

Supporting Information

Assessing Antigen Structural Integrity through Glycosylation Analysis of the SARS-CoV-2 Viral Spike

Juliane Brun^{1†}, Snežana Vasiljevic^{1†}, Bevin Gangadharan^{1‡}, Mario Hensen^{1‡}, Anu V. Chandran¹, Michelle L. Hill¹, J.L. Kiappes¹, Raymond A. Dwek¹, Dominic S. Alonzi¹, Weston B. Struwe^{2*}, Nicole Zitzmann^{1*}

¹Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford, United Kingdom

²Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, United Kingdom

†equal contribution

‡equal contribution

*Correspondence to: weston.struwe@chem.ox.ac.uk, nicole.zitzmann@bioch.ox.ac.uk

Table of Contents

1. Materials and Methods	4
1.1 Cell culture	4
1.2 Virus propagation	4
1.3 Expression and purification of the CR3022 antibody	4
1.4 SARS-CoV-2 S expression constructs	5
1.4.1 Non-stabilised spike (<i>S_{vaccine-antigen}</i>)	5
1.4.2 Recombinant, soluble spike glycoprotein (<i>S_{recombinant-trimer}</i>)	5
1.4.3 Recombinant S1 (<i>S1_{recombinant}</i>)	5
1.5 Recombinant SARS-CoV-2 S glycoprotein expression	5
1.6 Affinity purification of <i>S_{virus}</i> , <i>S_{recombinant-trimer}</i> , <i>S1_{recombinant}</i> and <i>S_{vaccine-antigen}</i>	5
1.7 Flow cytometry	6
1.8 SDS-PAGE and Western blot	6
1.9 In-gel protease digestion	6
1.10 In-solution digestion	7
1.11 Liquid chromatography	7
1.12 Mass spectrometry	7
1.13 Protein identification and sequence coverage	7
1.14 Mass photometry	8
1.15 Fluorescent labelling of N-linked glycans	8
1.16 HILIC-UHPLC	8
1.17 Exoglycosidase sequencing of N-linked glycans	9
2. Supplementary Figures	10
Figure S1. Mascot sequence coverage for <i>S_{virus}</i> and <i>S_{vaccine-antigen}</i>	10
Figure S2. Sequence alignment of S proteins	11
Figure S3. Site-specific glycosylation of <i>S_{recombinant-trimer}</i>	12
Figure S4. <i>S_{vaccine-antigen}</i> Western blot	13
Figure S5. MS/MS spectrum of the N74 glycopeptide from <i>S1_{vaccine-antigen}</i>	14
Figure S6. MS spectrum of N74 containing glycopeptide from <i>S1_{virus}</i>	15
Figure S7. MS/MS spectrum of the T678 glycopeptide from <i>S1_{vaccine-antigen}</i>	16
Figure S8. Purification and glycan analysis of <i>S1_{recombinant-monomer}</i>	17
Figure S9. Illustration of site-specific glycosylation changes on S1 glycoproteins	18
Figure S10. Flow cytometry analysis	19
Movie S1. Mass photometry of <i>S1_{vaccine-antigen}</i>	20
Movie S2. Mass photometry of <i>S_{recombinant-trimer}</i>	20
4. Supplementary Tables	21
Table S1. Mascot analysis	21

<i>Table S2. Raw data from FACS analysis.....</i>	<i>22</i>
---	-----------

1. Materials and Methods

Samples and buffers were filtered through 0.1 µm centrifugal filters (Ultrafree –MC- VV, MerckMillipore) prior to use. Initial stock concentrations of all proteins were measured using a DeNovix DS-11+ spectrophotometer. Protein solutions were diluted in PBS (Gibco DPBS (1X)) to obtain concentrations between 0.2 to 1 absorbance at 280 nm (low µM range). Pipetting precision was checked with a microbalance (Mettler Toledo, AT261).

1.1 Cell culture

African green monkey (*Chlorocebus sabaeus*) Vero E6 cells (ATCC CRL-1586) were grown in Dulbecco's Modified Eagle Medium (DMEM, Gibco) containing 4.5 g/L glucose and 2 mM L-glutamine supplemented with 10% heat inactivated fetal bovine serum (HI FBS, Gibco), 100 U/mL penicillin and 0.1 mg/mL streptomycin (Sigma). The human lung cancer cell line Calu-3 (kindly provided by Dr Anderson Ryan, Department of Oncology, University of Oxford, as part of an ongoing collaboration) was cultured in a 1:1 mixture of DMEM and Ham's F-12 medium (Gibco) supplemented with 10% HI FBS, 1% non-essential amino acids (Gibco), 1% sodium pyruvate (Gibco), 100 U/mL penicillin and 0.1 mg/mL streptomycin (Sigma). Incubations were carried out at 37 °C and 5% CO₂. The FreeStyle™ 293-F (HEK 293F, Thermo Fisher Scientific) cell line was cultured in Freestyle 293 expression media (ThermoFisher Scientific) and incubated at 37 °C, 8% CO₂ and 130 rpm.

1.2 Virus propagation

The SARS-CoV-2 England/02/2020 strain (GISAID: EPI_ISL_407073) was provided at passage one from Public Health England, Collindale. Viral stocks were obtained by infecting Vero E6 cells at a multiplicity of infection (MOI) of 0.01 in virus propagation medium (DMEM with 2% HI FBS) and harvested 72 h post infection (hpi). Culture media was centrifuged at 3200 x *g* for 5 min, aliquoted and stored at –80 °C. To obtain SARS-CoV-2 virions for endogenous spike glycoprotein purification, Calu-3 cells were infected with SARS-CoV-2 in Calu-3 media containing 2% HI FBS at a MOI of 0.1. Cell culture supernatant was harvested 72 hpi and centrifuged at 3200 x *g* for 5 min. Virus-containing supernatant was concentrated one-log - using a 100 kDa cut-off centrifugal filter (Amicon, Merck) - with subsequent inactivation of virions with a final concentration of 0.5% Triton X-100 (Sigma) by incubation for 30 min at 4 °C.

