Journal of proteome research

pubs.acs.org/jpr

Perspective

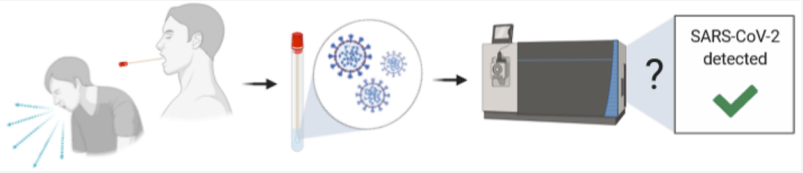
Perspective on Proteomics for Virus Detection in Clinical Samples

Marica Grossegesse, Felix Hartkopf, Andreas Nitsche, Lars Schaade, Joerg Doellinger, and Thilo Muth*

Cite This: *J. Proteome Res.* 2020, 19, 4380−4388

Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: One of the most widely used methods to detect an acute viral infection in clinical specimens is diagnostic real-time polymerase chain reaction. However, because of the COVID-19 pandemic, mass-spectrometry-based proteomics is currently being discussed as a potential diagnostic method for viral infections. Because proteomics is not yet applied in routine virus diagnostics, here we discuss its potential to detect viral infections. Apart from theoretical considerations, the current status and technical limitations are considered. Finally, the challenges that have to be overcome to establish proteomics in routine virus diagnostics are highlighted.

KEYWORDS: *virus proteomics, mass spectrometry, virus diagnostics, COVID-19, SARS-CoV-2, targeted proteomics, targeted mass spectrometry, peptide selection, parallel reaction monitoring, PRM*

1. INTRODUCTION

Application of Proteomics for Pathogen Detection

The current application of mass-spectrometry (MS)-based proteomics for infectious disease diagnostics highly depends on the pathogen type to be analyzed, for example, bacteria or viruses. Matrix-assisted laser desorption/ionization−time-of-flight (MALDI-ToF) MS has evolved as the method of choice for bacteria identification in clinical microbiology, as it is a fast and unbiased identification approach at a low cost. In conjunction with standardized sample preparation protocols and spectral libraries approved by the respective authorities, dedicated mass spectrometers enable the high-throughput identification of bacterial species. Currently, biotyping using MALDI-ToF MS is by far the most successful application of protein MS in clinical laboratories.[1−4] Unfortunately, this approach is not suited to identify viruses for two main reasons. First, viruses occur in a large cellular background and cannot be isolated purely by simple methods, as is the case for bacteria, which are routinely grown and isolated in colonies. Therefore, the sensitivity is highly limited by the dynamic range of the MS instrument. Second, viruses consist of much fewer proteins than bacteria and, hence, produce only a few if any signals in the narrow mass range of approximately 2−12 000 Da, which can be covered by whole-cell MALDI-ToF MS. The correlation analysis of spectra consisting of a few viral peaks in a large and varying human background is highly challenging. Although bacteria occur in a cellular background as well, the background can be significantly reduced by colony isolation. It remains elusive whether the combination of enrichment strategies and MALDI-ToF MS might be able to overcome the major current limitations and therefore extend the scope to viral diagnostics. Several years ago, the analysis of highly multiplexed polymerase chain reaction (PCR) by MS for the purpose of pathogen detection was commercialized under the names PlexID[5] and MassTag PCR.[6] However, these technologies have never become established in routine diagnostics because the analysis of multiplex PCRs has been adopted by DNA/RNA sequencing. In summary, MS currently has no impact on virus detection in the routine diagnostics of patient samples. However, given the current COVID-19 pandemic, a potential role of MS for virus detection is under evaluation, which can be seen in the growing number of publications and preprints dealing with the detection of SARS-CoV-2 by MS.[7−13] This Perspective aims at discussing the potential of MS-based proteomics for the diagnostics of viral infections, including the theoretical considerations, the current status, as well as the technical limitations and challenges.

Common Methods in Virus Diagnostics and Potential Advantages of Proteomics

Currently, real-time PCR as a genome-based detection method is the gold standard in virus diagnostics. Real-time PCR is, for example, also the method of choice for large-scale SARS-CoV-2 diagnostics currently done all over the world. For real-time

Special Issue: Proteomics in Pandemic Disease

Received: August 31, 2020
Published: October 22, 2020

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388

PCR diagnostics, the target virus has to be defined before the analysis because a small part of the virus genome is detected in a sequence-specific manner. In contrast with real-time PCR, next-generation sequencing (NGS) is a genome-based detection method that allows the comprehensive and untargeted analysis of the sample.[14] The advantage of NGS is that no prior knowledge of the target virus is necessary, but detection limits may vary dramatically.[15] Another method called AmpliSeq combines PCR with NGS, allowing targeted virus detection with higher sensitivity and multiplexing capacity.[16] Protein-based methods most often rely on antibodies for virus detection. These antigen-targeting immunoassays are performed in different formats, such as lateral flow assays, ELISA, or Western blot. The drawbacks in comparison with PCR are the long time needed for assay development, the lower sensitivity, and, moreover, potential issues with the specificity, as cross-reactivity against related virus species can occur. For example, antibodies against SARS-CoV-2 can be cross-reactive to other coronaviruses that only cause a common cold, depending on the epitope.[17] Nevertheless, virus diagnostics clearly benefit from immunoassays to detect viral proteins. For example, the p24 antigen, a viral capsid protein, is used to detect HIV infections by using ELISA.[18] Another example is the detection of the dengue virus (DENV) NS1 protein, a secreted viral protein, in serum or plasma samples. Notably, the NS1 protein is detectable in clinical samples even longer than viral RNA and hence allows detection in the time between viremia (the presence of the virus in the blood) and antibody production.[19] This demonstrates another advantage of protein-based virus detection, which is the higher compositional (not necessarily conformational) stability of certain proteins compared with that of nucleic acids.[20] Depending on the respective protein and sample storage, viral proteins might be less prone to degradation during sample transport compared with viral RNA, which could lead to fewer false-negative results. The current method of choice for sensitive and selective protein identification is MS-based proteomics. Apart from NGS, proteomics is the only technique with the potential to detect viruses in a comprehensive and untargeted manner, which would make it a powerful diagnostics approach.

