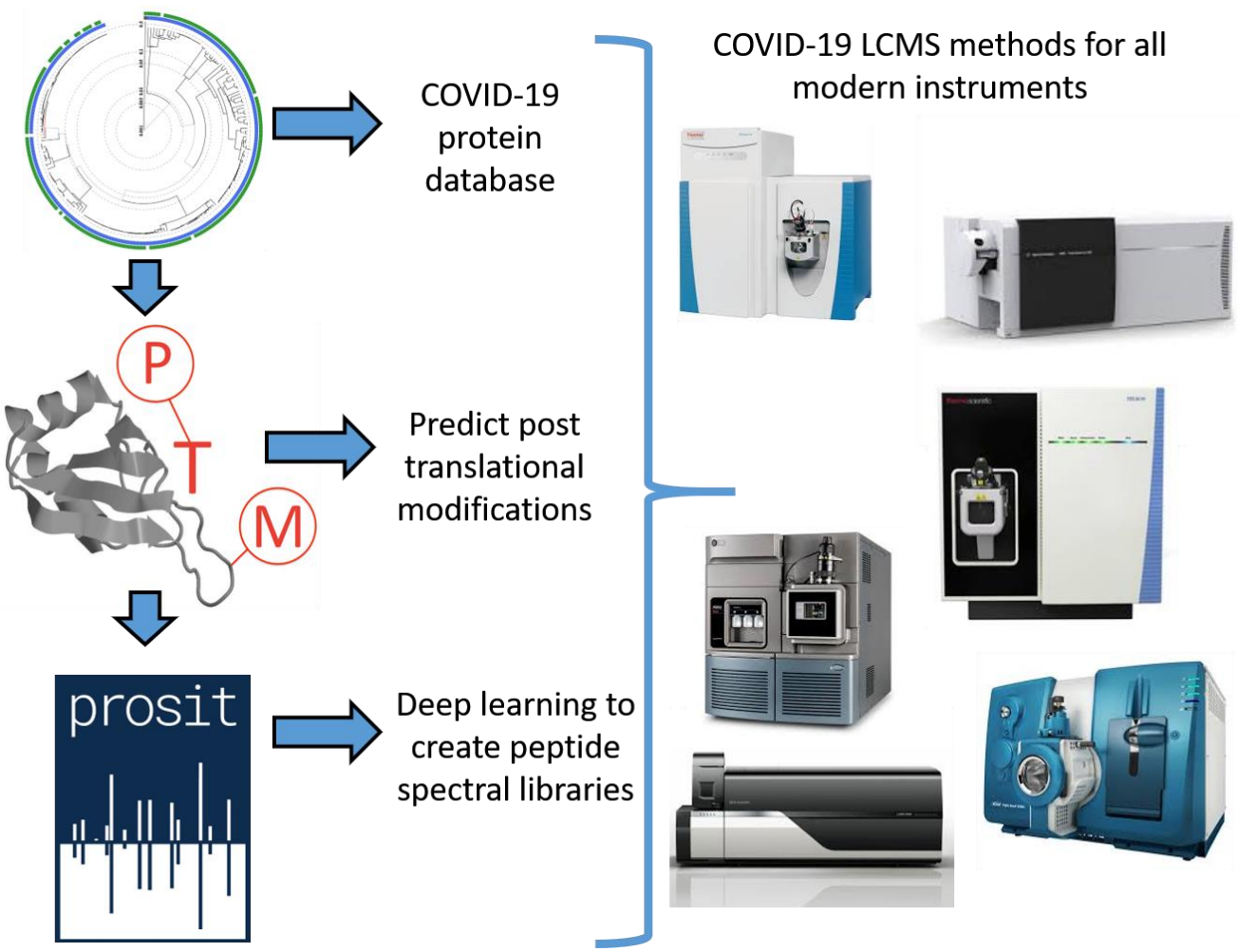
bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# *In silico* approach to accelerate the development of mass spectrometry-based proteomics methods for detection of viral proteins: Application to COVID-19

Conor Jenkins[1] and Ben Orsburn[2*]

[1]Hood College Department of Biology, Frederick, MD, USA; conor.jenkins@outlook.com
[2]University of Virginia School of Medicine, Charlottesville, VA, USA; orsburn@vt.edu
*Corresponding author




**Abstract**

The novel coronavirus disease first identified in 2019 in Wuhan, China (COVID-19) has become a serious global public health concern. One current issue is the ability to adequately screen for the virus causing COVID-2 (SARS-CoV-2). Here we demonstrate the feasibility of shotgun proteomics as a SARS-CoV-2 screening method, through the detection of viral peptides in proteolytically digested body fluids. Using *in silico* methods, we generated trypsin-based shotgun proteomics methods optimized for LCMS systems from 5 commercial instrument vendors (Thermo, SCIEX, Waters, Shimadzu, and Agilent). First, we generated protein FASTA files and their protein digest maps. Second, the FASTA files were used to generate spectral libraries based on experimental data. Third, transition lists were derived from the

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

spectral libraries using the vendor neutral and open Skyline software environment. Finally, we identified 17 post-translational modifications using linear motif modeling.

