Dynamic Biochemistry, Process Biotechnology and Molecular Biology ©2007 Global Science Books

plprot: The Plastid Proteome Database

Torsten Kleffmann[1] • Wilhelm Gruissem[2] • Sacha Baginsky[2*]

[1] Centre for Protein Research, University of Otago, Dunedin, New Zealand
[2] Institute of Plant Sciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland
Corresponding author: * **sbaginsky@ethz.ch**

ABSTRACT

Plant plastids develop and differentiate in a tissue-specific and signal-dependent manner. Each plastid type contains a distinct set of enzymes for specialized functions and metabolic activities. plprot was established as a plastid proteome database to provide information about the proteomes of chloroplasts, etioplasts, chromoplasts and the undifferentiated plastids from a tobacco BY2 cell culture and currently features 2793 protein entries. plprot furthermore provides an interactive rice etioplast proteome map that allows protein quantification and the analysis of proteome dynamics during light-induced etioplast to chloroplast conversion. plprot was designed to make all data readily accessible via a user friendly database interface and a BLAST-search module. plprot is available at http://www.plprot.ethz.ch.

THE PLASTID PROTEOME

Plastids are plant cell organelles that execute essential biosynthetic and metabolic functions including photosynthetic carbon fixation, the synthesis of amino acids, fatty acids, starch and secondary metabolites such as pigments (Neuhaus and Emes 2000; Lopez-Juez and Pyke 2005). A remarkable characteristic of plastids is their ability to develop and differentiate into specialized plastid types. Different plastid types can be distinguished by their structure, pigment composition (color) and functional differentiation. They are classified as elaioplasts in seed endosperm, chromoplasts in fruits and petals, amyloplasts in roots, etioplasts in dark-grown leaves, and chloroplasts in photosynthetically active leaf tissues (Neuhaus and Emes 2000; Lopez-Juez and Pyke 2005). All plastids carry out characteristic biosynthetic reactions in the plant cell and their energy metabolism allows distinguishing photosynthetic and non-photosynthetic plastid types. Photosynthetic chloroplasts synthesize sugar phosphates that are metabolized by the plastid oxidative metabolism to NADPH and ATP. Non-photosynthetic plastid types import sugar phosphates and ATP from the cytosol which is necessary to sustain their anabolic metabolism. This difference is often used to distinguish the energy metabolism of different plastid types as heterotrophic or autotrophic.

Similar to other cell organelles, plastids rely on the import of nucleus-encoded proteins from the cytosol in order to execute their manifold biosynthetic activities. Exact knowledge of which proteins enter different plastid types is rather scarce and current research efforts are directed towards identification of all plastid proteins, i.e. the plastid proteome. The first attempts to define the complete plastid proteome for *Arabidopsis thaliana* were based on plastid targeting prediction from the Arabidopsis genome sequences (Abdallah *et al.* 2000). More than 3600 proteins are predicted to localize to the Arabidopsis chloroplast (http://mips.gsf.de). Software-based prediction of protein targeting however is incomplete and has a limited reliability. For example, TargetP has a reported sensitivity of 85% (i.e. the rate of correct plastid prediction of true plastid proteins) and a specificity of 69% (i.e. the rate of correct prediction) (Emanuelsson *et al.* 2000). These values suggest that 15% of the true plastid proteins will escape detection, while 31% of the predicted plastid proteins are not localized to plastids. The great uncertainty of plastid targeting prediction has now shifted attention to proteomics, the systematic identification of all proteins that reside in plastids. In combination with targeting prediction tools and experimental validation, proteomics provides a unique opportunity to obtain complete and unbiased information about the complete set of plastid proteins.

