



University of Southern Denmark

## Biosaur

### An open-source Python software for liquid chromatography–mass spectrometry peptide feature detection with ion mobility support

Abdrakhimov, Daniil A; Bubis, Julia A; Gorshkov, Vladimir; Kjeldsen, Frank; Gorshkov, Mikhail V; Ivanov, Mark V

*Published in:*  
Rapid Communications in Mass Spectrometry

*DOI:*  
10.1002/rcm.9045

*Publication date:*  
2025

*Document version:*  
Accepted manuscript

*Citation for pulished version (APA):*  
Abdrakhimov, D. A., Bubis, J. A., Gorshkov, V., Kjeldsen, F., Gorshkov, M. V., & Ivanov, M. V. (2025). Biosaur: An open-source Python software for liquid chromatography–mass spectrometry peptide feature detection with ion mobility support. *Rapid Communications in Mass Spectrometry*, 39(S1), Article e9045.  
<https://doi.org/10.1002/rcm.9045>

Go to publication entry in University of Southern Denmark's Research Portal

#### Terms of use

This work is brought to you by the University of Southern Denmark.  
Unless otherwise specified it has been shared according to the terms for self-archiving.  
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.  
Please direct all enquiries to [puresupport@bib.sdu.dk](mailto:puresupport@bib.sdu.dk)

# Biosaur: An open-source Python software for LC/MS1 peptide feature detection with ion-mobility support

Daniil A. Abdrakhimov<sup>1,2</sup>, Julia A. Bubis<sup>1</sup>, Vladimir Gorshkov<sup>3</sup>, Frank Kjeldsen<sup>3</sup>, Mikhail V. Gorshkov<sup>1</sup>, Mark V. Ivanov<sup>1\*</sup>

<sup>1</sup>V.L. Talrose Institute for Energy Problems of Chemical Physics, N. N. Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, 38 Leninsky Pr., Bld. 2, Moscow, 119334, Russia

<sup>2</sup>Moscow Institute of Physics and Technology National Research University G. Dolgoprudny, Institutsky Lane 9, Dolgoprudnyj, RU 141701, Russia

<sup>3</sup>Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark

\*Corresponding author email: markmipt@gmail.com

## ABSTRACT

**RATIONALE:** One of the important steps in initial data processing of peptide mass spectra is the detection of peptide features in full-range mass spectra. Ion mobility offers advantages over previous methods performing this detection by providing additional structure-specific separation dimension. However, there is a lack of open-source software, which utilizes these advantages and detects peptide features in mass spectra acquired along with the ion mobility data using new instruments such as timsTOF and/or FAIMS-Orbitrap.

**METHODS:** Recently, a utility called Dinosaur was presented, which provides an efficient way for feature detection in peptide ion mass spectra. In this work we extended its functionality by developing Biosaur software to fully employ the additional information provided by ion mobility data. Biosaur was developed using the Python 3.8 programming language.

**RESULTS:** Biosaur supports the processing of data acquired using mass spectrometers with ion mobility capabilities, specifically timsTOF and FAIMS. In addition, it processes mass spectra obtained in negative ion mode and reports cosine correlation table for peptide features which is useful for differentiation between in-source fragments and semi-tryptic peptides.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/rcm.9045](https://doi.org/10.1002/rcm.9045)

**CONCLUSIONS:** Biosaur is a utility for detecting peptide features in LC/MS spectra with ion mobility and negative ion supports. The software is distributed with an open source APACHE 2.0 licence and freely available on Github at the following link <https://github.com/abdrakhimov1/Biosaur>.

Keywords: proteomics, mass spectrometry, feature detection, algorithm, software, ion mobility.

## INTRODUCTION

Quantitative analysis of complex protein mixtures is one of the main goals of mass spectrometry-based proteomics. The main challenge in obtaining accurate quantitation results is the complexity of the proteomic data. Indeed, the cellular proteome contains hundreds of thousands of proteins.<sup>1</sup> In the most popular realization of the proteome-wide analysis, the so-called bottom-up proteomics, these proteins are enzymatically digested, thus raising the sample complexity up to millions of distinct proteolytic peptide sequences per probe. To reduce this complexity, peptide separation is typically employed prior to mass spectrometry analysis.<sup>2,3</sup> To increase further the depth of proteome characterization the use of several orthogonal separation methods within the one experimental run has become a common approach.<sup>4</sup> Liquid chromatography in combination with ion mobility is one of such approaches that has become popular in recent years.<sup>5,6,7,8,9</sup> Ion mobility offers a number of advantages. First, it gives a third dimension of separation in addition to retention times and peptide masses, which further stretches the sample complexity in time. Second, numerous efforts and investments by the community in recent years have made the ion mobility technology an efficient and routinely utilized add-on option for a modern high-resolution mass spectrometer with a number of commercially successful realizations.<sup>10,11</sup> While the orthogonality of the ion mobility data with, e.g. peptide masses is limited,<sup>12,13</sup> it is evident that the additional separation of ions by the ion mobility factor allows high resolution spectra to be obtained at

