Journal of proteome research

Article

pubs.acs.org/jpr

MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra

Viktoria Dorfer,[†,⊥] Peter Pichler,[‡,∥,⊥] Thomas Stranzl,[‡,⊥] Johannes Stadlmann,[‡] Thomas Taus,[‡] Stephan Winkler,[†] and Karl Mechtler*[,‡,§]

[†]Bioinformatics Research Group, University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria

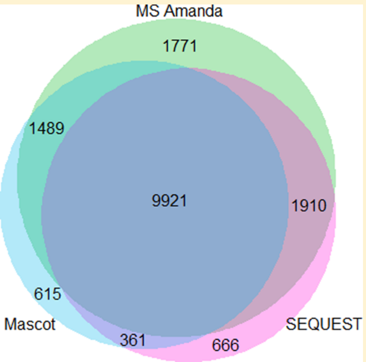
[‡]Protein Chemistry Facility, IMP, Research Institute of Molecular Pathology, Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[§]Protein Chemistry Facility, IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[∥]Wiener Krankenanstaltenverbund, Thomas-Klestil-Platz 7/1, 1030 Vienna, Austria

Supporting Information

ABSTRACT: Today's highly accurate spectra provided by modern tandem mass spectrometers offer considerable advantages for the analysis of proteomic samples of increased complexity. Among other factors, the quantity of reliably identified peptides is considerably influenced by the peptide identification algorithm. While most widely used search engines were developed when high-resolution mass spectrometry data were not readily available for fragment ion masses, we have designed a scoring algorithm particularly suitable for high mass accuracy. Our algorithm, MS Amanda, is generally applicable to HCD, ETD, and CID fragmentation type data. The algorithm confidently explains more spectra at the same false discovery rate than Mascot or SEQUEST on examined high mass accuracy data sets, with excellent overlap and identical peptide sequence identification for most spectra also explained by Mascot or SEQUEST. MS Amanda, available at http://ms.imp.ac.at/?goto=msamanda, is provided free of charge both as standalone version for integration into custom workflows and as a plugin for the Proteome Discoverer platform.



KEYWORDS: *tandem mass spectrometry, MS/MS, database search algorithm, high-resolution spectra, high mass accuracy, peptide identification, proteomics*

■ INTRODUCTION

Mass spectrometry (MS)-based proteomics has evolved into an indispensable approach in biological sample analysis.[1,2] In shotgun proteomics experiments, proteins are proteolytically cleaved to peptides, separated based on specific physico-chemical properties, and subsequently analyzed in a mass spectrometer.

Obtained spectra, containing mass-to-charge ratios of either charged peptides (MS^1) or fragment ions (MS/MS or MS^2) associated with respective ion intensities, are matched to candidate peptides, and a score dependent on an identification algorithm is assigned to each peptide spectrum match (PSM).

Scoring algorithms such as Mascot,[3] SEQUEST,[4] X-Tandem,[5] Andromeda,[6] OMSSA,[7] MyriMatch,[8] Phenyx,[9] or Morpheus[10] incorporate various strategies to evaluate the quality of a PSM. In particular, SEQUEST reports a cross-correlation score of the acquired mass spectrum matching a modeled peptide spectrum. In comparison, Mascot estimates the probability that a particular peptide spectrum match is a random event by probabilistic modeling. Other search engines are specifically designed for a particular purpose such as for the analysis of post-translationally modified peptides (e.g., ModifiComb[11] or InsPecT[12]).

