*Published online 8 May 2018* *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* ***W473–W478***
*doi: 10.1093/nar/gky353*

# xiSPEC: web-based visualization, analysis and sharing of proteomics data

**Lars Kolbowski[1,2], Colin Combe[1] and Juri Rappsilber[1,2,*]**

[1]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK and [2]Bioanalytics, Institute of Biotechnology, Technische Universität Berlin, 13355 Berlin, Germany

Received February 17, 2018; Revised April 17, 2018; Editorial Decision April 21, 2018; Accepted April 24, 2018

## ABSTRACT

**We present xiSPEC, a standard compliant, next-generation web-based spectrum viewer for visualizing, analyzing and sharing mass spectrometry data. Peptide-spectrum matches from standard proteomics and cross-linking experiments are supported. xiSPEC is to date the only browser-based tool supporting the standardized file formats mzML and mzIdentML defined by the proteomics standards initiative. Users can either upload data directly or select files from the PRIDE data repository as input. xiSPEC allows users to save and share their datasets publicly or password protected for providing access to collaborators or readers and reviewers of manuscripts. The identification table features advanced interaction controls and spectra are presented in three interconnected views: (i) annotated mass spectrum, (ii) peptide sequence fragmentation key and (iii) quality control error plots of matched fragments. Highlighting or selecting data points in any view is represented in all other views. Views are interactive scalable vector graphic elements, which can be exported, e.g. for use in publication. xiSPEC allows for re-annotation of spectra for easy hypothesis testing by modifying input data. xiSPEC is freely accessible at http://spectrumviewer.org and the source code is openly available on https://github.com/Rappsilber-Laboratory/xiSPEC.**

## INTRODUCTION

Mass spectra are the foundation of proteomics. Their analysis leads to identifications of peptides which in turn identify the proteins present in the sample (1,2). In the case of cross-linking experiments, linkage sites within the peptides are also identified (3,4). This provides proximity information for pairs of amino acid residues which can elucidate native protein structures (5) or protein–protein networks (6,7). In modern proteomics experiments thousands of spectra are generated, which necessitates automated algorithmic matching of spectra (search software). Nevertheless, humans still must be able to interact with spectra to remain in control of the identification process and investigate alternative hypotheses to those returned by automatic processing.

A typical proteomics dataset consists of two types of data: (i) mass spectra with associated data and (ii) peptides matched to the spectra by the search software. Both of these can come in different file formats depending on the manufacturer of the instrument or the developers of the search software, respectively. The multitude of file formats lead to an initiative for creating standardized formats for proteomics/mass spectrometry data by the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI). The HUPO-PSI standard format for encoding raw spectrometer output is mzML (8), with tools such as Proteowizard's MSconvert (9) being available to convert virtually every mass spectrometry raw data format to mzML. The existing HUPO-PSI standard format for reporting identifications mzIdentML (10) has recently been updated to version 1.2.0 (11), adding support for cross-linking data. A variety of tools have been developed to convert legacy formats to mzIdentML (9,12,13).

We strongly encourage the shift toward the use of community wide consistent standard formats. Therefore xiSPEC is fully compliant with the newest PSI standard formats mzML and mzIdentML. To provide backward compatibility, we additionally support the still widely used Mascot Generic Format (MGF) (14) for peak list data and identifications in a comma-separated format. To the best of our knowledge, the only existing PSI compliant tool for viewing and analyzing spectra is PRIDE Inspector (15), which has the downside of requiring download prior to use. It does not currently support cross-link data, is not designed for hypothesis testing by modifying the peptide-spectrum match data and does not provide scalable vector graphic (SVG) output. Proteomics data including cross-links can be visualized and shared through the browser-based MS-Viewer (16). However, MS-Viewer lacks support for the PSI standard identifications format (mzIdentML) and the amenities of modern web development. Lorikeet (https://github.

*To whom correspondence should be addressed. Tel: +49 30 314 72374; Email: juri.rappsilber@tu-berlin.de

com/UWPR/Lorikeet) is used for spectrum visualization in proXL (17), a web-based platform for the analysis of cross-linking data. Lorikeet is equivalent in functionality to MS-Viewer, but with the benefit of being open source.

xiSPEC version 1.0 (website and sourceforge release 2012) is a stand alone browser-based spectrum viewer that allows interrogating single spectra and their interpretation in interconnected views with SVG output for figure making. It has replaced the original Mascot spectrum viewer since Mascot v. 2.5. We present here xiSPEC version 2.0, an interactive tool for visualizing and analyzing mass spectrometry data in the browser. It supports data from standard proteomics and cross-linking experiments. The interactive design of xiSPEC follows the principle of multiple coordinated views (18). The user has all information available on a single web page and all views of the data are interconnected. We hope that xiSPEC's ease of use will positively impact on the proteomics community by enhancing data interrogation and sharing.

