DOI 10.1002/pmic.201700091 *Proteomics 17, 15–16*, 2017, 1700091

RESEARCH ARTICLE

multiplierz v2.0: A Python-based ecosystem for shared access and analysis of native mass spectrometry data

William M. Alexander[1,2], Scott B. Ficarro[1,2], Guillaume Adelmant[1,2] and Jarrod A. Marto[1,3,4]

[1] Department of Cancer Biology and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA, USA
[2] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA
[3] Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA, USA
[4] Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

The continued evolution of modern mass spectrometry instrumentation and associated methods represents a critical component in efforts to decipher the molecular mechanisms which underlie normal physiology and understand how dysregulation of biological pathways contributes to human disease. The increasing scale of these experiments combined with the technological diversity of mass spectrometers presents several challenges for community-wide data access, analysis, and distribution. Here we detail a redesigned version of *multiplierz*, our Python software library which leverages our common application programming interface (*mzAPI*) for analysis and distribution of proteomic data. New features include support for a wider range of native mass spectrometry file types, interfaces to additional database search engines, compatibility with new reporting formats, and high-level tools to perform post-search proteomic analyses. A GUI desktop environment, *mzDesktop*, provides access to *multiplierz* functionality through a user friendly interface. *multiplierz* is available for download from: https://github.com/BlaisProteomics/multiplierz; and *mzDesktop* is available for download from: https://sourceforge.net/projects/multiplierz/

Received: March 7, 2017
Revised: June 25, 2017
Accepted: June 28, 2017

Keywords:
Application programming interface (API) / Python

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

In a continued effort to achieve ever-more extensive coverage of the proteome, academic researchers and manufacturers have developed a myriad of mass spectrometer scan functions along with new or refined ionization and fragmentation techniques to improve detection, reproducibility, quantification accuracy and other analytical figures of merit [1–12]. As a result modern mass spectrometers are quite diverse, comprising a wide range of technologies and performance capabilities, along with a correspondingly diverse array of vendor-specific data formats. Taken together, the proprietary file formats, size, and high-dimensionality of mass spectrometry data, along with continued technological advances, complicates the task of ensuring shared access and reproducible analysis across the research community.

Based on the notion that instrument manufacturers are commercially incentivized to develop and maintain efficient binary file formats for their data, we proposed an open, common application programming interface (API) [13] as an alternative to common file formats [14] for shared access to native mass spectrometry binary files. APIs are a key element of modern software development as they enable expert programmers and novice users to utilize powerful software libraries or other computational infrastructure without intimate knowledge of the underlying implementation. A well-engineered common API for native mass spectrometry

Correspondence: Dr. Jarrod A. Marto, Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Longwood Center 2208, Boston, MA, 02115–5450, USA
E-mail: jarrod_marto@dfci.harvard.edu

Colour Online: See the article online to view Figs. 1, 2, 4 and 5 in colour.

Significance of the study

We introduce a redesign of our open-source, Python software library and desktop environment which leverages our recently described common application programming interface (*mzAPI*) for access, analysis, and distribution of mass spectrometry-based proteomic data. Significant extensions include support for a wider range of native mass spectrometry file types, database search engines, and reporting formats. Interactive GUIs are provided for common tasks, while the underlying Python architecture facilitates rapid prototyping of custom workflows.

data leverages the existing file indices to provide fast and efficient programmatic access to the binary file architecture, while also ensuring that users can leverage manufacturer data system- and instrument-specific functionality which may be absent from common file format definitions, or simply ignored by users when they utilize file extraction routines. In addition, working exclusively with vendor data files via a common API satisfies regulatory criteria and also minimizes file size 'bloat' which is a frequent by-product when native binary files are extracted to self-describing formats.

Using *mzAPI* as a foundation for transparent data access we developed *multiplierz*, a scriptable Python software library for access to and analysis of mass spectrometry data [15]. As an easy to learn scripting language, Python provides an ideal environment for rapid prototyping of data analytic tools [16], and has evolved to include powerful tools for visualization (matplotlib), data sharing (iPython), statistical analysis (numpy and scipy), and importantly provides a high level of interoperability with R (rpy2), the preferred open-source environment for rigorous analysis of large-scale molecular data [17, 18]. More recently several proteomics-focused software tools have leveraged Python in an effort to motivate wider acceptance and experimentation across the scientific community [19–22]. Here we describe improvements to the *multiplierz* library, including extension of our *mzAPI* implementation to accommodate a broader range of mass spectrometer file formats, and a library of new, native-Python proteomic analysis tools. This new codebase provides an accessible, open-source ecosystem for proteomic data access, analysis, and distribution.

