ProteomeXchange submissions via PRIDE

Juan Antonio Vizcaino [1], Attila Csordas [2]

- Proteins
- Beginner
- 1 hour

This course covers how to submit MS/MS data to ProteomeXchange via PRIDE. This course is aimed at people in the proteomics community (staff scientists, students, researchers, PIs, bioinformaticians) with data that they want to make available through public-domain databases of proteomics experiments.

Learning objectives:

- Have a basic knowledge of the resources PRIDE and PeptideAtlas (including the PASSEL component)
- Be able to submit mass spectrometry data to the ProteomeXchange consortium via PRIDE
- Be able to confidently use the ProteomeXchange submission tool
- Be aware of the complete and partial submission pipelines

What is ProteomeXchange?

ProteomeXchange (PX) provides a standard way for submitting mass spectrometry [3]-based proteomics [4] data to public-domain repositories. Once you have submitted your data to the PX entry point, it will be automatically disseminated to all other repositories in the consortium [1 [5]].

ProteomeXchange simplifies the submission and distribution of proteomics data.

It has a centralised infrastructure that allows:

- Data submission to a mass spectrometry-based proteomics data repository;
- Private access to data in the repository during the review process (for journal editors and reviewers);
- Data dissemination to the other members of the consortium once the corresponding manuscript is published;
- A straightforward way for all proteomics researchers to keep up to date with newly available datasets via the ProteomeXchange RSS feed [6].

Some scientific journals (including Proteomics [7], Molecular and Cellular Proteomics [8] and several journals from Nature Publishing Group and PLoS) recommend submission to proteomics repositories such as PRIDE.

Data resources in ProteomeXchange

In the ProteomeXchange (PX) consortium there are currently two kinds of proteomics [4] data resources:
Archival resources contain processed data [9] as published by the authors. The founding members of PX are PRIDE [10] and PeptideAtlas [11]. They were joined by MassIVE [12], jPOST [13] and iProX [14]. The PASSEL component of PeptideAtlas [11] is representative for selected reaction monitoring [15] (SRM) datasets (Figure 1).

Secondary data resources build on the primary data provided by submitters and are stored in archival resources. PeptideAtlas [11] and UniProt [16] are examples of secondary data resources that are members of the PX consortium (Figure 1).

After submission to an archival resource such as PRIDE, once they are made publicly available, files are propagated to the secondary data resources, from where they are made available to users via an RSS feed (Figure 1).

**Figure 1.** Basic workflow for MS/MS data in ProteomeXchange. *Raw data* [17] means mass spectrometer output files [18]. Figure taken from reference 1.

In addition to the databases, ProteomeCentral [19] is the resource that generates a unique identifier [20] for each ProteomeXchange dataset and, also, it constitutes a registry for all ProteomeXchange submissions (irrespective of the receiving repository). This queryable archive provides the users with an efficient way to identify datasets of interest. For instance, it is a way to monitor the re-use of particular datasets, and give an efficient way to monitor the volume and impact of the ProteomeXchange data exchange.

**What is PRIDE?**

The PRIDE [21] (Proteomics IDEntifications) database at the European Bioinformatics Institute (EBI) is a centralised, standards-compliant public database containing MS-based proteomics data. It was originally developed to provide the proteomics community with a public repository for peptide and protein identifications, together with the evidence supporting these identifications [2 [5], [22],[3].
Figure 2 The PRIDE Archive web interface as of December 2017.

What does PRIDE store?

PRIDE aims to represent the submitter’s original view of the data. PRIDE stores:

- **Processed identification results** [23]: peptide and protein identifications, including post-translational modifications (PTMs);
- **Mass spectrometer output files** [18]: both raw data [17] and peak list [24] files;
- Quantification values: peptide and protein expression values (if available);
- **Metadata** [25]: technical and biological metadata;
- Other files related to one submission: gel images, scripts, etc.

Why do we need PRIDE?

The PRIDE database is one of the main public repositories for proteomics data that have been generated by MS approaches. In addition, PRIDE is leading the ProteomeXchange consortium, accounting for >85% of all the submissions.
Various journals in the field strongly support, and some even mandate, deposition of MS-based proteomics data into a public proteomics repository such as PRIDE (or others from the ProteomeXchange consortium). This enables researchers to understand, and potentially reproduce, experiments described in a particular publication. It also allows you to reanalyse mass spectra data as protein sequence databases and search-engine toolkits improve.