Recent technological advance of instruments allows high-throughput identification of thousands of proteins,[13,14] which is a prerequisite for the challenging analysis of complete proteomes. Tackling the complete yeast proteome, the Mann group was able to detect more than 2000 proteins in 48 h in 2006.[15] Only a few years later, both the Mann group in 2012 as well as Coon and co-workers in 2013 described comprehensive analyses of the nearly complete yeast proteome at manifoldly decreased runtimes.[16,17] The continuous increase in throughput and precision enables the research community to address previously unsolvable scientific challenges, such as the in-depth analysis of mammalian proteomes.[18] Recent studies identified more than 10 000 human proteins in the proteome of a human cancer cell line, which is suggested to be close to completion.[19−21]

Technological development of instruments leads to more reliable data subsequently used by MS search engines for the assignment of potential peptides to spectra.[22] While newer instruments deliver potentially more MS/MS spectra per time unit, typically only up to 60% of these spectra are confidently assigned to peptides, suggesting a potential for improve-

Received: March 7, 2014

ment.[23,24] We further consider the emergence of high-resolution instruments with highly accurate mass recordings[25−27] as a stimulus for the development of peptide search algorithms particularly suitable to such data.

We here describe MS Amanda, a novel database search engine, specially developed for high-resolution tandem mass spectrometry data, taking advantage of high mass accuracy and considering fragment ion intensities. To show the general applicability of MS Amanda, the performance of the algorithm was evaluated on HCD, ETD, and CID fragmentation type data.

■ MATERIALS AND METHODS

MS Amanda Identification Algorithm

We have designed MS Amanda based on a binomial distribution function incorporating peak intensities and determining favorable outcomes (successes) and possible outcomes (sample space) in a specific manner. Our multithreaded implementation in C# incorporates the described identification algorithm.

During preprocessing, peaks corresponding to precursor ions are removed and an optional de-isotoping of fragment ions is applied (intensities of discarded isotopes are added to C12 peaks). In order to discriminate ion signals from noise, peak picking is performed. In each 100 Da window, the m most intense peaks are picked, where m is a value between 1 and 10. All possible values for m are tested, and the value representing the maximum PSM score is selected.[28,29]

Theoretical fragment ions of each candidate peptide, thus, of all peptides in the (forward or decoy) database that match the precursor mass of a certain spectrum considering a specific MS^1 mass tolerance, are matched to E, the set of picked peaks, allowing a given MS^2 mass tolerance (t). The first part of the scoring algorithm used in MS Amanda is based on a cumulative binomial distribution function defined as

$$P(n, p, N) = \sum_{k=n}^{N} \binom{N}{k} p^k (1-p)^{N-k} \quad (1)$$

that is, the probability to match at least n out of N peaks by chance. This formula assumes that the random variable denoting the number of matched peaks follows a binomial distribution as the sum of Bernoulli random variables X_i $\{i = 1,...N\}$. For each X_i, p is the probability to match one peak by chance (see formula 3). In our usage of the cumulative binomial distribution function, n is the number of matched peaks, and N is the number of picked peaks. We assume independence of the X_i.

The probability p to match one peak by chance is the fraction of the m/z range that is covered by the theoretical ions $f(pep)$ and the total mass window (first peak to last peak in the experimental spectrum) considering peak picking depth m. The covered m/z range of $f(pep)$ is based on fragment ion tolerance t, considering solely fragment masses in the mass range of the first peak ($e_1(s,m)$) and the last peak ($e_N(s,m)$) of spectrum s. Given the set F, which are all theoretical fragment ions $f(pep)$ within the mass of the first and the last picked peak of the experimental spectrum considering the fragment ion tolerance t

$$\begin{aligned} F(s, pep, m) = \{f(pep) | (e_1(s, m) - t) \\ \le f(pep) \\ \le (e_N(s, m) + t)\} \end{aligned} \quad (2)$$

probability p is defined as

$$p(s, pep, m) = \frac{(|F(s, pep, m)| \times 2t) - O(F(s, pep, m))}{(e_N(s, m) + t) - (e_1(s, m) - t)} \quad (3)$$

The overlap $O(F(s,pep,m))$ is the sum of all overlapping ranges in the theoretical spectrum F considering mass tolerance t. With peaks f_i sorted by m/z in ascending order, this overlap between consecutive peaks f_i and f_{i+1} is calculated as

$$o(f_i, f_{i+1}) = \begin{cases} 0 & f_{i+1} - f_i > 2t \\ (f_i + t) - (f_{i+1} - t) & \text{else} \end{cases} \quad (4)$$

$$O(F) = \sum_{i=1}^{|F|-1} o(f_i, f_{i+1}) \quad (5)$$

where $o(f_i, f_{i+1})$ is the overlap between two consecutive fragment ions f_i and f_{i+1}. For a graphical illustration see Supporting Information Figure S1.