2 Design and architecture

Our open-source ecosystem for analysis of native mass spectrometry data is comprised of four major sub-packages (Fig. 1): (i) mzAPI, for providing access to raw mass spectrometry data files [13]; (ii) mzSearch, for programmatically submitting data to various search engines; (iii) mzReport, for viewing and processing search result data; and (iv) mzTools, for post-acquisition data processing and analytic tasks. The most recent build (version 2.0) is written for Python 2 and is distributed under the GNU Lesser General Public License. Primary improvements and additions are described below,

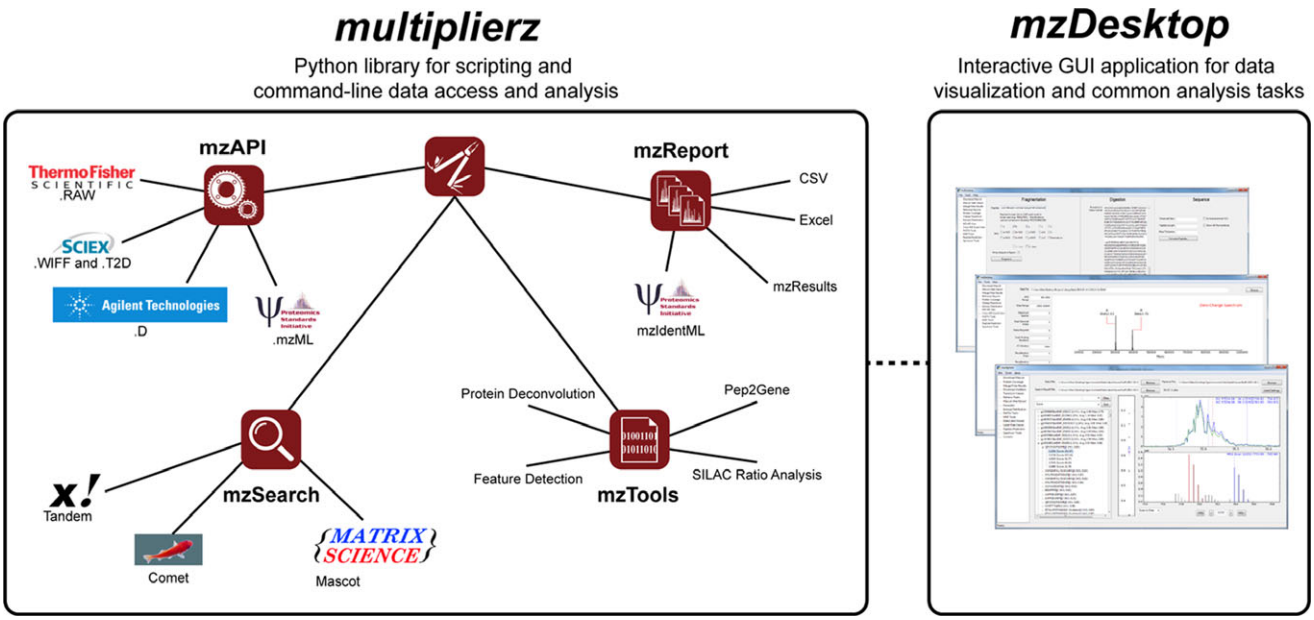

Figure 1. (left) *multiplierz* provides an open and extensible Python environment for rapid prototyping of mass spectrometry software tools and data analysis pipelines. (right) *mzDesktop* provides interactive GUIs for *multiplierz* functionality.


```
from multiplierz.mzAPIimport mzFile
import matplotlib.pyplot as pyplot

data = mzFile ('H9_MS_2.t2d')
mzs, intensities = zip(*data.scan())
pyplot.vlines(mzs, [0] * len(intensities), intensities)

pyplot.savefig('figure_1.svg')
pyplot.show()
```

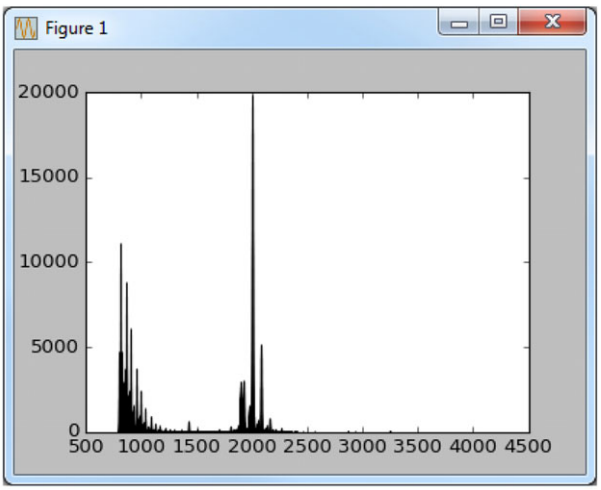

Figure 2. *multiplierz* script which provides single spectrum output for MALDI TOF-TOF (.T2D format) data.

with a complete overview of the tools provided in Supporting Information Table 1.

2.1 mzAPI

The mzAPI interface abstracts across many instrument-specific hardware, firmware, and acquisition capabilities to provide users with a powerful environment for programmatic access to mass spectrometry data. In addition to supporting high-dimension LC-MS/MS native data formats from ThermoFisher Scientific (.RAW), Danaher AB Sciex (.WIFF), and Agilent (.D), we have added support for simple MALDI experiments which often consist of individual scans (e.g. no time-dependent separation component). Figure 2 illustrates the function which returns a single spectrum from a 4800 MALDI TOF-TOF instrument (.T2D files). Internally, native data files are accessed through a set of proprietary Windows libraries provided by the respective instrument vendors and included in the *multiplierz* installer; as such mzAPI is currently limited to Windows data systems. To facilitate interoperability with existing workflows, we included support for the most recent definition of the mzML common file format [23]. In addition, we provide an embedded peak list extraction function for subsequent database searches.

