Fast and low-cost detection of SARS-CoV-2 peptides by tandem mass spectrometry in clinical samples

Cardozo, K.H.M. [#], Lebkuchen, A. [#], Okai, G.G. [#], Schuch, R.A. [#], Viana, L.G. [#], Olive, A.N., Lazari, C.S., Fraga, A.M. Granato, C.F.H., Carvalho, V.M.*

Grupo Fleury, São Paulo, SP Brazil

#Contributed equally.

*Corresponding author. Email: valdemir.carvalho@grupofleury.com.br.

Author contributions: conceived and designed the experiments: V.M.C., K.H.M.C., L.G.V., A.L., R.A.S. and G.G.O.; performed research: K.H.M.C., L.G.V., A.L., R.A.S., G.G.O. and V.M.C.; contributed the clinical samples and analytical data: A.N.O. and A.M.F.; analyzed data: K.H.M.C., L.G.V., A.L., R.A.S., G.G.O., V.M.C., C.S.L. and C.F.H.G.; wrote the paper: V.M.C., K.H.M.C., L.G.V., A.L., R.A.S. and G.G.O.; critically revised the manuscript: A.M.F. and C.F.H.G. All authors have read and agreed with the final version of the manuscript.

Abstract

The current outbreak of severe acute respiratory syndrome associated with coronavirus 2 (SARS-CoV-2) is pressing public health systems around the world, and large population testing is a key step to control this pandemic disease. Real-time reverse-transcription PCR (real-time RT-PCR) is the gold standard test for virus detection but the soaring demand for this test resulted in shortage of reagents and instruments, severely limiting its applicability to large-scale screening. To be used either as an alternative, or as a complement, to real-time RT-PCR testing, we developed a high-throughput targeted proteomics assay to detect SARS-CoV-2 proteins directly from clinical respiratory tract samples. Sample preparation was fully automated by using a modified magnetic particle-based proteomics approach implemented on a robotic liquid handler, enabling a fast processing of samples. The use of turbulent flow chromatography included four times multiplexed on-line sample cleanup and UPLC separation. MS/MS detection of three peptides from SARS-CoV-2 nucleoprotein and a ^{15}N-labeled internal global standard was achieved within 2.5 min, enabling the analysis of more than 500 samples per day. The method was validated using 562 specimens previously analyzed by real-time RT-PCR and was able to detect over 83% of positive cases. No interference was found with samples from common respiratory viruses, including other coronaviruses (NL63, OC43, HKU1, and 229E). The strategy here presented has high sample stability and low cost and should be considered as an option to large population testing.

INTRODUCTION

The outbreak of a novel coronavirus (SARS-CoV-2) was first identified in Wuhan, China at the end of 2019, posing a challenge to health systems worldwide [1]. The disease has rapidly spread around the world with over 2.9 million reported cases and more than 200,000 deaths confirmed as of April 28, 2020 [2].

Tests using real-time reverse transcription polymerase chain reaction (real time RT-PCR) were quickly developed after the viral genome sequence release [3]. Although highly sensitive, the soaring demand for this test caused a shortage of several reagents and instrumentation used for this method, severely limiting its applicability to large-scale screening [4-6].

Point-of-care tests are the most desirable alternative to ramp up large-scale population screening. Currently, hundreds of initiatives are on course to deliver serologic and antigen detection platforms. Serologic immunoassays targeting IgA, IgM, or IgG provide historic information about viral exposure, but their sensitivity in the acute phase of SARS-CoV-2 infections is still not well established. At the time of publishing, immunoassays directed towards antigen detection are being delivered but scarce information about their sensitivity and specificity is available. Immunoassays can be highly sensitive, simple to perform, and provide quick answers at a reasonable cost. However, they often suffer from interference because antibody recognition is not free of error and may be confounded by the presence of other molecules.

Mass spectrometry (MS) has become an essential tool in clinical laboratories and is the current gold standard for several clinical applications such as steroid hormone determination [7]. Recently, targeted proteomics, a derivative from proteomic-based mass spectrometry technology, has emerged as a viable alternative to immunoassays for protein analysis [8]. Generating targeted proteomics assays is faster than producing new antibodies and because mass spectrometry-based assays are intrinsically more specific than immunoassays, this technique can be an interesting addition to viral testing panels.

A few studies have described virus detection from clinical specimens. Majchrzykiewicz-Koehorst *et al.* [9] detected influenza A H3N2 and H1N1 in highly pure and concentrated samples obtained by culturing viruses in cell lines and spiked throat swabs. In addition to the time required to expand viruses in cultured cells, their strategy required 3 h to analyze 10 samples.

Foster and colleagues [10] detected human metapneumovirus (HMPV) directly from nasopharyngeal aspirates after concentration via size-exclusion chromatography, overnight trypsin digestion, and detection using 90-min multiple reaction monitoring assay.

This study aimed to develop a rapid, specific, and robust method to enable high-throughput screening to support large-scale SARS-CoV-2 clinical diagnostics. For this purpose, clinical respiratory tract samples were analyzed in a bottom-up proteomics workflow resulting in a spectral library used to generate a list of targeted peptides. The selected peptides were first refined by microflow chromatography coupled to high-resolution mass spectrometry and then exported to a high-throughput assay based on turbulent flow chromatography coupled to tandem mass spectrometry (TFC-MS/MS). The targeted proteomics approach was compared to an in-house real-time PCR method for SARS-CoV-2. Interference was assessed using clinical specimens from other coronavirus strains, rhinovirus, enterovirus, and influenza viruses.