**Table 1.** Resources provided as supplemental material with this submission.

| **Proteomics Databases** | **Description** |
|---|---|
| 2019-nCoVpFASTA1 | FASTA file of COVID-19 Protein Sequences |
| Human_plus_2019-nCoVpFASTA2 | Human UniProt FASTA plus COV-19 Proteins above |
| 2019-nCoV_Trypsin_Prosit_Spectral_Library-MSP-format.zip | Prosit derived spectral library of COV-19 peptides |
| | |
| **Instrument-Specific Run Parameters** | |
| **SRM Transition lists** | |
| Thermo TSQ Quantum/Vantage | Formatted for Thermo/Finnigan Legacy TSQ |
| Thermo TSQ Altis/Quantis/Fortis | Formatted for current generation Thermo TSQ |
| SCIEX Triple Quadrupole Systems | Contains specific SCIEX parameters as described in methods |
| Agilent all models | Formatted for SCIEX |
| Waters XEVO | Formatted for Waters and Waters XEVO |
| Shimadzu all models | Transitions in .txt format as required by vendor interfaced |
| PRM Window List for 3 peptides/protein | Monoisotopic mass for 3 peptides for building PRM methods |
| Example PRM methods for Thermo Q Exactive/Exploris/Fusion | Premade .meth files with all parameters for these systems |
| **Data Processing** | |
| PRM Example Processing Skyline Templates | Data processing template for PRM methods above |
| SRM Example Processing Skyline Templates | Data processing templates for SRM methods above |
| **Post Translational Modification Predictions** | |
| ModPred All Predicted Modifications Excel Sheet | A compiled list of all PTMs from ModPred output |
| ModPred All High confidence Predicted Modifications | A simplified sheet of all high confidence PTMs |
| ModPred Summary and counts of high confidence PTMs | A combined sheet for counting high confidence PTMs |

## Introduction

The detection of viral proteins in body fluids can be a rapid and specific diagnostic for infection in severe acute respiratory syndrome (SARS).[1–3] During the 2003 (SARS) outbreak, non-MS based methods of

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

protein detection proved to be more successful[4,5] than LCMS methods.[6–8] Non-MS based methods, such as western blots, enzyme-linked immunosorbent assays (ELISAs), and protein arrays, rely on antibodies for the detection of proteins. Given recent studies concerning high variability in antibody production, LCMS-based methods are an attractive alternative approach for the rapid identification of small molecules, proteins, and peptides in clinical settings where consistency is paramount.[9,10]

In the 15 years since the 2003 SARS outbreak, LCMS technology has experienced a revolution led primarily by increases in the speed, sensitivity, and resolution of MS instruments. Today, protein array and antibody-based methods are falling out of favor in both research and clinical diagnostics, due in large part to the improvements in LCMS technology.[11,12] A review of this growth by Grebe and Singh described a clinical lab with no LCMS systems in 1998 that completed over 2 million individual LCMS clinical assays in 2010.[13] Incremental improvements in rapid sample preparation techniques, chromatography, and data processing have also contributed to the increasing use of LCMS-based clinical testing. A 2013 study demonstrated the level of advance by identifying 4,000 yeast proteins in one hour of LCMS run time, identifying approximately 75 proteins/min at a rate 100 times faster than studies a decade prior.[14]

LCMS methods can measure protein quantity by either the intact protein (protein-centric) or the analysis of proteolytic peptides (peptide-centric). While the mass to charge (m/z) ratio of a peptide or protein (MS1) may be a specific diagnostic in some materials, the majority of LCMS methods employ tandem MS in which the peptide or protein parent ion is subjected to gas phase collision to produce fragment ions. The measurement of the fragment ions (MS2) has a higher specificity and lower level of false positives and is the method of choice in clinical diagnostics.[12,13] Protein-centric assays, such as intact parallel reaction monitoring or proteoform monitoring, performs well on small GTPases like KRAS, but larger proteins cannot currently be monitored with the same relative ease and accuracy.[15] For this reason, peptide-centric assays are currently the dominant proteomics approach.

In peptide-centric assays, commonly called “bottom-up” or “shotgun” proteomics, proteins are subjected to proteolytic cleavage to produce smaller peptide sequences prior to LCMS, often using chemically modified trypsin.[16–18] Proteolytic digestion of a protein mixture increases the total number of molecules present and thereby increases the relative background noise of the sample. However, the production of multiple peptides from the protein of interest typically results in the production of highly selective peptide targets. Through *a priori* selection of peptide targets that are biologically unique or in some way chemically distinct, extremely specific assays can be rapidly designed.[19,20] When multiple selective peptides are available for a given protein, the independent peptide measurements can be combined to provide replicate measurements to increase certainty in protein presence and relative abundance.

A universal in all LCMS assays is the rule of “breadth vs. depth,” simply stated, increasing the number of targets in an experiment results in a decrease in the overall sensitivity of each measurement. The most common example is that untargeted assays which may observe thousands of peptide ions per experiment invariably have a lower limit of detection (LOD) and quantification (LOQ) compared to assays where a smaller number of ions are targeted.[21–23] Improvements in each subsequent generation of hardware can mitigate this compromise, but the improvement is limited. Today, the only way to truly offset this rule is to increase the total LCMS run time. A 2014 study by Majchrzykiewicz-Koehorst *et al.* described an untargeted LCMS assay that could discern three viruses in samples, both *in vitro* and *ex vivo*. However, the run time for this experiment was six hours per sample using nanoESI-quadrupole TOF technology.[24]

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Targeted peptide-centric assays are advantageous when sensitivity is paramount over the quantity of identified targets. Targeted assays often rely on tandem MS with high speed, but relatively low accuracy, quadrupoles.[21] Quadrupoles can be used to select ions for fragmentation with quantification of fragment ions by other quadrupoles in single reaction monitoring (SRM). They can also be used in conjunction with high resolution systems in single ion monitoring (SIM) and parallel reaction monitoring (PRM). SRM relies on fragmentation which requires *a priori* the mass to charge ratio (m/z) of both the peptide ion of interest and its dominant ions produced during fragmentation. Commonly, the collision energy is optimized for each peptide fragment in order to maximize efficiency and signal. SIM uses higher resolution scans in lieu of fragmentation, which requires *a priori* only an approximate m/z ratio for the peptide. In SIM targeted assays, the peptide ion's exact mass is extracted post-run during data processing. Two studies have shown that SIM scans with ≥60,000 resolution produce selectivity comparable to SRM.[25,26] PRM targeted assays combine fragmentation and high resolution molecule monitoring. In contrast to SIM, isolated ions are subjected to fragmentation. Post-run data processing then calculates quantification from the high-resolution accurate mass of the fragment ions. For proteins with only one available peptide target, PRM is the most selective technique in modern proteomics.[27,28]