significantly reduced noise level.<sup>14,15</sup> Despite of this, the mass spectra still contain significant number of peaks which cannot be unambiguously differentiated into separate peptide features. The accuracy of the analysis depends largely on the outcome of a number of stages in the experimental data processing workflow from raw data handling down to the final post-search validation step. The former is crucially important. Indeed, the aim of this initial step is to detect peptide features in full-range mass spectra. Here, a peptide feature is a set of peaks in a mass spectrum which corresponds to the isotopic and chromatographic trace of the same peptide. The feature detection involves typically the ion intensities, measured retention times,  $m/z$  values, and the ion mobility data. Correctly, detected peptide features are further required in all steps of the data analysis including peptide identification and protein quantitation.

A number of feature detection algorithms have been proposed each having its own pros and cons.<sup>16,17,18,19,20,21</sup> Among them, the Dinosaur<sup>22</sup> algorithm is reportedly the most efficient. However, none of the recently proposed algorithms and developed feature detection tools takes into account the ion mobility data. In this work we attempted to fill this gap by developing a utility called Biosaur. The utility takes advantages of the Dinosaur algorithm with the latter's efficiency in feature detection, and has a number of added functionalities, including ion mobility and negative ion data support.

## METHODS

The Biosaur performance has been evaluated using publicly available (ProteomeXchange dataset PXD010012) timsTOF data.<sup>7</sup> The data were obtained for a 200-ng HeLa mammalian protein digest sample using a 120-min LC gradient and a timsTOF Pro mass spectrometer operated in Parallel

Accumulation-Serial Fragmentation (PASEF) mode.

Negative ion mode tests were performed for data obtained using ultra-short 5-min LC gradients (Dionex UltiMate 3000 HPLC system, Thermo Fisher Scientific, Waltham, MA, USA) and a high resolution Orbitrap FTMS mass spectrometer (Q-Exactive HF, Thermo Fisher Scientific). MS1 spectra of 200 ng of HeLa were acquired at a mass resolving power of 60,000 at  $m/z$  200 in both positive and negative ion modes. Details of the experimental setup were described elsewhere.<sup>23</sup> 5-min LC-MS1 and LC/MS/MS data obtained for HeLa in the latter study were used for in-source fragmentation and semi-tryptic peptide searches, as well as for the Biosaur performance comparison with other feature detection alternatives. The IdentiPy search engine<sup>24</sup> with default settings and an enabled semi-tryptic peptide search option was used. Biosaur was employed for MS1 feature detection.

## RESULTS

**Software design.** The principle of the Biosaur algorithm is shown schematically in **Figure 1**. It takes the input files in mzML format and provides the output files with detected peptide features in \*.tsv (tab-separated values) format. Contrary to the many feature detection algorithms that process peptide MS1 spectra one-by-one, Biosaur is based on the so-called “hill” concept introduced by Teleman et al.<sup>22</sup> The algorithm’s work starts with combining centroid ion peaks into hills, which are the groups of peaks with similar  $m/z$  values in consecutive MS1 scans. All peaks in the mass spectra from the first scan are converted into the constructed hills. After that, all peaks from the second scan are added to the current hills under construction if the difference between the hill’s  $m/z$  value and the added peak’s  $m/z$  value is within the user defined mass measurement

accuracy (MMA). The MMA is 8 ppm by default. The hill's  $m/z$  value is calculated as the average of the  $m/z$  values of the hill's peaks from the last 3 scans. If several peaks in the subsequent scan have  $m/z$  values close to that of one of the hills under construction, only the peak with the smallest mass difference is added to that hill. All peaks with  $m/z$  values far (by MMA) from the hills under construction are used to start new hills. After peak matching the Biosaur finds hills that were not updated with the new peaks for the two previous scans. These hills are then added to the group of finished hills and do not participate further in the next peak matching steps. The above procedure continues through all scans till the end of the run file. Upon reaching the last scan, all hills under construction are converted into the final group of finished hills. Finally, all finished hills with scan lengths less than the user defined minimum (3 scans by default) are discarded. As a result, the Biosaur collects a list of finished hills with the corresponding information about the peaks added to each of them, including  $m/z$  values, peak intensities and scan ids. It also calculates the hill  $m/z$  values as the weighted by intensity average of the  $m/z$  values of the peaks forming a particular hill.