1.3 Expression and purification of the CR3022 antibody

The recombinant monoclonal antibody CR3022 was transiently expressed in HEK293F cells. The cells were co-transfected with plasmids encoding the heavy- and light chain of CR3022 (a kind gift from the Krammer Laboratory, Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York) cloned into pFUSEss and pFUSEss2, respectively (1:2 molar transfection ratio) utilising FreeStyle™ MAX reagent (Invitrogen) and OptiPRO™ SFM (Gibco) following the manufacturer's protocol. Seven days post transfection, cell culture supernatant was harvested by centrifugation at 3000 x *g* for 45 min. Sodium phosphate was added to a final concentration of 20 mM, the pH adjusted to 7.2 and the sample sterile filtered. The solution was applied to a Protein A-Sepharose Fast Flow column (GE Healthcare) equilibrated in 20 mM sodium phosphate, pH 7.2. Antibody elution took place using 100 mM citric acid (pH 3.5) and the eluate pH adjusted to 7.2 using 1M Tris-NaOH. The sample was further purified by size exclusion chromatography using a Superdex™ 200 pg 16/600 column (GE Healthcare) equilibrated in 10 mM phosphate buffer, 2.7 mM potassium chloride 137 mM sodium chloride, pH 7.4 (PBS).

1.4 SARS-CoV-2 S expression constructs

1.4.1 Non-stabilised spike ($S_{\text{vaccine-antigen}}$). The sequence encoding SARS-CoV-2 (Addgene # 141382) was cloned from amino acids 2-1273 into a pENTR4-LPTOS vector (kindly provided by Dr Simon Draper, Jenner Institute, University of Oxford) in frame to an N-terminal secretion leader peptide tissue plasminogen activator (tPA), using InFusion cloning (Clontech). The plasmid also encodes a modified human cytomegalovirus major immediate early enhancer/promoter (IE CMV).

1.4.2 Recombinant, soluble spike glycoprotein ($S_{\text{recombinant-trimer}}$). The vector pCAGGS encoding for the SARS-CoV-2 spike glycoprotein was a kind gift from the Krammer Laboratory, Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York. In short, the SARS-CoV-2 spike glycoprotein (GenBank isolate: MN908947) residues 1-1213 were codon-optimised for mammalian cell expression; the polybasic cleavage site (RRAR, residues 682-685) were mutated to alanine and the prefusion conformation stabilizing mutations K986P and V987P (numbering according to wild type) were introduced. In addition, the construct has a C-terminal thrombin cleavage site, T4 fibrin trimerization domain and a hexahistidine tag.

1.4.3 Recombinant S1 ($S1_{\text{recombinant}}$). The nucleotides encoding for residues 1-682 of the SARS-CoV-2 spike glycoprotein (GenBank isolate: MN908947) were optimised for mammalian cell expression and subcloned into the mammalian expression vector pHLsec (AgeI/-KpnI linearised) using Gibson assembly reaction mix (NEB) following the manufacturer's protocol. All recombinant construct products were used to heat-shock transform *E. coli* DH5 α competent cells (Clontech). Cells were incubated at 37 °C for 16 hours under antibiotic selection (25 μ g/mL kanamycin or carbenicillin for $S_{\text{vaccine-antigen}}$, $S_{\text{recombinant-trimer}}$ or $S1_{\text{recombinant-monomer}}$, respectively) and single colonies were subsequently picked for DNA sequencing (Source Biosciences).

1.5 Recombinant SARS-CoV-2 S glycoprotein expression

Proteins were transiently expressed in HEK293F cells. The cells were transfected utilising FreeStyle™ MAX reagent (Invitrogen) and OptiPRO™ SFM (Gibco) following the manufacturer's protocol. Five days post transfection, cell culture supernatant was harvested by centrifugation at 3000 x *g* for 45 min. The sample was applied to an immunoaffinity chromatography column with immobilised CR3022 antibody as an antagonist (see below).

1.6 Affinity purification of S_{virus} , $S_{\text{recombinant-trimer}}$, $S1_{\text{recombinant}}$ and $S_{\text{vaccine-antigen}}$

CR3022 antibodies were immobilised on cyanogen bromide activated Sepharose 4 resin (GE Healthcare) according to the manufacturer's protocol (Instructions 71-7086-00 AF) for subsequent use in immunoaffinity purification. The culture supernatant of SARS-CoV-2 infected Calu-3 cells (in total: 56x10⁶ plaque forming units) was collected and the virus inactivated as previously described. Triton-X100 (0.5% w/v), carried over from the inactivation step, was diluted below its CMC (~ 0.02% w/v) using 20 mM Tris-HCl, 500 mM NaCl and a final concentration of 0.03% (w/v) *n*-dodecyl- β -D-maltoside (DDM, Sigma) added. Subsequently, the solution was loaded onto a CR3022-affinity column, washed with 10 column volumes (CV) of high-salt wash buffer (20 mM Tris-HCl, 500 mM NaCl, 0.03% DDM) and eluted with 3 M MgCl₂, 0.03% DDM, pH 7.2 (2 CV). The eluted protein was immediately buffer exchanged into 10 mM Tris-HCl, 75 mM NaCl, 0.03% DDM, pH 7.2 and further concentrated using a 30 kDa cut-off centrifugal filter (GE Healthcare). Recombinant spike glycoproteins ($S_{\text{vaccine-antigen}}$, $S_{\text{recombinant-trimer}}$ and $S1_{\text{recombinant}}$) were purified in the same manner by omitting the detergents (Triton-X100 and DDM) throughout the purification process.