Untargeted assays can be divided into two broad categories, data dependent (DDA) and data independent analyses (DIA). In DDA, the masses of all ions are observed in a relatively wide m/z range (MS1); the MS1 peptide ions meeting user-defined thresholds are subjected to fragmentation.[29] The resulting fragment ions are scanned, typically in a wide range for obtaining the most complete coverage of the peptide fragment ions. Common user-defined thresholds in DDA include selecting for peptides with: (1) sufficient signal for sequencing and (2) isotopic distributions matching models of typical peptides for that m/z.[30] The ion selection is performed in real time and automatically by the instrument, and has required less user input in method design with each generation of hardware.[31,32] DDA data analysis methods requires the least *a priori* information of all current LCMS methods. Although full *de novo* identification of peptide sequences is possible, it remains computationally demanding, thus typical post-instrument analysis relies on peptide search engines. Peptide search engines require the user to provide a database containing the protein sequences the user expects. As the MS1 and MS2 data is directly compared to this expected database, results are impacted by the accuracy of the database as well as on the availability of comprehensive proteome sequence information for the organism(s). Much of proteomics research relies on peptide search engines that predate the LCMS instrumentation by decades. Such software can consider multiple scenarios, such as alternative charge states, but cannot account for post translational modifications (PTMs) unless defined by the user *a priori*.[33] Recent developments in next-generation search engines (*i.e.*, MetaMorpheus[34], Bolt[35], Byonic[36], and MSFragger[37]) can identify PTMs without the user providing predicted PTMs, but these tools have yet to be widely adopted by the field.

Unlike DDA, DIA does not perform automatic real-time decision making processes. Instead, MS1 and MS2 scans are acquired with set mass widths (MS2 windows) that cover the entire peptide mass range of interest. The size of the MS2 windows vary based on the instrument speed and sensitivity, but the same instrument run method may be applied to any peptide experiment. Thus, DIA involves minimal optimization of the instrument run method, but this generalized data collection approach puts the onus on the post-run data analysis of the resulting MS2 fragments.

Large MS2 windows may display peptide ion fragments originating from hundreds of unique peptide ions. Peptides are detected in DIA MS2 windows by matching the experimental results against peptide fragment spectral libraries. Spectral libraries are annotated peptide fragments, typically produced by

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

previous DDA experiments. MS2 spectra selected for libraries typically contain complete fragment sequence coverage, with an MS2 fragment representing the product of breaking each peptide bond within the peptide of interest. In a common DIA workflow, a portion of all peptides from the study are pooled and chemically fractionated before being subjected to DDA LCMS analysis. The DDA experiments are used to create the spectral library for the experiment and each individual sample is separately analyzed by DIA. Quantification of the peptides and proteins occurs in the individual DIA experiments and the spectral libraries from the pools serve as a reference for identification of the peptides being quantified.[38–40]

While spectral libraries are essential for DIA, they can also be utilized in DDA and can results in the most specific SRM and PRM targeted assays. Peptide fragmentation follows specific energetic patterns, resulting primarily in fragments caused by separation at the peptide bond. It is therefore possible to create theoretical spectral libraries *in silico* from peptide sequence alone. Theoretical spectral libraries are especially useful when the biological samples are unavailable. New tools that employ deep learning algorithms have been demonstrated to produce theoretical MS2 spectra superior to previous prediction models and, in the absence of true experimental data, are the best resources currently available.[41,42] These deep learning algorithms can learn from vast libraries of experimental data to predict the fragmentation patterns of new peptide sequences that they are given. One such algorithm, PROSIT, uses the vast synthetic human peptide libraries, from the ProteomeTools project[43] for its training dataset. Due to the high quality of the 450,000 synthetic peptides experimentally fragmented in ProteomeTools to date, PROSIT has been demonstrated to create spectral libraries that are, in some cases, superior to experimentally derived in-house spectral libraries.[44]

In this study, we describe a *in silico* approach to develop LCMS screening assays of virus peptides in human samples. Currently, there is no publicly available proteomics data from COVID-19 infected human tissues and we are restricted from access to these materials. Our *in silico* developed materials facilitate both global and targeted analysis by providing all necessary materials for both DDA and DIA investigation of these materials through the production of FASTA databases, spectral libraries and a list of predicted PTMs investigators should consider when searching with historic peptide search engines. The peptide spectral libraries are further utilized to create transition lists optimized for hardware from 5 instrument vendors and complete PRM methods for all 3 quadrupole Orbitrap architectures. All materials and methods described herein are available as supplemental material to this publication as described in Table 1. The supplemental material includes protein sequences (FASTA files), predicted PTMs, theoretical MS2 spectral libraries, instrument methods and targeted method data processing templates.

Our work demonstrates not only the feasibility of this approach, but also its ability to rapidly develop methods even in the face of limitation of access to sample experimental data. We use the example of SARS-CoV-2 viral protein detection to underscore the utility of this approach in responding to an urgent public health crisis.

