The SARS-CoV-2 Spike Glycoprotein Directly Binds Exogeneous Sialic Acids: A NMR View

L. Unione, M. J. Moure, M. P. Lenza, I. Oyenarte, J. Ereño-Orbea*, A. Ardá*, J. Jiménez-Barbero*
# Table of Contents

1. Methods
   - Expression and purification of Pasteurella multica Sialytransferase (PmST1) S3
   - Expression and purification of mammalian (ST6GAL1) enzyme S3
2. Materials
   - Expression and purification of Pasteurella multica Sialytransferase (PmST1) S3
3. Preparative synthesis of CMP sialic acid
   - Scheme S1 S4
   - Figure S1 S4
4. Procedures for Enzymatic Synthesis
   - Procedure for the enzymatic synthesis of α2,3 Neu5Ac LacNAc by using PmST1 S5
     - Figure S2 S5
     - Figure S3 S6
   - Procedure for the enzymatic synthesis of α2,6 Neu5Ac LacNAc by using ST6Gal1 S6
     - Figure S4 S7
     - Figure S5 S7
5. Protein expression and purification of S and RBD proteins S8
6. NMR sample preparation and analysis
   - General information S8
     - Figure S6 S8
   - 2D STD-¹H,¹³C-HSQC S9
     - Figure S7 S9
   - Ligand epitope mapping analysis of 3´SLN* and 6´SLN* bound to the NTD S9
     - Figure S8 S10
   - Ligand epitope mapping analysis of 3´SLN* and 6´SLN* bound to the S protein S10
     - Figure S9 S10
     - Table S1 S11
     - Figure S10 S11
     - Table S2 S12
   - Figure S11. High resolution ¹H-¹⁵N HSQC spectra of ¹⁵RBD S12
7. References S14
1. Methods

Enzymatic reactions were monitored by mass spectrometry recorded on an UltraflieXtreme MALDI TOF/TOF Bruker Daltonics using 2,5-dihydroxybenzoic acid (DHB) as a matrix. Thin layer chromatography (TLC) was conducted on Silica gel 60 F254 (EMD Chemicals, Inc.) with detection by UV-absorption (254 nm) where applicable. TLC plates' visualization was accomplished by spraying with 10% sulphuric acid in ethanol and Hanessian's Stain, followed by charring at ~150 ºC. N-Acetyl-D-[UL-13C6] mannosamine and N-Acetylglactosamine uniformly 13C-labeled at the Gal residue were purchased from commercial source (Omicron Biochemicals) and used without further purification. Anion exchange chromatography was performed using a column (7 x 1.5 cm) packed with DEAE Sepharose Fast Flow resin (Cytiva), eluted with an ammonium bicarbonate solution as eluent.

2. Materials

Cytidine 5'-monophosphate-sialic acid (CMP-sialic acid) sugar nucleotide was prepared in one-pot enzymatic synthesis (Scheme S1).[1] Plasmids for enzyme's expression in E. coli were kindly provided by Prof. G. J. Boons (CCRC at the University of Georgia and Utrecht University). Those include recombinant N-Acetyl-D-neuraminic (Neu5Ac) acid aldolase (Aldolase, from Pasteurella multocida), CMP-sialic acid synthetase (CMP-Neu5Ac, from Neisseria meningitidis) and the Pasteurella multocida sialyltransase (PmST1, residues 26-412, Ppm0188Ph). The plasmids were cloned into pET23a (+) (GenScript). All enzymes were expressed in BL21 (D3) bacteria and cells and prepared in large scale as described elsewhere.[2,2] Alkaline phosphatase from Calf Intestinal was purchased from Invitrogen. The SARS-CoV-2 S1 NTD domain (aa 13-303) and the Anti-SARS-CoV-2 Spike NTD Neutralizing Antibody were purchased from Acris Biosystem (Cat No S1D-C52H6 and SPD-M121, respectively).

