bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Comprehensive characterization of N- and O- glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2.

Asif Shajahan, Stephanie Archer-Hartmann, Nitin T. Supekar, Anne S. Gleinich, Christian Heiss, and Parastoo Azadi*.

Complex Carbohydrate Research Center, The University of Georgia, Athens, GA 30602

*Corresponding Author:

Parastoo Azadi, Ph.D.

Analytical Service and Training Laboratory

Complex Carbohydrate Research Center,

University of Georgia,

315 Riverbend Road, Athens, GA 30602

Keywords: COVID-19 receptor, ACE2 glycosylation, SARS-CoV-2 receptor, ACE2 site mapping, hACE2 N-glycosylation.

Supplementary Data Included: Figures S1 – S11.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Emergence of COVID-19 pandemic caused by SARS-CoV-2 demanded development of new therapeutic strategies and thus the understanding the mode of viral attachment, entry and replication has become key aspect for such interventions. The coronavirus surface features a trimeric spike (S) protein that is essential in viral attachment, entry and membrane fusion. The S protein of SARS-CoV-2 binds to the human angiotensin converting enzyme 2 (hACE2) for the entry and the serine protease TMPRSS2 for S protein priming. The heavily glycosylated S protein is comprised of two protein subunits (S1 and S2), and the receptor binding domain within S1 subunit binds with to the hACE2 receptor. Even though hACE2 has been known for two decades and has been recognized as the entry point of several human coronaviruses, no comprehensive glycosylation characterization of hACE2 has been reported. Herein, we describe the quantitative glycosylation mapping on hACE2 expressed in human cells by both glycoproteomics and glycomics. We observed heavy glycan occupancy at all the seven possible N-glycosylation sites and surprisingly, detected three novel O-glycosylation sites. In order to deduce the detailed structure of glycan epitopes on hACE2 involved with viral binding, we have characterized the terminal sialic acid linkages, presence of bisecting GlcNAc and also the pattern of N-glycan fucosylation. We have conducted extensive manual interpretation of each glycopeptide and glycan spectra in addition to the use of bioinformatics tools to validate the hACE2 glycosylation. Elucidation of the site-specific glycosylation and its terminal orientations on the hACE2 receptor can aid in understanding the intriguing virus-receptor interactions and help in the development of novel therapeutics to circumvent the viral entry.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

In late 2019, the emergence of the highly transmissible coronavirus disease (COVID-19) led to a global health crisis within weeks and was soon declared a pandemic. The new underlying pathogen belongs to the family of the *coronaviridae* and has been named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), initially termed 2019 novel coronavirus (2019-nCoV) (Gorbalenya et al. 2020). More than four months into the pandemic, no specific vaccines or treatments have been approved for COVID-19 (World Health Organization (WHO) 2020; Li and De Clercq 2020). The elucidation of key structures involved in the transmission of SARS-CoV-2 will provide insights towards design and development of suitable vaccines and drugs against COVID-19 to curb the global public health crisis (Wang et al. 2020).

Located on the viral surface is the spike (S) protein, which attaches the SARS-CoV-2 pathogen to target cells in the human body. The trimeric spike protein belongs to the class I fusion proteins. Its two subunits S1 and S2 orchestrate its entry into the cell: First, the S1 subunit facilitates the attachment of the virus via its receptor binding domain (RBD) to the host cell receptor, and secondly the S2 subunit mediates the fusion of the viral and human cellular membranes (Hoffmann et al. 2020; Shang et al. 2020; Walls et al. 2020). The glycosylation pattern of the spike protein, which carries N-linked as well as O-linked glycosylation, has recently been the subject of several scientific studies (Shajahan et al. 2020; Watanabe et al. 2020; Zhang et al. 2020). Insight into the glycosylation pattern expands the understanding of the viral binding to receptors, the fusion event, the host cell entry, the replication as well as the design of suitable antigens for vaccine development. Our recent study on the site-specific quantitation of N-linked and O-linked glycosylation of the subunits S1 and S2 of the SARS-CoV-2 spike protein revealed the presence of one confirmed and one putative O-glycosylation sites at the RBD of subunit S1 (Shajahan et al. 2020). A number of amino acids within the RBD have been shown to be crucial determinants for the binding in general and the binding affinity in particular of the virus to the host

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cell receptors (Wan et al. 2020; Andersen et al. 2020). Several studies have identified the human angiotensin-converting enzyme 2 (hACE2, ACEH) as the key virus receptor of SARS-CoV-1 in the early 2000s (Li et al. 2003; Kuba et al. 2005). Moreover, hACE2 was identified recently as the receptor for cell entry of SARS-CoV-2 (Hoffmann et al. 2020; Zhou et al. 2020). However, the SARS-CoV-1 and SARS-CoV-2 pathogens exploit the hACE2 receptor for cell entry only, which is unrelated to its physiological function (Li 2013).

In 2000, hACE2 was identified and classified as a zinc metalloprotease, which enzymatically functions as a carboxypeptidase (Donoghue et al. 2000; Tipnis et al. 2000). This protease is a type-I transmembrane protein and includes an extracellular, a transmembrane and a cytosolic domain within a total of 805 amino acids (Jiang et al. 2014). Interestingly, is has been found that the transmembrane form can be cleaved to a soluble form of hACE2 (sACE2) that lacks the transmembrane and cytosolic domains but is enzymatically active, since the catalytic site as well as the zinc-binding motif lie within the extracellular region (Tipnis et al. 2000). hACE2 is secreted by endothelial cells and is involved in the renin-angiotensin system (RAS) (Donoghue et al. 2000). hACE2 can either cleave the decapeptide angiotensin I (Ang1-10) to the nonapeptide angiotensin 1-9 (Ang1-9) via its catalytic domain, or it can cleave angiotensin II to generate angiotensin 1–7 (Ang1-7), thus contributing to regulate the bioavailability of Ang1-9 and Ang1-7 (Jiang et al. 2014). The involvement in the RAS makes hACE2 a target for the treatment of blood pressure-related diseases (Crackower et al. 2002; Jiang et al. 2014).

The presence of seven potential N-glycosylation sites were experimentally verified through the treatment with exo-glycosidase PNGase F and monitoring the shift in the gel migration of hACE2 to the predicted molecular mass of ~85 kDa. (Tipnis et al. 2000; Li et al. 2003). To our knowledge, no in-depth thorough analysis of the glycosylation of hACE2 has been conducted - with exception of the published findings by Zhao et al. (2015). They investigated the N-linked glycans of hACE2 via sequential exo-glycosidase digestion and identified mainly biantennary N-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

linked glycans with sialylation and core fucosylation, as well as sialylated tri- and tetra-antennary N-glycans. Nevertheless, we could not find any reports with comprehensive site specific glycosylation or glycan structural study on hACE2, even on multiplatform glycosylation resource glygen.org (Kahsay et al. 2020). The study of glycosylation can help in understanding the key roles glycans play during the physiological function of hACE2 and more importantly, its involvement in facilitating viral binding.

Within this study we report the comprehensive and quantitative site-specific N-linked and O-linked glycan characterization of hACE2 by combination of glycoproteomics and glycomics. We identified glycosylation at all the seven potential N-glycosylation sites on hACE2 and also report three novel O-glycosylation sites. The N- & O-glycans were released from hACE2 and we characterized their structure along with location of fucosylation, bisecting GlcNAc and sialic acid linkages by MALDI-MS and ESI-MS^n.

Results

Mapping N-glycosylation on hACE2

We have procured recombinant hACE2 receptor expressed on human HEK 293 cells. The protein was expressed with a C-terminal His tag and comprises residues Gln18 to Ser740. We performed a combination of trypsin and chymotrypsin digestion on reduced and alkylated hACE2 and generated glycopeptides that each contain a single N-linked glycan site. The glycopeptides were subjected to high resolution LC-MS/MS, using a glycan oxonium ion product dependent HCD triggered CID program. The LC-MS/MS data were analyzed using a Byonic software search, each glycopeptide annotation was screened manually, and false detections were eliminated.

We observed heavy glycosylation at all seven predicted N-glycosylation sites of hACE2 (Figure 2, 3, S1-S7). Interestingly, complex-type glycans were much more abundant than high-mannose and hybrid type glycans across the N-glycosylation sites. We have quantified the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

relative intensities of glycans at each site by evaluating the area under the curve of each extracted glycopeptide peak on the LC-MS chromatogram. We discovered highly processed sialylated complex-type bi-antennary, tri-antennary and tetra-antennary glycans on all sites (Figure 4, 5).

Identification of O-Glycosylation on hACE2

Through a glycoproteomics approach, we for the first time have identified novel three O-glycosylation sites on hACE2 by searching the LC-MS/MS data for common O-glycosylation modifications. At present, no reports are available indicating the presence of O-glycosylation on hACE2. We have observed very strong evidence for the presence of O-glycosylation at sites Ser155, Thr496 and Thr730 on peptides SLDYNER, IVGVVEPVPHDETY and LGIQPTLGPPNQP, respectively, as the precursor masses, oxonium ions, neutral losses, and peptides fragments (b and y ions) were detected with high mass accuracy (Figure 2, 3, and 4).

Core-1 mucin type O-glycan GalNAcGalNeuAc$_2$ was observed as predominant glycan on sites Ser155 and Thr730 (Figure 2, 3, and 4). While majority of peptide containing Ser155 was found unglycosylated, almost 97 % of peptide with Thr730 was occupied by O-glycans GalNAcGalNeuAc (2 %) and GalNAcGalNeuAc$_2$ (95 %) (Figure S8, 3, 4). Thr496 was shown to contain only a single HexNAc modification as the spectra showed characteristic oxonium ions (204.09) and neutral losses (Figure S9). Since a single HexNAc modification can come from Tn antigen (GalNAc) or O-GlcNAc modification we could not currently assign the type of monosaccharide occupied at this site, and further experiments will be needed to confirm the identity of this HexNAc. Nevertheless, only 0.4 % of this site was found occupied. Considering that the peptide IVGVVEPVPHDETY does not have any N-glycosylation consensus site and that the O-GlcNAc modification does not typically occur with N- or mucin type O- glycosylation, it is likely that this site is occupied by the Tn antigen (GalNAc).

