

RESEARCH ARTICLE

MS Amanda 2.0: Advancements in the standalone implementation

Viktoria Dorfer¹  | Marina Strobl¹ | Stephan Winkler¹  | Karl Mechtler^{2,3,4} 

¹Bioinformatics Research Group, University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria

²Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, Vienna, 1030, Austria

³Institute of Molecular Biotechnology (IMBA), Austrian Academy of Sciences, Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, Vienna, 1030, Austria

⁴Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, Vienna, 1030, Austria

Correspondence

V. Dorfer, University of Applied Sciences Upper Austria, Bioinformatics Research Group, Softwarepark 11, 4232 Hagenberg, Austria.
Email: viktoria.dorfer@fh-hagenberg.at

Funding information

Austrian Science Fund, Grant/Award Number: ERA-CAPS I 3686; Horizon 2020 Framework Programme, Grant/Award Number: 823839; University of Applied Sciences Upper Austria, Grant/Award Number: Basic Research (PPI-ID)

Rationale: Database search engines are the preferred method to identify peptides in mass spectrometry data. However, valuable software is in this context not only defined by a powerful algorithm to separate correct from false identifications, but also by constant maintenance and continuous improvements.

Methods: In 2014, we presented our peptide identification algorithm MS Amanda, showing its suitability for identifying peptides in high-resolution tandem mass spectrometry data and its ability to outperform widely used tools to identify peptides. Since then, we have continuously worked on improvements to enhance its usability and to support new trends and developments in this fast-growing field, while keeping the original scoring algorithm to assess the quality of a peptide spectrum match unchanged.

Results: We present the outcome of these efforts, MS Amanda 2.0, a faster and more flexible standalone version with the original scoring algorithm. The new implementation has led to a 3–5× speedup, is able to handle new ion types and supports standard data formats. We also show that MS Amanda 2.0 works best when using only the most common ion types in a particular search instead of all possible ion types.

Conclusions: MS Amanda is available free of charge from <https://ms.imp.ac.at/index.php?action=msamanda>.

1 | INTRODUCTION

For decades, mass spectrometry has been known as the primary method to analyze proteins in biological samples.^{1,2} A considerable amount of effort has been spent on instruments, technology and also on algorithm development.^{3–7} Different techniques have evolved to identify peptides in mass spectra from bottom-up mass spectrometry experiments, namely *de novo* identification, database search and spectrum library search. A plethora of different algorithms exist for each analysis category,^{8–10} but despite the increasing popularity of spectrum library search in the last years,^{11–13} database search is still often the method of choice when it comes to identifying peptides in mass spectra.¹⁴

In a database search, each spectrum is compared with a list of peptide candidates from a protein database where the peptide mass matches the precursor mass with a certain tolerance. For each peptide candidate a theoretical spectrum, i.e., all potential fragment ions that could occur in a mass spectrum, is calculated, compared with the experimental spectrum and a score is calculated. The peptide candidate with the highest score is then reported.⁸

The score is an essential part of a search engine, one component that distinguishes different algorithms from each other. In a good search engine, the score for each peptide is constructed in a such a way that false identifications can be discriminated from correct identifications, i.e., the higher the score for a peptide spectrum match (PSM), the more likely the PSM is correct.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Rapid Communications in Mass Spectrometry* published by John Wiley & Sons Ltd.

However, not only a good scoring scheme is essential for a good search engine, but also ease of use and especially maintenance and future development. The scoring scheme of a search engine can be brilliant, but if the code is not maintained and regularly updated to eradicate errors or improve user experience, the algorithm will at some point no longer be used.

In 2014, we published the peptide identification algorithm MS Amanda,¹⁵ which has been accepted and widely used by the proteomics community.^{16–21} Since then, we have worked hard to constantly maintain the software and incorporated user feedback and feature requests, while retaining the original scoring algorithm. In 2018, we released an improved version of MS Amanda available in Thermo Fisher Proteome Discoverer that is able to identify and validate chimeric spectra.²²

In this paper, we summarize our improvements for the standalone version of MS Amanda, namely:

- Increase in search speed
- Support of multiple spectra and database files
- Support of standardized input and output formats
- Support of common ion types in UVPD spectra
- Improvements in usability

