

PRIDE: The Proteomics Identifications database

www.ebi.ac.uk/pride

Partners

PRIDE has been developed through a collaboration between EMBL-EBI, Ghent University in Belgium and Manchester University in the UK.

Need help?

PRIDE is hosted by the EMBL-European Bioinformatics Institute.

Support: www.ebi.ac.uk/pride/userManual.do

pride-support@ebi.ac.uk

Phone: +44 (0)1223 492610

Fax: +44 (0)1223 494468

Henning Hermjakob
EMBL-EBI

Wellcome Trust Genome Campus
Hinxton, Cambridge
CB10 1SD, UK

Support

PRIDE is supported by the Wellcome Trust, the European Molecular Biology Laboratory (EMBL) and by the European Commission under SLING, grant agreement number 226073 (Integrating Activity) within Research Infrastructures of the FP7 Capacities Specific Programme.

High-throughput proteomics has enabled the identification – and publication – of an ever-increasing number of proteins. The Proteomics Identifications Database (PRIDE) has become one of the main repositories of proteomics data derived from mass spectrometry. Various journals in the field are now strongly supporting and even mandating deposition of proteomics data in proteomics repositories in general, and PRIDE in particular. The new PRIDE Converter submission tool has made data submission easy and straightforward. Datasets are stored in PRIDE without modification or re-analysis.

What is PRIDE?

PRIDE offers a freely available web-based query interface, a BioMart interface, a user-friendly data upload facility, and a documented application programming interface for direct computational access. The complete PRIDE database, source code, data, and support tools can also be downloaded and installed locally.

PRIDE stores three types of information: MS and MS/MS mass spectra as peak lists, the derived peptide and protein identifications (IDs), and any associated metadata.

The database supports identifications from both LC-MS-based and gel-based techniques. Processed peak lists arising from mass spectrometry (MS^N) can be included. Post-translational modifications (both natural and artefactual) can be annotated on the identified peptide in a PRIDE dataset.

The PRIDE team implement the data standards and formats being developed by the Human Proteomics Organization's Proteomics Standards Initiative (HUPO-PSI; <http://psidev.info>). PRIDE currently supports data exchange using the mzData XML standard and PRIDE will eventually allow data transfer using the new mzML and mzIdentML formats. The controlled vocabularies being developed by PSI are also being leveraged by PRIDE, in conjunction with the Ontology Lookup Service (OLS; www.ebi.ac.uk/ols), to support diverse proteomics datasets in a manner that can be effectively queried by the user.

As a consequence of providing a data submission service in a diverse field, PRIDE contains protein identifications based on a large number of different protein sequence databases. Incorporating the use of the Protein Identifier Cross Reference Service (PICR; www.ebi.ac.uk/Tools/picr) into PRIDE ensures that all submitted protein identifications are regularly mapped to the latest accessions from a large number of different databases.

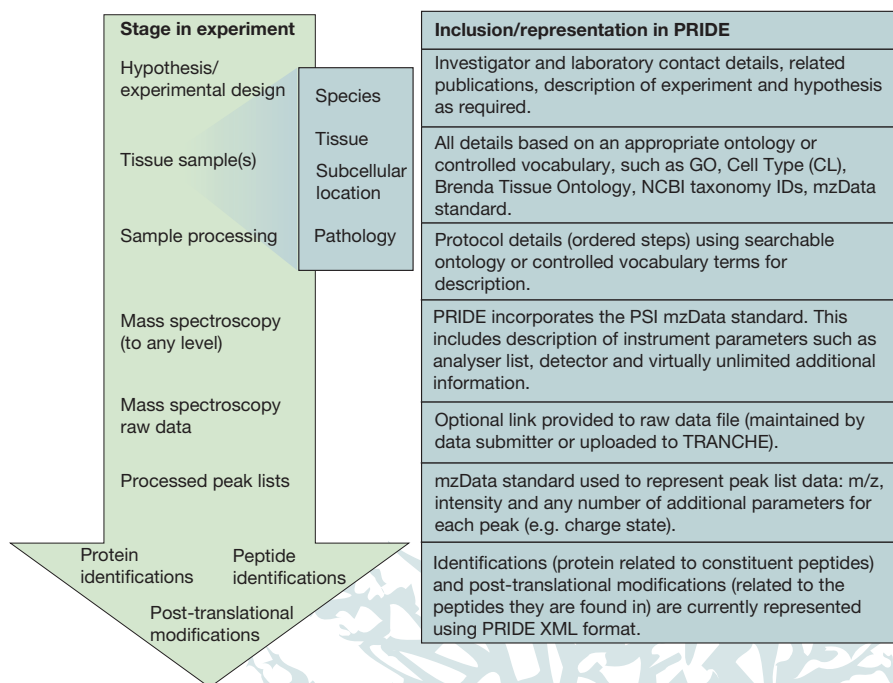
What can I do with PRIDE?