Comparison of glycosylation sites of human ACE2 with other related species.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We aligned the amino acid sequence of human, bat, pig, cat and mouse ACE2 and compared the glycosylation sites among these species. Figure 5 schematically shows the N- and O-linked glycosylation sites for human, bat, pig, cat and mouse ACE2. Whilst displaying overall sequence similarities of about 67 percent, human ACE2 possesses 7, bat ACE2 (Chinese rufous horseshoe bat) also 7, cat ACE2 a total of 9, porcine ACE2 8 and murine ACE2 only 6 potential N-glycosylation sites. Our sequence alignment studies showed that human ACE2 five N-glycosylation sites share similarities with bat ACE2 but only three N-glycosylation sites with mouse ACE2, which is not susceptible to SARS-CoV-1 binding. Pig ACE2 showed four similar sites with human whereas cat ACE2 showed five sites which are similar (Figure 5).

N- and O- glycomics analysis on hACE2

N- and O-glycomic study of the hACE2 receptor were performed through methods described previously (Shajahan, Heiss, et al. 2017). Briefly N-glycans were released by treating the reduced – alkylated hACE2 protein with the exo-glycosidase PNGase F. The released N-glycans were separated by passing through a C18 solid phase extraction (SPE) cartridge, and de-N-glycosylated proteins on the cartridges were eluted with 80 % acetonitrile with 0.1 % formic acid. The O-linked glycans were then released from hACE2 by reductive *β*-elimination. The released N- and O- glycan fractions were then permethylated, a derivatization method that, in addition to increasing sensitivity, allows for further structural and linkage characterization *via* MS^n fragmentation.

Sequencing of the N-glycans obtained from hACE2 by MALDI-MS and ESI-MS^n indicated a highly diverse pool of glycoforms comprised of high mannose, hybrid, and complex-type structures (Figure S10, 6, Table 1,). The glycans were predominantly complex-type and were primarily (~60%) biantennary, with triantennary and quaternary structures also present in high amounts. More than 85% of the structures were fucosylated, and roughly half of the structures

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were sialylated. The glycan structures were annotated based on high resolution precursor mass, ESI-MS^n fragmentation and the common biosynthetic pathways of mammalian N-glycans.

Core versus antennae fucosylation on hACE2 N-glycans

MS/MS sequencing of the observed N-glycans was conducted with an automated top down program, collecting MS^2 spectra of the highest intensity peaks with collision-induced-dissociation (CID). This helped confirm overall structure (for example hybrid forms versus complex-type) and to determine placement of the fucose. We observed that, while most of the fucosylated structures were primarily core-fucose, small amounts of the same glycoform appeared to be fucosylated on the antennae (Figure 7A). This can be determined by a diagnostic terminal GalGlcNAcFuc fragment of m/z 660 [$M+Na^+$], which corresponds to the fragmentation of the antenna (Figure 7A) appearing as a b ion. The corresponding y ion m/z 1402 [$M+Na^+$] is also found, confirming this structure. Additional glycoforms displaying two fucoses (core and antennal) were also observed among these structures (Figure 6, Table 1).

Identifying the sialic acid linkages on hACE2 N-glycans by ESI-MS^n

A separate MS^n procedure was used to determine the linkages of sialic acids in the complex-type N-glycans in hACE2. The method, which uses direct infusion of the permethylated glycans in a lithium carbonate and methanol solution, fragments down the sialylated arm of a complex-type N-glycan, then fragments the remaining galactose. The cross-ring fragments from this MS^5 method provided diagnostic ions corresponding to 2→3 or 2→6 linked sialic acids (Figure 7B, S10) (Shajahan, Heiss, et al. 2017). While 2→3 linked sialylated N-glycans were the major ones, most structures were a combination of 2→3 and 2→6 linked, and a few were only 2→3 linked. No glycoform was found to be solely 2→6 linked. However, sialylated N-glycans were lower abundant glycoforms, and it is possible that detection limits affected this finding.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Determination of bisecting GlcNAc on hACE2 N-glycans by ESI-MS^n

To determine presence of bisecting GlcNAc, an MS^n strategy to trim down the permethylated N-glycan to the trimannose core was utilized (Ashline et al. 2015). Fragmentation to the core yields a 3-substituted structure at m/z 852 [M+Na$^+$], that would correspond to either a triantennary structure or a biantennary structure with a bisecting GlcNAc. Fragmentation of this ion would yield an m/z of 444 [M+Na$^+$] if the structure was bisecting, or m/z 458 [M+Na$^+$] if it was not. As shown in Figure 7C, both m/zs of 444 and 458 were observed, indicating that both tri-antennary structures, as well as bisecting structures were present in this sample. This supports earlier findings by Zhao et al. who reported both bisecting and higher-order complex-type glycans (Zhao et al. 2015).

Detection and confirmation of O-glycans by MALDI-MS and ESI-MS^n

The released O-glycans from hACE2 were sequenced by both MALDI-MS and ESI-MS^n after permethylation (Figure S11, 7D). The signal intensity was very low, and mostly the disialylated Core-1 O-glycan structure was observed. This is supported by our O-glycoproteomics findings, which showed that O-glycosylation occurs only on three sites, disialylated Core-1 is major glycoform and only one site showed higher occupancy. This could explain the reason for the low signal intensity of O-glycans during the glycomics experiment. The MS^2 fragmentation of the O-glycan confirms its structure as a sialylated Core-1 O-glycan (Figure 7D).

Discussion

hACE2 acts as a receptor for human coronaviruses SARS-CoV-1 and SARS-CoV-2, as well as human coronavirus NL63/HCoV-NL63 (Li et al. 2003; Hoffmann et al. 2020; Hofmann et al. 2005). Recent evidences indicate that the molecular and structural features of SARS-CoV-2 RBD resulted in tight hACE2 binding in comparison to the earlier counterpart SARS-CoV-1 (Andersen et al. 2020; Hoffmann et al. 2020).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

According to recent cryo-EM studies on SARS-CoV-2, the binding of S protein to the hACE2 receptor primarily involves extensive polar residue interactions between RBD and the peptidase domain of hACE2 (Hoffmann et al. 2020; Walls et al. 2020). The RBD located in the S1 subunit of SARS-CoV-2 S protein undergoes a hinge-like dynamic movement which enhances the capture of the spike protein RBD with hACE2. This enhanced affinity for the human ACE2 receptor is predicted to be 10–20-fold higher for SARS-CoV-2 than SARS-CoV-1, which may be responsible for the increased transmissibility of the new virus (Wrapp et al. 2020; Yan et al. 2020). The protease domain of ACE2 interacts with the RBD of coronaviruses, and thus soluble ACE2 (sACE2), which are devoid of neck and transmembrane domains are capable of binding with RBD, neutralizing infection (Yan et al. 2020; Hofmann et al. 2004).

ACE2 is expressed in most vertebrates. The ACE2 variants from human, bat, domestic pig, domestic cat and mouse are all comprises of 805 amino acid residues (Figure 5). A comparison of their ACE2 sequences reveals a sequence identity of about 67 %. SARS-CoV-1, as well as SARS-CoV-2 can infect a wide variety of organisms, including but not limited to humans, palm civets, cats and bats. *In vitro* virus infectivity studies conducted by Zhou et al. (2020) indicated that SARS-CoV-2 is able to exploit the ACE2 proteins from humans, bats, pigs and civets to infect the cell cultures expressing the respective receptor. However, cell culture expressing murine ACE2 was not infected. This is based on a varying degree of receptor recognition. Receptor recognition is largely determined by two factors, (i) the binding specificity and (ii) the binding affinity of the RBD of the virus' spike protein to the cell entry receptor ACE2. This attachment step has been identified as a crucial limiting step for infection, as well as cross-species infection. Two regions within the amino acid sequence of hACE2 have been identified as critical for the interaction with the spike protein. These regions - called 'hotspots' - are located around amino acid residue 31 (hotspot 31) and residue 353 (hotspot 353) (Hou et al. 2010; Li 2013). Our comparison of N-glycosylation sites across species indicated that human ACE share some

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

glycosylation site with other species. The study of correlation of glycosylation sites between the species and their susceptibility to viral binding can help in understanding the key sites involved (Figure 5). According to the data published by Qiu et al. (2020), the overall sequence identity of the ACE2 between different organisms and the hACE2 cannot be directly translated into prediction of transmissibility. For example, mouse ACE2 matches with a higher sequence similarity than bat and pig, but the murine ACE2 cannot be exploited by the coronavirus, the study found (Qiu et al. 2020).

A recent *in silico* study by Stawiski et al. (2020) identified the glycosylation at Asp90 as an important modification that partly disrupts the interaction of the coronavirus with the ACE2 receptor. Therefore, mutations at either N90 or T92 removes the glycosylation motif and makes the unglycosylated variant prone to interactions with SARS-CoV-2. Also, the elucidation of the structure of murine ACE2 would be beneficial for the investigation of insusceptible organisms. It might in turn provide valuable insight into the molecular structures that make humans prone to infection with SARS-CoV-2 via hACE2.

