Poster B-31

MS-PED: A Management System for Proteomics Experimental Data

Authors:
W. Yan (*Bioinformatics User Facility, University of California, Los Angeles, CA*)
S. Than (*School of Dentistry, University of California, Los Angeles, CA*)
R. Qiao (*School of Medicine, University of Southern California, Los Angeles, CA*)
D.S. Parker (*Department of Computer Science, University of California, Los Angeles, CA*)
D. Wong (*School of Dentistry, University of California, Los Angeles, CA*)

Short Abstract: MS-PED: A Management System for Proteomics Experimental Data
W. Yan1, S. Than2, R. Qiao3, D.S. Parker4, D. Wong2 1Bioinformatics User Facility, 4Department of Computer Science, and 2School of Dentistry, University of California, Los Angeles, CA; 3School of Medicine, University of Southern California, Los Angeles, CA.
MS-PED is a web-based management system for proteomics experimental data. It differs from existing platforms in handling both bottom-up and top-down proteomics experimental data, its support for a PEDRo-like level of information detail, its integration of tools with the database, its portable web-based design, and its commitment to open-source availability.

Long Abstract:
MS-PED: A Management System for Proteomics Experimental Data
 W. Yan1, S. Than2, R. Qiao3, D.S. Parker4, D. Wong2 1Bioinformatics User Facility, 4Department of Computer Science, and 2School of Dentistry, University of California, Los Angeles, CA; 3School of Medicine, University of Southern California, Los Angeles, CA.
 Introduction: Rapid advances in the technology of Mass Spectrometry (MS) make it feasible to perform large-scale identification of proteins present in a cell, tissue or even an organism. The increasing scale of biological research using MS-based proteomics technology has led to rapid accumulation of vast amounts of data generated by proteomics experiments. Consequently, it has become increasingly important and necessary to develop data management system for storage, query and mining of data from proteomics experiments. In the present study, we introduce an efficient and user-friendly management system for proteomics experimental data (MS-PED) developed for the Human Salivary Proteome project at UCLA.
 Results: MS-PED is a web-based data storage and analysis software package containing features for storing, annotating and mining proteomics experimental data. It captures and stores information involved in proteomics experiments, including sample source, sample collection, sample separation, as well as MS experiment and MS result analysis for peptide and protein identification. It takes both gel-based and liquid-chromatography (LC)-based proteomics experimental data and accommodates top-down and bottom-up proteomics. The back-end relational database design of MS-PED is in symphony with the proposed standard from proteomics initiatives and guidelines for publication of protein and peptide identified through MS. Web-based graphics interfaces and applications are developed that allows users to enter, edit, view, query, search, and download data. Multiple features, such as uploading MS analysis results and templates for repeatedly-used protocols, are built into the web interfaces to facilitate data entry and reduce manual data entry. An automatic tool allows users to download and disseminate proteomics

experimental data from MS-PED to other proteomics databases in XML format, the data entry format commonly adopted by those proteomics databases, such as PRIDE and PEDRo databases. Security of data in MS-PED is managed and authorized through an administrative web interface. In addition, MS-PED provides tools to annotate protein identifications using other biological databases, such as Primary Protein Sequence databases and the Gene Ontology database. Mining tools are developed for comparison of data from different experiments. All computing tools for construction of MS-PED are open-source. (The system runs in a Linux environment with MySql as the relational database server, Apache as the web server, Perl and Javascript as the main scripting languages.) After initial set-up, the system can be run and used without substantial computational support.
 Conclusion: Unlike existing proteomics databases, which were designed to store either Gel-based or LC-based proteomics experimental data derived from bottom-up approach, MS-PED stores both kinds of experimental data from bottom-up as well as top-down approach. It stores information in a level of detail similar to PEDRo, but where PEDRo separates its data capture tools from the database, all tools built under MS-PED are web-based and run under the same platform. Users enter experimental data into MS-PED without conversion to a predefined format such as excel or XML and the entered data can be viewed, edited, searched and mined under the same interface as the data entry. Specific tools are designed and implemented in MS-PED to increase the efficiency of the data entry. Besides, multiple search and mining functions are provided in MS-PED. The system is currently being used to support the human salivary proteome project at UCLA (http://www.hspp.ucla.edu).
 MS-PED is supported by a grant from the National Institute of Dental and Craniofacial Research (NIH UO1 DE16275).