Viral Protein Amounts in Clinical Samples

Real-time PCR is extremely sensitive, with limits of detection (LODs) of less than a single viral genome copy per microliter of the reaction mixture, depending on the assay. For example, a commonly used real-time PCR assay for the detection of SARS-CoV-2 has an LOD of 3.9 copies in a 5 μL reaction mixture.[21] Real-time PCR owes its sensitivity to target (nucleic acid) amplification. However, currently, there is no amplification method for proteins in proteome-based diagnostics; therefore, the overall sensitivity of the detection method itself has to be accordingly high. The advantage is that viral proteins are generally more abundant than DNA or RNA in a virus particle. Although the copy numbers of virion proteins have been determined in highly purified virus preparations, for example, for adenovirus,[22] influenza virus[23] and poxviruses,[24] quantification is not transferable to clinical samples. This is because clinical samples do not consist of isolated virus particles. Apart from virus proteins incorporated in virions, additional protein amounts produced inside the cell or secreted virus proteins are present in clinical samples, depending on the sample material. The DENV NS1 protein amounts in serum, for example, range between about 2.5 and 20 pmol/mL.[25] Another example is the HIV p24 antigen amount, which is generally analyzed by ELISA. To obtain CE certification, the LOD for HIV p24 antigen ELISA is 2 IU/mL WHO standard, corresponding to ~10.3 pg/mL (430 amol/mL).[26] Currently, LC-MS is not able to provide sensitivities in the low picogram per milliliter range in undepleted or unfractionated plasma.[27] Therefore, the enrichment of viruses, proteins, or peptides prior to analysis could be an option to increase the sensitivity of proteomics to a range that seems to be necessary to compete with the current technologies used for virus detection.[28] However, because quantitative virus proteomics of clinical samples has hardly been done yet, the amounts of most viral proteins in clinical samples remain elusive.

2. CURRENT STATUS OF MS-BASED VIRUS DETECTION IN CLINICAL SAMPLES

Beginning in 2015, Foster and colleagues applied multiple reaction monitoring (MRM) on a Waters Xevo TQ-S triple quadrupole instrument to detect human metapneumoviruses (HMPV) from eight clinical samples.[30] HMPV causes respiratory tract infections, and hence nasal airway fluid (nasopharyngeal aspirate) was used as the sample material. Using specific peptides, it was even possible to differentiate genetic lineages of HMPV by MRM. Next, in 2017, Schlatzer and colleagues used selected reaction monitoring (SRM) on a Thermo TSQ Quantum Ultra mass spectrometer to detect latent HIV-1 reservoirs in CD4+ T-cells from four patients with no PCR-detectable viral loads.[31] To enhance the sensitivity, they applied immunoprecipitation of the HIV Gag protein during sample preparation. In a publication from 2019, Wee and colleagues published the first parallel reaction monitoring (PRM) assay on an Orbitrap Q Exactive mass spectrometer to detect viruses from serum samples.[25] They were able to identify and differentiate serotypes of DENV from as little as 4 μL of serum. Differentiation of the four serotypes is important in DENV diagnostics because secondary infection with another serotype is associated with severe courses of disease. As shown, DENV serotypes can be differentiated by serotype-specific peptides, demonstrating the advantages of highly selective PRM-based virus detection. Finally, PRM outperformed PCR in serotyping in the late-acute phase when viral RNA amounts decrease.

Because of the COVID-19 pandemic, most recent publications about MS-based virus detection from clinical samples deal with SARS-CoV-2. It has to be noted that some of these publications are currently only available as preprints and should be treated with according caution. SARS-CoV-2 has been analyzed by MS from respiratory patient samples, like bronchoalveolar lavage,[32] nasopharyngeal and oropharyngeal swabs,[11,12] gargle solution,[8] and nasal wash.[11] However, the results vary greatly. Whereas it has been reported that the targeted analysis of SARS-CoV-2 by PRM does not reach the sensitivities of PCR so far,[8,11,32] it was possible to detect SARS-CoV-2, even from samples with higher Ct values, using parallel accumulation serial fragmentation (PASEF).[12] Interestingly, the authors state that using PRM, the sensitivity could be possibly further enhanced, even below the sensitivity of real-time PCR.[12] The different results obtained for MS-based SARS-CoV-2 detection, on the one hand, can result from the inherent heterogeneity of respiratory samples but, on the other hand, can also result from different sample preparations and MS analysis procedures. Moreover, Ct values between different

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388

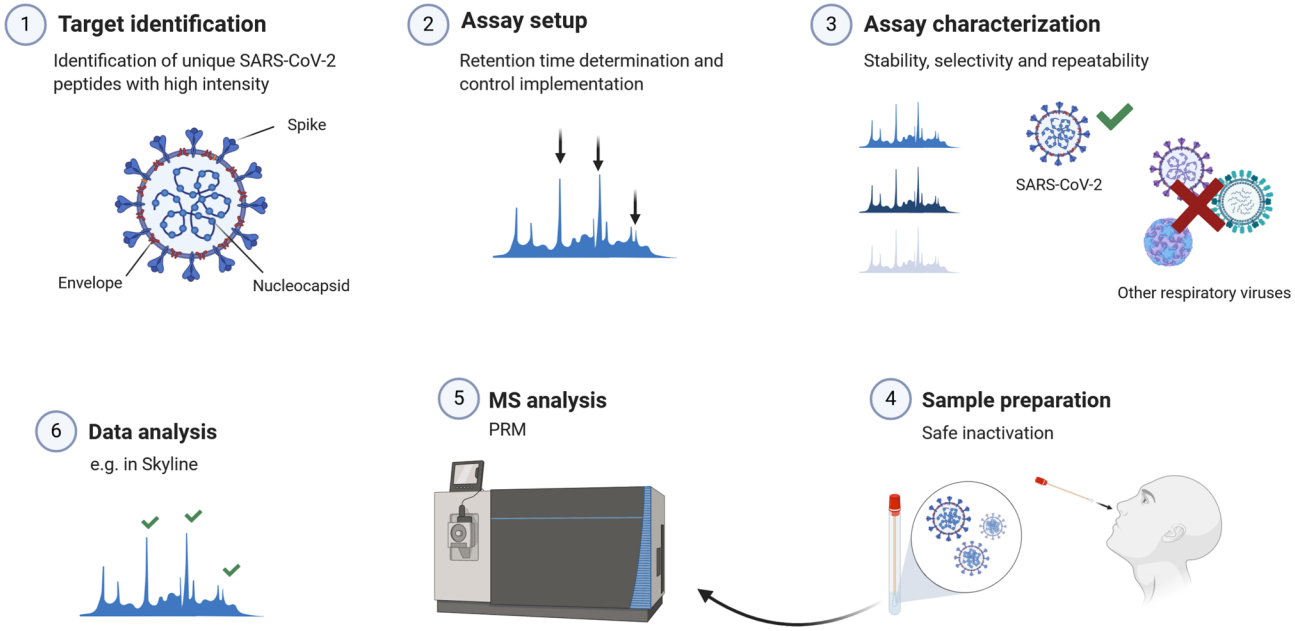


Figure 1. Overview of an exemplarily targeted assay development for SARS-CoV-2 detection from clinical samples. PRM has been applied to detect SARS-CoV-2 from respiratory samples.[8,11,32] Because workflows for assay development are quite diverse in detail, the basic steps are summarized and extended by the authors' experience, for example, control implementation and assay characterization.

