

MS_Piano: A Software Tool for Annotating Peaks in CID Tandem Mass Spectra of Peptides and N-Glycopeptides

Xiaoyu Yang,* Pedatsur Neta, Yuri A. Mirokhin, Dmitrii V. Tchekhovskoi, Concepcion A. Remoroza, Meghan C. Burke, Yuxue Liang, Sanford P. Markey, and Stephen E. Stein



Cite This: <https://doi.org/10.1021/acs.jproteome.1c00324>



Read Online

ACCESS |



Metrics & More



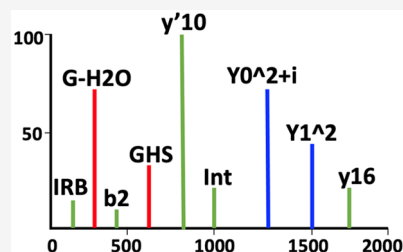
Article Recommendations



Supporting Information

ABSTRACT: Annotating product ion peaks in tandem mass spectra is essential for evaluating spectral quality and validating peptide identification. This task is more complex for glycopeptides and is crucial for the confident determination of glycosylation sites in glycoproteins. MS_Piano (Mass Spectrum Peptide Annotation) software was developed for reliable annotation of peaks in collision induced dissociation (CID) tandem mass spectra of peptides or N-glycopeptides for given peptide sequences, charge states, and optional modifications. The program annotates each peak in high or low resolution spectra with possible product ion(s) and the mass difference between the measured and theoretical m/z values. Spectral quality is measured by two major parameters: the ratio between the sum of unannotated vs all peak intensities in the top 20 peaks, and the intensity of the highest unannotated peak. The product ions of peptides, glycans, and glycopeptides in spectra are labeled in different class-type colors to facilitate interpretation. MS_Piano assists validating peptide and N-glycopeptide identification from database and library searches and provides quality control and optimizes search reliability in custom developed peptide mass spectral libraries. The software is freely available in .exe and .dll formats for the Windows operating system.

KEYWORDS: peak annotation, peptide fragmentation, glycopeptide fragmentation, proteomics, glycoproteomics, mass spectrometry, software, peptide identification, glycopeptide identification



INTRODUCTION

Tandem high and low resolution mass spectrometry have become routine for identifying peptides, their modifications, and precursor proteins in proteomic studies.^{1–4} This technique is also used to identify glycosylation sites in glycoproteins in complex biological samples.^{5,6} Two principal means of identifying peptides and glycopeptides in bottom-up proteomic studies are by searching protein sequence databases^{7–10,34–36} and mass spectral libraries.^{11–15} Combining these methods can enhance results.¹⁶ However, sometimes high scoring spectra have unassigned ions due to contamination or incorrect identification. This is especially true for data collected from highly complex protein digests containing thousands of spectra, each of which may contain peaks from cofragmentation of ions present in m/z fragmentation windows. In such cases, annotating peaks in mass spectra with reliable characterized product ions can provide additional confidence of correct and context-consistent identification. This is particularly useful for library searching, where spectral quality can be ensured by identifying the origin of all peaks in each library spectrum prior to its use as a reliable reference. Prior annotation enables users to quickly understand the significance of any differences in query and library spectra.

Software tools for automatically annotating product ion peaks in tandem mass spectra have been limited and few software tools have been reported. One of these is an online

tool, MS-Product.¹⁷ Such programs usually provide useful predicted fragmentation ions, but do not annotate experimental spectra, verify identification correctness, or evaluate spectrum quality. They require users to manually match the theoretical m/z values with the experimental ones or write additional programs for implementing this match. Other studies have noted commonly encountered product ions,^{18,19} but these are not available in currently used software. Several software programs are available (e.g., free software PDV,³⁰ TOPPView,³¹ and the commercial software Proteome Discoverer³²) as spectral visualization tools, but these do not provide glycopeptide product ion annotation.