2.2 mzSearch

A new feature in this version of *multiplierz* is mzSearch, a package to abstract the interface details for protein identification software; mzSearch currently supports three popular engines (Mascot, Comet and XTandem) and is readily extensible to enable complex consensus searchers. A scripted mzSearch session proceeds in three steps; an mzSearch object is initialized that loads a file of parameters for the target search engine; optionally, parameters are accessed or changed by reference to corresponding attributes of the object; finally, a search method is called against one or more data files, which invokes the search engine associated with the object.

2.3 mzReport

The third Python module is mzReport, which enables users to easily generate, interrogate, and distribute results from proteomic experiments; mzReport supports multiple formats, including simple CSV or Excel files, interactive .mzD SQLite files [24], as well as the evolving mzIndentML standard [25, 26]. As we [24] and others [27–29] have previously discussed, rapid technological developments in proteomic research make it difficult to establish stable reporting standards while also maintaining the flexibility to browse the underlying data and test alternative hypotheses as new tools become available. The architecture of our *multiplierz* environment provides seamless integration with native data files via mzAPI, enabling use, distribution, and further development (through Python) of dynamic reporting formats [24].

2.4 mzTools

The mzTools package enables users to rapidly prototype new utilities for various aspects of mass spectrometry data analysis. Of particular interest, we have further optimized and expanded our recently described Pep2Pro utility [30]. This new tool (now termed Pep2Gene) enables rapid annotation of peptides with respect to their source proteins, protein isoforms, and parent genes based on an application-/database-specific sequence-to-gene dictionary. Briefly, this is an optimized form of the k-mer-to-protein mapping database algorithm previously described, whereby a lookup table is compiled that maps all 4-mers present in a sequence database to the corresponding set of source proteins; decomposing a peptide into successive 4-mers and identifying the corresponding protein sets by lookup vastly reduces the search space required to perform peptide-to-protein mapping. From there, a protein-to-gene mapping function is performed, using data obtained from the Uniprot KnowledgeBase [31, 32]. Creation of a Pep2Gene database for the human proteome (e.g. UniProt) takes less than 30 min. on a modern desktop

computer, after which annotating ≥50 000 peptides, as is typical in large-scale proteome experiments, can be completed in a few minutes. As a result our implementation of Pep2Gene is well-suited for deployment in data analytic pipelines built around modest processing power. Inspired by efforts in label-free quantification [33, 34], we have also implemented a sophisticated algorithm to detect and catalog MS and MS/MS spectral features; this algorithm provides a foundation for relative peptide quantification by SILAC (*multiplierz* currently supports duplex and triplex labeling schemes) or label-free analyses. We provide documentation for a full *multiplierz* peptide identification and SILAC quantitation workflow performed on a third-party data set (*multiplierz* Github repository); our analysis agreed well with the original quantification performed via Proteome Discoverer [35]. For intact protein mass spectrometry applications, we implement spectral deconvolution as originally described by Fenn [36] and Marshall [37] (Fig. 3).

2.5 mzDesktop

In its original implementation, *multiplierz* served primarily as a GUI-based application to expose mzAPI and provide a modest set of tools for basic analysis of mass spectrometry data.

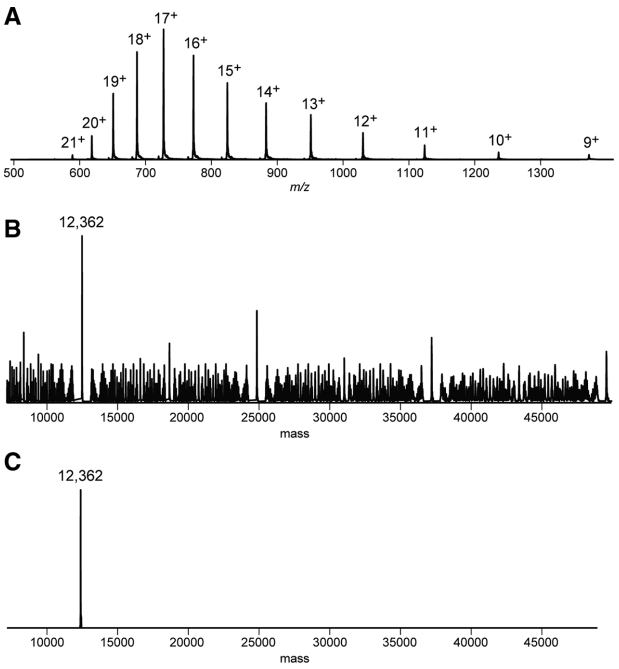

Figure 3. (A) low resolution mass spectrum containing multiple charge states of cytochrome C obtained by electrospray ionization on a ThermoFisher LTQ linear ion trap. (B) Same spectrum after deconvolution based on Mann and Fenn's algorithm [36] or (C) the Zscore method described by Zhang and Marshall [37]. Both methods yield a final charge-reduced peak within 0.4 Da of the average molecular weight for cytochrome C.

