HUPO Highlights

The HUPO initiative on Model Organism Proteomes, iMOP

Alexandra M. E. Jones[1], Ruedi Aebersold[2,3], Christian H. Ahrens[4,5], Rolf Apweiler[6], Katja Baerenfaller[7], Mark Baker[8], Emøke Bendixen[9], Steve Briggs[10], Philip Brownridge[11], Erich Brunner[4,5], Michael Daube[4,5], Eric W. Deutsch[12], Ueli Grossniklaus[4,13], Joshua Heazlewood[14], Michael O. Hengartner[4,5], Henning Hermjakob[6], Marko Jovanovic[4,5], Craig Lawless[11], Günter Lochnit[15], Lennart Martens[16,17], Christian Ravnsborg[18], Sabine P. Schrimpf[4,5], Yhong-Hee Shim[19], Deni Subasic[4,5], Andreas Tholey[20], Klaas van Wijk[21], Christian von Mering[4,5,22], Manuel Weiss[4,5] and Xue Zheng[4,5]

[1] The Sainsbury Laboratory, Norwich Research Park, Norwich, UK
[2] Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland
[3] Faculty of Science, University of Zurich, Zurich, Switzerland
[4] Quantitative Model Organism Proteomics, University of Zurich, Zurich, Switzerland
[5] Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland
[6] European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
[7] Department of Biology, ETH Zurich, Zurich, Switzerland
[8] Chemistry & Biomolecular Sciences, Macquarie University, NSW, Australia
[9] Department of Animal Science – Integrative Physiology, Tjele, Denmark
[10] UCSD, Gilman Drive, La Jolla, CA, USA
[11] Faculty of Life Sciences, University of Manchester, UK
[12] Institute for Systems Biology, Seattle, WA, USA
[13] Institute of Plant Biology, University of Zurich, Zurich, Switzerland
[14] Joint BioEnergy Institute, Lawrence Berkeley National Laboratory and Physical Biosciences Division, Berkeley, CA, USA
[15] Justus-Liebig University, Faculty of Medicine, Institute of Biochemistry, Protein Analytics, Giessen, Germany
[16] Department of Medical Protein Research, VIB, Gent, Belgium
[17] Department of Biochemistry, Ghent University, Gent, Belgium
[18] Thermo Fisher Scientific, Copenhagen, Denmark
[19] Department of Bioscience and Biotechnology, Konkuk University, Seoul, Korea
[20] Christian-Albrechts-Universität zu Kiel, Kiel, Germany
[21] Department of Plant Biology, Emerson Hall, Cornell University, Ithaca, NY, USA
[22] Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland

HUPO Human Proteome Organisation Highlights

The community working on model organisms is growing steadily and the number of model organisms for which proteome data are being generated is continuously increasing. To standardize efforts and to make optimal use of proteomics data acquired from model organisms, a new Human Proteome Organisation (HUPO)

initiative on model organism proteomes (iMOP) was approved at the HUPO Ninth Annual World Congress in Sydney, 2010. iMOP will seek to stimulate scientific exchange and disseminate HUPO best practices. The needs of model organism researchers for central databases will be better represented, catalyzing the integration of proteomics and organism-specific databases. Full details of iMOP activities, members, tools and resources can be found at our website http://www.imop.uzh.ch/ and new members are invited to join us.

Keywords:
Data standardization / Human Proteome Organisation / Model organisms / Proteomics initiative

1 Introduction

The principle behind the use of model organisms is that fundamental biological processes are common and that by studying an experimentally amenable species one can make valid inferences to better understand a less studied species. Perhaps, the most famous example is given by Gregor Mendel's peas, which he carefully selected based on the phenotypes of the traits he was interested in, and from which he derived the general principles of inheritance. Likewise, modern model organisms are generally chosen to be easily cultivated in the laboratory, to be amenable to experimental manipulation and to have rather short generation times. A historical perspective is provided by Müller and Grossniklaus [1]. Several model organisms were amongst the first to have fully sequenced and annotated genomes, and in parallel many of these systems also developed extensive systematic genetic resources (e.g. knock-out lines, facile transformation, RNAi libraries). The use of model organisms can thus permit experiments that would be impossible in other organisms, such as genetic and transgenic manipulations, see [2] for an illustrative review.