Proteomics is now indispensable for plant research and several proteomics studies were conducted with different plant cell organelles (reviewed in Peck 2005; Agrawal and Rakwal 2006; Baginsky and Gruissem 2006; Rossignol *et al.* 2006). To date, most of the proteome analyses with plastids were carried out with autotrophic chloroplasts with the aim of describing the static chloroplast proteome in as much detail as possible (reviewed in van Wijk 2004; Baginsky and Gruissem 2004, 2006). The major focus on chloroplasts has two disadvantages. First, chloroplast proteome analyses suffer from the presence of highly abundant photosynthetic proteins which dominate the proteome and obscure the detection of low abundance proteins (Baginsky *et al.* 2005). This is particularly obvious in cases where the same set of proteins are repeatedly identified and reported, while the detection rate of new proteins is low (e.g. Peltier *et al.* 2006). Such analyses essentially confirm that the mass of chloroplast proteins is involved in photosynthesis (Peltier *et al.* 2006).

A second disadvantage is the lack of information about proteome dynamics during the differentiation and development of different plastid types. Proteomics is the approach of choice to analyze these processes, since e.g. transcriptional profiling provides only very limited information about the quantitative events at the protein level (e.g. Beloštotsky and Rose 2005). An additional layer of regulation that is essential for plastid function is the protein and metabolite import capacity of different plastid types that cannot be inferred from transcriptional profiling (Neuhaus and Emes 2000; Weber *et al.* 2005). Different protein targeting routes exist and comprehensive information on the regulation of protein targeting during plastid differentiation is missing (Jarvis and Robinson 2004). A promising way to circumvent both shortcomings is the proteome analysis of different plastid types and the dynamics during plastid type transition. Heterotrophic plastids for example do not contain highly abundant photosynthetic proteins allowing the detection of low abundance proteins.

Received: 27 November, 2006. Accepted: 1 January, 2007.

Techniques Paper

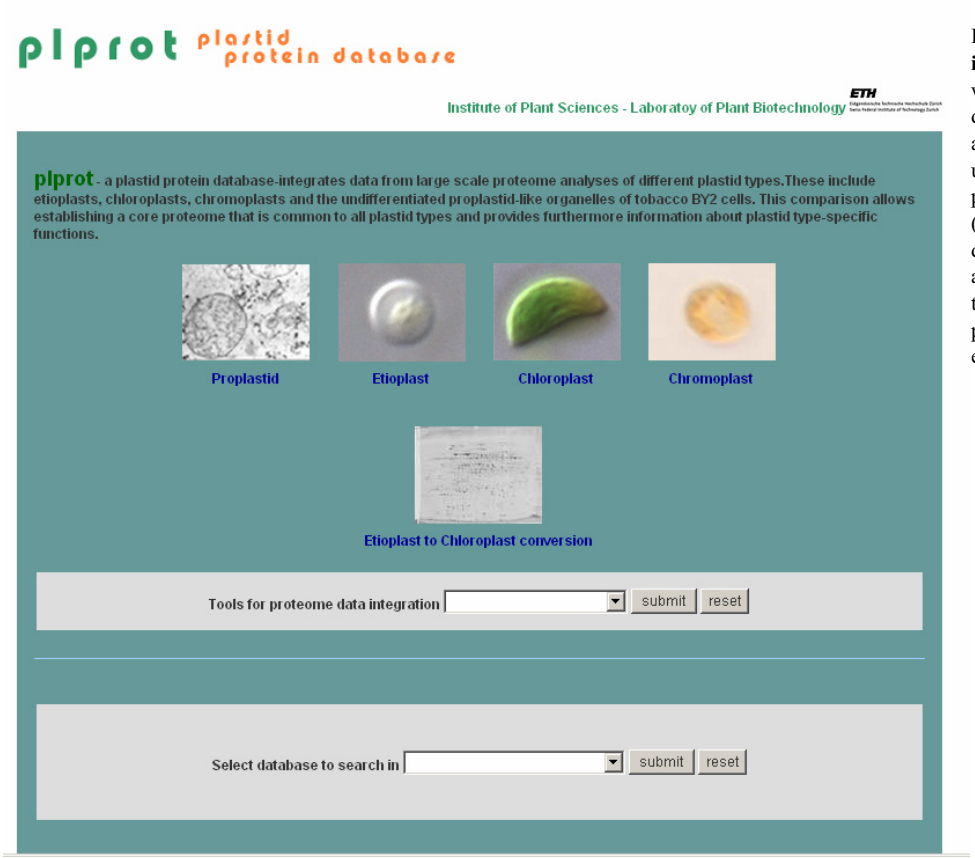


Fig. 1 Screenshot of the plprot database interface. plprot features two modules with selection menus (grey panels). Module one contains general information about the different plastid types that were used for the proteome analyses that were performed in house. The second module (lower grey panel) contains all individual databases for the different experiments and plastid types and also one database that contains all identified proteins (complete plastid proteome with currently 2793 entries).