O-glycosylation is initiated by the α-glycosidic attachment of N-acetylgalactosamine (GalNAc) to the hydroxyl group of serine or threonine, and mucin type O-glycosylation is the most common type in higher eukaryotes (Van den Steen et al. 1998; Brockhausen and Stanley 2015). O-glycans are involved in protein stability and function, and have been suggested to play roles in mediating pathogenic binding with human receptors (Mayr et al. 2018). Our analysis confirmed the presence of O-glycosylation at sites Ser155, Thr496 and Thr730 (Figure 2, 3, 4, and 7D). The frequency of occurrence of proline residues is higher adjacent to O-glycosylation sites (Thanka Christlet and Veluraja 2001). The evidence for the O-glycosylation assignments at each of the detected sites on hACE2, particularly Thr730, was strengthened by the presence of proline adjacent to the O-glycosylation residues.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Human and avian virus Hemagglutinins (HAs), including 2009 pandemic H1N1, were shown to bind glycans with sulfation, fucosylation and internal sialylation. It has been reported that the human pandemic H1N1 influenza viruses shift the preference from α2→6-linkages in sialylated Galβ(1,4)Glc/GlcNAc O-glycans to α2→3-linkages in sialylated Galβ(1,3)GalNAc (Chandrasekaran et al. 2008). The sialylated core 1 glycoforms are involved in the life cycle of influenza A virus and plays crucial role during infection (Mayr et al. 2018). The full-length S proteins of SARS-CoV-2 and SARS-CoV-1 share almost 76% identity in amino acid sequences, whereas the N-terminal domains (NTDs) show only 53.5% of homology. The NTDs of different coronavirus S proteins bind with varying avidity to different sugars, the NTD of MERS-CoV prefers α2→3-linked sialic acid over α2→6-linked sialic acid with sulfated sialyl-Lewis X (with antennal fucose) being the preferred binder (Park et al. 2019; Li et al. 2017). Nevertheless, we observed relatively higher level of 2→3 linked sialylated N-glycans on hACE2 and thus the sialic acid linkage could favor the binding of coronavirus S proteins (Figure 7B).

S glycoproteins of coronaviruses HCoV-OC43 and Bat CoV mediates attachment to oligosaccharide receptors by interacting with 9-O-acetyl-sialic acid (Tortorici et al. 2019). No sialic acid binding preference of SARS-CoV-1 or SARS-CoV-2 has been reported and whether the sialic acid linkages on the hACE2 receptor affect virus entry remains to be determined. We have searched for 9-O-acetyl-sialic acid on the hACE2 glycans by extracting the masses of its oxonium ions on the LC-MS/MS spectra but could not detect any evidence for its presence. The expression of 9-O-acetyl-sialic acid depends on the expression level of the sialate O-acetyltransferase gene (CasD1), and only 1 to 2 % of total sialic acid is 9-O-acetylated in HEK93 cells (Barnard et al. 2019). This could be the reason for the lack of 9-O-acetyl-sialic acid on the glycans from hACE2 we used for this study as it is produced from HEK293 cells. We have detected both core fucosylation and antennal fucosylation on hACE2 N-glycans (Figure 7A). Moreover, we found evidence of bisecting GlcNAc on the hACE2 N-glycans (Figure 7C). The implication of such

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

glycan epitopes of hACE2 provide a better understanding of viral binding preferences and can guide the research for the development of suitable therapeutics.

Our comprehensive N- and O-glycosylation characterization of hACE2 expressed in a human cell system through both glycoproteomics and glycomics provides key understanding to elucidate the roles glycans play in the function of hACE2 and, more importantly, how these glycans mediate pathogenic invasion. We have conducted extensive manual interpretation strategies for the assignment of each glycopeptide, glycan and linkage structures in order to eliminate false detections. We are currently exploring the protein polymorphism on hACE2 and how the glycosylation profile varies in the variants which alter the glycosylation sites as it is reported that natural ACE2 variants that are predicted to alter the virus-host interaction and thereby potentially alter host susceptibility (Stawiski et al. 2020).

Detailed glycan analysis is important for the development of hACE2 or sACE2 based therapeutics which are suggested as a therapeutic measure to neutralize the viral pathogens (Hofmann et al. 2004; Yan et al. 2020). Evaluation of glycosylation on glycoprotein therapeutics produced from various human and non-human expression systems is critical from the point of view of immunogenicity, stability as well as therapeutic efficacy (Sola and Griebenow 2010; Beck, Cochet, and Wurch 2010). The understanding of complex sialylated N-glycans and sialylated mucin type O-glycans, on hACE2, along with their linkage and structural isomerism provides the basic structural knowledge that is useful for elucidating their interaction with viral surface protein and can aid in future therapeutic possibilities.

Materials and methods

Dithiothreitol (DTT), iodoacetamide (IAA), and iodomethane were purchased from Sigma Aldrich (St. Louis, MO). Sequencing-grade modified trypsin and chymotrypsin were purchased from Promega (Madison, WI). Peptide-N-Glycosidase F (PNGase F) was purchased from New

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

England Biolabs (Ipswich, MA). All other reagents were purchased from Sigma Aldrich unless indicated otherwise. Data analysis was performed using Byonic 2.3 software and manually using Xcalibur 4.2. The purified human angiotensin converting enzyme hACE2 (Cat. No. 230-30165) was purchased from RayBiotech (Atlanta, GA).

Reduction, alkylation and protease digestion of hACE2 for glycoproteomics

The purified hACE2 (20 µg) expressed on HEK293 cells in 50 mM ammonium bicarbonate solution was reduced by adding 25 mM DTT and incubating at 60 °C for 30 min. The protein was further alkylated by the addition of 90 mM IAA and incubating at RT for 30 min in dark. Subsequently, the protein was desalted by 10kDa centrifuge filter and digested by sequential treatment with trypsin and chymotrypsin by incubating for 18 h at 37 °C during each digestion step. The digest was filtered through 0.2 µm filter and directly analyzed by LC-MS/MS.

N- and O-linked glycan release, purification, and permethylation

N- and O-linked glycans were released by following the methods described previously (Shajahan, Heiss, et al. 2017). N-linked glycans were released from about 80 µg of reduced and alkylated (as mentioned previously) hACE2 sample by treatment with PNGase F at 37˚C for 16 hours. The released N-glycans were isolated by passing the digest through a C18 SPE cartridge with a 5% acetic acid solution (3 mL) and dried by lyophilization.

The remaining de-N-glycosylated hACE2 protein which contain O-glycans were then eluted from the column using 80 % aqueous acetonitrile with 0.1 % formic acid (3 mL). The O-glycans were released from the peptide backbone by reductive *β*-elimination (Shajahan, Heiss, et al. 2017). The eluted hACE2 were treated to a solution of 19mg/500 µL of sodium borohydride in a solution of 50mM sodium hydroxide. The solution was heated to 45 ˚C for 16 hours, then neutralized with a solution of 10% acetic acid. The sample was desalted on a hand-packed ion exchange resin (DOWEX H+) by eluting with 5% acetic acid and dried by lyophilization. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

borates were removed by the addition of a solution of methanol:acetic Acid (9:1) and evaporation under a steam of nitrogen.

The released N- and O-linked glycans were then permethylated using methods described elsewhere (Shajahan, Heiss, et al. 2017).

Data acquisition of protein digest samples using nano-LC-MS/MS

The glycoprotein digests were analyzed on an Orbitrap Fusion Tribrid mass spectrometer equipped with a nanospray ion source and connected to a Dionex binary solvent system (Thermo Fisher, Waltham, MA). Pre-packed nano-LC column (Cat. No. 164568, Thermo Fisher, Waltham, MA) of 15 cm length with 75 µm internal diameter (id), filled with 3 µm C18 material (reverse phase) were used for chromatographic separation of samples. The precursor ion scan was acquired at 120,000 resolution in the Orbitrap analyzer and precursors at a time frame of 3 s were selected for subsequent MS/MS fragmentation in the Orbitrap analyzer at 15,000 resolution. The LC-MS/MS runs of each digest were conducted for 180 min in order to separate the glycopeptides. The threshold for triggering an MS/MS event was set to 1000 counts, and monoisotopic precursor selection was enabled. MS/MS fragmentation was conducted with stepped HCD (Higher-energy Collisional Dissociation) product triggered CID (Collision-Induced Dissociation) (HCDpdCID) program. Charge state screening was enabled, and precursors with unknown charge state or a charge state of +1 were excluded (positive ion mode). Dynamic exclusion was also enabled (exclusion duration of 30 secs).

Data analysis of glycoproteins

The LC-MS/MS spectra of combined tryptic /chymotryptic digest of hACE2 were searched against the FASTA sequence of hACE2 using the Byonic software 2.3 by choosing appropriate peptide cleavage sites (semi-specific cleavage option enabled). Oxidation of methionine, deamidation of asparagine and glutamine, and possible common human N-glycans and O-glycan

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

masses were used as variable modifications. The LC-MS/MS spectra were also analyzed manually for the glycopeptides with the support of the Thermo Fisher Xcalibur 4.2 software. The HCDpdCID MS^2 spectra of glycopeptides were evaluated for the glycan neutral loss pattern, oxonium ions and glycopeptide fragmentations to assign the sequence and the presence of glycans in the glycopeptides.

N- and O-linked glycomic profiling by MALDI-MS

The permethylated N- and O-glycans were dissolved in 20 µL of methanol. 0.5 µL of sample was mixed with equal volume of DHB matrix solution (10 mg/mL in 1:1 methanol-water) and spotted on to a MALDI plate. MALDI-MS spectra were acquired in positive ion and reflector mode using an AB Sciex 5800 MALDI-TOF-TOF mass spectrometer.

N- and O-linked glycomic profiling by DI-ESI-MS^n

A solution of the permethylated N-glycans from hACE2 were diluted into a solution of 1 mM sodium hydroxide/50% MeOH and directly infused (0.5 µL/min) into an Orbitrap Fusion Tribrid mass spectrometer equipped with a nanospray ion source. An automated program was used to collect a full MS then MS^2 of the highest-intensity peaks. The top 300 peaks collected from a m/z range of 800-2000 were fragmented by CID, with a dynamic exclusion of 60 sec. The total run time was 20 minutes. The full MS was collected in the Orbitrap at a resolution of 120,000 while the MS^2 spectra were collected in the Orbitrap at 60,000. A similar automated program was used for the collection of O-glycan data, with a m/z range adjusted to 600-1600.