PCR assays are not necessarily comparable, which is why the viral genome copies per volume is a more suited specification to allow better comparability between different publications.

In summary, until now, mainly targeted proteomics methods have been used in experimental approaches to detect viruses from clinical samples. On the basis of the cited publications/preprints and our own experience, we present in Figure 1 an overview of a potential workflow for the development of a targeted assay for SARS-CoV-2. The steps include: target peptide identification, assay setup and characterization, sample preparation, the MS run, and data analysis. PRM as a targeted data acquisition mode is further discussed as follows. However, targeted proteomics methods have not yet been applied in routine diagnostics, underlining the currently low impact of MS in virus diagnostics.

Concerning the data analysis, tools are mostly developed for bacteria identification or metaproteomics. However, some of them can still be directly applied to virus samples. For example, the recently published taxonomic identification tool TaxIt[33] was also tested on virus samples (e.g., cowpoxvirus and adenovirus strains) that have been measured by shotgun proteomics. TaxIt aims to overcome the problem of inaccurate species and strain resolution when proteomics data are searched against reference databases that lack taxonomic depth. In a first step, TaxIt applies reference sequence data for the identification of species candidates and, in a second step, it performs an automated acquisition of relevant strain sequences for low-level classification.

For designing targeted proteomics SRM/PRM assays, algorithms and software tools have been developed to support the user-defined selection of peptides. For example, Skyline[34] is one of the most widely used software tools in the field. During targeted assay design with Skyline, one can select unique peptides by loading a background proteome. However, it does not consider any sequence homologies between related organisms during the peptide selection process. Picky[35] presents a web-based application and method designer for targeted assays but currently only supports human and mouse sequences while being based on synthetic peptide data from the human-focused ProteomeTools project.[36] PeptidePicker[37] is a web-based application for selecting peptides after submission of the protein accession identifier and was specifically designed for human and mouse proteomes. As a specific use case of targeted metaproteomics, the Unique Peptide Finder of the UniPept web application[38] allows the selection of unique peptides for user-defined taxa and can therefore be interesting for the proteome-based analysis of viral samples. Finally, to fill the gap of MS-based virus diagnostics, Purple[39] presents a stand-alone software by which taxon-specific tryptic peptides can be selected for targeted proteomics assays in the context of virus diagnostics. Purple comes with a graphical user interface that makes it possible for the user to perform a specific homology-based peptide candidate selection across various taxonomic levels. In addition, it provides automated filtering against backgrounds of varying complexity. In summary, there are several tools facilitating viral peptide identification and data analysis for targeted analysis. However, there is currently no community-wide standard procedure for the development of targeted assays for the detection of virus peptides.

3. TECHNICAL LIMITATIONS AND CHALLENGES

Target Selection for Virus Detection

One of the first questions to ask when planning a proteomics experiment for performing virus diagnostics is which target protein(s) to choose. Because virus particles generally do not incorporate every encoded viral protein in the virus particle itself, a common approach in virology is to differentiate structural and nonstructural viral proteins. Structural viral proteins include, for example, proteins of the virus envelope and nucleocapsid, whereas nonstructural proteins may play a role, for example, in immunomodulation, transcription, or replication. Because structural proteins are generally more

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388

Table 1. Target Identification for PRM-based SARS-CoV-2 Detection

reference	method	targets spike protein	targets nucleocapsid protein
Bezstarosti et al.[7]	DDA of culture supernatant of infected cells; uniqueness not considered	N/A	ADETQALPQR
			EITVATSR
			GFYAEGSR
Cardozo et al.[11]	1. DDA of respiratory samples from SARS-CoV-2-infected patients	FNGIGVTQNVLYENQK[a]	ADETQALPQR[a]
	2. Identification of unique peptides by blastp against SwissProt UniProt	LQDVVNQNAQALNTLVK[a]	AYNVTQAFGR[a]
		LQSLQTYVTQQLIR[a]	DGIIWVATEGALNTPK[a]
		VGGNYNYLYR[a]	IGMEVTPSGTWLTYTGAIK[a]
		VYSTGSNVFQTR[a]	ITFGGPSDSTGSNQNGER[a]
			GQGVPINTNSSPDDQIGYYR[a]
			NPANNAAIVLQLPQGTTLPK[a]
			QQTVTLLPAADLDDFSK[a]
			WYFYYLGTGPEAGLPYGANK[a]
conservative approach using Purple[39]	1. *In silico* digestion of SARS-CoV-2 proteomes	DLPQGFSALEPLVDLPIGINITR[a]	ITFGGPSDSTGSNQNGER[a]
	2. Selection of unique peptides	FDNPVLPFNDGVYFASTEK[a]	
	3. Exclusion of nonconserved peptides	GWIFGTTLDSK[a]	
	4. Filter by protein and length	LPDDFTGCVIAWNSNNLDSK[a]	
		MFVFLVLLPLVSSQCVNLTTR[a]	
		TQLPPAYTNSFTR[a]	
		TQSLLIVNNATNVVIK[a]	
		VCEFQFCNDPFLGVYYHK[a]	
		VYSTGSNVFQTR[a]	
		VYSSANNCTFEYVSQPFLMDLEGK[a]	
Gouveia et al.[41]	1. DDA of SARS-CoV-2-infected Vero cells	FQTLLALHR[a]	ADETQALPQR[a]
	2. Unique peptide identification by NCBInr BLAST and GISAID data	GWIFGTTLDSK	AYNVTQAFGR[a]
	3. Identification of potential modifications and missed cleavages	HTPINLVR[a]	GFYAEGSR
		LQSLQTYVTQQLIR[a]	GPEQTQGNFGDQELIR
			IGMEVTPSGTWLTYTGAIK
			NPANNAAIVLQLPQGTTLPK[a]
			WYFYYLGTGPEAGLPYGANK[a]
Orsburn et al.[9]	1. *In silico* digestion of SARS-CoV-2 proteome	CVNFNFNGLTGTGVLTESNK[a]	ADETQALPQR[a]
	2. Multistep selection of diagnostic peptides	DIADTTDAVR[a]	DQVILLNK[a]
	3. Selection of highly abundant proteins using published MS data	DLPQGFSALEPLVDLPIGINITR[a]	QLQQSMSSADSTQA[a]
	4. Remove targets from highly variable regions	FDNPVLPFNDGVYFASTEK[a]	
		GVYYPDK[a]	
		GWIFGTTLDSK[a]	
		HTPINLVR[a]	
		IADYNYK[a]	
		LPDDFTGCVIAWNSNNLDSK[a]	
		MFVFLVLLPLVSSQCVNLTTR[a]	
		NIDGYFK[a]	
		QIAPGQTGK[a]	
		SWMESEFR[a]	
		TQLPPAYTNSFTR[a]	
		TQSLLIVNNATNVVIK[a]	
		VCEFQFCNDPFLGVYYHK[a]	
		VGGNYNYLYR[a]	
		VQPTESIVR[a]	
		VYSSANNCTFEYVSQPFLMDLEGK[a]	
		VYSTGSNVFQTR[a]	
		YNENGTITDAVDCALDPLSETK[a]	
Zecha et al.[32]	1. Analysis of SIL peptides	FASVYAWNR	ADETQALPQR
	2. DDA of cell culture supernatant	FLPFQQFGR[a]	AYNVTQAFGR
	3. Identification of unique peptides against complete UniProt database	GIYQTSNFR[a]	GFYAEGSR
		LQSLQTYVTQQLIR[a]	GQGVPINTNSSPDDQIGYYR
		VGGNYNYLYR	IGMEVTPSGTWLTYTGAIK