2 | METHODS

2.1 | Performance improvements

The first issue we tackled was search speed. In the original version of MS Amanda, it was important to us that the algorithm could run on any machine, independent of the available CPU cores and RAM. Two parameters controlled how many spectra could be processed at once and how many proteins could be searched at the same time, thus defining the speed and – indirectly – the required memory. In addition, already digested protein databases were re-used in subsequent searches – provided that the digestion parameters, i.e., digestion enzyme type or number of missed cleavages, matched. While this is still true for the new version, we changed the way in which digested FASTA files are stored on the hard disk. In contrast to the first version where we used compressed plain text, we now work with binary encodings. In the first version each protein was digested and its peptides stored individually. Although this allowed for fast database digestion, the subsequent file operations to read the digested peptides were identified as a major bottleneck. We changed this implementation and now peptides with the same sequence are grouped and stored only once. Additional mapping files are generated to keep track of the connection between peptides and proteins. Although the grouping and generation of mapping files takes additional time, the decreased number of files that have to be read still significantly reduces the runtime (see section 3).

While these changes have significantly improved search speed, there was still room for improvement on operating systems other than Windows. For us it was essential that MS Amanda runs on all

commonly used operating systems. As MS Amanda is implemented in C#, this was only possible using the mono framework by the time of publication in 2014. While the mono framework was a great way to start, we could still see that the algorithm could not use the full potential of its parallelized implementation on Linux and macOS systems.

In 2016, Microsoft released a new framework, .NET Core, able to run on any operating system. We therefore ported MS Amanda from .NET Framework to .NET Core (which works cross-platform) to make it available on Windows, macOS and Linux without requiring parallel development. We have tested these performance improvements by using three replicates of HeLa cell lysates measured on a Thermo Fisher QExactive+ (PXD007750, Dataset A²²).

In addition, users have reported great results achieved using MS Amanda on phosphorylated data sets and it has frequently been used to identify modified peptides.^{23–25} We therefore analyzed four phospho-enriched HeLa cell lysates from the Chorus Project²⁶ (<https://chorusproject.org/>, identifier 1,374, DDA files only) and compared results from MS Amanda 2.0 with a search engine thoroughly accepted by the community: X!Tandem²⁷ (version 5.0.1). Analyses were performed using SearchGUI²⁸ (version 4.0.18), using a Human Swiss-Prot database (2020-12) including common contaminants. Searches were performed using the following parameters: 10 ppm precursor mass tolerance, 0.02 Da fragment mass tolerance, carbamidomethyl (C) as fixed, oxidation (M) and phosphorylation (S,T,Y) as variable modifications, refinement set to false for X!Tandem. All other parameters were left at defaults. Calculation of the false discovery rate (FDR) was performed within PeptideShaker²⁹ (version 1.16.13).

In addition, we executed comparative performance tests using the HeLa cell lysates also utilized for the runtime analysis (PXD007750, Dataset A²²), applying the same parameters and using the same modification settings except for phosphorylation.

2.2 | Support of multiple spectra and database files

When trying to identify peptides in mass spectra using database search, it is essential to include common contaminants in the list of potential peptide candidates. In the original version of MS Amanda, the algorithm could only handle a single FASTA file. However, these contaminants are normally stored in a separate file, making it necessary to combine the protein database that will be used for the search and the contaminations database prior to starting the search. As this is impractical for users and a possible source of errors, MS Amanda now also accepts a folder containing all FASTA files the spectra should be compared with. The same holds for spectra files. Nowadays, mass spectrometry experiments do not consist of single result files but rather comprise multiple biological and technical replicates or different instrument settings that are compared. We therefore also changed the implementation such that now multiple spectra files can be queued for search at once.

2.3 | Support of standardized input and output formats

Considerable effort has been put in by the HUPO PSI standardization community to guarantee and enhance communication between tools and algorithms by providing standard data formats for mass spectra and its (peptide) identification results, namely .mzML³⁰ and .mzIdentML.^{31,32} We strongly support these efforts as this increases the usability and versatility of algorithms. Providing support for standardized data formats can easily support dissemination of tools and boost utilization of developed tools. We thus enabled MS Amanda to read and write these standard data formats in addition to the file formats supported by the original publication, i.e., .mgf as input file format and .csv as output file format.

