BRUKER

Recombinant SARS-CoV-2 Receptor Binding Domain: Comprehensive Top-Down Sequence Confirmation, Curation and O-Glycosylation Site Determination

Know your SARS-CoV-2 antigen structure before it´s use in research and diagnostics

Abstract

The COVID pandemic dramatically influences our life and the demand remains very high for recombinant SARS-CoV-2 surface proteins for diagnostics, vaccine development and research. The receptor binding domain (RBD) of the spike (S) protein mediates the interaction with the ACE2 receptor on host cells. As its structural features,

such as sequence variations or post translational modifications like N- and O-linked glycosylation may modulate its function, well characterized recombinant RBDs are essential. Therefore, we have recently per-formed an in-depth structural and functional charac-terization of RBDs expressed in Chinese hamster ovary (CHO) and human embryonic kidney (HEK293) cells [1].

Here we describe the structural characterization of the native RBDs expressed in CHO and in HEK293 cells. Intact mass analysis after glycan-enzymatic dissection and Top-Down se-quence analysis provided a comprehensive annotation of the RBD sequences used to clone the SARS-CoV-2 RBD and the localization of the O-glycosylation sites therein.

Keywords:
Covid, SARS-CoV-2 RBD, impact II QTOF, rapifleX, timsTOF fleX, mass spectrometry, Top-Down sequencing, O-glycosylation, BioPharma Compass

LUMC Leids Universitair Medisch Centrum

Authors: Christoph Gstöttner[1], Tao Zhang[1], Anja Resemann[2], Arndt Asperger[2], Detlev Suckau[2], Manfred Wuhrer[1], Elena Domínguez-Vega[1]; [1]Center for Proteomics and Metabolomics, Leiden University Medical Center, Albinusdreef 2, 2333ZA Leiden, The Netherlands; [2]Bruker Daltonics GmbH & Co. KG, Bremen, Germany.

Detailed functional analysis [2] and the glycan structure identification and profiling [1,3] utilized the elucidated RBD sequences established here.

Introduction

The coronavirus is an enveloped single-stranded RNA virus consisting of three different structural membrane proteins, the envelop (E) protein, the membrane (M) protein and the spike (S) glycoprotein. The S glycoprotein is heavily glycosylated and carries 22 N-glycosylation sites. Of particular interest is the RBD of the S protein, which mediates the interaction between virus and ACE2 receptor on the host cells. It carries two N-linked glycans at positions Asn-331 and Asn-343, and one or two O-linked glycosylation sites (Thr-323/Ser-325) were described to be occupied. Further findings suggest that glycosylation is involved in the binding to the receptor, emphasizing the importance of the assessment of RBD glycosylation.

Recombinant S proteins, including the RBD domain, are essential tools in the fight against COVID-19 [2]. However, it is important to realize that structural characteristics - including glycosylation - may differ between different biotechnologically produced proteins and in respect to their natural forms. Considering the relevance of RBD glycosylation on ACE2 binding and recognition by neutralizing anti-bodies, the use of well-characterized RBD proteins is crucial. Still, in previous Bottom-Up studies the localization of the O-glycosylation site at Thr-5 and/or Ser-7 was not successful. Thus, O-glycosylation was neglected completely [4].

Here we describe an in-depth Top-Down sequence characterization of two commercially available SARS-CoV-2 RBDs expressed in HEK293 or CHO cells. To obtain sequence information under optimal conditions, both RBDs were deglycosylated,

reduced and initially mass analyzed at the intact level followed by Top-Down Sequencing using MALDI in-source-decay (MALDI-ISD) mass spectrometry (MS) [5] to establish their protein sequences and putative non-glycan modifications. With the established sequences and the help of glycosidase treatment, the O-glycosylation site was assigned.

Experimental

Samples

SARS-CoV-2 RBDs (with a C-terminal His$_6$-tag added) were expressed in CHO and in HEK293 cells (InVivo Biotech Services) as described previously [1].