## Methods and Materials

### *Coronavirus FASTA databases*

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

At the date of this writing, only theoretical protein sequences for SARS-CoV-2, are available. These sequences are being acquired and annotated and the result of translation of genomic sequence information. All sequences in this study were obtained from NCBI accession: txd2697049, https://www.ncbi.nlm.nih.gov/protein/?term=txid2697049). Using Proteome Discoverer 2.4 (Thermo), the protein sequences were combined into a single protein FASTA database (2019-nCOVpFASTA1; Supplemental Information), and added to human proteome sequences (UniProt SwissProt Human database; downloaded 2/15/2020) to produce a database including both human and COVID-19 protein sequences (Human_plus_2019-nCOVpFASTA2; Supplemental Information).

*Publicly available proteomics data from human samples infected with other coronaviruses*

Publicly available experiments on other coronavirus experiments were found by searching the ProteomeXchange Consortium web interface (http://www.proteomexchange.org/).[45,46] Clarification of the identity of unpublished data from Pacific Northwest National Laboratory (PNNL) was provided by Dr. Michael Monroe.

*Creation of peptide spectral libraries exclusive to SARS-CoV-2 with Prosit deep learning*

The 2019-nCOVpFASTA1 (Supplemental Information) was converted to PROSIT peptide format with the EncyclopeDIA[47] software, resulting in an *in silico* peptide digestion (parameters: Charge Range = 2-3; Max missed cleavage = 1, m/z range = 400-1,500; default NCE = 30eV; default charge = +2). PROSIT peptide fragmentation prediction libraries were generated using the PROSIT online web portal (https://www.proteomicsdb.org/prosit/) with the spectral library modeling interface (options: prediction model = Prosit_2019_intensity; iRT prediction model = Prosit_2019_irt). The libraries were exported in NIST MSP text format for use in Skyline.[48,49] While preparing this manuscript, Proteome Sciences reported the creation of PROSIT libraries for use in the ScaffoldDIA software (cite ScaffoldDIA?) (https://support.proteomesoftware.com/hc/en-us/articles/360035151172-Prosit-Derived-Spectral-Libraries-for-Scaffold-DIA-Searches).

*PRM and SRM method development*

For SRM transitions, the 2019-nCOVpFASTA1 (Supplemental Information) was imported into Skyline v20.1.0.28 (University of Washington) along with the PROSIT tryptic peptide spectral library. Peptide settings and transitions were optimized within Skyline to reflect the vendor optimization requirements. For Agilent systems, the 20ms default dwell time was selected for the transition settings. For SCIEX instruments, the same dwell time was utilized as well as automatic optimization of the declustering potential and compensation voltage from the transition settings menu. For Waters, Thermo, and Shimadzu systems, no further settings were required for transition list generation. All transition lists were exported as unscheduled 15min methods. For PRM methods, three peptides were selected for each protein due to the increased time per scan relative to SRM methods. Instrument-specific Skyline files are included in the Supplemental information as described in Table 1.

*Prediction of PTMs*

PTMs were predicted for the 2019-nCOVpFASTA1 proteins (Supplemental Information) using the ModPred web interface (www.modpred.org; accessed 1/31/2020).[50] All PTMs available at the date of analysis were selected as theoretical sites, and the Basic non evolutionary model was applied. ModPred ranks each PTM as high, medium, or low confidence as previously described.[50] The ModPred web

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

interface can accept a maximum of 5,000 amino acids. In order to analyze YP_009724389.1 the predicted sequence had to be divided into five sequences, using a 100 amino acid overlap to avoid disrupting potential large motifs. The results of ModPred were compiled into a single spreadsheet with all modifications of all confidence levels. A second sheet was created that contained only the high confidence PTMs predicted, as well as a final summary for the counting of predicted high confidence PTM occurrence, provided as Supplemental Material as described in Table 1.

## Results

### *Theoretical peptides*

In shotgun proteomics, proteins are first digested into smaller peptide fragments that are more easily detected by the instrument. Given its widespread use, high efficiency and speed of digestion, we chose to develop methods that exclusively use the proteolytic enzyme trypsin, which produces “tryptic peptides." Sequencing grade trypsin exhibits high efficiency cleavage at unmodified (1) arginine and (2) lysine residues unless followed by a proline. Trypsin also has the advantage of leaving a terminal basic residue at the cleavage site, which increases the likelihood of complete fragment ion coverage from the charged terminal.[18] Due to these reasons, trypsin is utilized unless the protein sequence has an abnormally high or low number of lysine or arginine residues. A very high frequency of the residues (such as Lysine-rich proteins) will create very short peptides that could be uninformative for protein identification. A very low frequency of the residues will create very large peptides, or undigested (“intact”) proteins in some cases, that are difficult to detect and fragment. Our theoretical trypsin digest of the 2019-nCoVpFASTA1 database produced tryptic peptides with average lengths of 8 to 18 amino acids. These results indicate that trypsin digestion is an appropriate choice for detection of these viral proteins.

### *Example targeted methods*

To widen use of our developed methods, we used the Skyline software to create transition lists optimized specifically for each of the Skyline-compatible triple quadrupole instruments. While minor modifications are required for SCIEX, Agilent, and Shimadzu instruments, Waters Xevo and Thermo instruments use identical parameters for transition list design. Most modern triple quadrupole instruments are capable of 500 SRMS/sec and fully permit the use of 2,000 transition lists, as provided here. For older instruments that lack this scan speed or that require higher dwell times, the transition lists included in the supplemental methods may be reduced by the end user accordingly.