2.1 Expression and purification of Pasteurella multocida Sialyltransferase (PmST1)[2]

PmST1 sialyltransferase was expressed in BL21 bacterial cells grown in LB broth media with Ampicillin (100 µg/mL) until cells reached an OD600 between 0.6-0.8. Protein expression was induced with 0.2 mM isopropylthio galactopyranoside (IPTG), and the cell culture was incubated at 20 ºC with rapid shaking (190 rpm) for 20 hours. Cells were harvested by spinning at 5500 rpm for 20 minutes at 4 ºC. The cell pellet was resuspended 1:1 (1 mg cell pellet: 1 mL buffer) in lysis buffer (100 mM Tris-HCl pH 8.0 and 0.1% triton X-100). Enzyme was released by sonication in presence of 1 mM of PMSF and 1 tablet of EDTA-free SIGMAFAST™ protease inhibitor cocktail. Cells were subsequently disrupted by eight 30 s, 10W sonication pulses applied at 60 s intervals. Extract was ultracentrifuged at 35000 rpm for 30 minutes at 4 ºC and the supernatant was loaded onto a Ni-NTA affinity column (GE Healthcare) chromatography pre-equilibrated with buffer A (50 mM Tris-HCl pH 7.5, 150 mM NaCl). The column was washed with a step gradient of 0 and 25 mM imidazole added to buffer A, at ten column volumes each. The His6-PmST1 eluted with ten column volumes of buffer B (50 mM Tris-HCl pH 7.5, 150 mM NaCl and plus 500 mM imidazole). Purified The purified enzyme was dialyzed against buffer 20 mM Tris-HCl pH 7.5 and stored in 10% glycerol at -80 ºC.

2.2 Expression and purification of mammalian Glycosyl Transferase (ST6GAL1) enzyme

The DNA of ST6Gal1 fused with GFP was cloned into pHL1-secl[3] plasmid and synthesized by Genescript Biotech. The sequence was codon optimized for mammalian cell expression. A 6x His-tag was added to the C-terminus of the protein. ST6Gal1 was transiently transfected in HEK293S Gn Tl- suspension cells (ATCC) using FectoPro (Polyplus Transfections) transfection agent in a 1:1 ratio at 0.8-1 x 10^6 cells/mL density. Cells were incubated in Minitron Pro shaker (Infors HT) with gentle agitation (130 rpm) at 37°C, with 70% humidity, for 6-7 days. Then, cells were harvested by centrifugation and the supernatant was charged in a HisTrap Ni-NTA column (GE Healthcare). The eluted protein was concentrated and separated in a Superdex 200 Increase size exclusion column (GE Healthcare) in the final buffer containing 50 mM Tris pH 7.4 and 150 mM NaCl. Fractions corresponding to the pure protein were concentrated to 1 mg/ml using Amicon concentrators (Merck Millipore) with 10 kDa cut-off and stored at -80 ºC.
3. Preparative synthesis of CMP-sialic acid

The preparative synthesis of labeled CMP-sialic acid was carried out in a Tris-HCl buffer (100 mM, pH 8.8) containing MgCl₂ (20 mM), labeled N-Acetyl-D-[UL-¹³C⁶]mannosamine (20 mg), sodium pyruvate [¹³C₃] (5 eq), CTP (1 eq), the recombinant sialic acid aldolase, and the CMP-sialic synthetase. The reaction mixture was incubated at 37 °C for 18 h in an incubator with shaking (140 rpm). The reaction was stopped by the addition of EtOH when TLC (elution system: EtOH/1M NH₄HCO₃; 7:3 vol/vol) analysis indicated reaction completion. The precipitate was removed by centrifugation. The supernatant was concentrated and used without further purification.

3.1. Scheme S1

Scheme S1. One-pot two-enzyme synthesis of CMP-sialic acid from ¹³C labeled ManNAc with a sialic acid aldolase and CMP-sialic acid synthetase.

3.2. Figure S1

Figure S1. 2D ¹H,¹³C-HSQC NMR (400 MHz, 298 K, 10% D₂O) spectrum of labeled CMP-sialic acid from the reaction mixture.
4. Procedures for Enzymatic Synthesis

4.1 Procedure for the enzymatic synthesis of α2,3 Neu5Ac LacNAc by using PmST1

Commercial N-Acetyl-D-[UL-13C6Gal]lactosamine acceptor and CMP-sialic acid (1.5 eq) were dissolved at a final acceptor concentration of 10 mM in a Tris-HCl buffer solution (100 mM, pH 8.5) containing MgCl₂ (20 mM). Recombinant PmST1 (200 µg per µmol acceptor) was added, and the reaction mixture was incubated overnight at 37 ºC with moderate shaking. Reaction progress was monitored by MALDI-TOF MS, and additional PmST1 was used until no starting material was detected. The reaction mixture was centrifuged using a Vivaspin® Sartorius ultrafiltration device (10 kDa MWCO) to remove enzymes, and the supernatant was subjected to anion exchange chromatography (see Methods Section). Fractions containing product were combined and lyophilized to give the desired product.

![Chemical Structure](image)

MALDI TOF-MS m/z calcd for C_{12}H_{13}C_{13}N_{2}O_{19} (M-1H)^- exact 686.2745, found 686.3002.