1.7 Flow cytometry

HEK293F cells transfected with constructs for $S_{\text{recombinant-trimer}}$ or $S_{\text{vaccine-antigen}}$ were harvested 48 hours post transfection by centrifugation at $400 \times g$ for 5 min, and stained using LIVE/DEAD fixable near-IR dead cell stain kit (Life Technologies) prior to fixation in 2% paraformaldehyde (Electron Microscopy Sciences) for 20 min to exclude dead cells. Cells were blocked in fluorescence activated cell sorting (FACS) block (PBS with 0.5% w/v BSA, 5 mM EDTA, 5% HI goat serum (Thermo Fisher Scientific) and 5% HI normal human plasma (Thermo Fisher Scientific) for 20 min, followed by permeabilization with 0.5% saponin (Sigma) in FACS block for 20 min on ice. Cells were first surface stained with an antibody recognizing S1 (CR3022; 20 $\mu\text{g/mL}$) or with an antibody recognizing S2 (Genetex #GTX632604; 10 $\mu\text{g/mL}$) for 40 min on ice. Primary antibodies were detected using 10 $\mu\text{g/mL}$ of the appropriate Alexafluor 488-conjugated antibody (Invitrogen #A-11013 or #A-11008, respectively). Flow cytometry acquisition was performed using an Attune NxT Flow Cytometer (Thermo Fisher) and gated for singlet live cells. A minimum of 10,000 gated events were identified and results were analysed using the software tool FlowJo V10.

1.8 SDS-PAGE and Western blot

Recombinant and endogenous proteins were analysed using 4-12% NuPAGE Bis-Tris SDS-PAGE gel (Invitrogen) to determine the protein expression in either the supernatant or the cell pellet. To determine the expression of $S_{\text{vaccine-antigen}}$ in HEK293F cells, cells were harvested 4 days post transfection as described above. The resulting cell pellet was resuspended in lysis buffer (1% DDM in PBS); the suspension was sonicated on ice three times for 5 min using pulses of 0.7s on and 0.3s off with an applied 65% intensity. After protein separation using SDS-PAGE, gels were either stained with SimplyBlue SafeStain (Invitrogen) for 1 hour and de-stained in distilled water overnight, or used for Western blot analysis (iBlot™, Thermo Fisher Scientific). Blots were incubated with the primary antibodies listed in the Flow Cytometry section and additionally against S2 (Genetex #GTX635693), and respective HRP secondary antibodies (GE Healthcare #NA933, Promega #W4011, Sigma #71045).

1.9 In-gel protease digestion

The SDS-PAGE bands corresponding to spike glycoprotein were excised from the gel and cut into pieces. The gel pieces were washed twice with water, twice with 50%v/v acetonitrile followed by 100%v/v acetonitrile. The spike glycoprotein sequence was separately digested *in silico* with three sequencing grade proteases (trypsin, chymotrypsin and alpha lytic protease) to confirm peptides of suitable length (between 7 to 25 amino acids) covering the 22 potential N-glycosylation sites. Gel pieces were rehydrated with 100 mM ammonium bicarbonate, 100 mM Tris-HCl 10 mM calcium chloride pH 8.0 or 100 mM ammonium bicarbonate for digests with trypsin, chymotrypsin and alpha lytic protease, respectively. Gels were dehydrated with 100 %v/v acetonitrile and dried down in a centrifugal vacuum concentrator before rehydration for 45 min at 56°C with 10 mM dithiothreitol in 100 mM ammonium bicarbonate, 100 mM Tris-HCl 10 mM calcium chloride pH 8.0 or 100 mM ammonium bicarbonate for digests with trypsin, chymotrypsin and alpha lytic protease, respectively. Proteins were alkylated with 55 mM iodoacetamide in the respective buffer for 30 min in the dark. The gel pieces were washed twice with the respective buffer followed by the buffers in 50% v/v acetonitrile before drying in a centrifugal vacuum concentrator. Gel pieces were rehydrated on ice for 45 min using 12.5 ng/ μL of trypsin, chymotrypsin (Promega) and alpha lytic protease (Sigma) in 100 mM ammonium bicarbonate, 100 mM Tris-HCl 10 mM calcium chloride pH 8.0 or 100 mM ammonium bicarbonate, respectively. Proteins were digested at 37 °C for 16 hours. Gels were washed in the respective buffer in 50% v/v acetonitrile before extracting the peptides

twice with 5% v/v formic acid and acetonitrile. Peptide extracts were dried, resuspended in 0.05% v/v TFA using a water bath sonicator and then run by LC-MS.

1.10 In-solution digestion

Endogenous spike (8 µg) purified by CR3022 was dried down and denatured with 8 M urea containing 200 mM ammonium bicarbonate. Proteins were reduced with tris(2-carboxyethyl)phosphine (1 mM final concentration) in 100 mM ammonium bicarbonate for 1 hour at 56 °C followed by alkylation with iodoacetamide (4.4 mM final concentration) in 100 mM ammonium bicarbonate for 1 hour in the dark. Urea was diluted to 1 M with 100 mM ammonium bicarbonate and then the proteins were digested with 1:100 w/w alpha lytic protease:protein for 16 hours at 37 °C. An additional 1:100 w/w alpha lytic protease:protein was added and the digestion was continued for 5 h. The peptide containing sample was dried, resuspended in 0.05% v/v TFA using a water bath sonicator and then run by LC-MS with online desalting.

1.11 Liquid chromatography

Peptides were separated on a Dionex Ultimate 3000 nano UHPLC system (Thermo Scientific). A nano analytical C18 reversed phase column (PepMap) was used with dimensions 75 µm x 15 cm, 2 µm particle size (Thermo Scientific) with a flow rate of 300 nL/min at 35 °C. The mobile phase used was solvent A: 0.1% v/v formic acid in LC-MS grade water, and solvent B: 0.1% v/v formic acid in 80% v/v acetonitrile. The gradient used to separate peptides on the analytical column was: 2% B (0-4.5 min), 2-30% B (4.5-40 min), 30-95% B (40-55 min), 95% B (55-59 min), 95-2% B (59-60 min) and 2% B (60-70 min) for column equilibration.