PRM methods monitor multiple transitions simultaneously but at a time cost. The highest scan speed currently available in Orbitrap instruments is 48 scans per second and is only available on the Exploris 480 system (data not shown). In order to achieve maximum sensitivity, higher fill times are often required for these instruments. We chose to utilize three peptides/protein for these methods. Alternative peptides can be selected from the Skyline files provided (Supplemental Information) or by selecting peptide mass targets from the other transition lists.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

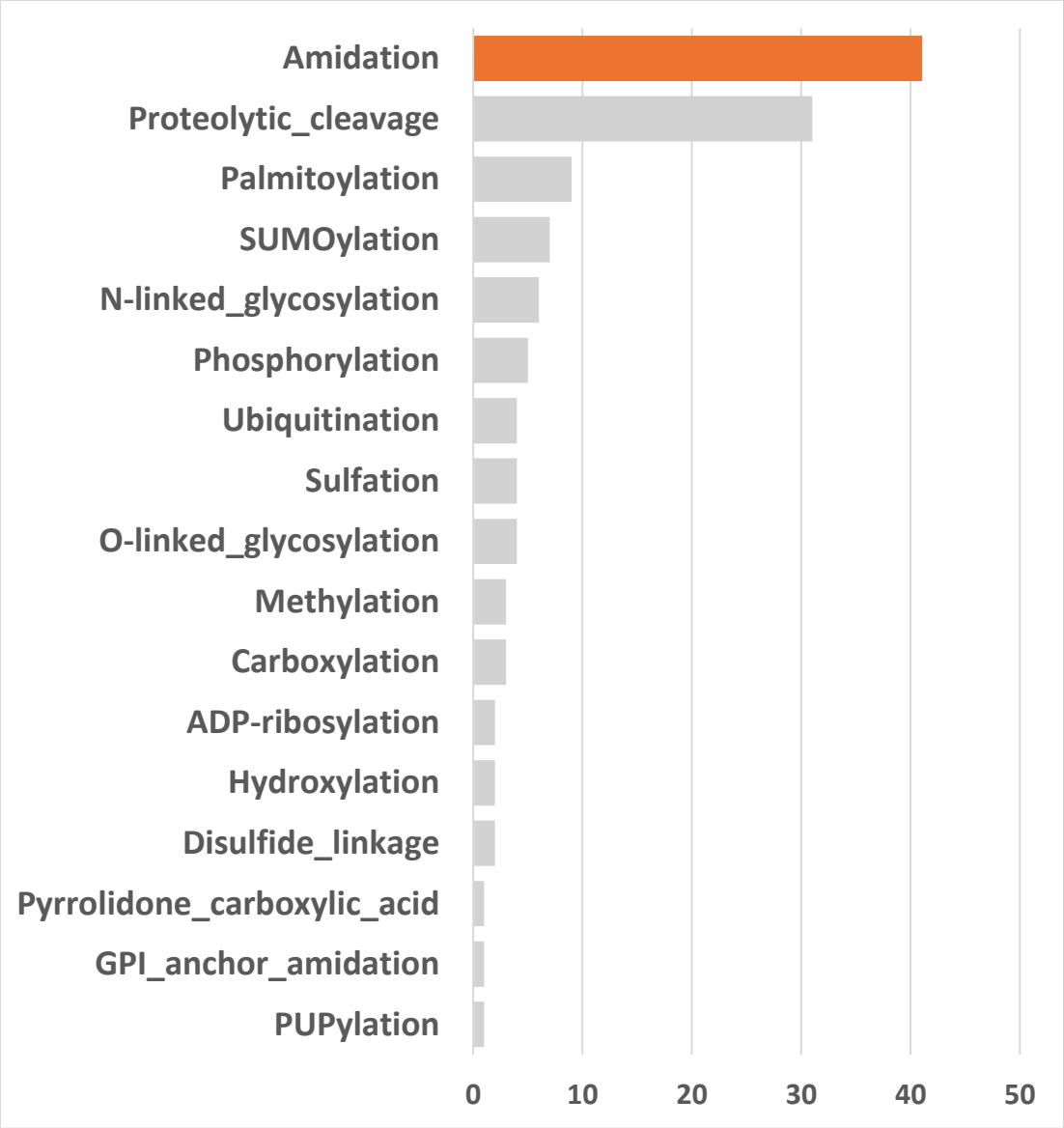



**Figure 1.** The number of high confidence prediction sites for each PTM by the ModPred program.

*Untargeted methods*

The Prosit spectral libraries (Supplemental Material) enable the interrogation of DIA data and may be used for DDA experiments that employ tools such as the MSPepSearch (NIST).[51] DDA data requires only the protein FASTA file and a list of PTMs that may be present in the sample. Our analysis using ModPred predicted 17 possible PTMs (Figure 1). Amidation was the most frequent predicted PTM, but there is no known biological mechanism that we could derive from a survey of the literature. Palmitoylation, the second most frequent predicted PTM, is a well characterized viral PTM with critical functions in human immunodeficiency virus (HIV), human herpes virus (HHV), and influenza virus infectivity.[52–54]

**Table 2.** Currently unpublished data from PNNL on MERS-COV infected human cells.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

| Description | ProteomeXchange Identifier |
|---|---|
| Calu-3 proteome response to an infectious clone of Middle Eastern Respiratory Syndrome coronavirus | PXD002358 |
| Calu-3 metabolome response to an infectious clone of Middle Eastern Respiratory Syndrome coronavirus | PXD002359 |
| Calu-3 lipidome response to an infectious clone of Middle Eastern Respiratory Syndrome coronavirus | PXD002360 |
| Calu-3 proteome response to an infectious clone of H7N9 Influenza virus and mutant H7N9 viruses | PXD002385 |
| Calu-3 metabolome response to an infectious clone of H7N9 Influenza virus and mutant H7N9 viruses | PXD002362 |

*LCMS datasets for other coronaviruses*

The ProteomeXchange Consortium is an open-access platform for the rapid sharing of proteomics data. The prominent proteomic technical journals, the Journal of Proteome Research and Molecular and Cellular Proteomics, strictly require that all unprocessed instrument files and processed results are made publicly available through these services. While attempting to obtain coronavirus data, we identified unpublished metabolomics, lipidomics, and proteomics data generated from human samples infected with Middle Eastern respiratory syndrome (MERS-CoV). We provide this data, as well as a list of other proteomic studies of note, for use in comparative studies with SARS-CoV-2.

