARS-CoV-1	Angiotensin-converting enzyme 2 (ACE2) (18)	
SARS-CoV-2	ACE2 (18)	9- O-Acetylated α2-8-linked sialosides (7)

Table 1: Overview of some CoVs including their proteinaceous and glycan receptors

In addition to proteinaceous cell surface receptors, some coronaviruses, specifically embecoviruses such as HCoV-HKU1 and HCoV-OC43, primarily recognize glycans modified by N-terminal sialic acids as their host receptor (15,19). These sialylated glycans exist in various forms and different linkages, most are α-linked (20). While sialic acids exist in many different forms, their synthesis is tightly regulated by various factors that include tissue type, species, and physiological conditions. This, together with the glycan expression in different tissues plays a key role in determining the host range, targeted tissues and pathogenesis of CoVs (7,21). The role of sialylated glycans in determination of cell, tissue, and host tropism of CoVs can be supplemented by the fact that MERS-CoV, which infects the lower and upper respiratory tract, is preferential for α2,3-linked sialosides. This preference is likely determinative for MERS-CoV tissue and host tropism since α2,3-linked sialosides are highly expressed on the surface of human tracheal epithelial cells, i.e., the lower and upper respiratory tract. (22,23).

Interestingly, it was recently observed that CoV binding of sialylated glycans is rarely an exclusive event with a singular purpose in the context of viral attachment and entry mechanisms. For example, MERS-CoV has a dual receptor dependency which means it relies on simultaneous S protein binding of a proteinaceous and a glycan receptor as an entry and attachment receptor, respectively (17). Additionally, SARS-CoV-2 was observed to also contain a sialic acid binding site in the NTD of the S protein (24). Further experiments confirmed that SARS-CoV-2 binds α2,8-linked sialic acids in the NTD binding site for which the role of the interaction is not yet clear and requires further investigation but allows for future elucidation of the SARS-CoV-2 infection mechanism (7).
Similarly, binding of sialylated glycans has been shown to enable dual receptor binding capabilities in HCoV-HKU1. It was previously thought that HCoV-HKU1 binds sialylated glycans as primary and only cell surface receptors. However, it was recently shown that HCoV-HKU1 binding of α2,8-linked 9-O-acetylated disialosides initiates a conformational change in the virus' S protein, allowing for a formation of an allosteric opening. Consequently, the S protein gains the ability of binding secondary receptors. HCoV-HKU1 was therefore found to use transmembrane protease serine 2 (TMPRSS2) as a protein-based entry receptor, similarly to some other coronaviruses. (15,25). TMPRSS2 is typically found alongside other receptors, like ACE2, on the cell membrane. It plays a crucial role by cleaving the S protein, a process that facilitates the fusion of the viral and host membranes. As such, it plays a key role in in the infection mechanism of CoVs (26). Other coronaviruses have been observed to also rely on the proteolytic activity of the TMPRSS2 in membrane fusion, including SARS-CoV-1and MERS-CoV, among others (15).

Because of the known ability of CoVs to interact with sialic acids and utilize them to aid in attachment and infection, sialic acids and other sialosides have gained the interest of researchers in hopes to better understand the infection mechanism of these pathogenic viruses further and to potentially utilize them as pharmacological targets. Further research is also needed into the repertoire of sialylated glycans that CoV can engage since only a portion is known. The goal of this project is to elucidate this role and the intricate relationship between host glycans and viral

infections by conducting S protein binding studies with mammalian cells expressing various glycan and proteinaceous receptors.

Handling CoVs is dangerous, expensive, and otherwise complicated due to their categorization as biosafety level 3 pathogens (27). To address this, we have produced pseudoviruses which are non-replicative viral particles that do not envelop any infectious RNA but do express envelope proteins, such as the S protein, making them relevant for our binding studies (28). Pseudoviruses were produced using an HIV-based lentiviral system which has been previously successfully pseudotyped for the SARS-CoV-2 S protein by Crawford et al. (28).
We pseudotyped the particles for the S protein of several CoVs. Namely, MERS-CoV, HCoV-HKU1, SARS-CoV-1, and SARS-CoV-2.
Infection ability of the pseudoviruses was tested on HEK ACE2 TMPRSS2 and Rhabdomyosarcoma (RD) cell lines using the luciferase assay. Additionally, susceptibility of cell lines to infection was tested by measuring the interaction between CoV S protein domains and cell lines with Fluorescence activated Cell Sorting (FACS). For this HEK ACE2 TMPRSS2, RD, Madin-Darby canine kidney (MDCK), and lung adenocarcinoma (CALU3) cell lines were used.

Materials & methods

Plasmid and recombinant DNA production

Pseudovirus producing plasmids were made and transfected into JM109 cells to make a pseudoviral particle-producing DNA. A lentiviral HIV pHDM vector plasmid codon optimized for the SARS-CoV-2 S protein was used as a backbone (BEI Resources NR-52514) and was pseudotyped for just the S1 region of the S protein (Appendix 1A) and then for the entire S protein (Appendix 1B). Additionally, the vector plasmid contains a codon luciferase green – a reporter protein used in the luciferase assay for the testing of the infection of mammalian cells by the pseudoviruses.

The initial experiments were conducted using pseudoviruses pseudotyped for the S1 region of the S protein only. The S2 region of the S protein was a preexisting part of the backbone sequence and was derived from SARS-CoV-2. This strategy was replaced by pseudo typing for the entire S protein (S1+S2) in later experiments. The existing S1 region or the whole S protein encoding inserts present in the backbone plasmid were removed in order to replace them with the desired inserts of the CoVs of interest. This was done with enzymatic restriction with NheI (NEB R3131L) and PacI (NEB R0547L). An additional PacI restriction site was inserted into the backbone plasmid with PCR at the end of the S2 region of the SARS-CoV-2 S protein (Appendix 2) because it was not previously present and was needed to remove the entire S protein encoding insert.

The S1 region or the whole S protein encoding inserts of different CoVs were derived with NheI/PacI restriction from already available protein producing plasmids. Alternatively, the inserts can be obtained with PCR amplification and subsequent gel extraction, according to the manufacturers protocol, in the absence of necessary restriction sites in the plasmid (29). Thermocycler C1000 Touch by BioRad was used for PCR reactions together with primers from Appendix 3.

The S1 region or the whole S protein inserts and the backbone plasmids were combined by either ligation (T4 DNA ligase M0202L) or Gibson assembly, depending on whether the inserts were derived using restriction or PCR amplification, respectively. The plasmids were then transfected into JM109 bacteria (Promega L200B). The bacterial cultures were first grown on LB agar agar plates with the addition of ampicillin at 50 µg/mL. To obtain working solutions of purified pseudovirus producing DNA, mini or maxiprep were conducted. The colonies were grown in preparation for mini or maxiprep in liquid LB medium with the same concentration of ampicillin. Final checks of the DNA

were done with a NheI/PacI digestion control or DNA sequencing. The concentration of the DNA was measured on the Nanodrop at A280nm (Thermo Nanodrop 2000).

