

# GLYCOSYLATION OF THE RECEPTOR BINDING DOMAIN OF COVID-19 VIRUS SPIKE PROTEIN

## INTRODUCTION

The coronavirus 2019 (COVID-19) pandemic has quickly become a global socio-economic crisis. It is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). From the first case of the disease in Wuhan, China in early January to the date of June 29th, 2020, the virus has infected over 10 million people and caused over 500,000 deaths worldwide.

The virus utilizes its membrane spike (S) protein to bind to angiotensin-converting enzyme 2 (ACE-2) to initiate infection. The S protein is composed of two functional subunits, S1 and S2. S1 contains a receptor binding domain (RBD) and is responsible for the initial attachment of the virus to the surface of host cells, and S2 is responsible for membrane fusion that triggers entry of the virus into the host cells. Each S protein is extensively glycosylated with host-derived glycans at 22 N-linked glycosylation sites and at least two O-glycan sites (Shajahan, A., Supekar, N.T., *et al.* 2020, Watanabe, Y., Allen, J.D., *et al.* 2020). The virus acquires glycosylation by hijacking the glycosylation machinery in the host cell reticulum-Golgi intermediate compartments when offspring virions are formed. Therefore, the S protein has its unique glycosylation

sites but also carries the signature glycosylation patterns of the host cells. Glycosylation on the S protein helps to increase stability and solubility of the protein. More importantly, glycosylation serves to camouflage the immunogenic epitopes on the protein, enhancing the virus' ability to evade the host immune response.

The S protein is also a principal target for vaccine development and a critical component of serological assays. To help further research on SARS-CoV-2 therapeutics and diagnostics, we have expressed the RBD of the S protein in several common host cell lines, including HEK293 cells, CHO cells, and Sf21 cells. Since these host cells are known to have distinctive glycosylation capabilities, RBD proteins expressed from these cells are expected to exhibit different glycosylation features. In this report, we analyze the glycosylation using direct fluorescent glycan labeling (DFGL) tools we developed earlier (Wu, Z.L., Person, A.D., *et al.* 2019). The results may provide quick guidance for selecting the most relevant recombinant RBD protein for different purposes.

## MATERIALS AND METHODS

### RECOMBINANT PROTEINS AND FLUORESCENT SUBSTRATES

CMP-Cy5-Sialic acid, recombinant human ST3Gal1, ST6Gal1, and *C. perfringens* Neuraminidase were from R&D Systems, Bio-Techne. All SARS-CoV-2 spike receptor binding domain proteins were expressed and purified using Bio-Techne's proprietary methods.

### DFGL LABELING

Labeling reactions were performed using 2 µg of the target protein mixed with 0.2 nmol CMP-Cy5-Sialic Acid and 0.5 µg of a sialyltransferase in a 30 µL buffer of 25 mM Tris, 10 mM MnCl<sub>2</sub>, pH 7.5. Reactions were incubated at 37° C for 60 minutes. To assess the sialylation level of the proteins, samples were pretreated with 0.1 µg of recombinant *C. perfringens* Neuraminidase. The reaction was then separated by SDS-PAGE and imaged by a traditional protein imaging station via trichloroethanol (TCE) staining and a fluorescent imager (FluorChem R, ProteinSimple, Bio-Techne).

# RESULTS

To investigate the various glycan structures on the Spike RBD proteins, we first probed O-glycans with or without pretreatment of neuraminidase or endoglycosidase EF using the O-glycan-specific sialyltransferase ST3Gal1 (Figure 1). Addition of the fluorescent Cy5-conjugated Sialic acid by ST3Gal1 allows for simple detection of O-Glycans using a fluorescent imager. *C. perfringens* Neuraminidase removes existing terminal sialic acid residues from glycoproteins and opens sites on O-glycans for the addition of Cy5-conjugated Sialic acid by ST3Gal1 (Figure 1C). Endoglycosidase EF is an O-glycosidase that efficiently removes simple core-1 type O glycans but leaves more complex core-2 and -3 type O-glycans unaffected (1C). First, using ST3Gal1, we labeled non-sialylated O-glycans and detected fluorescence on all four RBD proteins (Figure 1, - lanes). Pretreatment with neuraminidase significantly increased the labeling on HEK293 de-

rived Spike RBD protein while little to no change was observed on the CHO- and insect-derived Spike RBD proteins (Figure 1, Neu lanes). Densitometry analysis on the HEK293 derived Spike RBD revealed a 71% increase in fluorescence suggesting that 71% of the O-glycans on the HEK293 cell derived protein are sialylated. Next, neuraminidase and endoglycosidase EF pretreatment revealed differences in O-glycan patterns of the four Spike RBD proteins. A major difference is that the O-glycans on HEK293 Spike RBD are resistant to O-glycosidase treatment indicating that the O-glycans on HEK293 expressed protein are elongated (Figure 1, HEK293, EF lane). The labeling on CHO, Sf21 and Tn5 expressed proteins was completely removed by EF (indicated with yellow arrows), suggesting that RBD proteins expressed in these cells only contain simple core-1 type O-glycans.