RESULTS

Selection of target peptides for SAR-Cov-2 detection by microflow chromatography coupled to high-resolution mass spectrometry

Respiratory tract specimens previously analyzed by an in-house real-time RT-PCR method were directly processed by a shotgun proteomics protocol to investigate the presence of predicted peptides from the UniProt SARS-CoV-2 pre-release (downloaded on March 13, 2020). To increase the probability of detection of SARS-CoV-2 proteins, samples with low cycle threshold values (Ct), which correspond to high viral load, were selected. Simple ethanol precipitation was used to concentrate proteins from nasopharyngeal/oropharyngeal swab specimens conserved in virus transport media. Protein pellets were lysed by SDS, reduced, alkylated, and digested by trypsin. Data-dependent acquisition (DDA) analyses revealed the presence of 119 unique peptides from eight proteins out of the 14 proteins predicted by UniProt SARS-CoV-2 (Table 1). Nucleoprotein (NCAP_WCPV) accounted for 23.5% of the identified peptides and had a sequence coverage of 72.3%. The DDA experiments were used to create a spectral library for PRM design (Fig. 1). After filtering out missed cleavages and cysteine-containing peptides, 17 peptides were selected: nine from the nucleoprotein, five from the spike glycoprotein, two from the membrane protein, and one from the protein 3a (Supp. fig. 1). Using a first-round 60-min PRM acquisition, nucleoprotein peptides were found to exhibit approximately 80-fold higher relative intensities compared to peptides from other proteins and were thus selected for SARS-CoV-2 detection. A fast PRM method was achieved with a 9-min microflow chromatography separation using eight targeted peptides from nucleoprotein protein (Supp. fig. 2).

The specificity of the targeted peptides was also confirmed by blastp against the UniProt database. The presence of mutations in the targeted peptides was excluded after inspecting 18 individual nucleoprotein sequences that included all local sequences deposited in GISAID (Supp. fig. 3).

Turbulent flow chromatography coupled to triple quadrupole MS detection

To determine the best conditions for analysis of targeted peptides by turbulent flow chromatography, several conditions of loading, analyte transfer, washing, and re-equilibration were evaluated. For the loading step, different organic solvent contents (0, 5, and 10%) and formic acid concentrations (0, 0.1, and 0.5%) were tested of which the most efficient condition was 0.5% formic acid. The transfer step was the most critical; usual elution by 100% organic solvent plug used in the focus mode did not yield good results. Different organic proportions were analyzed, and best analyte transfer efficiency and peak shapes were achieved with 200 µL 60/40 0.5% formic acid/acetonitrile. The width of the transfer window was also evaluated in 6-s increments and 96 s provided the most efficient transfer for the selected peptides. It was observed that moving the flow reduction in the loading pumps forward by 6 s before they are connected during the transfer step improved peak shape. A preliminary analysis indicated the presence of carryover after injection of samples with a high virus load. To investigate the source of carryover, the injection syringe was removed. The contamination persisted indicating it was not related to the syringe or injection port. Next, tubing from the injector to the VIM module was replaced and the carryover was still observed. Finally, changing the TurboFlow

column revealed it as the main source of contamination. Thus, several tests were performed to reduce TFC column contamination. Different organic solvents were essayed, including methanol, acetonitrile, isopropanol, acetone, dimethyl sulfoxide, and trifluoroethanol. The incorporation of alternate flushing of the TFC column with 20% DMSO/2% TFE in water followed by an organic solvent mixture (acetonitrile/isopropanol/acetone, 40:40:20, v/v) proved to be the most efficient in reducing the carryover. The concurrent analytical separation in the UPLC column was achieved with a multistep linear gradient. Finally, the elution loop was filled, the analytical column was flushed with 80% acetonitrile, and both columns were equilibrated for the following injection. Final chromatographic conditions as well as valve-switching programming are shown in Table 2.

The transition list for the selected peptides for SARS-CoV-2 was exported from the 9-min PRM Skyline method and imported into the TraceFinder Instrument Setup module. Among the initial eight target peptides there were important differences in polarity, and the most hydrophobic (DGIIWVATEGALNTPK, IGMEVTPSGTWLTYTGAIK, and WYFYYLGTGPEAGLPYGANK) were more efficiently focused onto the analytical column. Because they yielded sharp peaks and therefore better signal-to-noise ratios, these peptides were selected for high-throughput analysis. The peptide HSGFEDELSEVLENQSSQAELK from the ^{15}N-labeled protein used as a global internal standard was included because it was detected within the chromatographic window for SARS-CoV-2. Three transitions were monitored for each peptide (Table 3). By eluting the target peptides within a 2.5 min window it was possible to multiplex by 4 for a total run time of 10.5 min per sample (Fig. 2). Thus, this system can process more than 500 samples per day.

Automated sample preparation

From the initial nonautomated SP3 sample processing, several variables from bottom-up proteomics were investigated to reduce processing time of disulfide bonds reduction, alkylation, lysis, and digestion steps. The elimination of the alkylation step did not affect the detection of the target peptides. To assess the efficiency of protein capture on the magnetic beads, as well as other downstream steps such as digestion efficiency and LC-MS/MS detection, a fully ^{15}N-labeled protein (chromogranin A) was introduced in the first step of sample preparation. Just after the ethanol step for protein precipitation on the particles, a lysis buffer was added and the mixture was heated to 65 °C, inactivating the virus. The optimized non-automated protocol was implemented in the Hamilton Robotics Microlab STARlet liquid-handler aiming for full automation of sample preparation (Fig. 3). The digestion step, which bottlenecks sample preparation, was reduced to 2 h with no loss in sensitivity to the targeted peptides resulting in a 3-h processing time for 96 samples.