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388

Table 1. continued

reference	method	targets spike protein	targets nucleocapsid protein
		VQPTESIVR	NPANNAAIVLQLPQGTTLPK
		VVVLSFELLHAPATVCGPK	QQTVTLLPAADLDDFSK
		VYSTGSNVFQTR[a]	

[a]Unique/most suitable for SARS-CoV-2 diagnostics according to publication/preprint.

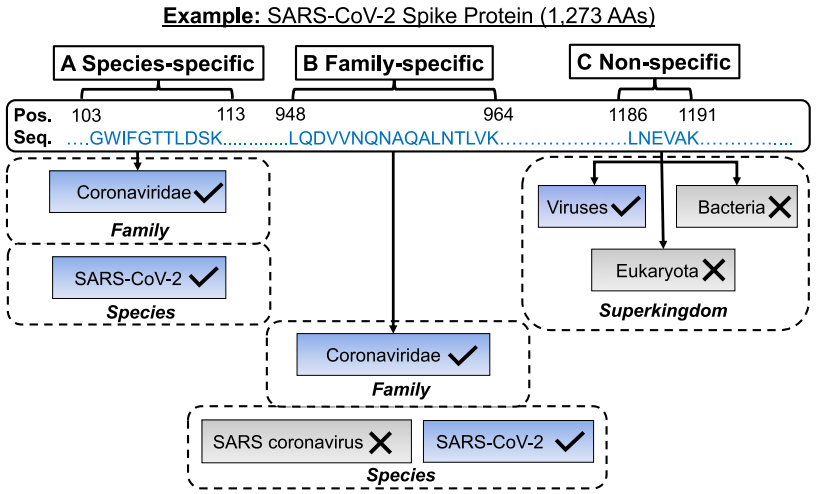


Figure 2. Identification of unique SARS-CoV-2 peptides on family and species levels. The spike protein is used as an example to illustrate the different levels of uniqueness.

abundant in virus particles, they are probably most suitable for MS-based detection, which can be seen in the SARS-CoV-2 nucleocapsid protein as the favored target for detection.[8,9,12] In contrast with structural proteins, nonstructural viral proteins may be actively secreted from the infected cell, making them potentially more detectable without cell lysis. As is known for the secreted DENV NS1 protein, viral proteins can be detectable in the sample for a longer time than the viral genome,[25] but structural protein amounts might correlate better with intact virus particle amounts in the sample. For example, it has been shown that the detection of the structural p24 HIV antigen correlates with viral RNA concentrations.[40] Nevertheless, it seems conclusive that structural viral proteins could also be detectable for a longer time than the viral genome because of damaged virus particles, proteins released from virus-induced cell lysis, and the overall protein stability.[20] In the end, the choice of which proteins to use for virus detection has to be evaluated based on the detectable peptides in real sample material.

A critical point to consider in MS-based virus detection is the identification of viral peptides with good stability and detectability. Whereas it is advantageous to choose highly abundant viral peptides that give rise to high signal intensities, peptide uniqueness is of equal importance. This is because virus peptides can be shared with other species, eukaryotes, or bacteria;[41] moreover, closely related viruses can considerably differ in their pathogenicity and virulence (the degree of pathogenicity). For example, *Zaire ebolavirus* is highly virulent, whereas the related *Reston ebolavirus* species is not pathogenic for humans.[42] In the case of SARS-CoV-2, for example, closely related coronaviruses could lead to false-positive results if peptide uniqueness is not considered.[41] Virus diagnostics by real-time PCR are generally done on the species level or, in the case of screening assays, on the family or genus level. For some viruses, like the influenza virus or DENV, it makes sense to further differentiate genotypes and serotypes. Nevertheless, for some viruses, it might not be feasible to detect conserved unique peptides on the species level. In this case, one could combine different peptides to cover all variants, for example, as was done for HIV detection by MRM.[31] The strategies that have been applied for viral target peptide identification often involve the data-dependent analysis (DDA) of cell-culture-derived viruses.[11,25,30−32] However, this poses some limitations because not all viruses can be grown in a cell culture, and safety measures have to be taken for highly pathogenic viruses. For example, for the propagation of SARS-CoV-2, a biosafety level 3 laboratory is required according to the WHO laboratory biosafety guidance related to COVID-19.[43] Additionally, it should be considered that viruses adapt in cell culture, potentially leading to proteomic changes.[44] The solution can be to identify target peptides *in silico*, as was recently suggested in a preprint from Orsburn and colleagues.[9] In a rigorous selection process, they identified unique tryptic peptides of SARS-CoV-2 that can potentially be used for detection. As stated by Bauer and colleagues,[45] it is important here to select references and databases that can cover the dynamic nature of a pandemic and include newly emerging mutations (e.g., D614G mutation in the spike protein of SARS-CoV-2[46]) that are not present in the first sequenced/available samples. The challenge is still the assay characterization that cannot be done *in silico*. For example, according to the Clinical Proteomic Tumor Analysis Consortium (CPTAC) assay characterization guidance, a "fit-for-purpose" diagnostics targeted proteomics assay has to be characterized by multiple experiments, like response curve, repeatability, selectivity, stability, and reproducible detection of the endogenous analyte.[47] A selection of strategies applied for SARS-CoV-2 target peptide identification for PRM experiments is shown in Table 1. The strategies are highly diverse and range from no consideration of peptide uniqueness[7] to a multistep selection process.[9] Nevertheless, the