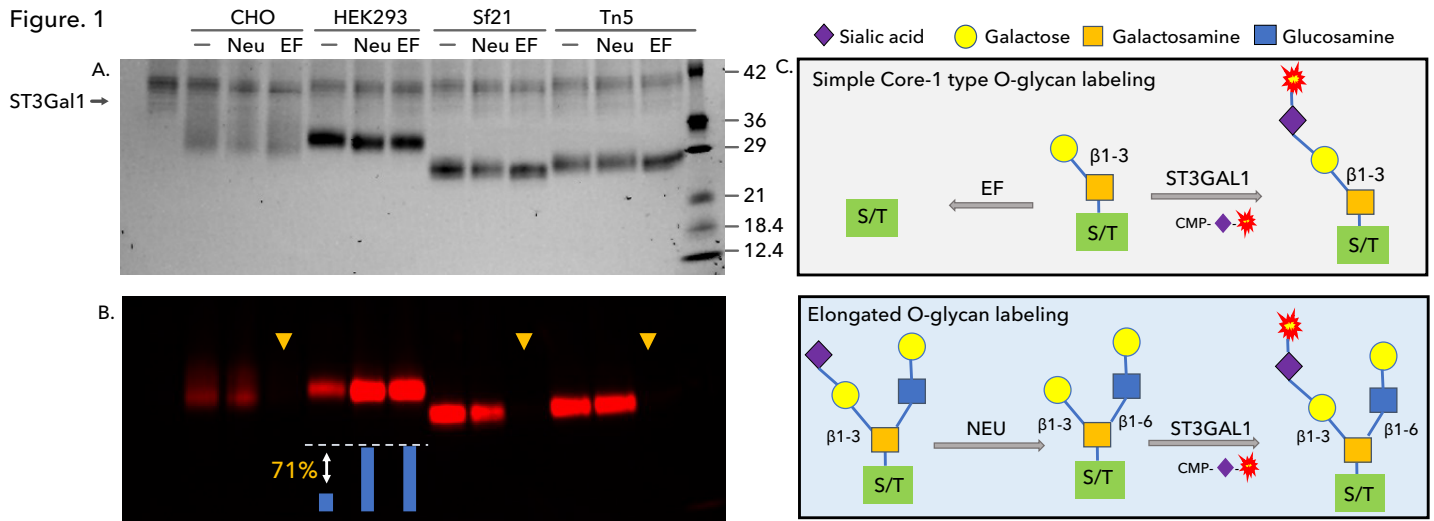


Figure 1. SARS-CoV-2 Spike RBD expressed in CHO, HEK293, Sf21 and Tn5 cells were probed for O-glycans with CMP-Cy5-Sialic acid and ST3Gal1. SARS-CoV-2 Spike RBD proteins were directly labeled or labeled after pretreatment with *C. perfringens* neuraminidase alone (Neu) or in combination with endoglycosidase EF (EF), an O-glycosidase that is more specific for Core-1 O-glycans. The samples were separated by SDS-PAGE and imaged by TCE staining (A) and fluorescent imaging (B). Models depicting enzymatic reactions and fluorescent labeling are shown (C). The labeling on CHO, Sf21 and Tn5 cell expressed proteins was completely removed by EF (yellow arrows), suggesting that RBD proteins expressed in these cells only contain simple Core-1 related O-glycans. Labeling on HEK293 cell expressed protein was resistant to EF, suggesting that the protein contains other types of O-glycans. Neuraminidase treatment resulted in a 71% increase in labeling of the protein expressed in HEK293 cells, suggesting that 71% of the protein was sialylated.

Next, we probed complex N-glycans on these proteins with CMP-Cy5-Sialic acid using the N-glycan-specific sialyltransferase ST6Gal1 and high-mannose N-glycans with GDP-AF555-fucose using FUT8 simultaneously (Figure 2). MGAT1 was also added into the labeling reactions to convert the high-mannose N-glycan Man5 to the substrate of FUT8. The results indicate that mammalian cell expressed proteins contain mainly complex N-glycans and

insect cell expressed proteins only contain high-mannose gly- cans. Interestingly, only the HEK293 expressed Spike RBD protein was found to be significantly sialylated on N-glycans, consistent with the O-glycan observations. The sialylation level of the N-gly- cans on the HEK293 expressed is 33.5% calculated by densitome- try analysis.

