

## APPENDIX H.2



### HUPO Proteomics Standards Initiative (PSI) Activities report 07/2009 to 06/2010



#### Co-Chairs

Henning Hermjakob, EMBL-EBI, hhe@ebi.ac.uk  
Eric Deutsch, ISB, edeutsch@systemsbiology.org  
Randall K. Julian, rkjulian@indigobio.com

#### Mission statement

The HUPO Proteomics Standards Initiative (PSI) defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification.

#### Summary

In the reporting period, the PSI has been highly successful. One PSI-related EU grant (PSIMEx, 1.5 M Euro) is currently active, another one, ProteomeXchange (1.8 M Euro) is expected to start 1/1/2011. Two workshops were held, and 12 PSI-related journal articles published (9 “content” manuscripts, 3 meeting reports/reviews [1-3]).

#### Organization

##### Work groups

The main organizational unit of the Proteomics Standards Initiative is the work group. Currently, there are the following work groups:

- [Molecular Interactions \(MI\)](#)
- [Mass Spectrometry \(MS\)](#)
- [Proteomics Informatics \(PI\)](#)
- [Protein Separation \(PS\)](#)
- [Protein Modifications \(MOD\)](#)

##### Inter-group activities

To ensure consistency between work group deliverables, the PSI also maintains the following inter-group activities:

- Controlled vocabularies
- [Steering Committee](#)
- MIAPE (Minimum Information About a Proteomics Experiment)

##### Deliverables

The standard deliverables of each work group are

- Minimum Information Specification: For the given domain, this specifies the minimum information required for the useful reporting of experimental results in this domain.
- Formal exchange format for experimental results in the domain. This will usually be an XML format, capable of representing at least the Minimum Information, and normally significant additional detail.
- Controlled vocabularies.

- Support for implementation of the standard in publicly available tools.

### **The PSI document process**

With increasing investments in PSI standard implementations by instrument providers, software developers, and end users, we developed a formal process for the development and maintenance of PSI standards. The process and its implementation on the PSI web site have been published by Vizcaino et al [4].

### **HUPO Initiatives Interaction**

The HUPO PSI has a subsidiary role supporting other HUPO projects and workgroups, aiming to ensure collaboration and consistency in the data management approaches taken by the HUPO proteomics initiatives. We have contributed to the data management of the HUPO Plasma Proteome (I), Brain, and Liver Proteome Projects, and participate in the proposed Model Organisms Initiative (iMOP). The HUPO PSI benefits greatly from these interactions, as regular interactions with HUPO experimentalists and analysts help to ensure that our own activities are not conducted in a void.

In the context of the EU ProteomeBinders project, an ongoing collaboration with the HUPO Antibody initiative had significant impact on the recently published “Community standard for the representation of protein affinity reagents (PSI-PAR)” [5, 6].

A successful interaction with the HUPO publications committee (joint workshops) has resulted in strong connections to key proteomics journals, highly valuable for dissemination and implementation of PSI standards.

Currently, major proteomics repositories represented in the PSI mass spectrometry work group (PRIDE, PeptideAtlas, Tranche) are defining the systematic data capture for the second phase of the Plasma Proteome Project (PPP), as well as for the joint HUPO project. A first large scale dataset for HUPO PPP II has been processed through the proposed pipeline. This preliminary work has contributed to the EU ProteomeXchange grant, which involves 12 partners from Europe and North America, and is expected to start 1/1/2011.

### **Activities**

#### **Workshops and Meetings**

- PSI spring meeting, Seoul, Korea, April 2010, 25 participants
- PSI workshop as part of the HUPO 2009 conference, Toronto, September, 35 participants [1].

