Mass Spectrometry Analysis of SARS-CoV-2 Nucleocapsid Protein Reveals Camouflaging Glycans and Unique Post-Translational Modifications

Zeyu Sun#, Xiaoqin Zheng#, Feiyang Ji, Menghao Zhou, Xiaoling Su, Keyi Ren, Lanjuan Li

Abstract
The devastating coronavirus disease 2019 (COVID-19) pandemic has prompted worldwide efforts to study structural biological traits of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and its viral components. Compared to the Spike protein, which is the primary target for currently available vaccines or antibodies, knowledge about other virion structural components is incomplete. Using high-resolution mass spectrometry, we report a comprehensive post-translational modification (PTM) analysis of nucleocapsid phosphoprotein (NCP), the most abundant structural component of the SARS-CoV-2 virion. In addition to phosphoryl groups, we show that the SARS-CoV-2 NCP is decorated with a variety of PTMs, including N-glycans and ubiquitin. Based on newly identified PTMs, refined protein structural models of SARS-CoV-2 NCP were proposed and potential immune recognition epitopes of NCP were aligned with PTMs. These data can facilitate the design of novel vaccines or therapeutics targeting NCP, as valuable alternatives to the current vaccination and treatment paradigm that is under threat of the ever-mutating SARS-CoV-2 Spike protein.

Keywords: glycosylation; mass spectrometry; nucleocapsid protein; post-translational modification; SARS-CoV-2

Introduction
The ongoing devastating coronavirus disease 2019 (COVID-19) pandemic is caused by the newly identified severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which can cause lower respiratory tract infections and atypical pneumonia.1-3 Compared to its close relative SARS-CoV, SARS-CoV-2 is more contagious.4-6 As in other coronaviruses, the SARS-CoV-2 virion is composed of the trimeric Spike protein (S), the membrane glycoprotein (M), and the envelope protein (E) on its surface, and nucleocapsid phosphoprotein (N, or NCP) in its core.7 Tremendous efforts have been made to elucidate structural and functional details of protein S due to its essential role in host attachment, thus making it the primary target for vaccine or antibody development.1-3,8-11 However, structural and functional details of other structural proteins of SARS-CoV-2 have not been investigated thoroughly.

Being the most abundant structural protein, NCP constitutes the coronavirus (CoV) core that encapsulates and stabilizes the viral genome.12 In infected cells, NCP plays a vital role in viral transcription and virion assembly.12-16 The multifunctional roles of NCP were also revealed by studies showing its ability to disrupt host innate immune responses17-19 and to manipulate host gene expression.20 At the early stage of SARS-CoV infections, NCP can trigger immediate and strong immune responses. Therefore, both NCP and anti-NCP antibodies can be detected in the serum of SARS-CoV patients, even at the early stage of infections.21 Studies have demonstrated that NCP of SARS-CoV or SARS-CoV-2 is a more sensitive indicator of early infections than IgG or viral RNA.22,23 Because SARS-CoV-2 NCP shares high homology with its SARS-CoV counterpart,24 it is presumed to have similar value for diagnostic applications.23,25-28 Moreover, the strong immunogenicity of NCP also makes it an attractive target to develop SARS-CoV vaccines as an alternative to protein S.29 Analogously, SARS-CoV-2 NCP has also been proposed as a candidate target for the development of...
vaccines and antiviral drugs against COVID-19. This idea was further prompted by increasing concerns over the efficacy of currently deployed vaccines or antibodies targeting SARS-CoV-2 protein S, which mutates rapidly worldwide. Encouragingly, recent studies have shown that SARS-CoV-2 patients are more likely to develop T cell responses against NCP than against protein S, and are able to generate durable B cell memory for both NCP and protein S.

The SARS-CoV NCP was previously found to contain several important post-translational modifications (PTMs), primarily in the form of phosphorylation. Phosphorylation of SARS-CoV-2 NCP was also recently proposed as a key factor in modulating nucleocapsid assembly. However, little is known about NCP decoration with other forms of PTMs that may regulate its functions or interactions with host and viral components. In particular, as bioinformatic analyses predict five possible glycosylation sequons (47NNT, 77NSS, 192NSS, 196NST, 269NVT) in NCP, it will be interesting to test if those sites contain glycans, which may contribute greatly to its immunogenicity and functions. In this study, we report a comprehensive analysis of the SARS-CoV-2 NCP PTM profile using high-resolution mass spectrometry, based on which a protein structural model was proposed to highlight possible biological or immunological implications. For the first time, we showed that SARS-CoV-2 NCP is decorated with N-glycans and ubiquitin in addition to phosphoryl groups.