2.4 | Support of common ion types in UVPD spectra

The first version of the MS Amanda algorithm supported ions occurring when using CID,³³ HCD,³⁴ ETD,³⁵ and EThcD³⁶ fragmentation. A fragmentation technique that has gained increasing attraction in recent years is ultraviolet photodissociation (UVPD).^{37,38} In addition to the common ions such as a, b, x, y, or z fragments, UVPD also often generates additional fragment ions such as a + 1, c, x + 1, or y – 1 ions.^{39,40} Consideration of these ion types for scoring is now also supported by MS Amanda. The Thermo Fisher Proteome Discoverer version of MS Amanda also features these ion types.

2.5 | Improvements in usability

To enhance usability, we changed the way how to call MS Amanda from the command line by introducing new command line arguments to be able to handle all new features. In addition, the order of input parameters is no longer essential as it was the case for the previous version of MS Amanda. While search parameters are still read from the settings .xml file, parameters such as the file or folder containing spectra, the FASTA file(s) or the desired output format are read as command line parameters (see Table 1 for all available options). Although these named parameters are in contrast to unix command

TABLE 1 All currently available command line parameters for the standalone version. Required parameters are given in bold

Parameter	Description
-s	Spectrum file or spectrum folder (.mgf .mzML)
-d	Protein peptide database file or folder (.fasta)
-e	MS Amanda settings file including all search settings (.xml)
-f	Output file format (1 2), 1: .csv (default), 2: .mzIdentML
-o	Output file name or folder (default: location of spectrum file)

line parameter conventions, where only optional parameters should use option names, we favor this approach due to its higher user-friendliness. To adhere to the Unix conventions we still support the previous command line call.

3 | RESULTS

3.1 | Performance comparison

We compared the search speed of the initial implementation of MS Amanda and of the currently available algorithm on all three operating systems. We used three replicates of HeLa cell lysates measured on a Thermo Fisher QExactive+ (see section 2), and compared runtimes of MS Amanda 1.0 and MS Amanda 2.0 on operating systems Windows, Linux and macOS. We used MS Amanda 1.0 (version 1.0.0.4484) and MS Amanda 2.0 (version 2.0.0.17442) on Windows and MS Amanda 1.0 (version 1.0.0.4485) using the mono framework (version 6.12.) and MS Amanda 2.0 (version 2.0.0.17442) on Linux and macOS, running them on systems with the following specifications:

- Windows 10 Pro, Intel® Xeon E3-1231v3, 3.4 GHz, 8 GB RAM, Samsung SSD 850 EVO
- Ubuntu 20.04, Intel® Xeon E3-1231v3, 3.4 GHz, 8 GB RAM, Samsung SSD 850 EVO
- macOS Big Sur, v. 11.2.1, Quad-Core Intel® Core i7, 2.8 GHz, 16 GB RAM, Apple SSD AP1024M

Linux and Windows were set up as virtual machines and benchmarks were performed on the same server, using libvirt version 5.10.0, pinned to 2 cores and 8 GB RAM each. For the first replicate a new digestion of the protein database was generated; the subsequent two replicates re-used the pre-digested database.

Figure 1 shows that, compared with the original version of MS Amanda, MS Amanda 2.0 runs on average more than three times faster on Windows and macOS systems and almost five times faster on Linux. Runtime comparison between operating systems is however only applicable for Linux and Windows, as they were run on the same machine. Data shown in Figure 1 are average runtimes of each replicate including the database digestion for the first file, i.e., a comparison of a fresh installation of both MS Amanda versions. On average, database digestion accounted for 8% of the total runtime for the old version (60, 63, and 13 seconds on Windows, Linux and Mac, respectively) but increased to on average 42% of the total runtime for the latest version (95, 97, 73 seconds for Windows, Linux, and Mac, respectively). Still, the performance gain through the adapted digestion handling outweighed these losses.

We also investigated the impact of porting MS Amanda from the .NET Framework to .NET Core, which makes the mono framework obsolete. We compared runtimes of the last version of MS Amanda prior to the .NET version change (MS Amanda 2.0 vs 2.0.0.14828) to the current version of MS Amanda. While this leads to a slight increase in runtime on the Windows operating system,