$P(n,p,N)$ indicates the reliability of a peptide spectrum match under the null hypothesis of a random match based on a binomial distribution. As a consequence, more reliable PSMs are characterized by a low probability (for randomly matching peaks). To improve the distinction between false and correct identifications, we additionally consider the intensities of the peaks: The calculated probability to match at least n out of N peaks by chance is weighted by the reciprocal of the explained ion current $eif(s,pep,m)$.

$$eif(s, pep, m) = \frac{\sum_{x \in M(s,pep,m)} I(x)}{\sum_{y \in E(s,pep,m)} I(y)} \quad (6)$$

$eif(s,pep,m)$ is the fraction of the sum of the intensities $I(M)$ of the matched peaks M ($|M| = n$) and the sum of the intensities $I(E)$ of all picked peaks E ($|E| = N$). The weighting rewards peptides matching more intense peaks over those matching less intense peaks.

Finally, the quality of the match of peptide pep with spectrum s is represented by the MS Amanda score $S(s,pep)$. The score $S(s,pep)$ is the basis for further false discovery rate (FDR) estimation.

$$S(s, pep) = \max_{m \in [1..10]}\left(-10 \times \log\left(\frac{P(s, pep, m)}{eif(s, pep, m)}\right)\right) \quad (7)$$

Data Sets

We compared the performance of MS Amanda based on four data sets: an HCD HeLa sample, a synthetic peptide library, a histone data set, and a CID HeLa sample. The HCD HeLa sample, published by Michalski et al.,[30] consists of three replicate measurements of tryptic peptides derived from one human cancer cell line. The synthetic peptide library, as described by Marx et al.,[31] is composed of more than 200 000 phosphorylated and nonphosphorylated peptides. Performance comparisons were based on provided HCD and ETD data. The histone data set is composed of four different preparations, namely, Histone II-A from calf thymus (Sigma), Histone III-S from calf thymus (Sigma), Histone IV from *Xenopus laevis*, recombinantly expressed in *Escherichia coli* (Upstate), and Core Histones from chicken erythrocytes (Millipore). The published CID HeLa sample[32] covers three replicates measured with a 1 h gradient (1 μg).

Histone Sample Preparation

Samples were reduced and alkylated using dithiothreotiol (DTT; 2 mM, final concentration) and methyl methanethiosulfonate (MMTS; 5 mM final concentration). Proteins were digested overnight with endoproteinase Glu-C (from *Staphylococcus aureus* V8, Sigma) in 100 mM ammonium bicarbonate at 37 °C.

Peptides were separated on a reversed-phase column (Acclaim PepMap RSLC column, 2 μ, 100 Å, 75 μm × 500 mm, Thermo Fisher) by a linear gradient from 0.8 to 32% acetonitrile in 0.1% formic acid over 30 min on an RSLC nano HPLC system (Dionex). The eluting peptides were directly analyzed using a hybrid quadrupole-orbitrap mass spectrometer (QExactive, Thermo Fisher). The QExactive mass spectrometer was operated in data-dependent mode, using a full scan (m/z range 350−2000, nominal resolution 140 000, target value 1×10^6) followed by MS/MS scans of the 12 most abundant ions. MS/MS spectra were acquired at a resolution of 17 500 using normalized collision energy 30%, isolation width of 2, and the target value was set to 5×10^4. Precursor ions selected for fragmentation (charge state 3 and higher) were put on a dynamic exclusion list for 10 s (dynamic exclusion tolerance is 10 ppm on QExactive by default). Additionally, the underfill ratio was set to 20%, resulting in an intensity threshold of 2×10^4. The peptide match feature and the exclude isotopes feature were enabled.