## IMPLEMENTATION

Essential to all the previously mentioned spectrum visualization tools is the ability to associate fragments of peptide sequences with peaks in the spectra. There are three ways the annotation of peaks with corresponding peptide fragments could occur. First, the annotated fragments could be recorded in the mzIdentML file, the specification allows for this. This has the benefit of allowing the spectrum viewer to show exactly those annotations that the search software used to derive the identification. Nevertheless, most search software do not include this information as it causes datasets to grow substantially. MS-Viewer and Lorikeet use an alternative approach by incorporating the annotation process into the spectrum viewer. At last, spectrum visualization and annotation can be separated into separate software components or services. This provides uniform annotation, separates concerns, eases the maintenance of both the annotation and visualization tools and allows the use of both independent of each other. An example of a stand-alone annotator is PRIDE-asap (19). PRIDE-asap does not support cross-linked peptides. Therefore, we use xiAnnnotator (https://github.com/Rappsilber-Laboratory/xiAnnotator).

xiSPEC itself consists of two major components: the data handling back-end and the interactive data visualization front-end. The backend data parser is written in python using the pymzML (20) (for mzML input) and pyteomics (21) (for mzIdentML input) packages and is also available as an open source project through GitHub (https://github.com/Rappsilber-Laboratory/xiSPEC_ms_parser). For fast data access into MGF files, xiSPEC uses an indexed based file reader we derived from pymzML. The parsed data gets written into a SQLite database. SQLite provides the benefit of having a single separate file for each dataset. This enables easy storage, deletion or compression of data. It is also cross-platform stable. On an annotation request, the data are read out from the database and converted into JSON. The annotation of spectra is done on-demand via API communication in JSON from the front-end to the Java application xiAnnotator running on a separate server. The front-end is written in JavaScript. The spectra data-visualization is based on D3 (22) to create SVGs. Event handling and synchronization between different views is achieved using the jQuery and Backbone JavaScript libraries. The interactive results tables are generated employing the DataTables jQuery library with server-side PHP processing for ordering, filtering and searching. This prevents processing of large datasets leading to prolonged load times and browser crashing.

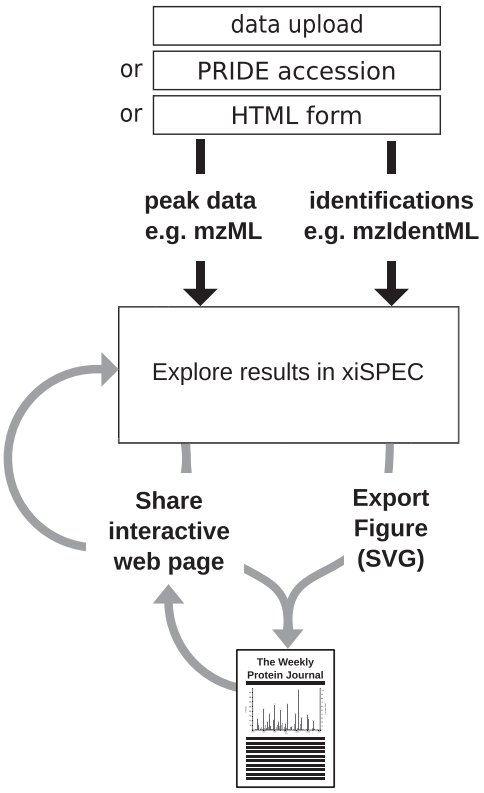


**Figure 1.** Overview of xiSPEC workflow. The input for xiSPEC are peak list data and peptide identifications. The user can either upload files directly to the xiSPEC server or select them from the PRIDE repository by providing the PXD accession number. For single spectra analysis data can be provided via HTML form input. Users can save datasets (publicly or password protected) and share them using a unique URL. Results can be exported as SVG for use in publications and presentations.

