HUPO Highlights

The initiative on Model Organism Proteomes (iMOP) Session

September 6, 2011, Geneva, Switzerland

Sabine P. Schrimpf[1,2], Christian von Mering[1,3],
Emøke Bendixen[4], Joshua L. Heazlewood[5], Dirk Bumann[6],
Gil Omenn[7] and Michael O. Hengartner[1,2]

[1] Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland
[2] Quantitative Model Organism Proteomics, University of Zurich, Zurich, Switzerland
[3] Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland
[4] Department of Animal Health and Bioscience, Faculty of Agricultural Sciences, Aarhus University, Tjele, Denmark
[5] Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[6] Infection Biology, Biozentrum, University of Basel, Basel, Switzerland
[7] Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

iMOP – the Initiative on Model Organism Proteomes – was accepted as a new HUPO initiative at the Ninth HUPO meeting in Sydney in 2010. A goal of iMOP is to integrate research groups working on a great diversity of species into a model organism community. At the Tenth HUPO meeting in Geneva this variety was reflected in the iMOP session on Tuesday September 6, 2011. The presentations covered the quantitative proteome database PaxDb, proteomics projects studying farm animals, *Arabidopsis thaliana*, as well as host–pathogen interactions.

Keywords:

Michael Hengartner (University of Zurich, Switzerland) opened the session outlining the history of iMOP (http://www.imop.uzh.ch/ and http://www. hupo.org/research/iMOP/). A first workshop to identify the necessity of a new initiative on model organism proteomes was organized in Amsterdam 2008. In 2010 iMOP was accepted as a new HUPO initiative. During the iMOP founding meeting on April 27–28, 2011, in Zurich, the goals of iMOP and the outline of planned activities were set. The first iMOP session integrated into the regular HUPO meeting took place in Geneva in September 2011. The goals of iMOP are: to integrate different model organism proteomics groups into a community and to promote interactions between them (e.g. the development of software tools to navigate databases and to compare proteomes); to integrate model organism communities better into HUPO (e.g. adopt HUPO standards and best practices, create synergies with other HUPO initiatives); and to promote proteomics within

the model organism communities (e.g. integrate and link proteome and organism-specific databases). iMOP welcomes scientists working on all non-human species, like classical ''model organisms'', agriculturally relevant species, medically relevant species, and ''non-model organisms'' (including metaproteomes). iMOP is aiming for a rich diversity of participating groups. The diversity of model organisms within iMOP was underlined at the HUPO meeting in Geneva. The four speakers of the iMOP session covered: computational analyses and databases for quantitative proteomics and the analysis of evolutionary properties; farm animals for maximum food production and as model organisms to understand human biology; approaches to analyze protein function in plant organelles; proteome analysis of host–pathogen interactions.

Christian von Mering (University of Zurich, Switzerland) introduced PaxDb ''a database dedicated to comparative proteome analysis in model organisms'' (http://pax-db.org/; Wang et al., in preparation). Quantitative data in PaxDb are derived from shotgun proteomics data by spectral counting but also from Western blotting or systematic GFP tagging. Currently, PaxDb covers 12 species (human, animals, microorganisms, and *Arabidopsis thaliana*). Driving questions to develop such a quantitative proteome database were: what is the absolute abundance of a given (human) protein, how are protein abundances changing during evolution, and is there an optimal abundance level? In PaxDb protein abundance histograms and lists of the most abundant proteins are generated for each species. Quantitative data from the original data sets are provided, but also – for each organism – a combined data set that often achieves demonstrably higher coverage and accuracy. Abundances are mapped between organisms via orthologous protein relations (from http://eggnog.embl.de/). Dr. von Mering presented different quantitative analyses: operons in *C. elegans* encode abundant genes, and protein abundance between different species correlates better than protein and mRNA abundance within a species [1, 2]. This led to the conclusion that mRNA levels may drift during evolution. A comparison of five species revealed that the protein abundance of the expressed core proteome is conserved [3] although with intrinsic variations: for example, the orthologs of the yeast protein GPA2 show a high variance, whereas the abundances of orthologs of the yeast protein UBC1 are currently among the best conserved in all eukaryotes.