### Figure 1.

The hills are gathered into  $^{13}\text{C}$  isotope clusters in the next stage of the algorithm's workflow. Biosaur detects potential isotopic clusters using cosine similarities between the RT profiles of monoisotopic and isotopic hills, cluster peaks abundances, masses, and the averaging model<sup>25</sup>. First, Biosaur goes through all the hills and determines those which have similar RTs and a mass shift equal to  $N \cdot 1.00335/\text{charge}$ , where  $N$  is the  $^{13}\text{C}$  isotope number from 1 up to 10 and charge is a potential charge from a user-defined range. User-defined thresholds for the minimal RT profile cosine correlation and mass accuracy are used. Biosaur does not check the  $N+1$  isotope if the  $N$

isotope was not found. The similarity of isotopic patterns is calculated using cosine correlation between theoretical and experimental intensity profiles. The averaging model is used for calculation of the peptide isotopes and the corresponding isotopic distributions. The  $^{13}\text{C}$  isotopic distribution is calculated by assuming that every 111.1254 Da contains 4.9384 carbon atoms. To reduce the number of false positive matches, cosine similarities are evaluated by filling zero values for the non-overlapping scans between the hills. Intensity weighted  $m/z$  values across all scans are reported as the monoisotopic mass-to-charge ratios of the finally determined peptide features. In addition, Biosaur reports intermediate metrics, such as the  $m/z$  values for all isotopic peaks, coefficients of variations between scans for each of these  $m/z$  values, experimental isotopic intensity distributions, etc. These values can be later used as the metrics of detected feature quality. For example, one can more accurately re-calculate correlation between theoretical and experimental isotopic intensity distributions during the following post-search analysis of the matching between peptide sequences and the feature. Note that an accurate theoretical profile can be used instead of the averaging approximation in that latter step.

### Figure 2.

At the third stage of the Biosaur's workflow, all potential isotopic clusters are sorted in decreasing order by the sum of the numbers of visible isotopes and cosine correlation for the corresponding average molecule (**Figure 2**). A list of all detected isotopic clusters is generated one-by-one and all hills that contributed to the already added clusters are removed from further consideration.

The key functionalities of the Biosaur algorithm are shown in Table 1. Biosaur has a *targeted* mode, in which it matches the results of identification of MS/MS spectra to the peptide

features. Current Biosaur version also supports X!Tandem,<sup>26</sup> Identipy,<sup>24</sup> MSFragger,<sup>27</sup> and Comet,<sup>28</sup> search engine outputs in pepXML format, as well as MSGF+<sup>29</sup> output in mzID format.

**Biosaur performance.** The evaluation of feature detection methods is usually focused mostly on the efficiency of the algorithm in extracting accurate MS1-based intensities or XICs for already identified MS/MS spectra. On the contrary, our main focus was on the efficiency in extracting all visible peptide features, with a minimal number of noise features at the same time. Direct estimation of how well the algorithm performs this task is not feasible and, thus, we employed the search engine that identifies proteins from the peptide features detected in MS1 spectra. We can then estimate the performance of the feature detection algorithm using the number of identified proteins. Recently, we introduced the DirectMS1 method for LC/MS1-only protein identification,<sup>23</sup> which is based on the comprehensively evaluated search engine ms1searchpy that identifies proteins from MS1-based peptide features. The experimental data were acquired in 3 technical replicates for 500 ng of HeLa using the Orbitrap Q Exactive HF in LC/MS1 mode and 5 min reversed-phased HPLC gradient separation. Three MS1 feature finders, Dinosaur, OpenMS and Biosaur, were used to process these data for comparison. Biosaur and Dinosaur feature detection parameters were default ones. The OpenMS centroided feature finder default parameters were slightly changed to make them closer to both the Biosaur and Dinosaur ones: maximal precursor charge and  $m/z$  accuracy were set to 6 and 0.02, respectively. The number of detected peptide features and the number of identified protein groups at 1% FDR by the DirectMS1 method are shown in Figure 3. Biosaur slightly outperforms Dinosaur and significantly outperforms OpenMS in both metrics. The data processing time for a single replicate on the same server was 1, 2 and 3 minutes for the Dinosaur, Biosaur and OpenMS centroided feature finder, respectively. Functionality differences between the compared tools are listed in Table1.

### Figure 3.

**Negative mode MS.** Negative ion mass spectrometry is considered as an efficient addition in many biological applications including proteomics, in which ionization in positive ion mode is traditionally the main way to study proteins. However, negative ion mode may present a valuable alternative, e.g. in phosphoproteomics and glycopeptides analyses.<sup>30,31</sup> Another useful feature of negative ion mass spectrometry of peptides is a reportedly lower noise level in the mass spectra.<sup>32</sup> However, the majority of proteomic research is still performed in positive mode due to the low efficiency of collisional fragmentation of negative peptide ions<sup>33,34,35</sup>. However, a number of approaches to characterize proteomes do not rely on tandem MS. For example, Accurate Mass and Time (tag)<sup>36</sup> and different MS1-only strategies<sup>37,38,39</sup> are based on the analysis of MS1 spectra, for which negative ion mode can be as efficient as the positive one. Combining positive and negative modes of mass spectra acquisition can improve the information content about peptide sequences and help further in, e.g. localization of modifications. Detection of negatively charged peptide features has been implemented in Biosaur. The data set obtained from LC/MS analysis of 200 ng of HeLa digest was used for testing this functionality. Biosaur successfully detected 72615 and 81152 peptide features in positive and negative modes of peptide ion mass spectra acquisition, respectively.