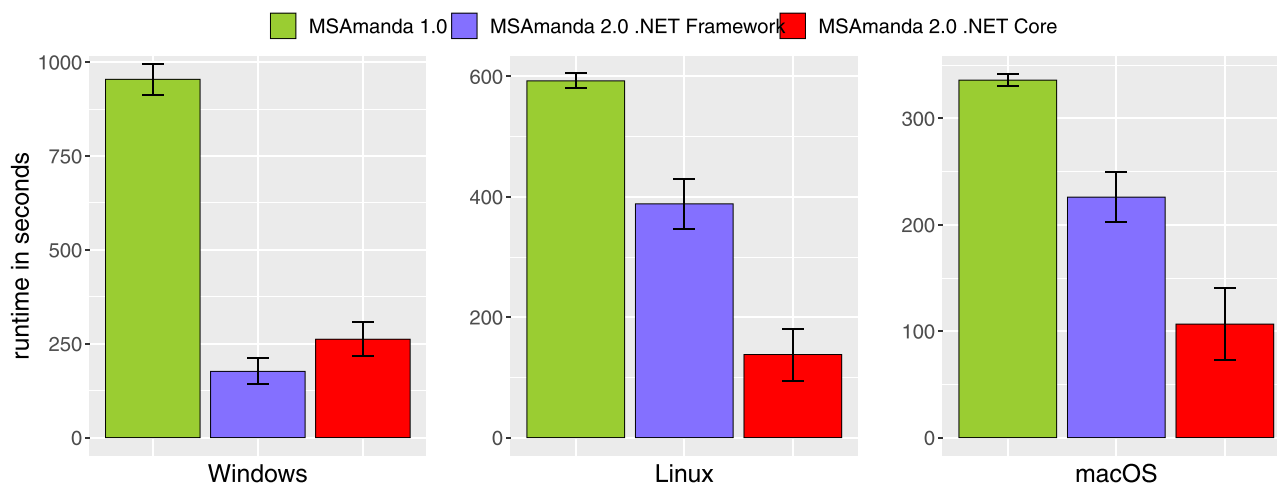


FIGURE 1 Search speed improvements of MS Amanda 2.0 ported to .NET Core (red) compared with the original implementation MS Amanda 1.0 (green) and MS Amanda 2.0 implemented on the .NET Framework (purple) on the operating systems Windows, Linux and macOS. MS Amanda 1.0 and MS Amanda 2.0.NET Framework were run with the help of mono (version 6.12) on Linux and macOS

the .NET Core versions running on Linux and macOS were 2 to 2.5 times faster than those versions using the mono framework. This fact and the advantage of just a single implementation for all operating systems further convinces us that .NET Core was a good choice.

When comparing the number of identified PSMs at 1% FDR of non-phosphorylated and phosphorylated HeLa data sets, we found that for the non-phosphorylated HeLa data set results from MS Amanda and X!Tandem are comparable (12,092 vs 12,179 PSMs at 1% FDR). However, MS Amanda outperforms X!Tandem on the phosphorylated data, as MS Amanda is able to identify on average 13% more PSMs at 1% FDR as compared with X!Tandem (13,175 vs 11,704 PSMs at 1% FDR).

3.2 | Identification results for UVPD spectra

Several groups have reported the common occurrence of $a + 1$, $x + 1$, and $y - 1$ ions in UVPD spectra.^{39,40} We wanted to investigate the applicability of these ion types to be used for scoring and tested various ion settings on HeLa samples measured on a Thermo Fisher QExactive using UVPD peptide fragmentation (PXD003109³⁹). In their manuscript, Fort and co-workers³⁹ compared UVPD and HCD fragmentation techniques and claimed that both techniques generated a comparable number of reliable identifications. Our results support these findings. In addition, the overlap of identified unique peptides between these two techniques matches the outcome of Fort and colleagues³⁹ (see Figure 2). However, we see that the identification quality strongly depends on the ion types considered to compare peptides to spectra. As we have seen during our research of the original MS Amanda publication, the MS Amanda algorithm works best when the most frequently seen ion types are used for scoring, in