Pseudovirus and cell line production

Pseudovirus particles and the cell lines used to test the infection of the pseudovirus particles were created in the manner described by Crawford et al. (28)
An HIV-based lentiviral system was used and pseudotyped with the S1 region or the whole S protein of SARS-CoV-1, SARS-CoV-2, HCoV-HKU1, and MERS-CoV. The pseudovirus producing DNA mixture consists of the following: the previously described pseudoviral DNA that encodes for the S1 region or the whole S protein, plasmids encoding for the HIV proteins necessary for virion formation (Hgmp2, Tat1b, and Rev1b), and the plasmid encoding for the reporter protein for luciferase green which will enable the measurement of the infection rate with the luciferase assay. The pseudovirus producing HEK 293T cell line was continuously grown in Dulbecco's modified Eagle medium (DMEM; Gibco) supplemented with 10% fetal calf serum (FCS; Sigma) and 25 units/mL penicillin and 0.025 mg/mL streptomycin (pen/strep; Sigma). Prior to transfection the cell line was seeded in separate 150 mm cell culture dishes and grown to 50-70 % confluency. Expression vectors were incubated with PEI in a 1:8 ratio (µg DNA/µg PEI) and DMEM for 20 min before addition to the cells. 5-6 hours after transfection, the medium was replaced with SFM medium (Gibco 11686-029 + 3 g Primatone + 3.6 g carbonate + 2 g glucose + 0.4 g valproic acid + 10 mL Glutamax + 15 mL DMSO). Cells were incubated for 48 hours before supernatants were collected. Alongside the pseudovirus producing cells, the infection ability of the pseudoviruses was tested on HEK 293 ACE2 TMPRSS2, RD, MDCK and CALU3 cells which were grown in the same conditions as the HEK 293T cell line.
The supernatant from each cell culture dish was collected in 15 mL tubes and centrifuged at 500 x g for 5 minutes to remove detached cells. The supernatants were further purified by passing them through a 0.45 µm filter. The filtered supernatant was concentrated using a Vivaspin-20 10.000MWCO tube at 3180xg at 4°C for around 30 minutes in total or until 1 mL of supernatant remained. The pseudoviruses were suspended in the remaining 1 mL of medium and aliquoted in cryo safe screw cap tubes before storing at -80 °C.

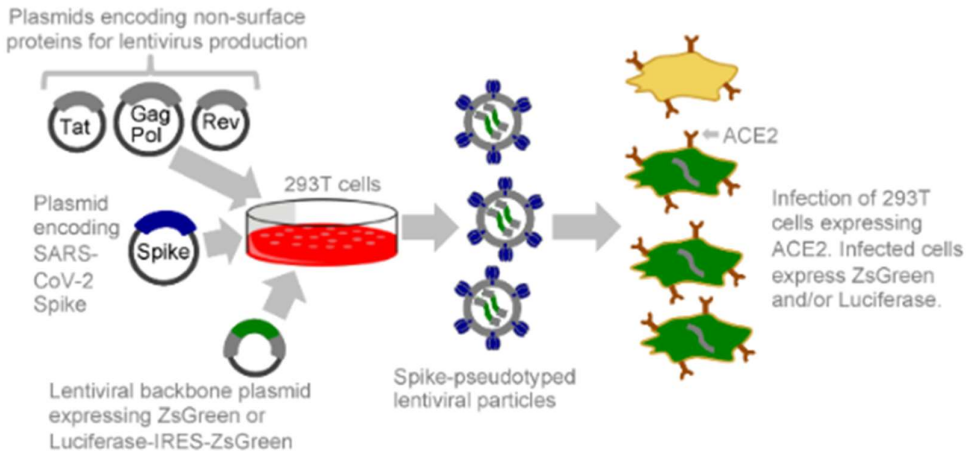


Figure 3: The principle behind pseudovirus production, formed from the spike protein DNA, reporter protein, and structural proteins Tat, Gag, and Rev. The pseudoviruses are used to infect 293T cells that express high levels of ACE2. (Image by Crawford at al. (28))

Testing the infection ability of pseudoviruses

Prior to infection testing, a 96-well plate was seeded with HEK ACE2 TMPRSS2 and RD cell lines. Firstly, a 96-well F-bottom white plate was coated with 100 µL poly-l-lysine (1/1000 in PBS) per well. The plate was sterilized under UV light for 20 minutes. The plate was then seeded with 1.25 × 10^ 4 293T-ACE2 or RD cells per well in DMEM medium and left to incubate at 37 °C and 5% CO_2. After overnight incubation, the number of cells was counted using a Countess II automated cell counter. The growth medium was gently removed without disturbing the cells and 100 µL of undiluted pseudovirus solution was added. Polybrene (TR-1003-G, Sigma Aldrich, St. Louis, MO, USA) was added to a final concentration of 5 µg/mL. The plate was set up in such a way that all the outer wells on the edges of the plate were filled with ~200 µL of PBS to reduce the evaporation of the growth medium from the wells containing the cells.
The cells were incubated for 48 hours. To test for infection, we used a Bright-glo luciferase assay system (Promega E2610). The assay was conducted according to the manufacturers protocol. We have added 10 µL of reagent per well after removing 100 µL of the growth medium from the wells (30). After addition of the reagent, the plate was incubated for two minutes at room temperature in the dark before being analyzed in the plate reader (POLARstar Omega). Three measurements were taken of the luciferase assay; once with the gain set to 3000, once with the gain set to 3600, and once with the gain adjusted to the well with the highest fluorescence. The graphs were made using the readout from the adjusted gain measurement and by normalizing for the value in the wells containing only the cells without any pseudovirus. The positive and negative controls were produced using the same protocol as the pseudoviruses of interest.

Testing the binding of spike protein domains to cell lines and protein expression

To test whether the prepared cell lines are susceptible to CoV S protein binding, a form of flow cytometry called FACS was used. To this end, S protein domains, namely NTD Fc proteins and RBD trimer proteins, were first produced with protein expression in HEK293T cells and then incubated with the cells of interest, as described by Tomris et al. (31).
HEK 293T cells were seeded in 150 mm culture dishes and incubated for 48 hours in DMEM medium at 37 °C and 5% CO_2. After incubation, the cells are transfected with 2 mL of the DNA/PEI solution (1:8) using 40 µg DNA per 150 mm dish. The cells were incubated with the transfection solution for 6 hours after which the medium was replaced with SFM medium. The cells were incubated for approximately five days before harvesting the protein from the supernatant.
The supernatant was collected and centrifuged for 10 minutes at 1200 rpm and then for 10 minutes at 3000 rpm in a swing-out rotor after transferring to a clean 50 mL tube. The supernatants were checked for protein expression by performing a Western blot. The protein is then purified from the supernatant by addition of 250 µL of Strep-tactin beads and incubation on a head-to-head roller at 4 °C overnight. The following morning, the 50 mL tubes were centrifuged for 10 minutes at 3000 rpm in a swing-out rotor. The supernatant was removed, and the beads washed with 800 µL of PBS in total while transferring them to a 1,5 mL Eppendorf tube. The tube was centrifuged for five minutes at 5000 rpm in a fixed-angle rotor and the supernatant removed, keeping the beads. Protein was eluted from the beads by adding 125 µL of the elution buffer pH 7.5 (0.1 M Tris buffer + 2.5 mM destioBiotin (Sigma D1411) + 150 mM NaCl + 1 mM EDTA), incubating for 10 minutes at room temperature, and centrifuging for five minutes at 5000 rpm in a fixed-angle rotor. The concentration of the protein in the supernatant was measured with Nanodrop A260nm. The elution was repeated until the concentration of the protein fell below 0.2 mg/mL.