**Ion mobility.** The primary focus of Biosaur development was extending its capabilities to the processing of ion mobility data. Currently, this extension allows processing data acquired using two recently introduced instruments, timsTOF Pro (Bruker Daltonik, Bremen, Germany) and the Orbitrap FTMS instrument with a FAIMS Pro interface (Thermo Fisher Scientific). In addition to

the workflow shown in Figure 1, MS1 spectra are grouped by different compensation voltage (CV) values when processing FAIMS information, and the features are then detected using the Biosaur's algorithm for every group as described above. The CV values are reported for detected peptide features in the output tables. Handling timsTOF data starts with combining ion peaks into the hills using ion mobility values. Ion peaks in consecutive MS scans are combined into hills only if the  $m/z$  and ion mobility values are close to each other (by default, 8 ppm and 0.1, respectively). The ion mobility values are also used to identify the monoisotopic peak among the possible candidates in the cluster. Biosaur successfully detected 115 000 peptide features in timsTOF data used in this work (see Methods section).

**Correlation matrix.** Another important feature of Biosaur is its ability to calculate the correlation matrix of retention time profiles for peptide features. Biosaur compares each detected feature with the rest and calculates cosine similarity between RT profiles. When the correlation value is greater than 0.5, it is reported to the correlation matrix. One of the possible applications of the matrix is differentiation between in-source fragments and semi-tryptic peptides. Semi-tryptic peptides are the results of proteolytic activities and may present important biological insights on the cell's functioning. On the contrary, in-source fragmentation is an artifact. In-source fragments are typically determined by comparing the retention times at which MS/MS spectra are recorded, or the retention times at the apexes of XICs corresponding to MS1 spectra for two peptide identifications in question. Simply speaking, for the pair of peptides PEPTIDER and TIDER, the latter will be reported as semi-tryptic if the retention time apexes are 10 and 20 min, and as in-source fragments if the retention times are 20 and 20 min. While this approach can work in such a trivial case, there can be more ambiguous situations, when retention times are close to each other,

as shown in Figure 4.

#### **Figure 4.**

Indeed, the RT profiles shown in Figure 4 have close retention time apexes and the peptides will probably be reported as in-source fragments if the above simple approach is used. However, the cosine correlations in cases shown in panels (A) and (C) are close to zero and show that these features are not in-source fragments.

#### **CONCLUSIONS**

Biosaur is a utility for detection of peptide features in LC/MS1 spectra. It has more functionalities than its predecessor Dinosaur. Specifically, its current version supports ion mobility data and negative mode MS, and reports additional information for peptide features which can be used for distinguishing semi-tryptic peptides from in-source fragments. Biosaur was developed in Python 3 programming language, is distributed under open source APACHE 2.0 licence, and is freely available on Github: <https://github.com/abdrakhimov1/Biosaur>.

#### **ACKNOWLEDGMENTS**

The work on Biosaur development, software performance evaluation and data processing was funded by the Russian Science Foundation, grant # 20-14-00229 to M.V.G. Experimental data were obtained at the University of Southern Denmark with support from the VILLUM Foundation (grant # 7292) and PRO-MS: Danish National Mass Spectrometry Platform for Functional Proteomics (grant # 5072-00007B).

## REFERENCES

1. Aebersold R, Agar JN, Amster IJ, et al. How many human proteoforms are there? *Nat Chem Biol.* 2018;14(3):206-214. doi:10.1038/nchembio.2576
2. Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann Appl Stat.* 2010;4(4):1797-1823. doi:10.1214/10-AOAS341
3. Xie F, Smith RD, Shen Y. Advanced proteomic liquid chromatography. *J Chromatogr A.* 2012;1261:78-90. doi:10.1016/j.chroma.2012.06.098
4. Zhang X, Fang A, Riley CP, Wang M, Regnier FE, Buck C. Multi-dimensional Liquid Chromatography in Proteomics. *Anal Chim Acta.* 2010;664(2):101-113. doi:10.1016/j.aca.2010.02.001
5. Hinzke T, Kouris A, Hughes R-A, Strous M, Kleiner M. More Is Not Always Better: Evaluation of 1D and 2D-LC-MS/MS Methods for Metaproteomics. *Front Microbiol.* 2019;10:238. doi:10.3389/fmicb.2019.00238
6. Burnum-Johnson KE, Zheng X, Dodds JN, et al. Ion Mobility Spectrometry and the Omics: Distinguishing Isomers, Molecular Classes and Contaminant Ions in Complex Samples. *Trends Anal Chem.* 2019;116:292-299. doi:10.1016/j.trac.2019.04.022
7. Meier F, Brunner A-D, Koch S, et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol Cell Proteomics.* 2018;17(12):2534-2545. doi:10.1074/mcp.TIR118.000900
8. Zheng X, Wojcik R, Zhang X, et al. Coupling Front-End Separations, Ion Mobility Spectrometry, and Mass Spectrometry For Enhanced Multidimensional Biological and Environmental Analyses. *Annu Rev Anal Chem.* 2017;10(1):71-92. doi:10.1146/annurev-anchem-061516-045212
9. D'Atri V, Causon T, Hernandez-Alba O, et al. Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry? *J Sep Sci.* 2018;41(1):20-67. doi:10.1002/jssc.201700919
10. Garabedian A, Benigni P, Ramirez CE, et al. Towards Discovery and Targeted Peptide Biomarker Detection Using nanoESI-TIMS-TOF MS. *J Am Soc Mass Spectrom.* 2018;29(5):817-826. doi:10.1007/s13361-017-1787-8
11. Bekker-Jensen DB, Martínez-Val A, Steigerwald S, et al. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol Cell Proteomics.* 2020;19(4):716-729. doi:10.1074/mcp.TIR119.001906
12. Morris CB, Poland JC, May JC, McLean JA. Fundamentals of Ion Mobility-Mass Spectrometry for the Analysis of Biomolecules. *Methods Mol Biol.* 2020;2084:1-31.