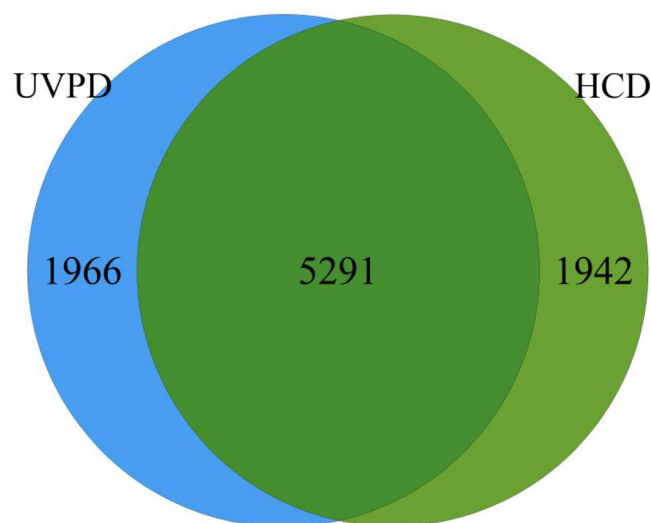


FIGURE 2 Overlap of unique peptides at 1% FDR for UVPD and HCD results of a single replicate

contrast to all potential ion types that might occur. For HCD, e.g., the highest number of identifications can be achieved when using b and y ions only. This is due to the probability score applied in MS Amanda. The more ion types are considered the more potential ion candidates are available – this holds also for random peptides that may lead to false identifications – and therefore the higher the probability to match random peaks by chance.

For UVPD spectra, we see a similar effect. Despite the fact that $x + 1$, $a + 1$, and $y - 1$ ions occur regularly in these spectra, they are still less common than a , b , or y ions. As depicted in Figure 3, using all these ion types that might occur in UVPD spectra decreases the number of identified PSMs at 1% FDR by 15%. Leaving out common

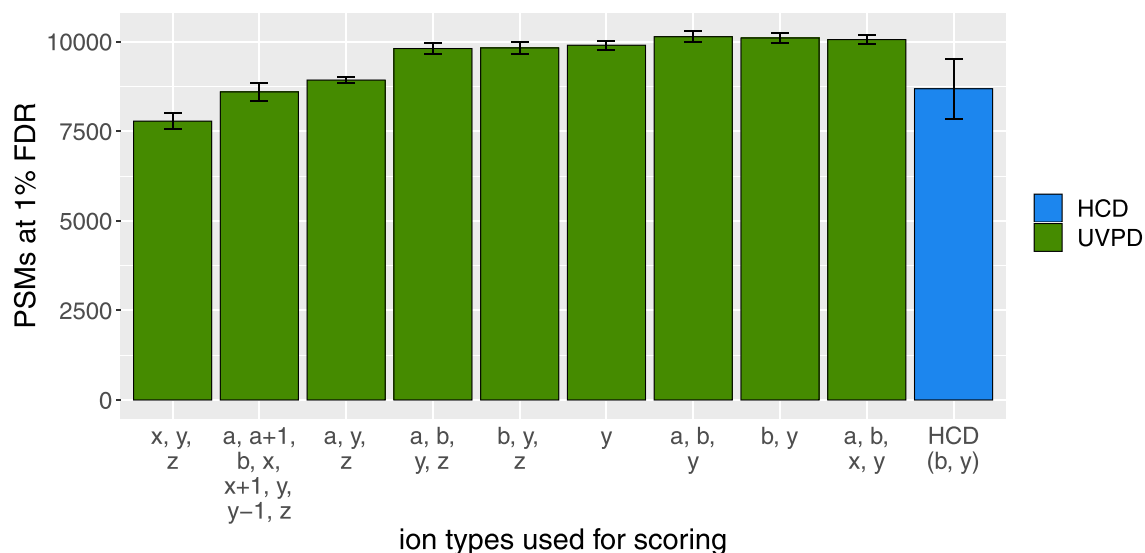


FIGURE 3 Impact of rare ion types: Considering ion types in the score that are rather rare has a huge impact on identification results

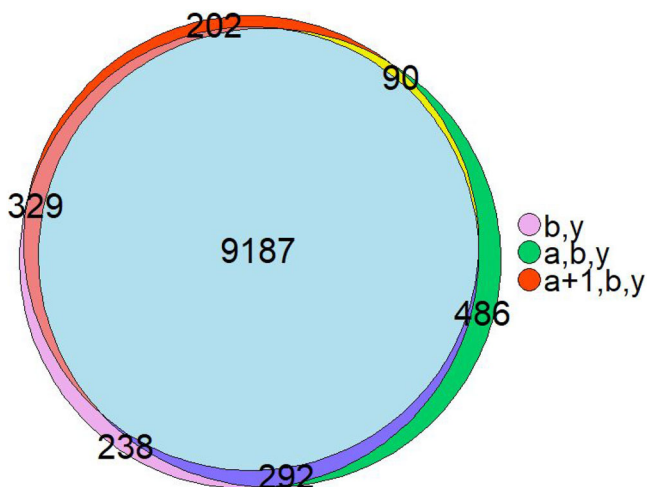


FIGURE 4 Overlap of PSMs at 1% FDR for different ion type settings when searching UVPD spectra. Including or excluding a/a + 1 ions has no significant impact on the search results