FACS was conducted on four different cell lines: HEK 293T ACE2 TMPRSS2, RD, MDCK, and CALU-3. The cells were suspended after trypsinization in a suitable volume of cold FACS staining buffer (0,5 % BSA, 2 mM EDTA in PBS). 100 µL of the cell solution was used per 96-well plate well and each well contained 50 000 cells. The NTDs were precomplexed to pA-LS nanoparticles by combining equal amounts of NTD and pA-LS. Specifically, 50 µg/mL of each was used. The solution was incubated for

one hour on ice. After incubation, 20 µL of the protein solution was added to each well. In parallel, 20 µL of 50 µg/mL trimeric RBDs was added to the remaining wells. The cells were incubated with the S protein domains for one hour at 4°C. The cells were then washed with 200 µL of ice-cold FACS buffer for five minutes while centrifuging in a swing-out rotor at 300 x g. The cells were incubated with a primary antibody StrepMAB-Classic HRP (IBA 2-1509-001) at a final concentration of 10 µg/mL at 4°C for 30 minutes in the dark. The cells were then washed twice with ice-cold FACS buffer for 5 minutes at 300 x g. The incubation step was repeated with goat anti-human IgG secondary antibody with Alexa Fluor™ 488 fluorescent tag (Invitrogen #A-11013) at the same final Ab concentration. Because of the fluorescent tag, the binding represented in the form of fluorescence can be detected on the flow cytometer. The cells were again washed twice after which the cells can be stored in darkness at 4°C in 100 µL of fresh ice-cold FACS buffer. The results were acquired within 24 hours. The binding was measured with a CytoFLEX LX cytometer. The fluorescence was measured on the B525 channel. Each cell line- protein combination was measured in triplicate and averaged.

Results

MERS-CoV, HCoV-HKU1, and SARS-CoV-1 S protein pseudotyped pseudoviruses do not infect cells compared to negative control on the luciferase assay

The primary assay to test the pseudoviruses' ability to infect mammalian cells was the Luciferase assay. Since the pseudoviruses express the reporter protein Luciferase green, the infected cells should produce more fluorescence than the non-infected cells and referenced to a positive and negative control. A pseudovirus expressing vesicular stomatitis virus glycoprotein G (VSV-G) was used alongside SARS-CoV-2 pseudovirus as positive control. VSV-G pseudotyped particles infect mammalian cells independent of any cell surface receptors (Figure 4). A pseudovirus particle that expresses no envelope proteins was used as a negative control since it should have no infection mechanism due to the absence of the S protein. Interestingly, in all graphs but one (Figure 4D), the pseudoviruses of interest have lower infection signal (i.e. fluorescence) than the no envelope pseudovirus which seems counter-intuitive. The luciferase assay was conducted twice: once with particles pseudotyped for the S1 region of the S protein (Figure 4A and 4B) and once pseudotyped for the entire S protein (Figure 4C and 4D).

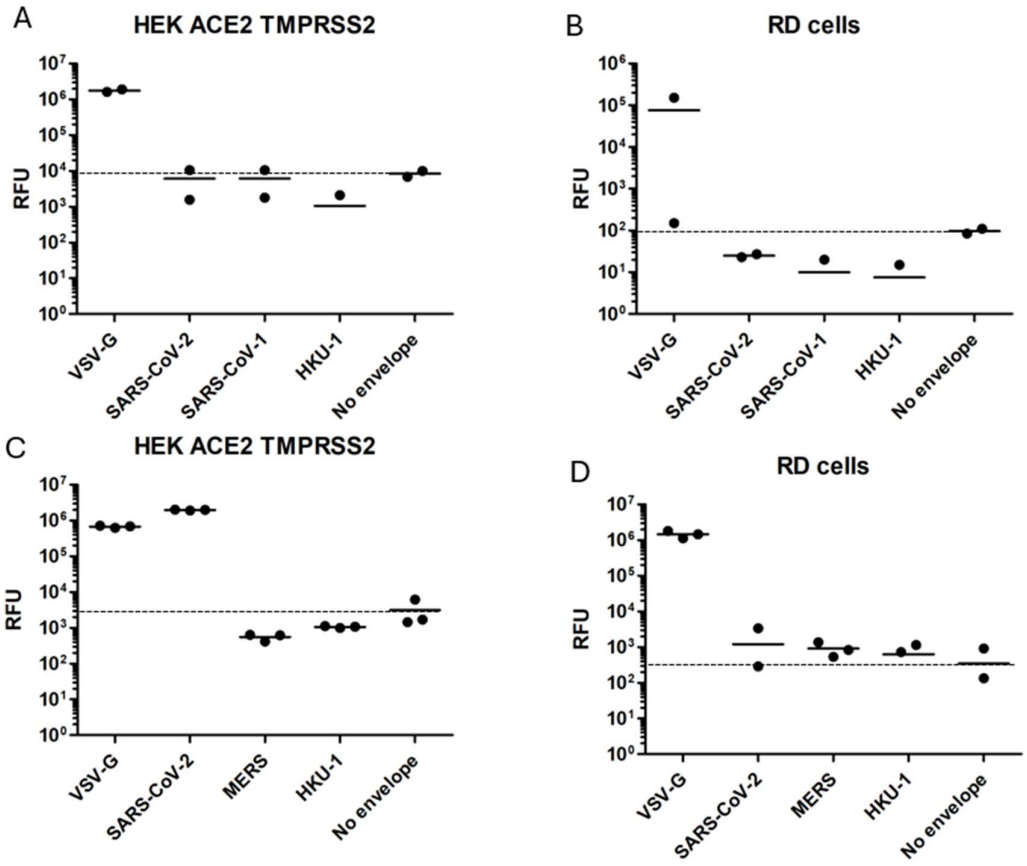


Figure 4: Luciferase assay results of the HEK ACE2 TMPRSS2 (A) and RD (B) cell lines after inoculation with the pseudovirus particles in the first round of experiments. The (C) and (D) results are from an identically conducted luciferase assay performed on pseudoviruses pseudo typed for the entire S protein.

