OSAC RESEARCH NEEDS ASSESSMENT FORM

Title of research need: Developing statistical and informatics tools for statistically sound forensic interpretations from peptide mass spectra

Describe the need:

Mass spectrometry-based proteomics has potential uses in a variety of forensic fields including species identification, body fluid and tissue identification, toxin identification, as well as human identification. However, all current proteomic peptide identification algorithms and methods are imperfect and sometimes give incorrect peptide sequences. This is more common with poor spectral data (low signal-to-noise) but can also occur with high-quality data. Thus, new methods are needed to understand and to address limitations in peptide identifications and interpretations derived from forensic proteomics assays.

The traditional method for addressing concerns about peptide identification in forensics is to use analytical standard peptides produced by reliable synthetic methods. In some cases, especially for quantification, these peptides are site-specific isotopically labeled with 13C, 15N, and/or 2H. In other cases, natural isotopic abundance is used. For methods that require such analytical standard peptides, establishment of procedures for synthesis, verification analysis (e.g., analytical purity, amino acid analysis, isotopic composition for quantitation), and logistical distribution should be developed. In particular, visual comparison of tandem mass spectra or presence of a small number of peptide-specific fragment ions may not be sufficient; orthogonal methods such as spectral similarity scores should be established.

Proteomics in the fundamental sciences effectively side-steps concerns about peptide identification by controlling the false discovery rate (FDR; essentially the proportion of incorrect identifications in a results set) using the target-decoy approach for FDR estimation. This approach estimates the number of incorrect peptide identifications in an experiment by comparing the data to peptide sequences assumed not to be present (i.e., decoys, often reversed or randomized versions of expected peptide sequences), and adjusting the score threshold to limit FDR to an acceptable level. FDR control methods work best for large results sets, but forensics methods often rely on a small number of markers. Moreover, forensic samples frequently consist of sample-limited materials, so the problem of a large search space, necessary when the contents of a sample are entirely unknown, may be compounded by a relatively small number of sample spectra to search - the combination of which naturally results in an atypically high false positive rate when using FDR. As such, adopting the current FDR methods used in the fundamental sciences for forensic applications is unsuitable and new methods are required.

Additionally, although analytical standard peptides are extremely important in forensic proteomics, there may eventually be untargeted methods that are demonstrated to perform well without them. There are several potential lines of research from other areas of proteomics that are relevant and could prove useful for such forensic proteomic applications, including but not limited to the following:

1. Organism/toxin/biofluid identification based on whole-dataset or large-dataset statistical analyses [1-4], possibly also using protein and peptide abundance profiles [5]. (Note: dealing appropriately with various batch effects is critical for these applications.)

2. Tiered searches and/or subset-based estimation of FDR [6, 7].

3. Better accounting for unexpected post-translational modifications (PTMs), especially those that lead to masses that can be confused with standard amino acid masses [8, 9].

4. Application of artificial intelligence/machine learning strategies for peptide identification [10-12].

5. Adaptation of strategies from single-cell proteomics for dealing with low-intensity spectra [13, 14].

6. Alternatives to target-decoy FDR estimation suggested by research in the related field of metaproteomics [15].

Keyword(s): Mass spectrometry-based proteomics, peptide identification, peptide mass spectra, false discovery rate (FDR), post-translational modifications (PTMs)

Submitting subcommittee(s): Forensic Proteomics Task Group **Date Approved:** 4/2/2026

Background Information:

1. Does this research need address a gap(s) in a current or planned standard? (ex.: Field identification system for on scene opioid detection and confirmation)

Drafting of relevant standards to address this gap are in progress.

2. Are you aware of any ongoing research that may address this research need that has not yet been published (e.g., research presented in conference proceedings, studies that you or a colleague have participated in but have yet to be published)?

No.

3. Key bibliographic references relating to this research need: (ex.: Toll, L., Standifer, K. M., Massotte, D., eds. (2019). Current Topics in Opioid Research. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-180-3)

1. Jenson, S. C.; Chu, F.; Barente, A. S.; Crockett, D. L.; Lamar, N. C.; Merkley, E. D.; Jarman, K.H., MARLOWE: Taxonomic Characterization of Unknown Samples for Forensics Using De Novo Peptide Identification. bioRxiv 2024, 2024.09.30.615220.

2. Jarman, K. H., Heller, N. C., Jenson, S. C., Hutchison, J. R., Kaiser, B. L. D., Payne, S. H., Wunschel, D. S., & Merkley, E. D. (2018). Proteomics Goes to Court: A Statistical Foundation for Forensic Toxin/Organism Identification Using Bottom-Up Proteomics. *Journal of proteome research*, *17*(9), 3075–3085. https://doi.org/10.1021/acs.jproteome.8b00212

3. Ogurtsov, A., Alves, G., Rubio, A., Joyce, B., Andersson, B., Karlsson, R., Moore, E. R. B., & Yu, Y. K. (2024). MiCId GUI: The Graphical User Interface for MiCId, a Fast Microorganism Classification and Identification Workflow with Accurate Statistics and High Recall. *Journal of computational biology : a journal of computational molecular cell biology*, *31*(2), 175–178. https://doi.org/10.1089/cmb.2023.0149