Data analysis and analytical validation

The sensitivity, specificity and positive and negative predictive values of the in-house targeted proteomic method were evaluated by comparing the different parameters extracted from Skyline and the combination between them, based on cutoffs that were determined using the receiver operating curve (ROC) against the gold standard real-time RT-PCR (Table 4). The area under the curve (AUC) for each parameter used for SARS-CoV-2 detection is presented in Supplemental figure 4. The AUC at the maximum efficiency point was similar for SN DGI and IGM/IS (0.883 and 0.886, respectively), whereas for SN IGM it was 0.912 (Supp. fig. 4). The

combination of the three parameters (SN IGM, SN DGI and IGM/IS) enabled to distinguish between positive and negative samples with a sensitivity of 83.8% and a specificity of 96.1% (Table 4). Repeatability measurements evaluated over five days revealed coefficients of variation (CVs) of 20% for a positive sample (Supp. fig. 5).

In order to determine the influence of temperature on the stability of targeted proteins, samples stored at 21 °C, 4 °C and -20 °C were compared to samples stored at -80 °C. Sterile saline samples were stable even at room temperature for up 5 days, whereas virus transport media samples required storage at 4 °C (Supp. table 4). The protein biomarker presented stability after treating samples at 90 °C for 5 minutes (Supp. table 5).

Pearson's correlation coefficients of classification parameters SN_DGI, SN_IGM, and IGM/IS were higher than 0.8. Moderate correlations (r = 0.5–0.7) were observed between viral loads estimated from real-time RT-PCR Cts and the classification parameters (Fig. 4).

The lack of cross-reactivity with other viruses was confirmed by testing specimens previously characterized by molecular testing. Targeted peptides from SARS-CoV-2 were absent in monoinfection cases (coronavirus 229E, coronavirus NL63, parainfluenza 1, parainfluenza 4, influenza A/H1-2009, human metapneumovirus, and respiratory sincycial virus) and coinfection cases (coronavirus HKU1/rhinovirus/enterovirus, rhinovirus/enterovirus, and rhinovirus/enterovirus/human metapneumovirus) (Supp. table 6).

DISCUSSION

In response to the urgent need of developing alternative diagnostic tests for the novel coronavirus (SARS-CoV-2) pandemic, we developed a method that satisfies the following requirements: specificity towards SARS-CoV-2; good sensitivity compared to the gold standard real-time RT-PCR; sample preparation of high-volume processing in the shortest time possible; and fast acquisition by multiplexing liquid chromatography coupled to tandem mass spectrometry. The first step into this process was to establish surrogate peptide targets for SARS-CoV-2 proteins. Specimens with high viral loads previously characterized by real-time RT-PCR enabled building a spectral library from data-dependent analysis that served as the basis for determining the most suitable peptide candidates. An initial 60-min PRM method explored 17 peptides from four proteins and revealed that nine nucleoprotein peptides were the most intense and therefore promising option for SARS-CoV-2 detection. Also, their sequences were unique among those deposited in UniProt and a local search for mutations showed that these regions are conserved.

Using microflow liquid chromatography coupled to hybrid quadrupole-orbitrap MS analysis we were able to complete sample acquisition in only 9 min. However, to achieve high-volume testing, we investigated alternatives to LC-MS/MS. Four-channel turbulent flow chromatography coupled to triple quadrupole MS detection was able to meet this requirement by increasing the speed of analysis and by incorporating an additional on-line solid-phase cleanup with the dual-column approach. TFC has enabled high-throughput determination of several important clinical biomarkers, such as small metabolites and proteins [11, 12]. The TurboFlow Cyclone-P HPLC 0.5 × 50 mm column used for turbulent flow chromatography was selected for its robustness while requiring lower flow rates to generate turbulence [9]. From the initial panel of eight selected peptides, the three most hydrophobic yielded the best results in TFC coupled to triple quadrupole MS detection. Two steps limit the speed of dual-column TFC assays, namely transfer and gradient elution [13]. During the transfer step, the extraction solvent previously stored in the loop is delivered to the extraction column combined with a weak mobile phase from the elution pump to focus the analytes onto the analytical column. The most hydrophobic peptides, DGIIWVATEGALNTPK, IGMEVTPSGTWLTYTGAIK, and WYFYYLGTGPEAGLPYGANK were most efficiently focused onto the analytical column and eluted within a 2.5 min window, enabling multiplex by 4 for a total run time of 2.5 min per sample. Thus, the system can process more than 500 samples in 24 hours.

Carryover is a ubiquitous phenomenon in liquid chromatography whereby the analytes from a previous injection are retained by adsorption in the flow path within the LC-MS system and detected in the subsequent injections. In the specific case of qualitative detection of SARS-CoV-2 by targeted proteomics, carryover may represent one of the greatest analytical challenges because viral load can range over eight orders of magnitude and erroneous biomarker identification can lead to false positive detections. Thus, a significant part of the efforts to develop this method was dedicated to minimizing carryover. After isolating the main sources of carryover within the LC-MS system, the TurboFlow column was found to be responsible for the largest part of the carryover observed. Of the cleaning solutions evaluated to reduce carryover in the TFC column, DMSO and TFE were the most effective. DMSO was previously indicated to reduce peptide adsorption when added to the sample solution and autosampler washing solutions [14]. Here, DMSO was included in the first autosampler washing solution and in the TFC mobile phase D. TFE was also previously reported to reduce peptide carryover in nano chromatography [15]. Here, a combination step alternating 20% DMSO and 2%

TFE in water with an organic cleaning solution (acetonitrile/isopropanol/acetone, 40:40:20, v/v) was used to reduce carryover resulting in a 4.5-min TFC column cleanup, which starts after analyte transfer to the chromatographic column. Nevertheless, as a carryover-free system is virtually unachievable, a rule was established in data processing where the succeeding two samples are reinjected when following a high-intensity sample.