Database Search Settings

Proteome Discoverer version 1.4.288 (PD) was used for peptide identifications. All data sets were searched with Mascot (version 2.2.1), SEQUEST (with probability score calculation) as provided in PD, and MS Amanda. Advanced search settings in PD were changed from default in order to store all PSMs in the result file (all cutoff filters and thresholds were disabled).

Searches for the HeLa and the histone data sets were performed with 7 ppm precursor mass tolerance and 0.03 Da fragment ion mass tolerance (0.5 for CID). Following Marx et al., we used 5 ppm precursor mass tolerance and 0.02 Da fragment mass tolerance for the synthetic peptide library. For HCD and CID, considered fragment ions were left at defaults for Mascot and SEQUEST, and set to *b* and *y* ions for MS Amanda. ETD searches with Mascot and MS Amanda were performed using *c*, *y*, *z* + 1, and *z* + 2 ions.

For the HeLa data sets, oxidation(M) was set as variable modification, carbamidomethyl(C) as fixed modification, and trypsin as enzyme allowing up to two missed cleavages. The peptide library was searched with oxidation(M) and phosphorylation(S,T,Y) as variable modifications and up to four missed cleavage sites for trypsin.

Variable modification settings for the histone data set were oxidation(M), phosphorylation(S,T,Y), methyl(K,R), dimethyl(K,R), trimethyl(K), and acetyl(K). Methylthio(C) was set as fixed modification, GluC (C-terminal cleavage after D or E) as enzyme, and two as the maximum number of missed cleavages.

Performance comparisons were based on 1% FDR.[33,34] We generated concatenated forward and reverse (decoy) protein databases with contaminants using MaxQuant Sequence Reverser (v1.0.13.13).[14] We searched the HeLa data sets against Swiss-Prot_human[35] (release 2013_10), merged the synthetic peptide sequences with Swiss-Prot_human for the peptide library, and searched the histone data against the complete Swiss-Prot (release 2013_10). For FDR calculation, peptides shorter than 7 amino acids were discarded and conservative FDR estimation was ensured by preferring the decoy peptide to an equally scored peptide. Peptide grouping for unique peptide level FDR estimation was solely based on the peptide sequence, and the highest score was kept for each peptide group.

■ RESULTS

We compared PSM and peptide identifications of MS Amanda to Mascot and SEQUEST, two search algorithms widely used for peptide identification in mass spectrometry. Performance of MS Amanda was evaluated on an HCD HeLa set (Figure 1), on a synthetic peptide library (Figure 2), a histone data set (Figure 3), and on a CID HeLa set. In addition to PSM identifications based on a forward decoy database approach at 1% FDR, we show results for unique peptides at 1% FDR in Supporting Information Table S1.

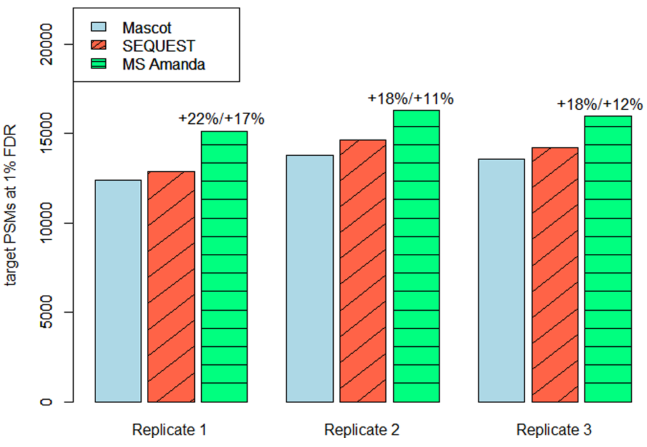


Figure 1. Performance comparison on HCD HeLa data set.[30] The previously published data set is composed of three replicates measured on a Thermo Fisher QExactive instrument. For all three replicates, consistently more PSMs were identified at 1% FDR (PSM level) with MS Amanda as compared to Mascot or SEQUEST.