- Retrieve complete sets of protein or peptide identifications and observed protein modifications that have been submitted to PRIDE in support of a published proteomics study.
- If there are one or more proteins that you are interested in, you can find out: 1) whether the proteins have been detected by mass spectrometry before; 2) if so, how the experiments were carried out; 3) information about the peptide sequences identified and possible protein modifications (PTMs).
- Construct complex queries using the PRIDE BioMart interface, specifying



Published by the
EMBL-EBI – a part
of the European
Molecular Biology
Laboratory

multiple filters and the precise subset of data for retrieval. In the current BioMart Central Portal (www.biomart.org/biomart/martview/), it is possible to retrieve data from PRIDE individually, but also to integrate information and build complex queries across PRIDE and other resources, such as Ensembl, UniProt, InterPro and Reactome.



Summary of data types captured in PRIDE.

Submitting data

Submission to PRIDE is open to any proteomics laboratory generating protein and peptide identification data. We encourage submission of identifications in support of your publications describing your work and findings.

PRIDE supports private data upload, allowing submitters to secure their data at the level of an individual user or as part of a collaboration. This allows data sharing among laboratories or with journal editors and peer reviewers without the data being available to the general public. As a result, PRIDE is now the recommended submission point for proteomics data for several journals such as *Nature Biotechnology*, *Nature Methods* and *Proteomics*.

The simplest way to submit data is to use the PRIDE Converter. This tool builds PRIDE XML files (the required submission format) and can be downloaded from <http://code.google.com/p/pride-converter/>. PRIDE Converter is platform independent, written in Java and open source.

PRIDE Converter can accommodate a large variety of input proteomics data formats, is suitable for both small and large data submissions, and it is easy to use, with a 'wizard-like' graphical user interface. The user needs to provide appropriate metadata using controlled vocabulary terms that are retrieved through an online connection to the OLS.

Once the PRIDE XML files are created, you can submit them through the PRIDE webpage (only recommended for files smaller than 15 MB) or to the EBI FTP server. If you need an FTP account, please email pride-support@ebi.ac.uk ●

Retrieving data from PRIDE

PRIDE data can be retrieved as PRIDE XML files, to allow external analysis of PRIDE data. PRIDE XML files for complete PRIDE experiments can be obtained directly from the PRIDE FTP service (<ftp://ftp.ebi.ac.uk/pub/databases/pride>). You can browse the contents of PRIDE and the details of individual experiments through a series of easily-navigable web pages. You can create customised data tables in Microsoft Excel format, using the PRIDE BioMart and you can perform simple comparisons of experiments in PRIDE using the Venn diagram tool, which compares experiments based upon the protein identifiers that they comprise.

Further reading

Vizcaíno J.A. *et al.* (2010) The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.* 38: D736-742

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

Barsnes, H. *et al.* (2009) PRIDE Converter: Making proteomics data sharing easy. *Nat. Biotechnol.* 27, 598-599

Jones, P. *et al.* (2008) PRIDE: new developments and new datasets. *Nucleic Acids Res.* 36, D878-D883

Côté R.G. *et al.* (2008) The Ontology Lookup Service: more data and better tools for controlled vocabulary queries. *Nucleic Acids Res.* 36, W372-W376