In addition, PRIDE provides proteomics data to other data resources such as UniProt [26].

An extra resource: ProteomeCentral

ProteomeCentral [19] (Figure 3) is an extra component of the consortium pipeline. It generates a unique identifier [20] for each ProteomeXchange data submission and acts as a registry for all ProteomeXchange submissions. This queryable archive provides an efficient way to identify and monitor public datasets of interest (for example, you can monitor the re-use of particular datasets). ProteomeCentral provides an efficient way to monitor the volume of data submitted.

Figure 3. The ProteomeCentral website as of September 2017

ProteomeCentral allows you to:

- Search original PX submissions (PXD identifiers) in participating repositories by experimental and biological metadata [25];
- Search for reprocessed datasets (RPXD identifiers);
- Easily query the archive and identify datasets of interest;
- Have a centralised resource to disseminate published data from the archival repository (PRIDE [10], PASSEL, MassIVE, jPOST, iProX) to the community via an RSS message [27] linked to a PX XML file.

- The PX XML file contains essential metadata information about the datasets and how to retrieve files that are part
How to submit data to ProteomeXchange via PRIDE

In order to submit your MS/MS data to ProteomeXchange via PRIDE [10], you should use the ProteomeXchange submission tool [28].

There are two pipelines for submitting to PX depending on the supporting information available for each submission:

- Complete submission
- Partial submission

The PX submission tool can be used in both cases. For larger datasets, it is possible to use a command line alternative which is based in a faster file transfer system called Aspera [29] (see details here [30]).

Each MS/MS data submission needs to include the following mandatory data types:

- **Metadata** [25] - Data that add context to the submitted experiment, such as what species the sample came from, are required as part of your submission to PRIDE/PX because they enable users to search for your experiments on the basis of the metadata;
- Peptide/protein identifications (processed results);
- Mass spectrometer output files [18] (raw data [17]).

A glossary of proteomics data formats used by are provided in Appendix 1 [31].

Once you have submitted your data, they can be held privately, allowing reviewers and journal editors access if desired. Since PRIDE is a public data repository, datasets will be made public upon the publication of the associated manuscript referring the particular PX identifier [20].

Next we will explore both pipelines.

## Complete submissions

You can use the PX submission tool [32] to submit your datasets to the complete submissions pipeline. You will need to provide all the raw data [17], and all applicable related metadata [25] to support your submission. All processed identification result files will need to be converted to mzIdentML [33] (version 1.1) format and the corresponding peak list [24] files must be provided as well.

There are different tools to convert or export files to mzIdentML (Check the previous section). To allow PRIDE to keep your submission secure you will need a PRIDE login. To get a username and password you will need to register [34].
ProteomeXchange MS/MS workflows for Complete Submissions

Figure 4. Overview of the ProteomeXchange MS/MS workflows for complete submissions.

Once the submission process is complete, you will be issued with a ProteomeXchange identifier [20] (accession number) and a permanent DOI (Digital Object Identifier [36]) that uniquely identifies your submission.

The submission is also manually checked by one of PRIDE’s curators. We consider this to be essential to ensure quality control of complete submissions. Please check the ProteomeXchange website [37] for more information.

The PX submission tool creates the appropriate relationships between the different types of files included in your submission (the raw data and the results are mandatory; the others are optional). It will allow you to add extra metadata for your dataset before submitting all the files to PRIDE, thereby producing more contextual datasets (Figure 4).

The complete list of data formats supported by PRIDE can be found in Appendix 2 [38].

Partial submissions

Although the option for Partial Submissions is available, the recommended first option is a Complete Submission (see previous page) as it significantly improves the reusability of your dataset. Partial Submissions should be used in case search results cannot be converted/exported to tmzIdentML files.

For newer methods, like DIA (data-independent acquisition) or MS Imaging [39], Partial Submissions are the only options. This might be also the suitable option for niche techniques and methods, like top-down proteomics [4].