4.1.1 Figure S2

![NMR Spectrum](image)

**Figure S2.** 1D ¹H NMR spectrum (¹C decoupled) of compound 1 (800 MHz, 298 K, D₂O).
4.1.2 Figure S3.

Figure S3. 2D $^1$H,$^{13}$C HSQC spectrum of 1, with constant-time evolution during CT$_1$ = 23 ms and resulting $^{13}$C($^{13}$C)$n$ carbon multiplicity editing in inverted sign for C($^1$C)-groups.

4.2 Procedure for the enzymatic synthesis of $\alpha$2,6 Neu5Ac LacNAc by using ST6Gal1

Commercial $N$-Acetyl-D-[UL-1$^{13}$C$_6$Gal]lactosamine acceptor and CMP-sialic acid (1.5 eq.) were dissolved at a final acceptor concentration of 10 mM Tris-HCl buffer solution (100 mM, pH 8.5) containing BSA (0.1% wt/wt). CIAP (10U uL$^{-1}$) and ST6GAL1 (4.4 μg per μmol acceptor) was added. The reaction mixture was incubated overnight at 37 °C with gentle shaking. Reaction progress was monitored by MALDI-TOF MS, and additional ST6GAL1 was added until no starting material was detected. The reaction mixture was centrifuged using a Vivaspin® Sartorius ultrafiltration device (10 kDa MWCO) to remove enzymes and BSA, and the supernatant was subjected to anion exchange chromatography (see Method Section). Fractions containing product were combined and lyophilized to give the desired product.

MALDI TOF-MS m/z calcd for C$_{12}$$^{13}$C$_{13}$H$_{41}$N$_2$O$_{19}$ (M-1H)$^-$ exact 686.2745, found 686.2900.
4.2.1 Figure S4

Figure S4. 1D $^1$H NMR spectrum ($^{13}$C decoupled) of compound 2 (800 MHz, 298 K, D$_2$O).

4.2.2 Figure S5.

Figure S5. 2D $^1$H,$^{13}$C HSQC spectrum of 2, with constant-time evolution during CT$_{1} = 23$ ms and resulting in $^{13}$C($^{13}$C)n carbon multiplicity editing of inverted sign for C(C)1 groups.
5. Protein expression and purification of S and RBD proteins

The extracellular domain of S protein from the Wuhan-Hu-1 strain at the prefusion state (BEI NR52394) and the SB domain of the RBD (amino acid residues 328-533) of the SARS-CoV-2 were produced by transient transfection of HEK293F cells (Thermo Fisher Scientific) at 0.8-1 x 10^6 HEK293F cells/mL using FectoPro (Polyplus Transfections) in a 1:1 ratio. Cells were grown at 37°C in a MiniTron Pro shaker (Infors HT) incubator for 5-7 days. For the expression of the ^15N-labeled RBD, the Free-Style media (Thermo Fisher Scientific) was supplemented with the U-^15N, 98% algal amino acid mixture (Cambridge Isotope Laboratories, Cat No NLM-2161-PK). Proteins were purified from the supernatant until homogeneity following a first step of affinity chromatography in a HisTrap Ni-NTA column (GE Healthcare) and a second step of size exclusion chromatography by Superose 6, for the S protein, and Superdex 75 Increase (GE Healthcare) in the case of RBD. S and RBD protein samples were concentrated using Amicon concentrators (Merck Millipore) with 100 and 10 kDa cut-off, respectively.

6. NMR sample preparation and analysis

6.1 General Information

Proteins and ligands were dissolved in a phosphate saline (PBS) buffer pH 7.4 in either 10% or 100% (vol/vol) D_2O. Shigemi NMR tubes with a diameter of 5 mm were used. A Crison Basic 20 (Crison Instruments SA, Barcelona, Spain) pH meter was used to check and adjust the sample pH by adding the required amount of NaOD and DCl. For STD NMR experiments with the S protein a ligand/protein ratio of 100:1 was used (S protein concentration was 9 μM). In case of RBD and NTD protein constructs, the ligand/protein ratio was 50:1 (protein concentration was ~20 μM). All the NMR spectra were acquired on an 800 MHz BRUKER AVANCE III spectrometer, equipped with a TCI cryo-probe with z-gradient coil and TopSpin 3.2.7 (BRUKER) software was employed for data acquisition and processing. A ^1H STD NMR spectrum for the sole S protein was also acquired (figure S6).

6.1.1 Figure S6.

![Image](https://example.com/image.png)

**Figure S6.** 1D ^1H-STD NMR spectrum of the sole S protein ectodomain of SARS-CoV-2. The spectrum was acquired with aromatic (7.0 ppm) protein irradiation. In black, the off-resonance spectrum and superimposed in orange, the ^1H-STD spectrum. Protein concentration was 9.0 μM.