plprot: THE PLASTID PROTEOME DATABASE

The two main concepts of plprot are to provide information about the proteomes of different plastid types and the proteome dynamics during plastid development and differentiation. plprot currently provides information about the proteome of *A. thaliana* chloroplasts, rice etioplasts, tobacco proplastids and bell pepper chromoplasts and currently features 2793 protein entries (Kleffmann *et al.* 2006; **Fig. 1**). In order to guarantee a coherently high data quality, we currently restrict the upload of data to our own results and those that have been carefully scrutinized by us. Authors who are interested to upload data into plprot are encouraged to contact the corresponding author and provide detailed information about MS-database search parameters and criteria for protein acceptance. Each protein entry in the database has a unique plprot entry code that links the protein to the organism, specific plastid type, and the publication in which the plastid protein identification was reported. The plprot-identifier has the general structure plp_xx_yyyyy. The xx-position identifies the organism (i. e. at-*Arabidopsis thaliana*, nt-*Nicotiana tabacum*, ca-*Capsicum annuum* and os-*Oryza sativa*). The five digit y-position represents the entry number. The unique plprot-identifier allows the user to compare proteome analyses of different plastid types and identify proteins that are common between datasets from different laboratories (Kleffmann *et al.* 2006).

The plprot user interface has two modules (**Fig. 1**). Module one contains a menu to select plastid type-specific information and tools for proteome data analysis. "Plastid-type information" provides a short description of each plastid type, their isolation and purification protocols, and additional information on selected proteins such as their observed solubility in the upstream protein fractionation procedure. The menu "tools for data integration" allows selecting the options "plastid type comparison" and "BLAST-search" (Altschul *et al.* 1997). The BLAST-search will identify all homologues for a protein of interest from the complete plastid dataset. A standard BLAST-search result is provided in **Fig. 2**, where we searched for homologues of *E. coli* thioredoxin reductase (Id: ZP_00724854.1; **Fig. 2**). Two homologues from chloroplasts and one from etioplasts were detected (e value< e^{-20}; **Fig. 2**). The BLAST-search result output contains all relevant information about the identified proteins, including the plprot identifier, the organism, the literature reference, and the BLAST e-value (**Fig. 2**). With this search option, the user has access to a comprehensive compilation of all plastid proteins that have been identified in different proteome studies.

We have used the BLAST-search function to cross-compare the proteomes of etioplasts, chloroplasts and BY-2 proplastids to establish the plastid core proteome that is essential for the functioning of all plastid types. This approach furthermore allowed us to detect plastid type-specific proteins that are indicative of plastid type specific functions. The results of the BLAST searches can be obtained from the "plastid type comparison" option in the first scroll-down menu (upper gray panel). The results can be downloaded from a figure of a triangle that shows the BLAST-search results in each of the fields as indicated. For example, the field CP-ET links the user to the BLAST-search results obtained from the BLAST comparison of the chloroplast dataset (CP) with the etioplast dataset (ET).

Most of the proteins that constitute the plastid core proteome are involved in carbohydrate and amino acid metabolism. Examples are glutamate synthase, cystein synthase, transketolase, phosphoglycerate kinase and glyceroldehyde 3-phosphate dehydrogenase. Amino acid and carbohydrate metabolism are housekeeping functions that are essential for the plant cell. In line with our expectations we find these functions overrepresented in the plastid core proteome. The plastid core proteome comprises furthermore a surprisingly high number of proteases suggesting that protein degradation may occur at a high rate in all plastid types, independent of the model organisms used for the analysis.

The second plprot-module holds the individual databases for the four plastid types and a merged database that contains all protein identifications that were reported from large scale plastid proteome analyses ("complete plastid", currently 2793 entries). We kept the possibility to search all databases individually because we provide specific information with all analyzed plastid proteomes.

