

MASC Proteomics- 2007 Subcommittee Report

Prepared by Wolfram Weckwerth (Chair, weckwerth@mpimpgolm.mpg.de), Harvey Millar (Co-chair, harvey.millar@uwa.edu.au), Sacha Baginsky (Co-chair, sbaginsky@ethz.ch)

Proteomics using *Arabidopsis thaliana* as a model system has made great progress in recent years. The throughput and accuracy of protein identification techniques relies strongly on the availability of whole genome sequences. The availability of both a complete genome sequence and high quality genome annotation makes *Arabidopsis* an ideal system to develop new proteomics technologies and identify candidate genes for development and differentiation of plant cells. Furthermore, *Arabidopsis* proteomics research may serve as a paradigm for other plant proteomics research projects and the advancement of plant proteomics in general. And finally, *Arabidopsis* proteomics enables the development of a unique platform for plant systems biology by integrating proteomics databases with existing transcriptomics and novel metabolomics and bioinformatics initiatives.

Due to the enormous complexity of a dynamic proteome, different approaches have to be combined to measure protein expression and dynamics, stress- and developmental responses, posttranslational protein modifications and protein interaction. Although the *A. thaliana* functional proteomics projects are progressing well in general, technical obstacles still remain to be overcome. The development of proteomics techniques for plant biology applications is showing strong progress, and some of these are probably groundbreaking methods that will be generally applicable to work in other model species. However, proteomics has not yet reached its goal. We are still in the process of initial data generation, and in the face of the enormous complexity of the dynamic proteome, substantial method development is still needed, especially in the fields of quantitative proteomics and the dissection of signalling pathways involving post-translational modification. Therefore, this working group devoted to *Arabidopsis* proteomics aims to combine the efforts of different research groups to develop programs which will consolidate databases, technique standards and experimentally validated candidate genes and functions.

The subcommittee plans to build a continuing presence at the International Conference on *Arabidopsis* Research (ICAR) in the form of a workshop. This would aim to discuss methods/practice of proteomics, to inform the community, make new contacts, and to develop initiatives for new activities of the subcommittee. In the future symposia with invited speakers will be organised where proteome analysis in *Arabidopsis* is a key theme. The first proteomics workshop took place at the 2004 Berlin meeting before the subcommittee was formed. The next workshop is planned for the 2007 ICAR in Beijing, and then on a regular basis.

Subcommittee activities in 2006/2007:

We have developed a proteomics subcommittee webpage that includes links to existing databases of *Arabidopsis* proteomics. The goal of this webpage is to disseminate standards for sample handling and data interpretation. Furthermore, we consider it a platform for the *Arabidopsis* community to exchange ideas and protocols and to spark discussions among interested researchers. Developed in the spirit to facilitate scientific networking, this webpage will foster interactions between plant proteomics researchers by providing contact information of proteomics laboratories and assembling information on meetings. Furthermore, we will crosslink the MASC efforts with a recently launched European COST initiative on Plant Proteomics using our webpage as an information exchange platform.

The website will be available at the TAIR website

(<http://www.arabidopsis.org/portals/masc/Subcommittees.jsp>)

Subcommittee members have been involved in several initiatives to develop databases and software tools for the exploitation of *Arabidopsis* proteomics data:

-A database for the subcellular location of *Arabidopsis* proteins (SUBA). The database houses large scale proteomic and GFP localisation sets from cellular compartments of *Arabidopsis*. (www.suba.bcs.uwa.edu.au).

-A repository of searchable MS/MS spectra as a protein/peptide-reference library for the *Arabidopsis thaliana* proteome integrating different levels of molecular organization including metabolites, pathways, and transcript expression (ProMex) (<http://promex.mpimp-golm.mpg.de/cgi-bin/peplib.pl>)

- Tools for the protein database-independent identification of MS/MS spectra in order to allow improving genome annotation with proteomics data

(<http://cvs.sourceforge.net/viewcvs.py/sashimi/qualscore/>)

These are only three examples and further activities and web resources will be found at the MASC proteomics website (<http://www.arabidopsis.org/portals/masc/Subcommittees.jsp>).

Subcommittee future goals:

- Establishment of a high density proteome map of *Arabidopsis thaliana* using different tissues and high-throughput mass spectrometry as a basis for future systems biology approaches
- Discuss international network grant proposals for plant proteomics.
- Develop standards for two-dimensional gel-electrophoresis, shotgun proteomics, and quantitative proteomics.
- Write guidelines for minimal requirements for different types of proteomic studies and distribute these to plant journals (especially: Plant Molecular Biology, Plant Physiology, Plant Cell, and the Plant Journal) for their consideration of publication standards – issues include experimental method, data analysis and call of identifications.
- Work towards a central international database of MS/ MS spectra derived from *Arabidopsis* samples.
- Collaborate with other interested parties on *Arabidopsis* proteomic data storage.
- Seek funding opportunities for the exchange of students and young researchers between different labs in order to develop a common training initiative

Members of the Subcommittee:

- Wolfram Weckwerth (Chair)
- Harvey Millar (Co-chair)
- Sacha Baginsky (Co-chair)
- Klaas van Wijk (Co-chair)
- Hans-Peter Braun
- Joshua Heazlewood
- Kathryn Lilley
- Hans-Peter Mock
- Sigrun Reumann
- Michel Rossignol
- Julian Whitelegge
- Jay Thelen
- Loïc Rajjou

Links for the MASC proteomics Webpage:

The primary goal is to build up a *Arabidopsis thaliana* proteomics resource. The MASC subcommittee has now assembled a list of proteomics databases from different laboratories.