## FUNCTIONALITY

### Data input

Users can provide data either by direct data upload of an identifications and peak list file(s) pair or by providing a PXD accession number to the PRIDE repository (23) and subsequent selection of the files from the list of project files (Figure 1). In the latter case, xiSPEC uses the PRIDE RESTful API (24) to retrieve the project files. After user selection the files are directly downloaded from the PRIDE FTP server to the xiSPEC backend server where they are processed. This relieves the user from downloading and then re-uploading the files, thus making data deposited in PRIDE accessible easier and independent of end-user in-

ternet connection speed. xiSPEC supports the PSI standard proteomics identifications file format mzidentML and additionally a simple csv file format for non-standard data. This format is described at http://spectrumviewer.org/help.php#csv (Supplementary Table S1). Supported peak list data formats are the PSI standard mzML and also MGF. xiSPEC supports compression archives in .gz and .zip formats. For single spectra analysis users can input data directly via an HTML form with interactive peptide preview. All three options are available through the Upload page of the website.

## Features

Opening a dataset in xiSPEC presents the user all available information on a single web page with inter-connected views (Figure 2). Sub elements of the website layout can be re-sized or hidden when they are not needed to accommodate different use cases and individual users' preferences.

Datasets can be saved for later access or for sharing with collaborators by using a unique URL. To save a dataset the user has to input a name for his dataset and chose whether it will be publicly available or private. If the user chooses private the user needs to choose a password that will be required to access his dataset. In this way the dataset can still be shared with collaborators or reviewers before making it publicly available.

Identification results are presented to the user in an interactive data table. Columns can be hidden to declutter the view of unnecessary information. Results are paginated to provide the user with the ability to view the results and plots at the same time. The order of results can be changed simply by clicking on the column name. If more than one score is present the user can select the score used for ordering via a drop down menu. Results can be filtered by string search. Additionally predefined filters are provided to toggle displaying decoy identifications, identifications not passing the threshold defined by the search software and to hide linear identifications (useful for cross-link datasets). If the identifications input file contains alternative explanations they can be easily accessed by switching to the 'alternative explanations' tab. This allows for quick comparison and manual reviewing of potential misidentifications. The protein column is automatically converted to a link to the UniProt (25) subpage for the corresponding protein if the accession number is present in the input data.

The underlying data of the selected identification is presented to the user in multiple interactively connected views. The two main ones being the annotated mass spectrum (Figure 2A) with a peptide sequence fragmentation key (Figure 2B). Matched fragment peaks (and their isotope cluster peaks) are visualized through color and fragment name labels. For cross-linking data, two different colors are used to differentiate the two peptides. Neutral-loss fragments are displayed in a lighter color. Additionally, spectra quality control (QC) plots are generated to allow for manual identification quality assessment. They show the error of matched fragments plotted over intensity (Figure 2C) and over acquired $m/z$ range (Figure 2D). All of these views are interactive SVG elements. Hovering over a data point in any view displays a tooltip with detailed information. Highlighting or selecting data points in any view is represented in all other views so that the information available in the different views can be leveraged together.

Detailed instructions to xiSPEC's features can be found on the help pages (http://spectrumviewer.org/help.php#features). They are described as text instructions while at the same time being displayed as GIFs to visually guide the user. xiSPEC includes the following features:

(i) Zooming into spectra and moving around the currently displayed section. The current zoom level ($m/z$ range) can be locked, i.e. temporarily disabling the zoom and move functionality in the current spectrum. The selected $m/z$ range stays in place when switching spectra to enable easy cross-spectra comparison of specific $m/z$ regions.
(ii) Changing appearance styles of the output by changing color schemes and highlight color (Figure 3A).
(iii) Toggle display of neutral loss fragment labels, to declutter the annotated spectrum.
(iv) Changing to absolute error values for the QC plots.
(v) Measuring distances between peaks in Thompson. The distance is converted to masses calculated for multiple charge states and possible amino acid matches are displayed (Figure 3C). This can be help analyzing unexplained peaks in the spectra.
(vi) Option to move labels for better visibility. This click and drag functionality automatically creates dashed lines to the corresponding peak which simplifies figure making (Figure 3C).
(vii) Adding new post-translational modifications (PSMs) or changing modification masses. This can be done through the data settings view, by typing non-uppercase characters into the peptide sequence input (Figure 3A). Modification masses can be changed in the modification table. Inserting a modification that is not present in the input data will result in adding a row to the table.
(viii) Changing modification position. Modification positions can be moved from one residue to another by first clicking on the modification in the peptide sequence fragmentation key view and then clicking on the destination residue (Figure 3B). Alternatively, the modification can be moved in the input sequence (Figure 3A).
(ix) Changing cross-linker positions can be done by simply clicking on the cross-linker line in the peptide sequence fragmentation key view and selecting the new destination residue (Figure 3D). Another way is to move the cross-link symbol (#) in the input sequence (Figure 3A). Specifications on the data input syntax can be found in the help pages.
(x) Modifying precursor data, namely the precursor charge state and the peptide sequence (amino acids and modifications) (Figure 3A).
(xi) Changing fragment ion types considered. Ion types currently supported are unfragmented precursor (peptide) ion, b, c, y, z ions (Figure 3A).
(xii) Changing the permitted error tolerance for matching fragment peaks (Figure 3A).