doi:10.1007/978-1-0716-0030-6\_1

13. Michaelievski I, Kirshenbaum N, Sharon M. T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *J Vis Exp.* 2010;(41). doi:10.3791/1985
14. Swearingen KE, Moritz RL. High-field asymmetric waveform ion mobility spectrometry for mass spectrometry-based proteomics. *Expert Rev Proteomics.* 2012;9(5):505-517. doi:10.1586/epr.12.50
15. Pfammatter S, Bonneil E, McManus FP, et al. A Novel Differential Ion Mobility Device Expands the Depth of Proteome Coverage and the Sensitivity of Multiplex Proteomic Measurements. *Mol Cell Proteomics.* 2018;17(10):2051-2067. doi:10.1074/mcp.TIR118.000862
16. Listgarten J, Emili A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics.* 2005;4(4):419-434. doi:10.1074/mcp.R500005-MCP200
17. Röst HL, Sachsenberg T, Aiche S, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods.* 2016;13(9):741-748. doi:10.1038/nmeth.3959
18. Hoopmann MR, Finney GL, MacCoss MJ. High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. *Anal Chem.* 2007;79(15):5620-5632. doi:10.1021/ac0700833
19. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008;26(12):1367-1372. doi:10.1038/nbt.1511
20. Gutierrez M, Smith R. XFlow: An algorithm for extracting ion chromatograms. *PLoS One.* 2020;15(10):e0227659. doi:10.1371/journal.pone.0227659
21. Olivon F, Grelier G, Roussi F, Litaudon M, Touboul D. MZmine 2 Data-Preprocessing To Enhance Molecular Networking Reliability. *Anal Chem.* 2017;89(15):7836-7840. doi:10.1021/acs.analchem.7b01563
22. Teleman J, Chawade A, Sandin M, Levander F, Malmström J. Dinosaur: A Refined Open-Source Peptide MS Feature Detector. *J Proteome Res.* 2016;15(7):2143-2151. doi:10.1021/acs.jproteome.6b00016
23. Ivanov MV, Bubis JA, Gorshkov V, et al. DirectMS1: MS/MS-Free Identification of 1000 Proteins of Cellular Proteomes in 5 Minutes. *Anal Chem.* 2020;92(6):4326-4333. doi:10.1021/acs.analchem.9b05095
24. Levitsky LI, Ivanov MV, Lobas AA, et al. IdentiPy: An Extensible Search Engine for