Figure 1: Sequence of the SARS CoV-2 RBD expressed in CHO cells with added His$_6$ tag (blue). The N- and putative O-glycosylation sites are highlighted in green and red, respectively. For the pQ-HEK293-RBD analysis, an N-terminal pyro-Glu residue had to be added (see Figure 6).
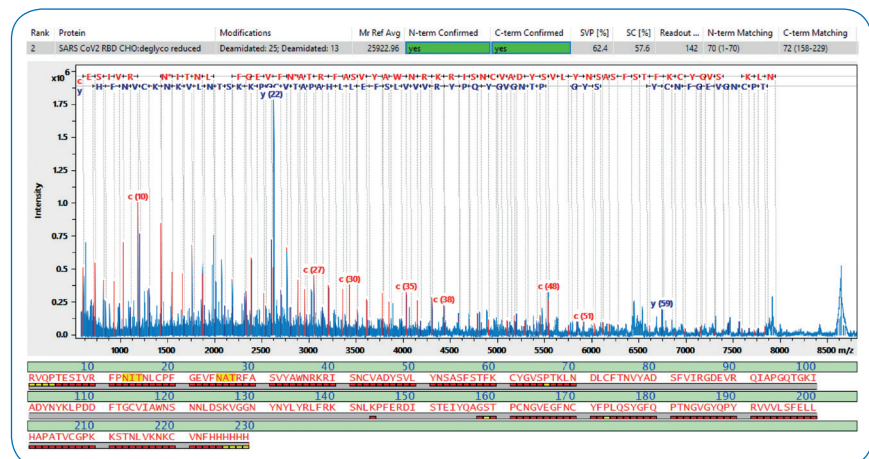

Figure 2: MALDI-ISD Top-Down sequencing spectrum (obtained on the rapifleX in sDHB matrix) of fully deglycosylated and reduced RBD expressed in CHO cells. N-terminal c-ions and C-terminal y-ions are annotated in spectrum and sequence, confirming ~70 residues from both termini and 62 % of the overall sequence.
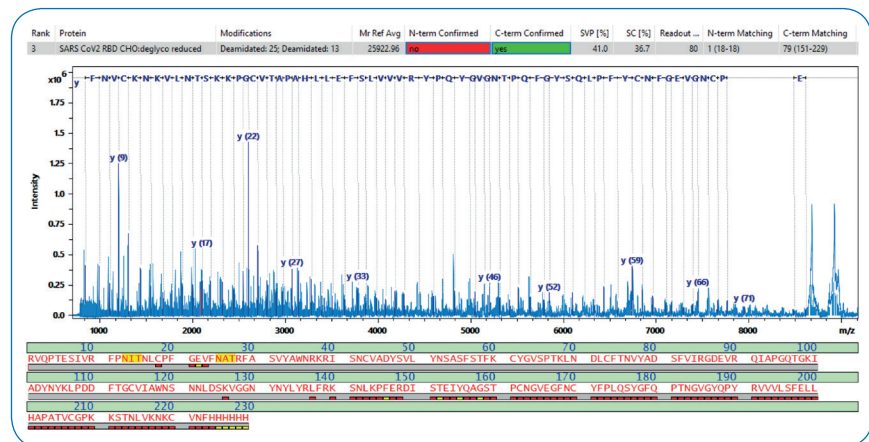

Figure 3: Top-Down sequencing spectrum of fully deglycosylated and reduced RBD expressed in HEK293 cells. N-terminal c-ions of the expected sequence did not match, only the C-terminal y-ions are annotated in spectrum and sequence, confirming 72 C-terminal residues, incl. the His$_6$-tag.

Glycosidase Treatment

N-linked glycans were removed using PNGAse F (Promega) following the instructions of the manufacturer.

For enzymatic cleavage of O-linked glycans the endo-α-N-acetylgalactosaminidase OglyZOR (Genovis) was used, which predominantly hydrolyzes core 1 type O-glycans. To further enhance deglycosylation, the sialidase SialEXO (Genovis) was applied simultaneously as recommended by the manufacturer.