1.12 Mass spectrometry

Peptides from the nano LC were analysed on a benchtop Q Exactive hybrid quadrupole-Orbitrap mass spectrometer using the Nanospray Flex ion source. Prior to data acquisition, the mass spectrometer was calibrated for mass accuracy according to the manufacturer's recommendations using a positive ion calibration solution injected at 5 µl/min into a heated electrospray ionisation (HESI) probe (Thermo Scientific). The conditions for data dependent acquisition (DDA) were as follows: chromatographic peak width was set at 12 s and the Full MS conditions were with a resolution of 70,000, AGC target of 3e6, maximum IT (injection time) of 60 ms, scan range of 375 to 1500 m/z. The dd-MS2 conditions were with a resolution of 17,500, AGC target of 1e5, maximum IT of 60 ms, loop count of 10 (i.e. Top 10), isolation window of 2.0 m/z, fixed first mass of 100.0 m/z and normalised collision energy (NCE) was stepped at 27, 30 and 33 in a high-energy collision dissociation (HCD) cell. The data-dependent (dd) settings were with minimum intensity threshold of 3.3e4 ions, charge exclusion: unassigned, 1, >8, peptide match: preferred, dynamic exclusion: 20 s.

1.13 Protein identification and sequence coverage

The acquired *.raw files were converted to Mascot generic files (*.mgf) using MSConvertGUI 64-bit (ProteoWizard version: 3.0.19172-57d620127). A peak-picking filter was set with MS levels 1-2. The Mascot server software (Matrix Science, London) was used to search the *.mgf files against the SwissProt database (containing SPIKE_SARS2 spike glycoprotein P0DTC2 in Version 2020_03 with 562,755 sequences released on 17th June 2020). The search parameters in Mascot were set as follows: MS/MS Ion Search with either trypsin, chymotrypsin or 'none' for alpha lytic protease (with up to 3 missed cleavages). Carbamidomethyl (C) was set as a fixed modification and Oxidation (M) as a variable modification. Taxonomy was set to 'All'. Peptide mass tolerance and fragment mass tolerance were both set as ± 10 ppm and the peptide charge was set as +2, +3 and

+4. Data format and instrument were set as Mascot generic and Q Exactive, respectively. A decoy database was searched in each case with the false discovery rate (FDR) adjusted to 1%. Sequence coverage was determined using only peptides with the statistical strongest assignment in bold red (i.e. highest-scoring match for a given MS/MS spectra and highest scoring protein in which the peptide sequence appears).

Glycopeptide data was analysed via Byonic™ v3.10.10 and Byologic™ v3.10-52x64 (Protein Metrics Inc.). Byonic search settings were as follows: Precursor mass tolerance = 5, Fragment mass tolerance = 20, FDR = 1% with a glycan library consisting of 132 N-glycan and 9 O-glycan structures. Additional permissible modifications were oxidation (Met/Trp), deamidation (Asn/Gln), N-terminal Gln to PyroGln and Cys carbamidomethyl (fixed modification). Glycopeptides search criteria was set for a minimum score ≥ 200 and <0.001 Pep2D value and data was inspected manually in Byonic to ensure quality of MS/MS spectra and precursor monoisotopic peak assignments. Byologic was used to obtain the area of each extracted ion chromatogram, which were exported and used to quantify site-specific glycosylation at each site.

1.14 Mass photometry

Mass photometry measurements were done as described in detail ³⁸. Briefly, borosilicate microscope coverslips were cleaned via sonication in Milli-Q H₂O, isopropanol, Milli-Q H₂O and dried under a nitrogen gas flow. Sample chambers were assembled using silicone gaskets (CultureWell™ reusable gasket, 3mm diameter x 1 mm depth, Grace Bio-Labs). Coverslips were placed on the sample stage and a single gasket was filled with 10 μ l phosphate buffered saline (PBS) to find focus and ensure a low signal-to-noise background. Immediately prior to each sample measurement, S1_{vaccine-antigen} and S_{recombinant-trimer} were diluted from a 1 μ M stock (object concentration) to 100 nM, of which 2 μ l was added to the 10 μ l PBS in a single unused gasket, and data were recorded for 90 seconds. MP settings were as follows: frame averaging = 50, frame rate = 955 Hz and 4x4 pixel binning. Data was analysed using Discovery MP v2.3dev7 (Refeyn Ltd).

1.15 Fluorescent labelling of N-linked glycans

Glycans of the peptides from the in-gel trypsin digests were enzymatically released by PNGase F (NEB) for 16 hours at 37 °C. Released glycans were labelled with 2-aminoanthranilic acid (2-AA) as previously described ³⁹. Briefly, glycans were resuspended in 30 μ l of water followed by addition of 80 μ l of labelling mixture (30 mg/ml 2-AA and 45 mg/ml sodium cyanoborohydride in a solution of sodium acetate trihydrate [4% w/v] and boric acid [2% w/v] in methanol). Samples were then incubated at 80 °C for 1 hour. Excess label was removed using Spe-ed Amide-2 cartridges (Applied Separation), as previously described ³⁹.

1.16 HILIC-UHPLC

Fluorescently labelled glycans were resolved by hydrophilic interaction liquid chromatography-ultra performance liquid chromatography (HILIC-UHPLC) using a 2.1 mm \times 10 mm Acquity BEH Amide Column (1.7 μ m particle size) (Waters, Elstree, UK). The mobile phase used was solvent A: 50 mM ammonium formate, pH 4.4, and solvent B: acetonitrile. The gradient used to separate fluorescently labelled glycans on the analytical column was: time = 0 min ($t = 0$): 22.0% A, 78.0% B (flow rate of 0.5 ml/min); $t = 38.5$: 44.1% A, 55.9% B (0.5 ml/min); $t = 39.5$: 100% A, 0% B (0.25 ml/min); $t = 44.5$: 100% A, 0% B (0.25 ml/min); $t = 46.5$: 22.0% A, 78.0% B (0.5 ml/min), $t = 48$: 22.0% A, 78.0% B (0.5 ml/min). Fluorescence was measured using an excitation wavelength of 360 nm and a detection wavelength of 425 nm.