Here, we present our newly developed software tool, MS_Piano (Mass Spectrum Peptide Annotation), that rapidly annotates peaks in tandem mass spectra of both peptides and N-glycopeptides with product ions based on peptide sequence, charge state, and modifications using a variety of rules developed to minimize errors. This software also provides a measure of spectral quality to evaluate the validity of

Received: April 19, 2021

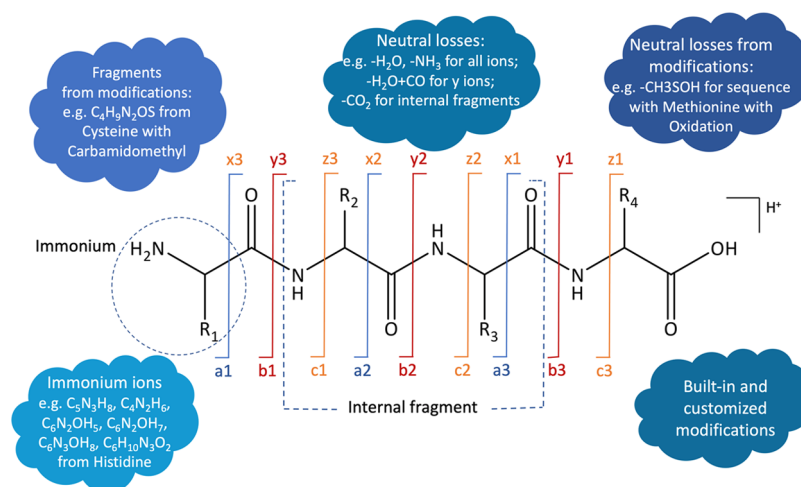


Figure 1. Varieties of product ions and fragments used in MS_Piano (c, x, and z ions are not used for annotating CID mass spectra).

identifications. It is easy to use for proteomic and glycoproteomic data analysis and freely available as a Windows application in both .exe and .dll formats.

SOFTWARE FEATURES

MS_Piano annotates peaks in a tandem high or low resolution mass spectrum of a peptide or N-glycopeptide based on a presumed sequence, charge state, and optional modifications.

Peptide Annotation

Peaks in a tandem mass spectrum are annotated with precursor (p), product ion y, b, a, immonium ions,¹⁸ internal fragment ions, neutral losses or gains from the precursor and product ions, as well as their isotopic ions. A peak is not labeled as an internal fragment if it is assigned with any of the above ions for simplicity, and to avoid internal fragment complexity. MS_Piano follows common peptide fragmentation and product ion nomenclature^{28,29} and excludes c, x, and z ions for annotating collision induced dissociation (CID) spectra. Product ion annotations and examples are shown in Figure 1. The software provides more than 100 built-in modifications and used more than 800 modifications from unimod.²⁰ Users can add their own modifications in a separate file "mod_added.txt". Detailed information on name, formula, and amino acid residues of all the built-in modifications, immonium and fragment ions, neutral losses and gains, iTRAQ (isobaric tags for relative and absolute quantitation) and TMT (Tandem Mass Tag) fragments are provided in the Supporting Information (Table S1–S5).

For each peptide's sequence, charge state and optional modifications, MS_Piano uses the example format in Figure 2.

An example of an annotated high-resolution mass spectrum in Figure 3 shows that MS_Piano annotates not only y, b, immonium, and internal fragment ions, but also iTRAQ and its

reporters, the neutral losses of b, y, and internal fragments. Two examples of low-resolution ion trap mass spectra annotated with MS_Piano are shown in Figure S1 in the Supporting Information.

Glycopeptide Annotation

We extended MS_Piano to annotate peaks of CID tandem mass spectra of N-glycopeptides. We use 6 common monosaccharides found in mammals as basic units for glycan composition (Table 1). These short symbols enable their compact representation on computer screens. We also use "G": to distinguish glycosylation from other peptide modifications. For example, in Figure 4, YHYNGTLLDGLFDSSYSR/3_2(0,Y,TMT)(3,N,G:G2H9), G₂H₉ is the glycan on amino acid Asparagine (N) at location 3 counting from N-terminus starting from 0.

In addition to the above-described product ions and fragments for peptide annotation, the following 3 product ion types were added for N-glycopeptide annotation:

1. Peptides: Product ions y', b', and a' ions are y, b, and a ions that have lost N-glycosylation, respectively. These ions plus various sugars are also used for annotation. For example, in Figure 4, b'₄ is b₄ ion without glycan G₂H₉, and b'₄+G is b'₄ with a G modification.
2. Glycans: Glycans and those with various losses; e.g., G, G-H₂O, G-2H₂O in Figure 4 and GHS in Figure 6.
3. Glycopeptides: Peptide sequence with various glycans. For brevity, the capital letter Y is followed by numbers (Y0 to Y10) to annotate common N-glycopeptide fragments from high-mannose glycans (Table 2); e.g., isotopic Y0 at charge 2 (annotated as "Y0Δ2+i") in Figure 4.