Beginning with the release described herein we have renamed the GUI application '*mzDesktop*' to distinguish it from the broader *multiplierz* project. In general, most functions now available in the *multiplierz* Python package have GUI-based equivalents available through *mzDesktop*: the Data Viewer (Fig. 4) provides direct access to native mass spectrometry files via mzAPI; database search engine interfaces powered by the mzSearch API can be used to manually launch protein/peptide identification jobs; mzReports-compatible files (as well as mzIdentML) can be opened by the *mzDesktop* file reader, and so on. We have also included several GUI-specific functions. One example is the Peptide Coverage viewer which allows users to quickly ascertain the degree of protein sequence coverage supported by a given set of PSMs and generate presentation-quality graphics (Supporting Information Fig. 1).

3 Analysis of complex mass spectrometry datasets

We have developed novel high-performance liquid chromatography (LC) assemblies for fully automated single- and multiple-dimension peptide separation coupled directly to mass spectrometry for in-depth interrogation of mammalian proteomes [38–43]. Optimization of total peak capacity for these high-performance platforms requires systematic analysis of peptide elution profiles across multiple separation dimensions. Moreover, the use of multiplexed isotope labeling reagents to enable higher throughput, quantitative analysis may alter the physicochemical behavior of tryptic peptides.

In an effort to understand how iTRAQ [44] and TMT [45] reagents may influence peptide elution in 3-dimension RP-SAX-RP-MS/MS experiments, we digested HeLa cell protein lysates and then created equal aliquots of unlabeled, iTRAQ 4-Plex, iTRAQ 8-Plex, and TMT 6-Plex labeled tryptic peptides. All peptides were mixed and then analyzed by 3D-RP-SAX-RP at a depth of 32 fractions (.RAW data files available for download at: ftp://massive.ucsd.edu/MSV000081101). Our *multiplierz* pipeline output the subset of 3127 sequence-unique tryptic peptides identified in common across labeled and unlabeled peptides. We generated XICs via random file access with mzAPI (32 .RAW files, 120 GB) to identify the fraction corresponding to the apex elution for each peptide (documentation available on the *multiplierz* Github repository). Next, these data were collated to create two-dimensional plots to illustrate the number of peptides identified in each first and second dimension fraction (Fig. 5A–E). As expected the addition of each isobaric tag increased peptide retention compared to unlabeled analogues (Fig. 5F). Approximately one-half of iTRAQ-labeled peptides were more strongly retained compared to their unlabeled analogs. Surprisingly peptides labeled with TMT multiplex reagents exhibited a dramatic shift in retention, requiring higher salt and acetonitrile concentrations for elution from the first- and second-dimension columns, respectively. This effect is presumably

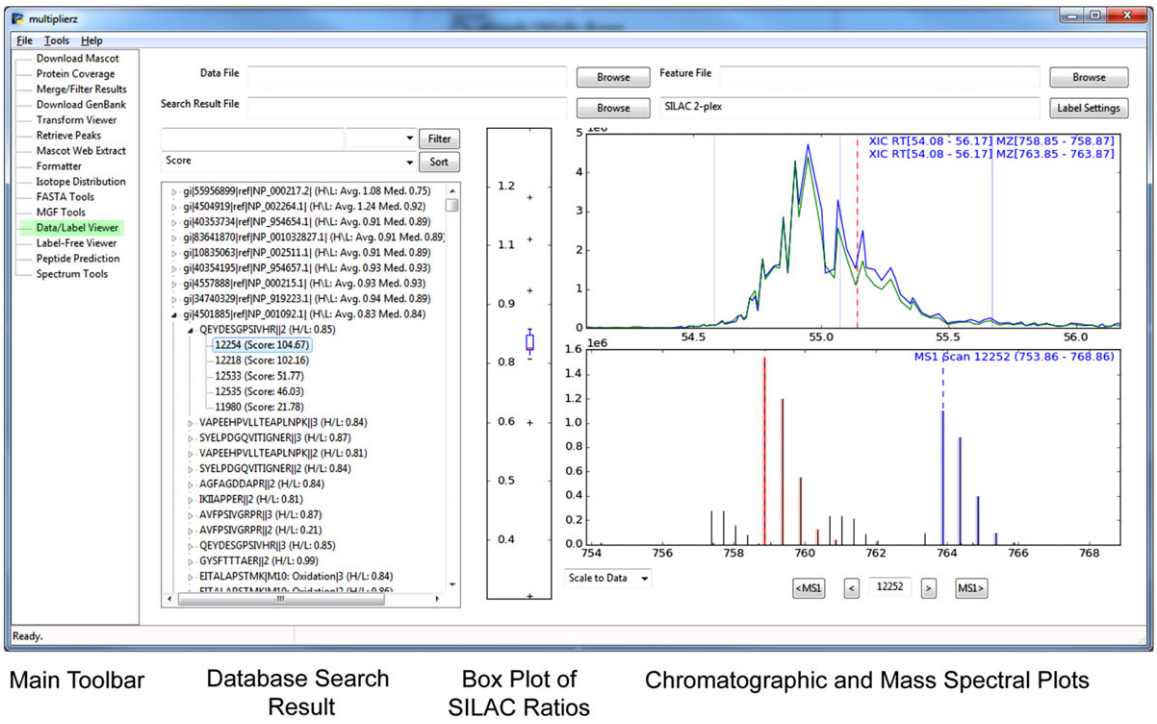

Figure 4. The Data Viewer is a feature-rich GUI included in *mzDesktop* which leverages mzAPI for direct access to native mass spectrometry data files.

due to the greater basicity and hydrophobicity of the dimethylpiperidine in the TMT scaffold compared to iTRAQ's methylpiperazine moiety. This example highlights the ease with which *multiplierz* can support custom data workflows; in this specific case providing a readily interpretable yet powerful quantitative visualization which informs adjustment of experimental parameters to maintain separation power for label-based or label-free proteomic strategies.