To enable the analysis of hundreds of samples with a short turnaround time, we implemented a modified automated single-pot solid-phase-enhanced sample preparation (SP3). SP3 is based on the binding of proteins to paramagnetic beads in the presence of an organic solvent followed by extensive washings and digestion [16]. Recently, Müller and colleagues described an automated SP3 implemented on an Agilent Bravo platform, which was applied to cell culture and formalin-fixed paraffin-embedded tissue sections [17]. Their method required 3.5 h to process 96 samples until reaching the digestion step, which required an additional 4 h. With our strategy, all steps can be completed in the robotic platform and the resulting microplate is ready to be processed by TFC-MS/MS. However, one offline centrifugation step was included to accelerate the magnetic bead collection. The use of a robotic liquid handler not only reduced processing time but also decreased the risk of infection for the laboratory personnel.

Although most targeted proteomics strategies use quantitative methods, the same approach may also be applied to qualitative or semi-quantitative analysis. Respiratory tract samples are intrinsically heterogeneous compared to biological matrices such as plasma and urine. Several factors influence virus load in respiratory tract samples such as method of collection, anatomical collection site (e.g., nasopharyngeal and/or oropharyngeal), type of swab, sampling at a given diagnostic window, and individual variability [18, 19]. Here, we focused on a dependable approach to successfully detect the virus in respiratory tract specimens rather than determine the amount of proteins in these clinical specimens. Even though data on specificity and sensitivity of immunoassays targeted to virus proteins are still scarce, targeted proteomics most likely present higher specificity. Moreover, selectivity is an intrinsic feature for mass spectrometry-based tests and combined with unique peptide sequences increases the potential application for this strategy [20, 21].

The diagnostic accuracy of the targeted proteomic method was enhanced by combining three parameters: SN IGM, SN DGI and IGM/IS. Although our method is less sensitive than real-time RT-PCR, it detects 83.8% of positive cases with high specificity (96.1%). The intrinsic real-time RT-PCR sensitivity is currently not possible to be surpassed by any protein analysis, mainly because proteins cannot be multiplied. However, in a scenario where large-scale population testing is needed and the supply for real-time RT-PCR reagents and instruments is insufficient to cover this demand [4, 5], targeted proteomics provides an alternative to complement real-time RT-PCR testing. Currently, the demand for tests based on LC-MS/MS dropped in clinical laboratories due to the several levels of social distancing necessary to limit the spread of SARS-CoV-2 around the world. Therefore, there are thousands of idle LC-MS/MS instruments that could be used for SARS-CoV-2 testing.

An additional positive aspect of targeted proteomics over real-time RT-PCR is the analyte stability. While it is recommended to store specimens for RNA testing at -70 °C right after sample collection [22], the targeted protein is stable in saline solution at room temperature for up to 5 days and therefore samples can be collected even in situations where their storing at -70 °C is not an option, like in remote areas. The targeted peptides can be detected after thermal inactivation at 90 °C for 5 minutes. It has been demonstrated that disinfection at 80 °C for 1 minute was effective to reduce coronavirus infectivity [23]. Thus, inserting the thermal

inactivation step would not affect the results and it would be an extra safe step to reduce the biologic risk for infection of laboratory staff during testing. Lastly, our targeted proteomics SARS-CoV-2 testing is approximately half the cost of real-time RT-PCR.

In conclusion, the application of automated sample preparation and multiplexing turbulent flow chromatography coupled to triple quadrupole mass spectrometry is a feasible alternative for detecting SARS-CoV-2 in clinical respiratory tract samples in large scale at a population level. Our strategy enables high-volume testing in a short turnaround time and can be combined with other tests currently used for detection of COVID-19 infection to help control the pandemic.

METHODS

Clinical samples

Respiratory tract samples (nasopharyngeal swab, oropharyngeal swab, and nasal wash) were collected in virus transport media or sterile saline solution and stored at −80 °C [22]. All specimens used in this study were previously analyzed by an in-house real-time RT-PCR method implemented according to WHO guidelines [24]. Ten positive samples with low-cycle threshold values were used to create 5 pools for shotgun bottom-up proteomics analysis. Cross-reactivity was evaluated against specimens of other human coronaviruses (HCoV-HKU1, HCoV-229E, and HCoV-NL63), Influenza A (H1N1), respiratory syncytial virus (RSV), human metapneumovirus (HMPV), parainfluenza virus types 1 and 4, and rhinovirus/coronavirus HKU1/enterovirus coinfection previously characterized by Biofire® FilmArray® Respiratory Panel (bioMérieux, Marcy-l'Étoile, France). This study included only specimens collected as part of standard diagnostic protocols that would normally be discarded. Patient identification was not recorded or registered, and only proteins related to SARS-CoV-2 were investigated.

Sample preparation

Nonautomated sample processing

Two hundred microliters of clinical specimens was transferred to 1.5-mL conical polypropylene tubes. Proteins were precipitated by the addition of 5 volumes of ethanol, followed by storage at −80 °C for 30 min and centrifugation at 4 °C, 14000 rpm for 15 min. Proteins were digested with trypsin using a modified single-pot solid-phase-enhanced sample preparation (SP3) protocol described by Hughes [16]. Briefly, after careful removal of supernatant by aspiration, the pellets were suspended in 50 µL of lysis buffer (1% SDS, 5 mM DTT in 50 mM TEAB pH 8.5) and lysed and reduced at 85 °C, 2000 rpm for 20 min in a thermomixer. Cysteine residues were carbamidomethylated with the addition of 5 mM iodoacetamide followed by incubation in the dark for an additional 20 min. Next, 20 µL of Sera-Mag magnetic carboxylate modified particles (GE Healthcare, Little Chalfond, UK), prepared by combining equal volumes of hydrophilic and hydrophobic particles washed with an equal volume of water and resuspended in water to reach a final concentration of 2 mg/mL, was added to the tubes followed by one volume of ethanol, and the mixture was incubated in a thermomixer at room temperature, 1000 rpm for 10 min. The beads were immobilized on a magnetic rack, the supernatant was removed, and the beads were washed three times with 200 µL of 80% ethanol. A total of 75 µL of trypsin (Gold Mass Spectrometry Grade; Promega, Madison, WI, USA) at 66.7 µg/mL in 50 mM TEAB pH 8.5 was added and the mixture was incubated overnight at 37 °C, 500 rpm in a thermomixer. Lastly, 5 µL of 10% TFA in water was added, the beads were immobilized on a magnetic rack, and the tryptic digest transferred to total recovery glass vials.