Results

Landscape of PTMs on SARS-CoV-2 NCP

To portray the PTM landscape, liquid chromatography with tandem mass spectrometry (LC-MSMS) data of peptides, derived by digestion of SARS-CoV-2 NCP with multiple proteases, were subjected to iterative spectra-database matching for common PTMs (Figure 1). Detailed PTM identification can be found in Supplementary Table 1, http://links.lww.com/IMD/A9. Our data unveiled that SARS-CoV-2 NCP is decorated with a plethora of PTMs (Figure 2A), including acetylation (K342, K375), succinylation (K387, K388), di-Gly (K169, K374, K388), phosphorylation (S2, S26, S176, S180, S184, T265, T379, T391, T393, S410, S416) and methylation (E136) as summarized in Table 1. All domains of NCP were decorated with PTMs, with the majority of PTMs clustered in the N terminal domain (NTD) and C tail of NCP. As NCP is expected to be heavily modified by phosphorylation, this study identified five novel NCP phosphorylation sites in addition to previous archived sites, amounting to a total of 27 identified phosphorylation sites out of 80 potential sites predicted by GPS 5.0 in SARS-CoV-2 NCP (Figure 2B, Supplementary Table 2, http://links.lww.com/IMD/A10). Example mass spectra of sequences containing phosphorylation identification can be found in Supplementary Figure 1, http://links.lww.com/IMD/A12. Other PTMs, such as crotonylation (K), farnesylation (K/Nterm), myristoylation (K/Nterm), palmitoylation, or prenylation (C) were not found in this study.

Confirmation of ubiquitination of SARS-CoV-2 NCP

The di-Gly residue on lysine can be the result of tryptic digestion either by ubiquitin or ISG15 modifier. To resolve this ambiguity, SARS-CoV-2 NCP expressed in 293T cells was immunoprecipitated and then immunoblotted with either anti-ubiquitin or anti-ISG15 antibodies. As compared to the negative control sample (IgG immunoprecipitated), SARS-CoV-2 NCP was clearly modified by ubiquitin rather than by ISG15 (Figure 2C).

Determination of N-glycosylation on SARS-CoV-2 NCP

To characterize glycosylation, glycopeptides from SARS-CoV-2 NCP digested with multiple proteases were subjected to hydrophilic interaction liquid chromatography-solid phase extraction (HILIC SPE). To determine glycosylated sites, peptides

---

Figure 1. Overall workflow for comprehensive PTM survey of SARS-CoV-2 NCP using multi-enzyme digestion and enrichment techniques. DTT: dithiothreitol; FASP: filter aided sample preparation; HILIC: hydrophilic interaction liquid chromatography; IAA: iodacetamide; NCP: nucleocapsid phosphoprotein; PTM: post-translational modification; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; SIMAC: sequential elution from IMAC (immobilized metal ion affinity based chromatography).
were deglycosylated by PNGase F in H₂O¹⁸ creating a +2.98 Da mass shift to mark glycosylated sites. Out of five possible N-X-S/T glycosylation sites, two sites (N77, N269, Table 1) in NCP were confirmed by intact glycopeptide characterization, while no evidence of N-glycosylation was found for the other three potential sites (N47, N192, and N196). Example mass spectra showing evidence of glycosylation were shown in Supplementary Figure 2A and 2B, http://links.lww.com/IMD/A12.

Intact glycopeptides were also analyzed directly by LC-MSMS to resolve the highly heterogeneous glycan components. Our analysis identified 55 unique N-glycopeptides, deemed as a unique amino acid sequence with a unique N-glycan composition (Supplementary Table 3, http://links.lww.com/IMD/A11). Glycans in both sites (N77, N269) were confirmed by intact glycopeptide profiling. The composition analysis by pGlyco (Figure 3A) suggested that N77 was preferentially occupied by complex type N-glycans, with 13 possible configurations comprising nearly 95% of total LC-MS intensity of all glycopeptides involving site N77, while the remaining 5% intensity was contributed by a hybrid N-glycan. According to LC-MS intensity, the most dominant glycan on N77 was Fuc2Hex3HexNAc6, which is predicted to be a fucosylated bi-antennary complex. In comparison, nearly 79% of LC-MS intensity of all N269-related glycopeptides can be attributed to six high-mannose oligosaccharides, while the remaining 21% was contributed by complex and hybrid N-glycans. The most dominant glycan on N269 was Hex3HexNAc2. So far, no evidence of sialic acid conjugation to NCP glycans was found at either site. Example N-glycopeptides spectra can be found in Supplementary Figure 2C and 2D, http://links.lww.com/IMD/A12.