4 Discussion

Increased breadth and depth of mass spectrometry-based studies continues to drive proliferation of commercial and open-source software tools for customized data analysis, and distribution [27, 46]. Efforts in the former were led by Pedrioli in 2004 through extraction of native data to an XML-based common file format [14]. Human-readable formats provide cross-platform compatibility and can live in-perpetuity independently from the source data and original instrument manufacturer. These formats continue to evolve and are supported by a diverse ecosystem of software tools and frameworks for bioinformatic analysis [47–58]. These advantages are tempered by the technical difficulties of establishing efficient schemes for random access to high-dimension data incorporated within XML-formatted files. As a result, recent iterations of XML common file formats for mass spectrometry have sought to establish an acceptable compromise between raw-/meta-data content and performance for sequential versus random scan/file access within the system memory typically available to a broad range of end-users [23, 59–61]. The increasing acquisition speed of modern mass spectrometers coupled with the trend toward more comprehensive studies will accentuate these challenges, and fuel the debate regarding the technical merit of human-readable formats for complex numerical data [13, 62–64].

As an alternative to common file formats we have promoted the use of a common API [13, 15] for mass spectrometry in order to leverage optimized indexing and architecture of the native data files. Our strategy is well-aligned with trends across a wide array of scientific fields, wherein the domain-specific computational tasks are codified into a series of reusable software libraries; for instance, the particular demands of data analysis in astronomy led to the development of AstroPy, a library facilitating access to efficient binary data formats, unit- and coordinate-aware computation, image convolution tools, and other related capabilities [65]. It should be noted that our emphasis on a common API sacrifices the permanent archival nature of human-readable formats, although the future maintainability of workflows based around either extraction (XML) or abstraction (API) of native data files relies on continued support from mass spectrometry instrument manufacturers. While community-defined standards in proteomics have coalesced around XML-formatted files, we note that an increasing set of software tools [49, 66, 67] designed for mass spectrometry now tout an ability to access and analyze