Original glycoform assignments were made based on full-mass molecular weight. Additional structural details were determined by MS^n and modeling with GlycoWorkbench 2 software.

Sialic Acid Linkage Analysis by IT-MS spectrometry including MS^n fragmentation

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A solution of the permethylated N-glycans from hACE2 were diluted into a solution of 1 mM lithium carbonate/50% MeOH and directly infused (0.5 µL/min) into an Orbitrap Fusion Tribrid mass spectrometer equipped with a nanospray ion source. The sialic acid-containing N-glycans (determined by MALDI-TOF-MS and ESI-MS^n experiments) were probed with MS^n analysis described previously (Shajahan, Supekar, et al. 2017; Anthony et al. 2008; Lin et al. 2015). Isolation was conducted in the quadrupole while detection was conducted in the IT. The isolation width of each fragmentation was 2 mass units and the maximum injection time was 100 ms. More than 300 spectra were collected for each glycoform, then spectrally averaged.

Determination of bisecting GlcNAc by IT-MS spectrometry including MS^n fragmentation

A solution of the permethylated N-glycans from hACE2 were diluted into a solution of 1 mM sodium hydroxide/50% MeOH and directly infused (0.5 µL/min) into an Orbitrap Fusion Tribrid mass spectrometer equipped with a nanospray ion source. The neutral complex-type N-glycans (determined by MALDI-TOF-MS and ESI-MS^n experiments) were probed with MS^n analysis described previously (Ashline et al. 2015). Because of the extra fragmentation steps, sialylated N-glycans were not probed since they yielded too low of a signal. Isolation was conducted in the quadrupole while detection was conducted in the IT. The isolation width of each fragmentation was 2 mass units and the maximum injection time was 100ms. More than 300 spectra were collected for each glycoform, then spectrally averaged.

Acknowledgements

Financial support from the US National Institutes of Health (S10OD018530) is gratefully acknowledged. This work was also supported in part by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award DE-SC0015662 to DOE - Center for Plant and Microbial Complex Carbohydrates at the Complex Carbohydrate Research Center, USA. We thank Rupali Mahadik (CCRC, UGA) for help with the retrieval of hACE2 data from glygen.org.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abbreviations

Spike – S, receptor binding domain - RBD, coronavirus disease - COVID-19, human angiotensin-converting enzyme 2 - hACE2, hemagglutinins - HAs, human embryonic kidney - HEK, dithiothreitol – DTT, iodoacetamide – IAA, Peptide-N-Glycosidase F - PNGase F, solid phase extraction - SPE, acetonitrile - ACN, Higher-energy Collisional Dissociation – HCD, product triggered – pd, Collision-Induced Dissociation – CID, IT-MS – ion trap-mass spectrometry.

Contributions

P.A. and A.S. conceived of the paper; A.S. performed glycoproteomics and glycomics sample processing, S.A.H. conducted glycomics data acquisition, A.S., S.A.H., N.S. and A.G. performed data analysis; everyone contributed toward writing the paper; P.A. and C.H. monitored the project.

Competing interests

The authors certify that they have no competing interests.

Supplementary material: Annotated MS/MS spectra (S1 – S9) are incorporated as a separate supplementary material file.

The raw data files and search results can be accessed from glycopost repository - https://glycopost.glycosmos.org/preview/11623990675eab9b3c18219; pin: 1672

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Andersen, K. G., A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry. 2020. 'The proximal origin of SARS-CoV-2', *Nat Med*, 26: 450-52.

Anthony, R. M., F. Nimmerjahn, D. J. Ashline, V. N. Reinhold, J. C. Paulson, and J. V. Ravetch. 2008. 'Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc', *Science*, 320: 373-6.

Ashline, D. J., M. Duk, J. Lukasiewicz, V. N. Reinhold, E. Lisowska, and E. Jaskiewicz. 2015. 'The structures of glycophorin C N-glycans, a putative component of the GPC receptor site for Plasmodium falciparum EBA-140 ligand', *Glycobiology*, 25: 570-81.

Barnard, K. N., B. R. Wasik, J. R. LaClair, D. W. Buchholz, W. S. Weichert, B. K. Alford-Lawrence, H. C. Aguilar, and C. R. Parrish. 2019. 'Expression of 9-O- and 7,9-O-Acetyl Modified Sialic Acid in Cells and Their Effects on Influenza Viruses', *mBio*, 10.

Beck, A., O. Cochet, and T. Wurch. 2010. 'GlycoFi's technology to control the glycosylation of recombinant therapeutic proteins', *Expert Opin Drug Discov*, 5: 95-111.

Brockhausen, I., and P. Stanley. 2015. 'O-GalNAc Glycans.' in rd, A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger (eds.), *Essentials of Glycobiology* (Cold Spring Harbor (NY)).

Chandrasekaran, A., A. Srinivasan, R. Raman, K. Viswanathan, S. Raguram, T. M. Tumpey, V. Sasisekharan, and R. Sasisekharan. 2008. 'Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin', *Nat Biotechnol*, 26: 107-13.

Crackower, M. A., R. Sarao, G. Y. Oudit, C. Yagil, I. Kozieradzki, S. E. Scanga, A. J. Oliveira-dos-Santos, J. da Costa, L. Zhang, Y. Pei, J. Scholey, C. M. Ferrario, A. S. Manoukian, M. C. Chappell, P. H. Backx, Y. Yagil, and J. M. Penninger. 2002. 'Angiotensin-converting enzyme 2 is an essential regulator of heart function', *Nature*, 417: 822-8.

Donoghue, M., F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robison, R. Jeyaseelan, R. E. Breitbart, and S. Acton. 2000. 'A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9', *Circ Res*, 87: E1-9.

Gorbalenya, Alexander E., Susan C. Baker, Ralph S. Baric, Raoul J. de Groot, Christian Drosten, Anastasia A. Gulyaeva, Bart L. Haagmans, Chris Lauber, Andrey M. Leontovich, Benjamin W. Neuman, Dmitry Penzar, Stanley Perlman, Leo L. M. Poon, Dmitry Samborskiy, Igor A. Sidorov, Isabel Sola, and John Ziebuhr. 2020. 'Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group', *bioRxiv*: 2020.02.07.937862.

Hoffmann, M., H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche, M. A. Muller, C. Drosten, and S. Pohlmann. 2020. 'SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor', *Cell*.

Hofmann, H., M. Geier, A. Marzi, M. Krumbiegel, M. Peipp, G. H. Fey, T. Gramberg, and S. Pohlmann. 2004. 'Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor', *Biochem Biophys Res Commun*, 319: 1216-21.

Hofmann, H., K. Pyrc, L. van der Hoek, M. Geier, B. Berkhout, and S. Pohlmann. 2005. 'Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry', *Proc Natl Acad Sci U S A*, 102: 7988-93.

Hou, Y., C. Peng, M. Yu, Y. Li, Z. Han, F. Li, L. F. Wang, and Z. Shi. 2010. 'Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry', *Arch Virol*, 155: 1563-9.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Jiang, F., J. Yang, Y. Zhang, M. Dong, S. Wang, Q. Zhang, F. F. Liu, K. Zhang, and C. Zhang. 2014. 'Angiotensin-converting enzyme 2 and angiotensin 1-7: novel therapeutic targets', *Nat Rev Cardiol*, 11: 413-26.

Kahsay, R., J. Vora, R. Navelkar, R. Mousavi, B. Fochtman, X. Holmes, N. Pattabiraman, R. Ranzinger, R. Mahadik, T. Williamson, S. Kulkarni, G. Agarwal, M. Martin, P. Vasudev, L. J. G. Castro, N. Edwards, W. Zhang, D. A. Natale, K. Ross, K. F. Aoki-Kinoshita, M. P. Campbell, W. S. York, and R. Mazumder. 2020. 'GlyGen data model and processing workflow', *Bioinformatics*.

Kuba, K., Y. Imai, S. Rao, H. Gao, F. Guo, B. Guan, Y. Huan, P. Yang, Y. Zhang, W. Deng, L. Bao, B. Zhang, G. Liu, Z. Wang, M. Chappell, Y. Liu, D. Zheng, A. Leibbrandt, T. Wada, A. S. Slutsky, D. Liu, C. Qin, C. Jiang, and J. M. Penninger. 2005. 'A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury', *Nat Med*, 11: 875-9.

Li, F. 2013. 'Receptor recognition and cross-species infections of SARS coronavirus', *Antiviral Res*, 100: 246-54.

Li, G. D., and E. De Clercq. 2020. 'Therapeutic options for the 2019 novel coronavirus (2019-nCoV)', *Nature Reviews Drug Discovery*, 19: 149-50.

Li, W., R. J. G. Hulswit, I. Widjaja, V. S. Raj, R. McBride, W. Peng, W. Widagdo, M. A. Tortorici, B. van Dieren, Y. Lang, J. W. M. van Lent, J. C. Paulson, C. A. M. de Haan, R. J. de Groot, F. J. M. van Kuppeveld, B. L. Haagmans, and B. J. Bosch. 2017. 'Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein', *Proc Natl Acad Sci U S A*, 114: E8508-E17.

Li, W., M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan. 2003. 'Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus', *Nature*, 426: 450-4.

Lin, N., J. Mascarenhas, N. R. Sealover, H. J. George, J. Brooks, K. J. Kayser, B. Gau, I. Yasa, P. Azadi, and S. Archer-Hartmann. 2015. 'Chinese hamster ovary (CHO) host cell engineering to increase sialylation of recombinant therapeutic proteins by modulating sialyltransferase expression', *Biotechnol Prog*, 31: 334-46.

Mayr, J., K. Lau, J. C. C. Lai, I. A. Gagarinov, Y. Shi, S. McAtamney, R. W. Y. Chan, J. Nicholls, M. von Itzstein, and T. Haselhorst. 2018. 'Unravelling the Role of O-glycans in Influenza A Virus Infection', *Sci Rep*, 8: 16382.

