# QUALITY CONTROL AND STANDARDIZATION EFFORTS AT THE PROTEOMICS IDENTIFICATIONS (PRIDE) DATABASE

*J.A. Vizcaíno[1], R. Coté[1], A. Csordas[1], A. Fabregat[1], J. Foster[1], D. Ovelleiro[1], F. Reisinger[1], D. Ríos[1], R. Wang[1], H. Hermjakob[1]*

[1] EMBL-EBI, Wellcome Trust Genome Campus, Cambridge, UK.

The PRoteomics IDEntifications database (PRIDE, http://www.ebi.ac.uk/pride) at the European Bioinformatics Institute (EBI) is one of the main repositories of mass spectrometry (MS) derived proteomics data. PRIDE stores three different kinds of information: MS and MS/MS mass spectra as peak lists, the derived peptide and protein identifications, and any associated metadata. A major aim for PRIDE is the support of the peer-review process in the proteomics community.

PRIDE is entirely dependent on data submissions, as detailed proteomics data cannot be curated from existing literature. Currently, PRIDE is working towards establishing Quality Control of the data at two levels:

A new stand-alone tool, PRIDE Inspector (http://code.google.com/p/pride-toolsuite/) allows the users to examine their data before the submission is done. Different views of the data are provided: metadata, spectrum, peptide and protein centric. Additionally, the major strength of the tool is the possibility to perform a first assessment on data quality, since a variety of charts based on the data are automatically generated. Highly important for journal reviewers and editors, it will facilitate the thorough review of submitted data in the prepublication stage.

A new resource called PRIDE-Q (for Q-rated) will contain only the highest-quality data from PRIDE. The relationship between the current PRIDE repository, and the planned PRIDE-Q resource is very similar to that between UniProtKB/SwissProt and UniProtKB/TrEMBL.

One of the most challenging tasks of PRIDE as a data repository is to handle files coming from a wide variety of experimental approaches, and/or software tools. This is why PRIDE has always committed to use community data standards as formulated by the HUPO-PSI (Proteomics Standards Initiative, http://www.psidev.info). In 2011 PRIDE will adopt the emerging new data standards mzML (for MS) and mzIdentML (for protein and peptide identifications).