The presence of $m/z = 197.0445$ indicates the loss of C₂H₉NO₃ loss from NeuAc;²¹ the loss of C₂H₆O₃, CH₆O₃ and C₂H₄O₂ serve to annotate HexNAc.²² After careful manual examination, the following annotations were added to the software: C₇H₅O₂, C₈H₇O₃, C₁₁H₅NO, C₉H₇O₄, C₉H₁₀NO₄, and C₉H₁₁O₆.

As a critical step in MS_Piano, when a peak in a spectrum could be annotated with multiple product ions, they are prioritized in the following order: precursor (p), Y, glycopeptides, glycan oxonium ions, y or b, y' or b', immonium ions or fragments from modifications, a or a' ions. Prioritizing peak annotations facilitates optimal library searching results.

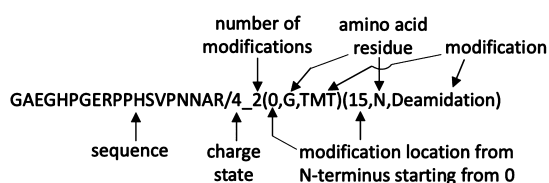


Figure 2. An example format for a peptide sequence, charge state, and optional modifications.

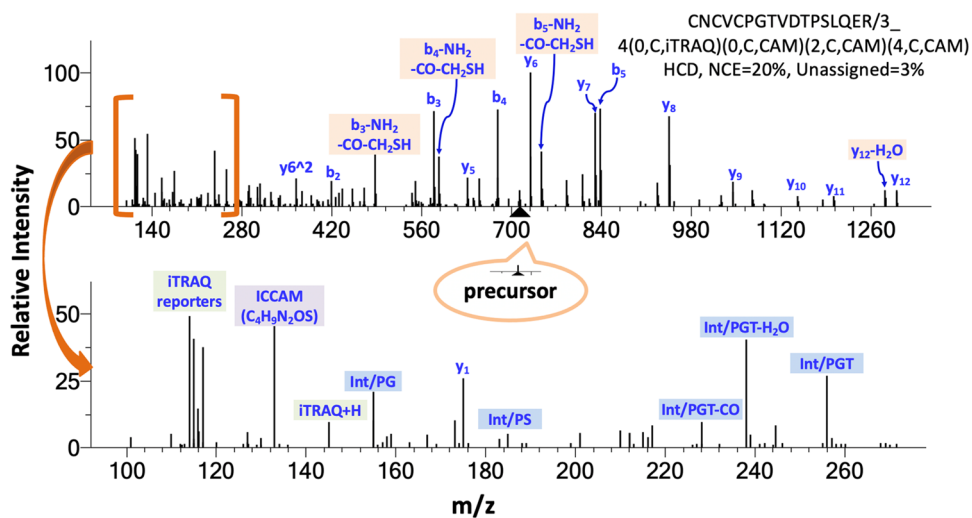


Figure 3. An example high resolution higher-energy collision-induced dissociation (HCD) mass spectrum with normalized collision energy (NCE) 20% of a peptide with iTRAQ and carbamidomethyl (CAM) modifications at charge 3 annotated with MS_Piano. The highlighted annotations are neutral losses, iTRAQ and its reporter ions, immonium ion (labeled with “I” in the beginning of the annotation, e.g., ICCAM), and internal fragments (labeled with “Int” in the beginning of the annotation, e.g., Int/PG). A product ion with multiple charge states is labeled with “^” before charge value, e.g., y6^2 is y6 at charge state 2. The “Unassigned” value is the fraction of unannotated peak intensities among the top 20 peaks.

Table 1. Glycans for Annotation

abbreviation	sugar	full name
G	HexNAc	N-Acetylhexosamine
H	Hex	Hexose
F	Fuc	Fucose
S	NeuAc	Neuraminic acid
Sg	NeuGc	N-Glycolyl neuraminic acid
P	Pent	Pentose
So ^a	SO ₃	
Po ^a	HPO ₃	

^aNonsugar modifications.

An example spectrum of an N-glycopeptide in Figure 4 shows that MS_Piano annotates the fragments of glycans, peptides, and glycopeptides, enabling the rapid validation of a glycopeptide identification.