Performance of MS Amanda

For HCD data, the numbers of identified PSMs by Mascot, SEQUEST, and MS Amanda are depicted in Figure 1 for the HeLa data set and Figure 2(A,B) for the synthetic peptide library. Results for the histone data set are shown in Figure 3. We report identified PSMs in the synthetic peptide library separately for nonphosphorylated (Figure 2A) and phosphorylated (Figure 2B) peptides.

Consistently higher quantities of PSM identifications were observed for MS Amanda as compared to both Mascot and SEQUEST for all high-resolution data sets. In the three HCD HeLa replicates, we identified between 11 and 22% more PSMs with MS Amanda compared to Mascot and SEQUEST.

While SEQUEST performed slightly better than Mascot on the nonphosphorylated peptide library subset (2A), the reciprocal situation was observed on the phosphorylated peptide library subset (2B). Still, MS Amanda outperformed both search engines for both subsets by 4−22%.

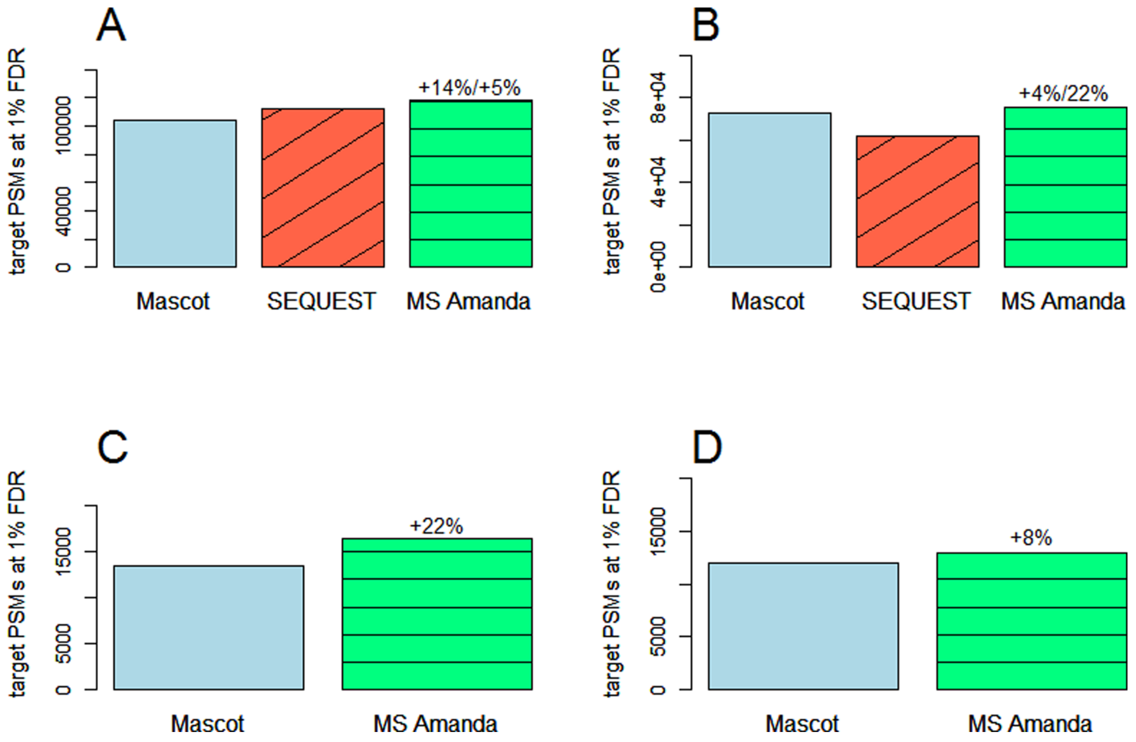


Figure 2. Identified PSMs in a synthetic peptide library comprising HCD and ETD data.[31] Applying MS Amanda led to the highest number of identified PSMs on the HCD data set for both nonphosphorylated (A) and phosphorylated (B) peptides. A similar performance increase was observed on the ETD data set for nonphosphorylated (C) and phosphorylated (D) peptides.