**Conclusions**

Using *in silico* methods, we have developed methods for the detection of SARS-CoV-2 in human samples. *In vitro* validation of this method is required and outside the scope of this paper given our lack of access to such samples. We have provided the minimum materials for data processing for both DDA and DIA untargeted proteomics methods with FASTA databases, spectral libraries and by predicting relevant PTMs for consideration. To broaden the number of labs that can apply our methods, we optimized run parameters for widely used LCMS systems compatible with Skyline, representing instruments from five companies. We will continue to refine these resources and post updates to these methods to LCMSmethods.org and invite researchers anywhere in the world to contact us for assistance in further optimization to address this emerging threat.

**Acknowledgements**

We thank Dr. Matthew Monroe at Pacific Northwest National Laboratory (PNNL) for assistance curating unpublished data in the ProteomeXchange/MASSIVE public repository

(1) Kammila, S.; Das, D.; Bhatnagar, P. K.; Sunwoo, H. H.; Zayas-Zamora, G.; King, M.; Suresh, M. R. A Rapid Point of Care Immunoswab Assay for SARS-CoV Detection. *J. Virol. Methods* **2008**. https://doi.org/10.1016/j.jviromet.2008.05.023.

(2) Cho, S. J.; Woo, H. M.; Kim, K. S.; Oh, J. W.; Jeong, Y. J. Novel System for Detecting SARS

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Coronavirus Nucleocapsid Protein Using an SsDNA Aptamer. *J. Biosci. Bioeng.* **2011**. https://doi.org/10.1016/j.jbiosc.2011.08.014.

(3) Che, X. Y.; Hao, W.; Wang, Y.; Di, B.; Yin, K.; Xu, Y. C.; Feng, C. Sen; Wan, Z. Y.; Cheng, V. C. C.; Yuen, K. Y. Nucleocapsid Protein as Early Diagnostic Marker for SARS. *Emerg. Infect. Dis.* **2004**. https://doi.org/10.3201/eid1011.040516.

(4) Yip, T. T. C.; Cho, W. C. S.; Cheng, W. W.; Chan, J. W. M.; Ma, V. W. S.; Yip, T. T.; Lau Yip, C. N. B.; Ngan, R. K. C.; Law, S. C. K. Application of ProteinChip Array Profiling in Serum Biomarker Discovery for Patients Suffering from Severe Acute Respiratory Syndrome. *Methods Mol. Biol.* **2007**. https://doi.org/10.1007/978-1-59745-304-2_20.

(5) Yip, T. T. C.; Chan, J. W. M.; Cho, W. C. S.; Yip, T. T.; Wang, Z.; Kwan, T. L.; Law, S. C. K.; Tsang, D. N. C.; Chan, J. K. C.; Lee, K. C.; et al. Protein Chip Array Profiling Analysis in Patients with Severe Acute Respiratory Syndrome Identified Serum Amyloid A Protein as a Biomarker Potentially Useful in Monitoring the Extent of Pneumonia. *Clin. Chem.* **2005**. https://doi.org/10.1373/clinchem.2004.031229.

(6) Ren, Y.; He, Q. Y.; Fan, J.; Jones, B.; Zhou, Y.; Xie, Y.; Cheung, C. Y.; Wu, A.; Chiu, J. F.; Peiris, J. S. M.; et al. The Use of Proteomics in the Discovery of Serum Biomarkers from Patients with Severe Acute Respiratory Syndrome. *Proteomics* **2004**. https://doi.org/10.1002/pmic.200400897.

(7) Zhang, L.; Zhang, Z. P.; Zhang, X. E.; Lin, F. S.; Ge, F. Quantitative Proteomics Analysis Reveals BAG3 as a Potential Target To Suppress Severe Acute Respiratory Syndrome Coronavirus Replication. *J. Virol.* **2010**. https://doi.org/10.1128/jvi.00213-10.

(8) Ying, W.; Hao, Y.; Zhang, Y.; Peng, W.; Qin, E.; Cai, Y.; Wei, K.; Wang, J.; Chang, G.; Sun, W.; et al. Proteomic Analysis on Structural Proteins of Severe Acute Respiratory Syndrome Coronavirus. In *Proteomics*; 2004. https://doi.org/10.1002/pmic.200300676.

(9) Voskuil, J. L. A. Commercial Antibodies and Their Validation. *F1000Research* **2014**. https://doi.org/10.12688/f1000research.4966.2.

(10) Voskuil, J. L. A. The Challenges with the Validation of Research Antibodies. *F1000Research* **2017**. https://doi.org/10.12688/f1000research.10851.1.

(11) Geyer, P. E.; Kulak, N. A.; Pichler, G.; Holdt, L. M.; Teupser, D.; Mann, M. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2016**. https://doi.org/10.1016/j.cels.2016.02.015.

(12) Grebe, S. K. G.; Singh, R. J. Clinical Peptide and Protein Quantification by Mass Spectrometry (MS). *TrAC - Trends in Analytical Chemistry*. 2016. https://doi.org/10.1016/j.trac.2016.01.026.