As a result, you will be issued with a PX identifier [20] but not with a DOI. In addition your dataset will not be fully
integrated in PRIDE, but will be searchable based on metadata [25], and the corresponding files will be made available on the PRIDE FTP [40] server to download.

The Partial Submission pipeline requires you to have the raw data [17] (mass spectrometer output files [18]), search engine output files [41] (output from search engine or analysis pipeline) and a PRIDE login, so you will need to register [42] (Figure 5). Other files types can be included optionally (quantification output files, peak list [24] files, scripts, gel images, etc). Your username and password will allow PRIDE to securely process your submission.

Figure 5. ProteomeXchange MS/MS data workflow for partial submissions.

The PX submission tool creates appropriate relationships between the different file types used in a Partial Submission, i.e. the raw data and search engine output files. For bigger submissions, a faster file transfer command-line alternative is available (see here [30]).

Metadata requirements for MS/MS submissions

Proteomics data is substantially enriched when it is accompanied by sufficient metadata. The repositories in PX are quite flexible about how much experimental metadata they will accept. However, there are some minimal requirements. Below we summarise what you must supply when submitting your data to PX. We recommend that you add extensive metadata in the interest of making your dataset as useful as possible in the future.

The following details are mandatory for any PX submission:

- Contact name and e-mail for the submission. The contact details of the data submitters need to be provided, allowing interested users to contact the original authors if desired;

- Lab Head or Principal Investigator;

- Project title of the PX dataset;

- Project description: it could be considered as the abstract information of the dataset (provided as free text);
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

- A summary for each of the Sample and Data Processing Protocols (provided as free text);
- Experiment type (chosen from a drop-down menu);
- Keywords: A list of keywords that describe the content and type of the experiment being submitted. Multiple entries should be comma separated;
- Sample annotation:
  - Species. At least one NEWT ontology Controlled Vocabulary (CV) term is mandatory per dataset;
  - Tissue. At least one 'BRENDA Tissue' Ontology (BTO) CV term is mandatory per dataset;
  - Instrument. At least one 'Mass Spectrometry' (MS) CV term is mandatory per dataset.

Quantification method (if applicable).

The following details are optional:

- Sample annotation:
  - Cell type. Use the “Cell Type” ontology (CL);
  - Disease. Use the “Human Disease” ontology (DOID).

- Dataset details:
  (a) Your dataset is part of a bigger project/effort (for instance the Human Proteome Project or ‘PRIME-XS’). It is a way to tag your dataset to enable grouping this way;
  (b) There is already a PubMed ID associated with it (the data has been already published);
  (c) Your dataset represents a reanalysis of an earlier public PX dataset;
  (d) There are other related “omics” datasets (for instance transcriptomics, metabolomics data present in other repositories) that can be associated.
Submission to ProteomeXchange via PRIDE using the PX submission tool

The following slideshow provides a step-by-step walk-through of the submission process using the ProteomeXchange submission tool. Use the numbers at the top right of the panel below to move through the steps.

ProteomeXchange submission tool

The PX submission tool is a stand-alone desktop application that handles both complete and partial MS/MS proteomics submissions to ProteomeXchange (Figure 6). This tool will guide you through the whole submission process.

To run the tool, choose the [download](https://www.ebi.ac.uk/training/online) option directly on [PRIDE Submission Guidelines](https://www.ebi.ac.uk/training/online) page.

The main tool to submit data is the PX Submission Tool than can be downloaded by clicking on the button below.

![Download PX Submission Tool](download.png)

Create a PRIDE account

To start your submission, you will be prompted to enter details for a PRIDE user account. If you haven't already done so, you will need to create a new [PRIDE account](https://www.ebi.ac.uk/training/online) (Figure 7):
Creating a PRIDE [10] account is free and only takes a few minutes to register. You should receive an automated confirmation email more-or-less straight away and be able to use your new account. Please contact us by pride-support [at] ebi.ac.uk (email) if your login information is not valid after 24 hours following registration.

**Select your workflow**

When you launch the ProteomeXchange submission tool, you will be prompted to choose between complete submission and partial submission (Figure 8).
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Choose submission option below

**Complete Submission** *(recommended)*

Use this option if you can provide your identification results in either mzIdentML or mzTab, also legacy PRIDE XML format. It will then be possible to fully integrate the results in PRIDE and visualise them (e.g. as required by MCP).