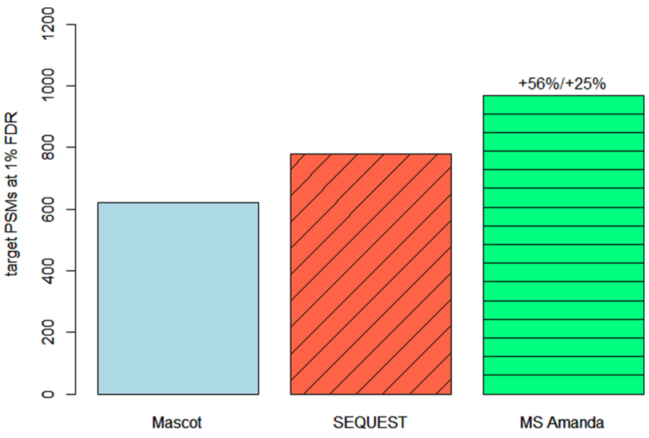


Figure 3. Performance comparison of identified PSMs in a histone data set. We used four different histone preparations originating from three species and measured them on a Thermo Fisher QExactive mass spectrometer. HCD raw files were combined for peptide identification. At 1% FDR, we identified more PSMs with MS Amanda as with Mascot and SEQUEST.

For the histone data set, we identified 620 target PSMs with Mascot and 778 with SEQUEST. By applying MS Amanda we identified 969 PSMs, which corresponds to a performance increase in identified PSMs of 56 and 25%, respectively.

We further analyzed the performance of MS Amanda to Mascot on the peptide library ETD data subset. Both search algorithms identified considerably more PSMs than SEQUEST, a comparison with SEQUEST on the ETD subset was therefore omitted.

In accordance with our analysis of the HCD data subset, we report both PSMs of nonphosphorylated (Figure 2C) and phosphorylated (Figure 2D) peptides. While we identified 13 489 PSMs of nonphosphorylated peptides with Mascot in the ETD data, we found notably more PSMs (16 400) with MS Amanda, which is a 22% increase in identified PSMs at 1% FDR. For the phosphorylated subset, we found a comparable trend. Here, we identified 12 016 PSMs with Mascot and 12 979 PSMs with MS Amanda (an increase of 8%).

Benchmarking MS Amanda, Mascot, and SEQUEST on the low-resolution CID data reported comparable performance for all three search engines, with slightly higher PSM identification rates for MS Amanda (1–5%; see Supporting Information Table S2).

We list the numbers of identified PSMs for all three high-resolution data sets in Supporting Information Table S2. In Supporting Information Table S1, we show identified unique peptides at 1% FDR (peptide level) for the HCD and CID HeLa data set and for the HCD and ETD peptide library data sets. The limited number of proteins in the histone data set did not allow for accurate peptide level FDR estimation. On these data, we only report PSM level FDR estimation.

For completeness, we also compared the performance of MS Amanda with the noncommercial search engine Morpheus, a recently described search algorithm which was also specifically designed for high mass accuracy MS^2 spectra (see Supporting Information Table S3).