Parameters for Evaluating Spectral Quality

In addition to possible product ion(s), each peak in a spectrum is annotated with the mass difference between the measured and theoretical *m/z* values. The default mass tolerance range is set as 20 ppm and can be customized by users by adding the option such as “-r 10ppm” or “-r 0.4 Da” for high and low resolution spectra, respectively. Furthermore, spectrum quality

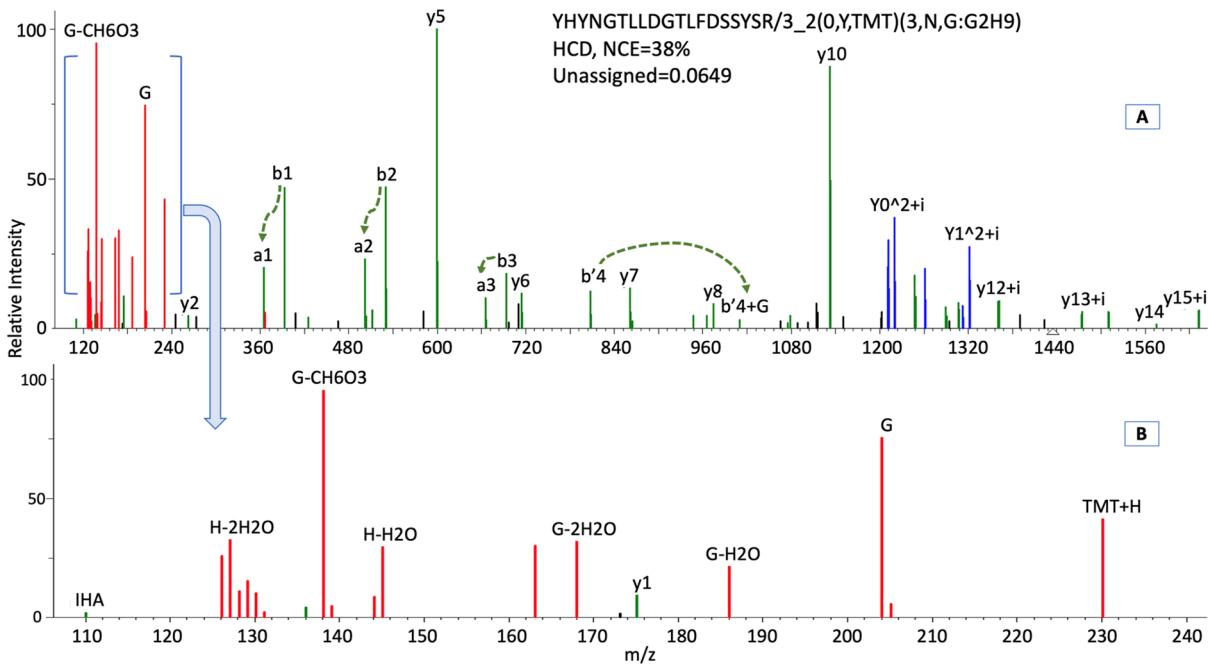


Figure 4. An example of an N-glycopeptide spectrum, acquired on high-resolution HCD with normalized energy (NCE) 38%, annotated with MS_Piano. Product ions a1, a2, a3, and b'4+G are confirmed with b1, b2, b3, and b'4 respectively. Peaks were labeled in different colors for product ions of glycans (red), peptides (green), and glycopeptides (blue).

Table 2. N-Glycopeptide Fragments for Annotation

abbreviation	fragments
Y0	no glycans
Y1	Y0+{G}
Y2	Y0+{G2}
Y3	Y0+{G2H}
Y4	Y0+{G2H2}
Y5	Y0+{G2H3}
Y6	Y0+{G2H4}
Y7	Y0+{G2H5}
Y8	Y0+{G2H6}
Y9	Y0+{G2H7}

is measured with the fraction of unannotated peak intensities among the top 20 peaks (Unassigned_all), the intensity of the highest unassigned peak (max_unassigned_ab, % of base peak), and number of unassigned peaks in top 20 peaks (top_20_num_unassigned_peaks) or number of unassigned peaks (num_unassigned_peaks), etc. By using these parameters, users can easily assess spectrum quality to verify peptide identification.

SOFTWARE DESCRIPTION

MS_Piano was developed in Microsoft Visual C++ 2015 and is released in 2 formats: .exe and .dll (peptide annotation only). The .exe software can be directly used without installation on the Windows operating system. The command is simple; e.g., MS_Piano C:\in.msp C:\out.msp.

Data Input and Output

The NIST text file format (msp) is used for input and output of m/z vs intensity lists. The following screenshots (Figure 5) illustrate an example spectrum of a peptide in an msp file annotated with MS_Piano software. In the input file, the following information is required for annotation: peptide

sequence, modifications (optional), charge state, precursor m/z value (Parent, optional), number of peaks (Num peaks) and a peak list with m/z and intensity values. The peptide information is listed in the Name line. In the output file, MS_Piano annotates peaks with product ions and mass difference between experimental and theoretical m/z value. The output file also provides the parameters described above to indicate spectral quality.

