EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Exploratory Workshop Scheme

Standing Committee for Life, Earth and Environmental Sciences (LESC)

ESF Exploratory Workshop on

Model Organism Proteomics

Zurich, Switzerland, 11-13 April 2007

Scientific Report

Co-sponsored by:

C-MOP
Center for
Model Organism Proteomes

University of Zurich

SystemsX.ch
The Swiss Initiative in Systems Biology

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Convenor:

Professor Dr. Ruedi Aebersold
Swiss Federal Institute of Technology
ETH-Hönggerberg
Institute for Molecular Systems Biology
Wolfgang-Pauli-Str. 16
CH-8093 Zürich
Phone: +41-44-633 3170
E-mail: rudolf.aebersold@imsb.biol.ethz.ch

Co-Convenors:

Dr. Erich Brunner
Center for Model Organism Proteomes
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich
Phone: +41-44-635 3126
E-mail: erich.brunner@molbio.uzh.ch

Dr. Sabine Schrimpf
Institute of Molecular Biology
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich
Phone: +41-44-635 3112
E-mail: sabine.schrimpf
@molbio.uzh.ch

1. Executive summary

Systems biology is an integrative, inter-disciplinary approach of diverse areas including biology, physics, and the computational sciences and is thus collaborative by nature. Operationally systems biology is concerned with the comprehensive and quantitative (dynamic) analysis of networks of interacting the elements that constitute the living cell. As essentially all cellular processes and most molecular networks consist of or contain proteins, the systematic analysis of the proteins expressed by a cell – the analysis of the proteome- is of particular importance for systems biology. The ESF exploratory workshop (11-13 April, 2007) held in Zurich focused on “Model Organism Proteomics”, a sub discipline within systems biology. To achieve such a challenging goal as the proteome analysis of a whole organism – which has not been accomplished so far – sophisticated techniques and complex and costly infrastructure are required. In particular, the high throughput processing of samples as well as data acquisition, storage, evaluation and dissemination require a close collaboration with researchers in different fields.

To compile and discuss the currently available bioinformatics and biochemical techniques, to explore recent developments and to plan (future) activities in the field, we were able to gather 26 European researchers from 9 countries specialized in the different disciplines of systems biology and proteomics at the University of Zurich. In a two-day meeting, the participants discussed the current status of the field and planned the formulation of future research proposals related to model organism proteomics.

The focus of the workshop was on three topics:

i) proteome annotation of model organisms – what are the available and most promising techniques for whole proteome annotation
ii) the use of such proteome catalogs – for genome annotation and technology development
iii) databases/data mining – databases to store and mine such information

For each of these topics three specific presentations built the basis for the discussion groups following in the afternoon. The respective experts in the filed explored the topics in round table discussions.

The second day, representatives of the individual discussion groups presented the outcome, followed by a lively discussion of all the participants together, with the aim to outline a future application to the EC framework program FP 7.

The following statements resulted from these discussions:

i) The project the participants agreed upon will be “enabling time resolved deep proteomics”. The core technique will be proteotypic peptides (PTPs), a concept that has been pioneered by the SystemsX.ch glue project Model Organism Proteomes. PTPs are tryptic peptides, which are most likely identified applying the current mass spectrometric methods, and which are unique for a specific protein. Sophisticated predictions programs are available to define PTPs (considering their physicochemical properties) for all proteins of all organisms. These peptides can be synthesized chemically and spiked to the samples for quantitative comparisons. This implies the generation of PTP libraries with *in silico* and real compounds like peptides and phosphopeptides for different model organisms (*D. melanogaster, C. elegans, A. thaliana, S. cerevisiae, microbes, mus musculus,* and eventually *homo sapiens*).

ii) Around the core technique there was on one side the input placed comprising separation science and mass spectrometry development for an accurate quantitative measurement and kinetics, and the generation of complete proteome maps. On the other side were the applications and biological questions based on the expertise of the workshop participants like e.g. growth signaling and biomarker discovery.

2. Scientific content

2.1 Talks

After the welcome by Ruedi Aebersold, the ESF representative Constantinos Doukas gave an overview of the objectives and funding possibilities provided by the ESF.
The following scientific presentations were split into the three different sessions

- proteome annotation
- databases / data mining

- use of proteome maps.