(13) Grebe, S. K. G.; Singh, R. J. LC-MS/MS in the Clinical Laboratory - Where to from Here? *Clin. Biochem. Rev.* **2011**.

(14) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. *Mol. Cell. Proteomics* **2014**. https://doi.org/10.1074/mcp.M113.034769.

(15) Ntai, I.; Fornelli, L.; DeHart, C. J.; Hutton, J. E.; Doubleday, P. F.; LeDuc, R. D.; van Nispen, A. J.; Fellers, R. T.; Whiteley, G.; Boja, E. S.; et al. Precise Characterization of KRAS4b Proteoforms in Human Colorectal Cells and Tumors Reveals Mutation/Modification Cross-Talk. *Proc. Natl. Acad.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

*Sci. U. S. A.* **2018**. https://doi.org/10.1073/pnas.1716122115.

(16) Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond Trypsin. *FEBS Journal*. 2015. https://doi.org/10.1111/febs.13287.

(17) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six Alternative Proteases for Mass Spectrometry-Based Proteomics beyond Trypsin. *Nat. Protoc.* **2016**. https://doi.org/10.1038/nprot.2016.057.

(18) Kolsrud, H.; Malerod, H.; Ray, S.; Reubsaet, L.; Lundanes, E.; Greibrokk, T. A Critical Review of Trypsin Digestion for LC-MS Based Proteomics. In *Integrative Proteomics*; 2012. https://doi.org/10.5772/29326.

(19) Zauber, H.; Kirchner, M.; Selbach, M. Picky: A Simple Online PRM and SRM Method Designer for Targeted Proteomics. *Nature Methods*. 2018. https://doi.org/10.1038/nmeth.4607.

(20) Lawrence, R. T.; Searle, B. C.; Llovet, A.; Villén, J. Plug-and-Play Analysis of the Human Phosphoproteome by Targeted High-Resolution Mass Spectrometry. *Nat. Methods* **2016**. https://doi.org/10.1038/nmeth.3811.

(21) Pino, L. K.; Searle, B. C.; Yang, H.-Y.; Hoofnagle, A. N.; Noble, W. S.; MacCoss, M. J. Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in Quantitative Proteomics. *J. Proteome Res.* **2020**. https://doi.org/10.1021/acs.jproteome.9b00666.

(22) Galitzine, C.; Egertson, J. D.; Abbatiello, S.; Henderson, C. M.; Pino, L. K.; MacCoss, M.; Hoofnagle, A. N.; Vitek, O. Nonlinear Regression Improves Accuracy of Characterization of Multiplexed Mass Spectrometric Assays. *Mol. Cell. Proteomics* **2018**. https://doi.org/10.1074/mcp.RA117.000322.

(23) Meyer, J. G.; Schilling, B. Clinical Applications of Quantitative Proteomics Using Targeted and Untargeted Data-Independent Acquisition Techniques. *Expert Review of Proteomics*. 2017. https://doi.org/10.1080/14789450.2017.1322904.

(24) J.A., M.-K.; E., H.; H., T.; A.G., H.; A.L., de J.; M.C., V.; N.J.A., S.; J., van der P.; F.E.J., C.; N.J., S.; et al. Rapid and Generic Identification of Influenza A and Other Respiratory Viruses with Mass Spectrometry. *Journal of virological methods*. 2015.

(25) Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics* **2012**. https://doi.org/10.1074/mcp.O112.019802.

(26) Higgs, R. E.; Butler, J. P.; Han, B.; Knierman, M. D. Quantitative Proteomics via High Resolution MS Quantification: Capabilities and Limitations. *Int. J. Proteomics* **2013**. https://doi.org/10.1155/2013/674282.

(27) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **2012**. https://doi.org/10.1074/mcp.O112.020131.

(28) Gallien, S.; Bourmaud, A.; Kim, S. Y.; Domon, B. Technical Considerations for Large-Scale Parallel Reaction Monitoring Analysis. *J. Proteomics* **2014**. https://doi.org/10.1016/j.jprot.2013.10.029.

(29) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews*. 2013. https://doi.org/10.1021/cr3003533.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(30) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**. https://doi.org/10.1021/acs.analchem.7b04424.

(31) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**. https://doi.org/10.1146/annurev-anchem-071114-040325.

(32) Hecht, E. S.; Scigelova, M.; Eliuk, S.; Makarov, A. Fundamentals and Advances of Orbitrap Mass Spectrometry. In *Encyclopedia of Analytical Chemistry*; 2019. https://doi.org/10.1002/9780470027318.a9309.pub2.

(33) Prakash, A.; Majumder, S.; Ahmad, S.; Varkey, M.; Anish, T. A.; Jenkins, C.; Rigby, M.; Orsburn, B. Detection and Verification of 2.3 Million Cancer Mutations in NCI60 Cancer Cell Lines with a Cloud Search Engine. *J. Proteomics* **2019**. https://doi.org/10.1016/j.jprot.2019.103488.

(34) Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M. Enhanced Global Post-Translational Modification Discovery with MetaMorpheus. *J. Proteome Res.* **2018**. https://doi.org/10.1021/acs.jproteome.7b00873.

(35) Prakash, A.; Ahmad, S.; Majumder, S.; Jenkins, C.; Orsburn, B. Bolt: A New Age Peptide Search Engine for Comprehensive MS/MS Sequencing Through Vast Protein Databases in Minutes. *J. Am. Soc. Mass Spectrom.* **2019**. https://doi.org/10.1007/s13361-019-02306-3.

(36) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinforma.* **2012**. https://doi.org/10.1002/0471250953.bi1320s40.

(37) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry-Based Proteomics. *Nat. Methods* **2017**. https://doi.org/10.1038/nmeth.4256.