Visualization

Tandem mass spectra with peaks annotated with reliable product ions in the output file can be viewed (Figures 4, 6) in the MS Search program¹² to facilitate validation and generate figures for publications and presentations. The annotated peaks with fragments of peptides, glycans, and glycopeptides in a mass spectrum can be displayed in different colors by adjusting color and fonts in the MS Search version 2.5 or later.

SOFTWARE PERFORMANCE, TESTING, AND APPLICATION

MS_Piano.exe program was tested for performance on a desktop computer, Windows 10 Enterprise with Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz with 64 GB memory. As an example of performance, MS_Piano took <1.5 h to annotate 1 million high resolution mass spectra of peptides from protein digest samples by processing 4 files in parallel, each containing 250 000 identified spectra. Peptide length was 10–30, charge state from 2 to 5, with 0–4 modifications, and 30–500 peaks in 95% of these testing spectra. These spectra were first converted to msp files (mass vs intensity peak lists in text format) from the MS-GF+⁹ searching results. For glycopeptide spectra, the annotation time increases with the number of sugars in the glycans. MS_Piano was tested and refined by annotating the msp files converted from MS-GF+⁹ searching output files of all the data generated in Study 3 of the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH);

```

input:
Name: GAEGHPGERPPHSVPNNAR/4_2(0,G,TMT)(15,N,Deamidation)
Comment: Parent=553.0310
Num peaks: 138
130.1348 258.77
155.0821 64.03
430.2168 55.74
712.3787 48.43
764.8841 1000.6
765.3854 733.71
...

output:
Name: GAEGHPGERPPHSVPNNAR/4_2(0,G,TMT)(15,N,Deamidation)
Comment: Parent=553.0310 Theo_mz=553.0320 Theo_mz_diff=-1.7ppm
Unassigned_all_20ppm=0.3109 Unassigned_20ppm=0.1597
num_unassigned_peaks_20ppm=70 max_unassigned_ab_20ppm=0.33
top_20_num_unassigned_peaks_20ppm=4/20
Num peaks: 138
130.1348 258.77 "TMT130N/-0.2ppm"
155.0821 64.03 "Int/PG/3.9ppm"
430.2168 55.74 "b15-NH3^4/-9.0ppm"
712.3787 48.43 "?"
764.8841 1000.6 "y14^2/-0.4ppm"
765.3854 733.71 "y14^2+i/-0.5ppm"
...
```

Figure 5. An example spectrum of a peptide in an msp file annotated with MS_Piano.