In the first session Lukas Huber reported on techniques how to study the EGFR-Ras-MAPKinase growth signaling pathway in subcellular compartments. He described the application of chemical genetics and the enrichment of phosphopeptides. Jeroen Krijgsveld introduced labeling and non-labeling methods for the identification and quantification of orthologous stem cell proteins and of certain classes of proteins in different model organisms. Kris Gevaert talked about COFRADIC, a method for isolation of tagged N-termini of proteins, and about the use of SILAC to study protein processing in Jurkat cells.

The first speaker of the second session, Andrey Alexeyenko, introduced FunCoup, a statistical framework of data integration for finding functional coupling between proteins (http://funcoup.sbc.su.se). Data of different sources are collected and probabilistically evaluated in a Bayesian network. Gos Micklem explained InterMine and FlyMine (www.flymine.org), an integrated database of genomic, expression and protein data for *D. melanogaster*, *Anopheles* and *C. elegans*. Intermine is a general-purpose object-oriented data warehouse system developed as part of the FlyMine project and made available as stand-alone open-source software (www.intermine.org). Rolf Apweiler presented PSI standards (proteomics standards initiative) for mass spectrometry and molecular interactions: IntAct (www.ebi.ac.uk/intact/), and IMEX (://imex.sf.net). He further pointed out that PSI is an open community initiative and that there is a call for participation (www.psidev.info).

The third session was opened by Karl Mechtler, who talked about the usage of iTRAQ stable isotope labeling and multiple reaction monitoring (MRM) mass spectrometry for protein quantification. He described extended LC-MS/MS runs and software tools, which reduce the number of single peptide identifications. Anja Persson presented the human protein atlas project (www.proteinatlas.org). She described the strategies for generation of antibodies by PrEST (protein epitope signature tags). To date 3574 antibodies were approved and annotated. Erich Brunner gave an overview of the “Center for Model Organism Proteomes” (www.mop.uzh.ch), which is a glue project of SystemsX.ch, the Swiss initiative in

Systems Biology (www.systemsx.ch). Currently the proteomes of 3 different model organisms (*D. melanogaster, C. elegans, A. thaliana*) are being analyzed in collaboration with the Functional Genomics Center Zurich (www.fgcz.ethz.ch).

2.2 Discussions

It turned out that the discussions in all three groups followed similar paths. The two fundamental topics were biological application and technology development.

- The discussion of the “proteome annotation” group put an emphasis on the time aspect. Biological problems should not be solved by analysis of a complete proteome at a static time point, but by quantification of changes in a subproteome over time.
- The “databases / data mining” group concluded that PTP should be the core technology and the long-term goals to study humans. The resources like the synthesized peptides, good PTP prediction algorithms, and the data standardized pipelines should be made available to the scientific community.
- The group “use of proteome maps” emphasized the importance of the analysis of complete proteomes for genome annotation. The available human antibodies, which were designed by homology to other species, could be used to check conserved disease genes.

These issues were deepened further during the discussion session with all participants together:

To achieve the identification of an entire proteome of a model organism, methods beyond the current shotgun approach have to be established. To gain information on proteins, which have not been seen yet (e.g. low abundant ones), the idea was to make use of PTPs. Especially in combination with accurate mass tags (AMT) PTPs could become a new milestone in proteome analysis. For that aim, the PTPs have to be reliably predicted and libraries of synthetically synthesized peptides have to be established to spike samples for absolute quantitative analyses. In general, the test system should be triggerable, dynamic, and perturbable to detect differences. Model organisms are used to develop the technology to finally tackle higher organisms. Other data types like gene expression studies, metabolomics, and monitoring of phosphorylation events complement the proteomics data. The analysis of different model organisms and systems opens up the focus on the evolutionary aspect by revealing conservation on protein level and thus functional annotations. The

availability of antibodies allows the combination of imaging and proteomics. Long-term goals are to establish interaction maps and modeling projects.