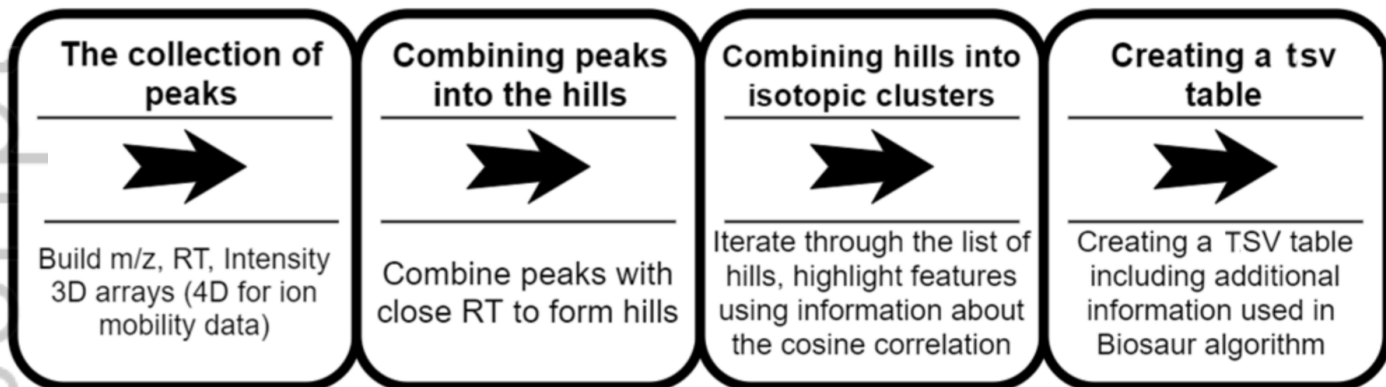
- Protein Identification in Shotgun Proteomics. *J Proteome Res.* 2018;17(7):2249-2255. doi:10.1021/acs.jproteome.7b00640
25. Senko MW, Beu SC, McLafferty FW. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J Am Soc Mass Spectrom.* 1995;6(4):229-233. doi:10.1016/1044-0305(95)00017-8
  26. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics.* 2004;20(9):1466-1467. doi:10.1093/bioinformatics/bth092
  27. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods.* 2017;14(5):513-520. doi:10.1038/nmeth.4256
  28. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics.* 2013;13(1):22-24. doi:10.1002/pmic.201200439
  29. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun.* 2014;5:5277. doi:10.1038/ncomms6277
  30. Engholm-Keller K, Larsen MR. Technologies and challenges in large-scale phosphoproteomics. *Proteomics.* 2013;13(6):910-931. doi:10.1002/pmic.201200484
  31. Leymarie N, Zaia J. Effective use of mass spectrometry for glycan and glycopeptide structural analysis. *Anal Chem.* 2012;84(7):3040-3048. doi:10.1021/ac3000573
  32. Hua Y, Wainhaus SB, Yang Y, et al. Comparison of negative and positive ion electrospray tandem mass spectrometry for the liquid chromatography tandem mass spectrometry analysis of oxidized deoxynucleosides. *J Am Soc Mass Spectrom.* 2001;12(1):80-87. doi:10.1016/S1044-0305(00)00191-4
  33. Steinborner ST, Bowie JH. A comparison of the positive- and negative-ion mass spectra of bio-active peptides from the dorsal secretion of the Australian red tree frog, *Litoria rubella*. *Rapid Commun Mass Spectrom.* 1996;10(10):1243-1247. doi:10.1002/(SICI)1097-0231(19960731)10:10<1243::AID-RCM639>3.0.CO;2-I
  34. Riley NM, Rush MJP, Rose CM, et al. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). *Mol Cell Proteomics.* 2015;14(10):2644-2660. doi:10.1074/mcp.M115.049726
  35. Brinkworth CS, Dua S, McAnoy AM, Bowie JH. Negative ion fragmentations of deprotonated peptides: backbone cleavages directed through both Asp and Glu. *Rapid Commun Mass Spectrom.* 2001;15(20):1965-1973. doi:10.1002/rcm.457
  36. Conrads TP, Anderson GA, Veenstra TD, Pasa-Tolić L, Smith RD. Utility of accurate mass tags for proteome-wide protein identification. *Anal Chem.* 2000;72(14):3349-3354. doi:10.1021/ac0002386

37. Ivanov MV, Tarasova IA, Levitsky LI, et al. MS/MS-Free Protein Identification in Complex Mixtures Using Multiple Enzymes with Complementary Specificity. *J Proteome Res.* 2017;16(11):3989-3999. doi:10.1021/acs.jproteome.7b00365
38. Zimmer JSD, Monroe ME, Qian W-J, Smith RD. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev.* 2006;25(3):450-482. doi:10.1002/mas.20071
39. Moruz L, Hoopmann MR, Rosenlund M, Granholm V, Moritz RL, Käll L. Mass fingerprinting of complex mixtures: protein inference from high-resolution peptide masses and predicted retention times. *J Proteome Res.* 2013;12(12):5730-5741. doi:10.1021/pr400705q