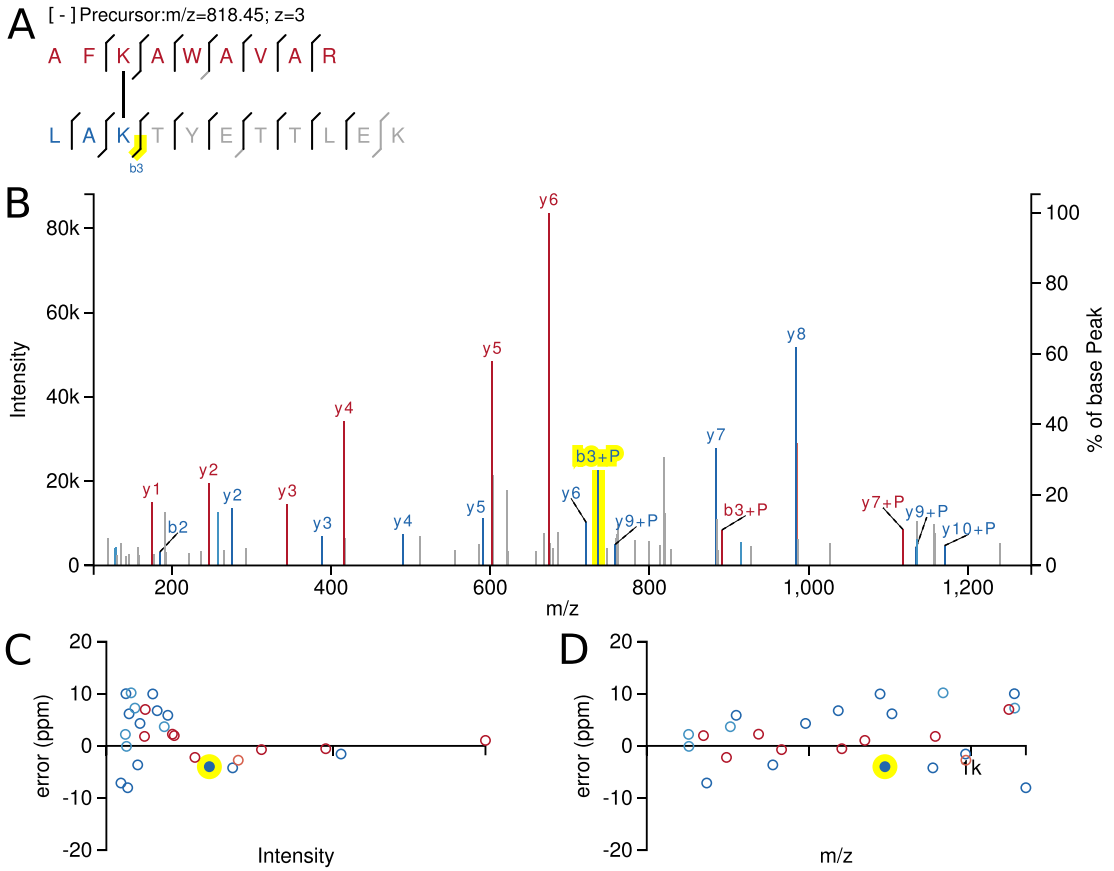


**Figure 2.** SVG output of xiSPEC's views for a cross-linked peptide example. (**A**) Peptide sequence with fragmentation key. Amino acid residues in one-letter code. Lines show matched peptide fragments. Grayed-out residues are not included in currently selected fragment. (**B**) Annotated mass spectrum. Matched peaks are colored and labeled (though labels of neutral-loss fragments are hidden in this example). (**C**) and (**D**) show spectra QC plots. Each point represents a matched peak of the mass spectrum. (C) Fragment match error over peak intensity. (D) Match error over $m/z$. Coloring (red and blue) is used to differentiate between the two peptides. Neutral-loss fragments are displayed in a lighter color. Yellow highlight is the currently selected fragment (interconnected between all views).

