



European Bioinformatics Institute

PRIDE: The PRoteomics IDentifications database

www.ebi.ac.uk/pride

Why do we need PRIDE?

The advent of high-throughput proteomics has enabled the identification of ever-increasing numbers of proteins. Correspondingly, the number of publications centred on these protein identifications has increased dramatically. However, the mechanism for publishing these identified proteins has lagged behind in technical terms. Long lists of identifications are either published directly with the article, resulting in a voluminous and rather tedious read, or are included on the publisher's website as supplementary information. In either case, these lists are typically provided as PDF documents with a custom-made layout, making it practically impossible for computer programmes to interpret them, let alone efficiently query them.

What is PRIDE?

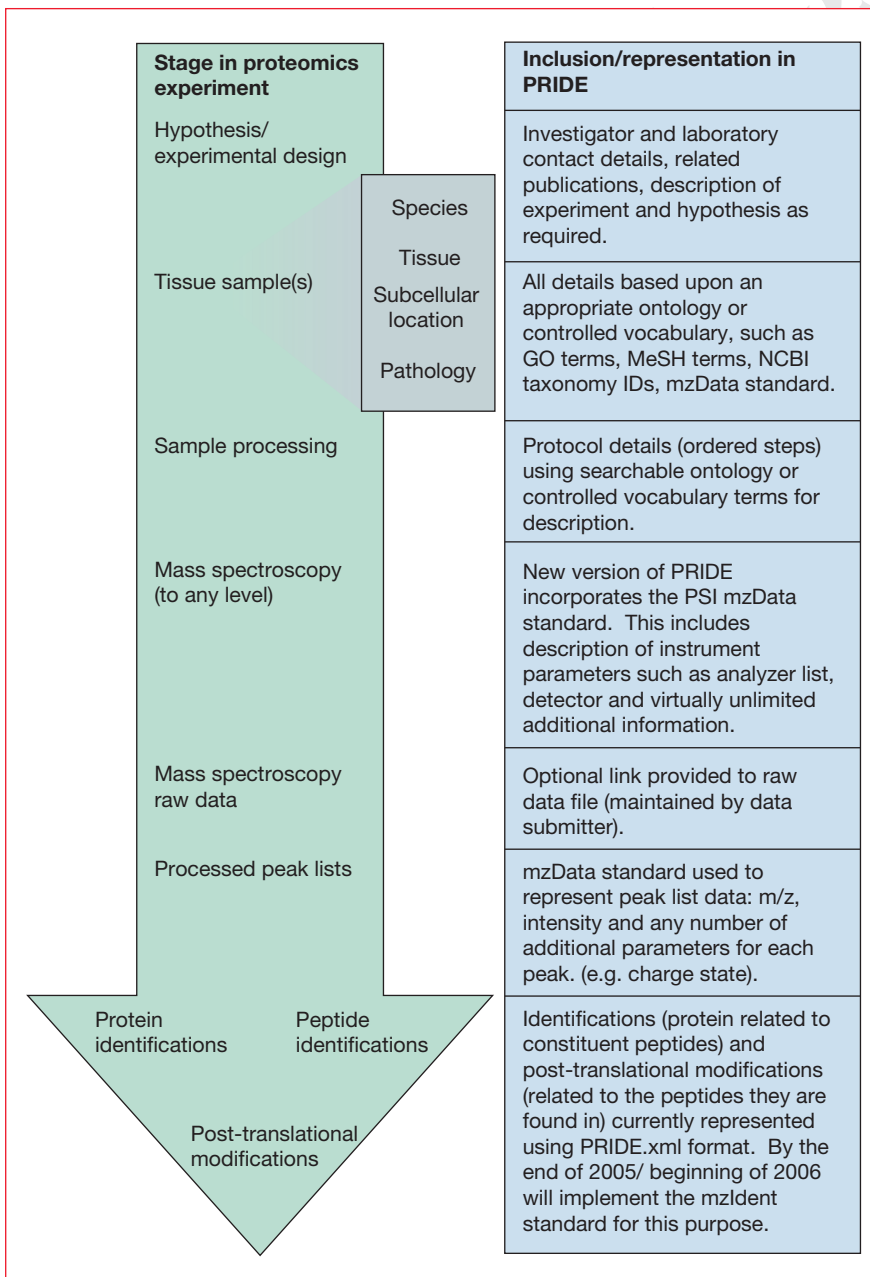
PRIDE, the open source 'PRoteomics IDentifications database', has been developed through a collaboration between EMBL-EBI and Ghent University to alleviate these

needs. PRIDE offers a web-based query 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 are freely available for web access or download and local installation.

