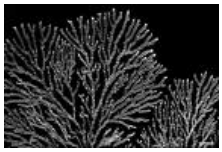
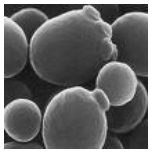
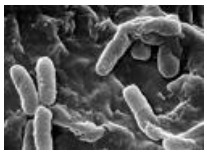
Initiative on Model Organism Proteomes (iMOP)







iMOP initiative committee

Zurich, Switzerland, July 2010

Submitted to the HUPO Executive Committee on the 27th of July 2010

Executive summary

The model organism community is growing steadily and the number of comprehensive model organism proteome data being generated is continuously increasing. To standardize efforts and to make optimum use of the proteomics data acquired in model organisms, a new HUPO initiative in model organism proteomes (iMOP) is proposed. iMOP will stimulate the scientific exchange within this community, and HUPO best practices and standards can be brought to the model organism researchers. Their needs towards the central databases will be better represented, catalyzing the integration of proteomics and organism-specific databases.

1. Vision and mission

a. Model organism proteomes

Model organisms have been used for decades to study embryogenesis, development and other fundamental biological processes. They are easy to grow in the laboratory and have rather short generation times. The conservation of disease pathways between model organisms and human allows model organisms to be used to study disease pathogenesis and to develop drugs. Many model organisms are important commensals or pathogens and thus very closely interacting with the human proteome. The conservation between model organisms and crop plants is leading to improved crop performance.

Historically, research on model organisms depended largely on genetic approaches. More recently, however, large-scale approaches such as genomics and proteomics have become an essential part of model organism research. Unfortunately, the community of model organism researchers is highly fragmented, first because they work on many different model organism species, and second because of their wide distribution around the world. Often, groups that start proteomic projects in model organisms have a previous expertise in genomics but are new in the field of proteomics. Moreover, many of these groups face similar technical challenges in their efforts to characterize the proteome of their favorite model organism: small size, difficulties in obtaining pure cell populations, integration of proteomics data into already well-established organism-centered databases.

In order to better address these challenges, we propose to establish a new HUPO initiative on model organism proteomes (iMOP) 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 are also applied in the future to all model organisms. The iMOP initiative will not focus on a

single species but rather will bring together model organism proteomicists to reach the critical mass necessary to get influence and impact within the scientific community.

b. Scientific objectives

iMOP will pursue the following objectives:

- **Integrate different model organism research groups into a model organism proteomics community and promote interaction between them**

The vision of iMOP is to join groups covering a broad spectrum of model organisms and methods into a common model organism proteomics community. iMOP will facilitate the combination of resources and efforts of different labs. Scientific interactions within iMOP will be encouraged through the exchange of visiting researchers between the labs, as well as workshops during the regular HUPO meetings and specific iMOP meetings.

- Adopt HUPO standards and best practices

In analogy to already running HUPO initiatives, the use of best practices will be established within the model organism community. Data storage and exchange using established platforms will be facilitated by following the PSI guidelines for standardized data acquisition and analysis.

- Integrate and link proteome and organism-specific databases

Several proteomics data repositories exist encompassing different species. Proteomics data are also being integrated into model organism-specific databases. To facilitate access to all proteome data, iMOP will help to efficiently link proteome-centric and organism-based databases. Such linked model organism datasets will be a valuable resource of information also for human proteome projects.

- Develop software tools to navigate databases

Since most biologists are interested in using rather than generating proteomics data, it is important that proteomics data will be easily accessible, easily manipulated and analyzed, and that different datasets will be easily compared. iMOP will help develop software tools to navigate the databases and to extract and compare sub-datasets (Figure 1). 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.