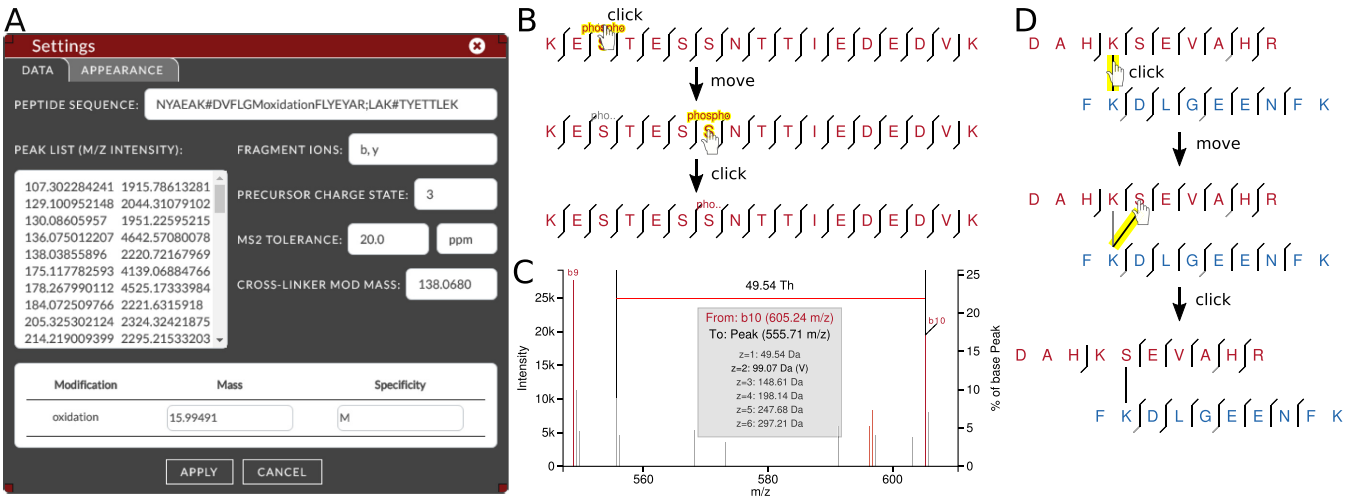


**Figure 3.** xiSPEC feature examples. (**A**) Settings view. Peptide input data can be modified in the displayed tab. Appearance customization can be done through the 'appearance' tab. (**B**) Changing PSM modification positions. (**C**) Use of measuring tool in zoomed-in excerpt of spectrum. Measure distance between peaks with automatic calculation and amino-acid residue matching for multiple charge states. (**D**) Changing cross-linker position.

(xiii) Reverting back to original annotation after modifying input data at the click of a button (background color changes to visualize changed data).

### Data output

xiSPEC offers the user the option to download single plots in publication quality as easily modifiable vector graphics (Figure 2). Additionally, xiSPEC allows sharing of visualized datasets at the ease of just sharing an unique URL, open or password protected. Anyone using a modern web browser can access it without the need to install third party software.

## USE CASES

We imagine xiSPEC bringing a positive impact to individuals from a variety of different backgrounds working with mass spectrometry proteomics data. Users can be divided into two main groups: (i) providers of data and (ii) users of data.

Providers of data are for example staff of mass spectrometry core facilities. They are not necessarily interested in interpretation details, but have the obligation of communication, i.e. sharing the data with their users. Authors of publications that include mass spectrometry data also need a way to make their annotated data available to their reviewers and readers. In fact many proteomics field guidelines include making annotated mass spectra of published results available (26–28), which can be achieved using xiSPEC by either sharing the datasets unique URL or downloading plots for use in publication. When still looking at data scientists may benefit from the possibility of interactive sharing with collaborators, other scientists or the community. Another example are teachers and lecturers who want to give their students access to view and work with mass spectrometry data. Removing the extra step of having to download additional software lowers the entry barrier significantly.