The original motivation behind the development of PRIDE was to provide a common data exchange format and repository to support proteomics publications. However, the original purpose has grown with PRIDE, with the hope that PRIDE will provide a reference set of tissue-based identifications for use by the proteomics community.

The PRIDE data model has been designed with flexibility in mind. The current iteration supports identifications from both LC-MS-based and gel-based techniques. Processed peak lists arising from MS, MS-MS and higher MS levels are supported. Post-translational modifications (both natural and artefactual) can be included in the data set.

PRIDE is closely linked to the Human Proteomics Organization's Proteomics Standards Initiative (HUPO-PSI; psidev.sourceforge.net). It is intended that PRIDE will allow data transfer using the mzData and mzIdent XML schemata, which are being developed as part of PSI. The controlled vocabularies being developed by the PSI General Proteomics Standards workgroup will increase the power of PRIDE to support diverse proteomics data sets in a manner that can be effectively queried by the user.



What can I do with PRIDE?

PRIDE allows you to retrieve the complete set of protein identifications for a publication, along with the supporting peptide identifications and hyperlinks to further information. You can also query PRIDE to find all relevant data sets for a particular protein of interest or for a given tissue. The data you have retrieved can be formatted in two different ways: one is the familiar HTML format, which is easily readable and hyperlinked. The other format is XML, which can be readily parsed by simple scripts or computer programmes.

You can also submit your identifications to PRIDE. This obviates the need to publish long, impractical PDF tables as supplementary information; the accession number(s) for your data set(s) and a reference to the PRIDE database URL are all you need to include in your publications.

Submitting data to PRIDE

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 a collaboration. This allows data sharing among laboratories or with journal editors and peer reviewers without the data being available to the general public. Whenever you decide to make your data publicly available (e.g. after a publication has been accepted), it is simply a matter of clicking a button to share your findings with the community. It is, after all, the aim of PRIDE to create an open, public repository of protein identifications.

Submission of data to PRIDE can be accomplished by completing (free) registration with the PRIDE system and following the 'Upload Data' hyper-link on the left of the PRIDE web page.

The submission format is an XML document following the current PRIDE XML schema. This schema can be downloaded from the PRIDE homepage by following the 'How To Submit Data to PRIDE' hyper-link. Please contact the PRIDE team if you require any support with exporting your data to this format.

Retrieving data from PRIDE

The data in PRIDE can be retrieved as a brief summary (we are currently in the process of expanding the power of the statistical reporting functions of PRIDE), as a PRIDE.xml file or as a formatted HTML table. The database can be searched by following the 'Search PRIDE' hyper-link on the left of the PRIDE homepage.

Further reading

Martens, L. *et al.* (2005) **PRIDE: The PRoteomics IDentifications database** *Proteomics* (in press)

Orchard S. *et al.* (2005) **Further steps towards data standardisation: the Proteomic Standards Initiative.** *Proteomics* 5, 337-339

Support

PRIDE is supported through BBSRC 'ISPIDER' and HUPO Plasma Proteome Project funding as well as an EU Marie Curie fellowship.

Contacting PRIDE

URL: www.ebi.ac.uk/pride
e-mail: Philip Jones (pjones@ebi.ac.uk) or
Lennart Martens (lennart.martens@UGent.be)
Phone: +44 (0) 1223 492610
Fax: +44 (0) 1223 494468

Post: Philip Jones
EMBL-European Bioinformatics Institute,
Wellcome Trust Genome Campus,
Cambridge,
CB10 1SD, UK