ion types, however, is even worse, as this yields 23% less identifications. Therefore, for MS Amanda 2.0 it is best to search only for the most common ion types also in UVPD spectra. We assume this might be similar for other search engines using probability-based scores. In addition, we also compared the identified PSMs at 1% FDR when a and a + 1 ions were included or excluded as ion type. The comparison has been made on a spectrum-by-spectrum basis as proposed by Agten and co-workers.⁴¹ Figure 4 reveals that the difference in identifications for these settings is negligible, indicating that solely b and y ions could be used here as ion types in the search.

4 | CONCLUSIONS

Valuable software in general is not only defined by powerful algorithms but also by continuous maintenance and development. This is of course also true for mass spectrometry software. Several years ago, we published our peptide identification algorithm MS Amanda, showing that we are able to outperform algorithms very frequently used by the community. In this manuscript we show that the MS Amanda implementation has advanced, reacting to user needs and feature requests. The latest standalone version of MS Amanda, MS Amanda 2.0, has numerous improvements compared with the first published version, including increase in search speed, support of multiple FASTA and spectra files, support of standardized formats (.mzML and .mzIdentML), new ion types occurring in UVPD spectra, and usability improvements. In addition, MS Amanda 2.0 has been ported from the .NET Framework to .NET Core 3.1, being able to run on all operation systems without the usage of the mono framework. With that change and other improvements, we have shown that the latest version of MS Amanda is now 3.2–4.3 times faster than the initial version, depending on the operating system used. MS Amanda is now even more flexible and widely applicable to all sorts of mass spectrometry data. The standalone version of MS Amanda can also be used within SearchGUI²⁸ and results can be analyzed using PeptideShaker.²⁹

Of course, the further development of MS Amanda is an ongoing endeavor. We are currently working on supporting chimeric spectra identification published as the CharmRT workflow also in the standalone version. In addition, we are working on an automated pin file generation to be able to validate MS Amanda results with Percolator.⁴²

ACKNOWLEDGEMENTS

Work in the Mechtler lab was financially supported by EPIC-XS, project number 823839 within the Horizon 2020 Framework Programme of the European Union, and the ERA-CAPS I 3686 project of the Austrian Science Fund. Work at the University of Applied Sciences (FH OÖ) was funded by the Basic Research Program of the University of Applied Sciences Upper Austria (PPI-ID).