Users of data, e.g. biologists who receive mass spectrometry data from core facilities often have no dedicated software installed. Installing and getting offline software to work constitutes an additional hurdle, often involving frustration from steep learning curves. We believe that xiSPEC with its ease of use, user-centered and browser-based approach can simplify both lives of biologists and core facility members. By providing intuitive and easy to use tools for testing of hypothesis (measuring tool, re-annotation with modified parameters and QC plots) we also see a benefit for mass spectrometrists diving deeper into their data when looking at non-standard results.

## CONCLUSION

The importance of data sharing is widely appreciated in proteomics and secured by initiatives like ProteomeXchange. However, accessing, sharing and analyzing individual spectra from proteomic datasets is very cumbersome. xiSPEC aims to fill this gap by placing the user at the center of the interface design offering multiple view synchronization and a multitude of other features for hypothesis testing and sharing. As an actively developed open-source tool, it is open to community feature requests and contribution. We will be working toward seamless integration with online repositories such as PRIDE, UniProt or INTACT for users to interrogate primary data through simple web browsing.

## DATA AVAILABILITY

xiSPEC is an open source collaborative initiative available in the GitHub repositories (front-end: https://github.com/Rappsilber-Laboratory/xiSPEC; back-end: https://github.com/Rappsilber-Laboratory/xiSPEC_ms_parser). xiAnnotator is an open source collaborative initiative available in the GitHub repository (https://github.com/Rappsilber-Laboratory/xiAnnotator). All of which are freely available under the Apache License v2.0.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Jimi-Carlo Bukowski-Wills for creating xiSPEC 1.0 (accessible through http://legacy.spectrumviewer.org/; source code available at http://sourceforge.net/projects/spectrumviewer/), Lutz Fischer for creating the Java application backend service for the annotation of spectra data (https://github.com/Rappsilber-Laboratory/xiAnnotator) and Martin Graham for his help with JavaScript libraries and input on software design.

## FUNDING

Einstein Foundation; Wellcome Trust Senior Research Fellowship [103139 to J.R.]; Wellcome Trust Multi-user Equipment Grant [108504]; Wellcome Centre for Cell Biology [203149]. Funding for open access charge: Wellcome Centre for Cell Biology [203149].
*Conflict of interest statement.* None declared.

## REFERENCES

1. Rappsilber,J. and Mann,M. (2002) What does it mean to identify a protein in proteomics? *Trends Biochem. Sci.*, **27**, 74–78.
2. Nesvizhskii,A.I. and Aebersold,R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics*, **4**, 1419–1440.
3. Rappsilber,J. (2011) The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.*, **173**, 530–540.
4. Walzthoeni,T., Leitner,A., Stengel,F. and Aebersold,R. (2013) Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.*, **23**, 252–260.
5. Belsom,A., Schneider,M., Fischer,L., Brock,O. and Rappsilber,J. (2016) Serum albumin domain structures in human blood serum by mass spectrometry and computational biology. *Mol. Cell. Proteomics*, **15**, 1105–1116.
6. Liu,F., Lössl,P., Scheltema,R., Viner,R. and Heck,A.J.R. (2017) Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification. *Nat. Commun.*, **8**, 15473.
7. Schweppe,D.K., Chavez,J.D., Lee,C.F., Caudal,A., Kruse,S.E., Stuppard,R., Marcinek,D.J., Shadel,G.S., Tian,R. and Bruce,J.E. (2017) Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 1732–1737.