Based on the structures of the most likely N-glycans derived in this study and the SARS-CoV-2 NCP structure models, we generated atomic models that represented the most likely spatial distribution of the N-glycans on the SARS-CoV-2 NCP (Figure 3B and 3C). Since the full structure of NCP has not been determined yet, currently available partial structure models of the N-terminal RNA-binding domain (7ACS) and C-terminal dimerization domain (6WZQ) were used to resolve glycans on site N77 and N296, respectively.

Alignment of PTMs with immunogenic epitopes on SARS-CoV-2 NCP

As PTMs, particularly in the form of glycans, can drastically change immunogenicity of a protein, we checked whether PTMs are located in potential epitope areas of SARS-CoV-2 NCP. However, because currently there is a lack of experimental data
on SARS-CoV-2 NCP immune recognition, we resorted to bioinformatic prediction tools to generate potential epitope sequences. Out of 11 B cell epitopes of NCP predicted by Bepipred, six involved phosphorylation, ubiquitination, succinylation, acetylation, and/or glycosylation (Table 2). In particular, multiple phosphorylation sites in the SR-rich domain (S176, S180, S184) and ubiquitination (K169) were found in epitope 165–216. Another C-tail epitope (358–402) was decorated by a variety of PTMs including phosphorylation (T379, T391, T393), ubiquitination (K374, K388), acetylation (K375), and succinylation (K387, K388). To highlight, glycosylation (N77) was also found in a B cell epitope (S9–105). In addition, as a viral core protein, the NCP is likely to be endogenously processed by CD8+ T cells and presented as major histocompatibility complex class I (MHC-I)-associated peptides. Out of 11 T-cell epitope peptides predicted by NetMHCpan, only two contain PTMs, including ubiquitination (K169) in epitope 165–173 and acetylation (K342) in epitope 338–346.

**Discussion**

In coronavirus-infected cells, NCP is mainly located in the cytoplasm. It is able to self-associate into an oligomer and bind to viral RNA via its NTD, thus forming the ribonucleoprotein complex that protects viral RNA and aids virus replication.12–15 On top of this, the multifunctional NCP has also been shown to participate in many host cellular processes, including gene transcription, signal transduction, apoptosis regulation, interferon inhibition, cytoskeleton reorganization, and cell cycle regulation.13,14,17,20,50,51 Given its importance, a detailed survey of PTMs may help to extend our understanding of the structural details and biological functions of SARS-CoV-2 NCP.

CoV NCP is widely known as a phosphoprotein.7,43,44 Multiple phosphorylation sites have been mapped to the SR-rich region in SARS-CoV NCP, which have been shown to play roles in the polymerization and translocation of NCP and its binding to viral RNA and host heteronuclear ribonucleoproteins.44,52,53 However, previous LC-MS-based studies have revealed more phosphorylation sites outside of the SR-rich region.47,48 Our data also support that SARS-CoV-2 NCP is heavily decorated by phosphoryl groups across all domains. Compared with previous studies,47,48 our data have identified five novel phosphorylation sites, all located in the C-tail region where phosphorylation has not been discovered before. The difference of phosphorylation profiles between our study and two previous studies can probably be attributed to differential bias of enrichment techniques, and varying cellular consequences of viral infection and NCP transfection. Given that there is as many as 80 potential phosphorylation sites in SARS-CoV-2 NCP, and the technical bias of the peptide-centric LC-MS approach against hydrophilic phosphorylated sequences, more phosphorylation sites are likely to be mapped in future studies with complementary techniques or different expression systems. Future studies are also warranted to explore virological or immunological implications of the complex NCP phosphorylation profile, particularly for those sites outside of the SR-rich region, such as the C-tail region, which has been demonstrated to have high antigenicity on SARS-CoV-2 NCP.58 Other than phosphorylation, additional forms of PTMs such as acetylation, methylation, and succinylation were also found in SARS-CoV-2 NCP. Acetylation can reduce the charge state and increase the hydrophobicity of modified sites. Further studies are needed to investigate possible biological roles of these PTMs.