PSM Overlap

To show the validity of our approach, we investigated the overlap in target PSM identification for all three search algorithms. Analyzing one replicate of the HCD HeLa data set (MS Amanda 15 091 PSMs, Mascot 12 386 PSMs, SEQUEST 12 858 PSMs), 9921 spectra were commonly identified by all three search engines (Figure 4). While MS Amanda identified considerably more unique PSMs than compared search engines, the capability of MS Amanda to identify large fractions of

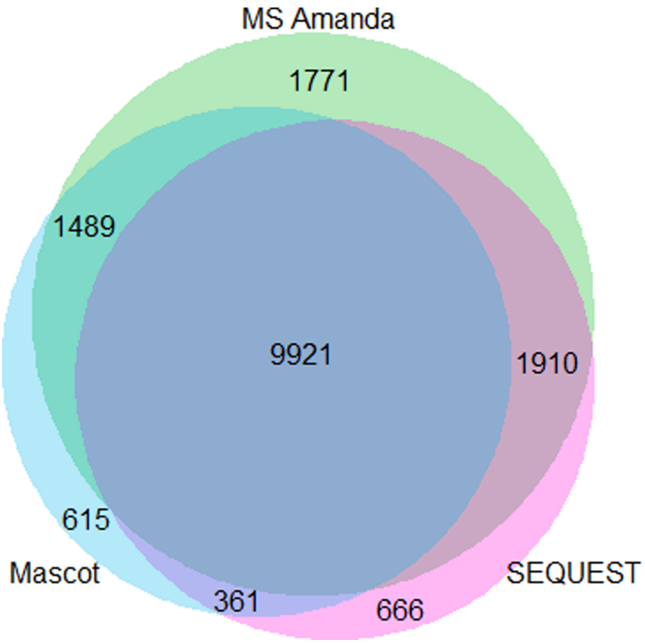


Figure 4. Overlap of target PSMs based on one HCD HeLa replicate. MS Amanda explains large fractions of PSMs also identified by Mascot and SEQUEST. Further, our algorithm explains many peptides otherwise uniquely identified by either Mascot or SEQUEST.

peptides found by either Mascot or SEQUEST is noteworthy; 92% of the PSMs identified by Mascot and further 92% of those identified by SEQUEST are reliably found by MS Amanda, while only 80% of PSMs identified by SEQUEST and 83% of PSMs identified by Mascot are also found by the respective other search engine. This highlights that MS Amanda is remarkably capable of explaining spectra otherwise uniquely identified by either Mascot or SEQUEST.

■ DISCUSSION

Current state-of-the-art mass spectrometers provide highly accurate m/z data of both intact peptides and fragment ions. These instruments were not readily available at the time when Mascot and SEQUEST were developed. Still, Mascot and SEQUEST are among the most widely used search engines and perform generally well for both low- and high-resolution data. Here we present MS Amanda, a peptide identification algorithm shown to outperform these established search engines on examined data sets.

MS Amanda is based on a cumulative binomial distribution function, which estimates the probability to match n out of N peaks by chance. In our implementation of the cumulative distribution function, N is the number of picked peaks, and n the number of matching peaks (formula 1). We consider this strategy beneficial for spectra where the number of theoretical fragment ions is large (e.g., for spectra with many different types of neutral loss peaks). In addition, our estimation of the probability p to match one peak by chance (formula 3) provides the advantage that fragment ion tolerances can be specified in parts per million. Further, our scoring system considers the intensities of all matched peaks for reporting the score of each potential peptide spectrum match.

We found that MS Amanda provides an increased peptide identification performance in comparison to the well-established search engines Mascot and SEQUEST, as highlighted both for HCD and ETD data sets (increase in PSMs between 11 and 22% on the HCD HeLa set). The number of detected PSMs in a data set correlates with the number of unique peptides. More identified PSMs lead to potentially more identified peptides, which subsequently influences protein scoring and potentially increases the number of identified proteins. While MS Amanda uniquely identified many additional PSMs, our search engine further incorporates large fractions of PSMs otherwise uniquely reported by either Mascot or SEQUEST.

We suggest MS Amanda as particularly well-suitable for high-resolution data sets, as we observed a substantial performance gain for HCD and high mass accuracy ETD data. In addition, by showing small but consistent improvements for CID data, we further highlight its general applicability. We want to emphasize the performance of MS Amanda on our modification-rich histone data set, where we observed a 24−56% increase in identified PSMs. This observation suggests that one possible explanation for the increased performance might be that MS Amanda is particularly well-suited for the identification of peptides of large mass and higher charge state (charge states +4 to +8 constitute almost middle-down data).