The results of the luciferase assay report that there is no infection measured in either cell type with SARS-CoV-1 or HKU1 when compared to the no envelope pseudovirus (Figure 4A and 4B). Notably, the positive control SARS-CoV-2 also did not show any signs of infection (Figure 4A), which is unexpected, especially in HEK ACE2 TMPRSS2 cells given that cell line is known to be highly susceptible to SARS-CoV-2 infection. As can be seen in the Figure, the graphs of the first luciferase assay contain less data points. Firstly, these experiments were done in duplicate as opposed to the triplicates in the second luciferase assay. Secondly, some data points are outside graph limits due to large differences between duplicates. Large errors, together with ineffective positive controls, reduce the reliability of these results.

We conducted the luciferase assay for the second time (Figure 4C and 4D), this time with MERS instead of SARS-CoV-1. Another change was introduced during the second attempt of this assay which is the homogenization of the S protein. In other words, pseudoviruses used in the first luciferase assay were only pseudotyped for the S1 region of the S protein and the S2 region remained the same for all pseudoviruses as it was a part of the backbone, and it originated from a SARS-CoV-2. In the second round of experiments, both S1 and S2 regions of the S protein belonged to the same type of CoV. These results (Figure 4C and 4D) showed notable improvement compared to the initial experiment in Figure 4A and 4B. The variability and errors between the triplicate samples were reduced, improving the reliability of these results.

Moreover, SARS-CoV-2 successfully infected the HEK ACE2 TMPRSS2 cells, as seen in Figure 4A, serving as a solid positive control for the assay. However, despite the improved consistency in results,

there was still no evidence of infection in the cells by the pseudoviruses of interest when compared to the negative control.

Protein expression and testing of different cell lines for susceptibility to infection using FACS

At this point, we conducted FACS to investigate the lack of infection observed in the luciferase assays. FACS was performed to determine whether the NTD and RBD S protein domains from MERS, and SARS-CoV-2 can bind the target cells or rather their cell surface receptors. This would give us a clearer picture of the cells' susceptibility to infection.
binding of CoV-NTDs to sialylated glycans required multivalency to increase the binding signal. To do this, we precomplexed the purified NTDs with a nanoparticle known as pA-LS. It is a nanoparticle scaffold produced by combining domain B of protein A (pA) of *Staphylococcus aureus* with Lumazine Synthase (LS) (31). The trimeric RBDs display increased binding compared to the dimeric NTDs, explained by additional available receptor binding sites. Consequently, no nanoparticle precomplexing was necessary for the RBDs (31).

MERS-CoV and HCoV-HKU1 RBDs were the only S protein domains produced during this project since the other domains were already expressed, purified, and available from previous projects. However, all the domains were produced in the same manner.
Both MERS-CoV and HCoV-HKU1 RBDs were expressed well (Figure 5A) and were well concentrated after Strep-tactin bead purification, a Coomassie blue gel of which is visible in Figure 5B. The final concentrations of the purified proteins were 0,175 mg/mL and 0,4 mg/mL for MERS-CoV and HCoV-HKU1, respectively.

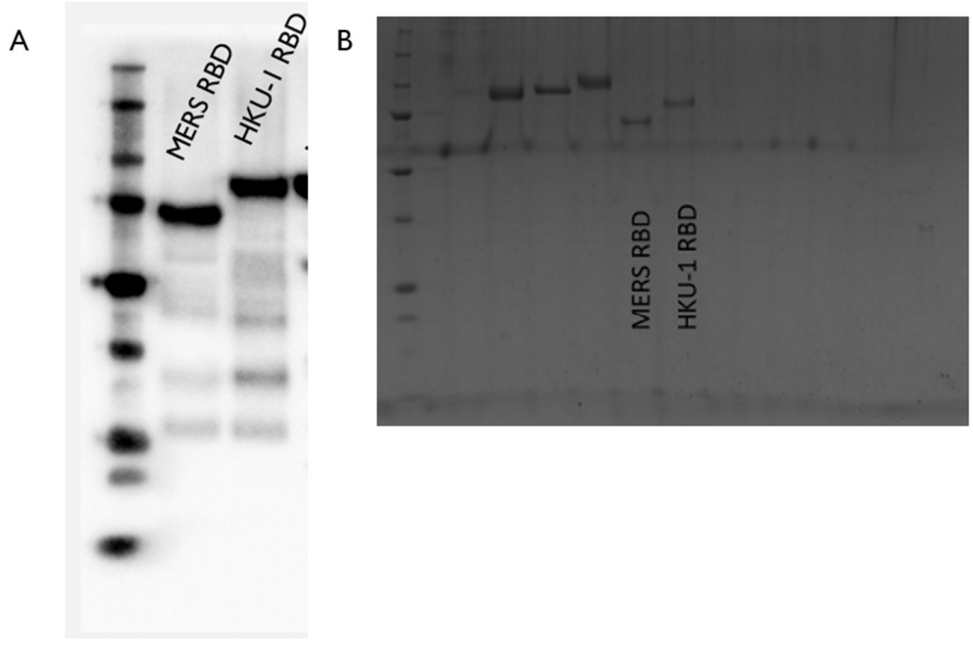


Figure 5: Western blot (A) and subsequent Coomassie blue (B) images of the HEK293T expressed RBD trimers of MERS-CoV and HCoV-HKU1

We incubated the S protein domains with four different cell lines based on their expected susceptibility to infection by the CoVs, namely, HEK ACE2 TMPRSS2, RD, MDCK, AND CALU3 cell lines (Table 2).

HEK 293T cells that overexpress ACE2 and TMPRSS2 were expected to be susceptible to MERS-CoV, SARS-CoV-2, and HCoV-HKU1 infections. RD cells do not express the necessary proteinaceous cell surface receptors so no binding was expected to be observed in these cells from the MERS-CoV and SARS-CoV-2 (32–34).
RD cell line is commonly used for detection of viruses and was expected to be susceptible to HCoV-HKU1 infection due to the cell surface sialylated glycans known to be expressed by this cell line (31).
Susceptibility to MERS-CoV infection was expected in MDCK cells as these cells express α2,3- and α2,6-linked sialic acid receptors (33).
CALU-3 cell line was anticipated to be susceptible to MERS-CoV, SARS-CoV-2, and HCoV-HKU1 infections due to their expression of both TMPRSS2 and DPP4 receptors, with DPP4 being one of the primary receptors used by MERS-CoV and TMPRSS2 known to be utilized by the other CoVs for infection (34–36).

Cell line	**MERS-CoV**	**HCoV-HKU1**	**SARS-CoV-2**
HEK ACE2 TMPRSS2	+ (32)	+ (15)	+ (35)
RD	-	+ (31)	- (37)
MDCK	+ (33)	- (36)	- (37)
CALU3	+ (34)	+ (36)	+ (35)

Table 2: Expected susceptibility of four different cell lines to the CoVs of interest that the pseudoviruses were pseudotyped for.