6.2. 2D STD-^1H,^13C-HSQC

Recently we presented a general description for the 2D STD ^1H,^13C-HSQC, which is a combination of the conventional ^1H, ^1H-STD with a 2D ^1H,^13C-HSQC. Briefly, the initial ^1H, ^1H-STD module was implemented with saturation by a train of 4 x 90º PC9_4 shaped pulses
(with 1 ms separation) during δ9 = 2 s, applied at either the methy H peak (0.84 ppm) or the aromatic H peak (7.00 ppm) with the non-saturation cut-off set at 1.18 ppm or 6.66 ppm, respectively, (i.e. distance Δν = 0.34 ppm), resulting in a PC9_4 pulse length of 33 ms at 800 MHz field strength. Protein saturation was alternated with offresonant irradiation (at -25 ppm) in successive scans. The STD spectrum was then constructed by simple subtraction of both 2D ¹H,¹³C-HSQC spectra. Blank 2D STD-¹H,¹³C-HSQC experiment of the S glycoprotein alone (figure S7A), the α2,3- (figure S7B) and α2,6- (figure S7C) sialyl N-Acetyllactosamine trisaccharides with ¹³C-labeled Gal and Neu5Ac were acquired as control.

6.2.1. Figure S7

**Blank 2D STD-¹H-¹³C HSQC NMR**

![S protein](image1.png)  ![3´SLN*](image2.png)  ![6´SLN](image3.png)

**Figure S7.** 2D STD-¹H-¹³C HSQC NMR experiments of the sole S protein ectodomain of SARS-CoV-2 (A), the sole 3´SLN* (B) and the sole 6´SLN* (C). The spectra were acquired with aromatic (7.0 ppm) irradiation. In black, the off-resonance spectrum and superimposed in orange, the STD-HSQC spectrum. Protein concentration was 9.0 μM, ligand concentration was 1.0 mM

6.3 Ligand epitope mapping analysis of 3´SLN* and 6´SLN* bound to the NTD

The NTD was mixed with 3´SLN* and 6´SLN*, and 2D STD-¹H,¹³C-HSQC NMR experiments were acquired. The resulting spectra showed the presence of STD NMR signals, demonstrating that the NTD contains a Sia binding site. Ligand epitope mapping analysis (figure S8) was performed by integration of the ¹H,¹³C-HSQC NMR cross peaks and assigning the value of 100% to the highest value, which was used to calculate the relative STD intensity of the other ¹H,¹³C-HSQC NMR cross peaks from the 2D STD ¹H,¹³C-HSQC experiment.
Figure S8. 2D STD-$^1$H-$^{13}$C HSQC NMR experiments for the interaction of the NTD of the S protein of SARS-CoV-2 with (A) 3´SLN* and (B) 6´SLN*. On the left: NMR spectra with aromatic (7.0 ppm) protein irradiation. In black, the off-resonance spectrum and superimposed in orange, the STD-HSQC spectrum. Protein concentration was 20.0 μM, ligand concentration was 1.0 mM. Ligand epitope mapping is presented on the right as relative STD intensities.

6.4 Detailed Ligand epitope mapping analysis of 3´SLN* and 6´SLN* bound to the S protein

Herein, we provide the relative and absolute STD intensities as deduced by the analysis of the 2D STD-$^1$H-$^{13}$C-HSQC NMR experiments acquired for the S glycoprotein in presence of 3´SLN* and 6´SLN*. The plot bar (figure S9) and the corresponding table1 refer to the aliphatic protein irradiation. The plot bar (figure S10) and the corresponding table2 refer to the aromatic protein irradiation.
Figure S9. Plot bar of the relative STD intensities as derived by 2D STD-\(^{1}H-^{13}C\) HSQC NMR experiments, with aliphatic protein irradiation, for the interaction of the S protein of SARS-CoV-2 with 3’S LN* and 6’S LN*. Protein concentration was 9.0 μM, ligand concentration was 1.0 mM.

