Received: 22 August 2020 | Revised: 8 February 2021 | Accepted: 17 February 2021 | Accepted article online: 22 February 2021

DOI: 10.1002/pmic.202000226

Proteomics
Proteomics and Systems Biology

TECHNICAL BRIEF

Application of spectral library prediction for parallel reaction monitoring of viral peptides

Marica Grossegesse[1] | Andreas Nitsche[1] | Lars Schaade[2] | Joerg Doellinger[1,3]

[1] Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Highly Pathogenic Viruses (ZBS 1), Berlin, Germany

[2] Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin, Germany

[3] Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Proteomics and Spectroscopy (ZBS 6), Berlin, Germany

Correspondence
Marica Grossegesse, Highly Pathogenic Viruses (ZBS 1), Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Seestraße 10, 13353 Berlin, Germany.
Email: grossegessem@rki.de

*Correction added on 5 April 2021, after first online publication: DEAL statement.

Abstract

A major part of the analysis of parallel reaction monitoring (PRM) data is the comparison of observed fragment ion intensities to a library spectrum. Classically, these libraries are generated by data-dependent acquisition (DDA). Here, we test Prosit, a published deep neural network algorithm, for its applicability in predicting spectral libraries for PRM. For this purpose, we targeted 1529 precursors derived from synthetic viral peptides and analyzed the data with Prosit and DDA-derived libraries. Viral peptides were chosen as an example, because virology is an area where in silico library generation could significantly improve PRM assay design. With both libraries a total of 1174 precursors were identified. Notably, compared to the DDA-derived library, we could identify 101 more precursors by using the Prosit-derived library. Additionally, we show that Prosit can be applied to predict tandem mass spectra of synthetic viral peptides with different collision energies. Finally, we used a spectral library predicted by Prosit and a DDA library to identify SARS-CoV-2 peptides from a simulated oropharyngeal swab demonstrating that both libraries are suited for peptide identification by PRM. Summarized, Prosit-derived viral spectral libraries predicted in silico can be used for PRM data analysis, making DDA analysis for library generation partially redundant in the future.

KEYWORDS
parallel reaction monitoring, PRM, SARS-CoV-2, tandem mass spectra prediction, virus proteomics

Parallel reaction monitoring (PRM) is a mass spectrometry method that is applied for sensitive and selective target analysis. In PRM mode, the target peptide is filtered by the quadrupole, fragmented, for example, with higher energy collisional dissociation, and the fragment ions are detected in a high-resolution mass analyzer, such as an orbitrap. PRM data is most often analyzed in the Skyline software environment, which has been developed for targeted mass spectrometry data analysis [1]. For correct peptide identification, a combination of multiple parameters is used. While heavy labeled peptides co-eluting with the peptide of interest are the gold standard, retention time, precursor mass, and matching of fragment ions to a spectral library are commonly applied in order to identify unambiguously the peptide of interest. The similarity of the measured fragment spectrum to a library spectrum is calculated by normalized spectral contrast angle and called dotp value in Skyline. Until now, spectra for spectral libraries had to be generated in separate analysis prior to the actual PRM run, using data-dependent analysis (DDA). However, recently the in

Abbreviations: PRM, parallel reaction monitoring; DDA, data-dependent acquisition; NCE, normalized collision energy; RT, retention time