To the best of our knowledge, this is the first time ubiquitination has ever been identified in a CoV protein. Other
than causing protein degradation events, ubiquitination also triggers signaling or protein translocation. The locations of the newly identified ubiquitination sites on SARS-CoV-2 NCP are interesting: site K169 resides immediately before the SR-rich domain, while sites K374 and K388 both sit within the predicted bipartite nuclear localization signal domain (NLS-BP, KKKKADETQALPQRQKKQ), which is responsible for NCP translocation from the cytoplasm to the nucleus. Notably, alignment analysis (Supplementary Figure 3A, http://links.lww.com/IMD/A12) suggests that all three ubiquitination sites are strictly conserved in NCP of SARS-CoV-2 (K169, K374, K388), SARS-CoV (K170, K375, K389) and bat CoV HKU3 (K169, K374, K388). In fact, ubiquitination, rather than ISGylation, can also be detected by immunoblotting of NCP from SARS-CoV (Supplementary Figure 3B, http://links.lww.com/IMD/A12). The impact of these ubiquitination events on the cellular location and

Figure 3. SARS-CoV-2 NCP decorated with glycans. A: Relative LC-MS intensity of four major N-glycan categories on both sites. B: The most likely glycan structure on site N77 was added to the NTD RNA-binding domain SARS-CoV-2 NCP based on NTD binding model with dsRNA (Protein Data Bank ID: 7ACS, sequence 44-180). C: The most likely glycan structure on site N269 was added to the CTD dimerization domain SARS-CoV-2 NCP based on model of four CTD subunits polymerized (Protein Data Bank ID: 6WZQ, sequence 247-364), with each subunit colored differently. All NCP sequences were represented in cartoon mode while glycans were represented in sphere mode. CTD: C terminal domain; dsRNA: double-stranded RNA; LC-MS: liquid chromatography with mass spectrometry; NCP: nucleocapsid phosphoprotein; NTD: N terminal domain; PTM: post-translational modification; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.
functions of NCP remains to be investigated. It will also be interesting to know which E3 ligase is responsible for the ubiquitination of NCP. Interestingly, E3 ligase TRIM25 was recently found to be an anti-viral host determinant, and an interactor of SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV). Therefore, future studies are needed to test if SARS-CoV-2 NCP can be ubiquitinated by TRIM25.

Glycosylation is typically found on the viral surface or envelope proteins, and it has been shown to be a critical determinant for viral pathogenesis and immunogenicity, and thus has been considered as a key molecular aspect for antiviral treatment or vaccine development. For instance, CoV Spike proteins, including those of SARS-CoV and SARS-CoV-2, are heavily masked by glycan camouflage, with important implications on host attachment, immune responses, as well as virion assembly and budding. Glycosylation can also be found on CoV membrane proteins. However, so far there has been no report about glycosylation of non-surface proteins such as CoV NCP. In our study, glycosylation of SARS-CoV-2 NCP was evidenced indirectly by PNGase F treatment in H2O, which leaves an isotopic mark to the glycosylation locus. Alignment analysis of CoV sequences (Supplementary Figure 4) revealed that only N68 in NCP of the highly pathogenic MERS-CoV shares homology with site N77 in SARS-CoV-2 NCP. In contrast, potential glycosylation sites that are homologous to N269 of SARS-CoV-2 NCP can be found across multiple CoV species: SARS-CoV, H-CoV-229E, Bat-CoV-HKU3. To further confirm NCP glycosylation, the glycan component in these sites was determined directly by LC-MSMS as well. Typically, due to incomplete glycan maturation, viral protein glycans are mainly composed of short oligomannose branches, as in the case of site N269 in the C terminal domain domain. Nonetheless, site N77 in the NTD domain was almost completely occupied by complex glycans. These heterogeneous N-glycan patterns greatly extend conformational flexibility and epitope diversity of NCP and may profoundly contribute to the unique biological and clinical characteristics of SARS-CoV-2 as compared to other CoVs.

The predicted immuno-epitopes in SARS-CoV-2 NCP aligned with PTMs are shown in Table 2.

Table 2

| Start | End | Epitope sequence                          | Length | Score | PTM                |
|-------|-----|-------------------------------------------|--------|-------|--------------------|
| B cell epitope sequences predicted by Bepipred algorithm |
| 4     | 15  | NPQONRNAPRI                               | 12     | 0.60  | –                  |
| 17    | 48  | FGQPSDSTQNSQNGERS                         | 32     | 0.68  | Phosphorylation: S26|
| 59    | 105 | HIGEVLIKFPRGQGVPI                          | 47     | 0.56  | Glycosylation: N77  |
| 119   | 127 | AGLPYGWAN                                  | 9      | 0.54  | –                  |
| 137   | 163 | GALNTPKDDHTRINPA                           | 27     | 0.58  | –                  |
| 165   | 216 | TILPKGYAEAGGRRGS                           | 52     | 0.66  | Ubiquitination: K169; Phosphorylation: S176, S180, S184 |
| 226   | 267 | RNLQLEGSKGSGQOGQOQGVYTVKSSAAES             | 42     | 0.60  | Phosphorylation: S265|
| 276   | 299 | RGGQETQIDGMDQG                           | 24     | 0.56  | –                  |
| 343   | 348 | DPFND                                    | 6      | 0.53  | –                  |
| 358   | 402 | DAVYTPFRPTEPKKD                             | 45     | 0.62  | Ubiquitination: K374, K388; Acetylation: K375; Phosphorylation: T379, T391, T393; Succinylation: K387, K388 |
| 404   | 416 | SKLGQGSSMSSADS                            | 13     | 0.58  | Phosphorylation: S410, S416 |