4. Yang, H., Butler, E. R., Monier, S. A., Teubl, J., Fenyö, D., Ueberheide, B., & Siegel, D. (2021). A predictive model for vertebrate bone identification from collagen using proteomic mass spectrometry. *Scientific reports*, *11*(1), 10900. https://doi.org/10.1038/s41598-021-90231-5

5. Liu, W.; Butler, E.; Yang, H.; Fenyö, D.; Siegel, D., Informatics Approaches to Forensic Body Fluid Identification by Proteomic Mass Spectrometry. In Applications in Forensic Proteomics: Protein Identification and Profiling, American Chemical Society: 2019; Vol. 1339, pp 81-90.

6. Lin, A., Plubell, D. L., Keich, U., & Noble, W. S. (2021). Accurately Assigning Peptides to Spectra When Only a Subset of Peptides Are Relevant. *Journal of proteome research*, *20*(8), 4153–4164. https://doi.org/10.1021/acs.jproteome.1c00483

7. Kertesz-Farkas, A., Keich, U., & Noble, W. S. (2015). Tandem Mass Spectrum Identification via Cascaded Search. *Journal of proteome research*, *14*(8), 3027–3038. https://doi.org/10.1021/pr501173s

8. Bern, M., Kil, Y. J., & Becker, C. (2012). Byonic: advanced peptide and protein identification software. *Current protocols in bioinformatics*, *Chapter 13*, 13.20.1–13.20.14. https://doi.org/10.1002/0471250953.bi1320s40

9. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature methods*, *14*(5), 513–520. https://doi.org/10.1038/nmeth.4256

10. 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. https://doi.org/10.1038/s41592-019-0426-7

11. Yilmaz, M.; Fondrie, W.; Bittremieux, W.; Oh, S.; Noble, W. S., De Novo Mass Spectrometry Peptide Sequencing with a Transformer Model. In 39th International Conference on Machine Learning, Baltimore, Maryland, USA, 2022.

12. Buur, L. M., Declercq, A., Strobl, M., Bouwmeester, R., Degroeve, S., Martens, L., Dorfer, V., & Gabriels, R. (2024). MS²Rescore 3.0 Is a Modular, Flexible, and User-Friendly Platform to Boost Peptide Identifications, as Showcased with MS Amanda 3.0. *Journal of proteome research*, *23*(8), 3200–3207. https://doi.org/10.1021/acs.jproteome.3c0078

13. Boekweg, H., Van Der Watt, D., Truong, T., Johnston, S. M., Guise, A. J., Plowey, E. D., Kelly, R. T., & Payne, S. H. (2022). Features of Peptide Fragmentation Spectra in Single-Cell Proteomics. *Journal of proteome research*, *21*(1), 182–188. https://doi.org/10.1021/acs.jproteome.1c00670

14. Orsburn B. C. (2023). Time-of-Flight Fragmentation Spectra Generated by the Proteomic Analysis of Single Human Cells Do Not Exhibit Atypical Fragmentation Patterns. *Journal of proteome research*, *22*(3), 1003–1008. https://doi.org/10.1021/acs.jproteome.2c00715

15. Miura, N., & Okuda, S. (2023). Current progress and critical challenges to overcome in the bioinformatics of mass spectrometry-based metaproteomics. *Computational and structural biotechnology journal*, *21*, 1140–1150. https://doi.org/10.1016/j.csbj.2023.01.015

4. Review the annual operational/research needs published by the National Institute of Justice (NIJ) at https://nij.ojp.gov/topics/articles/forensic-science-research-and-development-technology-working-group-operational#latest? Is your research need identified by NIJ?

This research need would be included under the following Operational Requirements published by the NIJ in February 2024:
Forensic Biology, Scientific Research: Foundational research related to the discriminatory power and sensitivity of alternate biological analyses (e.g., proteomics, microbiome, plants, animals) to associate individuals with crime scene evidence.

5. In what ways would the research results improve current laboratory capabilities?

This research would provide forensic practitioners with tools for appropriate and statistically sound use of proteomic data applied to forensic problems.

6. In what ways would the research results improve understanding of the scientific basis for the subcommittee(s)?

A thorough investigation of methods of peptide identification from proteomic mass spectrometric data, as applied to forensically relevant questions, will provide a robust foundation for the advancement of forensic proteomic mass spectrometry and development of statistically sound forensic proteomic methods.

7. In what ways would the research results improve services to the criminal justice system?

Methods to understand and minimize the limitations in peptide identification based on interpretations derived from forensic proteomics assays would result in robust and reproducible analysis of forensic proteomic data and statistically sound forensic interpretations.

8. Status assessment (I, II, III, or IV): I

	Major gap in current knowledge	Minor gap in current knowledge
No or limited current research is being conducted	I	III

Existing current research is being conducted	II	IV

This research need has been identified by one or more subcommittees of OSAC and is being provided as an informational resource to the community.