Not all biological processes and pathways are common to all life, so the relevance of a model organism might be limited to relatively closely related species, such as mouse models of human diseases, or the weedy *Arabidopsis thaliana* for crop plants. Conversely, evolutionary insights can be gained through comparative studies of diverse organisms, for example the innate immune systems of plants or animals [3], the conservation of histone modifications [4] or comparative functional analysis of worm and fly proteomes [5]. Indeed, such is the diversity of model organisms, and the rapidity of genome sequencing and genetic tool generation, that it is becoming difficult to draw a line between a 'model' and a 'non-model' organism. While neither farm animals nor plant pathogenic fungi spring to mind when considering proteomic experiments, both areas have been recently explored by proteomic tools [6, 7]. Moreover, many organisms interact with other organisms, either as important commensals or as pathogens [8, 9]. Thus, the knowledge of these organisms' proteomes is also of significant importance for understanding their interactions and full proteome.

At present, the analysis of proteomic data (particularly MS-based data) relies heavily on the availability of the appropriate annotated genome sequence for database-dependent spectrum assignments. Thus, well-established model organisms benefit from their long history of genetic resources and, for some of these organisms, large-scale proteomics studies have become an essential part of the research endeavour as exemplified recently by a special issue in *Journal of Proteomics* [10]. In some cases, proteomics has contributed to the identification of open reading frames and splice events [11, 12]. Many species benefit from stable locus identifiers, widely accepted 'gold-standard' databases, and extensive proteome coverage [13]. With the ability to use

metabolic labelling in several live model organisms, highly advanced proteomic techniques are becoming available, a few examples of recent work or reviews include: Arabidopsis [14], the SILAC fly [15], Chlamydomonas [16] and *Caenorhabditis elegans* [17]. Whilst there are some expert proteomic groups who work on many organisms, such expertise in proteomics is not easily reproduced by species specialists. Several model organisms serve as important tools for the development of novel analytical approaches, as their complexity is often reduced. An example is the use of the human pathogen *Leptospira interrogans* to determine the cellular concentration of proteins [18]. Furthermore, the availability of well-defined biological material (e.g. bacterial strains or cloned plants) permits reproduction and comparison of analytical techniques between many laboratories [19].

Scientists involved in research using model organisms tend to view themselves as members of specific species research communities or addressing a specific biological question, rather than as 'model organism researchers'. Some of these research communities are highly cooperative and well organized with community meetings and shared resources, including databases for proteomics (e.g. Arabidopsis [20, 21] and Drosophila [22]). While other endeavours have had a technical focus, e.g. a database of phosphorylation sites in model organisms [23] or a shared research focus e.g. a recent initiative for plant proteomics has been launched [24]. However, in other cases, communities and resources are more fragmented and proteomic studies are less advanced. In those cases, groups that initiate new proteomic projects may face basic technical challenges in their efforts to characterize the proteome of their favourite model organism. These can range from issues of a small community, difficulties in obtaining pure cell populations and integration of proteomics data into already well-established organism-centric databases.

In order to better address these challenges, we have established a new Human Proteome Organisation (HUPO) initiative on model organism proteomes (iMOP; http://www.imop.uzh.ch) to create a global network of experimental and bioinformatics groups interested in model organism proteomes. The initiative will ensure that the same principles, protocols and standards used in all current HUPO initiatives will also be applied to all model organisms. Importantly, the iMOP initiative will not focus on a single species but will bring together model organism proteomics researchers to reach the critical mass necessary to gain sufficient traction and impact within the scientific community. In these times of rapid genome sequencing, iMOP intends to operate a broadly inclusive policy; researchers of any 'non-human' organism are welcome to join, regardless of how 'well established' their organism is as a model.

2 Vision and objectives

The vision of iMOP is to create a community of researchers by providing resources and tools to facilitate the exchange of data, by sharing protocols and by providing best practice guidelines for model organism proteomics. This vision encompasses at least three levels: (i) interaction between model organism researchers and the main body of HUPO; (ii) interaction between proteomic researchers studying different organisms; and (iii) outreach or promotion of proteomics to the wider model organism communities. Due to these interactions (Fig. 1) iMOP will both promote a better visibility of model organism communities within HUPO, as well as a better appreciation of proteomics within the various model organism communities. Increased interactions among model organism proteomics researchers are also expected to lead to the development of joint resources and standards that should ultimately lead to improved and innovative science.