In addition to a PXD identifier, a permanent Digital Object Identifier (DOI) will be provided to uniquely identify the dataset.

**Partial Submission**

You should only choose this option if your search results cannot be converted to mzIdentML, mzTab or legacy PRIDE XML. Identifications will not be integrated in PRIDE. However, files will be made available to download (and may be visualised with other external tools).

A PXD identifier will be provided to uniquely identify the dataset, but not a DOI.

---

Figure 8 The ProteomeXchange submission tool's welcome screen.

You can choose between two main submission types depending on the availability of mzIdentML [33] or PRIDE XML [47] files as "Result" files for complete submissions. The recommended submission subtype is a complete submission, but alternatively partial submissions are accepted as well. For details about this, please see the PRIDE Help Submission section [48].

**Prerequisites**

You’ll be asked to prepare material and details about your study and dataset to be uploaded (Figure 9).
Log-in with PRIDE account

After choosing your submission type, you will be prompted to log in with your PRIDE [10] account (Figure 10). Please enter your registered email address and password to continue.
Figure 10 PX submission tool, PRIDE account log-in

**ORCID id and country information**

Next, you can optionally provide your ORCID identifier [49] and also supply the country information for statistical purposes (Figure 11).
Figure 11 Optional: providing ORCID identifier and source of origin

**Dataset details**

Here mandatory details about the dataset are required, such as a title for your submission, experimental and data processing protocol, and experiment type (Figure 12).
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Figure 12 PX submission tool, dataset details.

Please note that the description and protocol fields need at least 50 characters to be entered in order to continue.

Add your files

On the 'Add files' screen, select all the files you wish to submit and choose which file type they belong to (Figure 13). The tool will automatically guess what the type is for each file, which you can manually change. You can also remove any files you have accidentally included.

Depending on what type of submission you have made, different files will be mandatory or optional. An example of a complete submission with an mzIdentML [33] result file, along with the related peak list [24] file, and raw data [17] file is shown in the figure below.

Again, for details about the submission types, please see the PRIDE Help submission section [48].
Figure 13 The PX submission tool's ‘Add files’ screen. Here, files that are going to be part of the submission are selected and tagged.

For a full list of all the different file types, see Appendix 2 [50].

Create relationships between files

Raw files need to be mapped to the corresponding result (complete submission) or search engine output files [41] (partial submission). The tool will try to do this automatically depending on the file naming structure, and this can be edited manually using the “+ Relation” button for each result or search output file (Figure 14).
Experiment details

Further experimental information is required, either individually for each result file (complete submissions), or overall for the dataset (partial submissions).

In Figure 15, below, an example for a complete submission is shown. In order to reach this screen, the "Annotate" button for the result file was clicked (similarly to the previous step for mapping relationships). A partial submission will go straight to this screen instead, after performing the mapping for relationships.

To enhance your dataset you should add as much available metadata [25] as possible.
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Figure 15 PX submission tool, experiment details. Example of a complete submission.

Mandatory information needed for a complete submission: species, tissue, and instrument.

A partial submission also requires modifications to be entered \[4^{[22]}\].

Ontology \[51\] terms have to be selected from each of the drop-down menus which have the most common options listed. If a required term is not present: select the bottom "Other..." option from the drop-down which will then allow you to search across various ontologies to select a desired term.

Add Lab Head information
Lab Head (Principal Investigator) details are required to continue (Figure 16).
Figure 16 PX submission tool, lab head details.

The mandatory fields are: name, email, and affiliation.

Additional details

Additional details is an unassuming title but actually annotating your data at this step might give a boost to its reusability. For instance you can indicate what kind of bigger projects your data is a part of or link to associated other omics datasets that were part of your study too using the same samples. (Figure 17).
Final checks and data submission

Once all the relevant information has been added you will have an opportunity to do a final check before you start the submission (Figure 18).

After you click "Submit", your files will be uploaded to PRIDE [10]. This will take some time depending on how large your dataset is, and how fast your internet connection is. When all files have finished transferring to PRIDE, you will receive an e-mail including a submission reference. This is not an accession [35] number to be quoted in manuscripts. It should be used in all communications between you (the submitter) and the PRIDE team. An automated confirmation e-mail will also be sent to your registered e-mail address.