MHC-I epitope sequences predicted by Net MHC pan EL algorithm

| Start | End | Epitope sequence                          | Length | Score | PTM                |
|-------|-----|-------------------------------------------|--------|-------|--------------------|
| 112   | 121 | YLLQGPEAGL                                | 10     | 0.229 | –                  |
| 158   | 167 | VQEGPGVTLL                                | 10     | 0.197 | K169               |
| 165   | 173 | TILPGQEPY                                 | 9      | 0.095 | –                  |
| 220   | 230 | ALLLDRLLNLQ                                | 11     | 0.956 | –                  |
| 226   | 234 | RNLQLEGSKM                                | 9      | 0.126 | –                  |
| 305   | 313 | AGFAPASA                                  | 9      | 0.176 | –                  |
| 316   | 324 | GMSRIGMEV                                 | 9      | 0.422 | –                  |
| 330   | 339 | WLHYGQKRL                                 | 10     | 0.066 | –                  |
| 336   | 346 | KLQDQMPN                                  | 9      | 0.352 | Acetylation: K342  |
| 351   | 359 | ILNQHDA                                   | 9      | 0.125 | –                  |
| 397   | 407 | AAADDIEFSKQI                              | 11     | 0.191 | –                  |

Bepipred score were calculated for each amino acid, and score for a peptide (selection threshold >0.5) were calculated as average of all amino acids. For MHC peptides prediction, the panEL prediction score was provided, and peptides of top 2% ranking score (corresponding to >0.095 in our analysis) were selected.

MHC-I: major histocompatibility complex class I; NCP: nucleocapsid phosphoprotein; PTM: post-translational modification; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.
The ongoing global efforts to generate vaccines or antibodies for SARS-CoV-2 primarily focus on protein S as the target antigen.\(^7,14,77\) Compared to surface protein S that mutates at a higher rate, NCP is more conserved across CoV species.\(^24\) Therefore, NCP is a more attractive target for the development of universal CoV vaccines. This idea is further prompted by recent studies showing NCP is capable of inducing long-term humoral immune responses in SARS-CoV-2 patients.\(^40-42\) As a viral core protein that is highly expressed in infected cells, NCP is likely to induce immune responses containing cytotoxic T lymphocytes,\(^78,79\) providing complementary protection in addition to neutralizing antibodies. This is underscored by the recent discovery showing SARS-CoV-2 patients developed stronger T cell responses against NCP than against protein S.\(^40\) Furthermore, current anti-CoV drug developments primarily target replicase proteins, therefore development of “cocktail” antiviral therapies can be achieved by targeting NCP.\(^30\) The comprehensive PTM profile, along with the refined structural models of SARS-CoV-2 NCP proposed in this study, can provide guidance to produce new vaccines or therapeutics to join the current vaccine portfolio. The overall workflow for comprehensive PTM survey can be found in Figure 1.

### Materials and methods

#### Expression and purification of SARS-CoV-2 NCP

HEK293T (Human embryonic kidney, ATCC\(^a\) CRL-3216\(^b\)) cells were grown in Dulbecco modified Eagle medium ( Gibco BRL, Grand Island, NY, USA) supplemented with 10% fetal bovine serum ( Gibco BRL). The full-length SARS-CoV-2 NCP gene (GenBank number QHD43419.1) was cloned into the pcDNA3.1 vector with a C-terminal 2 X Strep tag. The NCP gene (GenBank number QHD43419.1) was cloned into the pcDNA3.1 vector with a C-terminal 2 X Strep tag. The full-length SARS-CoV-2 NCP gene (GenBank number QHD43419.1) was cloned into the pcDNA3.1 vector with a C-terminal 2 X Strep tag. The full-length SARS-CoV-2 NCP gene (GenBank number QHD43419.1) was cloned into the pcDNA3.1 vector with a C-terminal 2 X Strep tag. NCP was expressed in HEK293T cells (ATCC\(^c\) CRL-3022\(^d\)). After 40–60 hours, the cells were lysed in 1 mL cold immunoprecipitation (IP) buffer (50mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA) supplemented with 0.5% Nonidet P 40 Substitute (NP-40, Solarbio) and 1× complete EDTA-free protease inhibitor cocktail (Roche, Switzerland, 1187358001) for 30 minutes at 4°C. The lysates were then centrifuged for 30 minutes at 15000×g. The remaining lysate was incubated with 50 μL of Strept-Tactin Sepharose beads (IBA Lifesciences, Germany) in 0.6 mL IP Buffer for 2 hours with constant rotation at 4°C. The beads were washed three times with 1 mL of IP buffer containing 0.05% NP-40 followed by one final wash in detergent-free IP buffer in a fresh detergent-free tube. Proteins were eluted by agitating beads in 40 μL IP buffer supplemented with 2.5 mM D-Desfodiobiotin (IBA Lifesciences) on a vortex mixer at room temperature for 30 minutes.

