OSAC RESEARCH NEEDS ASSESSMENT FORM

Title of research need: Development of algorithms and software to address forensically-relevant questions in proteomics

Describe the need:

Many software packages currently exist for the analysis of both targeted and non-targeted proteomic data. However, while these software packages do provide powerful tools for assigning peptide sequences to mass spectra, they do not provide additional information useful to forensic practitioners. For instance, in the case of human identification using germline single amino acid polymorphism (SAPs) contained in genetically variant peptides (GVPs), information such as random match probability and linkage disequilibrium are crucial. For species identification, phylogenetic specificity is an important factor. For body fluid or tissue identification, background distribution of protein expression in different body fluids or tissues is a useful metric of protein specificity. There are also common aspects that are essential for all forensic proteomic methods such as false identification rates and the degree of specificity of a peptide sequence as it relates to either the protein, tissue, or organism of interest. While some software does exist for these forensic applications, they are typically in-house, lack intuitiveness, and require knowledge and skill in coding languages.

Three stages of research are required to develop useful tools for forensic practitioners. The first is to develop core algorithms to address the needs for identification (i.e., human, species, and body fluids or tissues), the second is to develop code packages that house these algorithms for bioinformatic use, and the third is to develop user-friendly graphical user interface (GUI) software for forensic technicians to use.

Examples of specific areas where these algorithms and resulting software could be developed include (1) random match probability calculation as derived from DNA databases using inferred non-synonymous single nucleotide polymorphism (nsSNP) frequencies; (2) false positive/false negative detections; (3) estimation of variation from multiple sources, such as using different sample preparation, instrumentation, peptide identification software, or associated analytical parameters; (4) assigning background distributions to body fluid-specific peptides; and (5) assigning phylogenetic homology metrics to species-specific peptides.

The goal of these research projects would be to automatically generate statistically-grounded reports from common proteomic input data and be amenable with all acquisition methods including data dependent acquisition, data independent acquisition, and targeted methods. This level of research will assist in making forensic proteomic data analysis more user-friendly and improve accessibility to laboratories currently in need of these capabilities.

Keyword(s): Proteomics, algorithms, software, mass spectra, data analysis

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. Jarman, K. H., & Merkley, E. D. (2019). The statistical defensibility of forensic proteomics. In *Applications in forensic proteomics: Protein identification and profiling* (pp. 203-228). American Chemical Society.
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. Woerner, A. E., Hewitt, F. C., Gardner, M. W., Freitas, M. A., Schulte, K. Q., LeSassier, D. S., Baniasad, M., Reed, A. J., Powals, M. E., Smith, A. R., Albright, N. C., Ludolph, B. C., Zhang, L., Allen, L. W., Weber, K., & Budowle, B. (2020). An algorithm for random match probability calculation from peptide sequences. *Forensic science international. Genetics*, *47*, 102295. https://doi.org/10.1016/j.fsigen.2020.102295
4. Yang, H., Zhou, B., Deng, H., Prinz, M., & Siegel, D. (2013). Body fluid identification by mass spectrometry. *International journal of legal medicine*, *127*(6), 1065–1077. https://doi.org/10.1007/s00414-013-0848-1

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?

Since there is no easy-to-use commercially or publicly available software for the forensic interpretation of proteomic data, current laboratory capabilities are limited by the extent of prior work they or others have conducted. The ability of a laboratory to define and interpret statistical framework is also typically limited to whatever forensic application they are interested in. The development of algorithms and software for forensic proteomics would dramatically improve capabilities of existing forensics laboratories as it would remove barriers of laboratories currently unable to employ these methods and create equivalency in data interpretation between different laboratories.

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

Positive research results would improve the understanding of the scientific basis of forensic proteomic mass spectrometry by demonstrating the abilities and limitations of each of the multiple scientific disciplines involved in proteomic mass spectrometry (biology, biochemistry, genetics and population statistics) in evaluating and establishing standards for forensic data analysis and interpretation.

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

The largest impact of this type of research would be in assisting crime laboratories to successfully interpret proteomic data. Currently there are no software packages that generate forensic reports. As a consequence, data analysis and subsequent interpretation is a large impediment on laboratories looking to implement these methods. Having accessible software for application in either body fluid identification, species identification, or human identification will allow for the potential growth of forensic proteomics. Further, accessible software would be especially impactful if the resulting algorithms and software are capable of producing output from common proteomic data analysis platforms including both targeted and non-targeted data.

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.