Your submission will then be processed by expert biocurators in the PRIDE team, who may need to contact you in order to improve the quality of the provided metadata [25]. Once everything is corrected (if applicable), your dataset will be loaded into the PRIDE database. After this happens you will receive another automated e-mail with a unique ProteomeXchange identifier [20] (PXD accession number) and the private reviewer account details. If you had made a complete submission, you will also receive a data DOI [52].
ProteomeXchange submissions via PRIDE

Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Figure 18 PX submission tool, submission summary.

In the top-right corner of this screen there is a button called "Export summary file". This will save all the meta-data for your dataset into a summary "submission.px" file. This is useful if making a "bulk" submission through the PX Submission tool, or a direct manual submission to PRIDE [10] via our private Aspera [29]/FTP [40] server. Such manual submissions are useful when dealing with large datasets that cannot be easily managed with the PX Submission tool alone. Full details about this can be found on the PRIDE Help [53] section.

There are a few options possible post-submission, e.g. making the dataset public. These are listed in our PRIDE Help FAQ [54].

Summary

- ProteomeXchange provides a standard way to submit MS-based proteomics data to public repositories.

- It is developed by the ProteomeXchange consortium, a collaboration among the groups that develop the main MS-based proteomics repositories.
• Submitting proteomics data to ProteomeXchange is a simple way of ensuring that your data is made available to the research community and disseminated to the major data proteomics data resources.

• Submission to a public repository is recommended, if not mandated, by many of the major proteomics journals as part of the publication of a paper.

• ProteomeXchange ensures that published (archival), raw and reprocessed (secondary) data are made available, disseminated and appropriately cross-linked.

• PRIDE is the initial submission point of MS/MS data.

• There are two pipelines for submission of MS/MS data: one for complete submissions (containing raw data and PRIDE XML or mzIdentML files) and one for partial submissions (containing raw data and search output files). The ProteomeXchange submission tool can be used for both routes.

• PASSEL is the initial submission point of SRM data.

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Learn more

Find out more

• Proteome Xchange concept [55] - For more background knowledge on this project.

• PRIDE/PX tutorial document [56] - Detailed information about how to perform submissions.

• PRIDE - Quick tour [57] - For a quick overview on PRIDE see the Quick tour.

Get help and support on ProteomeXchange

Where to get help and support
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

- More information about ProteomeXchange [58] and PRIDE [59] is available on their websites.
- PRIDE Help [53] details submitting data, generating data, and searching PRIDE. There is also an up-to-date FAQ [60].
- You can also contact the PRIDE [61] team by pride-support [at] ebi.ac.uk (email).

References

1. Vizcaíno J.A. et al. (2014) ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination. Nature Biotechnol 32(3): 223-226.

2. Vizcaíno J.A. et al. (2013) The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. [62] Nucleic Acids Res. 41. D1063-1069.

3. Vizcaíno, J. A. et al. (2009) A guide to the Proteomics Identifications database proteomics repository. [63] Proteomics 9. 4276-4283.

4. Phan, I. Q. et al. (2003) NEWT, a new taxonomy portal. [64] Nucleic Acids Res. 31. 3822-3823.

5. Montecchi-Palazzi, L. et a. (2008) The PSI-MOD community standard for representation of protein modification data. [65] Nature Biotechnology 26. 864-866.

Other recommended reading

- Deutsch, E.W., et al. (2017). The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition [66]. Nucleic Acids Res. 45(D1):D1100-D1106.
- Vizcaíno, J.A., et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. [67] 44(D1):D447-56.

Appendix 1: Proteomics data formats

Proteomics data is available in a variety of formats, the ones used by Pride and ProteomeXchange are defined here:

| File name                        | File content                                                                 |
|----------------------------------|-----------------------------------------------------------------------------|
| **Mass spectrometry output files ('Raw' data)** | This is the data and metadata generated by the original profile mode scans or may be applied, such as centroiding, applied. They may be available as mass spectrometry spectra in a standardised format (see below). It is important that all the scans generated |
### mzXML - developed at the Institute of Systems Biology (ISB)
mzXML is a successor to mzData, originally developed by the HUPO Proteomics Standards Initiative (PSI).