#### Sample preparation for PTM survey

NCP was digested into peptides according to a modified filter aided sample preparation protocol.\(^90\) Briefly, NCP was first reduced by 5 mM dithiothreitol in 0.5% Rapigest SF and 50 mM ammonium bicarbonate (ABC) in a 10 kD molecular weight cut-off spin column (Millipore, Germany) at 56°C for 15 minutes, and then alkylated by 15 mM iodoacetamide for 45 minutes in the dark. The sample was centrifuged at 12000×g for 10 minutes and washed twice by 50 mM ABC. Protein was resuspended in 50 mM ABC with 0.1% Rapigest SF and digested overnight at 37°C by sequence-grade trypsin, endoproteinase Lys-C, and chymotrypsin (all Promega, WI, USA) with a protein: protease ratio of 1:30. All peptides were desalted using C18 stage tips, dried by SpeedVac before direct LC-MSMS analyses or glycosylation characterization. The overall workflow for comprehensive PTM survey can be found in Figure 1.

#### Sample preparation for protein glycosylation characterization

Peptides derived from nucleocapsid protein by multiple protease digestion were used to enrich intact N-glycopeptides by house-made zwitterionic hydrophobic interaction liquid chromatography (ZIC-HILIC) stage-tip. Briefly, stage-tip packed with Exsil Pure ZIK HILIC 5 μm beads (Dr. Maisch GmbH, Germany, 6136918) was pre-conditioned with 0.1% trifluoro-acetic acid (TFA) in 80% acetonitrile (ACN). Peptides in 0.1% TFA in 80% ACN were loaded to the beads and unbound flow-through fraction was collected. The column was washed twice with 0.1% TFA in 80% ACN. Glycopeptides were eluted first by 0.1% TFA and then by 30 mM ABC. Other than direct analyses by LC-MSMS, glycopeptides were also dried by Speed Vac and deglycosylated by PNGase F (NEB, 1:100) overnight at 37°C in 50 mM ABC in pure H2O.\(^38\)

#### LC-MSMS experiments

Peptides (500 ng) were separated by C18 nano-column using an Ultimate 3000 nanoflow liquid chromatography system (Thermo Scientific, MA, USA) and analyzed by the Q-Exactive HFX mass spectrometry (Thermo Scientific) operated under the data-dependent mode. The MS spectra were recorded by Xcalibur software 2.3 (Thermo Scientific). All operation parameters for nanoLC and MS systems are detailed in the Supplementary Methods based on previously published methods, http://links.lww.com/IMD/A12.\(^64\)

#### Bioinformatics

All MS raw data files of directed peptides analyses were searched by MaxQuant (version 1.6.10.43) against the Uniprot SARS-CoV-2 sequences. Mass tolerance of 4.5 ppm was set for the search. Trypsin/chymotrypsin/LysC with up to two missed cleavages was set. Carbamidomethylation on Cys was set as fixed modification, while oxidation on Met and O\(^18\) deamidation on Asn were set as variable modification. Identification was further screened by N-glycosylation possible sequons (N-X-S/T, X\#P). To explore additional PTM forms, phosphorylation (S/T), acetylation (K), methylidyne (K/R/E), succinylation (K), crotonylation (K), di-Gly (K), farnesylation (K/Nterm), myristoylation (K/Nterm), palmitoylation or prenylation, and oxidation on proline were separately specified in individual searches. Peptide level 1% FDR was set to filter the results. Confident identification of PTM was based on localization probability of 95%.