ORCID

Viktoria Dorfer  <https://orcid.org/0000-0002-5332-5701>

Stephan Winkler  <https://orcid.org/0000-0002-5196-4294>

Karl Mechtler  <https://orcid.org/0000-0002-3392-9946>

REFERENCES

- Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003;422(6928):198-207. <https://doi.org/10.1038/nature01511>
- Angel TE, Aryal UK, Hengel SM, et al. Mass spectrometry-based proteomics: Existing capabilities and future directions. *Chem Soc Rev*. 2012;41(10):3912-3928. <https://doi.org/10.1039/c2cs15331a>
- Deutsch EW, Orchard S, Binz PA, et al. Proteomics Standards Initiative: Fifteen years of progress and future work. *J Proteome Res*. 2017;16(12):4288-4298. <https://doi.org/10.1021/acs.jproteome.7b00370>
- Chalmers MJ, Gaskell SJ. Advances in mass spectrometry for proteome analysis. *Curr Opin Biotechnol*. 2000;11(4):384-390. [https://doi.org/10.1016/S0958-1669\(00\)00114-2](https://doi.org/10.1016/S0958-1669(00)00114-2)
- Michalski A, Damoc E, Hauschild J-P, et al. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics*. 2011;10(9):M111.011015. <https://doi.org/10.1074/mcp.M111.011015>
- Cappadona S, Baker PR, Cutillas PR, Heck AJR, Van Breukelen B. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids*. 2012;43(3):1087-1108. <https://doi.org/10.1007/s00726-012-1289-8>
- Perez-Riverol Y, Wang R, Hermjakob H, Müller M, Vesada V, Vizcaino JA. Open source libraries and frameworks for mass spectrometry based proteomics: A developer's perspective. *Biochim Biophys Acta Proteins Proteomics*. 2014;1844(1 PART A):63-76. <https://doi.org/10.1016/j.bbapap.2013.02.032>
- Verheggen K, Raeder H, Berven FS, Martens L, Barsnes H, Vaudel M. Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows. *Mass Spectrom Rev*. 2017;39(3):292-306. <https://doi.org/10.1002/mas.21543>
- Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int J Mol Sci*. 2020;21(8). <https://doi.org/10.3390/ijms21082873>
- Tabb DL. The SEQUEST family tree. *J Am Soc Mass Spectrom*. 2015; 26(11):1814-1819. <https://doi.org/10.1007/s13361-015-1201-3>
- Deutsch EW, Perez-Riverol Y, Chalkley RJ, et al. Expanding the use of spectral libraries in proteomics. *J Proteome Res*. 2018;17(12): 4051-4060. <https://doi.org/10.1021/acs.jproteome.8b00485>
- Bittremieux W, Meysman P, Noble WS, Laukens K. Fast open modification spectral library searching through approximate nearest neighbor indexing. *J Proteome Res*. 2018;17(10):3463-3474. <https://doi.org/10.1021/acs.jproteome.8b00359>
- Schweppe DK, Chavez JD, Navare AT, et al. Spectral library searching to identify crosslinked peptides. *J Proteome Res*. 2016;15(5): 1725-1731. <https://doi.org/10.1021/acs.jproteome.6b00014>
- Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC. A critical review of bottom-up proteomics: The good, the bad, and the future of this field. *Proteomes*. 2020;8(3) MDPI AG:1-26. <https://doi.org/10.3390/proteomes8030014>
- Dorfer V, Pichler P, Stranzl T, et al. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J Proteome Res*. 2014;13(8):3679-3684. <https://doi.org/10.1021/pr500202e>
- Schittmayer M, Fritz K, Liesinger L, Griss J, Birner-Gruenberger R. Cleaning out the litterbox of proteomic scientists' favorite pet: Optimized data analysis avoiding trypsin artifacts. *J Proteome Res*. 2016;15(4):1222-1229. <https://doi.org/10.1021/acs.jproteome.5b01105>
- Walker C, Ryu S, Giannone RJ, Garcia S, Trinh CT. Understanding and eliminating the detrimental effect of thiamine deficiency on the oleaginous yeast *Yarrowia lipolytica*. *Appl Environ Microbiol*. 2020;86(3). <https://doi.org/10.1128/AEM.02299-19>
- Willems P, Fijalkowski I, Van Damme P. Lost and found: Re-searching and re-scoring proteomics data aids genome annotation and improves proteome coverage. *mSystems*. 2020;5(5). <https://doi.org/10.1128/mSystems.00833-20>
- Saiz-Fernández I, Milenković I, Berka M, et al. Integrated proteomic and metabolomic profiling of *Phytophthora cinnamomi* attack on sweet chestnut (*Castanea sativa*) reveals distinct molecular reprogramming proximal to the infection site and away from it. *Int J Mol Sci*. 2020;21(22):1-19. <https://doi.org/10.3390/ijms21228525>
- Leienacker L, Jung PS, Krecioch I, et al. LSD 1 inhibition induces differentiation and cell death in Merkel cell carcinoma. *EMBO Mol Med*. 2020;12(11):e12525. <https://doi.org/10.15252/emmm.202012525>
- Ruiz-May E, Altúzar-Molina A, Elizalde-Contreras JM, et al. A first glimpse of the Mexican fruit fly *Anastrepha ludens* (Diptera: Tephritidae) antenna morphology and proteome in response to a proteinaceous attractant. *Int J Mol Sci*. 2020;21(21):1-26. <https://doi.org/10.3390/ijms21218086>
- Dorfer V, Maltsev S, Winkler S, Mechtler K, Charmer RT. Boosting peptide identifications by chimeric spectra identification and retention time prediction. *J Proteome Res*. 2018;17(8):2581-2589. <https://doi.org/10.1021/acs.jproteome.7b00836>
- Locard-Paulet M, Bouyssie D, Froment C, Burlet-Schiltz O, Jensen LJ. Comparing 22 popular phosphoproteomics pipelines for peptide identification and site localization. *J Proteome Res*. 2020;19(3): 1338-1345. <https://doi.org/10.1021/acs.jproteome.9b00679>
- David G, Fogeron ML, Montserret R, et al. Phosphorylation and alternative translation on wheat germ cell-free protein synthesis of the DHBV large envelope protein. *Front Mol Biosci*. 2019;6:138. <https://doi.org/10.3389/fmolb.2019.00138>
- Jiang D, Borg M, Lorković ZJ, et al. The evolution and functional divergence of the histone H2B family in plants. *PLoS Genet*. 2020;16(7):e1008964. <https://doi.org/10.1371/journal.pgen.1008964>
- Searle BC, Lawrence RT, MacCoss MJ, Villén J. Thesaurus: Quantifying phosphopeptide positional isomers. *Nat Methods*. 2019; 16(8):703-706. <https://doi.org/10.1038/s41592-019-0498-4>
- Craig R, Beavis RC. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*. 2004;20(9):1466-1467. <https://doi.org/10.1093/bioinformatics/bth092>
- Barsnes H, Vaudel M. SearchGUI: A highly adaptable common interface for proteomics search and de novo engines. *J Proteome Res*. 2018;17(7):2552-2555. <https://doi.org/10.1021/acs.jproteome.8b00175>
- Vaudel M, Burkhart JM, Zahedi RP, et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotechnol*. 2015; 33(1):22-24. <https://doi.org/10.1038/nbt.3109>
- Martens L, Chambers M, Sturm M, et al. mzML – A community standard for mass spectrometry data. *Mol Cell Proteomics*. 2011;10(1): R110.000133. <https://doi.org/10.1074/mcp.R110.000133>