For example, the BY2-database contains information

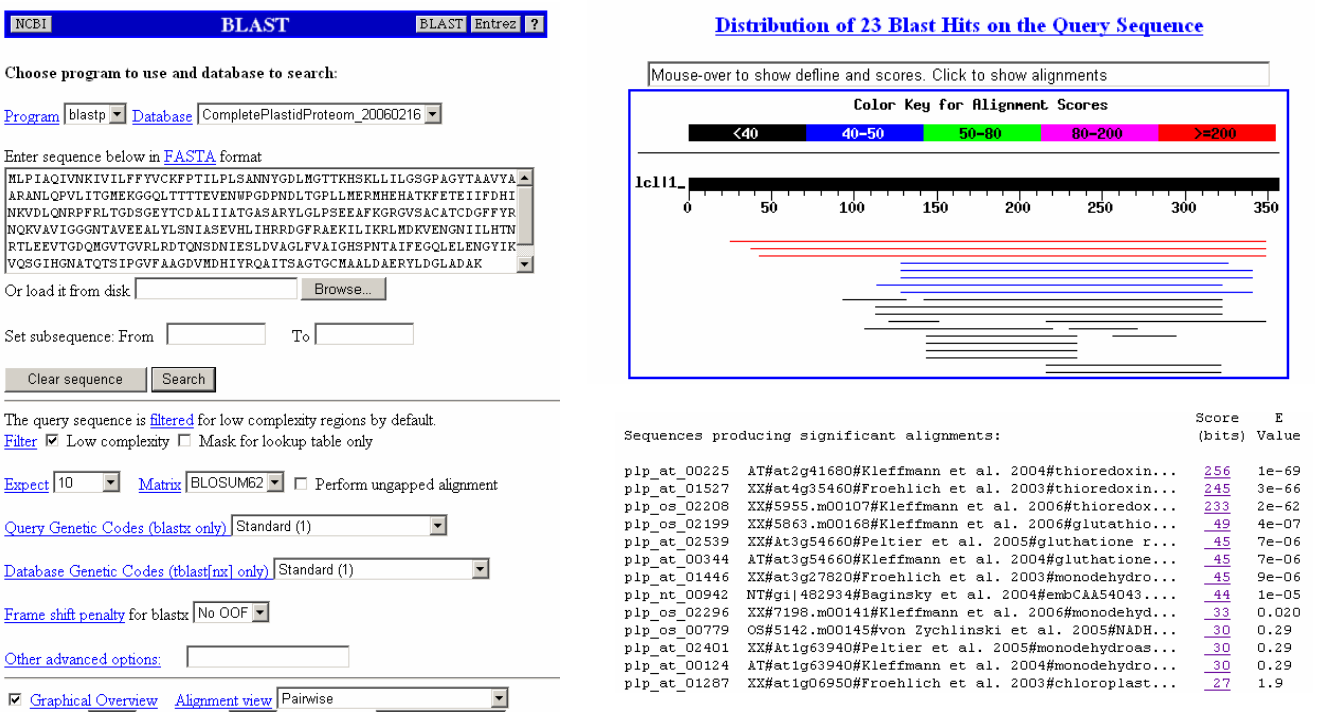


Fig. 2 Screenshot of the BLAST search interface (left panel) and a BLAST search result (right panel). In this example, the sequence of thioredoxin reductase was searched against the complete plastid proteome database. The result (right panel) provides information about all identified homologues that were identified in the different plastid types and from different laboratories.