For the detection of the O-linked glycosylation site (threonine or serine), samples were treated with PNGAse F and SialEXO to remove the complex N-glycosylation and reduce O-glycan heterogeneity to the core structures.

MALDI Sample Preparation and Analysis for rapifleX Measurements

Deglycosylated samples were reduced with DTT (30 min, 50°C). Two μL of the reduced sample (approx. 40 pmole/μL) were spotted on a hydrophilic anchor of an MTP BigAnchor sample plate (Bruker) and incubated for 2 min. The remaining droplet was removed, the spot zip-washed using 0.1% TFA in water and let dry. Subsequently, 1 μL of sDHB matrix (Bruker) solution (25 μg/μL in 50% AcCN/0.1% TFA in water) was deposited on the sample spot and let dry. MALDI-ISD spectra were acquired with a rapifleX MALDI-TOF MS instrument (Bruker) in positive reflector ion mode using a method optimized for ISD acquisition, provided by the manufacturer.

MALDI-ISD spectra were processed in flexAnalysis (Compass for flexSeries 2.0) using a smoothing step (5 cycles at 0.15 Da) and TopHat baseline subtraction. Peaks were picked using the SNAP2 and SNAP algorithms up to m/z 5000 and 10,000, respectively. The spectra were calibrated using a MALDI-ISD spectrum from 50 pmoles bovine ubiquitin prepared with sDHB. For interpretation and annotation of ISD and intact mass spectra BioPharma Compass 2021 was used (Bruker).

MALDI Sample Preparation and Analysis for timsTOF fleX Measurements

A 0.25 μL aliquot of the sample solution was spotted on a MTP AnchorChip 384 BC MALDI plate (Bruker) and was dried down and on-target zip-washed with 0.8 μL of cold washing buffer (1% TFA, 10 mM $NH_4H_2PO_4$). Washed samples were overlayed with 0.5 μL sDHB matrix solution (25 g/L in 50% AcCN, 0.1% TFA in water) and let dry at ambient air. For mass calibration, 0.5 μL aliquots of a suspension of milled red phosphorous in acetone were spotted on calibration spots without any MALDI matrix. Monoisotopic phosphorous cluster ion peaks were used for calibration.

MALDI-ISD spectra were acquired on the timsTOF fleX (Bruker) in positive ion mode. They were accumulated from 10,000 - 30,000 laser shots by rastering along the sDHB matrix outer rim (500 - 1,000 shots per raster position). For more details see [6].

Data were processed in the Bruker DataAnalysis 5.3 software using the SNAP2 peak finder. Top-Down sequences were analyzed in BioPharma Compass 2021 and Biotools 3.2 SR7 software.
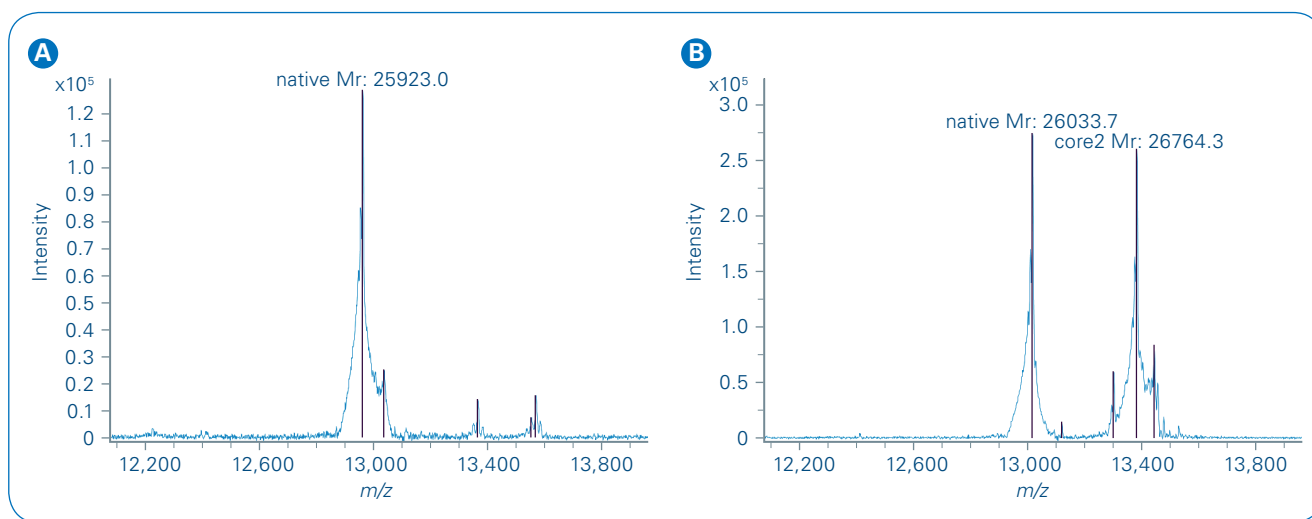