Automated sample processing

Automated sample preparation was achieved on a Hamilton Robotics Microlab STARlet liquid handling system (Hamilton Company, Reno, NV, USA) equipped with eight pipetting channels, 96-channel multi-probe head, labware gripper, and an automated heater shaker. The robotic

liquid handler was modified with a HEPA (high-efficiency particulate arrestance) filter connected to an exhaustion pump. Programming and operation were achieved using Hamilton Robotics Venus Three software. Two hundred microliters of clinical specimens was transferred to a 96-deep-well plate. Next, 30 µL of fully ^{15}N-labeled chromogranin A at 15 µg/mL (internal standard to verify protein capture, digestion, LC separation, and MS detection) was dispensed into the plate followed by 50 µL of Sera-Mag™ carboxyl modified magnetic particles suspension in water at 1 mg/mL. One volume of ethanol was then dispensed into the plate and the mixture was agitated at 900 rpm for 5 min. Samples were lysed and reduced with 50 µL of lysis buffer and incubated at 65 °C at 1000 rpm for 5 min. After lysis, an additional 100 µL of water and 150 µL of ethanol were added and the plate was agitated at 900 rpm for 5 min. The plate was spun down to accelerate the bead separation process and transferred to a Magnum™ EX Universal Magnet Plate (Alpaqua, Beverly, MA, USA) for 5 min where the supernatant was removed. The immobilized beads were washed once with 800 µL of 80% ethanol and twice with 200 µL of 80% ethanol. A solution of TPCK-treated trypsin (75 µL at 65 µg/mL; Sigma-Aldrich, St. Louis, MO, USA) in 50 mM TEAB pH 8.5 was dispensed and the plate was incubated at 37 °C, 1000 rpm for 2 h. Lastly, 10 µL of 10% TFA in water was added to each well and after short mixing, the beads were immobilized on a magnetic rack. Tryptic digests were transferred to a Protein LoBind Deepwell plate 96/500 µL. The target plate was removed from the robotic liquid handler and stored at −20 °C until analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Untargeted LC-MS/MS analysis

Target peptide selection was achieved on an UltiMate® 3000 Nano LC system coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) via an EASY-Spray™ source operating in positive ion mode. The UltiMate® 3000 Nano system was fitted with a PepMap100 C18 5 µm, 0.3 × 5 mm sample trapping pre-column and a PepMap RSLC C18 2 µm, 150 µm × 15 cm analytical column (Thermo Fisher Scientific).

Data-dependent acquisitions were obtained injecting 10 µL of the tryptic digest loaded in the trapping column with 0.1% TFA at 150 µL/min for 3 min. For chromatographic separation, the flow rate was 1.5 µL/min and the column was maintained at 45 °C; solvent A was 1% DMSO, 0.1% formic acid in LC/MS grade water and B was 1% DMSO, 0.1% formic acid in acetonitrile. A 60-min linear gradient was used as follows: 3–20% B for 50 min, 20–40 45% B for 10 min, 40–90% B for 2 min. Source parameters were set as follows: spray voltage = 2.2 kV, capillary temperature = 275 °C, and S-lens RF level = 50. The MS spectra were acquired with 120,000 mass resolution (m/z 200) from m/z 350 to 1650, AGC target of 3×10^6, and maximum IT of 60 ms. The MS/MS spectra were acquired for the 15 most intense ions of each MS scan (TopN = 15) with 15,000 mass resolution (m/z = 200), an isolation window of m/z 1.6, automatic gain control (AGC) target of 2×10^5, maximum injection time (IT) of 60 ms, and (N)CE = 27. Single-charged ions and those with more than six charges were excluded and a 20-s dynamic exclusion was used. The signal at m/z 401.92272 from DMSO was used as a lock mass.

Selection of target peptides

Data-dependent acquisition raw files were processed by the MaxQuant software version 1.6.14 [25] and searched against the UniProt SARS-CoV-2 pre-release (downloaded on March 13,

2020). Mass tolerance values for MS and MS/MS and the false discovery rate were set at 20 ppm and 1%, respectively. Carbamidomethylation of cysteine was set as a fixed modification, methionine oxidation and N-terminal acetylation as variable modifications. "Include contaminants" was left unchecked to not interrogate human proteins.

Skyline (version 20.1.1.32) [26] was used to build a spectral library from data processed by MaxQuant. The first set of candidate peptides was established importing UniProt SARS-CoV-2 pre-release into Skyline. Only peptides matching the library, fully digested and with no cysteine residues were included. Filtered peptides were exported into an isolation list to construct a parallel reaction monitoring (PRM) method for the mass spectrometer (Supp. table 1). Chromatographic and ion source parameters were identical to those described above. PRM data were acquired with 120,000 mass resolution (m/z 200), AGC target of 3×10^4, maximum IT of 250 ms, isolation window of m/z 1.6, and (N)CE = 27. Positive and negative samples were analyzed by the PRM method loaded into Skyline and the number of targets was reduced to the top 17 most intense ones across positive samples and absent in negative samples.