(38) Kelstrup, C. D.; Bekker-Jensen, D. B.; Arrey, T. N.; Hogrebe, A.; Harder, A.; Olsen, J. V. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J. Proteome Res.* **2018**. https://doi.org/10.1021/acs.jproteome.7b00602.

(39) Prakash, A.; Peterman, S.; Ahmad, S.; Sarracino, D.; Frewen, B.; Vogelsang, M.; Byram, G.; Krastins, B.; Vadali, G.; Lopez, M. Hybrid Data Acquisition and Processing Strategies with Increased Throughput and Selectivity: PSMART Analysis for Global Qualitative and Quantitative Analysis. *J. Proteome Res.* **2014**. https://doi.org/10.1021/pr5003017.

(40) Doerr, A. DIA Mass Spectrometry. *Nat. Methods* **2014**. https://doi.org/10.1038/nmeth.3234.

(41) Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth, F. P.; Gygi, S. P. Intensity-Based Protein Identification by Machine Learning from a Library of Tandem Mass Spectra. *Nat. Biotechnol.* **2004**. https://doi.org/10.1038/nbt930.

(42) Yang, Y.; Liu, X.; Shen, C.; Lin, Y.; Yang, P.; Qiao, L. In Silico Spectral Libraries by Deep Learning Facilitate Data-Independent Acquisition Proteomics. *Nat. Commun.* **2020**. https://doi.org/10.1038/s41467-019-13866-z.

(43) Zolg, D. P.; Wilhelm, M.; Schmidt, T.; Médard, G.; Zerweck, J.; Knaute, T.; Wenschuh, H.; Reimer, U.; Schnatbaum, K.; Kuster, B. ProteomeTools: Systematic Characterization of 21 Post-Translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MS/MS) Using Synthetic Peptides. *Mol. Cell. Proteomics* **2018**. https://doi.org/10.1074/mcp.tir118.000783.

(44) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. Prosit: Proteome-Wide Prediction of Peptide Tandem Mass Spectra by Deep Learning. *Nat. Methods* **2019**. https://doi.org/10.1038/s41592-019-0426-7.

(45) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange Provides Globally Coordinated Proteomics Data Submission and Dissemination. *Nature Biotechnology*. 2014. https://doi.org/10.1038/nbt.2839.

(46) Ternent, T.; Csordas, A.; Qi, D.; Gómez-Baena, G.; Beynon, R. J.; Jones, A. R.; Hermjakob, H.; Vizcaíno, J. A. How to Submit MS Proteomics Data to ProteomeXchange via the PRIDE Database. *Proteomics* **2014**. https://doi.org/10.1002/pmic.201400120.

(47) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat. Commun.* **2018**. https://doi.org/10.1038/s41467-018-07454-w.

(48) 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. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**. https://doi.org/10.1093/bioinformatics/btq054.

(49) Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; et al. Platform-Independent and Label-Free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline. *Mol. Cell. Proteomics* **2012**. https://doi.org/10.1074/mcp.M112.017707.

(50) Pejaver, V.; Hsu, W. L.; Xin, F.; Dunker, A. K.; Uversky, V. N.; Radivojac, P. The Structural and Functional Signatures of Proteins That Undergo Multiple Events of Post-Translational Modification. *Protein Sci.* **2014**. https://doi.org/10.1002/pro.2494.

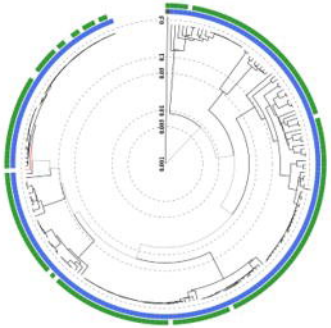
(51) Zhang, Z.; Burke, M.; Mirokhin, Y. A.; Tchekhovskoi, D. V.; Markey, S. P.; Yu, W.; Chaerkady, R.; Hess, S.; Stein, S. E. Reverse and Random Decoy Methods for False Discovery Rate Estimation in High Mass Accuracy Peptide Spectral Library Searches. *J. Proteome Res.* **2018**. https://doi.org/10.1021/acs.jproteome.7b00614.

(52) Serwa, R. A.; Abaitua, F.; Krause, E.; Tate, E. W.; O'Hare, P. Systems Analysis of Protein Fatty Acylation in Herpes Simplex Virus-Infected Cells Using Chemical Proteomics. *Chem. Biol.* **2015**. https://doi.org/10.1016/j.chembiol.2015.06.024.

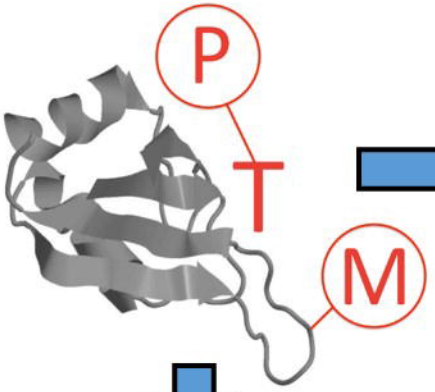
(53) Veit, M. Palmitoylation of Virus Proteins. *Biol. Cell* **2012**. https://doi.org/10.1111/boc.201200006.

(54) Veit, M.; Serebryakova, M. V.; Kordyukova, L. V. Palmitoylation of Influenza Virus Proteins. In *Biochemical Society Transactions*; 2013. https://doi.org/10.1042/BST20120210.

bioRxiv preprint doi: https://doi.org/10.1101/2020.03.08.980383; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



COVID-19 protein database




Predict post translational modifications

prosit

Deep learning to create peptide spectral libraries

COVID-19 LCMS methods for all modern instruments