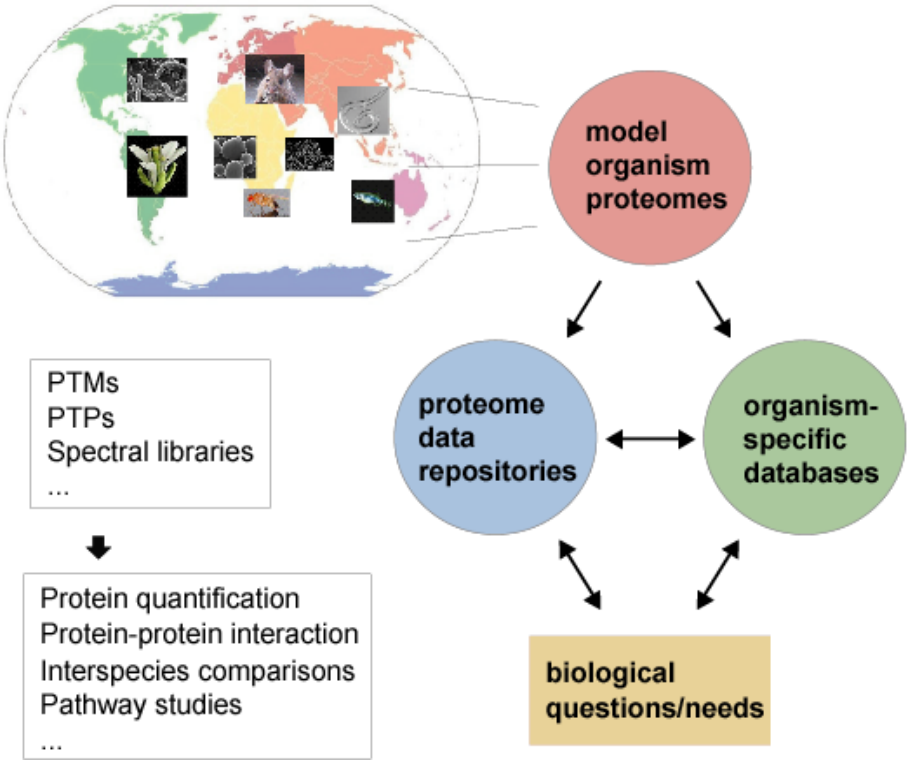


Figure 1:
Integrative function of a new HUPO initiative in model organism proteomes (iMOP). Different groups working on different model organisms all over the world contribute their mass spectrometry data. iMOP promotes the integration and linkage of proteome data into central proteome and organism-specific databases. Based on the requirements and needs of the community, tools to access specific biological data will be generated and integrated. Sub-datasets can be extracted to address a variety of biological questions and problems that are also applicable to human research projects.

2. Methodology

a. Action plan and phases

Phase I – project planning

iMOP ambassadors (see section 4) will be recruited who help to develop the iMOP program and to promote iMOP within their respective model organism communities. The ambassadors will agree upon a common procedure of data storage and data exchange and ensure that projects are run according to PSI standards. They will establish a network which includes model organism-specific database curators and proteome database curators in order to link proteomics-specific and model organism-specific databases.

Phase II - research

The needs of the model organism community towards a model organism proteome database will be identified. Based on these needs, plug-ins and user-friendly navigation tools to extract the required information will be generated. In a later research phase, community projects can be started focusing on comprehensive targeted analyses of specific pathways in different model organisms.

Phase III – commercialization outcomes

We do not expect any specific commercial applications to be spun out of this initiative, as we plan to make both data and analysis tools freely available. We do expect however that this work will support research and drug development within various biotech, pharmaceutical, and agricultural companies.

b. Timelines

12 months milestones and deliverables

Milestones:

- ambassadors recruited
- community needs regarding proteome databases identified
- first workshop held

Deliverables:

- agreement of collaboration signed
- funding agencies identified
- grant applications submitted
- iMOP website set up
- first workshop report

24 months milestones and deliverables

Milestones:

- database navigation tools developed
- annual meetings and workshop established

Deliverables:

- funding and financing secured
- workshop/progress reports

36 months milestones and deliverables

Milestones:

- community projects – e.g. focus on diseases, specific pathways or posttranslational modifications

Deliverables:

- research papers
- workshop/progress reports

c. Phases

see section 2 a

d. What we propose in terms of scientific / peer review

Annual reports will be submitted to all members and to HUPO leadership.

3. Leadership

Michael Hengartner (University of Zurich) will initially chair for a two-year period. Klaas van Wijk (Cornell University), Steve Briggs (University of California), and Andreas Tholey (University of Kiel) will be co-chairs. The chair and co-chairs are expected to rotate on a regular basis and will be elected by all initiative members. Sabine Schrimpf (University of Zurich) will co-ordinate the initiative.

4. Co-applicants and their involvement

Several international co-applicants working on different model organisms have been identified during the workshop at the 7th HUPO meeting in Amsterdam:

Pierre-Alain Binz (Genebio, Geneva, Switzerland; pierre-alain.binz@genebio.com)

Data analysis standards

Steve Briggs (University of California, La Jolla, U.S.A.; sbriggs@ad.ucsd.edu)

Plant proteomics/cell adaptation to environmental changes

Michael Hecker, Dörte Becher (University of Greifswald, Germany; Hecker@uni-greifswald.de; dbecher@uni-greifswald.de)

Unicellular organisms/prokaryotes (Bacillus subtilis and related gram-positive pathogens e.g. Staphylococcus aureus)

Mike MacCoss (University of Washington, U.S.A.; maccoss@uw.edu)

C. elegans genome annotation

Andreas Tholey (Christian-Albrechts-University Kiel, Germany; a.tholey@iem.uni-kiel.de)

Techniques for proteome analysis, posttranslational modifications, host-pathogen and host-microbiome interactions