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388

nucleocapsid protein of SARS-CoV-2 is the favored target protein as a result of its abundance. In the context of the previously mentioned publications and those in Table 1, we implemented a conservative *in silico* approach to select suited SARS-CoV-2 tryptic peptides with a focus on conserved and unique sequences in the scope of all biological kingdoms. As a data basis, the proteomes of 17 937 isolates were available on GISAID[48] on May 12, 2020. These were compared against the UniProtKB[29] April 22, 2020 release using Purple in version 0.4.2 with a homology threshold of 80%. Purple utilizes the SARS-CoV-2 proteomes as a target and the reviewed and unreviewed sequences of the UniProtKB as background. It selects unique tryptic target peptides that are not present in the background. An overview of the different levels of uniqueness of SARS-CoV-2 peptides originating in the spike protein is shown in Figure 2. With 1877 peptides provided by Purple, we calculated the relative occurrences of the peptides in the isolates. Peptides with a length of 7 to 30 amino acids from the spike and nucleocapsid protein that are present in 95% of the isolates were extracted. These 10 peptides of the spike protein and 1 peptide of the nucleocapsid protein are listed in Table 1. A closer look at the collection of peptides in Table 1 shows that some peptides are present independent of the chosen method, for example, the peptide ADETQALPQR. This peptide belonging to the nucleoprotein has also been identified by DDA as a potential SARS-CoV-2 target with high abundance in purified virus preparations and nasopharyngeal swab samples.[10] It should be further highlighted that for clinical diagnostics a database tailored for clinical use cases makes more sense than including the whole UniProtKB into the analysis. This decreases the needed computational power and would have no negative effect on the false-positive rate in the real-world application of clinical diagnostics. But still, standardized methods and workflows for viral target peptide identification have to be implemented in the future. It is important to have an understanding of how to integrate the various levels of uniqueness into the development of these new methods.

Sample Preparation

Viruses are categorized into four risk groups according to pathogenicity, transmissibility, and availability of countermeasures. Depending on the risk group, appropriate safety measures have to be taken. Hence, for personal protection, virus inactivation has to be guaranteed during proteomics sample preparation. In real-time PCR diagnostics, the first step is the inactivation by sample lysis, including cells and viruses. Lysis is also generally the first step during sample preparation for proteomics. Enveloped viruses are generally easier to inactivate than nonenveloped viruses because the lipid-based envelope is more fragile. Because the inactivation of viruses depends on the virion composition and sample material, it can be assumed that not all lysis buffers used in proteomics are able to inactivate viruses. Hence, inactivation has to be validated for the respective virus to be analyzed. In the case of untargeted analysis, inactivation should be validated for multiple viruses covering different virion properties. Inactivation efficiency further depends on the sample material itself; for example, virions in a nasopharyngeal swab are probably easier to lyse than virions in a biopsy-derived piece of tissue. A good starting point for virus inactivation is heating the sample in ≥1% sodium dodecyl sulfate (SDS) for 5 min at 95 °C, for example, as is done in the FASP[49] and STrap protocol.[50] Heating the sample in SDS is a validated method for the inactivation of viruses up to risk group four, for example, ebola virus.[51] The drawback of this method is that SDS has to be removed prior to MS analysis, which considerably prolongs the sample processing. A novel sample preparation technique called SPEED uses pure trifluoroacetic acid (TFA) for lysis, making detergent removal redundant, which streamlines the sample processing.[52] Moreover, for some viruses, inactivation can be achieved by heating the sample prior to MS sample preparation, for example, to 65 °C for 30 min for the inactivation of SARS-CoV-2.[12] However, there is no study that has systematically investigated virus inactivation by proteomics sample preparation methods, which is why this has to be evaluated in the future.

Throughput

One of the biggest challenges for proteomics to become a competitive method for virus detection is sample throughput. Even the most advanced LC-MS-based proteomics platforms are only able to analyze up to 200 samples per day (Scanning Swath, EvoSep). In comparison, a conventional real-time PCR cycler can analyze between 96 and 1536 samples in 60−90 min and costs <10% of a high-end mass spectrometer. Recently, it was shown that multiplexing of LC pumps can increase the sample throughput of proteomics for SARS-CoV-2 detection to ∼500 samples per day.[11] This study represents an enormous advancement because it enables almost 100% MS utilization. However, the throughput of the system still lags behind PCR-based diagnostics. Theoretically, MALDI-ToF MS provides the throughput needed for the targeted detection of a virus, even in a pandemic situation. But again, its potential remains elusive, although proof-of-principle studies have been published as preprints.[53,54]

Detection Methods

In MRM, multiple fragment ions resulting from a single peptide precursor are detected in a triple quadruple mass spectrometer. This technique is well established for routine applications of LC-MS due to its sensitivity and robustness but has rarely been used to detect viruses in clinical samples.[30] PRM assays, on the contrary, are easier to establish because all fragment ions are recorded simultaneously and therefore do not have to be preselected. The potential of PRM for the detection of virus peptides in clinical samples has been demonstrated for DENV in serum samples[25] and SARS-CoV-2 from respiratory specimens.[8,11,32] PRM on Orbitrap-hybrid mass spectrometers has sensitivities in the low attomolar range, whereas the selectivity is ensured by high resolution and accurate mass. The number of peptides and hence the number of viruses that can be analyzed in a single PRM run is limited and depends on the chromatographic peak width, the cycle time, and the transient length. However, the multiplexing of up to 1000 targets in a single run has recently been shown.[55] This multiplexing capacity is theoretically sufficient to enable the detection of all human-pathogenic virus species in a single run. Nevertheless, with prior knowledge of the symptoms (e.g., respiratory symptoms), one can reduce the number of targeted viruses (e.g., targeting only respiratory viruses). Another disadvantage of (scheduled) PRM is the retention time dependence. Retention times can vary with chromatographic performance and also with the sample background. In the worst case, peptides shift out of the retention time window and are no longer detectable. The use of stable isotope-labeled (SIL) standard peptides as an internal control can avoid these