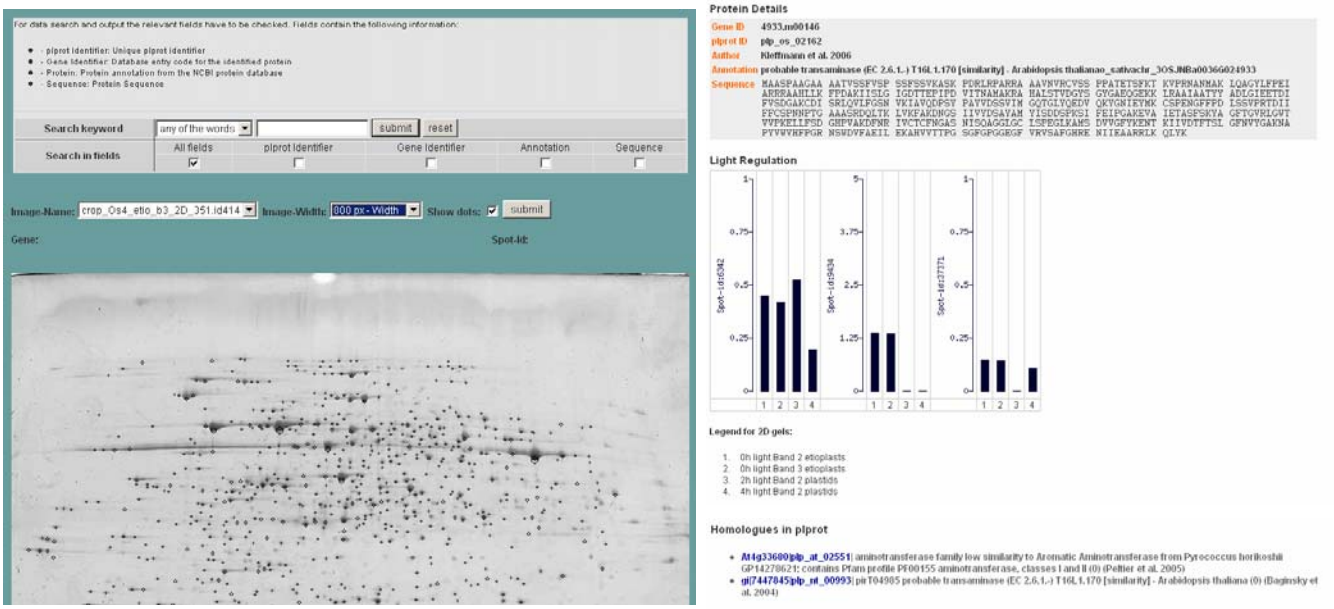


Fig. 3 The etioplast proteome map. Based on 2-dimensional gel electrophoresis, we have constructed an etioplast proteome map that provides quantitative information for all identified proteins. The etioplast proteome map can be queried via a search form (left panel) for identifier (plprot identifier), annotation or sequence. The number of hits, if any, is indicated right to the search input field and spots corresponding to correct hits are encircled in green (note: it is advisable to enlarge the 2D for an easier detection of the protein hit). Activating the URL underlying a protein spot will provide all available information about the protein, including its regulation by light, i.e. its intensity after 2 and 4 hours illumination (right panel).

about the solubility-based extraction step from which the protein was identified, i.e., osmotic shock, 8M urea wash or detergent solubilization. This information is useful to asses the solubility of proteins and their membrane attachment *in vivo*. For most of the identified Arabidopsis chloroplast proteins we provide RNA expression levels determined by Affymetrix ATH1 GeneChip® analysis (Kleffmann *et al.* 2004). Assessing the transcript levels of the identified proteins was useful to estimate the depth of our proteome analysis and the abundance of a protein in the chloroplast (Baginsky *et al.* 2005). All features mentioned above are provided together with manually annotated and curated protein functions, making plprot a unique resource of information about plastid proteins. In this respect plprot considerably differs from PPDB, a chloroplast protein database that is available at the Boyce Thompson Institute (Friso *et al.* 2004).

During the conversion of different plastid types, signifycant adaptations take place at the proteome level to support plastid-type specific functions. We have analyzed the light-induced conversion of etioplasts into photosynthetically active chloroplast by 2-dimensional gel electrophoresis, and made all data available in plprot. Activating the 2D-gel icon in the first plprot module (**Fig. 1**) opens a description of the