Figure. 2

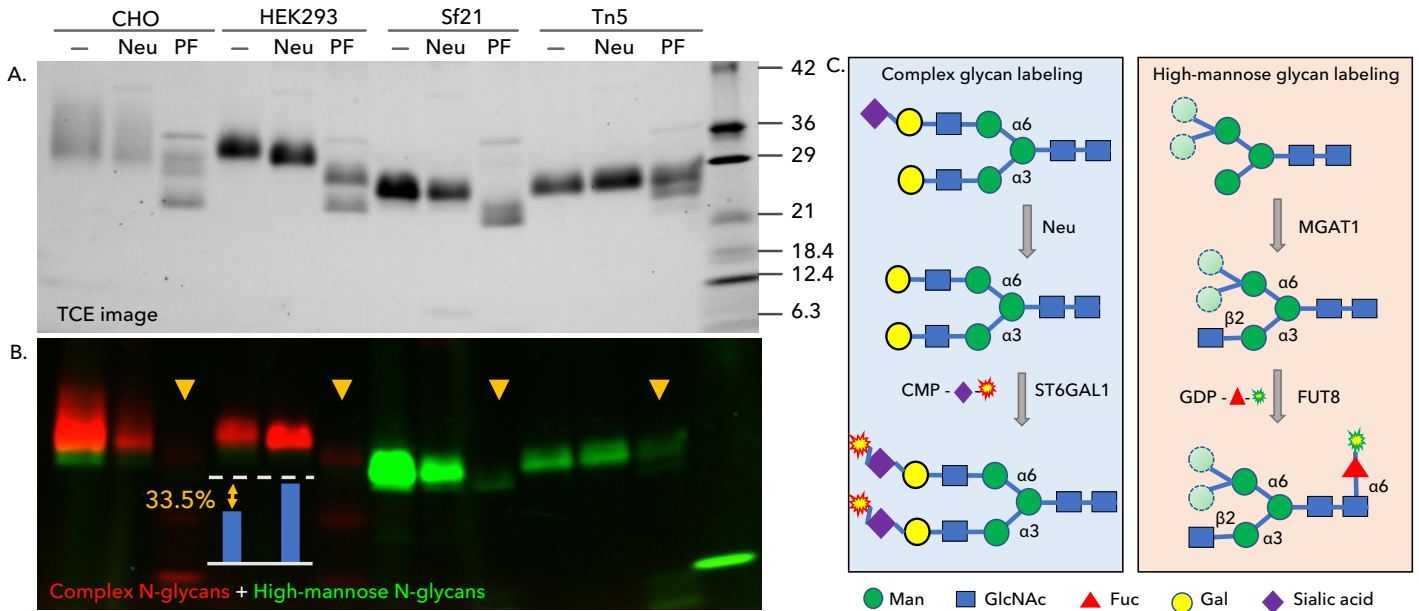


Figure 2. SARS-CoV-2 S1 RBD expressed in CHO, HEK293, Sf21 and Tn5 cells were probed for complex and high-mannose N-glycans with CMP-Cy5-Sialic acid/ST6Gal1 (red) and GDP-AF555-Fucose/FUT8 (green), respectively. The various Spike RBD proteins were labeled directly before or after pretreatment with *C. perfringens* neuraminidase (Neu) or PNGase F (PF). The samples were separated on SDS-PAGE and imaged by TCE staining (A) and fluorescent imaging (B). Models depicting enzymatic reactions and fluorescent labeling are shown (C). CHO and HEK293 expressed proteins were largely labeled by Cy5-Sialic acid, suggesting that they primarily contain complex N-glycans. Sf21 and Tn5 cell expressed proteins were only labeled by AF555-fucose, suggesting that they only contain high mannose N-glycans. In all cases, labeling was sensitive to PNGase F treatment (indicated by arrow heads), confirming that labeling was N-glycan specific. HEK293 cell expressed protein exhibited 33.5% sialylation suggested by densitometry analysis. Densitometry analysis was not performed on other cases as the amounts of the protein loaded varied due to protein precipitation caused by the pretreatments.