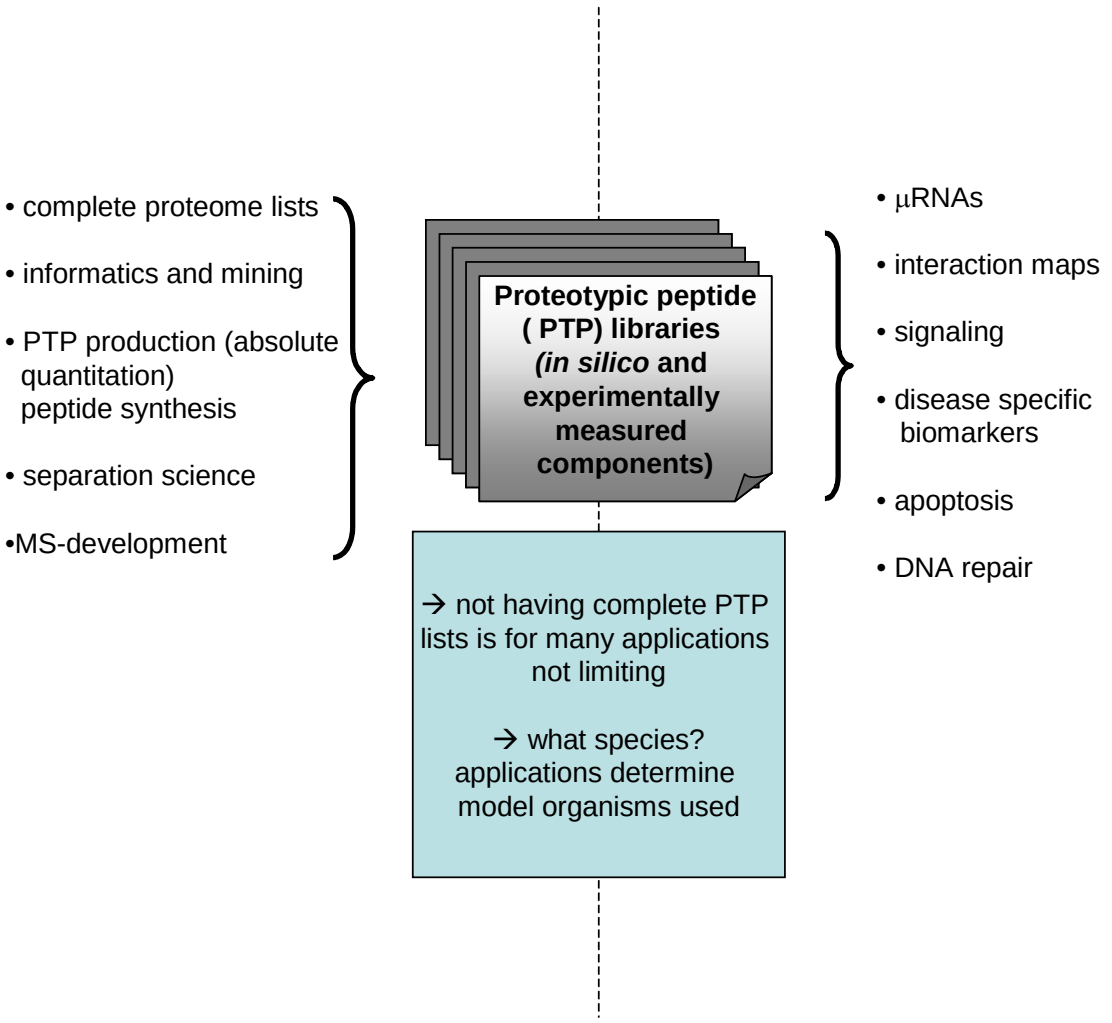
2.3 Application outline

The application outline was based on the strengths of the scientists gathered for this workshop: model organisms, protein identification technologies, and disease-related pathways. As a working title for the application was chosen:

"Enabling time resolved deep proteomics"

PTP libraries and *in silico* compounds like peptides and phosphopeptides should form the core of this project. The PTPs will be used for different organisms, like *D. melanogaster, C. elegans, A. thaliana, M. musculus*, and eventually *H. sapiens*. The core technology is surrounded by the input on one side and the biological applications on the other side. Components of the input side are technology development for separation science and accurate quantitative MS measurement. This requires the establishment of complete proteome maps and comparison to transcriptomics data, and informatics efforts for data mining. For chemical production and distribution of the PTPs as well as with respect to instrumentation and software, collaborations with the biotech industry could be established.

The potential biological applications were based on the expertise of the workshop participants. An important aspect mentioned was to put the emphasis on following changes over time and kinetics. Biological applications could be miRNA, analysis of the phosphoproteome, signaling mechanisms, DNA damage and repair, apoptosis and disease marker discovery. Additional fields mentioned were the identification of protein complexes and networks, protein-DNA interactions and membrane proteins. Possible extensions could be microbes and host-pathogen interactions. However, it should be noted that the resources and technologies generated by the planned project will be generically applicable to a wide range of research questions in a multitude of species.