1.17 Exoglycosidase sequencing of N-linked glycans

Endoglycosidase H (Endo H, NEB) was used for quantitation of oligomannose structures. Digestions were performed at 37 °C for 16 hours, according to manufacturer's instructions. The digested glycans were purified using a PVDF protein-binding membrane plate (Millipore) prior to HILIC-UHPLC analysis. Data processing was performed using Empower 3 software. The percentage abundance of oligomannose-type glycans was calculated by integration of the relevant peak areas before and after Endo H digestion, following normalisation.

2. Supplementary Figures

A		B	
1	MFVFLVLLPL VSSQCVNLT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS	1	MFVFLVLLPL VSSQCVNLT RTQLPPAYTN SFTRGVYYPD KVFRSSVLHS
51	TQDLFLFFFS NVTWFHAIHV SGTNGTKRFD NPVLFPNDGV YFASTKSNL	51	TQDLFLFFFS NVTWFHAIHV SGTNGTKRFD NPVLFPNDGV YFASTKSNL
101	IRGWIFGTTL DSKTQSLIV NNATNVVIVK CEFQFCNDPF LGVYYHKNK	101	IRGWIFGTTL DSKTQSLIV NNATNVVIVK CEFQFCNDPF LGVYYHKNK
151	SWMESEFRVY SSANNCTFEY VSQPFIMDLK GKQGNFKNLR EFVFKNIDGY	151	SWMESEFRVY SSANNCTFEY VSQPFIMDLK GKQGNFKNLR EFVFKNIDGY
201	FKIYSKHTPI NLVRDLPGQF SALEPLVDLP IGINITRFQT LLALHRSYLT	201	FKIYSKHTPI NLVRDLPGQF SALEPLVDLP IGINITRFQT LLALHRSYLT
251	PGDSSSGWTA GAAAYVGYL QPRTFLLYKN ENGTITDAVD CALDPLSETK	251	PGDSSSGWTA GAAAYVGYL QPRTFLLYKN ENGTITDAVD CALDPLSETK
301	CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPEGE VFNATRFASV	301	CTLKSFTVEK GIYQTSNFRV QPTESIVRFP NITNLCPEGE VFNATRFASV
351	YAWNRRKISN CVADYSVLIN SASFSTFKCY GVSPTKLNLI CFTNVYADSF	351	YAWNRRKISN CVADYSVLIN SASFSTFKCY GVSPTKLNLI CFTNVYADSF
401	VIRGDEVRFQI APGQTKGIAD YNYKLPDDFT GCVIAWNSNN LDKVGGNYN	401	VIRGDEVRFQI APGQTKGIAD YNYKLPDDFT GCVIAWNSNN LDKVGGNYN
451	YLRLFRKSN LKPFERDIST EIIYQAGSTPC NGVEGFNCYF PLQSYGFQPT	451	YLRLFRKSN LKPFERDIST EIIYQAGSTPC NGVEGFNCYF PLQSYGFQPT
501	NGVGYPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN FNFNGLTGTG	501	NGVGYPYRV VVLSFELLHA PATVCGPKKS TNLVKNKCVN FNFNGLTGTG
551	VLTSNKKFL PFQPFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSUITP	551	VLTSNKKFL PFQPFGRDIA DTTDAVRDPQ TLEILDITPC SFGGVSUITP
601	GINTSNQAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL	601	GINTSNQAV LYQDVNCTEV PVAIHADQLT PTWRVYSTGS NVFQTRAGCL
651	IGAETHVNSY ECDIPIGAGI CASYQTQTN PRARSVASQ SIAYTMSLG	651	IGAETHVNSY ECDIPIGAGI CASYQTQTN PRARSVASQ SIAYTMSLG
701	AENSVAYSN SIAIPTNFTI SVTTEILPVS MTKTSVDCM YICGDSSTCS	701	AENSVAYSN SIAIPTNFTI SVTTEILPVS MTKTSVDCM YICGDSSTCS
751	NLLQYGSFC TQINRALTGI AVEQDKNTQE VFAQVKQIYK TPIKDFGGF	751	NLLQYGSFC TQINRALTGI AVEQDKNTQE VFAQVKQIYK TPIKDFGGF
801	NFSQILPDP KPSKRSPFIED LLFNKVTIAD AGFIKQYGDG LGDIAARDLI	801	NFSQILPDP KPSKRSPFIED LLFNKVTIAD AGFIKQYGDG LGDIAARDLI
851	CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM	851	CAQKFNGLTV LPPLLTDEMI AQYTSALLAG TITSGWTFGA GAALQIPFAM
901	QMAYRFNGIG VTQNVLYENQ KLIANQFNISA IGKIQDSLSS TASALGKLQD	901	QMAYRFNGIG VTQNVLYENQ KLIANQFNISA IGKIQDSLSS TASALGKLQD
951	VVNQNAQALN TLVKQLSSNF GAISSVLNDI LSRLDKVEAE VQIDRLITGR	951	VVNQNAQALN TLVKQLSSNF GAISSVLNDI LSRLDKVEAE VQIDRLITGR
1001	LQSLQTYVTQ QLIRAAEIRA SANLAATKMS ECVLGQSKRV DFCGKGHYLM	1001	LQSLQTYVTQ QLIRAAEIRA SANLAATKMS ECVLGQSKRV DFCGKGHYLM
1051	SFPQSAPHGV VFLHVTYVPA QEKWFTTAPA ICHDGKAHFP REGVVFVNGT	1051	SFPQSAPHGV VFLHVTYVPA QEKWFTTAPA ICHDGKAHFP REGVVFVNGT
1101	HWFTVQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP LQPELDSFKE	1101	HWFTVQRNFY EPQIITTDNT FVSGNCDVVI GIVNNTVYDP LQPELDSFKE
1151	ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL	1151	ELDKYFKNHT SPDVDLGDIS GINASVVNIQ KEIDRLNEVA KNLNESLIDL
1201	QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCMTSC CSCLKGCCSC	1201	QELGKYEQYI KWPWYIWLGF IAGLIAIVMV TIMLCMTSC CSCLKGCCSC
1251	GSCKKFDDED SEPVLKGVKL HYT	1251	GSCKKFDDED SEPVLKGVKL HYT