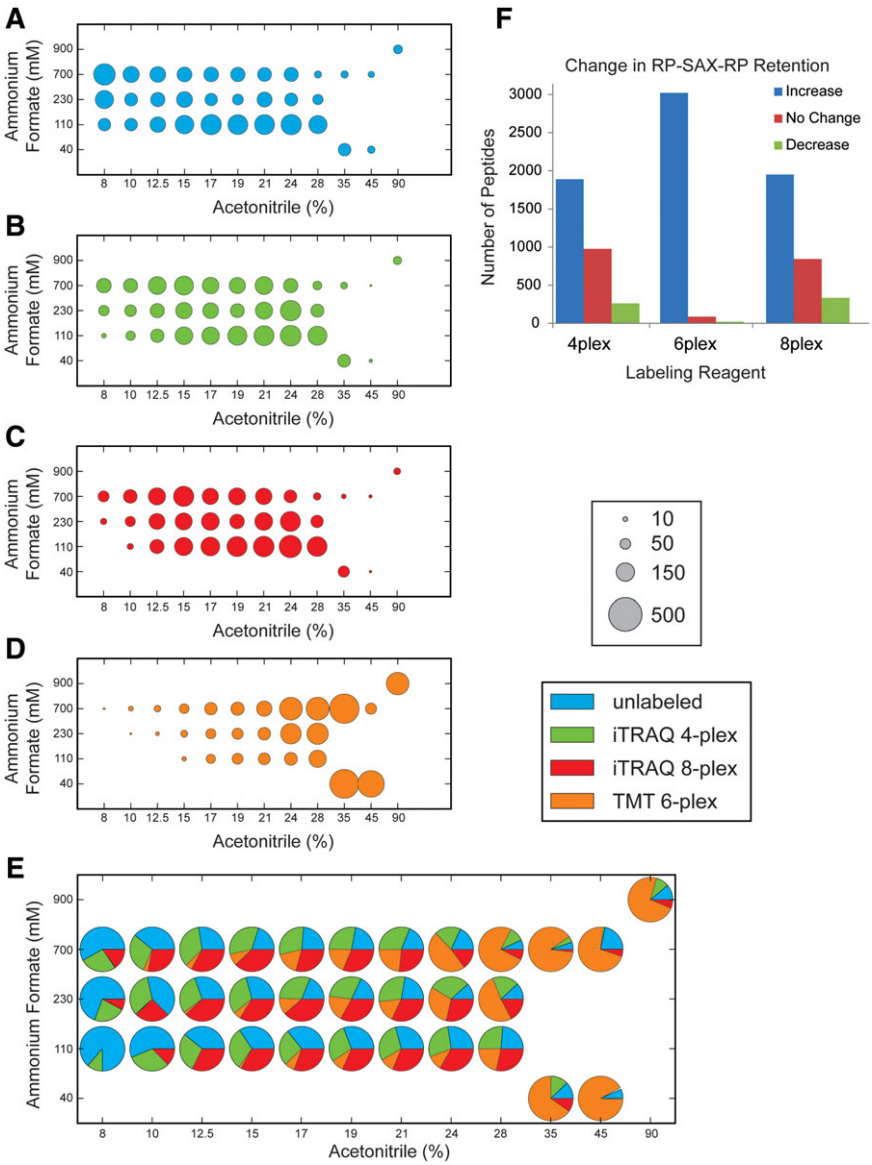

Figure 5. *multiplierz* analysis of changes peptide retention on RP-SAX-RP resulting from use of TMT or iTRAQ isobaric stable isotope labels. The set of sequences identified in common for (A) unlabeled, (B) iTRAQ 4-plex labeled, (C) iTRAQ 8-plex labeled, and (D) TMT 6-plex labeled peptides are plotted as a function of first- (x-axis, high-pH RP) and second-dimension (y-axis, high-pH SAX) elution conditions. For each plot the number of peptides in each fraction is represented by a circle of proportional diameter. (E) The relative contribution of labeled and unlabeled peptides to the total identified in each 3D fraction is represented as a multi-colored pie chart. (F) Bar charts illustrating the quantitative change in peptide retention across 3D fractions as a result of labeling with iTRAQ or TMT reagents.

native data files directly, suggesting that a significant subset of researchers are productive using proteomic workflows start-to-finish in the Windows environment. Consistent with this trend, *multiplierz* and *mzDesktop* leverage our common API to provide unfettered access to native mass spectrometry data files, and we will continue to expand the set of file formats supported by mzAPI.

High-level scientific software libraries are playing an increasingly important role at the intersection of complex instrumentation, high-dimensional data, and federated computational resources. As currently implemented, *multiplierz* provides all the components necessary to build a complete front-to-back data analysis workflow, through an interface designed with novice programmers in-mind; common tasks can be performed in several lines of Python code, and scripts can be easily generalized across different instruments and search applications. Moreover, we designed *multiplierz* to minimize conceptual overhead; details of complex tasks are encapsulated within functions, which fulfill well-understood roles at each stage of proteomic data analysis [28, 68]. Importantly, native Python data types are returned whenever possible. Experienced programmers can use the algorithms and tools provided as a jumping-off point to write additional functions specifically suited to their requirements, or to build workflows that integrate other scientific computing capabilities available in Python. Finally, by enabling the creation of concise, transparent, and complete data processing scripts, *multiplierz* allows the details of an analysis to be easily archived and shared, facilitating experimental transparency and reproducibility. In addition to adding capabilities for desktop analysis and increasing support for other native data formats, our future development efforts will include porting

GUI features available in *mzDesktop* to our *mzServer* web resource [69] to enable remote, web-based, and platform-independent access to native mass spectrometry data files.

multiplierz is available for download from: https://github.com/BlaisProteomics/multiplierz; and *mzDesktop* is available for download from: https://sourceforge.net/projects/multiplierz/

This work was supported by the Dana-Farber Strategic Research Initiative and Barr Program in Basic Research, in addition to the N.I.H. (CA188881 and CA178860), the Honorable Tina Brozman Foundation for Ovarian Cancer Research, and the Michael J. Fox Foundation.

The authors have declared no conflict of interest.

5 References

[1] Bodzon-Kulakowska, A., Suder, P., Imaging mass spectrometry: instrumentation, applications, and combination with other visualization techniques. *Mass Spectrom. Rev.* 2016, *35*, 147–169.

[2] Brodbelt, J. S., Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* 2014, *43*, 2757–2783.

[3] Hu, Q., Noll, R. J., Li, H., Makarov, A. et al., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* 2005, *40*, 430–443.

[4] Medzihradszky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M. et al., The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* 2000, *72*, 552–558.

[5] Morris, H. R., Paxton, T., Dell, A., Langhorne, J. et al., High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* 1996, *10*, 889–896.

[6] Nikolaev, E. N., Kostyukevich, Y. I., Vladimirov, G. N., Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: theory and simulations. *Mass Spectrom. Rev.* 2016, *35*, 219–258.

[7] Qi, Y., Volmer, D. A., Electron-based fragmentation methods in mass spectrometry: an overview. *Mass Spectrom. Rev.* 2017, *36*, 4–15.

[8] Rodriguez-Suarez, E., Whetton, A. D., The application of quantification techniques in proteomics for biomedical research. *Mass Spectrom. Rev.* 2013, *32*, 1–26.

[9] Schneider, B. B., Nazarov, E. G., Londry, F., Vouros, P. et al., Differential mobility spectrometry/mass spectrometry history, theory, design optimization, simulations, and applications. *Mass Spectrom. Rev.* 2016, *35*, 687–737.

[10] Schwartz, J. C., Senko, M. W., Syka, J. E., A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2002, *13*, 659–669.

[11] Syka, J. E., Marto, J. A., Bai, D. L., Horning, S. et al., Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* 2004, *3*, 621–626.

[12] Xia, Y., McLuckey, S. A., Evolution of instrumentation for the study of gas-phase ion/ion chemistry via mass spectrometry. *J. Am. Soc. Mass Spectrom.* 2008, *19*, 173–189.

[13] Askenazi, M., Parikh, J. R., Marto, J. A., mzAPI: a new strategy for efficiently sharing mass spectrometry data. *Nat. Methods* 2009, *6*, 240–241.