With its remarkably consistent performance and provided as downloadable version (standalone and integrated in PD), we believe that our ready-to-use implementation is of particular value for the proteomics community. MS Amanda is available at http://ms.imp.ac.at/?goto=msamanda.

■ ASSOCIATED CONTENT

Supporting Information

Additional tables and figure as discussed in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mechtler@imp.ac.at. Phone: 0043 1 79730. Fax: 0043 1 798 7153.

Author Contributions

⊥V.D., P.P., and T.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF) (TRP 308-N15, P24685-B24, and SFB F3402) and the European Community's Seventh Framework Programme (PRIME-XS (262067) and MeioSys (222883-2)).

■ REFERENCES

(1) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198−207.

(2) Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem. Soc. Rev.* **2012**, *41*, 3912−3928.

(3) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551−3567.

(4) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid

sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.

(5) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466−1467.

(6) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794−1805.

(7) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958−964.

(8) Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* **2007**, *6*, 654−661.

(9) Colinge, J.; Masselot, A.; Giron, M.; Dessingy, T.; Magnin, J. OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **2003**, *3*, 1454−1463.

(10) Wenger, C. D.; Coon, J. J. A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *J. Proteome Res.* **2013**, *12*, 1377−1386.

(11) Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **2006**, *5*, 935−948.

(12) Tanner, S.; Shu, H.; Frank, A.; Wang, L.-C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. InsPecT: identification of post-translationally modified peptides from tandem mass spectra. *Anal. Chem.* **2005**, *77*, 4626−4639.

(13) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18132−18138.

(14) 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*, 1367−1372.

(15) De Godoy, L. M. F.; Olsen, J. V.; de Souza, G. A.; Li, G.; Mortensen, P.; Mann, M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **2006**, *7*, R50.

(16) Nagaraj, N.; Kulak, N. A.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **2012**, *11*, M111.013722.

(17) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The one hour yeast proteome. *Mol. Cell. Proteomics* **2013**, 1−23.

(18) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell* **2013**, *49*, 583−590.

(19) Munoz, J.; Low, T. Y.; Kok, Y. J.; Chin, A.; Frese, C. K.; Ding, V.; Choo, A.; Heck, A. J. R. The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol. Syst. Biol.* **2011**, *7*, 550.

(20) Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **2011**, *7*, 549.

(21) Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Pääbo, S.; Mann, M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **2011**, *7*, 548.

(22) Walsh, G. M.; Rogalski, J. C.; Klockenbusch, C.; Kast, J. Mass spectrometry-based proteomics in biomedical research: emerging technologies and future strategies. *Expert Rev. Mol. Med.* **2010**, *12*, e30.

(23) Käll, L.; Vitek, O. Computational mass spectrometry-based proteomics. *PLoS Comput. Biol.* **2011**, *7*, e1002277.

(24) Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 1785−1793.

(25) Olsen, J. V.; de Godoy, L. M. F.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4*, 2010−2021.

(26) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4*, 709−712.

(27) Andrews, G. L.; Simons, B. L.; Young, J. B.; Hawkridge, A. M.; Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* **2011**, *83*, 5442−5446.

(28) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24*, 1285−1292.

(29) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13417−13422.

(30) Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **2011**, *10*, M111.011015.

(31) Marx, H.; Lemeer, S.; Schliep, J. E.; Matheron, L.; Mohammed, S.; Cox, J.; Mann, M.; Heck, A. J. R.; Kuster, B. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat. Biotechnol.* **2013**, *31*, 557−564.

(32) Köcher, T.; Pichler, P.; Swart, R.; Mechtler, K. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. *Nat. Protoc.* **2012**, *7*, 882−890.

(33) Moore, R. E.; Young, M. K.; Lee, T. D. Qscore: an algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 378−386.

(34) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207−214.

(35) UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **2013**, *41*, D43-7.