

From data to discovery: The essential role of computational tools in proteomics

In the ever-evolving landscape of scientific inquiry, the saying “software is eating the world,” popularized in Silicon Valley over a decade ago, rings truer than ever before. This aphorism, initially indicative of the transformative power of software in reshaping industries and everyday life, has found a significant echo in the realm of science. Akin to a master chef who artfully combines a variety of raw ingredients to concoct a delightful meal, in proteomics, bioinformatics serves as the critical skill set that distills complex, raw data into digestible, insightful knowledge. This editorial aims to showcase the breadth of innovation and inquiry encapsulated in this special issue of *Proteomics*, dedicated to computational mass spectrometry and proteomics, and underline the indispensable role of advanced computational tools in deciphering the molecular intricacies of life itself.

Proteomics research, a cornerstone of ‘omics studies, provides a panoramic view into the molecular and cellular mechanisms underpinning life. Through the analysis of proteins, their structures, functions, and interactions with various molecules, proteomics endeavors to unravel the complex molecular tapestry of biological systems. The manuscripts featured in this special issue illuminate the wide scope of scientific knowledge that can be gleaned from proteomics experiments, made possible only through the employment of sophisticated computational tools and bioinformatics analyses.

Echoing recent advancements in artificial intelligence, several papers in this issue delve into the application of machine learning tools for enhancing the analysis of mass spectrometry-based proteomics data. For instance, Adams et al. offer a comprehensive review on utilizing predicted peptide properties like spectral similarity, retention time, and ion mobility features to refine immunopeptidomics data analysis [1]. In a similar vein, Siraj et al. discuss the enhancement of protein–nucleic acid cross-links detection through the prediction of fragment ion intensities and retention time [2]. Peptide property prediction, a task that has become increasingly commonplace in recent years, enables accurate and sensitive rescoring of spectrum assignments in bottom-up proteomics data. The contributions in this special issue demonstrate that this strategy is particularly potent in realms that exhibit non-standard and highly complex spectral data, such as immunopeptidomics and protein–RNA crosslinking mass spectrometry.

Further, Joyce and Searle’s review on computational approaches for phosphoproteomics identification and localization presents the future potential of using predicted peptide properties for interpreting phosphopeptide positional isomers and disambiguating chimeric spectra containing multiple isomeric peptides that differ only in the

phosphorylation location [3]. Additionally, Picciani et al. introduce the Oktoberfest tool, leveraging the Prosit peptide property prediction model to create simulated spectral libraries and rescore peptide–spectrum matches, thereby providing a convenient tool to use these predictions for extracting more information from mass spectrometry experiments [4].

This issue also highlights manuscripts that explore the varied applications of computational approaches in mass spectrometry. Gomez-Zepeda et al. present the HowDirty software, a tool for evaluating contamination in mass spectrometry data, emphasizing the critical, yet often overlooked, aspect of quality control [5]. Agten et al. delve into the prediction of isotopic abundances with a highly accurate statistical model, addressing another underappreciated aspect of mass spectrometry data analysis [6]. Carr et al. introduce a novel spectral averaging algorithm that enhances the signal-to-noise ratio in MS1 spectra for top-down proteomics, leading to improved identification of proteoforms [7]. Lastly, Lin et al. tackle the challenging issue of false discovery rate control in proteomics data analysis, comparing various strategies that have been implemented in their Crema tool [8].

The contributions within this special issue underscore an important trend: the transition of complex algorithmic advancements into open-source and user-friendly software implementations for the scientific community. This endeavor, often underrecognized in the traditional scientific environment that typically rewards novel discoveries over software development, merits profound appreciation. The authors’ efforts in crafting graphical user interface and command-line tools exemplify a commendable dedication to making computational tools accessible and usable, bridging the gap between advanced algorithms and practical applications.

In conclusion, the manuscripts in this special issue not only introduce groundbreaking computational methods but also embody the collaborative spirit and openness that define the computational proteomics community. It is with great pride that I have read these contributions, which not only advance our understanding of the proteome but also reinforce the notion that, much like in broader society, software indeed is eating the world in science. Just as a chef skillfully combines ingredients to create a dish greater than the sum of its parts, bioinformatics and computational tools blend complex data into a coherent understanding of life’s molecular foundations. This special issue stands as a testament to the power of computational mass spectrometry and proteomics in advancing our quest for knowledge, underscoring the transformative impact of software in the scientific world.

Wout Bittremieux 

Department of Computer Science, University of Antwerp, Antwerpen,
Belgium

Correspondence

Wout Bittremieux, Department of Computer Science, University of
Antwerp, Antwerpen, Belgium.

Email: wout.bittremieux@uantwerpen.be

ORCID

Wout Bittremieux  <https://orcid.org/0000-0002-3105-1359>

REFERENCES

1. Adams, C., Laukens, K., Bittremieux, W., & Boonen, K. (2023). Machine learning-based peptide-spectrum match rescoring opens up the immunopeptidome. *Proteomics*, 2300336. <https://doi.org/10.1002/pmic.202300336>
2. Siraj, A., Bouwmeester, R., Declercq, A., Welp, L., Chernev, A., Wulf, A., Urlaub, H., Martens, L., Degroev, S., Kohlbacher, O., & Sachsenberg, T. (2024). Intensity and retention time prediction improves the rescoring of protein-nucleic acid cross-links. *Proteomics*, unpublished.
3. Joyce, A. W., & Searle, B. C. (2023). Computational approaches to identify sites of phosphorylation. *Proteomics*, 2300088. <https://doi.org/10.1002/pmic.202300088>
4. Picciani, M., Gabriel, W., Giurcoiu, V.-G., Shouman, O., Hamood, F., Lautenbacher, L., Jensen, C. B., Müller, J., Kalhor, M., Soleymaniyi, A., Kuster, B., The, M., & Wilhelm, M. (2023). Oktoberfest: Open-source spectral library generation and rescoring pipeline based on ProSIT. *Proteomics*, 2300112. <https://doi.org/10.1002/pmic.202300112>
5. Gomez-Zepeda, D., Michna, T., Ziesmann, T., Distler, U., & Tenzer, S. (2024). HowDirty: An R package to evaluate molecular contaminants in LC-MS experiments. *Proteomics*, 2300134. <https://doi.org/10.1002/pmic.202300134>
6. Agten, A., Vilenne, F., Prostko, P., & Valkenburg, D. (2023). A compositional data model to predict the isotope distribution for average peptides using a compositional spline model. *Proteomics*, 2300154. <https://doi.org/10.1002/pmic.202300154>
7. Carr, A. V., Bollis, N., Pavek, J. G., Shortreed, M. R., & Smith, L. M. (2024). Spectral averaging with outlier rejection algorithms to increase identifications in top-down proteomics. *Proteomics*, unpublished.
8. Lin, A., See, D., Fondrie, W. E., Keich, U., & Noble, W. S. (2024). Target-decoy false discovery rate estimation using Crema. *Proteomics*, unpublished. <https://doi.org/10.1002/pmic.202300084>