[14] Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M. et al., A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.* 2004, *22*, 1459–1466.

[15] Parikh, J. R., Askenazi, M., Ficarro, S. B., Cashorali, T. et al., multiplierz: an extensible API based desktop environment for proteomics data analysis. *BMC Bioinformatics* 2009, *10*, 364.

[16] Lutz, M., *Programming Python*, O'Reilly Media 1996.

[17] Ihaka, R., Gentleman, R., R: a language for data analysis and graphics. *J. Comput. Graph. Statist.* 1996, *5*, 299–314.

[18] Gentleman, R., Carey, V., Bates, D., Bolstad, B. et al., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004, *5*, R80.

[19] Rost, H. L., Schmitt, U., Aebersold, R., Malmstrom, L., pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. *Proteomics* 2014, *14*, 74–77.

[20] Bald, T., Barth, J., Niehues, A., Specht, M. et al., pymzML–Python module for high-throughput bioinformatics on mass spectrometry data. *Bioinformatics.* 2012, *28*, 1052–1053.

[21] Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., Gorshkov, M. V., Pyteomics–a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.* 2013, *24*, 301–304.

[22] Kremer, L. P., Leufken, J., Oyunchimeg, P., Schulze, S. et al., Ursgal, universal Python module combining common bottom-up proteomics tools for large-scale analysis. *J. Proteome Res.* 2016, *15*, 788–794.

[23] Martens, L., Chambers, M., Sturm, M., Kessner, D. et al., mzML–a community standard for mass spectrometry data. *Mol. Cell. Proteomics* 2011, *10*, R110 000133.

[24] Webber, J. T., Askenazi, M., Marto, J. A., mzResults: an interactive viewer for interrogation and distribution of proteomics results. *Mol. Cell. Proteomics* 2011, *10*, M110.003970–M003110.003970.

[25] Eisenacher, M., mzIdentML: an open community-built standard format for the results of proteomics spectrum identification algorithms. *Methods Mol. Biol.* 2011, *696*, 161–177.

[26] Jones, A. R., Eisenacher, M., Mayer, G., Kohlbacher, O. et al., The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol. Cell. Proteomics* 2012, *11*, M111 014381.

[27] Perez-Riverol, Y., Alpi, E., Wang, R., Hermjakob, H. et al., Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics* 2015, *15*, 930–949.

[28] Perez-Riverol, Y., Wang, R., Hermjakob, H., Muller, M. et al., Open source libraries and frameworks for mass spectrometry based proteomics: a developer's perspective. *Biochim. Biophys. Acta* 2014, *1844*, 63–76.

[29] Ghali, F., Krishna, R., Lukasse, P., Martinez-Bartolome, S. et al., Tools (viewer, library and validator) that facilitate use of the peptide and protein identification standard format, termed mzIdentML. *Mol. Cell. Proteomics* 2013, *12*, 3026–3035.

[30] Askenazi, M., Marto, J. A., Linial, M., The complete peptide dictionary–a meta-proteomics resource. *Proteomics* 2010, *10*, 4306–4310.

[31] Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C. et al., UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2004, *32*, D115–D119.

[32] The UniProt Consortium, UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017, *45*, D158–D169.

[33] Wiener, M. C., Sachs, J. R., Deyanova, E. G., Yates, N. A., Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal. Chem.* 2004, *76*, 6085–6096.

[34] Wang, W., Zhou, H., Lin, H., Roy, S. et al., Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* 2003, *75*, 4818–4826.

[35] Wu, X., Zahari, M. S., Renuse, S., Nirujogi, R. S. et al., Phosphoproteomic analysis identifies focal adhesion kinase 2 (FAK2) as a potential therapeutic target for tamoxifen resistance in breast cancer. *Mol. Cell. Proteomics* 2015, *14*, 2887–2900.

[36] Mann, M. M., Chin Kai, Fenn, John B., Interpreting mass spectra of multiply charged ions. *Anal. Chem.* 1989, *61*, 1702–1708.

[37] Zhang, Z., Marshall, A. G., A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra. *J. Am. Soc. Mass Spectrom.* 1998, *9*, 225–233.

[38] Ficarro, S. B., Zhang, Y., Carrasco-Alfonso, M. J., Garg, B. et al., Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. *Mol. Cell. Proteomics* 2011, *10*, O111 011064.

[39] Ficarro, S. B., Zhang, Y., Lu, Y., Moghimi, A. R. et al., Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal. Chem.* 2009, *81*, 3440–3447.

[40] Zhou, F., Cardoza, J. D., Ficarro, S. B., Adelmant, G. O. et al., Online nanoflow RP-RP-MS reveals dynamics of multicomponent Ku complex in response to DNA damage. *J. Proteome Res.* 2010, *9*, 6242–6255.

[41] Zhou, F., Lu, Y., Ficarro, S. B., Adelmant, G. et al., Genome-scale proteome quantification by DEEP SEQ mass spectrometry. *Nat Commun* 2013, *4*, 2171.

[42] Zhou, F., Lu, Y., Ficarro, S. B., Webber, J. T. et al., Nanoflow low pressure high peak capacity single dimension LC-MS/MS platform for high-throughput, in-depth analysis of mammalian proteomes. *Anal. Chem.* 2012, *84*, 5133–5139.