Figure S1. Mascot sequence coverage for S_{virus} and S_{vaccine-antigen}. The overall sequence coverage for (A) S_{virus} and (B) S_{vaccine-antigen} is shown for the three combined proteases (trypsin, chymotrypsin and alpha lytic protease) used for bottom-up MS. Over the sequence of S1, peptides covering 588 amino acids for S_{virus} (88% coverage) and 586 amino acids for S_{vaccine-antigen} (87% coverage) were observed. Red bold: Peptides identified in Mascot. Blue bold: Potential N-glycosylation sites. Underlined: S1. Italics: S2. Not italics/not underlined: Signal peptide.

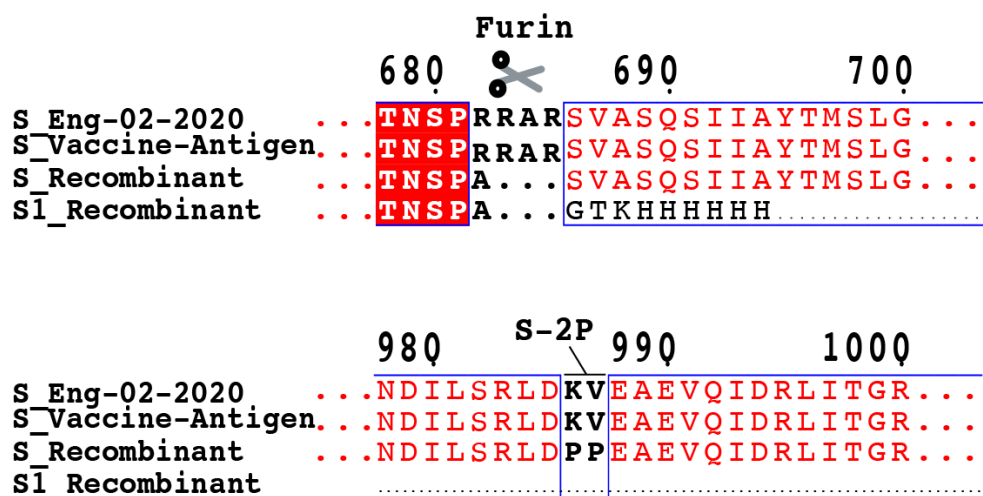


Figure S2. Sequence alignment of S proteins. Alignment of amino acid residues 678-700 and 978-1000 covering the furin cleavage site (top) and PP stabilisation site (bottom) of the constructs used in this study. EsPript was used for sequence alignment (ESPript - <http://esprpt.ibcp.fr>).

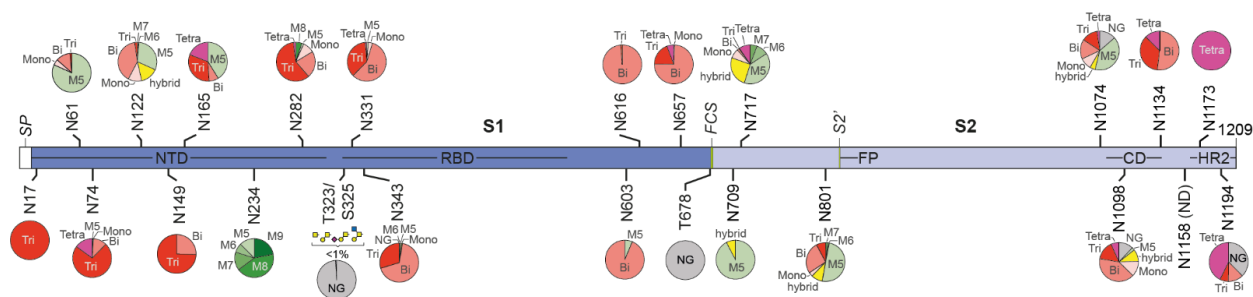


Figure S3. Site-specific glycosylation of S recombinant trimer.

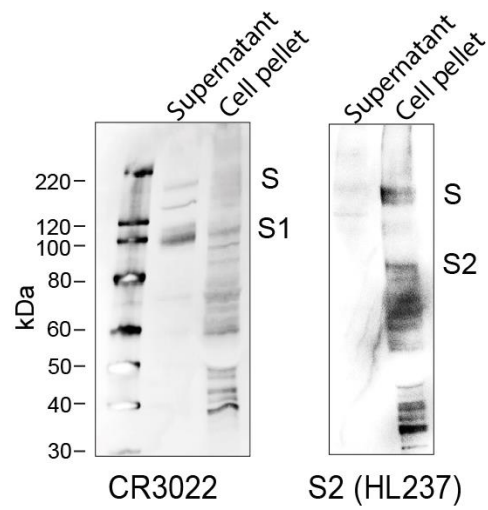


Figure S4. $S_{\text{vaccine-antigen}}$ Western blot. Western blot of unpurified $S_{\text{vaccine-antigen}}$ probed with anti-S1 (CR3022) and anti-S2 (HL237) antibody.