To reveal N-glycosylation forms, MS raw files from ZIC-HILIC fractioned peptides were analyzed by pHYCO (version 2.2.2)\(^81\) using the following parameters: carboxamidomethylation on Cys was set as fixed modification, while oxidation on Met was set as variable modification; two missed trypsin/chymotrypsin/ LysC digestion cleavages; mass error of 5 ppm and 15 ppm for MS and MS2. The MS2 spectra were annotated by built-in pHYCO.gdb glycan structure database\(^81\) to identify glycan fragments.
Immunogenic epitopes prediction

Identified PTM positions were aligned with possible immunogenic epitopes on SARS-CoV-2 NCP predicted by bioinformatic tools. Briefly, predictions of B cell epitopes were performed by Bepipred (v2.0) available via IEDB database (http://tools.iedb.org/main/bcel). Sequences with average R score higher than 0.5 were considered as potential linear B cell epitopes. In addition, T cell epitopes likely presented by MHC-I were predicted by NetMHCpan (v4.1b). Sequences with top 2% ranking score were considered as potential T cell epitopes. Redundant or nested sequences were discarded.

Immunoblotting analysis

To confirm protein ubiquitination or ISGylation status, SARS-CoV-2 NCP expressed in 293T cells was immunoprecipitated by Strept-Tactin Sepharose beads (IBA Lifesciences, 2-1502-001), separated by 4%–15% gradient sodium dodecyl sulfate-polyacrylamide gel electrophoresis before transferred to PVDF membranes. Immunoblotting was performed by Strept-Tactin-HRP conjugate antibody (IBA Lifesciences, 2-1502-001, 1:1000), anti-Ubiquitin (CST 3933, 1:1000), anti-ISG15 (CST 2758, 1:1000), and anti-GAPDH (CST 5174, 1:1000) antibody and visualized by chemiluminescence detection kit.

Acknowledgments

The authors thank the proteomics and metabolomics platform in the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases at Zhejiang University for mass spectrometry analyses.

References

[1] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579(7798):265–269.
[2] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20(5):e534–e536.
[3] Zhou F, Yu T, Du R, et al. Clinical course and risk factors of mortality of 187 patients in Wuhan, China: a retrospective cohort study. Lancet 2020;395(10223):1054–1062.
[4] Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected Wuhan in China. JAMA 2020;2020;323(11):1061–1066.
[5] Zaat Z, van der Peet-Walberg AJ, van der Werf K, et al. Role of the virus entry receptors ACE2 and TMPRSS2 in SARS-CoV-2 entry. J Virol 2020;94(13):e02471–e02420.
[6] Hoffmann M, Kleine-Wedell K, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181(2):271–280.e8.
[7] Walls AC, Park YJ, Tortorelli MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020;181(2):281–292.e6.
[8] Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the SARS-CoV-2 spike protein. Science 2020;367(6483):1088–1093.
[9] Hurst KR, Ye R, Goebel SJ, Jayaprakash M, Masters PS. An interaction between the nucleoside protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. J Virol 2010;84(19):10276–10288.
[10] Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein forms and functions. Antiviral Res 2014;103:39–50.
[11] McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. Virus Res 2016;216(8):2991–3000.
[12] Dinnes DC, Chalupska D, Silhan J, et al. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. PLoS Pathog 2020;16(12):e1009100.
[13] Savastano A, Ibanez de Opakuza A, Rankovic M, Zweckstetter M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. Nat Commun 2020;11(1):6041.
[14] Hu Y, Li W, Gao T, et al. The severe acute respiratory syndrome coronavirus nucleocapsid inhibits type I interferon production by interfering with TRIM25-mediated RIG-I ubiquitination. J Virol 2017;19(8):e02143–e02216.
[15] Li JY, Liao CH, Wang Q, et al. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Res 2020;286:1907974.
[16] Chen K, Xiao F, Hu D, et al. SARS-CoV-2 nucleocapsid protein interacts with RIG-I and represses RIG-mediated IFN-beta production. Viruses 2020;13(1):47.
[17] Tsai TL, Lin CH, Lin CN, Lo CY, Wu HY. Interplay between the poly(A) tail, poly(A)-binding protein, and coronavirus nucleocapsid protein regulates gene expression of coronavirus and the host cell. J Virol 2018;92(23):e01162–e01218.
[18] Che XY, Qiu LW, Pan YX, et al. Sensitive and specific monoclonal antibody-based capture enzyme immunoassay for detection of nucleocapsid antigen in sera from patients with severe acute respiratory syndrome. J Clin Microbiol 2004;42(6):2629–2635.
[19] Liu L, Liu W, Zheng Y, et al. A preliminary study on serological assay for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 238 admitted hospital patients. Microbes Infect 2020;22(4):206–211.
[20] Guo L, Ren L, Yang S, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). Clin Infect Dis 2020;71(15):778–785.
[21] Basu BV, Brown OR. Comparative analysis of Coronavirusidaceae nucleocapsid and surface glycoprotein sequences. Front Bioeng Biotechnol 2020;5:1894–1900.
[22] McAndrews KM, Dowlatshahi DP, Dai J, et al. Heterogeneous antibodies against SARS-CoV-2 spike receptor binding domain and nucleocapsid with implications on COVID-19 immunity. JCI Insight 2020;5(18):e142386.
[23] Li T, Wang L, Wang H, et al. Serum SARS-CoV-2 nucleocapsid protein: a sensitivity and specificity early diagnostic marker for SARS-CoV-2 infection. Front Cell Infect Microbiol 2020;10:470.
[24] Cha WN, Tan CW, Foo R, et al. Serological differentiation between COVID-19 and SARS infections. Emerg Microbes Infect 2020;9 (1):1497–1505.
[25] Liu D, Wu F, Chen Y, et al. Comparative research on nucleocapsid and spike glycoprotein as the rapid immunodetection targets of COVID-19 and establishment of immunoassay strips. Mol Immunol 2021;131:e1–12.
[26] Datta NK, Mazumdar K, Gorai P. The nucleocapsid protein of SARS-CoV-2: a target for vaccine development. J Virol 2020;94(13):e00647–e00720.
[27] Kumar A, Kumar P, Saumya KU, Kapuganti SK, Bhardwaj T, Giri R. Exploring the SARS-CoV-2 structural proteins for multi-epitope vaccine development: an in-silico approach. Expert Rev Vaccines 2020;19(9):887–898.
[28] Ong E, Wong MU, Huffman A, He Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. Front Immunol 2020;11:1581.
[29] Kwarteng A, Assiedu E, Sakyi SA, Assiedu SO. Targeting the SARS-CoV-2 nucleocapsid protein for potential therapeutics using immuno-informatics and structure-based drug discovery techniques. Biomed Pharmacother 2020;132:e01194.
[30] Ray M, Sarkar S, Rath SN. Drugability for COVID-19: in silico discovery of potential drug compounds against nucleocapsid (N) protein of SARS-CoV-2. Genomics Inform 2020;18(4):e43.
[39] dos Santos WG. Impact of virus genetic variability and host immunity for the success of SARS-CoV-2 vaccines. Biomed Pharmacother 2021;136:112722.