Figure 4: MALDI mass spectra of the reduced and deglycosylated RBDs expressed in (A) CHO and (B) HEK293 cells. The [M+2H]²⁺ ions are annotated with their respective neutral Mr [Da]. The CHO-RBD molecular weight corresponds to the expected Mr while the HEK293-RBD is 110.8 Da higher than expected. The presence of an additional Hex$_2$HexNAc$_2$ is deduced from the peak at 26764.3 Da.

Results and discussion

CHO-expressed RBD (CHO-RBD) was treated with PNGase F to fully remove N-glycans, with OglyZOR to remove core 1 O-glycans and DTT to reduce the disulfide bonds in order to determine the molecular weight and sequence of the RBD [Figures 1, 2].

The MALDI-ISD spectrum confirmed 62% of the full sequence including 70 N-terminal and 72 C-terminal residues. The single $[M+3H]^{3+}$ peak at approximately m/z 8640 represents the fully deglycosylated sequence.

In contrast, the corresponding MALDI-ISD spectrum of the HEK293-RBD solely matched the C-terminal sequence, indicating sequence aberrations or further modifications near the N-terminus (Figure 3). The peak doublet at ~m/z 8900 suggests the additional presence of core 2 O-glycosylation at only one of the putative glycosylation sites. OglyZOR together with SialEXO entirely removed any existing core 1 O-glycan but truncated the core 2 structure to only leave $Hex_2HexNAc_2$. We'll return to the subject later. To better characterize the N-terminal structure variation, we acquired intact mass spectra of the CHO- and HEK293-RBDs (Figure 4).

The intact RBD MALDI mass spectra were analyzed using the $[M+2H]^{2+}$ ions as they provide the most accurate results in our experience with antibody subunits. The observed m/z 12962.5 of the CHO-RBD resulted in an experimental Mr=25923.0 Da confirming the theoretical Mr=25922.96 Da (Figure 4A). In contrast, the peak at m/z 13071. 9 resulted in Mr=26033.7 Da for the HEK293-RBD, 110.8 Da higher than expected (Figure 4B). A second peak at m/z 13383.2 yielded an Mr=26764.3 Da, which was 730.6 Da higher, corresponding to the molecular weight of the core 2 O-glycan structure $Hex_2HexNAc_2$.

To resolve the N-terminal ambiguity with the +111 Da mass offset, we added the pro-peptide sequence to HEK293-RBD (pro-HEK293-RBD) as it had previously been used for cloning and re-analyzed the MALDI-ISD spectrum, allowing for protein clipping and for N-terminal pyro-glutamylation in BioPharma Compass´ Top-Down sequencing workflow.