3. Contribution to the future direction of the field

The collaborative project will be split into biological and technical aspects. The strength and novelty of such an application will be that the participants will share their specific knowledge centered around the PTP technology. The benefits of such a consortium for the analysis of different model organisms are e.g. the synergies to study evolutionary aspects and to integrate the acquired data into existing data repositories. It makes proteomics a powerful tool and facilitates its penetration into clinical studies and disease related questions. Such a network of collaborations certainly strengthens the influence of the European research community in the field of systems biology and proteomics worldwide.

4. Final programme

Wednesday, 11 April 2007
Arrival of participants

Thursday, 12 April 2007

8.30 – 8.35:	Welcome by **Ruedi Aebersold**
8:35 – 9:00:	Introduction by **Constantinos Doukas** - ESF representative

Session “proteome annotation”

9.00 – 9.30:	**Lukas Huber** Organelle- and phosphoproteomics of EFG-receptor/MAP kinase signalling
9.30 – 10:00:	**Jeroen Krijgsveld** Qualitative and quantitative proteomics in Drosophila and other model organisms
10:00 – 10.30:	**Kris Gevaert** Targeted peptide-centric proteomics by COFRADIC

Coffee break

Session “databases / data mining”

10.45 – 11.15:	**Andrey Alexeyenko** FunCoup: networks of functional coupling as a large-scale data integration project in eukaryotes
11.15 – 11.45:	**Gos Micklem** FlyMine and InterMine: data and tool integration in a flexible query environment
11.45 – 12.15:	**Rolf Apweiler** Development and implementation of proteomics standards and services

Lunch

Session “use of proteome maps”

13.15 – 13.45:	**Karl Mechtler** Quantitative transgenic mouse proteomics
13.45 – 14.15:	**Anja Persson** Applications of the human protein atlas project
14.15 – 14.45:	**Erich Brunner** Center for Model Organism Proteomes

Thursday afternoon – 18.30:
Discussion groups

Proteome annotation
Discussion leader: Sabine Schrimpf

Databases / data mining
Discussion leader: Ruedi Aebersold

Use of proteome maps
Discussion leader: Erich Brunner

Friday, 13 April 2007

8.30 – 12.00:

Presentation/discussion of the results of the 3 discussion groups

Lunch

13.00 – 16.00:

Conclusion and application outline

End of workshop at 16 pm.

5. Final list of participants

Professor Dr. Alfred Nordheim
Institute for Cell Biology,
Department of Molecular Biology
University of Tuebingen
Auf der Morgensphonele 15
D-72076 Tuebingen
Germany
Phone: +49-7071 297 8898
Fax: +49-7071 29 53 59
Email: alfred.nordheim@uni-tuebingen.de

Professor Dr. Arne Elofsson
Center for Biomembrane Research and
Stockholm Bioinformatics Center
Dep. of Biochemistry and Biophysics
University of Stockholm
Svante Arrhenius väg 12
S-10691 Stockholm
Sweden
Phone: +46-(0)8-16 4672
Fax: +46-(0)8-15 3679
Email: arne@bioinfo.se

Dr. Anja Persson
KTH-Royal Institute for Technology
Department of Biotechnology
Alba Nova University Center
Roslagstullsbacken 21
S-10691 Stockholm
Sweden
Phone: +46-8-55378017
E-mail: anja@biotech.kth.se

Dr. Karl Mechtler
IMP - Research Institute of Molecular Pathology
Dr. Bohr-Gasse 7
A-1030 Wien
Austria
Phone: +43 (1) 79044 4280
Fax: +43 (1) 7987153
Email: mechtler@imp.univie.ac.at

Professor Dr. Christian von Mering
Institute of Molecular Biology
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland
Phone: +41-44-635 3147
E-mail: mering@molbio.unizh.ch

Professor Dr. Benno Schwikowski
Institut Pasteur
25-28 rue du Dr. Roux
F-75015 Paris
France
Phone: +33 1 4568-8620
Fax: +33 1 4061-3704
E-mail: benno@pasteur.fr

Professor Dr. Peter James
Lund University
Protein chemistry
Sölvegatan 33a
S-221 84 Lund
Sweden
Phone: +46 46 222 1496
E-mail: peter.james@elmat.lth.se

Dr. Adrian Honegger
ETH Zurich
Wolfgang-Pauli-Strasse 16
HPT C 74
CH-8093 Zurich
Switzerland
Phone: +41 44 633 36 94
E-mail: adrian.honegger@systemsx.ch