Park, Y. J., A. C. Walls, Z. Wang, M. M. Sauer, W. Li, M. A. Tortorici, B. J. Bosch, F. DiMaio, and D. Veesler. 2019. 'Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors', *Nat Struct Mol Biol*, 26: 1151-57.

Qiu, Y., Y. B. Zhao, Q. Wang, J. Y. Li, Z. J. Zhou, C. H. Liao, and X. Y. Ge. 2020. 'Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2', *Microbes Infect*.

Shajahan, A., C. Heiss, M. Ishihara, and P. Azadi. 2017. 'Glycomic and glycoproteomic analysis of glycoproteins-a tutorial', *Anal Bioanal Chem*, 409: 4483-505.

Shajahan, A., N. T. Supekar, C. Heiss, M. Ishihara, and P. Azadi. 2017. 'Tool for Rapid Analysis of Glycopeptide by Permethylation via One-Pot Site Mapping and Glycan Analysis', *Anal Chem*, 89: 10734-43.

Shajahan, Asif, Nitin T. Supekar, Anne S. Gleinich, and Parastoo Azadi. 2020. 'Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2', *bioRxiv*: 2020.04.01.020966.

Shang, Jian, Gang Ye, Ke Shi, Yushun Wan, Chuming Luo, Hideki Aihara, Qibin Geng, Ashley Auerbach, and Fang Li. 2020. 'Structural basis of receptor recognition by SARS-CoV-2', *Nature*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Sola, R. J., and K. Griebenow. 2010. 'Glycosylation of therapeutic proteins: an effective strategy to optimize efficacy', *BioDrugs*, 24: 9-21.

Stawiski, Eric W., Devan Diwanji, Kushal Suryamohan, Ravi Gupta, Frederic A. Fellouse, J. Fah Sathirapongsasuti, Jiang Liu, Ying-Ping Jiang, Aakrosh Ratan, Monika Mis, Devi Santhosh, Sneha Somasekar, Sangeetha Mohan, Sameer Phalke, Boney Kuriakose, Aju Antony, Jagath R. Junutula, Stephan C. Schuster, Natalia Jura, and Somasekar Seshagiri. 2020. 'Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility', *bioRxiv*: 2020.04.07.024752.

Thanka Christlet, T. H., and K. Veluraja. 2001. 'Database analysis of O-glycosylation sites in proteins', *Biophys J*, 80: 952-60.

Tipnis, S. R., N. M. Hooper, R. Hyde, E. Karran, G. Christie, and A. J. Turner. 2000. 'A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase', *J Biol Chem*, 275: 33238-43.

Tortorici, M. A., A. C. Walls, Y. Lang, C. Wang, Z. Li, D. Koerhuis, G. J. Boons, B. J. Bosch, F. A. Rey, R. J. de Groot, and D. Veesler. 2019. 'Structural basis for human coronavirus attachment to sialic acid receptors', *Nat Struct Mol Biol*, 26: 481-89.

Van den Steen, P., P. M. Rudd, R. A. Dwek, and G. Opdenakker. 1998. 'Concepts and principles of O-linked glycosylation', *Crit Rev Biochem Mol Biol*, 33: 151-208.

Walls, A. C., Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler. 2020. 'Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein', *Cell*.

Wan, Y., J. Shang, R. Graham, R. S. Baric, and F. Li. 2020. 'Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus', *J Virol*, 94.

Wang, N., J. Shang, S. Jiang, and L. Du. 2020. 'Subunit Vaccines Against Emerging Pathogenic Human Coronaviruses', *Front Microbiol*, 11: 298.

Watanabe, Yasunori, Joel D. Allen, Daniel Wrapp, Jason S. McLellan, and Max Crispin. 2020. 'Site-specific analysis of the SARS-CoV-2 glycan shield', *bioRxiv*: 2020.03.26.010322.

World Health Organization (WHO). 2020. 'Coronavirus, Licence: CC BY-NC-SA 3.0 IGO'. https://www.who.int/health-topics/coronavirus#tab=tab_1.

Wrapp, D., N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan. 2020. 'Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation', *Science*, 367: 1260-63.

Yan, R., Y. Zhang, Y. Li, L. Xia, Y. Guo, and Q. Zhou. 2020. 'Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2', *Science*, 367: 1444-48.

Zhang, Yong, Wanjun Zhao, Yonghong Mao, Shisheng Wang, Yi Zhong, Tao Su, Meng Gong, Dan Du, Xiaofeng Lu, Jingqiu Cheng, and Hao Yang. 2020. 'Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins', *bioRxiv*: 2020.03.28.013276.

Zhao, X., F. Guo, M. A. Comunale, A. Mehta, M. Sehgal, P. Jain, A. Cuconati, H. Lin, T. M. Block, J. Chang, and J. T. Guo. 2015. 'Inhibition of endoplasmic reticulum-resident glucosidases impairs severe acute respiratory syndrome coronavirus and human coronavirus NL63 spike protein-mediated entry by altering the glycan processing of angiotensin I-converting enzyme 2', *Antimicrob Agents Chemother*, 59: 206-16.

Zhou, P., X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, and Z. L. Shi. 2020. 'A pneumonia outbreak associated with a new coronavirus of probable bat origin', *Nature*, 579: 270-73.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and Tables

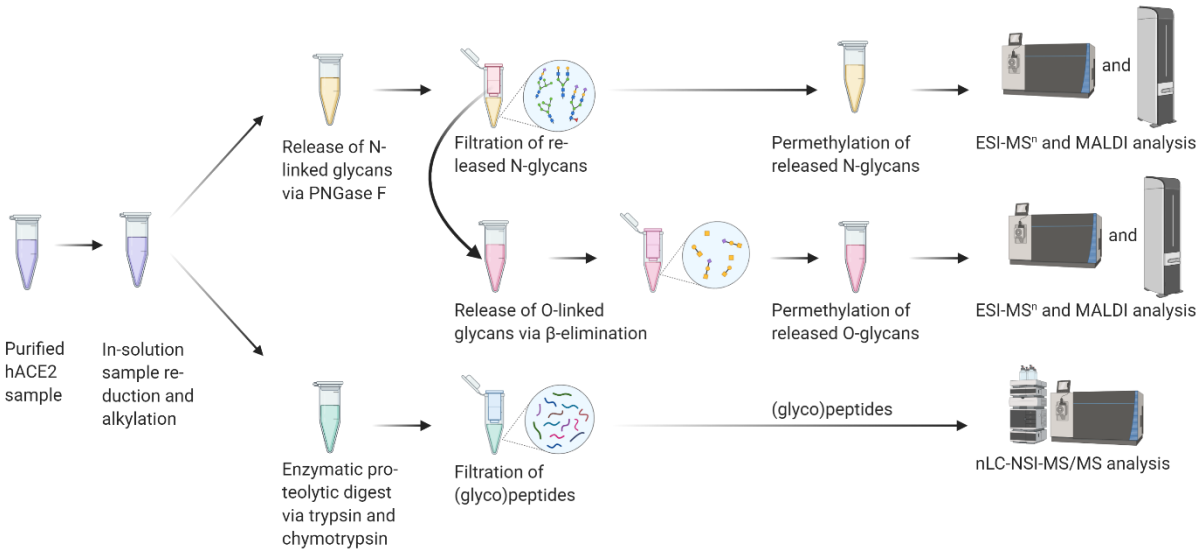


Figure 1: The N- and O-glycosylation on recombinantly expressed hACE2 from HEK293 cells were comprehensively analyzed by glycoproteomics and glycomics. The reduced (DTT), alkylated (IAA) hACE2 were treated separately with proteases (trypsin and chymotrypsin) for glycopeptide analysis and PNGase F for N-glycan release. Subsequently, the protease digests were analyzed by nLC-NSI-MS/MS, and the released N-glycans were analyzed by MALDI-MS and ESI-MS^n after permethylation. The de-N-glycosylated protein fraction was subjected to *β*-elimination for the O-glycan release and the released O-glycans were also analyzed by MALDI-MS and ESI-MS^n after permethylation (created with biorender.com).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