#### **Standards development**

The following standard documents/guidelines have been publicly released/published by the PSI in the reporting period:

- Proteomics Informatics:
  - mzIdentML: While mzML [7] is a format for mass spectra representation, mzIdentML describes the next step, the process and results of mass spectra analysis, leading to peptide and protein identifications. mzIdentML 1.0.0 was released on August 20, 2009.
- Molecular Interactions:
  - PAR: Standard Format for Protein Affinity Reagents: With support from the EU ProteomeBinders project, we have extended the existing PSI MI format for molecular interactions to application in a related domain, the representation of Protein Affinity Reagents, for example antibodies. The PSI-PAR format was released in the last reporting period, the corresponding journal publication published in this reporting period [6].
  - The corresponding Minimum Information for a Protein Affinity Reagent (MIAPAR) has also been published in this reporting period [5].

- PSICQUIC: PSI Common QUery InterfaCe: PSICQUIC is an API for computational access to molecular interaction data. It has been developed by the PSI MI group and is now supported by 12 molecular interaction data sources, among them DIP, IntAct, MINT, ChEMBL, and STRING. As of August 2010, 12 million (redundant) binary interactions are accessible through PSICQUIC [publication in preparation].
- Protein Separations: After a significant preparation period, the Separations work group published a series of three articles in Nature Biotechnology:
  - MIAPE-GEL: “Guidelines for reporting the use of gel image informatics in proteomics” [8]
  - MIAPE-CC: “Guidelines for reporting the use of column chromatography in proteomics” [9]
  - MIAPE-CE: “Guidelines for reporting the use of capillary electrophoresis in proteomics” [10]

### **Tools and applications**

Among the PSI related publications in this reporting period, one group deserves special mention, namely those which concern tools and implementations related to PSI standards, rather than the “standards” publications themselves. These range from the PSI semantic validator [11] via a “Semi-automatic tool to describe, store and compare proteomics experiments based on MIAPE compliant reports” [12] to “A general pipeline for quality and statistical assessment of protein interaction data using R and Bioconductor”[13]. Also worth mentioning in this section is the implementation of mzIdentML in the latest version of the highly popular Mascot search engine ([http://www.matrixscience.com/help/export\\_help.html#MZIDENTML](http://www.matrixscience.com/help/export_help.html#MZIDENTML)).

### **Journal collaboration**

An ongoing dialog with journal editors has been continued, with a satellite meeting being organised at the HUPO Congress 2009, Toronto between representatives from 6 domain specific and 2 of the Nature Publishing group journals and the PSI Steering Committee. This meeting addressed the status of the current MIAPE guidelines and discussed any updates and amendments the journals felt to be appropriate. Largely as a result of this ongoing process, the journals MCP and Proteomics have both significantly increased their requirements for reporting of data supporting publications in this reporting period.

### **Grants**

- **ProteomeXchange**  
The EU project “International Data Exchange and Data Representation Standards for Proteomics” is expected to start 1/1/2011. It comprises 12 partners, thereof 2 US partners and one publisher, and has an overall volume of 1.8 M Euro. As indicated in the title, the two aims are to establish an international proteomics data exchange, and further develop PSI standards, in particular for quantitative proteomics.
- **PSIMEx**  
The EU project “Proteomics Standards Initiative and International Molecular Exchange – Systematic Capture of Published Molecular Interaction Data (PSIMEx)” started in January 2009. The project proposal comprises 14 partners from Europe, China, and the USA, and has an overall volume of 1.5 M Euro. Primary aim is the regular exchange of molecular interaction data between major data providers, based on HUPO PSI standards.
- **jmzIdentML**  
The BBSRC has funded an 18 month project to develop an API and viewer for the mzIdentML standard. The API will help proteome informatics developers to embed support within their tools for mzIdentML and is expected to be used within PRIDE for data import. The viewer will provide

simple access to the standard for bench scientists, enabling data curation and simplifying submission to repositories.

