Poster G-14

Storage, management and exchange of experimental and validated proteomics data considering HUPO-PSI standards

Authors:
Klaus-Peter Pleissner (*Max Planck Institute for Infection Biology , Berlin, Germany*)
Benjamin Arndt (*University of Applied Sciences, Berlin, Germany*)
Daniel Paape (*Max Planck Institute for Infection Biology , Berlin, Germany*)
Toni Aebischer (*Max Planck Institute for Infection Biology , Berlin, Germany*)
Monika Schmid (*Max Planck Institute for Infection Biology , Berlin, Germany*)
Robert Stein (*Max Planck Institute for Infection Biology , Berlin, Germany*)
Peter R. Jungblut (*Max Planck Institute for Infection Biology , Berlin, Germany*)

Short Abstract: Comprehensive data management and intelligent data analysis are nowadays indispensable preconditions in proteomics research. By the investigation of the proteome of Leishmania mexicana amastigotes we demonstrate the versatile application of information management and databases for experimental and validated proteomics data considering the requirements of the HUPO-Proteome Standardization Initiative (PSI).

Long Abstract:
Comprehensive data management (storage, organisation, exchange) and intelligent data analysis are nowadays indispensable preconditions in proteomics research. Therefore, Proteome Information Management Systems (PIMS) for easily feasible acquisition of experimental data, dissemination of validated data using web-accessible proteome databases, data exchange by versatile standardized formats according to HUPO-Proteome Standardization Initiative (PSI) [1] and analysis tools for proteins, gels, LC/MS spectra etc. must be applied. Here we describe our approach for this task on the basis of publicly available and self-developed tools and systems following the sample data flow starting from protein sample solution and ending with standardized data exchange possibilities of mass spectra. For the routine acquisition of experimental samples, protocols, 1-D, 2-D image we are using the P.I.M.S. of the Babraham Institute [2] which is a flexible LIMS based on a client/server structure and MySQL. If the experiment consists of many gels produced under different conditions (treated vs. untreated) they are processed by the gel image analysis system PDQuest to detected differentially regulated proteins. The subsequent protein identification is realized by MALDI-TOF/TOF-MS (4700 Proteomics Analyzer, Applied Biosystems) which also enables to export peptide mass fingerprints (PMF) in the standardized mzData XML-format [3]. Fulfilling the requirements of PSI we are using the comprehensive open source system Proteios [4] for which we have written import tools enabling either the connection of the PDQuest or TopSpot spot numbers on the gels with the PMFs of identified proteins in mzData-format. Beside other information on experiments Proteios also contains links to raw MS spectra which reside on our public web-server [5]. The Proteios system is hierarchically organized according to sample data flow. Both systems (P.I.M.S. and Proteios) complement one another. As an example of application of our workflow we processed the proteome of Leishmania mexicana amastigotes containing 2-DE gels with 192 identified protein spots with the precise m/z-values and peak heights data in

mzData-format for each spot. This information together with raw spectra allows to reproduce the identification of proteins by other researchers to confirm the results and is a prerequisite of manuscript submission to journals [6].
The experiment-driven approach aforementioned is a reasonable complement to our web-accessible proteome database system [7] containing of validated proteomics data (2D-gels with identified protein spots, protein annotations, mass spectrometry data etc.) of more than 19 organisms and strains with over 4250 identified protein spots in thirty-five 2-DE gels.

[1] Orchard, S., Hermjakob, H. and Apweiler, R. (2003) Proteomics, 3, 1374-1376.
[2] http://www.bioinformatics.bbsrc.ac.uk/projects/pims/
[3] http://psidev.sourceforge.net/ms/
[4] Garden, P., Alm, R. and Hakkinen, J. (2005) Bioinformatics, 21, 2085-2087.
[5] http://web.mpiib-berlin.mpg.de/proteios/ms/
[6] Jones, P., Cote, R.G., Martens, L., Quinn, A.F., Taylor, C.F., Derache, W., Hermjakob, H. and Apweiler, R. (2006) Nucleic Acids Res, 34, D659-663.
[7] http://www.mpiib-berlin.mpg.de/2D-PAGE/