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388

false-negatives.[32] However, it has to be noted that the use of SIL peptides for virus detection could also lead to false-positive results, as non- or low-purified heavy-labeled peptides can contain low levels of their unlabeled counterparts.[56]

Some of the drawbacks of PRM could be overcome by data-independent acquisition (DIA) in the future. DIA measurements record all peptide fragments above the sensitivity threshold of the mass spectrometer using predefined isolation windows. DIA is well suited for the reliable detection of peptides in large sample cohorts because it circumvents the stochastic precursor sampling of DDA. DIA is not restricted to a certain number of targets and is less influenced by retention time shifts than PRM. Recent developments enable the generation of *in silico* peptide libraries[57,58] and the measurement of proteomes using short gradients at increasing depth.[59,60] It is important to understand that PRM and DIA were mainly developed for protein quantification in large sample cohorts. However, for virus detection, accurate and precise quantification is much less important than it is for most other proteomic applications. This opens up possibilities for method optimization to improve the identification of virus proteins. For example, it has been shown that decreasing the number of data points per peak in DIA improves the peptide identification while negatively affecting the quantification at a certain point.[61,62]

Run-to-Run Contaminations

Peptide carryover between subsequent runs is a well known phenomenon in LC-MS/MS and presents a huge challenge for virus detection.[11,63] The dynamic range of viral proteins in clinical samples can span the whole range of the proteome. Subsequent LC-MS/MS measurements after samples with high viral loads are therefore prone to lead to false-positive results because it is very challenging to obtain no signal for a high-intensity peptide in the following run.[11] In contrast, MALDI-ToF MS has the advantage of being free of carryover between samples because of the lack of a separation system. In many proteomic applications, a low level of carryover can be tolerated because it only slightly affects the quantification accuracy and precision. However, in virus diagnostics, all detectable levels of carryover are highly problematic. To exclude contamination across runs in LC-MS/MS experiments, one can run a blank or a wash run between clinical samples.[11] However, this impairs the sample throughput. Promising results for a similar application were shown using an Evosep One LC system, which uses a disposable StageTip for sample loading, for the detection of antibiotic resistances. The authors performed PRM measurements on bacterial isolates and claimed that no peptide carryover was observed for any targeted peptide.[64]

4. CONCLUSIONS

Currently, MS might seem to be an attractive alternative to PCR-based methods, for example, when PCR reagents are getting scarce or when testing capacities have to be rapidly expanded. Furthermore, proteomics would add another level of confirmation to complement or, in some cases, even outperform real-time PCR results. However, there are still some major hurdles to clear before MS can be applied in routine virus diagnostics concerning, for example, throughput, sensitivity, and data analysis. Finally, whereas further technical developments pave the potential way for MS to become an alternative method for virus detection in the clinic, standardization procedures within the community should play a major role.

AUTHOR INFORMATION

Corresponding Author

Thilo Muth − *Section eScience (S.3), Federal Institute for Materials Research and Testing, Berlin 12205, Germany;* orcid.org/0000-0001-8304-2684; Email: Thilo.Muth@bam.de

Authors

Marica Grossegesse − *Centre for Biological Threats and Special Pathogens, Highly Pathogenic Viruses (ZBS 1), Robert Koch Institute, Berlin 13353, Germany;* orcid.org/0000-0002-9369-8203

Felix Hartkopf − *Microbial Genomics (NG 1), Robert Koch Institute, Berlin 13353, Germany; Section eScience (S.3), Federal Institute for Materials Research and Testing, Berlin 12205, Germany*

Andreas Nitsche − *Centre for Biological Threats and Special Pathogens, Highly Pathogenic Viruses (ZBS 1), Robert Koch Institute, Berlin 13353, Germany*

Lars Schaade − *Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin 13353, Germany*

Joerg Doellinger − *Centre for Biological Threats and Special Pathogens, Proteomics and Spectroscopy (ZBS 6), Robert Koch Institute, Berlin 13353, Germany;* orcid.org/0000-0001-8309-082X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00674

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Ursula Erikli for copy editing. We acknowledge the omics.pnl.gov website (https://omics.pnl.gov/software/protein-digestion-simulator). This work was supported by the German Research Foundation (DFG) (MU 4430/2-1 to T.M.).

ABBREVIATIONS

PRM, parallel reaction monitoring; DDA, data-dependent acquisition; DIA, data-independent acquisition; MRM, multiple reaction monitoring

REFERENCES

(1) Nomura, F. Proteome-based bacterial identification using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS): A revolutionary shift in clinical diagnostic microbiology. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854* (6), 528−37.

(2) Schubert, S.; Kostrzewa, M. MALDI-TOF MS in the Microbiology Laboratory: Current Trends. *Curr. Issues Mol. Biol.* **2017**, *23*, 17−20.

(3) Seng, P.; et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* **2009**, *49* (4), 543−51.

(4) Welker, M.; et al. An update on the routine application of MALDI-TOF MS in clinical microbiology. *Expert Rev. Proteomics* **2019**, *16* (8), 695−710.