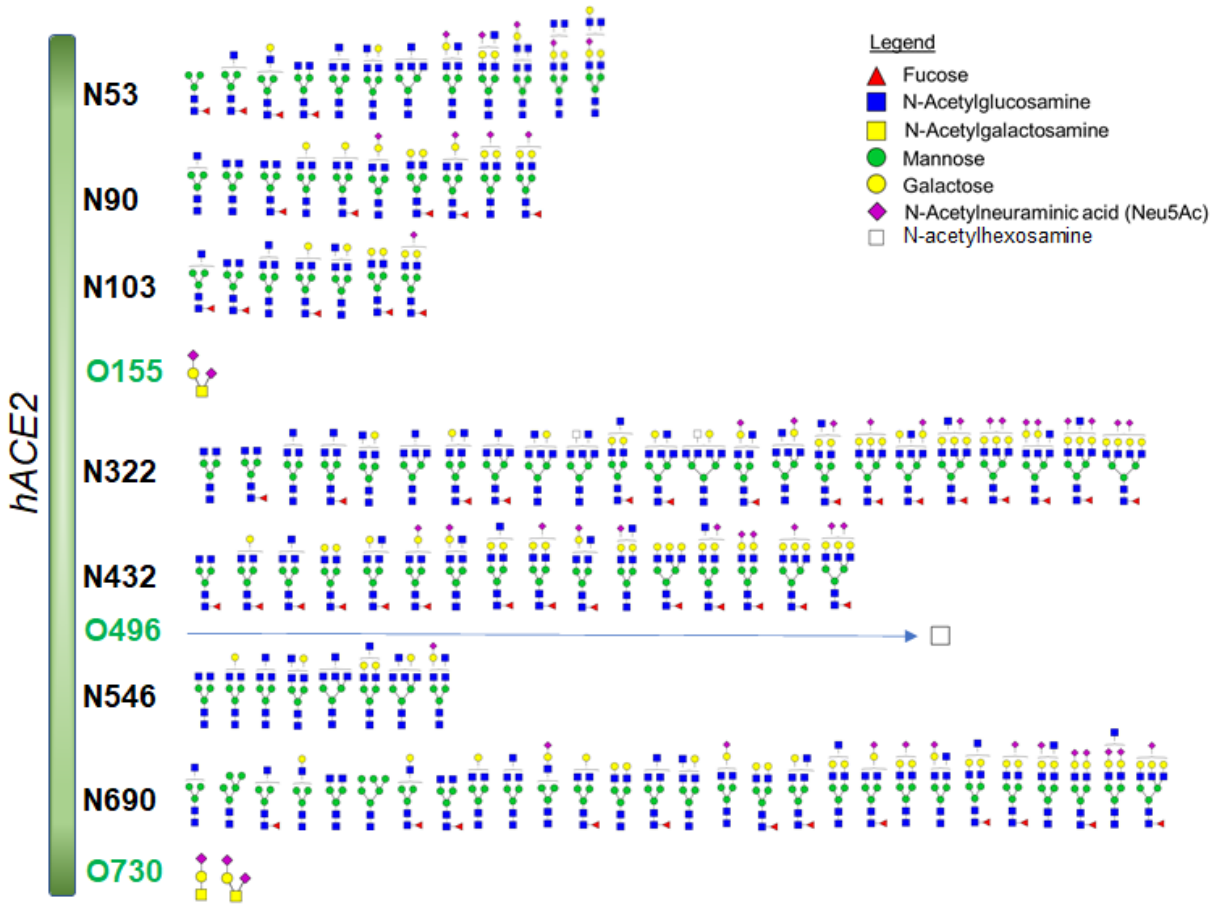


Figure 2: Glycosylation profile on hACE2 characterized by high-resolution LC-MS/MS. All the seven potential N-glycosylation sites were found occupied along with three O-glycosylation sites bearing core-1 type O-glycans and single HexNAc. Mostly complex type glycans were observed in all N-glycosylation sites. Some N-glycosylation sites were partially glycosylated. Monosaccharide symbols follow the SNFG (Symbol Nomenclature for Glycans) system (Varki, A., Cummings, R.D., et al. 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

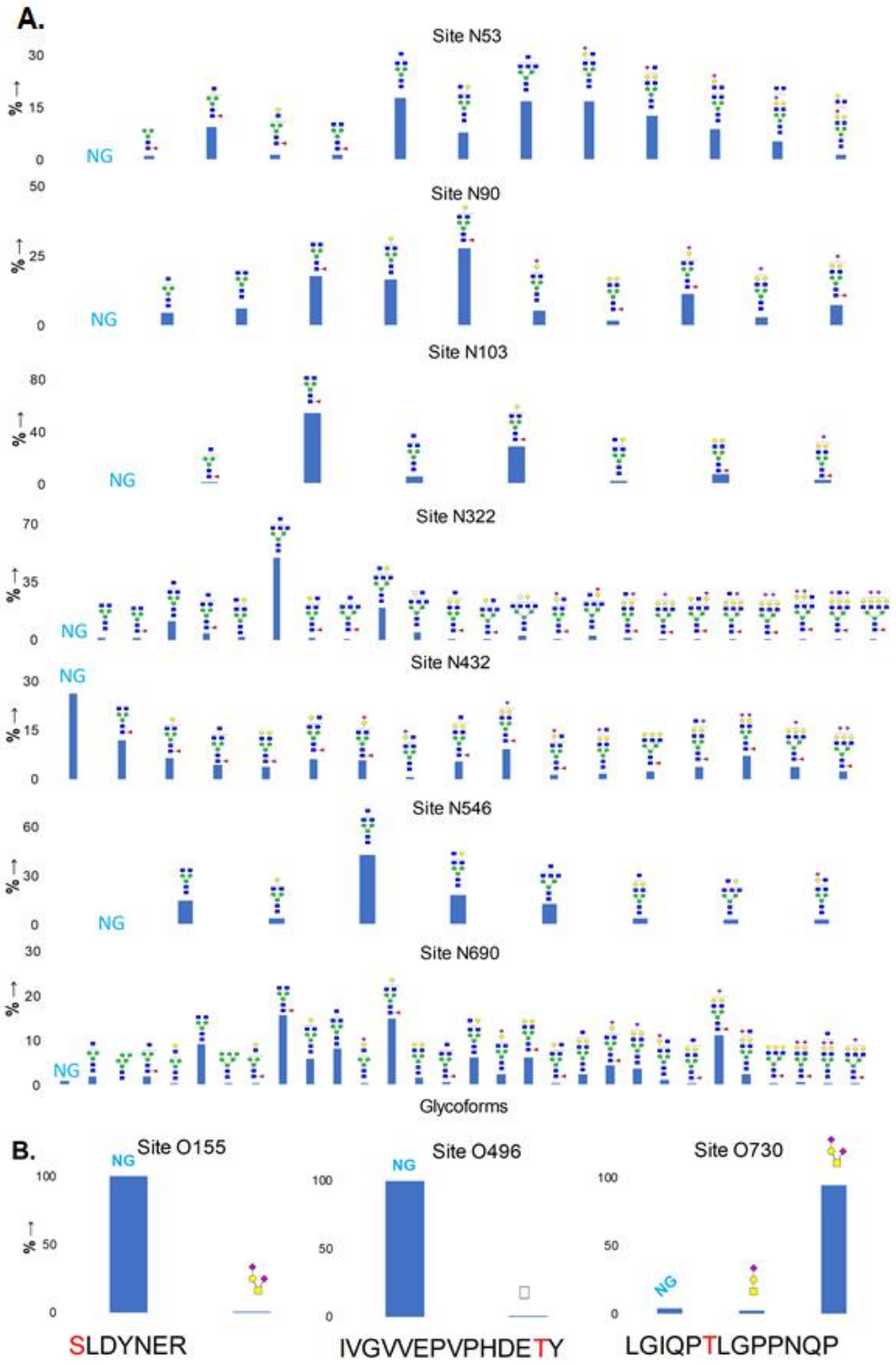


Figure 3: Quantitative glycosylation profile of glycans on hACE2 characterized by high-resolution LC-MS/MS. **A.** relative ratio of non-glycosylated peptide and glycoforms detected on seven N-glycosylation sites; **B.** relative ratio of non-glycosylated peptide and glycoforms detected on three O-glycosylated sites. RA – Relative abundances. NG- non-glycosylated peptide. Monosaccharide symbols follow the SNFG system (Varki, A., Cummings, R.D., et al. 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

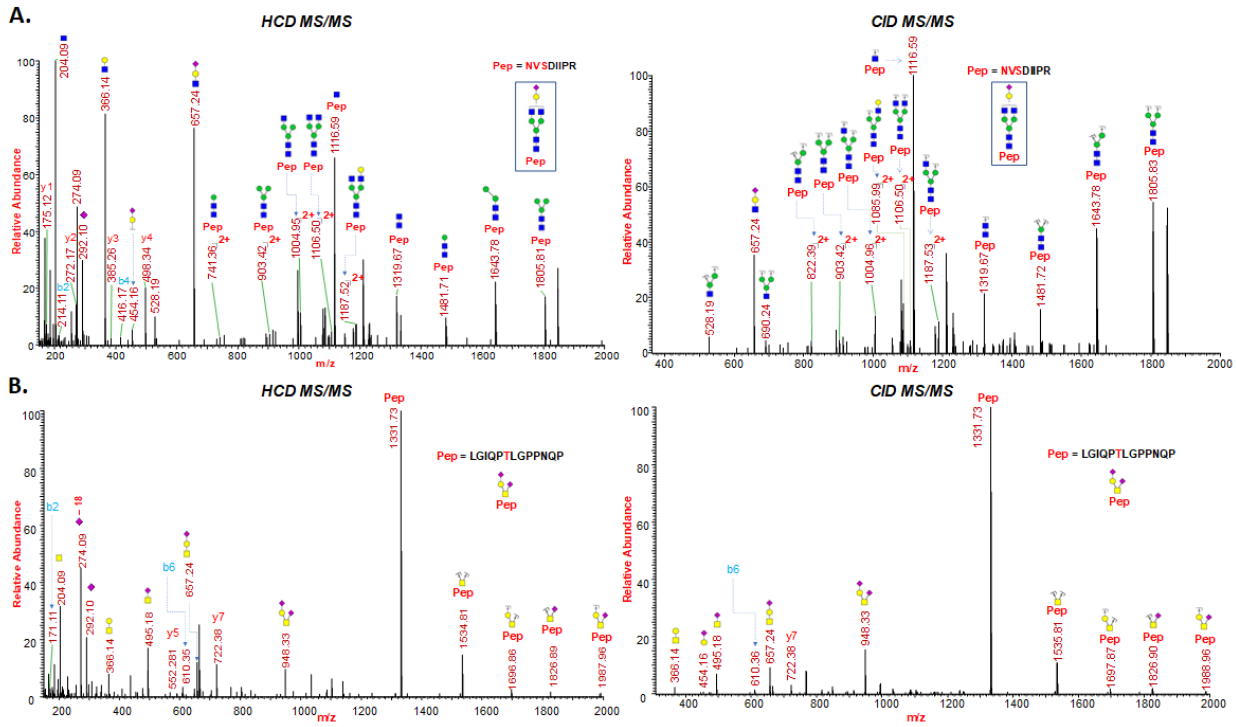


Figure 4: A. Representative HCD and CID MS/MS spectra of intact N-glycopeptide NVSDIIPR with assigned N-glycan ($GlcNAc_2Man_3GlcNAc_2Gal_1NeuAc_1$) at N690, **B.** Representative HCD and CID MS/MS spectra of intact O-glycopeptide with assigned O-glycan ($GalNAc_1Gal_1NeuAc_2$) at O730.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