SUBA <http://www.plantenergy.uwa.edu.au/applications/suba/index.php>

The SubCellular Proteomic Database (SUBA) houses large scale proteomic and GFP localisation sets from cellular compartments of Arabidopsis. It also contains precompiled bioinformatic predictions for protein subcellular localisations from a set of 10 different prediction tools. Complex relational queries can be performed between these experimental and predicted datasets to find and collate evidence for the subcellular location of Arabidopsis proteins.

PPDB

<http://ppdb.tc.cornell.edu/>

PPDB is a specialized proteome database with a focus on plant plastids. The main objective is to provide a centralized, curated, data deposit for in-house experimentally determined plastid and other proteins in *Arabidopsis thaliana* and maize (*Zea mays*), their annotated functions, as well as their experimental and predicted molecular properties and subcellular localization. Most recently, PTMs (e.g. N-terminal acetylation, deamidation, proline hydroxylation) determined by high resolution, and high accuracy mass spectrometry are also reported for *Arabidopsis* proteins. Mass spectrometry details for all in-house identified proteins are provided as much as possible (individual ion and Mowse protein scores, mass errors, etc.), with the objective to allow the db visitor to judge the significance of protein and PTM identifications. The backbone of PPDB are all gene models in the Arabidopsis nuclear and organellar genomes as annotated by [TAIR](#), as well as all maize EST assemblies (ZmGI) and maize genome sequence assemblies (AZM) as annotated by [TIGR](#). These accessions are linked to each other via directional Blast alignment. Published proteomics data set collected from the literature are uploaded and linked to the identified accessions.

ProMEX <http://promex.mpimp-golm.mpg.de/cgi-bin/peplib.pl>

ProMEX is a mass spectral library, consisting of tryptic peptide product ion spectra from protein samples of *Arabidopsis thaliana*, *Medicago truncatula*, *Chlamydomonas reinhardtii*, *Tomato* and *Potato*. The database serves as a reference and can be used for protein identification in uncharacterized samples. Protein identification by ProMEX is linked to other molecular levels of biological organization such as metabolite (<http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html>), pathway (<http://www.genome.jp/kegg/pathway.html>) and transcript data (http://csbdb.mpimp-golm.mpg.de/csbdb/dbxp/ath/ath_xpmsgq.html). The database is further connected to annotation and classification services via [BioMoby](#). ProMEX allows also text searches based on AGI code, metadata such as experimental information of the samples, and mass spectrometric instrument parameters.

ARAPEROX <http://www.araperox.uni-goettingen.de/>

This database does not only contain all Arabidopsis orthologs of known peroxisomal matrix and membrane-associated enzymes, metabolite transport proteins and the PEX proteins (peroxisome biogenesis proteins) but also a large number of novel Arabidopsis proteins that carry a putative peroxisome targeting signal type 1 (PTS1) or type 2 (PTS2) and may be targeted to plant peroxisomes. Data compilation of diverse information is intended to facilitate the identification of unknown proteins that have an important enzymatic or regulatory function in photorespiration, fatty acid β -oxidation, N and S metabolism, and in the biosynthesis or degradation of plant hormones.

SEED-PROTEOME <http://www.seed-proteome.com>

We are interested in determining the biochemical and genetic mechanisms that control seed dormancy, seed longevity, and that regulate the transition from quiescence to highly active metabolism during germination and seedling establishment. To this end, we are developing a proteome analysis of the model plant *Arabidopsis thaliana* (Chibani et al, 2006 ; Gallardo et al., 2001, 2002a,b ; Job et al., 2005 ; Rajjou et al., 2004, 2006a,b). This global proteomics approach would allow characterization of the accumulation pattern of key metabolic enzymes in dormant seeds, after-ripened non-dormant seeds, and germinating seeds.

AMPP <http://www.gartenbau.uni-hannover.de/genetik/AMPP>

An *Arabidopsis thaliana* mitochondrial proteome project was started for a comprehensive investigation of mitochondrial functions in plants. As a first step the proteome of mitochondria isolated from suspension cell cultures was characterized by high resolution 2D gel electrophoresis and mass spectrometry (Kruft et al. 2001, Plant Physiol.127, 1694-1710; Werhahn and Braun 2002, Electrophoresis 23, 640-646). Currently the proteomes of selected *Arabidopsis* knock-out mutants concerning nuclear genes of mitochondrial proteins are under investigation in our laboratory.

GABI http://gabi.rzpd.de/projects/Arabidopsis_Proteomics/

2-DE gel images and supplementary tables to the *Arabidopsis thaliana* plant tissue proteome:

1. Overview of 2DE gel images of 8 different *Arabidopsis thaliana* tissues (inflorescence, stem, primary leaf, leaf, seed, seedling, silique, root).
2. Overview and download of 2-DE gel images and result tables from MALDI-TOF MS analysed *Arabidopsis thaliana* tissues (primary leaf, seedling, silique).

NASC <http://proteomics.arabidopsis.info/>

Find within this database a set of proteomics experiments and their related data.
The experiments were made at the Cambridge Centre for Proteomics

ARAMEMNON <http://aramemnon.botanik.uni-koeln.de/>

ARAMEMNON is a database about all putative membrane proteins of Thale Cress (*Arabidopsis thaliana*), Rice (*Oryza sativa*) and about some 4800 putative membrane proteins of ~270 other seed plants.