plprot plastid protein database

Institute of Plant Sciences - Laboratoy of Plant Biotechnology ETH

Protein Details

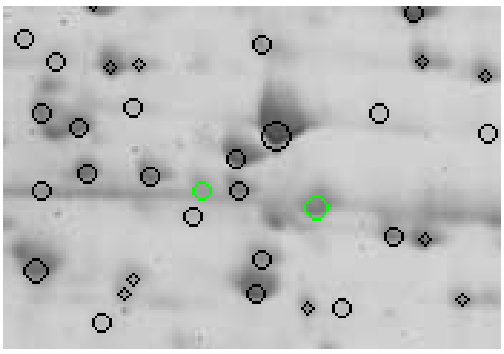
Gene ID	at4g16155
plprot ID	plp_at_00416
Author	Kleffmann et al. 2004
Annotation	dihydrolipoamide dehydrogenase 2, plastidic (lipoamide dehydrogenase 2) (ptlpd2) identical to plastidic lipoamide dehydrogenase from Arabidopsis thaliana [gi:7159284]
Sequence	MQSVLSLSFS QASLPLANRT LCSSNAAPST PRNLRFCGLR REAFCFSPSK QLTSCRFHIQ SRRIEVSAAA SSSAGNGAPS KSFDYDLIII GAGVGGHGAA LHAVEKGLKT AIIEGDVVGG TCVNRGCVPS KALLAVSGRM RELQNEHHMK AFGLQVSAAG YDRQGVADHA SNLATKIRNN LTNSMKALGV DILTGFGAVL GPQKVKYGDN IITGKDIIIA TGSVPFVPKG IEVDGKTVIT SDHALKLESV PDWIAIVGSG YIGLEFSDVY TALGSEVTFI EALDQLMPGF DPEISKLAQR VLINTRKIDY HTGVFASKIT PAKDGKPVLI ELIDAKTKEP KDTLEVDAAL IATGRAPFTN GLGLENINVT TQRGFIPVDE RMRVIDGNGK LVPHLYCIGD ANGKLMLAHA ASAQGISVVE QVTGRDHVLN HLSIPAACFT HPEISMVGLT EPQAREKAEK EGFKVSIAKT SFKANTKALA ENEGEGLAKM IYRPDNGEIL GVHIFGLHAA DLIHEASNAI ALGTRIQDIK LAVHAHPTLS EVVDELFKAA KVDSPASVTA QSVKVTV

Homologues in plprot

- at3g16950|plp_at_00281| dihydrolipoamide dehydrogenase 1, plastidic (lipoamide dehydrogenase 1) (ptlpd1) identical to plastidic lipoamide dehydrogenase from Arabidopsis thaliana [gi:7159282] (0) (Kleffmann et al. 2004)
- at3g16950|plp_at_01434| dihydrolipoamide dehydrogenase 1, plastidic (lipoamide dehydrogenase 1) (ptlpd1) identical to plastidic lipoamide dehydrogenase from Arabidopsis thaliana [gi:7159282] (0) (Froehlich et al. 2003)
- at4g16155|plp_at_01499| dihydrolipoamide dehydrogenase 2, plastidic (lipoamide dehydrogenase 2) (ptlpd2) identical to plastidic lipoamide dehydrogenase from Arabidopsis thaliana [gi:7159284] (0) (Froehlich et al. 2003)
- 3004.m00143|plp_os_02085| dihydrolipoamide dehydrogenaseo_sativachr_5P0676G053004 (0) (Kleffmann et al. 2006)
- 2800.m00138|plp_os_00676| dihydrolipoamide dehydrogenaseo_sativachr_1P0487E112800 (0) (von Zychlinski et al. 2005)
- 3000.m00123|plp_os_00700| dihydrolipoamide dehydrogenaseo_sativachr_5P0431G053000 (0) (von Zychlinski et al. 2005)
- at3g54660|plp_at_00344| gluthatione reductase, chloroplast nearly identical to SPP42770 Glutathione reductase, chloroplast precursor (EC 1.8.1.7) (GR) (GRASE) {Arabidopsis thaliana} (2e-045) (Kleffmann et al. 2004)
- At3g54660|plp_at_02539| gluthatione reductase, chloroplast nearly identical to SPP42770 Glutathione reductase, chloroplast precursor (EC 1.8.1.7) (GR) (GRASE) {Arabidopsis thaliana} (2e-045) (Peltier et al. 2005)
- 5863.m00168|plp_os_02199| glutathione reductaseo_sativachr_3OJ1607A125863 (3e-043) (Kleffmann et al. 2006)
- gi|482934|plp_nt_00942| embCAA54043.1 glutathione reductase (NADPH) [Nicotiana tabacum] (1e-036) (Baginsky et al. 2004)

Fig. 4 Example for a plprot search result for dehydrolipoamide dehydrogenase. Provided is the author and the study in which the protein identification was reported, the sequence, and all homologues in plprot that were identified in the different plastid types.