Absence of simultaneous RBD and NTD binding of mammalian cell lines in CoV S protein domains & susceptibility of cell lines to S protein binding

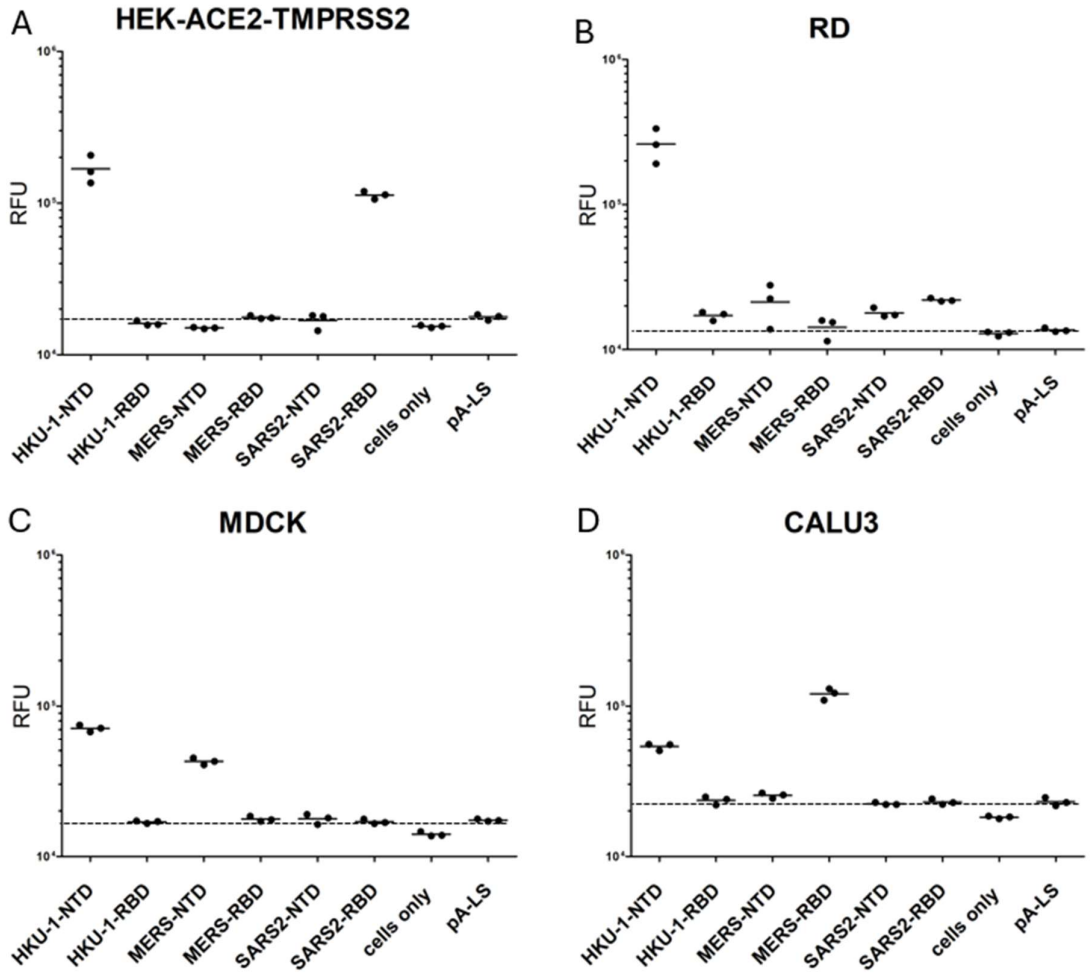


Figure 6: FACS results of the NTD or RBD binding of HEK ACE2 TMPRSS2 (A), RD (B), MDCK (C), AND CALU3 (D) cell lines

The RBDs and NTDs of HCoV-HKU1, MERS-CoV, and SARS-CoV-2 were tested for the ability for binding cell surface receptors.

This was done as a subsequent analysis following the absence of binding observed in the luciferase assay. This subsequent analysis was done to investigate whether the S protein domains can support the pseudovirus-receptor interaction which would potentially explain the absence of infection seen in the previous assay. The measurements were obtained with FACS which was enabled by the use of Alexa Fluor 488 goat anti-human IgG as a secondary antibody. Consequently, we were able to sort cells based on their expression of the fluorescent tag. Since the S protein domains contained a strep-tag, only those cells that were bound by the anti-strep primary mAb (and then by the fluorescently tagged secondary antibody), would be detected with FACS. This allowed us to separate the cells based on their engagement by the S protein domains.

As part of the FACS analysis, we included negative controls consisting of cells without the addition of spike protein domains and cells treated with only the nanoparticle- not precomplexed with any domains. The results show that only the HKU1 N-terminal domain (NTD) was able to bind to all the tested cell lines. However, the RBD of the same CoV did not show any binding in any cell line (Figure 6). This is a trend that was consistent across all the CoVs; even when one domain demonstrated binding, none of the pseudoviruses contained both domains capable of binding to the cells simultaneously.

SARS-CoV-2 RBD showed significant binding in HEK ACE2 TMPRSS2 cells (Figure 6A), which was expected due to the presence of ACE2 and TMPRSS2, the primary receptors utilized by SARS-CoV-2. Consequently, we can also observe the SARS-CoV-2 RBD as a positive control in this cell line. We expected to see some MERS-CoV RBD binding in this cell line as well, considering the findings described in literature (Table 2), however, this was not the case.

The significant loss of binding of SARS-CoV-2 RBD in RD cells we observed (Figure 6B) was anticipated, given that these cells do not express the necessary proteinaceous receptors to support this interaction. Notably, there is also no significant binding by the HCoV-HKU1 NTD in the same cell line, which is unexpected considering the previously described preference of this domain from this CoV for RD cells in literature. In MDCK cells (Figure 6C) we expected and saw the MERS-CoV NTD binding due to the presence of sialylated glycans. The binding of the HKU1 NTD was observed in MDCK cells, as was in all cell lines, although literature found this CoV to be preferential for a differently linked sialylated glycan (15,36).

Finally, for CALU-3 cells, we anticipated that domains from all three pseudoviruses would bind due to the expression of TMPRSS2 and DPP4 receptors, both of which are involved in the infection mechanism of all three CoVs (15,16,34–36). However, only the MERS-CoV RBD and HCoV-HKU1 NTD were able to interact with these cells, while the other domains displayed no binding activity compared to the negative controls.

The binding analysis showed clear interactions between some spike protein domains and different cell lines. While some domains were able to bind, the absence of dual binding for both the NTDs and RBDs indicates a possible limitation in their ability to infect cells.

Discussion

This report provides an overview of pathogenic human CoVs, focusing on the infection mechanism of MERS-CoV, HCoV-HKU1, SARS-CoV-1, and SARS-CoV-2. The burden that currently circulating CoVs as well as the threat of newly emerging, potentially pandemic future CoVs present to society clearly highlight the need for research into therapeutic targets.
An important aspect of the report is the differentiation between proteinaceous and sialylated glycan receptors. The previously and recently established ability of different CoVs to engage both types of cell surface receptors underscores the complexity of CoV pathogenesis (7,31). Particular interest and the ultimate goal of this project is the examination of the S protein- sialic acid interactions and their implication for the CoV viral entry. Elucidating this relationship can aid in understanding multiple approaches for viral attachment disruption and provide novel pharmaceutical targets.