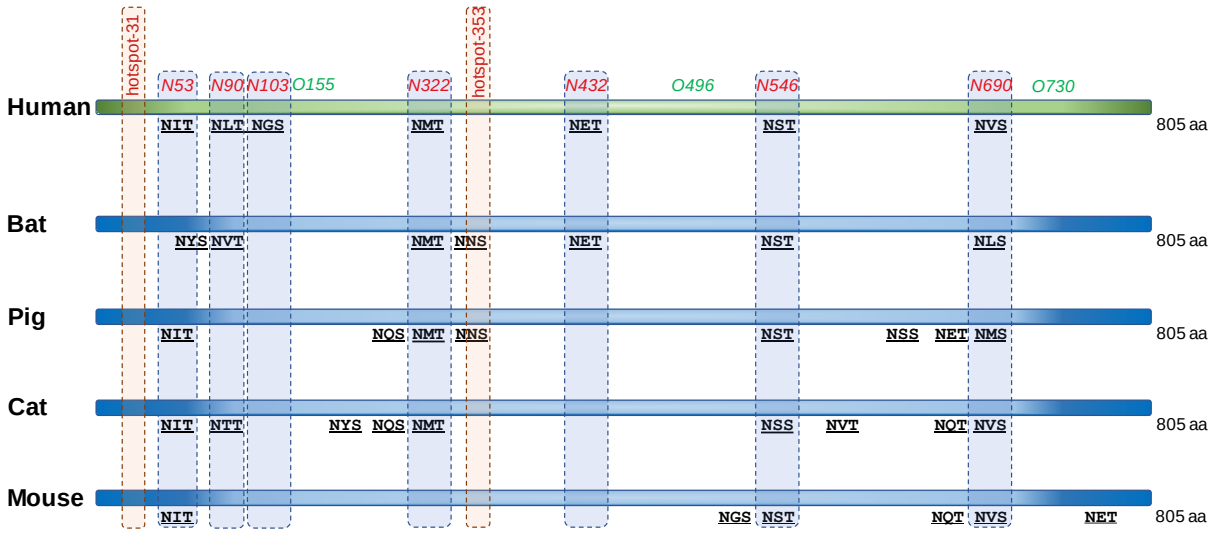


Figure 5: Schematic amino acid sequence alignment of glycosylation sites of ACE2 receptor from human, bat, pig, cat and mouse (Uniprot identifiers human [Q9BYF1], bat [U5WHY8], pig [K7GLM4], cat [Q56H28], mouse [Q8R0I0]). The ACE2 sequences between all five organisms share about 67 % sequence identity. The number of N-linked glycosylation motifs (-NXS/T-) accounts for 7 potential sites in the human ACE2 (highlighted in light blue), 7 potential N-linked glycosylation sites in the bat ACE2, 8 potential N-glycan sites in the porcine ACE2, 9 potential sites in the feline ACE2, and 6 potential N-linked glycosylation sites in the murine ACE2. The two regions to be critical for the interaction with the coronavirus (hotspot-31 at Lys31 and hotspot-353 at Lys353) are highlighted (light red).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

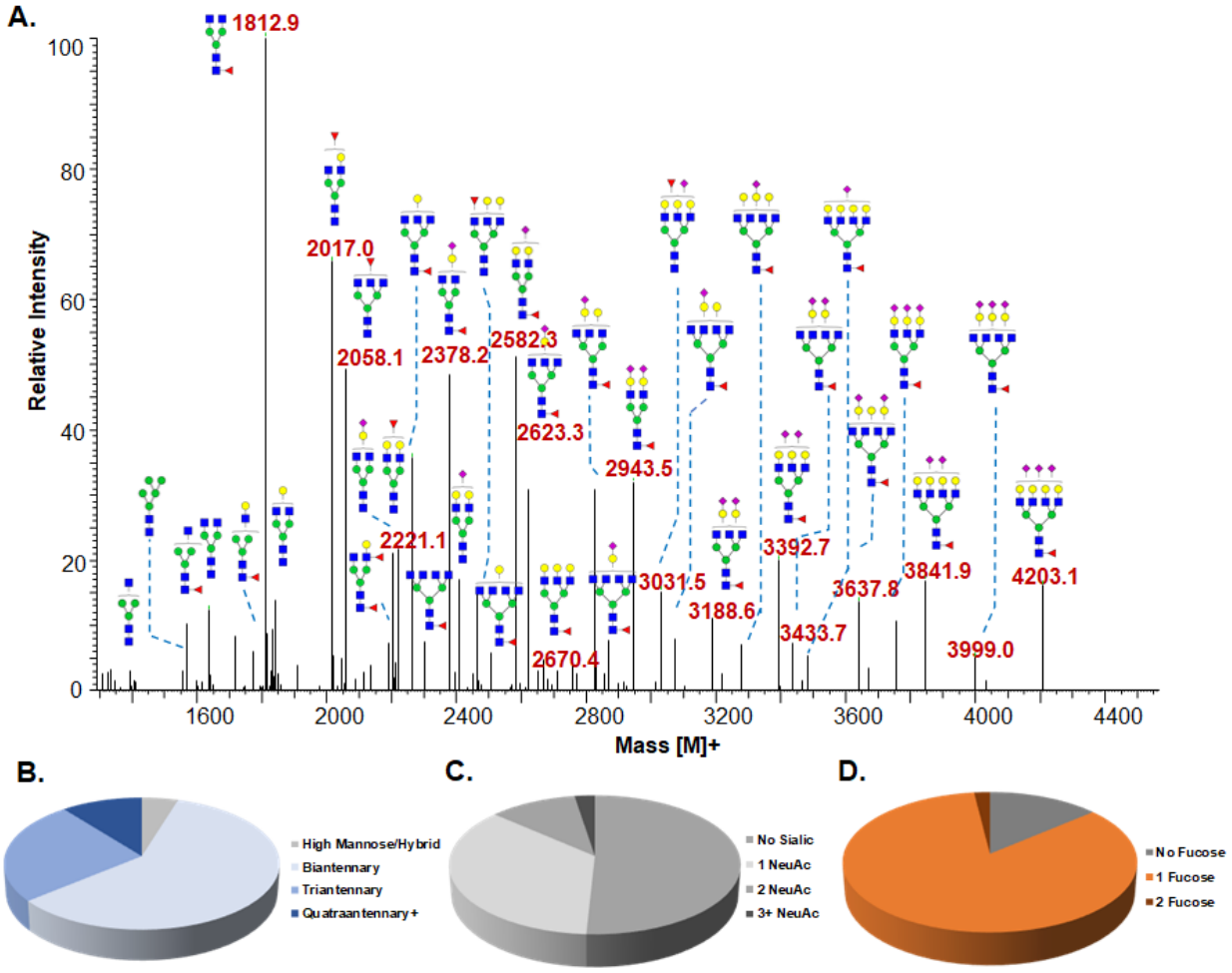


Figure 6: N-glycomics profiling of hACE2: N-glycans were released, purified, and permethylated prior to analysis with ESI-MS^n. **A**. Deconvoluted profile of permethylated N-glycans from hACE*2* (Only major glycoforms are shown). Masses are shown as deconvoluted, molecular masses. The N-glycan structure were confirmed by ESI-MS^n; **B.** Percentage breakdowns of N-glycan class: High mannose/Hybrid 5%; Biantennary 59%; Triantennary 25%; Tetraantennary 11%. **C.** Percentage breakdowns of Sialyation: Nonsialylated 51%; 1 NeuAc 36%; 2 NeuAc 11%; 3< NeuAc 3%. **D.** Percentage breakdown of Fucosylation: No Fucose 14%; 1 Fucose 84%; 2 Fucose 2%.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

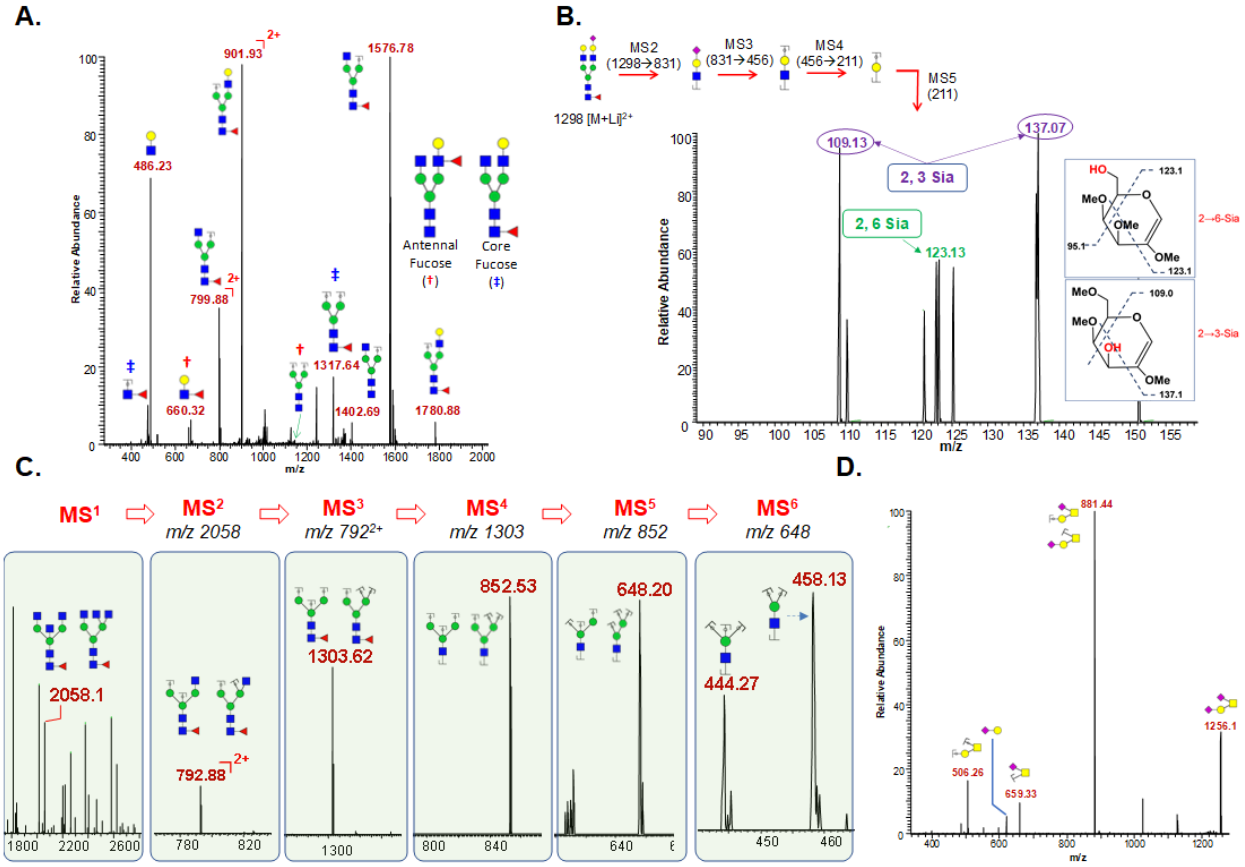
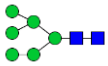
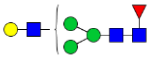
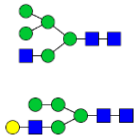
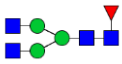
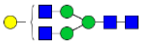
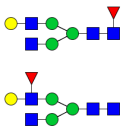
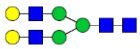
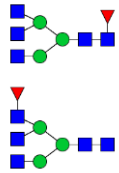
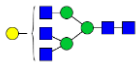
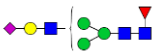