These data formats can be used to represent processed peak lists, as well as raw data. In addition to the mass spectra, they contain detailed metadata that gives context to the information.

### Processed peak lists

Heavily processed form of mass spectrometry data, usually derived from raw data files via various (semi-) automatic steps, e.g.: centroiding, deisotoping and charge deconvolution. These files are formatted in plain text, with typical formats like dta, pkl, ms2 or mgf.

### Protein/peptide identifications

Proteomics mass spectra can be matched to peptides or proteins, resulting in identifications for those spectra. Typically a spectrum is considered to have been identified if the score attributed to a peptide match qualifies against an a priori or a posteriori defined threshold. In the initial identification will consist of a peptide list, a list of proteins from the identified peptides is a discernible process with its own input and output files, or it can be implicit in the overall identification software.

### Search engine output files

These files contain the data and metadata generated by the software (called search engines) used for performing the identification and quantification of peptides and proteins. Each search engine has its own specific output file format. The outputs are typically formatted in either plain text or XML.

- **mzIdentML** [69] - provides a common format for the export of identification results from any search engine.

To allow a full representation of the processed results in the PRIDE database and the PX tool, the search engine output files need to be converted to PRIDE XML possible, but the necessary tools are not available.

### Protein/peptide quantification

Protein/peptide expression values can also be obtained from an MS-based proteomics experiment and then this data and metadata is used for performing the quantification analysis of peptides and proteins.

### Quantification software output files

The data and metadata generated by the software used for performing the quantification analysis of peptides and proteins.

- **mzQuantML** [72] - provides a common format for the export of quantification results from any search engine.

- **mzTab** [73] - represents both identification and basic quantification results.

### Metadata

A term used to describe data that provides additional information about a particular data set. This information can include how, when and where the data set was generated and what standards were used. The addition of metadata such as peptide and protein identifications and quantification of their expression values gives meaning to the raw data.
Appendix 2: Formats supported by PRIDE for PX MS/MS submissions

a) as raw data

Formats supported:

- mzML, mzXML, mzData. These files must not be heavily processed to be considered ‘raw’
- Thermo .RAW, ABSCIEX .wiff, .wiff.scan, Agilent .d/, Waters .raw/
- imzML, Shimadzu .run/, Bruker .baf, .yep

All peak lists formats (mgf, dta, ms2, pkl) can be supported but they will not be considered raw data. They will be considered as ‘peak list processed files’ or simply ‘peak’.

b) as processed identification 'results'

mzIdentML (version 1.1): There are a number of tools that can export mzIdentML 1.1. Formats supported this way:

- Mascot [74].dat (direct export functionality available from Mascot 2.4) from Matrix Science [75], see screenshots [76]
- Scaffold [77] (direct export functionality available) by Proteome Software [78]. From version 4.0, see tutorial [79]
- MS-GF+ [80] from UCSD Computational Mass Spectrometry Group [81]. The default output option is mzid 1.1 (direct export functionality available)
- ProteinPilot [82] group files from AB SCIEX. See screenshots [83]
- Sequest .out files (using the ProCon tool [84])
- Proteome Discoverer [85] .msf files (using the ProCon tool [84])
- ProteinScape [86] 2.1 (Bruker) database content (using the ProCon tool [84])
- Phenyx (direct export functionality available)
- Trans-Proteomic Pipeline [87] (pep.xml files). The idConvert tool from can be downloaded from ProteoWizard, or is bundled with the TPP directly starting with version 4.6.3
- OpenMS [88] output
- MIAPE MSI Extractor output (ProteoRed [89], Madrid)
- PAnalyzer output: Tool to perform protein inference analysis [90]
- Output files from Myrimatch, Pepitome (spectral library search), TagRecon and IDPicker
- Tandem XML (using mzidLibrary [91])
- OMSSA .csv (using mzidLibrary [91])
- All accompanying peak lists formats
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

3) as search engine output files

Only those data formats that cannot be converted/exported to mzIdentML can be submitted as partial submissions.

4) quantification output files

Any format is supported.