The innovative use of pseudoviruses in this research allows for a safer investigation of highly pathogenic CoVs. As described, the pseudoviral particles were produced according to an established protocol by Crawford et al. (28). In their report, Crawford describes a protocol for the design, production, and purification of a SARS-CoV-2 S protein pseudotyped non-replicative particle. We have attempted to imitate said protocol with the addition of pseudo typing for the S protein of HCoV-HKU1, MERS-CoV, and SARS-CoV-1. Initially, there was some difficulty with producing particles pseudotyped for the S1 region of the S protein. The luciferase assay, which we used to test the pseudovirus' infection rate of the mammalian cells, showed no infection by any S1 protein pseudotyped viruses, including the SARS-CoV-2 which was meant to act as a positive control. Additionally, it is important to note the significant variability observed between duplicate samples in this experiment, most notably in the RD cell line. These high differences may have contributed to the discrepancies in the infection results, warranting another attempt in this experiment with freshly prepared pseudoviruses, differently pseudo typed, and more careful pipetting.
Indeed, a reduction in variability between triplicates was seen and the function of SARS-CoV-2 pseudovirus as a positive control was obtained by more careful pipetting, preparation of fresh pseudoviruses, and pseudo typing for the entire S protein instead of the S1 region only. This way of pseudo typing was decided on following a theory that the S1 regions and the S2 regions of the S protein were incompatible since they belonged to different viruses. While the SARS-CoV-2 did successfully act as a positive control in HEK ACE2 TMPRSS2 cells in the second round of experiments, none of the other pseudoviruses were able to infect cells,

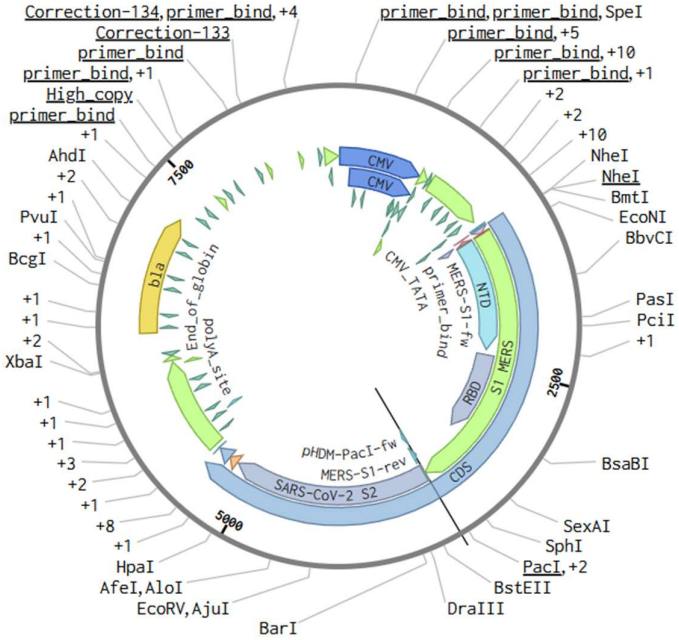
This approach has resulted in valuable insights, although the unexpected results from the luciferase assays raise questions about receptor expression, the reliability of certain cell lines, and the design of the pseudovirus particles, among other limitations of this project. Following the undesirable results of the luciferase assay, FACS analysis was conducted to investigate the ability of CoV NTDs and RBDs to bind mammalian cell lines. Potentially, the lack of RBD and NTD binding of cell surface receptors might explain the absence of infection in the luciferase assay.
While some S protein domains successfully bound to specific cell surface receptors, there was a clear absence of simultaneous binding of both the NTD and the RBD by any CoV which might offer an insight on why there was no infection observed by these pseudoviruses.
SARS-CoV-2 pseudovirus was so far the only pseudovirus we successfully pseudotyped for and observed the infection in the mammalian cells. Additionally, the SARS-CoV-2 RBD was able to bind HEK ACE2 TMPRSS2 cells, while the NTD of the same CoV did not bind the same cell line. Clearly, simultaneous binding by both S protein domains is not an absolute requirement for attachment and entry, at least not in all CoVs.
On the other hand, HCoV-HKU1 NTD was able to bind all cell lines, while no infection was detected by this pseudovirus. One possible explanation for this is the previously described receptor duality

where the inability of the HCoV-HKU1 RBD to participate in binding as well was terminal for the infectivity of the pseudovirus.
Additionally, HCoV-HKU1 NTD binding to MDCK cell line was detected despite the preference of this CoV for a different sialylated glycan as an attachment receptor. Namely, the MDCK cell line expresses α2,3- and α2,6-linked sialylated glycans while the HCoV-HKU1 is preferential for α2,8-linked sialic acid receptors (15,33).
Based on literature finings, we have expected to detect MERS-CoV RBD binding in HEK ACE2 TMPRSS2 cells due to its previously demonstrated affinity for the TMPRSS2 receptor (32). Contrary to these findings, we have not detected any interaction between this cell line and either of the MERS-CoV S protein domains. However, Shirato et al. (32), that demonstrated this affinity, did not use the HEK 293T cell line but rather a different one that just overexpressed the TMPRSS2 receptor. These findings potentially point to still unknown infection mechanisms of these CoVs.
The most likely explanation for the absence of pseudovirus infection and interaction with the cell surface receptors again regards the S protein domains' incompatibility. Even if the engagement of receptors by both S protein domains simultaneously is not imperative, proper functioning and compatibility of both domains is a prerequisite for the proper functioning of the S protein as a whole. While the S1 and S2 domains of the S protein were homogenized per pseudovirus, the TMD remained identical in all pseudoviruses, and it belongs to SARS-CoV-2. The importance of the TMD in proper orientation and therefore functioning of the other S protein domains has been previously demonstrated (6). Consequently, improper compatibility of the foreign TMD with the rest of the S protein domains could have had functional consequences for the S protein.