[43] Zhou, F., Sikorski, T. W., Ficarro, S. B., Webber, J. T. et al., Online nanoflow reversed phase-strong anion exchange-reversed phase liquid chromatography-tandem mass spectrometry platform for efficient and in-depth proteome sequence analysis of complex organisms. *Anal. Chem.* 2011, *83*, 6996–7005.

[44] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B. et al., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 2004, *3*, 1154–1169.

[45] Thompson, A., Schafer, J., Kuhn, K., Kienle, S. et al., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 2003, *75*, 1895–1904.

[46] Wang, R., Perez-Riverol, Y., Hermjakob, H., Vizcaino, J. A., Open source libraries and frameworks for biological data visualisation: a guide for developers. *Proteomics* 2015, *15*, 1356–1374.

[47] Uszkoreit, J., Maerkens, A., Perez-Riverol, Y., Meyer, H. E. et al., PIA: an intuitive protein inference engine with a web-based user interface. *J. Proteome Res.* 2015, *14*, 2988–2997.

[48] Perez-Riverol, Y., Xu, Q.-W., Wang, R., Uszkoreit, J. et al., PRIDE Inspector Toolsuite: moving toward a universal visualization tool for proteomics data standard formats and quality assessment of ProteomeXchange datasets. *Mol. Cell. Proteomics* 2016, *15*, 305–317.

[49] Kessner, D., Chambers, M., Burke, R., Agus, D. et al., ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics.* 2008, *24*, 2534–2536.

[50] Deutsch, E. W., Albar, J. P., Binz, P. A., Eisenacher, M. et al., Development of data representation standards by the human proteome organization proteomics standards initiative. *J. Am. Med. Inform. Assoc.* 2015, *22*, 495–506.

[51] Sturm, M., Bertsch, A., Gropl, C., Hildebrandt, A. et al., OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinform.* 2008, *9*, 163.

[52] Deutsch, E. W., Mendoza, L., Shteynberg, D., Slagel, J. et al., Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics-Clin. Appl.* 2015, *9*, 745–754.

[53] Griss, J., Jones, A. R., Sachsenberg, T., Walzer, M. et al., The mzTab data exchange format: communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience. *Mol. Cell. Proteomics* 2014, *13*, 2765–2775.

[54] Walzer, M., Qi, D., Mayer, G., Uszkoreit, J. et al., The mzQuantML data standard for mass spectrometry-based quantitative studies in proteomics. *Mol. Cell. Proteomics* 2013, *12*, 2332–2340.

[55] Orchard, S., Binz, P. A., Borchers, C., Gilson, M. K. et al., Ten years of standardizing proteomic data: a report on the HUPO-PSI. *Spring Workshop*: April 12–14th, 2012, San Diego, USA. *Proteomics* 2012, *12*, 2767–2772.

[56] Yu, K., Salomon, A. R., HTAPP: high-throughput autonomous proteomic pipeline. *Proteomics* 2010, *10*, 2113–2122.

[57] Yu, K., Salomon, A. R., PeptideDepot: flexible relational database for visual analysis of quantitative proteomic data

and integration of existing protein information. *Proteomics* 2009, *9*, 5350–5358.

[58] Perez-Riverol, Y., Uszkoreit, J., Sanchez, A., Ternent, T. et al., ms-data-core-api: an open-source, metadata-oriented library for computational proteomics. *Bioinformatics* 2015, *31*, 2903–2905.

[59] Deutsch, E., mzML: a single, unifying data format for mass spectrometer output. *Proteomics* 2008, *8*, 2776–2777.

[60] Deutsch, E. W., File formats commonly used in mass spectrometry proteomics. *Mol. Cell. Proteomics* 2012, *11*, 1612–1621.

[61] Rost, H. L., Schmitt, U., Aebersold, R., Malmstrom, L., Fast and efficient XML data access for next-generation mass spectrometry. *PLoS One* 2015, *10*, e0125108.

[62] Lin, S. M., Zhu, L., Winter, A. Q., Sasinowski, M. et al., What is mzXML good for? *Exp. Rev. Proteomics* 2005, *2*, 839–845.

[63] Wilhelm, M., Kirchner, M., Steen, J. A., Steen, H., mz5: space- and time-efficient storage of mass spectrometry data sets. *Mol. Cell. Proteomics* 2012, *11*, O111 011379.

[64] Greenfield, P., Droettboom, M., Bray, E., ASDF: a new data format for astronomy. *Astronomy Comput.* 2015, *12*, 240–251.

[65] Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M. et al., Astropy: a community python package for astronomy. *Astronomy Astrophys.* 2013, *558*, 1–9.

[66] 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.

[67] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M. et al., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, *26*, 966–968.

[68] Smith, R., Mathis, A. D., Ventura, D., Prince, J. T., Proteomics, lipidomics, metabolomics: a mass spectrometry tutorial from a computer scientist's point of view. *BMC Bioinformatics* 2014, *15*, S9.

[69] Askenazi, M., Webber, J. T., Marto, J. A., mzServer: web-based programmatic access for mass spectrometry data analysis. *Mol. Cell. Proteomics* 2011, *10*, M110.003988.