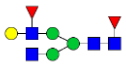
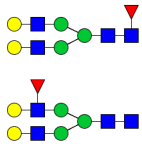
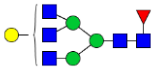
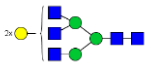
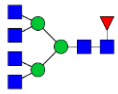
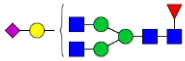
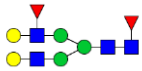
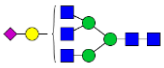
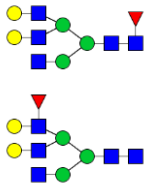
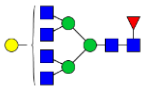
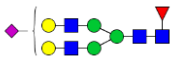
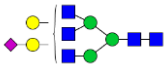
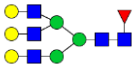
Figure 7: A. MS^2 Fragmentation of an N-glycan observed at m/z 1133 (sodiated; z=2). MS^2 Fragmentation reveals that there is a mixture of glycoforms that contain either core-linked or antennally-linked fucose; **B.** MS^5 Fragmentation of a sialylated N-glycan observed at m/z 1298 (lithium adduct; z=2). MS^5 Fragmentation breaks down the antennal arm to the sialylated galactose. Cross-ring fragmentation determines whether the sialyation was 2,3 or 2,6 linked; **C**: MS^6 Fragmentation of a complex-type N-glycan observed at m/z 1052 (sodium adduct; z=2). MS^6 Fragmentation breaks down the structure to the trimannose core. Fragmentation reveals the number of substituents, consistent with bisecting GlcNAc. **D:** MS^2 Fragmentation of an O-glycan observed at m/z 1256 (sodiated; z=1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

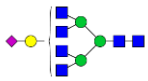
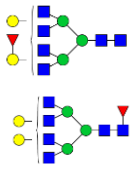
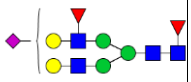
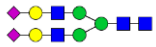
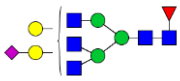
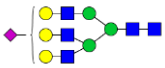
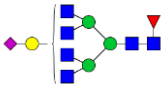
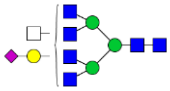
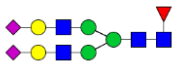
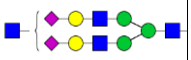
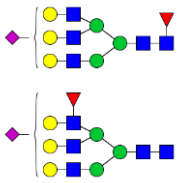
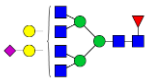
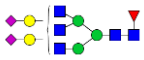
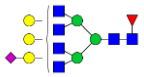
Table 1: Relative abundances and structures of N-glycans detected by glycomics analysis on hACE2. Masses of permethylated glycans are represented as molecular masses [M+] and the structures were assigned based on common mammalian N-glycosylation synthesis pathway and also by ESI-MS^n analysis.

Mass (M)	Mass Error	Composition	Proposed Structure	Relative Percent
1393.7	-0.002	$(HexNAc)_1 + (Man)_3(GlcNAc)_2$		0.65%
1556.8	-0.001	$(Hex)_2 + (Man)_3(GlcNAc)_2$		0.50%
1567.8	-0.002	$(HexNAc)_1$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		1.68%
1638.8	-0.004	$(HexNAc)_2 + (Man)_3(GlcNAc)_2$		2.01%
1760.9	0.002	$(Hex)_3 + (Man)_3(GlcNAc)_2$		0.23%
1771.9	-0.002	$(Hex)_1$ $(HexNAc)_1$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		0.88%
1801.9	-0.001	$(Hex)_2$ $(HexNAc)_1$ + $(Man)_3(GlcNAc)_2$		0.15%
1812.9	-0.004	$(HexNAc)_2$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		14.15%
1842.9	-0.003	$(Hex)_1$ $(HexNAc)_2$ + $(Man)_3(GlcNAc)_2$		2.08%
2017.0	-0.004	$(Hex)_1$ $(HexNAc)_2$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		9.62%
2047.0	-0.002	$(Hex)_2$ $(HexNAc)_2$ + $(Man)_3(GlcNAc)_2$		0.00%
2058.1	-0.004	$(HexNAc)_3$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		6.91%
2088.1	-0.004	$(Hex)_1$ $(HexNAc)_3$ + $(Man)_3(GlcNAc)_2$		0.24%
2133.1	-0.003	$(Hex)_1$ $(HexNAc)_1$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.67%

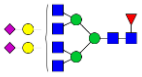
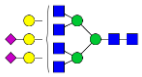
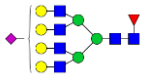
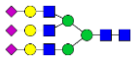
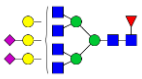
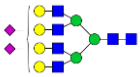
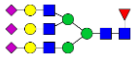
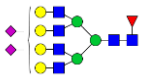
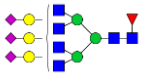
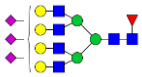
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2191.1	-0.003	$(Hex)_1$ $(HexNAc)_2$ $(Deoxyhexose)_2$ + $(Man)_3(GlcNAc)_2$		1.14%
2204.1	-0.004	$(Hex)_1$ $(HexNAc)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		2.84%
2221.1	-0.004	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		3.26%
2262.2	-0.004	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		1.94%
2292.2	-0.003	$(Hex)_2$ $(HexNAc)_3$ + $(Man)_3(GlcNAc)_2$		0.07%
2303.2	-0.006	$(HexNAc)_4$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		0.96%
2378.2	-0.004	$(Hex)_1$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		8.26%
2395.2	-0.008	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_2$ + $(Man)_3(GlcNAc)_2$		0.20%
2408.2	-0.004	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		2.47%
2449.2	-0.004	$(Hex)_1$ $(HexNAc)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.40%
2466.3	-0.005	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		2.16%
2507.3	-0.006	$(Hex)_1$ $(HexNAc)_4$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		0.88%
2582.3	-0.005	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		8.19%
2623.3	-0.005	$(Hex)_1$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		3.76%
2653.3	-0.005	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.50%
2670.4	-0.004	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		0.62%

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2694.4	-0.003	$(Hex)_1$ $(HexNAc)_4$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.15%
2711.4	-0.002	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_1$ + $(Man)_3(GlcNAc)_2$		0.48%
2756.4	-0.006	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_2$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.60%
2769.4	-0.004	$(Hex)_2$ $(HexNAc)_2$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.42%
2827.4	-0.005	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		3.51%
2857.4	-0.003	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.30%
2868.5	0	$(Hex)_1$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		1.06%
2939.5	-0.002	$(Hex)_1$ $(HexNAc)_5$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.02%
2943.5	-0.006	$(Hex)_2$ $(HexNAc)_2$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		3.98%
3014.5	-0.009	$(Hex)_2$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.02%
3031.5	0.001	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.15%
3072.6	-0.005	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		1.14%
3188.6	-0.005	$(Hex)_2$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		1.33%
3218.6	-0.009	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.36%
3276.7	-0.007	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.68%

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.01.071688; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3392.7	-0.008	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		2.10%
3433.7	-0.008	$(Hex)_2$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.57%
3463.7	-0.008	$(Hex)_3$ $(HexNAc)_4$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.13%
3480.7	-0.007	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_1$ + $(Man)_3(GlcNAc)_2$		0.57%
3579.8	-0.008	$(Hex)_3$ $(HexNAc)_3$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$		0.01%
3637.8	-0.008	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.91%
3667.8	-0.009	$(Hex)_4$ $(HexNAc)_4$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		0.17%
3753.9	-0.006	$(Hex)_3$ $(HexNAc)_3$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$		1.00%
3841.9	-0.007	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_2$ + $(Man)_3(GlcNAc)_2$		1.43%
3999.0	-0.007	$(Hex)_3$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$		0.33%
4203.1	-0.003	$(Hex)_4$ $(HexNAc)_4$ $(Deoxyhexose)_1$ $(NeuAc)_3$ + $(Man)_3(GlcNAc)_2$		1.17%