A homology search for targeted peptides was performed using blastp against SwissProt Uniprot [27] databases (retrieved on March 28, 2020).

SARS-CoV-2 sequences were downloaded from the GISAID (gisaid.org) platform on April 14, 2020. Searches were filtered from location (South America/Brazil) and only those sequences with full coverage on CDS coding region for nucleocapsid were included. The sequences were aligned together with NCBI Reference Sequence NC_045512.2 by the Clustal Omega server [28] and visualized in JalView alignment editor [29].

Fast separation PRM method

Fast PRM acquisitions were achieved with the following chromatographic separation process: samples were loaded into the trapping column with 0.1% TFA in water at 150 µL/min for 30 s; the flow rate was 1.5 µL/min and the column was maintained at 45 °C; solvent A was 1% DMSO, 0.1% formic acid in LC/MS grade water and B was 1% DMSO, 0.1% formic acid in acetonitrile. A 7-min linear gradient was used as follows: 24–25% B for 5 min, 25–80% B for 12 s keeping at 80% B for 30 s and returning to 24%. PRM data were acquired with 120,000 mass resolution (m/z 200), AGC target of 3×10^4, maximum IT of 250 ms, isolation window of m/z 1.6, and (N)CE = 27.

Turbulent flow chromatography coupled to triple quadrupole MS detection

A Transcend™ TLX-4 system consisting of four Dionex UltiMate 3000 quaternary pumps, four Dionex UltiMate 3000 binary pumps, one valve interface module (VIM), and one CTC PAL autosampler was coupled to a TSQ Altis™ Triple Quadrupole Mass Spectrometer fitted with a heated electrospray ionization (HESI) source (Thermo Fisher Scientific, San Jose, CA, USA). Aria MX (version 2.5, Thermo Fisher Scientific) was used to control the system and acquisition was done with TraceFinder software (version 4.1, Thermo Fisher Scientific). The TLX-4 system was fitted with four TurboFlow Cyclone-P HPLC 0.5 × 50 mm columns (Thermo Fisher Scientific) and four Acquity UPLC BEH C18, 1.7 µm, 2.1 mm × 50 mm columns (Waters, Milford, MA, USA). The mobile phase for the first dimension was 0.5% acid formic in water (mobile phase A), acetonitrile (mobile phase B), acetonitrile/isopropanol/acetone (40:40:20, v/v) (mobile phase

C), and 20% DMSO/2% TFE in water (mobile phase D). The mobile phase for the second dimension was 0.1% acid formic, 1% DMSO in water (mobile phase A) and 0.1% acid formic, 1% DMSO in acetonitrile (mobile phase B).

High-throughput screening acquisitions were obtained by injecting 25 µL of the tryptic digest sample onto the TurboFlow column with 0.5% acid formic in water at 1.2 mL/min. The flow was then reversed and slowed, and the retained peptides were eluted and transferred onto the analytical column. The total run time was 10 min, but multiplexing enabled a four-fold reduction in the overall analysis time. TSQ Altis™ optimized parameters were set as follows: spray voltage (kV): +4.0, sheath gas pressure (arb): 60, auxiliary gas pressure (arb): 15, sweep gas pressure (arb): 2, ion transfer tube temperature (°C): 300, vaporizer temperature (°C): 200, Q1 Resolution (FWHM): 2.0, Q3 Resolution (FWHM): 2.0, and CID gas (mTorr): 1.5. Peptides were detected using selected reaction monitoring (SRM) at a dwell time of 100 ms per transition. Collision energy (CE) and RF lens voltage (RF) for the three selected peptides and ^{15}N-labeled internal global standard were optimized using the Skyline optimization pipeline [30].

Data processing and interpretation

Data processing was done using Skyline. Briefly, the raw data were imported, peak integration was reviewed individually, and the results were exported to a spreadsheet along with the ratios of the internal global standard, targeted peptides areas, and background noise. A set of parameters was used to discriminate between positive and negative samples: signal-to-noise ratio (SNR) for peptides (IGMEVTPSGTWLTYTGAIK and DGIIWVATEGALNTPK) filtered by the limit of blank (LOB) and IGMEVTPSGTWLTYTGAIK area corrected by the ^{15}N-labeled global standard.

Data exploration was conducted using the Python/Scikit-learn library [31]. Viral loads were estimated using the equation $y = 3E+12e^{-0.693x}$, where x is the cycle threshold value and y is the estimated viral load (number of virus copies).

Analytical validation

Method analytical validation was based on the Clinical Laboratory and Standards Institute (CLSI) guidelines for evaluation of the qualitative method (EP12) [32]. Sensitivity and specificity were established by comparison with the in-house real-time RT-PCR method for SARS-CoV-2. The total number of samples analyzed for comparative studies was 562. Interference was assessed using clinical specimens from other human coronaviruses (HCoV-HKU1, HCoV-229E, and HCoV-NL63), influenza A (H1N1), rhinovirus, enterovirus, respiratory syncytial virus, human metapneumovirus (HMPV), and parainfluenza virus types 1 and 4. Limit of blank was estimated by measuring 38 replicates of a negative sample (no signal detected by real-time RT-PCR) and calculating the mean result and the standard deviation of signal-to-noise ratio multiplied by 1.645 (assuming a 95% confidence interval) (Supp. table 2). Limit of detection was calculated with the standard deviation of the signal-to-noise ratio (SN) from 20 replicates of pool samples with low viral load (as determined by real-time RT-PCR) multiplied by 1.645 (assuming a 95% confidence interval) and added to the limit of blank (Supp. table 3) [33]. Precision was evaluated using negative and positive pools over five days and two replicates per day (n = 10) and the statistical analysis was performed using MSstats plugin for Skyline [34]. The

system carryover was analyzed by injections of high viral load samples followed by three sequential injections of blank samples; the peptide area of blank samples was then compared to the peptide area of high-intensity samples. The stability study used sterile saline and virus transport media pools with low and high viral loads samples stored at 21 °C, 4 °C, -20 °C and -80 °C for 5 days. All conditions were analyzed in triplicate in the same run and t-tests were performed between each condition and controls stored at -80 °C. Stability after thermal inactivation was evaluated by heating samples at 90 °C and then comparing with non-heated samples. Positive and negative quality controls were included in each run. If the control material failed to yield the expected results, the run was rejected. Statistical analysis was performed using Excel, EP Evaluator v12, and Skyline software.