(5) Ecker, D. J.; et al. Ibis T5000: a universal biosensor approach for microbiology. *Nat. Rev. Microbiol.* **2008**, *6* (7), 553−8.
(6) Doellinger, J.; et al. Comparison of real-time PCR and MassTag PCR for the multiplex detection of highly pathogenic agents. *Mol. Cell. Probes* **2012**, *26* (5), 177−81.
(7) Bezstarosti, K.; Lamers, M. M.; Haagmans, M. L.; Demmers, J. A. A. Targeted Proteomics for the Detection of SARS-CoV-2 Proteins. *bioRxiv* **2020**, DOI: 10.1101/2020.04.23.057810.
(8) Ihling, C.; Tanzler, D.; Hagemann, S.; Kehlen, A.; Huttelmaier, S.; Arlt, C.; Sinz, A. Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *J. Proteome Res.* **2020**, DOI: 10.1021/acs.jproteome.0c00280.
(9) Orsburn, B. C.; Jenkins, C.; Miller, S. M.; Neely, B. A.; Bumpus, N. N. In silico approach toward the identification of unique peptides from viral protein infection: Application to COVID-19. *bioRxiv* **2020**, DOI: 10.1101/2020.03.08.980383.
(10) Gouveia, D.; Miotello, G.; Gallais, F.; Gaillard, J.-C.; Debroas, S.; Bellanger, L.; Lavigne, J.-P.; Sotto, A.; Grenga, L.; Pible, O. Proteotyping SARS-CoV-2 Virus from Nasopharyngeal Swabs: A Proof-of-Concept Focused on a 3 min Mass Spectrometry Window. *J. Proteome Res.* **2020**, DOI: 10.1021/acs.jproteome.0c00535.
(11) Cardozo, K. H. M.; Lebkuchen, A.; Okai, G. C.; Schuch, R. A.; Viana, L. G.; Olive, A. N.; Lazari, C. d. S.; Fraga, A. M.; Granato, C.; Carvalho, V. M. Fast and low-cost detection of SARS-CoV-2 peptides by tandem mass spectrometry in clinical samples. *Research Square* **2020**, DOI: 10.21203/rs.3.rs-28883/v1.
(12) Nikolaev, E. N.; Indeykina, M. I.; Brzhozovskiy, A. G.; Bugrova, A. E.; Kononikhin, A. S.; Starodubtseva, N. L.; Petrotchenko, E. V.; Kovalev, G. I.; Borchers, C. H.; Sukhikh, G. T. Mass-Spectrometric Detection of SARS-CoV-2 Virus in Scrapings of the Epithelium of the Nasopharynx of Infected Patients via Nucleocapsid N Protein. *J. Proteome Res.* **2020**, DOI: 10.1021/acs.jproteome.0c00412.
(13) Nachtigall, F. M.; Pereira, A.; Trofymchuk, O. S.; Santos, L. S. Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. *Nat. Biotechnol.* **2020**, *38*, 1168.
(14) Brown, J. R.; Bharucha, T.; Breuer, J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. *J. Infect.* **2018**, *76* (3), 225−240.
(15) Brinkmann, A.; Andrusch, A.; Belka, A.; Wylezich, C.; Hoper, D.; Pohlmann, A.; Nordahl Petersen, T.; Lucas, P.; Blanchard, Y.; Papa, A.; et al. Proficiency Testing of Virus Diagnostics Based on Bioinformatics Analysis of Simulated In Silico High-Throughput Sequencing Data Sets. *J. Clin. Microbiol.* **2019**, *57* (8), e00466−19.
(16) Brinkmann, A.; et al. Development and preliminary evaluation of a multiplexed amplification and next generation sequencing method for viral hemorrhagic fever diagnostics. *PLoS Neglected Trop. Dis.* **2017**, *11* (11), No. e0006075.
(17) van der Heide, V. SARS-CoV-2 cross-reactivity in healthy donors. *Nat. Rev. Immunol.* **2020**, *20* (7), 408.
(18) Cohen, M. S.; Gay, C. L.; Busch, M. P.; Hecht, F. M. The detection of acute HIV infection. *J. Infect. Dis.* **2010**, *202* (S2), S270−S277.
(19) Chuang, Y. C.; et al. Re-evaluation of the pathogenic roles of nonstructural protein 1 and its antibodies during dengue virus infection. *J. Biomed. Sci.* **2013**, *20*, 42.
(20) Buckley, M. Paleoproteomics: An Introduction to the Analysis of Ancient Proteins by Soft Ionisation Mass Spectrometry. *Paleogenomics* **2018**, 31−52.
(21) Corman, V. M; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K.; Bleicker, T.; Brunink, S.; Schneider, J.; Schmidt, M. L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* **2020**, *25* (3), 2000045.
(22) Benevento, M.; et al. Adenovirus composition, proteolysis, and disassembly studied by in-depth qualitative and quantitative proteomics. *J. Biol. Chem.* **2014**, *289* (16), 11421−30.
(23) Hutchinson, E. C.; et al. Conserved and host-specific features of influenza virion architecture. *Nat. Commun.* **2014**, *5*, 4816.
(24) Doellinger, J.; Schaade, L.; Nitsche, A. Comparison of the Cowpox Virus and Vaccinia Virus Mature Virion Proteome: Analysis of the Species- and Strain-Specific Proteome. *PLoS One* **2015**, *10* (11), No. e0141527.
(25) Wee, S.; et al. Multiplex targeted mass spectrometry assay for one-shot flavivirus diagnosis. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (14), 6754−6759.
(26) European Commission. Commission Decision of 3 February 2009 amending Decision 2002/364/EC on common technical specifications for in vitro-diagnostic medical devices (notified under document number C(2009) 565); 2009/108/EC; 2009.
(27) Geyer, P. E.; et al. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **2017**, *13* (9), 942.
(28) Renuse, S.; Vanderboom, P. M.; Maus, A. D.; Kemp, J. V.; Gurtner, K. M.; Madugundu, A. K.; Chavan, S.; Peterson, J. A.; Madden, B. J.; Mangalaparthi, K. K.; et al. Development of mass spectrometry-based targeted assay for direct detection of novel SARS-CoV-2 coronavirus from clinical specimens. *medRxiv* **2020**, DOI: 10.1101/2020.08.05.20168948.
(29) The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2018**, *46* (5), 2699.
(30) Foster, M. W.; et al. Targeted Proteomics of Human Metapneumovirus in Clinical Samples and Viral Cultures. *Anal. Chem.* **2015**, *87* (20), 10247−54.
(31) Schlatzer, D.; et al. A Targeted Mass Spectrometry Assay for Detection of HIV Gag Protein Following Induction of Latent Viral Reservoirs. *Anal. Chem.* **2017**, *89* (10), 5325−5332.
(32) Zecha, J.; et al. Data, reagents, assays and merits of proteomics for SARS-CoV-2 research and testing. *Mol. Cell. Proteomics* **2020**, *19*, 1503.
(33) Kuhring, M.; et al. TaxIt: An Iterative Computational Pipeline for Untargeted Strain-Level Identification Using MS/MS Spectra from Pathogenic Single-Organism Samples. *J. Proteome Res.* **2020**, *19* (6), 2501−2510.
(34) MacLean, B.; et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966−8.
(35) Zauber, H.; Kirchner, M.; Selbach, M. Picky: a simple online PRM and SRM method designer for targeted proteomics. *Nat. Methods* **2018**, *15* (3), 156−157.
(36) Zolg, D. P.; et al. Building ProteomeTools based on a complete synthetic human proteome. *Nat. Methods* **2017**, *14* (3), 259−262.
(37) Mohammed, Y.; et al. PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J. Proteomics* **2014**, *106*, 151−61.
(38) Mesuere, B.; et al. The unique peptidome: Taxon-specific tryptic peptides as biomarkers for targeted metaproteomics. *Proteomics* **2016**, *16* (17), 2313−8.
(39) Lechner, J.; et al. Purple: A Computational Workflow for Strategic Selection of Peptides for Viral Diagnostics Using MS-Based Targeted Proteomics. *Viruses* **2019**, *11* (6), 536.
(40) Brown, A. E.; et al. Quantitative relationship of circulating p24 antigen with human immunodeficiency virus (HIV) RNA and specific antibody in HIV-infected subjects receiving antiretroviral therapy. The RV43 Study Group. *J. Infect. Dis.* **1995**, *172* (4), 1091−5.
(41) Gouveia, D.; Grenga, L.; Gaillard, J.-C.; Gallais, F.; Bellanger, L.; Pible, O.; Armengaud, J. Shortlisting SARS-CoV-2 Peptides for Targeted Studies from Experimental Data-Dependent Acquisition Tandem Mass Spectrometry Data. *Proteomics* **2020**, *20* (14), No. 2000107.
(42) Hoenen, T.; et al. Ebola virus: unravelling pathogenesis to combat a deadly disease. *Trends Mol. Med.* **2006**, *12* (5), 206−15.
(43) WHO. Laboratory Biosafety Guidance Related to Coronavirus Disease 2019 (COVID-19), 2020. https://apps.who.int/iris/bitstream/handle/10665/331138/WHO-WPE-GIH-2020.1-eng.pdf (accessed 13.06.2020).
(44) Grossegesse, M.; et al. Combined Proteomics/Genomics Approach Reveals Proteomic Changes of Mature Virions as a Novel Poxvirus Adaptation Mechanism. *Viruses* **2017**, *9* (11), 337.
(45) Bauer, D. C.; et al. Supporting pandemic response using genomics and bioinformatics: A case study on the emergent SARS-