Figure. 3

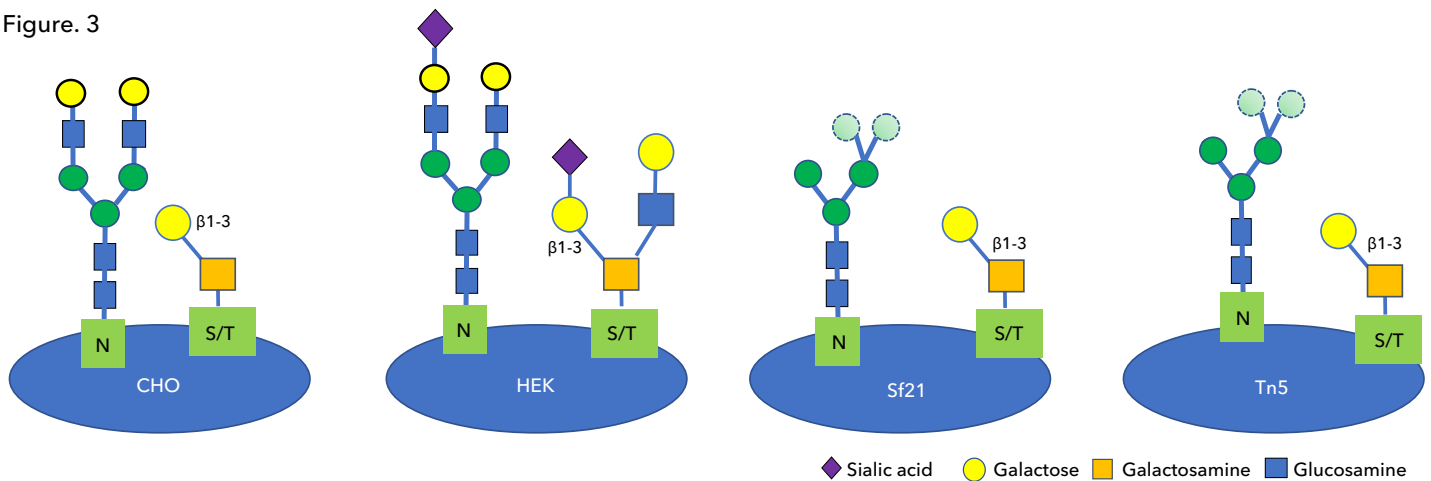


Figure 3: Summary on the glycans found on RBD proteins expressed in four major different host cells. Only representative glycans are depicted here for simplicity and clarity. Note that the identity of the O-glycans on HEK293 cell expressed protein, depicted as Core-2, is more speculative. The CHO expressed Spike RBD protein contains complex N-glycans and mainly Core-1 O-glycans. HEK293 expressed Spike RBD is depicted with complex N-glycans and Core-2 O-glycans both containing terminal sialic acid. Both Sf21 and Tn5 expressed Spike RBD proteins contain high mannose N-glycans and simple Core-1 O-glycans.

# SUMMARY

By DFGL probing we have shown major differences on the glycosylation of the SARS-CoV-2 Spike RBD protein expressed in four different key cell lines. Clear differences were found regarding the types of N- and O-glycans expressed on these RBD proteins (Figure 3). All proteins were found contain O-glycans, but only HEK293 cell expressed protein contained O-glycans other than Core-1 type. Only the HEK293 expressed protein was found to

have a significant level of sialylation. Complex N-glycans were only found on CHO and HEK293 expressed proteins. A significant level of high-mannose N-glycans were found on Sf21 and Tn5 cell expressed proteins, while a lower level of high-mannose N-glycans were found on HEK293 and CHO cell expressed proteins.

## REFERENCES

- Shajahan A, Supekar NT, Gleinich AS, Azadi P. 2020. Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*.
- Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*.
- Wu ZL, Person AD, Burton AJ, Singh R, Burroughs B, Fryxell D, Tatge TJ, Manning T, Wu G, Swift KAD, et al. 2019. Direct fluorescent glycan labeling with recombinant sialyltransferases. *Glycobiology*, 29:750-754.

## SARS-COV-2 SPIKE PROTEINS

PROTEIN	SPECIES	SOURCE	TAG	CATALOG #	BIOACTIVITY
SARS-CoV-2 Spike RBD	SARS-CoV-2	HEK293	His	10500-CV	Binds human ACE-2 (933-ZN)
	SARS-CoV-2	CHO	His	10534-CV	
	SARS-CoV-2	Tn5	His	10523-CV	

## ENZYMES

PROTEIN	SPECIES	SOURCE	TAG	CATALOG #	FUNCTION
ST3Gal1	Human	NS0	HA	6905-GT	O-glycan labeling
ST6Gal1	Human	NS0	His	7620-GT	N-glycan labeling
C. p Neuraminidase	Clostridium perfringens	<i>E. coli</i>	His	5080-NM	Remove sialic acid from glycoproteins
Endo EF	Enterococcus faecalis	<i>E. coli</i>	His	8886-GH	Remove Core-1 type O-glycans
PNGase F	Flavobacterium meningosepticum	<i>E. coli</i>	His	9109-GH	Remove N-glycans

## SUBSTRATES

SUBSTRATE	FLUORESCENT TAG	CATALOG #	FUNCTION
CMP-Cy5-Sialic acid	Cy5	ES-302	Donor substrate of Cy5 conjugated sialic acid