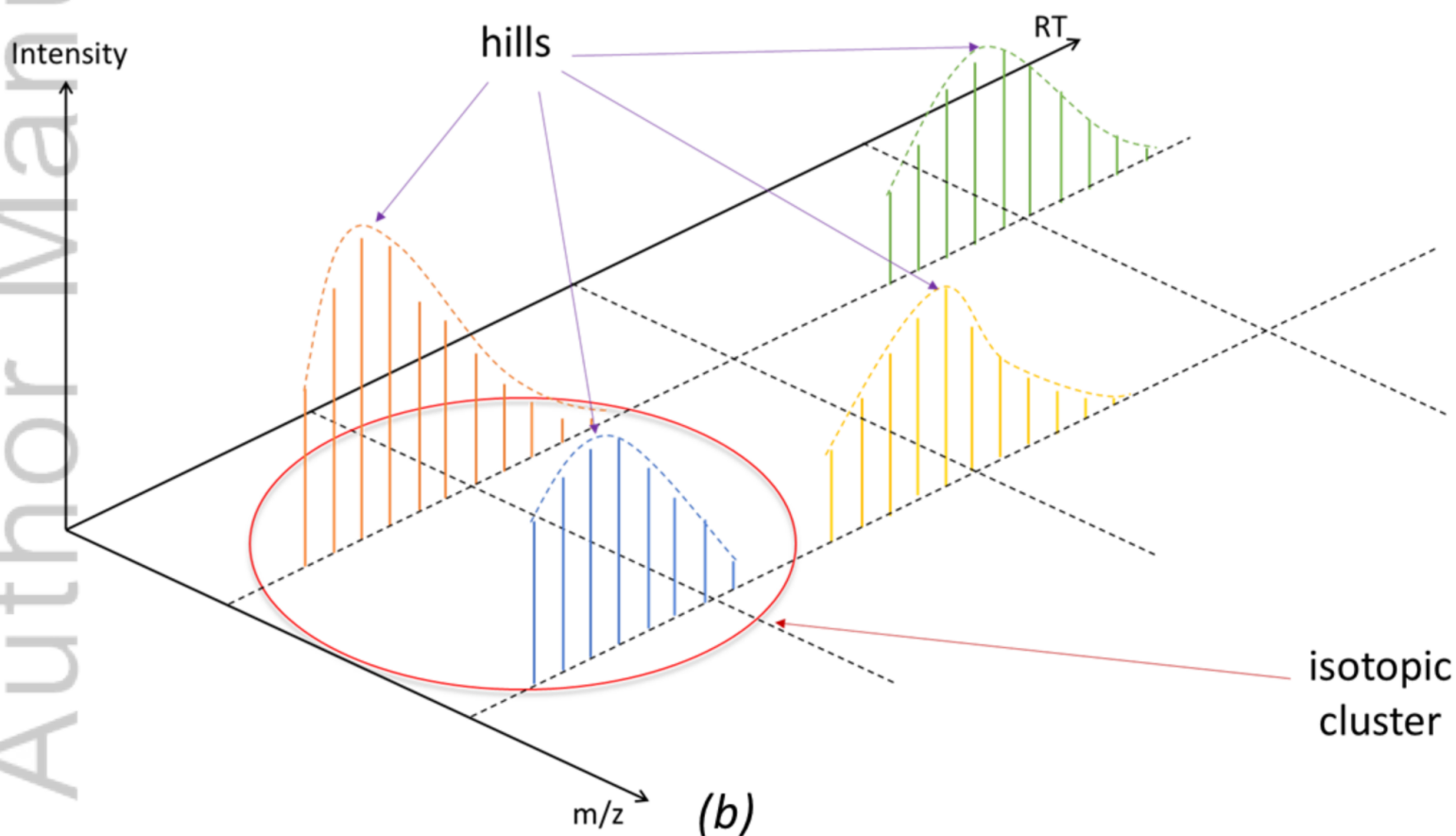
Table 1. Main functional difference between Biosaur and Dinosaur software.

	Biosaur	Dinosaur
Licence	APACHE 2	APACHE 2

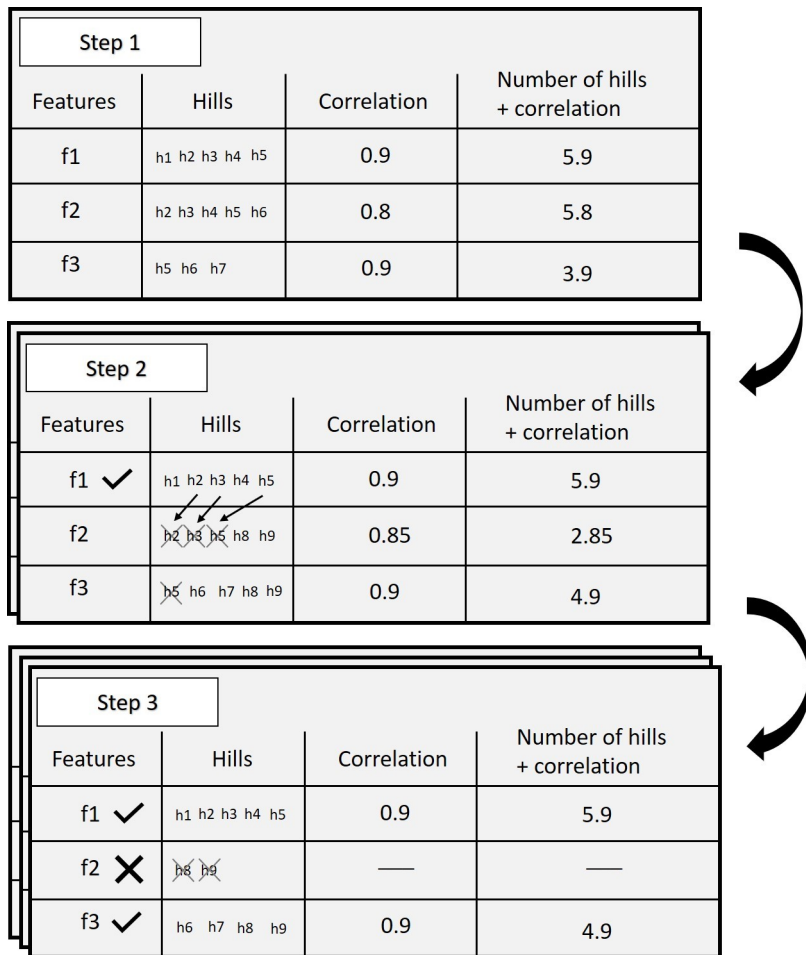
Language	Python 3	Scala
Profile input	-	+
Centroid input	+	+
Targeted mode for peptides identified in MS/MS searches	-	+
TimsTOF data	+	-
FAIMS data	+	-
Negative ion mode support	+	-
Correlation matrix for in-source fragment detection	+	-