31. Jones AR, Eisenacher M, Mayer G, et al. The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol Cell Proteomics*. 2012;11(7):M111.014381. <https://doi.org/10.1074/mcp.M111.014381>
32. Vizcaino JA, Mayer G, Perkins S, et al. The mzIdentML data standard version 1.2, supporting advances in proteome informatics. *Mol Cell Proteomics*. 2017;16(7):1275-1285. <https://doi.org/10.1074/mcp.M117.068429>
33. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci*. 1986; 83(17):6233-6237. <https://doi.org/10.1073/pnas.83.17.6233>
34. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007;4(9):709-712. <https://doi.org/10.1038/nmeth1060>
35. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci*. 2004;101(26):9528-9533. <https://doi.org/10.1073/pnas.0402700101>
36. Frese CK, Altelaar AFM, van den Toorn H, et al. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem*. 2012;84(22):9668-9673. <https://doi.org/10.1021/ac3025366>
37. Reilly JP. Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom Rev*. 2009;28(3):425-447. <https://doi.org/10.1002/mas.20214>
38. Brodbelt JS. Photodissociation mass spectrometry: New tools for characterization of biological molecules. *Chem Soc Rev*. 2014;43(8): 2757-2783. <https://doi.org/10.1039/c3cs60444f>
39. Fort KL, Dyachenko A, Potel CM, et al. Implementation of ultraviolet photodissociation on a benchtop Q Exactive mass spectrometer and its application to phosphoproteomics. *Anal Chem*. 2016;88(4):2303-2310. <https://doi.org/10.1021/acs.analchem.5b04162>
40. Dillillo M, De Graaf EL, Yadav A, Belov ME, McDonnell LA. Ultraviolet photodissociation of ESI- and MALDI-generated protein ions on a Q-Exactive mass spectrometer. *J Proteome Res*. 2019;18(1):557-564. <https://doi.org/10.1021/acs.jproteome.8b00896>
41. Agten A, Van Houtven J, Askenazi M, Burzykowski T, Laukens K, Valkenburg D. Visualizing the agreement of peptide assignments between different search engines. *J Mass Spectrom*. 2020;55(8): e4471. <https://doi.org/10.1002/jms.4471>
42. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*. 2007;4(11):923-925. <https://doi.org/10.1038/nmeth1113>

How to cite this article: Dorfer V, Strobl M, Winkler S, Mechtler K. MS Amanda 2.0: Advancements in the standalone implementation. *Rapid Commun Mass Spectrom*. 2021;35: e9088. <https://doi.org/10.1002/rcm.9088>