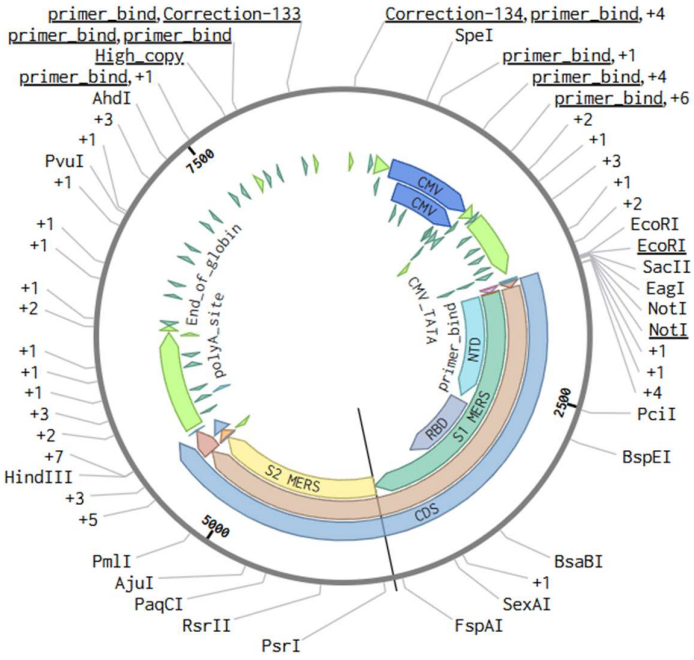
Looking ahead, the priority is to design and produce pseudoviruses that successfully infect mammalian cells. As already explained, this could very well be as simple as switching out the current TMD and homogenizing it with the rest of the S protein domains. However, should this approach fail, the design and method of production of the pseudoviruses might need to be reimagined, together with the cell lines used to produce and test the pseudoviruses.
The ability of the pseudoviruses to successfully infect mammalian cells is a prerequisite to any future steps and directions of this project. By satisfying this prerequisite, the pseudoviruses would enable us to test the binding of different cell lines, that either express or have the expression of various proteinaceous and/or glycan receptors knocked-out. The modifications of cell surface receptor expressions will result in increase, reduction or loss in pseudovirus binding. This would in turn provide us with a better understanding of CoV attachment and infection mechanism and enhance our strategies for combating viral infections.

Appendices

Appendix 1 –CoV pseudovirus pseudo typing on the example of MERS-CoV plasmid map

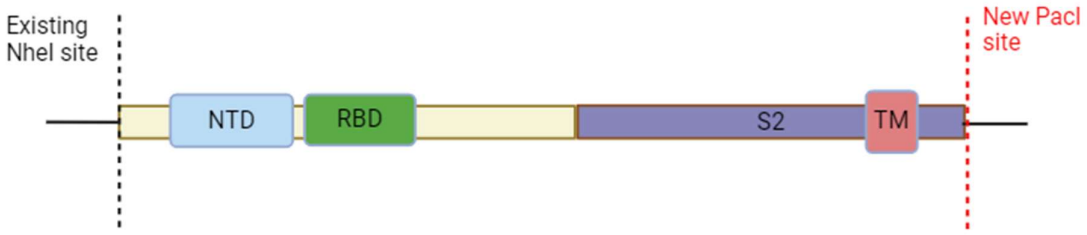


Appendix 1A: Benchling.com plasmid map of the S1 region of the S protein pseudo typed pseudovirus. The solid line represents a border between MERS-CoV S1 and SARS-CoV-2 S2.



Appendix 1B: Benchling.com plasmid map of the whole S protein pseudo typed pseudovirus. The solid line represents a border between MERS-CoV S1 and MERS-CoV S2.

Appendix 2 – Restriction site insertion



Name and number of primer	Primer sequence
52514-FL-PacI-fw #984	CAATATATCAAGTGGCCTTTAATTAAGTGGTATATTTGGC
52514-FL-rev #985	AGGCCACTTGATATATTGCTCGTAC

Appendix 2: position of insertion of the PacI site into the lentiviral backbone plasmid together with the primer sequences used.

Appendix 3 – PCR primers

Name and number of primer	Primer sequence
SARS-NTD-f #221 (for SARS-CoV-1 S1)	GGATGCTGGTCGCTTCCGTGCTAGCAGACCTGGATAGGTGCACC
SARS-S1_R #309 (for SARS-CoV-1 S1)	ATCTGCTTCATGGCGCTTAATTAACAGCAGGGACACGGTG
HCOV-hku-ntd-f #447 (for HcoV-HKU1 S1)	GGATGCTGGTCGCTTCCGTGCTAGCAgtgatcggcgacttcaactg
HCOV-hku-rbd-r #450 (for HcoV-HKU1 S1)	ATCTGCTTCATGCGCTTAATTAAatagttcacgcacacgcc

Appendix 3: PCR primers used for S1 amplification prior to gel extraction. The same primers are used for colony PCR.

References

1. Chen B, Er-Kang T, He B, Lejin T, Ruiying H, Shuangwen W, et al. Overview of lethal human coronaviruses | Signal Transduction and Targeted Therapy [Internet]. [cited 2024 Jan 22]. Available from: https://www.nature.com/articles/s41392-020-0190-2

2. Shi Y, Wang G, Cai X peng, Deng J wen, Zheng L, Zhu H hong, et al. An overview of COVID-19. J Zhejiang Univ Sci B. 2020 May;21(5):343–60.

3. Van Der Hoek L. Human Coronaviruses: What Do They Cause? Antiviral Therapy. 2007 Jan 1;12(4_part_2):651–8.

4. M Poutanen S. Human Coronaviruses - PMC [Internet]. [cited 2024 Feb 2]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152474/

5. Chakraborty C, Bhattacharya M, Sharma AR. Present variants of concern and variants of interest of severe acute respiratory syndrome coronavirus 2: Their significant mutations in S-glycoprotein, infectivity, re-infectivity, immune escape and vaccines activity. Reviews in Medical Virology. 2022;32(2):e2270.

6. Aliper ET, Efremov RG. Inconspicuous Yet Indispensable: The Coronavirus Spike Transmembrane Domain. International Journal of Molecular Sciences. 2023 Nov 16;24(22):16421.

7. Tomris I, Unione L, Nguyen L, Zaree P, Bouwman KM, Liu L, et al. SARS-CoV-2 Spike N-Terminal Domain Engages 9-O-Acetylated α2–8-Linked Sialic Acids. ACS Chem Biol. 2023 Apr 27;18(5):1180–91.

8. Baker AN, Richards SJ, Pandey S, Guy CS, Ahmad A, Hasan M, et al. Glycan-Based Flow-Through Device for the Detection of SARS-COV-2. ACS Sens. 2021 Oct 22;6(10):3696–705.

9. Acosta-Gutiérrez S, Buckley J, Battaglia G. The Role of Host Cell Glycans on Virus Infectivity: The SARS-CoV-2 Case. Advanced Science. 2022 Nov 23;10(1):2201853.

10. Thompson AJ, de Vries RP, Paulson JC. Virus recognition of glycan receptors. Current Opinion in Virology. 2019 Feb 1;34:117–29.

11. Perrotta F, Matera MG, Cazzola M, Bianco A. Severe respiratory SARS-CoV2 infection: Does ACE2 receptor matter? Respiratory Medicine. 2020 Jul 1;168:105996.

12. Milewska A, Zarebski M, Nowak P, Stozek K, Potempa J, Pyrc K. Human Coronavirus NL63 Utilizes Heparan Sulfate Proteoglycans for Attachment to Target Cells. Journal of Virology. 2014 Nov;88(22):13221.