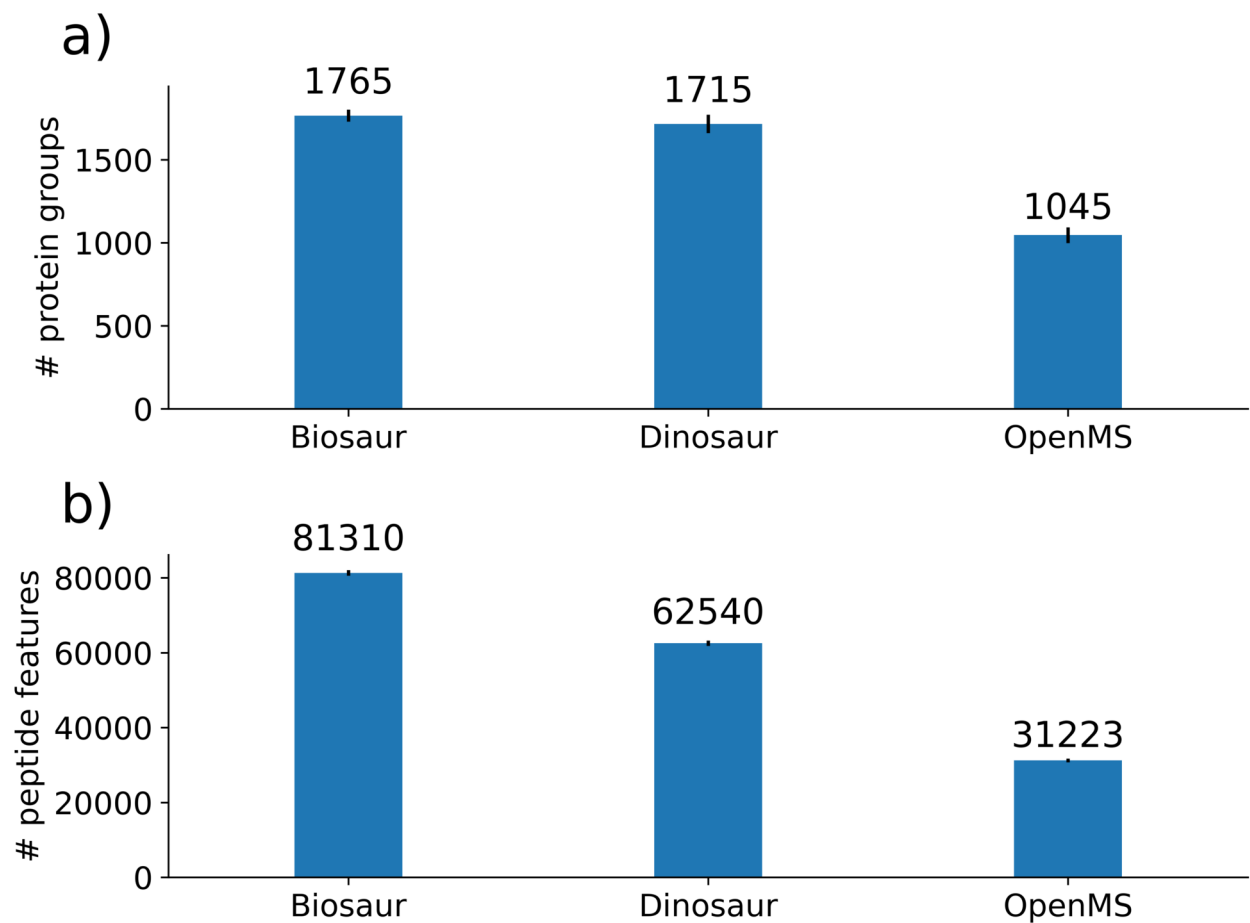
(a)



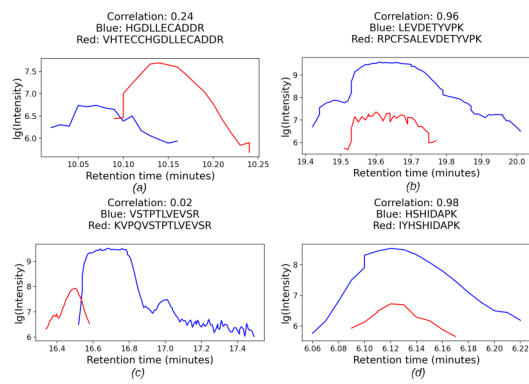
RCM\_9045\_Figure1.tif



RCM\_9045\_Figure2.tif



RCM\_9045\_Figure3.tif



RCM\_9045\_Figure4.tif