silico generation of spectral libraries has been introduced by different tandem mass spectra prediction algorithms [2–7]. In silico prediction of spectral libraries can be an important tool for targeted MS such as PRM, especially when library generation by DDA is not straightforward. Another advantage of in silico-generated libraries is that they do not have an inherent detection bias like DDA that can lead to the missing of target peptides. As a result of the SARS-CoV-2 pandemic, the analysis of viruses by MS is becoming increasingly popular [8,9]. In virology, tandem mass spectra prediction could be a valuable tool, since generation of viral spectral libraries is still challenging. This is due to the safety measures needed when working with certain viruses, the low abundance of viruses in the host background matrix or simply the unavailability of certain viruses. In this study, we chose to compare spectral libraries generated by DDA and predicted in silico by Prosit for their applicability to virus-targeted PRM. Prosit is a freely available (https://www.proteomicsdb.org/prosit/) tandem mass spectra prediction algorithm to generate spectral libraries compatible with Skyline [2]. The Prosit neural network algorithm has been trained on publicly available datasets from Proteome Tools [10] consisting of human peptide data. It has been shown that Prosit can be used for metaproteomic data analysis, suggesting its applicability to viral peptide spectra prediction. Moreover, Prosit has been shown to outperform other tools in predicting spectra for designated mass spectrometers like the Orbitrap Fusion and Lumos [11]. However, so far it has remained elusive how Prosit performs in virus-targeted PRM data analysis in comparison to classical DDA libraries, which is the main focus of this study.

For the first two experiments, we used synthetic peptides of diverse highly pathogenic viruses, namely Ebola virus, Marburg virus, Lassa virus, SARS coronavirus, variola virus, Crimean–Congo hemorrhagic fever virus, Machupo virus, and Eastern, Western, and Venezuelan equine encephalitis virus. All of these viruses have to be handled in high-containment laboratories with biosafety level 3 or 4. Without prior knowledge of the peptides' detectability by nLC-MS/MS, we ordered a total of 1569 crude synthetic viral peptides in six pools from JPT (Berlin, Germany). All peptides are specific for the respective virus. The peptide pools were analyzed in triplicate by DDA on an Orbitrap Q Exactive Plus, using a top 10 method. Raw files were searched with MaxQuant [12] against the respective virus database (UniProt) with a peptide FDR of 1%. Detailed MS analysis and MaxQuant parameters can be found in the supporting information and the data uploaded to ProteomeXchange Consortium. MaxQuant .msms output files were used to generate spectral libraries with BiblioSpec implemented in the Skyline environment. BiblioSpec uses the best-scored peptides for library generation. In the case of multiple spectra with identical scores, the one with the highest total ion current is selected. The respective DDA spectral libraries derived from the six peptide pools contained 1529 precursors belonging to 1026 peptides. In the following, we will refer to the DDA libraries as a single library although a library for each peptide pool was used. In the next step, these precursors were analyzed by PRM to ensure that every targeted peptide could theoretically be detected with the DDA spectral library. For this purpose, the 1529 precursors were divided into 38 unscheduled PRM runs with 40 precursors per run and a single run containing the remaining nine peptides. For PRM runs, the six synthetic peptide pools were used in the same way as for DDA runs. Peptide identification of PRM runs was done in Skyline, using the top six fragment ions of the DDA spectral library or the according Prosit-derived library (Prosit_2020_intensity_model). Collision energy calibration for Prosit was done by using six single DDA raw files, revealing that a normalized collision energy (NCE) of 25 on our Orbitrap Q Exactive Plus corresponds to a Prosit collisional energy of 32. A total of 1174 precursors were identified by using both libraries with dotp values of 0.85 or higher (Figure 1A). To ensure that identical peaks were analyzed with DDA- and Prosit-derived libraries, we compared the precursor retention times showing a correlation coefficient of 1 (Figure 1B). Comparing the dotp values of the 1174 precursors revealed even higher dopt values using the library generated by Prosit compared to DDA. The median dotp value with the Prosit-derived library was 0.96, while the median dotp values with the DDA-derived library was 0.94 (Figure 1C). Notably, the peptides with a length of six amino acids showed the lowest correlation to the Prosit spectral library (Figure 1D). This observation is conclusive, because Prosit has been trained on peptides with a length of 7–30 amino acids. This demonstrates that under the given experimental conditions Prosit performs very well for the prediction of tandem mass spectra between 7 and 21 amino acids in length, but is less suited for peptides of six amino acids. This observation is underlined by 77 peptides that were exclusively identified with the DDA library (Figure 1A,E). For the most part the peptides that could not be identified with Prosit had a length of six amino acids. Since short peptides are not very selective, they are generally not suited as PRM targets. As shown further in Figure 1A, 178 precursors could be exclusively identified with the Prosit library. Moreover notably, by using the Prosit library overall more fragment ions were identified, especially more doubly charged ones (Figure 1F).