Dr. Rolf Apweiler
EMBL Outstation - Hinxton
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SD
United Kingdom
Phone: +44 (0)1223 494 435
Fax: +44 (0)1223 494 468
E-mail: apweiler@ebi.ac.uk

Dr. Matthias Wilm
EMBL Heidelberg
Bioanalytical Research Group
Meyerhofstrasse 1
D-69 117 Heidelberg
Germany
Phone: +49 (0)6221 387 8560
Fax: +49 (0)6221 387 8306
E-mail: wilm@embl.de

Professor Dr. Kris Gevaert
Department of Medical Protein Research, Proteome Analysis and Bioinformatics Unit
Flanders Interuniversity Institute for Biotechnology
Faculty of Medicine and Health Sciences
Ghent University
A. Baertsoenkaai 3
B-9000 Gent
Belgium
Phone: +32-92649274
Fax: +32-92649496
E-mail: kris.gevaert@ugent.be

Dr. Henning Hermjakob
EMBL Outstation - Hinxton
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 1SD
United Kingdom
Phone: + 44 1223 49 4671
Fax: + 44 1223 49 4468
Email: hhe@ebi.ac.uk

Professor Dr. Jeroen Krijgsveld
Department of Biomolecular Mass Spectrometry
Utrecht University
Sorbonnelaan 16
NL-3584 CA Utrecht
The Netherlands
Phone: +31 30 253 9974
Fax: +31 30 251 8219
E-mail: j.krijgsveld@uu.nl

Professor Dr. Manuel Peitsch
Novartis Pharma AG
WKL-490.4.04
CH-4002 Basel
Switzerland
E-mail:
manuel.peitsch@pharma.novartis.com

Professor Dr. Nils Faergeman
University of Southern Denmark
Department of Biochemistry and Molecular Biology
Campusvej 55
DK-5230 Odense M
Denmark
Phone +45 6550 2453
E-mail: nils.f@bmb.sdu.dk

Dr. Gos Micklem
University of Cambridge
Department of Genetics/ Department of Applied Maths
Downing Street
Cambridge, CB2 3EH
United Kingdom
Phone: +44 (0) 1223 760 256
Fax : +44 (0) 1223 333 992
E-mail: g.micklem@gen.cam.ac.uk

Dr. Christian Ahrens
Functional Genomics Center
UNI ETH Zurich
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland
Phone: +41 44 635 39 30
Fax: +41 44 635 39 22
E-Mail: christian.ahrens@fgcz.ethz.ch

Professor Dr. Lukas Huber
Innsbruck Biocenter
Fritz-Pregl-Straße 3
A-6020 Innsbruck
Austria
Tel +43/512/9003-70171
Fax +43/512/9003-73100
E-mail: lukas.a.huber@i-med.ac.at

Dr. Andrey Alexeyenko
Stockholm Bioinformatics Center, Albanova
Stockholm University
Roslagstullsbacken 21
S-10691 Stockholm
Sweden
Phone: +46-8 55378586
Fax: +46-8 5537 8214
E-mail: andale@sbc.su.se

Dr. Sacha Baginsky
Institute of Plant Sciences
ETH Zurich
Universitätstrasse 2
CH-8092 Zurich
Switzerland
Phone: +41 44 632 38 66
E-mail:
sacha.baginsky@ipw.biol.ethz.ch

Dr. Jérôme Garin
CEA/Grenoble
iRTSV/EDyP (bat C3)
17 rue des martyrs
F-38054 Grenoble cedex 9
France
Phone: +33-4 38 78 96 56
Fax: +33-4 38 78 50 51
E-mail: jerome.garin@cea.fr

Dr. Marianne Tardif
CEA/Grenoble
iRTSV/EDyP (bat C3)
17 rue des martyrs
F-38054 Grenoble cedex 9
France
Phone: +33 4 38 78 11 14
Fax: +33 4 38 78 50 51
E-mail: Marianne.tardif@cea.fr

Dr. Bodo Lange
Max Planck Institute for Molecular Genetics
Ihnestraße 63-73
D-14195 Berlin
Germany
Phone: +49-30 8413 1645
Fax: +49-30 8413 1380
Email: lange_b@molgen.mpg.de

6. Statistical information of participants

26 participants (plus 1 ESF representative)
23 males
3 females

Country	Number of participants
Switzerland	8 (including 3 convenors)
United Kingdom	3
Austria	2
Belgium	1
Netherlands	1
Denmark	1
France	3
Sweden	4
Germany	3