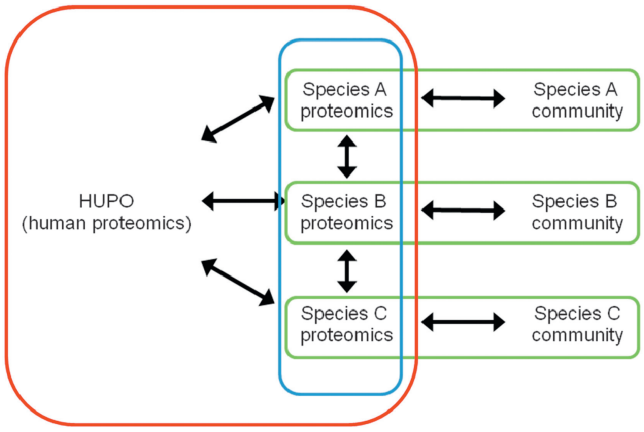

Figure 1. Interactions promoted by the iMOP initiative; interactions between HUPO and model organism communities (red box), interactions across model species groups to facilitate between-species comparisons (blue box) and interactions between proteomics researchers and the wider species-specific communities to promote proteomics as a key tool in these communities (green boxes).

Our specific objectives are:

(i) Integrate different model organism research groups into a model organism proteomics community and promote interaction between members.

Scientific interactions within iMOP will be encouraged both through the exchange of visiting researchers between the labs, as well as through workshops during the regular HUPO meetings and specific iMOP meetings. iMOP will also promote the flow of information into HUPO, for example by contributing to the comparative aspects of the Human Proteome Project [25].

(ii) Promote HUPO standards and best practices.

In analogy to established HUPO initiatives, the use of best practices will be adopted within the iMOP community. For example, standardized data acquisition, analysis, storage and exchange using established platforms will be facilitated by following HUPO PSI guidelines.

(iii) Integrate and link proteome and organism-specific databases.

Several proteomics data repositories encompassing different species exist (e.g. PRIDE [26], PeptideAtlas [27]). To facilitate access to all proteome data, iMOP will help to efficiently link proteome-centric and organism-based databases. iMOP will act as a catalyst to facilitate linking and integration of already established databases: it will not become another data repository itself. In this respect, it is already clear that meaningful sharing of proteome data will require a standardized file format and a consistent way for dealing with ambiguities, both in terms of protein inference and to support quantitative proteomics workflows [28]. Furthermore, software tools will need to be built to navigate databases efficiently and intuitively. Indeed, since most biologists are interested in using rather than generating proteomics data, it is important that proteomics data will be easily accessible, can easily be manipulated and analyzed and different data sets easily compared. iMOP will help develop the required software tools to navigate the databases and to extract and compare sub-data sets. Such tools

should, for example, allow for the global comparison of proteomes of different model organisms. Moreover, because many of these proteins have counterparts in human cells, these tools will allow for a comparative analysis of pathways and mechanisms conserved between model organisms and humans.

3 Organisation of iMOP

The founding meeting for iMOP took place on April 27 and 28, 2011 in Zurich, Switzerland. A number of issues were tabled including strategies on how to attract more groups working on species not yet represented within iMOP, and how to identify the biological needs of the model organism community. In addition, chairs were elected and the formal organization and structure of iMOP was undertaken. At the founding meeting of iMOP the participants unanimously elected Michael Hengartner (University of Zurich) as Chair and also appointed three co-chairs: Steve Briggs (UC San Diego), Andreas Tholey (Kiel University) and Klaas van Wijk (Cornell University). The co-ordinator of iMOP is Sabine Schrimpf (University of Zurich). The iMOP website is http://www.imop.uzh.ch/ and the organisation email is imop@imls.uzh.ch. A number of species are already well represented through the research focus of current iMOP members. Names and research interests of current members are available through the iMOP website.

4 Concluding remarks

Full details of iMOP activities, members, tools and resources can be found at our website http://www.imop.uzh.ch/ and new members are warmly invited to join our initiative.

The authors have declared no conflict of interest.

References

[1] Muller, B., Grossniklaus, U., Model organisms – a historical perspective. *J. Proteomics* 2010, *73*, 2054–2063.

[2] Gstaiger, M., Aebersold, R., Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat. Rev. Genet.* 2009, *10*, 617–627.

[3] Nurnberger, T., Brunner, F., Kemmerling, B., Piater, L., Innate immunity in plants and animals: striking similarities and obvious differences. *Immunol. Rev.* 2004, *198*, 249–266.