In the BioPharma Compass software, clipping analysis of the extended HEK293-RBD with variably modified N-terminal pyro-Glu yielded a 65 residue N-terminal sequence match (Figure 5, top). Thus, an unexpected cleavage site within the pro-peptide sequence was revealed, leaving an additional Gln residue with the HEK293-RBD, which was additionally pyroglutamylated (pQ-RBD-HEK293) (Figure 5, center). The presence of the $Hex_2HexNAc_2$ O-glycan in the N-terminal region was also observed resulting in a 59 residue N-terminal sequence readout (Figure 5, top). However, no specific fragments were observed distinguishing the putative O-glycosylation sites from each other, because of non-glycosylated N-terminal ion series interference in the spectrum.
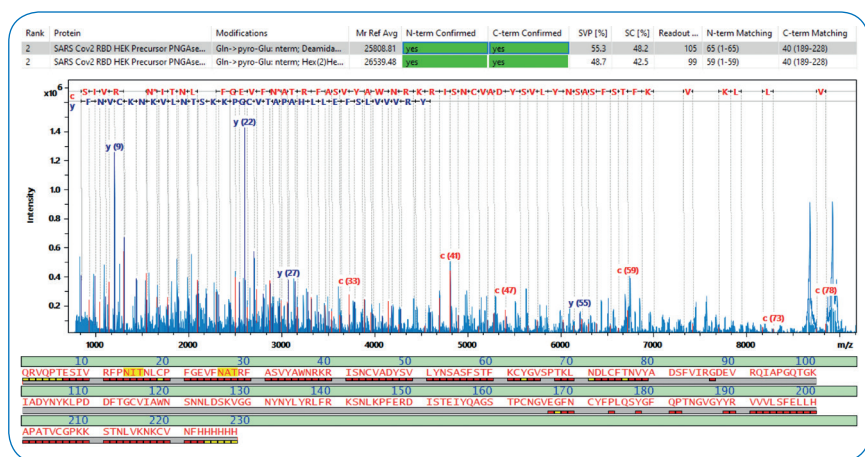

Figure 5: Top-Down sequencing analysis of fully deglycosylated and reduced RBD expressed in HEK293 cells; results table (top), spectrum (center), sequence (bottom). Clipping analysis of the RBD with added pro-peptide in BioPharma Compass yielded the best match with an RBD N-terminally extended by pyro-Glu. The addition of a variably modified $Hex_2HexNAc_2$ to the RBD also yielded a good match with slightly lower sequence coverage (see results table).
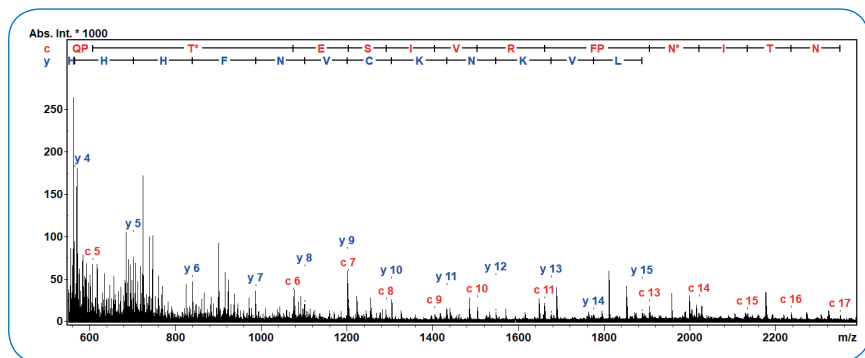

Figure 6: MALDI-ISD Top-Down sequencing spectrum of the pQ-HEK293-RBD, acquired on the timsTOF fleX with sDHB matrix after PNGase and SialEXO treatment and reduction. Just Thr-6 carried the $Hex_1HexNAc_1$ O-glycan – Ser-8 glycosylation was not observed.

To resolve the uncertainty, we combined PNGase F digestion with SialEXO – without addition of OglyZOR - in a second-round analysis, resulting in an RBD with desialylated core 1 and core 2 structures (365 and 730 Da). To clearly resolve Ser-8 from Thr-6 glycosylation, N-terminal c-ions down to m/z 600 need to be visible in the spectrum, which is difficult in axial MALDI-TOF-ISD spectra due of a high background below m/z 1000. The timsTOF fleX, however, with its high resolving power and mass accuracy - and a lower matrix cluster interference - allows to detect such low-mass peaks unambiguously.