ACKNOWLEDGMENTS

We thank Daniel Wagner Fonteles Alves for data analysis, Lucia Sivieri de Assis Rocha, Lina Yoshida, and David Domingues Pavanelli for helpful discussions, and Gabriela Rampazzo Valim, Luciana Guilhermino Pereira and Adauto Ayala for excellent technical assistance. We thank the editors at Publicase for reviewing this manuscript.

DATA AVAILABILITY

The mass spectrometry untargeted proteomics data (raw files and spectral library) have been deposited to the ProteomeXchange Consortium via the PRIDE [35] partner repository with the dataset identifier PXD019119 and 10.6019/PXD019119. Targeted analyses proteomics data (PRM 60- and 9-min analyses, SRM analyses and summary datasheet) are available through the Panorama repository [36] at the following link (https://panoramaweb.org/labkey/Fleury_SARS-Cov-2.url). All other data supporting the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

1. Wang, C., Horby, P.W., Hayden, F.G. & Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470-473 (2020).
2. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report - 90. (World Health Organization., Geneva, Switzerland.; 2020).
3. Zhang, Y.-Z., Vol. 2020. (ed. Virological) (http://virological.org/t/novel-2019-coronavirus-genome/319; 2020).
4. Akst, J. RNA Extraction Kits for COVID-19 Tests Are in Short Supply in US. in The Scientist Magazine (2020).
5. Baird, R.P. Why Widespread Coronavirus Testing Isn't Coming Anytime Soon. in The New Yorker (2020).
6. Esbin, M.N. et al. Overcoming the bottleneck to widespread testing: A rapid review of nucleic acid testing approaches for COVID-19 detection. *RNA* **rna.076232.120 [pii]** (2020).
7. Carvalho, V.M. The coming of age of liquid chromatography coupled to tandem mass spectrometry in the endocrinology laboratory. *J Chromatogr B Analyt Technol Biomed Life Sci* **883-884**, 50-58 (2012).
8. Method of the Year 2012. *Nat Methods* **10**, 1 (2013).
9. Majchrzykiewicz-Koehorst, J.A. et al. Rapid and generic identification of influenza A and other respiratory viruses with mass spectrometry. *J Virol Methods* **213**, 75-83 (2015).
10. Foster, M.W. et al. Targeted Proteomics of Human Metapneumovirus in Clinical Samples and Viral Cultures. *Anal Chem* **87**, 10247-10254 (2015).
11. Salameh, W.A., Redor-Goldman Mm Fau - Clarke, N.J., Clarke Nj Fau - Reitz, R.E., Reitz Re Fau - Caulfield, M.P. & Caulfield, M.P. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: total and free testosterone reference ranges.
12. Chen, Z.H. et al. Quantitative Insulin Analysis Using Liquid Chromatography-Tandem Mass Spectrometry in a High-Throughput Clinical Laboratory. *Clinical Chemistry* **59**, 1349-1356 (2013).
13. Zhou, S.L. et al. High-throughput biological sample analysis using on-line turbulent flow extraction combined with monolithic column liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **19**, 2144-2150 (2005).
14. Tamura, M. et al. Minimization of carryover for high-throughput liquid chromatography with tandem mass spectrometry analysis of 14 mycotoxins in corn grits. *Journal of Separation Science* **37**, 1552-1560 (2014).
15. Mitulovic, G. et al. Preventing Carryover of Peptides and Proteins in Nano LC-MS Separations. *Analytical Chemistry* **81**, 5955-5960 (2009).
16. Hughes, C.S. et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* **14**, 68-+ (2019).
17. Müller, T. et al. Automated sample preparation with SP3 for low-input clinical proteomics. *Mol Syst Biol* **16**, e9111 (2020).
18. Tang, Y.W., Schmitz, J.E., Persing, D.H. & Stratton, C.W. The Laboratory Diagnosis of COVID-19 Infection: Current Issues and Challenges. *J Clin Microbiol* (2020).
19. Lippi, G., Simundic, A.M. & Plebani, M. Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19). *Clin Chem Lab Med* (2020).

20. Arora, A. & Somasundaram, K. Targeted Proteomics Comes to the Benchside and the Bedside: Is it Ready for Us? *Bioessays* **41**, e1800042 (2019).
21. Sobsey, C.A. et al. Targeted and Untargeted Proteomics Approaches in Biomarker Development. *Proteomics* **20**, e1900029 (2020).
22. World Health Organization. Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020 (World Health Organization, Geneva; 2020).
23. Kampf, G., Voss, A. & Scheithauer, S. Inactivation of coronaviruses by heat. *J Hosp Infect* (2020).
24. Corman, V.M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* **25** (2020).
25. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).
26. Pino, L.K. et al. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* (2017).
27. Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515 (2019).
28. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
29. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).
30. MacLean, B. et al. Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry. *Analytical Chemistry* **82**, 10116-10124 (2010).
31. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).
32. Garrett, P.E., Lasky, F.D., Meier, K.L., Clark, L.W. & Clinical and Laboratory Standards Institute. User protocol for evaluation of qualitative test performance : approved guideline, Edn. 2nd. (Clinical and Laboratory Standards Institute, Wayne, Pa.; 2008).
33. Armbruster, D.A. & Pry, T. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev* **29 Suppl 1**, S49-52 (2008).
34. Choi, M. et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524-2526 (2014).
35. *Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res 47, D442-D450 (2019).*
36. *Sharma, V. et al. Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline. Mol Cell Proteomics 17, 1239-1244 (2018).*

Table 1. SARS-Cov-2 proteins identified by data-dependent analysis (DDA) [a.b].