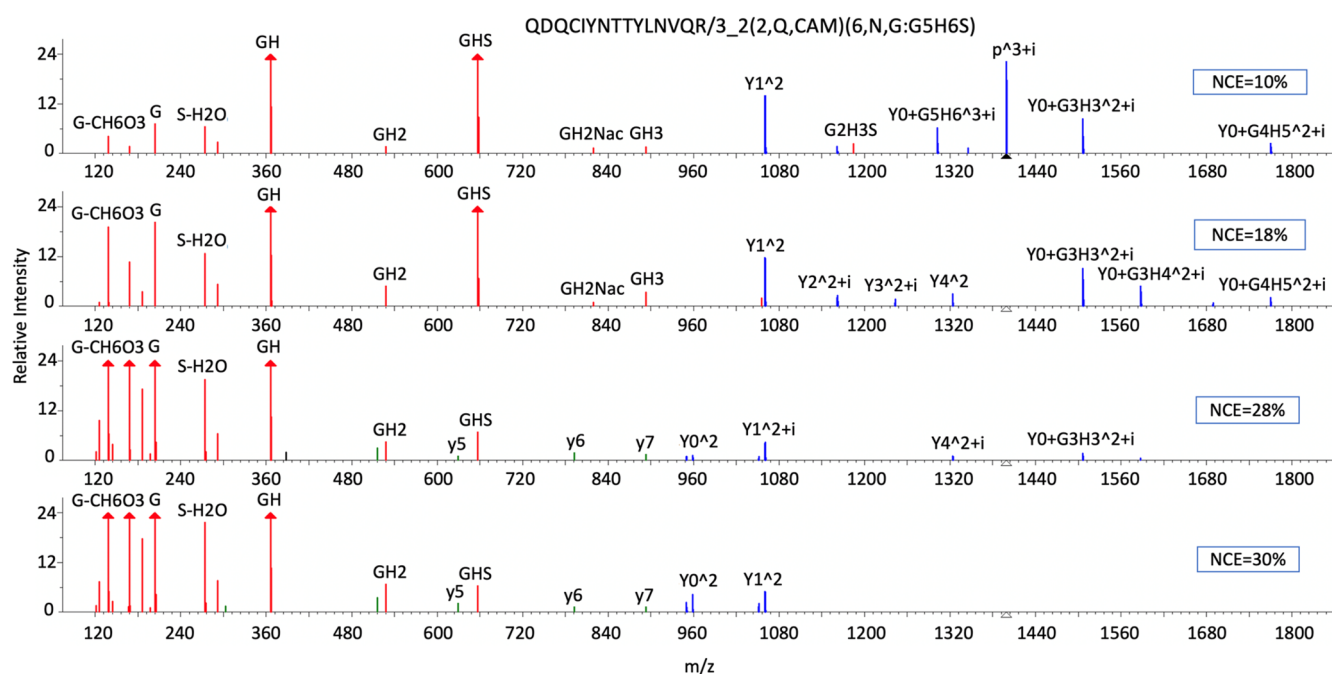


Figure 6. An example of energy dependence fragmentation of an N-glycopeptide (in Alpha-1-acid glycoprotein in plasma) in charge 3 annotated with MS_Piano. Peaks in 4 high-resolution HCD spectra at different collision energies were labeled in different colors for product ions of glycans (red), peptides (green), and glycopeptides (blue).

“Unassigned %” and other parameters described above together with MS-GF+ searching score were used to evaluate spectral quality and identification. This software has annotated 90 784 MS² spectra of all dipeptides, all tryptic tripeptides, and 1828 bioactive commercial peptides (purity >90%) in the NIST Tandem Mass Spectral Library.^{23,24} It has been used in annotating mass spectra of digests from iTRAQ and TMT labeled proteins²⁵ and human hair.²⁶ It is also used routinely to annotate high resolution mass spectra and evaluate spectral quality for optimal search for refining the NIST Peptide Tandem Mass Spectral Library¹³ with more than 4.3 million spectra. MS_Piano has been used to annotate mass spectra of protein digests of single glycoproteins (Figure 6) including the Spike Protein in SARS-CoV-2.

DISCUSSION AND CONCLUSIONS

MS_Piano enables the annotation of tandem mass spectra of both peptides and N-glycopeptides that contain virtually all known modifications. A key feature of MS_Piano is its capability to annotate one million peptide spectra in <2 h with reliable product ions, including some common but usually neglected fragments. The annotated spectra can be viewed with diverse ions displayed in different colors in NIST MS Search¹² browser for enhanced visual examination. Annotated spectra can be used to validate spectra and verify peptide identified by sequence database searching directly or embedded in other software packages to implement various proteomics data analysis software.²⁷ The output msp files of spectra annotated with MS_Piano can be used on Skyline³³ platform directly. MS_Piano provides a metric for spectrum quality and a reportable filter for constructing peptide mass spectral libraries. The software is helpful for understanding peptide fragmentation pathways. Its different formats provide flexibility for biologists and chemists to use .exe directly, or for advanced programmers to embed a functional .dll into other programs.

Due to the format complexity of raw data acquired on mass spectrometers from different manufactures and the peak lists lacking in searching results from various peptide and glycopeptide search engines, the simple text format (msp) is used for input, but users need to convert their files to this format to use this program. However, files can be combined from searching results from libraries, search engines, and de novo sequencing for annotation. We developed an msp file converter (free software convert2 msp, also available at the following MS_Piano download website) which quickly converts the results from free protein and glycoprotein sequence searching engine MSFragger³⁴ and pGlyco,³⁵ respectively, to msp files, and automatically connects to MS_Piano for spectral annotation to facilitate building-your-own libraries. More capabilities such as taking mzXML and pepXML together with raw data as input files, annotating mass spectra of negative mode, ETD and O-glycopeptides will be added to the software. Software, instructions, and examples can be freely downloaded at https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:ms_piano.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00324>.

Product ions used in MS_Piano; Table S1: Peptide modifications; Table S2: Immonium and fragment ions; Table S3: Neutral losses; Table S4: iTRAQ fragments; Table S5: TMT fragments; Figure S1: Two examples of low-resolution ion trap mass spectra of a peptide Bradykinin with sequence RPPGFSPFR at charge 2 and 3 respectively annotated with MS_Piano (XLSX)

AUTHOR INFORMATION

Corresponding Author

Xiaoyu Yang — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States; orcid.org/0000-0003-3371-9567; Email: xiaoyu.yang@nist.gov

Authors

Pedatsur Neta — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Yuri A. Mirokhin — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Dmitrii V. Tchekhovskoi — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Concepcion A. Remoroza — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States; orcid.org/0000-0003-1540-1635

Meghan C. Burke — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States; orcid.org/0000-0001-7231-0655

Yuxue Liang — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States; orcid.org/0000-0002-6430-915X

Sanford P. Markey — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Stephen E. Stein — Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jproteome.1c00324>

Notes

Certain commercial software or instruments are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the software or instruments identified are necessarily the best available for the purpose.