Demonstrating that Prosit can be used to predict synthetic viral peptides with a fixed NCE of 25, we next aimed to test its prediction capabilities for viral peptides analyzed with different collision energies. Different collision energies are particularly interesting for PRM, because they can be optimized for each precursor separately to ensure optimal fragmentation for identification. For this purpose, we selected 121 peptides with DDA library dotp value of $\geq$0.99 and at least four identified fragment ions. These 121 peptides were analyzed using PRM with five different NCEs (19, 22, 25, 28, and 31). Prosit spectral libraries adjusted for the respective NCEs were generated (Prosit_2020_intensity_model) and library matches were analyzed in Skyline. This resulted in a median dotp value between 0.93 and 0.95, depending on the NCE (Figure 2A). An example of the viral peptide spectra predicted at different NCEs is shown in Figure 2B. For the doubly charged peptide LTGSPCAAFIGDDNIVK at an NCE of 19, the doubly charged y13-ion has the highest intensity, while at an NCE of 25 or higher, the single charged y7-ion has the highest intensity in the spectrum. This data suggests that Prosit can be used for NCE optimization during PRM assay design. However, it has to be noted that Prosit calculates relative fragment intensities that do not allow a comparison of the sensitivity of different NCEs for target detection. This has to be considered when optimizing NCEs.