Uwe Völker (University of Greifswald, Germany; voelker@uni-greifswald.de)

Proteomics of prokaryotic model organisms and *in vivo* proteomics of bacterial pathogens

Ruedi Aebersold (ETH Zurich, Switzerland; aebersold@imsb.biol.ethz.ch)

Quantitative proteomics and tools for the central data repository

Erich Brunner (Q-MOP, University of Zurich, Switzerland, erich.brunner@imls.uzh.ch)

Drosophila melanogaster proteome

Christian Ahrens (Q-MOP, University of Zurich, Switzerland; christian.ahrens@imls.uzh.ch)

Bioinformatics, PTP prediction, and data analysis and integration

Günter Lochnit (Justus-Liebig University, Giessen, Germany;
Guenter.Lochnit@biochemie.med.uni-giessen.de)

C. elegans RNAi and protein analysis technologies

Michael Hengartner, Sabine Schrimpf (University of Zurich, Switzerland;
michael.hengartner@imls.uzh.ch, sabine.schrimpf@imls.uzh.ch)

C. elegans proteome analyses and inter-species comparative studies

Christian von Mering (University of Zurich, Switzerland; mering@imls.uzh.ch)

Bioinformatics and Systems Biology

Rolf Apweiler (EBI Hinxton, UK; apweiler@ebi.ac.uk)

PRIDE (MS database)

Henning Hermjakob (EBI Hinxton, UK; hhe@ebi.ac.uk)
PRIDE (MS database), IntAct (Molecular Interactions Database), and PSI

Lennart Martens (Ghent University and VIB, Belgium; lennart.martens@UGent.be)
Computational omics and Systems Biology

Alexey Nesvizhskii (University of Michigan, Ann Arbor, U.S.A.; nesvi@med.umich.edu)
Databases and statistical applications

Nitin Baliga (Institute for Systems Biology, Seattle, U.S.A.; nbaliga@systemsbiology.org)
Prokaryotic organisms and Systems Biology

Multinational Arabidopsis Steering Committee (MASCP):
Klaas van Wijk (Cornell University, Ithaca, U.S.A.; kv35@cornell.edu)
Chloroplast proteomics and biology
Alex Jones (The Sainsbury Laboratory, Norwich, UK; Alex.jones@tsl.ac.uk)
Katja Bärenfaller (ETH Zurich, Switzerland; katja.baerenfaller@ipw.biol.ethz.ch)

Yhong-Hee Shim (Konkuk University, Seoul, Korea; yshim@konkuk.ac.kr)
C. elegans proteome and germline proteome

5. Data sharing plan

iMOP will act as a catalyst to facilitate linking and integration of already established databases. It will not become another data repository itself. A meaningful sharing of proteome data requires a file format and consistent way for dealing with ambiguities. For sharing raw and XML data (minimally processed), Tranche could be the repository of choice. For sharing processed data (peptide and protein lists, quantification), Pride, GPMDB, and PeptideAtlas would be suitable. In the long run, a centralized data analysis will become standard. Software tools like those already established in BioMart (query-oriented data management system) to navigate model organism proteome data, and pipelines for the electronic transfer of large data files will be developed.

6. Budget requirements

The initial activities will be financed within running projects. Funding for the following activities has to be acquired:

- workshop during HUPO meetings — N/A
- specific iMOP meetings — up to 50.000 EUR
- training activities — up to 20.000 EUR
- initiative coordination, provided by home institution of chair and vice-chairs — up to 50.000 EUR / year
- software development – salaries for programmers — ~100.000 EUR / year

7. Fund-raising plan

A major advantage of having iMOP as an official HUPO initiative is that it facilitates the acquisition of additional funding. Salaries for programmers will be self-supported via 3rd party grants. Workshops and initiative meetings will be supported via funding programs such as COST. To finance training of young scientists, we will apply for an EU research training network grant. Coordinated research activities will be supported via the network's normal funding bodies like the human frontier program and EU-FP7 projects.

8. Interaction with other HUPO initiatives

There will be close interactions with PSI (www.psidev.info), and iMOP will serve as a link to other human proteome projects.

9. Education and training

We intend to set up a new EU research training network to transfer know-how between model organism and proteomics research groups. iMOP will encourage its members to attend training workshops at HUPO meetings and will help its members to offer proteomics training workshops at organism-specific meetings.

10. Workshops

Workshops during regular HUPO meetings and initiative specific meetings will be organized. If possible, the start-off of the new initiative will be discussed during the HUPO meeting in Sydney. Several of the co-applicants will participate.

11. Industrial partners

JPT peptide technologies, Berlin, Germany (Holger Wenschuh; wenschuh@jpt.com)

12. Ethics

We do not expect any ethical issues regarding this project, as it has mainly a networking and coordination function. iMOP encourages its members to generate data according to established standards.