13. Yeager CL, Ashmun RA, Williams RK, Cardellichio CB, Shapiro LH, Look AT, et al. Human aminopeptidase N is a receptor for human coronavirus 229E. Nature. 1992;357(6377):420.

14. Krempl C, Schultze B, Herrler G. Analysis of cellular receptors for human coronavirus OC43. Adv Exp Med Biol. 1995;380:371–4.

15. McCallum M, Park YJ, Stewart C, Sprouse KR, Addetia A, Brown J, et al. Human coronavirus HKU1 recognition of the TMPRSS2 host receptor. Cell. 2024 Aug 8;187(16):4231-4245.e13.

16. Raj VS, Mou H, Smits SL, Dekkers DHW, Muller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC - PubMed [Internet]. [cited 2024 Oct 17]. Available from: https://pubmed.ncbi.nlm.nih.gov/23486063/

17. Park YJ, Walls AC, Wang Z, Sauer MM, Li W, Tortorici MA, et al. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. Nat Struct Mol Biol. 2019 Dec;26(12):1151–7.

18. Nassar A, Ibrahim IM, Amin FG, Magdy M, Elgharib AM, Azzam EB, et al. A Review of Human Coronaviruses' Receptors: The Host-Cell Targets for the Crown Bearing Viruses. Molecules. 2021 Oct 26;26(21):6455.

19. Li Z, Lang Y, Liu L, Bunyatov MI, Sarmiento AI, de Groot RJ, et al. Synthetic O-acetylated sialosides facilitate functional receptor identification for human respiratory viruses. Nat Chem. 2021 May;13(5):496–503.

20. Stencel-Baerenwald JE, Reiss K, Reiter DM, Stehle T, Dermody TS. The sweet spot: defining virus–sialic acid interactions. Nat Rev Microbiol. 2014 Nov;12(11):739–49.

21. Li Y, Liu D, Wang Y, Su W, Liu G, Dong W. The Importance of Glycans of Viral and Host Proteins in Enveloped Virus Infection. Frontiers in Immunology. 2021 Apr 29;12:638573.

22. Li W, Hulswit RJG, Widjaja I, Raj VS, McBride R, Peng W, et al. Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. Proceedings of the National Academy of Sciences. 2017 Oct 3;114(40):E8508–17.

23. Kumlin U, Olofsson S, Dimock K, Arnberg N. Sialic acid tissue distribution and influenza virus tropism. Influenza and Other Respiratory Viruses. 2008;2(5):147–54.

24. Unione L, Moure MJ, Lenza MP, Oyenarte I, Ereño-Orbea J, Ardá A, et al. The SARS-CoV-2 Spike Glycoprotein Directly Binds Exogeneous Sialic Acids: A NMR View. Angew Chem Int Ed Engl. 2022 Apr 25;61(18):e202201432.

25. Saunders N, Fernandez I, Planchais C, Michel V, Rajah MM, Baquero Salazar E, et al. TMPRSS2 is a functional receptor for human coronavirus HKU1. Nature. 2023 Dec;624(7990):207–14.

26. Fraser BJ, Beldar S, Seitova A, Hutchinson A, Mannar D, Li Y, et al. Structure and activity of human TMPRSS2 protease implicated in SARS-CoV-2 activation. Nat Chem Biol. 2022 Sep;18(9):963–71.

27. Baxter VK, Montgomery SA. Preclinical coronavirus studies and pathology: Challenges of the high-containment laboratory. Veterinary Pathology. 2022 Apr 11;59(4):556.

28. Crawford KHD, Eguia R, Dingens AS, Loes AN, Malone KD, Wolf CR, et al. Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. Viruses. 2020 May 6;12(5):513.

29. GeneJET PCR Purification Kit [Internet]. [cited 2024 Oct 19]. Available from: https://www.thermofisher.com/order/catalog/product/K0701

30. Bright-Glo™ Luciferase Assay System Protocol [Internet]. [cited 2024 Oct 20]. Available from: https://nld.promega.com/resources/protocols/technical-manuals/0/bright-glo-luciferase-assay-system-protocol/

31. Tomris I, Kimpel ALM, Liang R, van der Woude R, Boons GJPH, Li Z, et al. The HCoV-HKU1 N-Terminal Domain Binds a Wide Range of 9-O-Acetylated Sialic Acids Presented on Different Glycan Cores. ACS Infect Dis [Internet]. 2024 Oct 12 [cited 2024 Oct 20]; Available from: https://doi.org/10.1021/acsinfecdis.4c00488

32. Shirato K, Kawase M, Shutoku M. Middle East Respiratory Syndrome Coronavirus Infection Mediated by the Transmembrane Serine Protease TMPRSS2 | Journal of Virology [Internet]. [cited 2024 Oct 20]. Available from: https://journals-asm-org.utrechtuniversity.idm.oclc.org/doi/full/10.1128/jvi.01890-13

33. Byrd-Leotis L, Lasanajak Y, Bowen T, Baker K, Song X, Suthar MS, et al. SARS-CoV-2 and other coronaviruses bind to phosphorylated glycans from the human lung. Virology. 2021 Oct 1;562:142–8.

34. Tao X, Mei F, Agrawal A, Peters CJ, Ksiazek TG, Cheng X, et al. Blocking of Exchange Proteins Directly Activated by cAMP Leads to Reduced Replication of Middle East Respiratory Syndrome Coronavirus. Journal of Virology. 2014 Apr;88(7):3902–10.

35. Pires De Souza GA, Le Bideau M, Boschi C, Wurtz N, Colson P, Aherfi S, et al. Choosing a cellular model to study SARS-CoV-2. Front Cell Infect Microbiol [Internet]. 2022 Oct 21 [cited 2024 Oct 20];12. Available from: https://www.frontiersin.org/journals/cellular-and-infection-microbiology/articles/10.3389/fcimb.2022.1003608/full

36. Chan JFW, Chan KH, Choi GKY, To KKW, Tse H, Cai JP, et al. Differential Cell Line Susceptibility to the Emerging Novel Human Betacoronavirus 2c EMC/2012: Implications for Disease Pathogenesis and Clinical Manifestation. The Journal of Infectious Diseases. 2013 Jun 1;207(11):1743–52.

37. Wang L, Fan X, Bonenfant G, Cui D, Hossain J, Jiang N, et al. Susceptibility to SARS-CoV-2 of Cell Lines and Substrates Commonly Used to Diagnose and Isolate Influenza and Other Viruses. Emerging Infectious Diseases. 2021 May;27(5):1380.

Use of GenAI statement

In this research project, I employed GenAI, specifically ChatGPT 4.0 mini, to simplify and save time in literature review. I asked ChatGPT to provide summaries of studies and reviews on glycoscience and virology. ChatGPT was also used in my writing phase, where it provided suggestions for improving the structure and cohesion of my text. I critically reviewed all AI-generated content, and I guarantee that the text is of high academic integrity. Using ChatGPT improved my efficiency and academic writing skills, without preventing me from writing my original report.