CoV-2 outbreak. *Transboundary Emerging Dis.* **2020**, *67* (4), 1453−1462.
(46) Daniloski, Z.; Guo, X.; Sanjana, N.E. The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. *bioRxiv* **2020**, DOI: 10.1101/2020.06.14.151357.
(47) Carr, S. A.; et al. Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **2014**, *13* (3), 907−17.
(48) Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **2017**, *1* (1), 33−46.
(49) Wisniewski, J. R.; et al. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6* (5), 359−62.
(50) Zougman, A.; Selby, P. J.; Banks, R. E. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics* **2014**, *14* (9), 1006−10.
(51) Haddock, E.; Feldmann, F.; Feldmann, H. Effective Chemical Inactivation of Ebola Virus. *Emerging Infect. Dis.* **2016**, *22* (7), 1292−4.
(52) Doellinger, J.; et al. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteomics* **2020**, *19* (1), 209−222.
(53) Iles, R. K.; Zmuidinaite, R.; Iles, J. K.; Carnell, G.; Sampson, A.; Heeney, J. L. A clinical MALDI-ToF Mass spectrometry assay for SARS-CoV-2: Rational design and multi-disciplinary team work. *medRxiv* **2020**, DOI: 10.1101/2020.08.22.20176669.
(54) Rocca, M. F.; Zintgraff, J. C.; Dattero, M. E.; Santos, L. S.; Ledesma, M.; Vay, C.; Prieto, M.; Benedetti, E.; Avaro, M.; Russo, M.; et al. A Combined approach of MALDI-TOF Mass Spectrometry and multivariate analysis as a potential tool for the detection of SARS-CoV-2 virus in nasopharyngeal swabs. *bioRxiv* **2020**, DOI: 10.1101/2020.05.07.082925.
(55) Grossegesse, M.; et al. Stable Isotope-Triggered Offset Fragmentation Allows Massively Multiplexed Target Profiling on Quadrupole-Orbitrap Mass Spectrometers. *J. Proteome Res.* **2020**, *19* (7), 2854−2862.
(56) Hoofnagle, A. N.; et al. Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry-Based Assays. *Clin. Chem.* **2016**, *62* (1), 48−69.
(57) Xu, R.; Sheng, J.; Bai, M.; Shu, K.; Zhu, Y.; Chang, C. A Comprehensive Evaluation of MS/MS Spectrum Prediction Tools for Shotgun Proteomics. *Proteomics* **2020**, 1900345.
(58) Gessulat, S.; et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16* (6), 509−518.
(59) Messner, C. B.; Demichev, V.; Bloomfield, N.; White, M.; Kreidl, M.; Ivosev, G.; Wasim, F.; Zelezniak, A.; Lilley, K. S.; Tate, S. Scanning SWATH acquisition enables high-throughput proteomics with chromatographic gradients as fast as 30 seconds. *bioRxiv* **2020**.
(60) Bekker-Jensen, D. B.; et al. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol. Cell. Proteomics* **2020**, *19* (4), 716−729.
(61) Doellinger, J.; Blumenscheit, C.; Schneider, A.; Lasch, P. Optimization of data-independent acquisition using predicted libraries for deep and accurate proteome profiling. *bioRxiv* **2020**, DOI: 10.1101/2020.03.02.972570.
(62) Bruderer, R.; et al. Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* **2017**, *16* (12), 2296−2309.
(63) Dolman, S.; et al. Investigation of carryover of peptides in nano-liquid chromatography/mass spectrometry using packed and monolithic capillary columns. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2013**, *912*, 56−63.
(64) Foudraine, D. E.; Dekker, L. J. M.; Strepis, N.; Bexkens, M. L.; Klaassen, C. H. W.; Luider, T. M.; Goessens, W. H. F. Accurate Detection of the Four Most Prevalent Carbapenemases in E. coli and K. pneumoniae by High-Resolution Mass Spectrometry. *Front. Microbiol.* **2019**, *10*, 2760.

https://dx.doi.org/10.1021/acs.jproteome.0c00674
J. Proteome Res. 2020, 19, 4380−4388