[4] Garcia, B. A., Hake, S. B., Diaz, R. L., Kauer, M. et al., Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.* 2007, *282*, 7641–7655.

[5] 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.

[6] Bendixen, E., Danielsen, M., Hollung, K., Gianazza, E., Miller, I., Farm animal proteomics – a review. *J. Proteomics* 2011, *74*, 282–293.

[7] Gonzalez-Fernandez, R., Prats, E., Jorrin-Novo, J. V., Proteomics of plant pathogenic fungi. *J. Biomed. Biotechnol.* 2010, *2010*, 932527.

[8] Becher, D., Hempel, K., Sievers, S., Zuhlke, D. et al., A proteomic view of an important human pathogen – towards the quantification of the entire *Staphylococcus aureus* proteome. *PLoS One* 2009, *4*, e8176.

[9] Malmstrom, L., Malmstrom, J., Aebersold, R., Quantitative proteomics of microbes: principles and applications to virulence. *Proteomics* 2011, *11*, 2947–2956.

[10] Ahrens, C. H., Schrimpf, S. P., Brunner, E., Aebersold, R., Model organism proteomics. *J. Proteomics* 2010, *73*, 2051–2053.

[11] Castellana, N. E., Payne, S. H., Shen, Z., Stanke, M. et al., Discovery and revision of *Arabidopsis* genes by proteogenomics. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 21034–21038.

[12] Baerenfaller, K., Grossmann, J., Grobei, M. A., Hull, R. et al., Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* 2008, *320*, 938–941.

[13] Ahrens, C. H., Brunner, E., Qeli, E., Basler, K., Aebersold, R., Generating and navigating proteome maps using mass spectrometry. *Nat. Rev. Mol. Cell. Biol.* 2010, *11*, 789–801.

[14] Wienkoop, S., Baginsky, S., Weckwerth, W., *Arabidopsis thaliana* as a model organism for plant proteome research. *J. Proteomics* 2010, *73*, 2239–2248.

[15] Sury, M. D., Chen, J. X., Selbach, M., The SILAC fly allows for accurate protein quantification in vivo. *Mol. Cell. Proteomics* 2010, *9*, 2173–2183.

[16] Rolland, N., Atteia, A., Decottignies, P., Garin, J. et al., Chlamydomonas proteomics. *Curr. Opin. Microbiol.* 2009, *12*, 285–291.

[17] Schrimpf, S. P., Hengartner, M. O., A worm rich in protein: quantitative, differential, and global proteomics in *Caenorhabditis elegans*. *J. Proteomics* 2010, *73*, 2186–2197.

[18] Malmstrom, J., Beck, M., Schmidt, A., Lange, V. et al., Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* 2009, *460*, 762–765.

[19] Denef, V. J., Shah, M. B., Verberkmoes, N. C., Hettich, R. L., Banfield, J. F., Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis. *J. Proteome Res.* 2007, *6*, 3152–3161.

[20] Weckwerth, W., Baginsky, S., van Wijk, K., Heazlewood, J. L., Millar, H., The multinational Arabidopsis steering subcommittee for proteomics assembles the largest proteome database resource for plant systems biology. *J. Proteome Res.* 2008, *7*, 4209–4210.

[21] 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 Phys.* 2010.

[22] Yamamoto, M. T., Drosophila genetic resource and stock center; The National BioResource Project. *Exp. Anim.* 2010, *59*, 125–138.

[23] Bodenmiller, B., Campbell, D., Gerrits, B., Lam, H. et al., PhosphoPep – a database of protein phosphorylation sites in model organisms. *Nat. Biotechnol.* 2008, *26*, 1339–1340.

[24] Agrawal, G. K., Job, D., Zivy, M., Agrawal, V. P. et al., Time to articulate a vision for the future of plant proteomics – a global perspective: an initiative for establishing the International Plant Proteomics Organization (INPPO). *Proteomics* 2011, *11*, 1559–1568.

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

[26] Vizcaino, J. A., Cote, R., Reisinger, F., Barsnes, H. et al., The proteomics identifications database: 2010 update. *Nucleic Acids Res.* 2010, *38*, D736–742.

[27] Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I. et al., The PeptideAtlas project. *Nucleic Acids Res.* 2006, *34*, D655–658.

[28] Qeli, E., Ahrens, C. H., PeptideClassifier for protein inference and targeted quantitative proteomics. *Nat. Biotechnol.* 2010, *28*, 647–650.