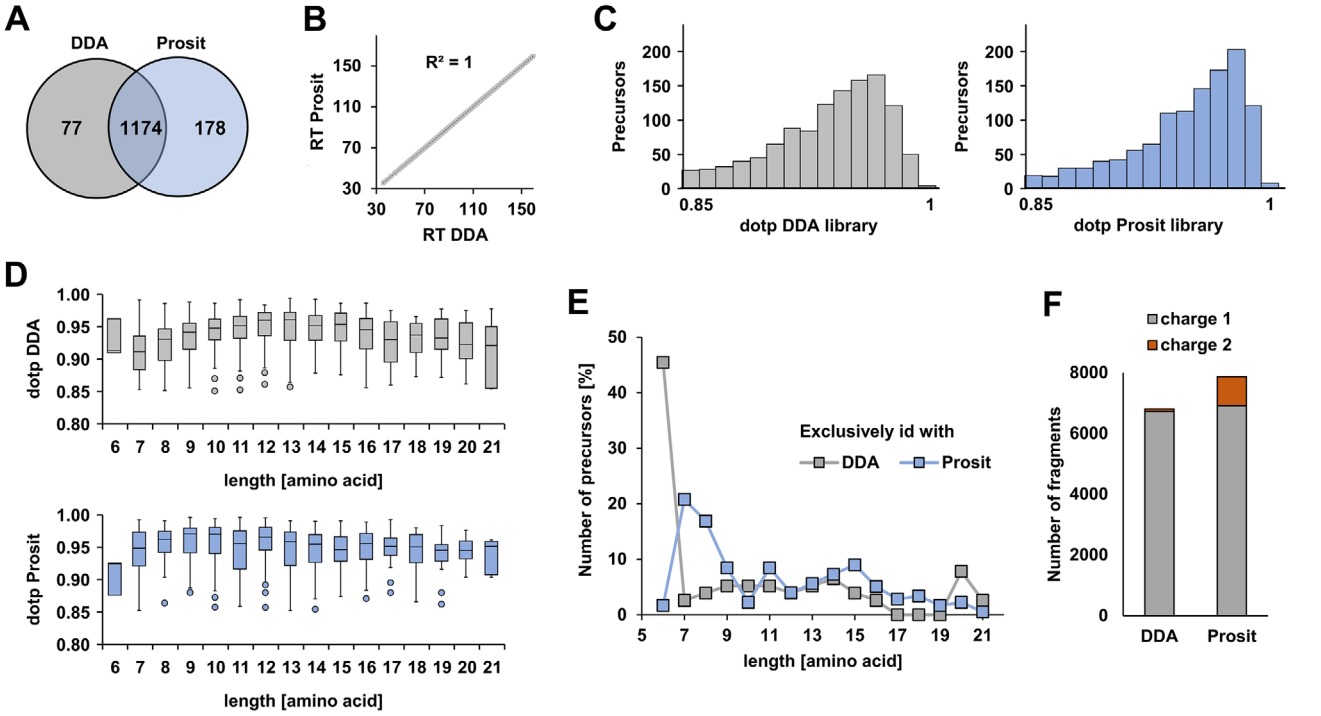

FIGURE 1 Comparison of viral peptide identifications by using spectral libraries generated by DDA and Prosit. A total of 1529 precursors belonging to 1026 synthetic viral peptides were analyzed by PRM, and the top six fragment ions were used for identification in Skyline. (A) Number of identified precursors with a minimum dotp value of 0.85. (B) Comparison of retention time (RT) of peptides identified with DDA and Prosit library. (C) Dotp values of 1174 precursors identified with DDA- and Prosit-derived library are shown in a histogram plot. (D) Dotp values of 1174 precursors identified with DDA- and Prosit-derived library sorted by peptide length. (E) Precursors identified exclusively with Prosit and DDA library sorted by peptide length. (F) Number of fragment ions with charge state one and two across all identified precursors