[40] Reynolds CJ, Swadling L, Gibbons JM, et al. Discordant neutralizing antibody and T cell responses in asymptomatic and mild SARS-CoV-2 infection. *Sci Immunol* 2020;5(49):eaaz3698.

[41] Hartley GE, Edwards ESJ, Au P, et al. Rapid generation of durable B cell memory to SARS-CoV-2 spike and nucleocapsid proteins in COVID-19 and convalescence. *Sci Immunol* 2020;5(49):eabz8891.

[42] Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* 2021;371(6529):eabc4640.

[43] Wang J, Ji J, Ye J, et al. The structure analysis and antigenicity study of the N protein of SARS-CoV. *Genomics Proteomics Bioinformatics* 2003;1(2):145–154.

[44] Peng TY, Lee KR, Tarn WY. Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J* 2008;275(16):4152–4163.

[45] Carlson CR, Asfaha JB, Ghent CM, et al. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol Cell* 2020;80(6):1102–1103.e4.

[46] Carlson CR, Asfaha JB, Ghent CM, et al. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol Cell* 2020;80(6):1102–1103.e4.

[47] Bouhaddou M, Memon D, Meyer B, et al. The global phosphorylation landscape of SARS-CoV-2 infection. *Cell* 2020;182(3):635–712.e19.

[48] Davidson AD, Williamson MK, Lewis S, et al. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Med* 2020;12(11):68.

[49] Wang C, Xu H, Lin S, et al. GPS 5.0: an update on the prediction of antigenic variation between 2019 novel coronavirus glycoprotein. *J Virol* 2020;94(16):e02360–2019.

[50] Kopecky-Bromberg SA, Martinez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *Old Pathol* 2010;84(17):557–566.

[51] Shajahan A, Supekar NT, Gleinich AS, Azadi P. Deducing the N- and O-glycosylation pro-

[52] landscape of SARS-CoV-2 infection. *Virusdisease* 2020;11(3):101–117.

[53] How to cite this article: Sun Z, Zeng X, Ji F, Zhou M, Su X, Ren K, Li L. Mass spectrometry analysis of newly emerging coronavirus HCoV-19 spike protein and human ACE2 reveals camouflage glycan and unique post-translational modifications. *Infect Microbiol* 2021;13(3):149–157. doi: 10.1097/IMM.000000000000073