A Search results for plp_os_02085



B Regulation of plp_os_02085

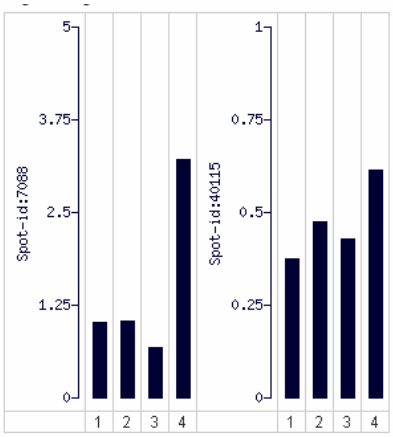


Fig. 5 Light regulation of dehydrolipoamide dehydrogenase. One homologue of dehydrolipoamide dehydrogenase was identified in etioplasts (see **Fig. 4**) (Kleffmann *et al.* 2006). With its entry number plp_os_02085 (see **Fig. 4**), we searched the etioplast 2D-map and retrieved one match. Clicking on the protein spot revealed that dehydrolipoamide dehydrogenase is distributed between two protein spots on the 2D-map and that it's abundance is higher in illuminated etioplasts. We can therefore conclude that dehydrolipoamide dehydrogenase is up-regulated by light. 1: etioplasts (Band 2); 2: etioplasts (Band 3); 3: 2 hours illuminated etioplasts (Band 2); 4: 4 h illuminated etioplasts (Band 2).

experimental strategy that underlies the acquisition of the quantitative proteome data. A direct entry to the interactive 2-D map is available in the second plprot module that contains the individual databases. Activating the option 2D Gel-ET/CP proteome map opens the 2D-gel map that can be searched with protein name, plprot identifier and database identifier, if known (note that the plprot identifier can best be accessed by the BLAST search option). A summary of search results (i.e. the number of detected proteins on the 2D-map that match the search criterion) for a specific search term are provided next to the input field. In case the protein has been detected, the corresponding spot(s) are encircled in green. Clicking on the protein spot opens all available information about the protein spot and its regulation by light. In cases where the same protein is distributed to different spots on the 2D gel, several different graphs will appear on the screen, each showing the individual regulation of each protein spot by light (**Fig. 3**).

FINDING A PROTEIN IN PLPROT

In order to search for a specific protein, we strongly suggest using the BLAST-search as the entry into plprot. This is the only "tolerant" search option that provides a number of hits ranked by percentage similarity. All other searches must provide a perfect hit in order to identify a protein in plprot. In case one or more homologues for a protein of interest are found in plprot, the reported plprot identifier for each homologue is the key to enter the complete plastid database or the individual databases for the different plastid types. Searching one of the plprot databases with the plprot identifier will give one positive result with a set of additional information as it is available for the different plastid types. The plprot identifier is connected with a hyperlink that provides access to amino acid sequence and a list of all homologues in plprot, i.e. from all plastid proteome studies that were conducted until to date. We show a representative search result in **Fig. 3** where we searched for plp_at_00416, a dihydrolipoamide dehydrogenase from *A. thaliana* (**Fig. 4**). A homologue of this protein has been identified in all different plastid types and by several different groups suggesting that this enzyme is part of the core proteome with essential functions for all plastid types analyzed to date.