8. Martens,L., Chambers,M., Sturm,M., Kessner,D., Levander,F., Shofstahl,J., Tang,W.H., Römpp,A., Neumann,S., Pizarro,A.D. *et al.* (2011) mzML–a community standard for mass spectrometry data. *Mol. Cell. Proteomics*, **10**, R110.000133.
9. Chambers,M.C., Maclean,B., Burke,R., Amodei,D., Ruderman,D.L., Neumann,S., Gatto,L., Fischer,B., Pratt,B., Egertson,J. *et al.* (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, **30**, 918–920.
10. Jones,A.R., Eisenacher,M., Mayer,G., Kohlbacher,O., Siepen,J., Hubbard,S.J., Selley,J.N., Searle,B.C., Shofstahl,J., Seymour,S.L. *et al.* (2012) The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol. Cell. Proteomics*, **11**, M111.014381.
11. Vizcaíno,J.A., Mayer,G., Perkins,S., Barsnes,H., Vaudel,M., Perez-Riverol,Y., Ternent,T., Uszkoreit,J., Eisenacher,M., Fischer,L. *et al.* (2017) The mzIdentML data standard version 1.2, supporting advances in proteome informatics. *Mol. Cell. Proteomics*, **16**, 1275–1285.
12. Ghali,F., Krishna,R., Lukasse,P., Martínez-Bartolomé,S., Reisinger,F., Hermjakob,H., Vizcaíno,J.A. and Jones,A.R. (2013) Tools (viewer, library and validator) that facilitate use of the peptide and protein identification standard format, termed mzIdentML. *Mol. Cell. Proteomics*, **12**, 3026–3035.
13. Mayer,G., Stephan,C., Meyer,H.E., Kohl,M., Marcus,K. and Eisenacher,M. (2015) ProCon—PROteomics CONversion tool. *J. Proteomics*, **129**, 56–62.
14. Perkins,D.N., Pappin,D.J., Creasy,D.M. and Cottrell,J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.
15. Wang,R., Fabregat,A., Ríos,D., Ovelleiro,D., Foster,J.M., Côté,R.G., Griss,J., Csordas,A., Perez-Riverol,Y., Reisinger,F. *et al.* (2012) PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat. Biotechnol.*, **30**, 135–137.
16. Baker,P.R. and Chalkley,R.J. (2014) MS-viewer: a web-based spectral viewer for proteomics results. *Mol. Cell. Proteomics*, **13**, 1392–1396.
17. Riffle,M., Jaschob,D., Zelter,A. and Davis,T.N. (2016) ProXL (Protein Cross-Linking Database): a platform for analysis, visualization, and sharing of protein cross-linking mass spectrometry data. *J. Proteome Res.*, **15**, 2863–2870.
18. North,C. and Shneiderman,B. (1997) A taxonomy of multiple window coordinations. *University of Maryland, Dept. of Computer Science Tech Report*.
19. Hulstaert,N., Reisinger,F., Rameseder,J., Barsnes,H., Vizcaíno,J.A. and Martens,L. (2013) Pride-asap: automatic fragment ion annotation of identified PRIDE spectra. *J. Proteomics*, **95**, 89–92.
20. Bald,T, Barth,J, Niehues,A., Specht,M., Hippler,M. and Fufezan,C. (2012) pymzML–Python module for high-throughput bioinformatics on mass spectrometry data. *Bioinformatics*, **28**, 1052–1053.
21. Goloborodko,A.A., Levitsky,L.I., Ivanov,M.V. and Gorshkov,M.V. (2013) Pyteomics–a Python framework for exploratory data analysis and rapid software prototyping in proteomics. *J. Am. Soc. Mass Spectrom.*, **24**, 301–304.
22. Bostock,M., Ogievetsky,V. and Heer,J. (2011) D3: data-driven documents. *IEEE Trans. Vis. Comput. Graph.*, **17**, 2301–2309.
23. Jones,P., Côté,R.G., Martens,L., Quinn,A.F., Taylor,C.F., Derache,W., Hermjakob,H. and Apweiler,R. (2006) PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res.*, **34**, D659–D663.
24. Reisinger,F., del-Toro,N., Ternent,T., Hermjakob,H. and Vizcaíno,J.A. (2015) Introducing the PRIDE Archive RESTful web services. *Nucleic Acids Res.*, **43**, W599–W604.
25. UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.*, **41**, D43–D47.
26. Wilkins,M.R., Appel,R.D., Van Eyk,J.E., Chung,M.C.M., Görg,A., Hecker,M., Huber,L.A., Langen,H., Link,A.J., Paik,Y.-K. *et al.* (2006) Guidelines for the next 10 years of proteomics. *Proteomics*, **6**, 4–8.
27. Bradshaw,R.A., Burlingame,A.L., Carr,S. and Aebersold,R. (2006) Reporting protein identification data: the next generation of guidelines. *Mol. Cell. Proteomics*, **5**, 787–788.
28. Deutsch,E.W., Overall,C.M., Van Eyk,J.E., Baker,M.S., Paik,Y.-K., Weintraub,S.T., Lane,L., Martens,L., Vandenbrouck,Y., Kusebauch,U. *et al.* (2016) Human proteome project mass spectrometry data interpretation guidelines 2.1. *J. Proteome Res.*, **15**, 3961–3970.