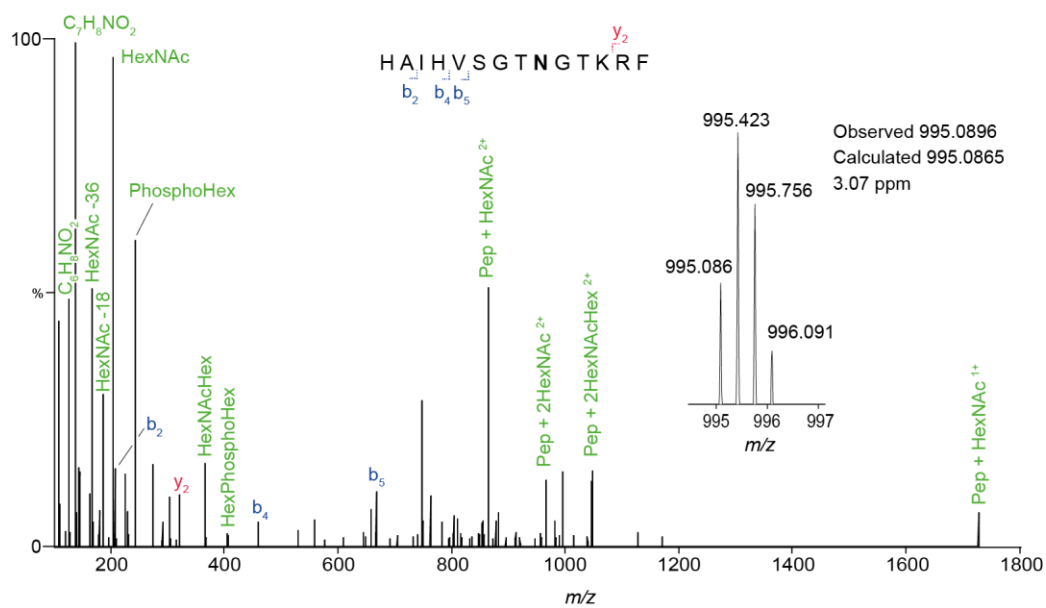


Figure S5. MS/MS spectrum of the N74 glycopeptide from S1_{vaccine-antigen}.

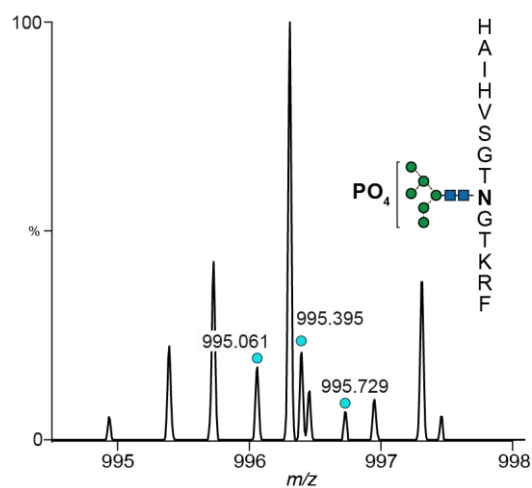


Figure S6. MS spectrum of N74 containing glycopeptide from S1_{virus}. MS spectrum showing the presence of the mannose-6-phosphate glycan at N74 on S1_{virus} (N-glycopeptide from chymotrypsin digestion). The isotopic distribution is labelled (blue circles).

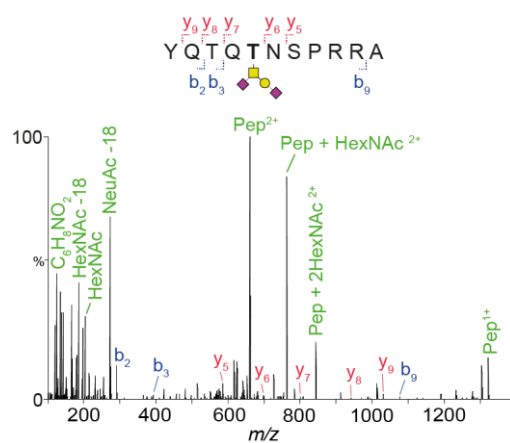


Figure S7. MS/MS spectrum of the T678 glycopeptide from S1_{vaccine-antigen}.