The search result presented in **Fig. 4** shows that one homologue of Arabidopsis dehydrolipoamide dehydrogenase was also identified on the rice etioplast proteome map (indicated with Kleffmann *et al.* 2006). In order to assess its abundance in etioplasts and its regulation by light, we searched the etioplast 2D-map with its plprot-identifier plp_os_02085 and retrieved two matching spots, which are encircled in green (**Fig. 5A**). Clicking on one of the protein spots revealed that dehydrolipoamide dehydrogenase is a protein of relatively low abundance that is distributed among two protein spots on the gel, both up-regulated by light (**Fig. 5B**).

OUTLOOK

plprot will be updated regularly as new proteins are being identified from different plastid types. Exploiting the natural dynamic range diversity of proteins in different plastid types allows deep insight into the plastid proteome and is the key to unravel the full proteome of this organelle. In the long term, we plan to include more quantitative information about the identified proteins and use the information about protein abundance and catalytic activity to provide network models for the plastid metabolism at different differentiation states. We will supplement these models by thorough metabolite measurements and include these data with the metabolic network model. We are convinced that plprot will continue to grow and become the benchmark resource for plant scientists who are interested in plastid proteomics and its implications for basic cell biology and organelle differentiation processes.

REFERENCES

Abdallah F, Salamini F, Leister D (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. *Trends in Plant Science* **5**, 141-142

Agrawal GK, Rakwal R (2006) Rice proteomics: a cornerstone for cereal food crop proteomes. *Mass Spectrometry Reviews* **25**, 1-53

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402

Baginsky S, Gruissem W (2004) Chloroplast proteomics: potentials and challenges. *Journal of Experimental Botany* **55**, 1213-1220

Baginsky S, Gruissem W (2006) *Arabidopsis thaliana* proteomics: from proteome to genome. *Journal of Experimental Botany* **57**, 1485-1491

Baginsky S, Kleffmann T, von Zychlinski A, Gruissem W (2005) Analysis of shotgun proteomics and RNA profiling data from *Arabidopsis thaliana* chloroplasts. *Journal of Proteome Research* **4**, 637-640

Belostotsky DA, Rose AB (2005) Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events. *Trends in Plant Science* **10**, 347-353

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005-1016

Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, Sun Q, Wijk KJ (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* **16**, 478-499

Jarvis P, Robinson C (2004) Mechanisms of protein import and routing in chloroplasts. *Current Biology* **14**, R1064-1077

Kleffmann T, Hirsch-Hoffmann M, Gruissem W, Baginsky S (2006) plprot: a comprehensive proteome database for different plastid types. *Plant Cell Physiology* **47**, 432-436

Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S (2004) The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Current Biology* **14**, 362-375

Leister D (2003) Chloroplast research in the genomic age. *Trends in Genetics* **19**, 47-56

Lopez-Juez E, Pyke KA (2005) Plastids unleashed: their development and their integration in plant development. *International Journal of Developmental Biology* **49**, 557-577

Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiology* **118**, 9-17

Miras S, Salvi D, Ferro M, Grunwald D, Garin J, Joyard J, Rolland N (2002) Non-canonical transit peptide for import into the chloroplast. *Journal of Biological Chemistry* **277**, 47770-47778

Neuhaus HE, Emes MJ (2000) Nonphotosynthetic metabolism in plastids. *Annual Review of Plant Physiology and Plant Molecular Biology* **51**, 111-140

Peltier JB, Cai Y, Sun Q, Zabrouskov V, Giacomelli L, Rudella A, Ytterberg AJ, Rutschow H, van Wijk KJ (2006) The oligomeric stromal proteome of *Arabidopsis thaliana* chloroplasts. *Molecular and Cellular Proteomics* **5**, 114-133

Peck SC (2005) Update on proteomics in Arabidopsis. Where do we go from here? *Plant Physiology* **138**, 591-599

Richly E, Leister D (2004) An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. *Gene* **329**, 11-16

Rossignol M, Peltier JB, Mock HP, Matros A, Maldonado AM, Jorrin JV (2006) Plant proteome analysis: A 2004-2006 update. *Proteomics* **20**, 5529-5548

van Wijk KJ (2004) Plastid proteomics. *Plant Physiology and Biochemistry* **42**, 963-977

Weber AP, Schwacke R, Flugge UI (2005) Solute transporters of the plastid envelope membrane. *Annual Review of Plant Biology* **56**, 133-164