### 6.4 2 Table S1.

Absolute and Relative STD intensities as derived by 2D STD-\(^{1}H-^{13}C\) HSQC NMR experiments analysis, with aliphatic protein irradiation.

|                        | 3’S LN* | 6’S LN* | 3’S LN* | 6’S LN* |
|------------------------|---------|---------|---------|---------|
| **Galactose**          | abs     | rel     | abs     | rel     |
| C1-H1                  | 0       | 0       | 0       | 0       |
| C2-H2                  | 0       | 0       | 0       | 0       |
| C3-H3                  | 0       | 0       | 0,16    | 15      |
| C4-H4                  | 0       | 0       | 0,25    | 25      |
| C5-H5                  | 0       | 0       | 0       | 0       |
| C6-H6s,r               | 0,07    | 12,3    | 0       | 0       |
| **N-Acetyl-Neuraminic Acid** |         |         |         |         |
| C3-H3eq                | 0,18    | 30      | 0,26    | 25      |
| C3-H3ax                | 0,31    | 55      | 0,22    | 20      |
| C4-H4                  | 0,47    | 80      | 0,93    | 100     |
| C5-H5                  | 0,57    | 100     | 0,63    | 70      |
| C6-H6                  | 0,32    | 55      | 0,18    | 20      |
| C7-H7                  | 0,25    | 45      | 0,23    | 25      |
| C8-H8                  | 0,09    | 15      | 0       | 0       |
| C9-H9s                 | 0       | 0       | 0,09    | 10      |
| C9-H9r                 | 0,13    | 25      | 0       | 0       |
Figure S10. Plot bar of the relative STD intensities as derived by 2D STD-$^1$H-$^{13}$C HSQC NMR experiments, with aromatic protein irradiation, for the interaction of the S protein of SARS-CoV-2 with 3' SLN* and 6' SLN*. Protein concentration was 9.0 μM, ligand concentration was 1.0 mM.

6.4.4 Table S2.
Absolute and Relative STD intensities as derived by 2D STD-$^1$H-$^{13}$C HSQC NMR experiments analysis, with aromatic protein irradiation.

| Galactose | 3'SLN* | 6'SLN* |
|-----------|--------|--------|
|           | abs    | rel    | abs    | rel    |
| C1-H1     | 0.05   | 5      | 0.26   | 20     |
| C2-H2     | 0.31   | 30     | 0      | 0      |
| C3-H3     | 0.16   | 15     | 0      | 0      |
| C4-H4     | 0.26   | 26     | 0.24   | 20     |
| C5-H5     | 0      | 0      | 0      | 0      |
| C6-H6s,r  | 0.04   | 5      | 0      | 0      |

| N-Acetyl-Neuraminic Acid | 3'SLN* | 6'SLN* |
|---------------------------|--------|--------|
|                           | abs    | rel    | abs    | rel    |
| C3-H3eq                  | 0.4    | 40     | 0.3    | 25     |
| C3-H3ax                  | 0.18   | 20     | 0.18   | 15     |
| C4-H4                    | 0.96   | 100    | 1.15   | 100    |
| C5-H5                    | 0.77   | 80     | 0.81   | 70     |
| C6-H6                    | 0.5    | 50     | 0.48   | 40     |
| C7-H7                    | 0.48   | 50     | 0.4    | 35     |
| C8-H8                    | 0.25   | 25     | 0.11   | 10     |
| C9-H9s                   | 0.26   | 25     | 0.09   | 10     |
| C9-H9r                   | 0.17   | 20     | 0      | 0      |
Figure S11. Zoom areas of the $^1$H-$^{15}$N HSQC of $^{15}$N-RBD in absence (black) and presence (orange) of 50 equivalents of 3’SLN. No chemical shift perturbations were observed.
7. References

[1] H. Yu, H. Yu, R. Karpel, X. Chen, *Bioorganic Med. Chem.* **2004**, *12*, 6427–6435.
[2] H. Yu, H. Chokhawala, R. Karpel, H. Yu, B. Wu, J. Zhang, Y. Zhang, Q. Jia, X. Chen, *J. Am. Chem. Soc.* **2005**, *127*, 17618–17619.
[3] A. R. Aricescu, W. Lu, E. Y. Jones, Acta Crystallogr. D. Biol. Crystallogr. **2006**, *62*, 1243–1250.
[4] M. P. Lenza, I. Oyenarte, T. Diercks, J. I. Quintana, A. Gimeno, H. Coelho, A. Díaz, F. Peccati, S. Delgado, A. Bosch, M. Valle, G. Millet, N. G. A. Abrescia, A. Palazón, F. Marcelo, G. Jiménez-Ósés, J. Jiménez-Barbero, A. Ardá, J. Erefo-Orbea, *Angew. Chemie Int. Ed.* **2020**, *59*, 23763–23771.
[5] M. J. Moure, A. Gimeno, S. Delgado, T. Diercks, G. J. Boons, J. Jiménez-Barbero, A. Ardá, *Angew. Chemie - Int. Ed.* **2021**, *60*, 18777–18782.

Author Contributions

LU, MJM, and MPL, equal contribution, investigation, formal analysis, writing original draft; IO: investigation. JEO, AA, JJB, investigation, project administration, funding acquisition, validation, writing final manuscript. JEO, AA, JJB leaders