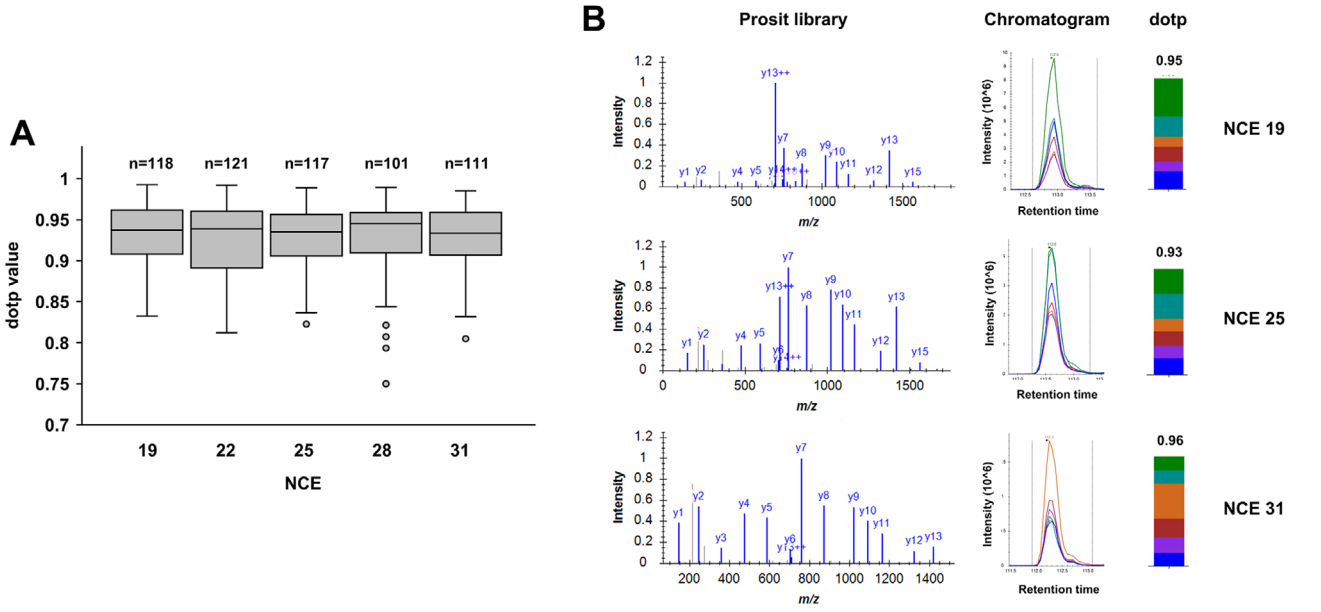

FIGURE 2 Performance of Prosit for the prediction of viral peptide tandem mass spectra at different collision energies. A total of 121 precursors were targeted in PRM mode at different normalized collision energies (NCE). Data was analyzed in Skyline with Prosit libraries for the respective NCE. (A) Boxplot of Prosit dotp values across different NCEs. Identification was done using the top six fragment ions. (B) Exemplarily, Prosit library spectra, chromatograms, and library matches (dotp values) for the doubly charged peptide LTGSPCAAFIGDDNIVK at different NCEs

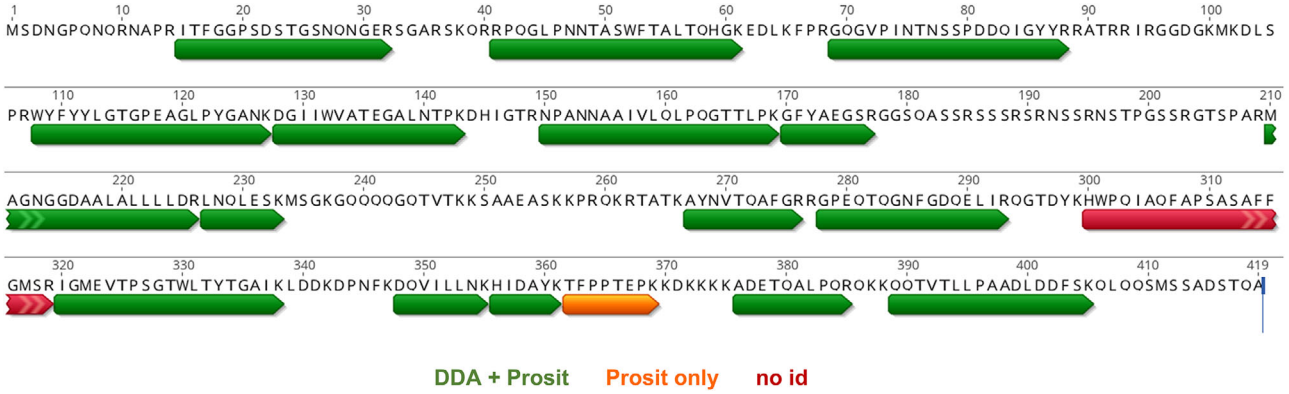

FIGURE 3 Detection of SARS-CoV-2 nucleoprotein peptides by PRM using DDA and Prosit-derived spectral libraries. Intact SARS-CoV-2 was spiked into a negative oropharyngeal swab to simulate a positive patient sample. A total of 18 viral peptides belonging to the nucleoprotein were targeted in a single PRM run and analyzed in Skyline using a DDA or Prosit-derived spectral library. Green: peptides detected using both libraries; orange: peptide detected exclusively with the Prosit library; red: peptide not detected. Protein sequence derived from UniProt (UniProtKB - P0DTC9 (NCAP_SARS2)