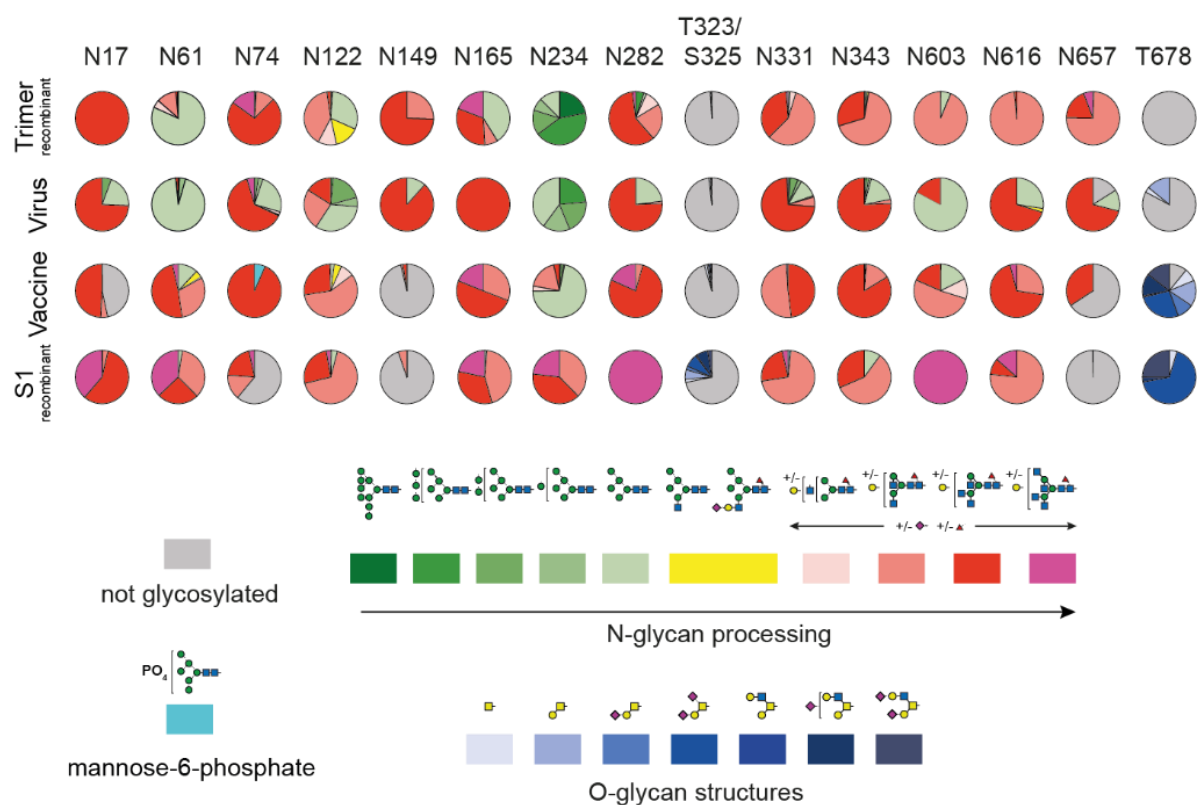


Figure S9. Illustration of site-specific glycosylation changes on S1 glycoproteins. Cumulative site-specific glycosylation changes across S1 subunits from recombinant-trimer, virus, vaccine-antigen and recombinant-monomer samples.

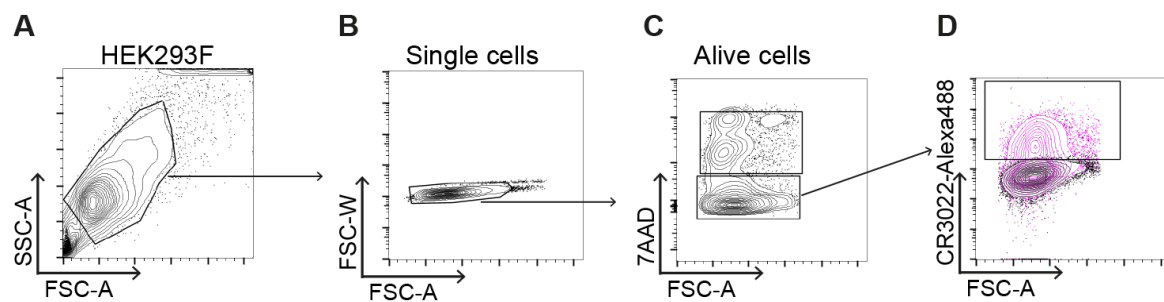


Figure S10. Flow cytometry analysis. Representative flow cytometry plots showing the gating strategy used to analyse the expression of S1 and S2. (A) HEK293F gate, excluding doublets shown in (B) and final gating for live cells (C). (D) Representative plot displaying non-transfected HEK293F cells (black dots) and S1-positive cells (pink).

3. Supplementary Movies

Movie S1. Mass photometry of S1_{vaccine-antigen}.

Movie S2. Mass photometry of S_{recombinant-trimer}.

4. Supplementary Tables

Table S1. Mascot analysis. Identification of SARS-CoV-2 spike, nucleoprotein and membrane protein in the S_{virus} sample (from alpha lytic protease in-solution digestion).

Protein identified	Accession number	Mascot score	Mass	Sequence coverage
Spike glycoprotein SARS-CoV-2	P0DTC2	1206	141088	49%
Nucleoprotein SARS-CoV-2	P0DTC9	105	45598	31%
Membrane protein SARS-CoV-2	P0DTC5	48	25130	26%

Table S2. Raw data from FACS analysis. Table corresponds to data presented in Figure 3.

Antibody	Staining	Sample	Experiment 1				Experiment 2			
			% HEK293F	% Single Cells	% Live	% Positive	% HEK293F	% Single Cells	% Live	% Positive
CR3022	total staining	mock-1	70	97.6	80.2	1.02	74.4	98.9	79	0.17
		mock-1	72	96.8	83.2	1.45	73.9	99.2	81.4	0.34
		Srecombinant trimer-1	69.7	96.4	77.1	35	72.8	99.2	78.6	21.7
		Srecombinant trimer-2	68.2	95.9	77.3	32.7	70.8	99.3	78.3	23.3
		Svaccine antigen-1	74.4	96.5	79.9	24.4	74.1	99.3	77.9	15.6
		Svaccine antigen-1	73	96.8	79.2	26.2	73.1	99.5	78.1	16.9
CR3022	surface staining	mock-1	74.8	97.2	83	0.25	74.2	99.6	81.4	1.41
		mock-1	69.9	95	83.2	0.27	75.8	99.5	83	1.6
		Srecombinant trimer-1	68.6	95	79.1	2.78	74.8	99.5	80.7	5.15
		Srecombinant trimer-2	70.3	95.9	80	1.99	72.5	98.4	83.3	5.62
		Svaccine antigen-1	74.6	96.6	78.7	13.4	76.1	99.4	80.1	12.9
		Svaccine antigen-1	74.6	96.2	79.9	13	76.4	99.7	80.5	14.2
S2 (1A9)	total staining	mock-1	72.8	97.5	81.9	0.25	68.9	98.6	79.1	0.054
		mock-1	72.2	97.1	82.2	0.31	72.9	99.1	83.5	0.15
		Srecombinant trimer-1	69.7	95.9	79.8	28.1	69.9	99.1	80.3	23.6
		Srecombinant trimer-2	67.4	95.9	78.3	23.4	72.6	99.2	79.7	21.1
		Svaccine antigen-1	75.1	96.5	80.7	30.3	73.2	99	79	28.6
		Svaccine antigen-1	74	96.6	81	29.4	73	99.4	79.4	29.5
S2 (1A9)	surface staining	mock-1	71.5	98.9	83.2	0.53	73.3	99.4	80.7	0.083
		mock-1	70.6	98.5	83	0.53	72.9	99.4	82.9	0.13
		Srecombinant trimer-1	66.4	97.7	78.2	9.15	72.7	99.4	81.4	4.49
		Srecombinant trimer-2	67.7	98.3	80.6	7.26	71.4	99.3	81	5.28
		Svaccine antigen-1	71.4	98.9	79.3	28.6	73.2	99.4	79.3	25.5
		Svaccine antigen-1	73	98.9	81.3	27.3	74.1	99.6	79.5	27.2