MALDI-ISD spectra obtained on the timsTOF fleX [6] ion mobility spectrometry QTOF (Figure 6) allowed to safely assign Thr-6 as only O-glycosylation site. The spectrum additionally matches the c-ion series of $Hex_2HexNAc_2$-modified Thr-6 (annotation not shown) and demonstrates the presence or core 1 and core 2 O-glycosylation at Thr-6.

The established pQ-HEK293-RBD sequence and the Thr-6 O-glycosylation site were a precondition to the subsequent in-depth glycan profiling of O- and N-glycans as described elsewhere [1,3].

Learn More

For more information about sample preparation, data acquisition and analysis by MALDI Top-Down Sequencing in a video visit https://www.youtube.com/watch?reload=9&v=JOtA5ewt_cw

For more information about the SARS-CoV-2 antigens used in this study please visit https://www.invivo.de/cell-line-development/sarscov-2-antigens/

Conclusion

- SARS-CoV-2 receptor binding domain (RBD) sequences expressed in CHO and in HEK293 cells were analyzed by MALDI Top-Down sequencing on the rapifleX TOF/TOF and the timsTOF fleX.

- The CHO-RBD sequence was confirmed and the HEK293-RBD was curated by automated clipping analysis of the HEK293-RBD pro-peptide in BioPharma Compass, yielding the pQ-HEK293-RBD sequence with an N-terminal sequence extension by pyro-Glu.

- The pQ-HEK293-RBD sequence was shown to populate a single O-glycosylation site at position Thr-6 according to MALDI-ISD spectra obtained on the timsTOF fleX - Ser-8 was not glycosylated. The results were prerequisite to in-depth N- and O-glycosylation analyses of RBDs expressed in CHO and HEK293 cells [1,3].

- Top-Down sequencing has been useful in resolving the near-terminal O-glycosylation site occupation in this work. The usefulness of Top-Down sequence analysis of RBDs was also observed in recent work [4], demonstrating the power of mass spectrometry beyond peptide mapping to quickly resolve difficult structural questions in biopharmaceutical research under the immense pressure due to the evolving pandemic.

Learn More

You are looking for further Information?
Check out the link or scan the QR code.

www.bruker.com/massspectrometry

References

[1] Gstöttner C, Zhang T, Resemann A, Ruben S, Pengelley S, Suckau D, Welsink T, Wuhrer M, Domínguez-Vega E (2021). *Structural and functional characterization of SARS-CoV-2 RBD domains produced in mammalian cells.* Anal Chem, Apr 19. doi: 10.1021/acs.analchem.1c00893

[2] Welsink T, Wolfenstetter S (2021), *SARS-CoV-2 antigens successfully produced with comparable quality from transient transfection and stable cell pool expression systems.* InVivo Application Note 1885914.

[3] Gstöttner C, Dominguez-Vega E, Wuhrer M (2021), *Proteoform characterization of recombinant SARS-CoV-2 Receptor Binding Domains.* Bruker Application Note LCMS-187

[4] Roberts DS, Mann MW, Melby JA, Larson EJ, Zhu Y, Brasier AR, Jin S, Ge Y (2021). *Structural O-Glycoform Heterogeneity of the SARS-CoV-2 Spike Protein Receptor-Binding Domain Revealed by Native Top-Down Mass Spectrometry.* BioRxiv, https://doi.org/10.1101/2021.02.28.433291

[5] Resemann A, Jabs W, Wiechmann A, Wagner E, Colas O, Evers W, Belau E, Vorwerg L, Evans C, Beck A, Suckau D (2016). *Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing.* MAbs., **8**(2): 318–330. doi: 10.1080/19420862.2015.1128607.

[6] Asperger A, Resemann A, Evers W, Goedecke N, Suckau D (2021), *Next Generation MALDI Top-Down Sequencing of protein biotherapeutics – Expanding the scope of timsTOF technology.* Bruker Application Note LCMS-186

● **Bruker Daltonics GmbH & Co. KG**

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com