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Drs. Lewis Geer, Guanghui Wang, Oleg Toropov, Zheng Zhang, Sergey Sheetlin and William Wallace who have provided suggestions, useful discussions, feedback, and technical support that have helped the improvement of the software.

REFERENCES

- (1) Patterson, S. D.; Aebersold, R. H. Proteomics: the first decade and beyond. *Nat. Genet.* **2003**, 33 (Suppl), 311–323.
- (2) Han, X.; Aslanian, A.; Yates, J. R., 3rd. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008**, 12 (5), 483–490.
- (3) Yu, Q.; Paulo, J. A.; Naverrete-Perea, J.; McAlister, G. C.; Canterbury, J. D.; Bailey, D. J.; Robitaille, A. M.; Huguet, R.; Zabrouskov, V.; Gygi, S. P.; Schweppe, D. K. Benchmarking the Orbitrap Tribrid Eclipse for Next Generation Multiplexed Proteomics. *Anal. Chem.* **2020**, 92 (9), 6478–6485.
- (4) Adhikari, S.; Nice, E. C.; Deutsch, E. W.; Lane, L.; Omenn, G. S.; Pennington, S. R.; Paik, Y. K.; Overall, C. M.; Corrales, F. J.; Cristea, I. M.; Van Eyk, J. E.; Uhlén, M.; Lindskog, C.; Chan, D. W.; Bairoch, A.; Waddington, J. C.; Justice, J. L.; LaBaer, J.; Rodriguez, H.; He, F.; Kostrzewa, M.; Ping, P.; Gundry, R. L.; Stewart, P.; Srivastava, S.; Srivastava, S.; Nogueira, F. C. S.; Domont, G. B.; Vandenbrouck, Y.; Lam, M. P. Y.; Wennersten, S.; Vizcaino, J. A.; Wilkins, M.; Schwenk, J. M.; Lundberg, E.; Bandeira, N.; Marko-Varga, G.; Weintraub, S. T.; Pineau, C.; Kusebauch, U.; Moritz, R. L.; Ahn, S. B.; Palmblad, M.; Snyder, M. P.; Aebersold, R.; Baker, M. S. A high-stringency blueprint of the human proteome. *Nat. Commun.* **2020**, 11 (1), 5301.
- (5) Sun, S.; Shah, P.; Eshghi, S. T.; Yang, W.; Trikannad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Höti, N.; Zhang, Z.; Chan, D. W.; Zhang, H. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **2016**, 34 (1), 84–88.
- (6) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, 369 (6501), 330–333.
- (7) 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 (18), 3551–3567.
- (8) 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 (5), 958–964.
- (9) Byonic. <https://www.proteinmetrics.com> (accessed in December 2020).
- (10) Kim, S.; Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **2014**, 5, 5277.
- (11) Toghi Eshghi, S.; Shah, P.; Yang, W.; Li, X.; Zhang, H. GPQuest: A Spectral Library Matching Algorithm for Site-Specific Assignment of Tandem Mass Spectra to Intact N-glycopeptides. *Anal. Chem.* **2015**, 87 (10), 5181–8.
- (12) NIST MS Search browser: <https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:nistmssearch> (accessed in December 2020).
- (13) NIST Libraries of Peptide Tandem Mass Spectra. <https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:start> (accessed in December 2020).
- (14) Lam, H.; Deutsch, E. W.; Edes, J. S.; Eng, J. K.; King, N.; Stein, S. E.; Aebersold, R. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **2007**, 7 (5), 655–667.
- (15) Lam, H.; Deutsch, E. W.; Edes, J. S.; Eng, J. K.; Stein, S. E.; Aebersold, R. Building consensus spectral libraries for peptide identification in proteomics. *Nat. Methods* **2008**, 5 (10), 873–875.
- (16) Fernández-Costa, C.; Martínez-Bartolomé, S.; McClatchy, D.; Yates, J. R., 3rd. Improving Proteomics Data Reproducibility with a Dual-Search Strategy. *Anal. Chem.* **2020**, 92 (2), 1697–1701.
- (17) MS-Product. <https://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct> (accessed in December 2020).
- (18) Liang, Y.; Neta, P.; Yang, X.; Stein, S. E. Collision-Induced Dissociation of Deprotonated Peptides. Relative Abundance of Side-Chain Neutral Losses, Residue-Specific Product Ions, and Comparison with Protonated Peptides. *J. Am. Soc. Mass Spectrom.* **2018**, 29 (3), 463–469.
- (19) Kilpatrick, L. E.; Neta, P.; Yang, X.; Simón-Manso, Y.; Liang, Y.; Stein, S. E. Formation of y + 10 and y + 11 ions in the collision-induced dissociation of peptide ions. *J. Am. Soc. Mass Spectrom.* **2012**, 23 (4), 655–663.
- (20) Unimod. <https://www.unimod.org/downloads.html> (accessed in December 2020).
- (21) Kleikamp, H. B. C.; Lin, Y. M.; McMillan, D. G. G.; Geelhoed, J. S.; Naus-Wiezer, S. N. H.; van Baaren, P.; Saha, C.; Louwen, R.