Contributors

Juan Antonio Vizcaino [1]
EMBL-EBI
PRIDE Group Coordinator
Juan is the Group Coordinator of EMBL-EBI's PRIDE resource for mass spectrometry proteomics data, and he has authored more than 65 papers. He joined the PRIDE team as a postdoc in 2006 and has been involved in numerous projects that span research, development of tools, and training and support for PRIDE users. Since 2011, he has been managing the ProteomeXchange consortium of proteomics resources.
He has a background in the life sciences, with a PhD in Molecular Biology from the University of Salamanca, Spain, a Masters in Microbiology and undergraduate degrees in Pharmacy and in Biochemistry.
**Tobias Ternent** [92]

EMBL-EBI  
Bioinformatician - Hermjakob group

Tobias is a curator for the PRIDE team (2013), and previously a curator for the ArrayExpress team (2011). Before joining the EMBL-EBI, he studied a Bachelor's in Computer Science from the University of Manchester (2005), and later a Master's in BioHealth Informatics in the same department (2010) which involved collaboration with the KNH-Centre for Biomedical and Forensic Egyptology. Along the way, his working experience has included a few software companies, and in particular managing clinical trial data for AstraZeneca for translational informatics.

**Attila Csordas** [2]

EMBL-EBI  
Bioinformatician - Vizcaino Team: Proteomics

**Source URL:** https://www.ebi.ac.uk/training/online/course/proteomexchange-submissions-pride

**Links**
[1] https://www.ebi.ac.uk/training/online/trainers/juan  
[2] https://www.ebi.ac.uk/training/online/trainers/acsordas  
[3] https://www.ebi.ac.uk/training/online/glossary/mass-spectrometry  
[4] https://www.ebi.ac.uk/training/online/glossary/proteomics  
[5] https://www.ebi.ac.uk/training/online/course/proteomexchange-submissions-pride/references  
[6] https://groups.google.com/group/proteomexchange/feed/rss_v2_0_msgs.xml  
[7] http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291615-9861/  
[8] http://www.mcponline.org/  
[9] https://www.ebi.ac.uk/training/online/glossary/processed-data  
[10] https://www.ebi.ac.uk/training/online/glossary/pride  
[11] http://www.peptideatlas.org/  
[12] https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp  
[13] https://jpost.org/  
[14] http://www.iprox.org/  
[15] https://www.ebi.ac.uk/training/online/glossary/selected-reaction-monitoring  
[16] http://www.uniprot.org/  
[17] https://www.ebi.ac.uk/training/online/glossary/raw-data  
[18] https://www.ebi.ac.uk/training/online/glossary/mass-spectrometer-output-files  
[19] http://proteomecentral.proteomexchange.org/cgi/GetDataset  
[20] https://www.ebi.ac.uk/training/online/glossary/identifier
ProteomeXchange submissions via PRIDE
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

[76] https://www.ebi.ac.uk/sites/ebi.ac.uk/files/groups/pride/Mascot_export.pdf
[77] http://www.proteomesoftware.com/products/scaffold/
[78] http://www.proteomesoftware.com/
[79] https://www.ebi.ac.uk/sites/ebi.ac.uk/files/groups/pride/Scaffold_Exporting%20mzIdentML.pdf
[80] http://proteomics.ucsd.edu/Software/MSGFPlus/
[81] http://proteomics.ucsd.edu/index.html
[82] http://www.absciex.com/products/software/proteinpilot-software
[83] https://www.ebi.ac.uk/sites/ebi.ac.uk/files/groups/pride/ProteinPilot_mzIdentML_export.pdf
[84] http://www.medizinisches-proteom-center.de/index.php/de/software-top/137-proteomics-conversion-tool-procon
[85] https://www.thermofisher.com/order/catalog/product/OPTON-30795
[86] https://www.bruker.com/products/mass-spectrometry-and-separations/ms-software/proteinscape/overview.html
[87] http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP
[88] https://www.openms.de/
[89] http://proteored.org/miape/
[90] https://code.google.com/p/ehu-bio/wiki/PAnalyzer
[91] http://code.google.com/p/mzidentml-lib/
[92] https://www.ebi.ac.uk/training/online/trainers/tobias