After showing that Prosit works well for synthetic viral peptides, we aimed to test it in a potential diagnostic PRM application example. Currently, as a result of the SARS-CoV-2 pandemic, MS-based proteomics is discussed as an alternative method for virus diagnostics. In particular PRM has been suggested as the method of choice to detect SARS-CoV-2 proteins from patient samples like naso- and oropharyngeal swabs [9,13-15,16,17]. To simulate a positive patient sample, we spiked cell culture-derived SARS-CoV-2 in a negative oropharyngeal swab to a final cT value of about 16. As it has been shown that the nucleoprotein (N protein) is the most abundant SARS-CoV-2 protein we used it as a target for detection by PRM [14]. Potential viral target peptides were identified by DDA of SARS-CoV-2 infected Calu-3 cells. The data was further used to generate a DDA library resulting in 27 precursors belonging to 18 peptides of the N protein. Retention times of these 18 peptides were identified by PRM of infected Calu-3 cells. The simulated positive swab sample was inactivated by addition of 20 µL 20% SDS to 200 µL sample volume and subsequent heating to 95°C for 10 min. Additionally to the N protein, we added 11 iRT peptides as a control resulting in 38 precursors targeted in a single PRM on an Orbitrap Q Exactive HF instrument using a 30 min gradient. Detailed sample preparation and LC-MS/MS parameters are described in the supporting information. The Prosit library was generated as described above. With both the DDA and the Prosit-derived library, a total of 16 out of 18 target peptides (21 out of 27 precursors) could be identified with a minimum dotp value of 0.91. The average dotp value using the DDA and the Prosit-derived library was 0.95 and 0.96, respectively. However, a single peptide (TFPPTEPK) could only be identified using the Prosit library (dotp 0.93), while using the DDA library the dotp value was 0.84 (Figure 3). These results demonstrate that Prosit also performs well when using shorter gradients (30 min), a more complex sample background (oropharyngeal swab) and intact virus particles instead of synthetic viral peptides.

Summarized, the identification of more than 1000 precursors derived from synthetic viral peptides and, moreover, 21 precursors derived from SARS-CoV-2 in an oropharyngeal swab demonstrates that Prosit can predict viral tandem mass spectra for PRM with at least equal performance compared to DDA-derived spectral libraries. When using Prosit, we could identify more fragment ions and, most notably, more doubly charged fragment ions compared to using DDA-derived libraries. Furthermore, we show that viral tandem mass spectra with different collision energies can be reliably predicted with Prosit. This study is the first large-scale application of Prosit—a spectral library prediction tool trained on human peptide data—for the generation of high-quality spectra of viral peptides for PRM data analysis. Nevertheless, from our data we conclude that Prosit can be routinely used to generate spectral libraries for viral PRM data analysis. In the future Prosit could facilitate PRM assay design, for example, in virus-targeted proteomics. Features like precursor intensity and charge state prediction would further facilitate assay design.

ACKNOWLEDGMENTS

The authors thank Ursula Erikli for copy editing.

Open access funding enabled and organized by Projekt DEAL.*

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data have been deposited to Zenodo (https://about.zenodo.org/) and are available under the following links:

https://doi.org/10.5281/zenodo.3988900 (DDA data)
https://doi.org/10.5281/zenodo.3995917 (PRM data)
https://doi.org/10.5281/zenodo.3988898 (PRM NCE data)
https://doi.org/10.5281/zenodo.4515285 (SARS-CoV-2_data)