Emøke Bendixen (Aarhus University, Denmark) presented in her talk on ''farm animals – a new generation of model organisms'' a variety of proteomics projects aiming at understanding food production biology on one hand and understanding human biology on the other hand. She outlined that in nature biological variation is based on natural selection, while in farm animals it originates to a wide extent from selective breeding aiming at maximizing milk and meat production. Thus, farm animals can serve as biological samples to study extreme physiology. Because of their similarities to humans, farm animals are valuable as model organisms. Particularly, the close similarity in genomes, physiology, and anatomy of pig and humans make pig a very useful model organism. The use of pig to study neurological diseases like ALS (amyotrophic lateral sclerosis) and Parkinson disease, and to study obesity and metabolic syndromes, is already well established. Dr. Bendixen presented data from a recent project that aimed at studying the influence of the microbiome on obesity in pigs using iTRAQ. A screening of 60 gut bacteria in co-culture with intestinal epithelia cells revealed that *Butyrovibrio fibrisolvens* particularly modulated lipid metabolism, immune response and inflammation, and histone release in gut cells in vitro. To test whether obesity traits can be influenced by the microflora also in vivo, new experiments are aiming to feed pigs with specific bacterial food supplements ('designer yoghurt'), sand key metabolic enzymes are quantified by targeted proteomics. Within milk production a particular challenge remains on fighting mastitis in lactating cows. Ongoing work on iTRAQ studies, and the development of SRM assays using QconCat peptides for mapping host response

proteins and biomarkers for mastitis directly in milk [4] was introduced. A great challenge in farm animal proteomics is to build up a community, an effort that is enhanced by the iMOP initiative.

Joshua Heazlewood (Lawrence Berkeley National Laboratory, USA) talked about ''mining post translational modifications in the endomembrane of the model plant *Arabidopsis thaliana*''. The biosynthesis of matrix polysaccharides for the construction of plant cell walls is currently not well characterized. To bypass inherent difficulties and to enable the analysis of such processes, he described three approaches and tools to characterize protein function: (i) proteomic analysis of the Golgi, (ii) MASCP Gator, a web platform to aggregate proteome data from online repositories, and (iii) ModHunter, a tool to screen for PTMs.

The Golgi proteome was enriched by FFE (free flow electrophoresis) to characterize proteins involved in cell wall biosynthesis: 425 proteins were identified including enzymes that produce and deliver sugars to the cell wall. These data were incorporated with the SUBA database (SUBcellular *Arabidopsis*; http:// suba.plantenergy.uwa.edu.au/) which houses localization information for proteins identified from a range of organelles (e.g. plasma membrane, vacuoles, and mitochondria). The SUBA resource currently contains localization information for over 10 000 proteins in *Arabidopsis* [5]. The MASCP Gator (http://gator.masc-proteomics.org/) is a web portal for the aggregation of online proteomics data sources (e.g. SUBA) [6] and allows simple extraction of information about a proteins function [7]. MASCP Gator contains information on identified peptides, phosphorylation, protein description, subcellular localization and organ/tissue evidence, and provides a visual overview of experimental information on *Arabidopsis* proteins. While *Arabidopsis* has the capacity to synthesize ca. 35 000 proteins, when considering protein modifications there are actually many 100 000's distinct functional states. The MASCP Gator is being extended to create the ModHunter, an undirected PTM screening tool based on extensive MS/MS data. Since the majority of MS/MS data matching occurs with no modifications, the MASCP Gator identifies gaps or 'hot-spots' in protein coverage by MS that could carry PTM's. The chemical removal of glycosylation sites from FFE isolated Golgi samples was shown as an example of the validation of ModHunter and the presence of gaps in global MS/MS data. The technique can be used for modification-based comparative proteomics without knowing the modification.

Dirk Bumann (University of Basel, Switzerland) talked about ''Salmonella proteome in infected host tissue''. The main objective is to elucidate pathogen in vivo properties as a basis for urgently needed new strategies for control of infectious diseases. *Salmonella* adaptations to hostile host environments were investigated in a mouse typhoid fever model. *Salmonella* were purified from infected spleen using flow cytometry and analyzed by shotgun proteomics. Absolute quantification of some 900 *Salmonella* proteins enabled reconstruction of *Salmonella* metabolism in host tissues, and provided a basis for building accurate in silico models with high predictive power for antimicrobial identification. This entire approach was initially focused on average *Salmonella* properties during infection. Close inspection of infected tissues, however, revealed remarkable heterogeneity of *Salmonella* microenvironments. As an example, about half of *Salmonella* were exposed to potent host inflammatory cells with multiple bacterial killing mechanisms. To investigate this issue, a growth rate biosensor was developed that revealed *Salmonella* subpopulations with a wide range of growth rates in the same infected tissue. Sorting and proteome analysis of these subpopulations suggested differential nutrient supply as one potential cause of *Salmonella* heterogeneity. Pathogen subpopulations with diverse properties could be important as they might require different antimicrobial approaches for effective control.