## Plans

- Activities
  - mzQuantML, standard format for quantitative mass spectrometry
  - TraML, standard format for SRM transitions
  - Update of MIAPE-MS specification, based on feedback and discussion among a group of journal editors and PSI representatives
  - PSICQUIC, the PSI Common Query Interface for molecular interactions
- Events
  - mzQuantML workshop EBI, 31/8/2010-1//9/2010
  - HUPO PSI workshop at the HUPO Sydney Conference, September 2010
  - HUPO PSI spring meeting 2011, April 4-6. Advanced Training Center, EMBL, Heidelberg, Germany

## Publications

Publications on “content” rather than meeting reports are marked with \*

### PSI Publications for the reporting period

1. Orchard, S., Albar, J., Deutsch, E.W., Eisenacher, M., Binz, P. & Hermjakob, H. implementing data standards: a report on the HUPOPSI workshop September 2009, Toronto, Canada. *Proteomics* 10, 1895-8(2010).
2. Orchard, S. Ending the "publish and vanish" culture: how the data standardization process will assist in data harvesting. *J. Proteome Res.* 8, 3219(2009).
3. Martínez-Bartolomé, S., Blanco, F. & Albar, J. Relevance of proteomics standards for the ProteoRed Spanish organization. *J Proteomics* 73, 1061-6(2010).
5. \*Bourbeillon, J., Orchard, S., Benhar, I., Borrebaeck, C., de Daruvar, A., Dübel, S. et al. Minimum information about a protein affinity reagent (MIAPAR). *Nat. Biotechnol.* 28, 650-3(2010).
6. \*Gloriam, D.E., Orchard, S., Bertinetti, D., Björling, E., Bongcam-Rudloff, E., Borrebaeck, C.A.K. et al. A community standard format for the representation of protein affinity reagents. *Mol. Cell Proteomics* 9, 1-10(2010).
7. \*Deutsch, E. mzML: a single, unifying data format for mass spectrometer output. *Proteomics* 8, 2776-7(2008).
8. \*Hoogland, C., O’Gorman, M., Bogard, P., Gibson, F., Berth, M., Cockell, S.J. et al. Guidelines for reporting the use of gel image informatics in proteomics. *Nat. Biotechnol.* 28, 655-6(2010).
9. \*Jones, A.R., Carroll, K., Knight, D., Maclellan, K., Domann, P.J., Legido-Quigley, C. et al. Guidelines for reporting the use of column chromatography in proteomics. *Nat. Biotechnol.* 28, 654(2010).
10. \*Domann, P.J., Akashi, S., Barbas, C., Huang, L., Lau, W., Legido-Quigley, C. et al. Guidelines for reporting the use of capillary electrophoresis in proteomics. *Nat. Biotechnol.* 28, 654-5(2010).
11. \*Montecchi-Palazzi, L., Kerrien, S., Reisinger, F., Aranda, B., Jones, A.R., Martens, L. et al. The PSI semantic validator: a framework to check MIAPE compliance of proteomics data. *Proteomics* 9, 5112-9(2009).
- 12.\* Martínez-Bartolomé, S., Medina-Aunon, J.A., Jones, A.R. & Albar, J.P. Semi-automatic tool to describe, store and compare proteomics experiments based on MIAPE compliant reports. *Proteomics* 10, 1256-60(2010).
13. \*Chiang, T. & Scholtens, D. A general pipeline for quality and statistical assessment of protein interaction data using R and Bioconductor. *Nat Protoc* 4, 535-46(2009).

**Relevant PSI references prior to the reporting period**

4. \*Vizcaíno, J.A., Martens, L., Hermjakob, H., Julian, R.K. & Paton, N.W. The PSI formal document process and its implementation on the PSI website. *Proteomics* 7, 2355-7(2007).
14. [Editors] Democratizing proteomics data. *Nat. Biotechnol.* 25, 262(2007).
15. [Editors] Thou shalt share your data. *Nature Methods* 5, 2009(2008).
16. [Editors] Credit where credit is overdue. *Nat. Biotechnol.* 27, 579(2009).
17. [Editors] Mind the technology gap. *Nat. Methods*, 765(2007).