ORCID

Marica Grossegesse https://orcid.org/0000-0002-9369-8203

REFERENCES

1. Maclean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., & Maccoss, M. J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, *26*(7), 966–968.
2. Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., Reimer, U., Ehrlich, H.-C., Aiche, S., Kuster, B., & Wilhelm, M. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, *16*(6), 509–518.
3. Degroeve, S., Maddelein, D., & Martens, L. (2015). MS2PIP prediction server: Compute and visualize MS2 peak intensity predictions for CID and HCD fragmentation. *Nucleic Acids Research*, *43*(W1), W326–W330.
4. Tiwary, S., Levy, R., Gutenbrunner, P., Salinas Soto, F., Palaniappan, K. K., Deming, L., Berndl, M., Brant, A., Cimermancic, P., & Cox, J. (2019). High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nature Methods*, *16*(6), 519–525.
5. Zhou, X.-X., Zeng, W.-F., Chi, H., Luo, C., Liu, C., Zhan, J., He, S.-M., & Zhang, Z. (2017). pDeep: Predicting MS/MS Spectra of peptides with deep learning. *Analytical Chemistry*, *89*(23), 12690–12697.
6. Guan, S., Moran, M. F., & Ma, B. (2019). Prediction of LC-MS/MS properties of peptides from sequence by deep learning. *Molecular and Cellular Proteomics*, *18*(10), 2099–2107.
7. Zeng, W.-F., Zhou, X.-X., Zhou, W.-J., Chi, H., Zhan, J., & He, S.-M. (2019). MS/MS spectrum prediction for modified peptides using pDeep2 trained by transfer learning. *Analytical Chemistry*, *91*(15), 9724–9731.
8. Ihling, C., Tänzler, D., Hagemann, S., Kehlen, A., Hüttelmaier, S., Arlt, C., & Sinz, A. (2020). Mass spectrometric identification of SARS-CoV-2 proteins from gargle solution samples of COVID-19 patients. *Journal of Proteome Research*, *19*(11), 4389–4392.
9. Zecha, J., Lee, C.-Y., Bayer, F. P., Meng, C., Grass, V., Zerweck, J., Schnatbaum, K., Michler, T., Pichlmair, A., Ludwig, C., & Kuster, B. (2020). Data, reagents, assays and merits of proteomics for SARS-CoV-2 research and testing. *Molecular and Cellular Proteomics*, *19*(9), 1503–1522.
10. Zolg, D. P., Wilhelm, M., Schnatbaum, K., Zerweck, J., Knaute, T., Delanghe, B., Bailey, D. J., Gessulat, S., Ehrlich, H.-C., Weininger, M., Yu, P., Schlegl, J., Kramer, K., Schmidt, T., Kusebauch, U., Deutsch, E. W., Aebersold, R., Moritz, R. L., Wenschuh, H. …, Kuster, B. (2017). Building ProteomeTools based on a complete synthetic human proteome. *Nature Methods*, *14*(3), 259–262.
11. Xu, R., Sheng, J., Bai, M., Shu, K., Zhu, Y., & Chang, C. (2020). A comprehensive evaluation of MS/MS spectrum prediction tools for shotgun proteomics. *Proteomics*, *20*(21-22), e1900345.
12. Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367–1372.
13. 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. (2020). Mass-spectrometric detection of SARS-CoV-2 virus in scrapings of the epithelium of the nasopharynx of infected patients via nucleocapsid N protein. *Journal of Proteome Research*, *19*(11), 4393–4397.
14. Cardozo, K. H. M., Lebkuchen, A., Okai, G. G., Schuch, R. A., Viana, L. G., Olive, A. N., Lazari, C. D. S., Fraga, A. M., Granato, C. F. H., Pintão, M. C. T., & Carvalho, V. M. (2020). Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts. *Nature communications*, *11*(1), 6201.
15. Grossegesse, M., Hartkopf, F., Nitsche, A., Schaade, L., Doellinger, J., & Muth, T. (2020). Perspective on proteomics for virus detection in clinical samples. *Journal of Proteome Research*, *19*(11), 4380–4388.
16. Gouveia, D., Grenga, L., Gaillard, J.-C., Gallais, F., Bellanger, L., Pible, O., & Armengaud, J. (2020). Shortlisting SARS-CoV-2 Peptides for Targeted Studies from Experimental Data-Dependent Acquisition Tandem Mass Spectrometry Data. *PROTEOMICS*, *20*(14), 2000107. https://doi.org/10.1002/pmic.202000107.
17. Grenga, L., & Armengaud, J. (2021). Proteomics in the COVID-19 Battlefield: First Semester Check-Up. *PROTEOMICS*, *21*(1), 2000198. https://doi.org/10.1002/pmic.202000198.

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202000226 in the Supporting Information section at the end of the article.

How to cite this article: Grossegesse, M., Nitsche, A., Schaade, L., & Doellinger, J. (2021). Application of spectral library prediction for parallel reaction monitoring of viral peptides. *Proteomics*, 21:e2000226. https://doi.org/10.1002/pmic.202000226