In the closing discussion, Christian von Mering suggested to merge iMOP projects within a common web platform like Gator (i.e. an aggregation of

projects without hosting). Michael Hengartner emphasized one of the goals of iMOP, the interaction with curators of organism-specific databases (e.g. FlyBase, WormBase) to include tracks with protein abundance information. Gil Omenn, chair of the HUPO Initiatives and newly elected chair of the HUPO Human Proteome Project, pointed out that all existing active HUPO initiatives will persist individually and also should contribute to the Human Proteome Project [8]. The broad interest in multi-omics platforms, the transmission of infectious agents to and from animals and humans, and agricultural proteomes and many other species related to climate change is sure to increase. Omenn mentioned specific relationships of iMOP to the Human Proteome Project, including studies dealing with agriculture and food production and their impact on humans, microbiome/metagenome studies of humans and our internal and external ecosystems, and the power of comparative model organism proteomics. By analogy to the Human Genome Project, proteomics in general and model organism proteomics in particular contribute to phylogenetics and enable evolutionary studies. There is emerging interest in enhancing the teaching of evolutionary biology in medical and public health schools [9, 10]. He also recommends that iMOP considers use of ProteomeXchange to link SwissProt/UniProt, EBI/PRIDE, ISB/PeptideAtlas, GPMdb, and Tranche, as adopted by the Human Proteome Project. PeptideAtlas reanalyses from the raw spectra results for any given species, organ, or biofluid proteome with uniform methodology and rigorous statistical cutoffs, using the TransProteomicPipeline and now the Cedar scheme for tiered protein identification [11].

The iMOP session ended with the announcement of the next iMOP meeting that will take place in Kiel, Germany, March 26–27, 2012, the organizer is Andreas Tholey (University of Kiel), followed by the 11th Annual HUPO World Congress of Proteomics iMOP Workshop in Boston, September 9–13, 2012.

References

[1] Schrimpf, S. P., Weiss, M., Reiter, L., Ahrens, C. H. et al., Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* 2009, *7*, e48.

[2] Laurent, J. M., Vogel, C., Kwon, T., Craig, S. A. et al., Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics* 2010, *10*, 4209–4212.

[3] Weiss, M., Schrimpf, S., Hengartner, M. O., Lercher, M. J., von Mering, C., Shotgun proteomics data from multiple organisms reveals remarkable quantitative conservation of the eukaryotic core proteome. *Proteomics* 2010, *10*, 1297–1306.

[4] Danielsen, M., Codrea, M. C., Ingvartsen, K. L., Friggens, N. C. et al., Quantitative milk proteomics – host responses to lipopolysaccharide-mediated inflammation of bovine mammary gland. *Proteomics* 2010, *10*, 2240–2249.

[5] Heazlewood, J. L., Verboom, R. E., Tonti-Filippini, J., Small, I., Millar, A. H., SUBA: the Arabidopsis subcellular database. *Nucleic Acids Res.* 2007, *35*, D213–218.

[6] Joshi, H. J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W. et al., MASCP Gator: an aggregation portal for the visualization of Arabidopsis proteomics data. *Plant Physiol.* 2011, *155*, 259–270.

[7] Tang, W., Kim, T. W., Oses-Prieto, J. A., Sun, Y. et al., BSKs mediate signal transduction from the receptor kinase BRI1 in Arabidopsis. *Science* 2008, *321*, 557–560.

[8] Legrain, P., Aebersold, R., Archakov, A., Bairoch, A. et al., The human proteome project: current state and future direction. *Mol. Cell. Proteomics* 2011, *10*, M111 009993.

[9] Omenn, G. S., Evolution in health and medicine Sackler colloquium: evolution and public health. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 1702–1709.

[10] Nesse, R. M., Bergstrom, C. T., Ellison, P. T., Flier, J. S. et al., Evolution in health and medicine Sackler colloquium: making evolutionary biology a basic science for medicine. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 1800–1807.

[11] Farrah, T., Deutsch, E. W., Omenn, G. S., Campbell, D. S. et al., A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* 2011, *10*, M110 006353.