Protein group	Entry name	Protein name	Unique peptides	Sequence coverage [%]	Mol. weight [kDa]
P0DTC2	SPIKE_WCPV	Spike glycoprotein	19	18.7	141.18
P0DTC3	AP3A_WCPV	Protein 3a	1	2.5	31.123
P0DTC4	VEMP_WCPV	Envelope small membrane protein	2	33.3	83.649
P0DTC5	VME1_WCPV	Membrane protein	5	26.6	25.146
P0DTC9	NCAP_WCPV	Nucleoprotein	28	72.3	45.625
P0DTD1; P0DTC1	R1AB_WCPV; R1A_WCPV	Replicase polyprotein 1ab; Replicase polyprotein 1a	39;24	12.2	794.05
P0DTD3	Y14_WCPV	Uncharacterized protein 14	1	9.6	80.496

[a] Proteins matched to SARS-CoV-2 (based on UniProt SARS-CoV-2 pre-release downloaded on March 13, 2020).
[b] The following SARS-CoV-2 proteins were not detected: protein 7a, non-structural protein 7b, non-structural protein 8, ORF10 protein, protein 9b, and uncharacterized protein 14.

Table 2. Chromatographic conditions for the loading and eluting pumps for the turbulent flow chromatography coupled to tandem mass spectrometry (TFC-MS/MS) analysis of SARS-Cov-2 nucleoprotein target peptides. Steps 1 to 3: sample loading; step 4: transfer of peptides onto the analytical column; steps 5 to 10: switching wash between organic solvents (C and D) to clean large particles and reduce carryover in the TurboFlow column, and peptide elution from analytical column; step 11: TurboFlow column equilibration and analytical column washing; step 12: system equilibration.

Step	Start (min)	Loading pump								Eluting pump			
		Flow	Grad	%A	%B	%C	%D	Tee	Loop	Flow	Grad	%A	%B
1	0	1.2	Step	100	–	–	–	====	out	0.3	Step	99	1
2	0.5	1.2	Step	100	–	–	–	====	out	0.3	Step	99	1
3	1	0.2	Step	100	–	–	–	====	out	0.3	Step	99	1
4	1.1	0.1	Step	100	–	–	–	T	in	0.3	Step	99	1
5	2.7	1.2	Step	60	40	–	–	====	in	0.3	Step	80	20
6	3.9	1.2	Step	–	–	–	100	====	out	0.3	Ramp	75.5	24.5
7	4.9	1.2	Step	–	–	100	–	====	out	0.3	Ramp	75.0	25
8	5.9	1.2	Step	–	–	–	100	====	out	0.3	Ramp	72.5	27.5
9	6.9	1.2	Step	–	–	100	–	====	out	0.3	Ramp	70	30
10	8.4	0.5	Step	100	–	–	–	====	out	0.3	Ramp	50	50
11	8.6	0.5	Step	100	–	–	–	====	out	0.3	Step	20	80
12	9.6	1.2	Step	100	–	–	–	====	out	0.3	Step	99	1

Loading: A (1–4) 0.5% formic acid in water; B (1–4) acetonitrile; C (1–4) isopropanol, acetonitrile, and acetone (40/40/20, v/v/v); D (1-4) 20% DMSO and 2% trifluoroethanol in water.
Eluting: A (1–4) 1% DMSO, 0.1% formic acid in water; B (1–4) 1% DMSO, 0.1% formic acid in acetonitrile.

Table 3. Selected reaction monitoring (SRM) parameters for determination of SARS-Cov-2 targeted peptides.

Peptide	Start time (min)	End time (min)	Precursor (m/z)	Product (m/z)	Collision energy (V)	RF Lens (V)
DGIIWVATEGALNTPK(+2)	1.7	2.3	842,948	930,489	25	160
				1001,526	25	
				1100,595	26	
				1286,674	25	
IGMEVTPSGTWLTYTGAIK(+2)	2.2	2.6	1013,021	1394,731	31	170
				1495,779	31	
				1594,847	31	
WYFYYLGTGPEAGLPYGANK(+2)	3.2	3.7	1134,044	649,33	21	190
				1173,589	21	
				1331,659	21	
HSGFEDELSEVLENQSSQAELK (heavy)(+3)	0.7	1.7	835,359	1017,462	27	170
				1143,423	27	
				1243,489	27	

Table 4. Test method performance based on different parameters to discriminate between positive and negative samples.

	Sensitivity%	Specificity%	PPV%	NPV%	95%CI
SN IGM	88.6	74.7	83.6	81.8	0.88 to 0.94
SN DGI	88.6	65.5	78.9	79.8	0.85 to 0.91
IGM/IS	89.2	69.4	80.9	81.5	0.86 to 0.91
Combination of parameters	83.8	96.1	96.9	80.3	0.86 to 0.91

PPV: positive predictive value; NPV: negative predictive value; SN IGM: signal-to-noise for peptide IGMEVTPSGTWLTYTGAIK; SN DGI: signal-to-noise for peptide DGIIWVATEGALNTPK; and IGM/IS: ratio of peptide IGMEVTPSGTWLTYTGAIK to the ^{15}N-labeled global internal standard