Sorokin, D. Y.; van Loosdrecht, M. C. M.; et al. Tackling the chemical diversity of microbial nonulosonic acids—A universal large-scale survey approach. *Chem. Sci.* **2020**, *11*, 3074–3080.

(22) Yu, J.; Schorlemer, M.; Gomez Toledo, A.; Pett, C.; Sihlbom, C.; Larson, G.; Westerlind, U.; Nilsson, J. Distinctive MS/MS Fragmentation Pathways of Glycopeptide-Generated Oxonium Ions Provide Evidence of the Glycan Structure. *Chem. - Eur. J.* **2016**, *22* (3), 1114–1124.

(23) Yang, X.; Neta, P.; Stein, S. E. Quality control for building libraries from electrospray ionization tandem mass spectra. *Anal. Chem.* **2014**, *86* (13), 6393–6400.

(24) Yang, X.; Neta, P.; Stein, S. E. Extending a Tandem Mass Spectral Library to Include MS2 Spectra of Fragment Ions Produced In-Source and MSn Spectra. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (11), 2280–2287.

(25) Zhang, Z.; Yang, X.; Mirokhin, Y. A.; Tchekhovskoi, D. V.; Ji, W.; Markey, S. P.; Roth, J.; Neta, P.; Hizal, D. B.; Bowen, M. A.; Stein, S. E. Interconversion of Peptide Mass Spectral Libraries Derivatized with iTRAQ or TMT Labels. *J. Proteome Res.* **2016**, *15* (9), 3180–3187.

(26) Zhang, Z.; Burke, M. C.; Wallace, W. E.; Liang, Y.; Sheetlin, S. L.; Mirokhin, Y. A.; Tchekhovskoi, D. V.; Stein, S. E. Sensitive Method for the Confident Identification of Genetically Variant Peptides in Human Hair Keratin. *J. Forensic Sci.* **2020**, *65* (2), 406–420.

(27) Bittremieux, W. spectrum_utils: A Python Package for Mass Spectrometry Data Processing and Visualization. *Anal. Chem.* **2020**, *92* (1), 659–661.

(28) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.

(29) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, *59* (21), 2621–2655.

(30) Li, K.; Vaudel, M.; Zhang, B.; Ren, Y.; Wen, B. PDV: an integrative proteomics data viewer. *Bioinformatics* **2019**, *35* (7), 1249–1251.

(31) Sturm, M.; Kohlbacher, O. TOPPView: an open-source viewer for mass spectrometry data. *J. Proteome Res.* **2009**, *8* (7), 3760–3763.

(32) Proteome Discoverer. <https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/proteome-discoverer-software.html> (accessed in June 2020).

(33) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966–968.

(34) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **2017**, *14* (5), 513–520.

(35) Liu, M. Q.; Zeng, W. F.; Fang, P.; Cao, W. Q.; Liu, C.; Yan, G. Q.; Zhang, Y.; Peng, C.; Wu, J. Q.; Zhang, X. J.; Tu, H. J.; Chi, H.; Sun, R. X.; Cao, Y.; Dong, M. Q.; Jiang, B. Y.; Huang, J. M.; Shen, H. L.; Wong, C. C. L.; He, S. M.; Yang, P. Y. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat. Commun.* **2017**, *8* (1), 438.

(36) Polasky, D. A.; Yu, F.; Teo, G. C.; Nesvizhskii, A. I. Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods* **2020**, *17* (11), 1125–1132.