

S41-023

Bioinformatic and proteomic analyses to identify novel leaf peroxisomal proteins

L. Babujee, S. Lemke, C. Rittstieg, S. Reumann

Albrecht-von-Haller-Institut für Pflanzenwissenschaften, Abtlg. Biochemie der Pflanze, Universität Göttingen, Untere Karspüle 2, D-37073 Göttingen, Germany, Fax: (49) 551-395749, Email: sreuman@gwdg.de

Keywords: Bioinformatics, peroxisomes, photorespiration, targeting

Introduction

Peroxisomes represent ubiquitous oxidative cell organelles, in which toxic reaction products are generated and detoxified efficiently. Leaf peroxisomes participate in photorespiration and play a key role in the recycling of P-glycolate. Moreover, leaf peroxisomes mediate fatty acid β -oxidation and are involved in the biosynthesis of glycine betaine. Because plant peroxisomes are difficult to isolate in high purity and sufficient quantity, our knowledge of their metabolic capability is mainly restricted to the most abundant enzymes. We try to identify less abundant and differentially expressed unknown matrix proteins by two complementary approaches: first, extraction of proteins with peroxisomal targeting sequences from the *Arabidopsis* genome and bioinformatic analyses of their subcellular localization, tissue-specific expression and postulated function; and second, proteomic analysis of leaf peroxisomes from *Spinacia oleracea* L. by 2-D gel electrophoresis and protein identification by mass spectrometry.

Materials and Methods

The program "Pattern Match" at the TAIR server (<http://www.arabidopsis.org/>) was used with the pattern [SACP][KRH][LIM] for extraction of PTS1-containing proteins out of the database of non-redundant *Arabidopsis* proteins. Leaf peroxisomes were isolated from 10-week old spinach plants using Percoll gradients as described by Reumann et al. (1995). Immobilized pH gradients with a linear range of 4-7 (Amersham Pharmacia) were rehydrated with protein sample solubilized in a solution containing urea and CHAPS as prescribed in the manual. IEF was performed in a Multiphor II System for a total of 50 kVh according to the manufacturer's instructions.

Results and Discussion

Bioinformatic analyses of the Arabidopsis genome for peroxisomal proteins

Most matrix proteins of peroxisomes from yeast, mammalian and plant cells are targeted to their destination by the C-terminal peroxisomal targeting signal 1 (PTS1), the so-called "SKL" motif ([SAC][KRH]L>; Gould et al. 1989; Hayashi et al., 1997). To detect novel proteins that are possibly localized in peroxisomes, a relatively conserved variant of the PTS1 motif ([SACP][KRH][LIM]>) was used to extract PTS1-containing ORFs out of

the *Arabidopsis* genome. In total, 228 *Arabidopsis* ORFs were identified (date of analysis: July 27, 2001). An analysis of the frequency of each possible amino acids of the PTS1 to occur at the specific position shows a strong preference for certain amino acids (Fig. 1). The two tripeptides SKL> and SRL> are most frequent and found in about 35% of all extracted ORFs, whereas the 14 sequences of the patterns [SAPC]H[IM]> and

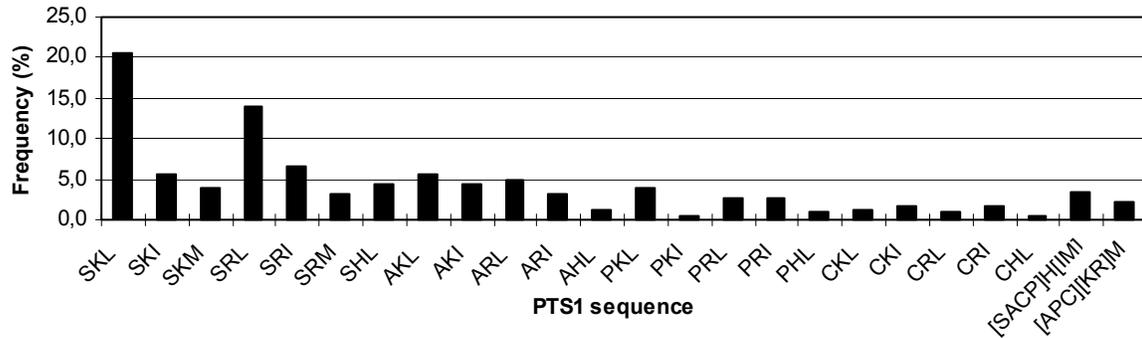


Fig. 1: Frequency of specific PTS1 tripeptides in *Arabidopsis* proteins.

228 ORFs containing the conserved pattern [SAPC][KRH][LIM]> were extracted out of the *Arabidopsis* genome, and the frequency of the tripeptides was analyzed.

[APC][KR]M> contribute only to a negligible portion of in total 6% of all extracted ORFs, indicating that these sequences play quantitatively only a minor role in targeting of *Arabidopsis* proteins to peroxisomes.

Despite the presence of a conserved variant of the PTS1 a certain number of false positive proteins of non-peroxisomal subcellular localization must be expected. True and some false positives are identified by three strategies. First, the identification of ESTs of other plants that correspond to the C-terminus of PTS1-containing *Arabidopsis* proteins and also contain a PTS1 followed by a stop codon is interpreted as supporting evidence for the peroxisomal localization of the protein. Second, fragment proteins are identified by homology analyses, and questionable intron/exon borders of ORFs are checked manually to identify predicted C-terminal PTS1 that are actually located in the middle of an ORF. Third, thanks to algorithms that have been deduced from targeting sequences from proteins of other cell compartments, the subcellular localization of many unknown *Arabidopsis* homologs can often be predicted with high accuracy (TargetP, Emanuelsson et al., 2000; PSORT: Nakai and Kanehisa, 1992). Even though dual targeting of proteins is known to occur, the presence of mitochondrial and plastidic targeting signals suggests that a less conserved variant of the PTS1 may be dominated by other targeting signals.

After elimination of false positives, the remaining PTS1-containing proteins that are targeted to peroxisomes with high probability fall into five different groups depending on the presence of homologs in other plants and their sequence similarity shared with other proteins. Group I and II contain known peroxisomal proteins and homologs the genes of which have previously been cloned from *Arabidopsis* or other plants. In the groups III to V a large number of proteins are found, many of which represent good candidates for novel proteins of plant peroxisomes. Group III contains novel isoforms of peroxisomal proteins that probably have a different function, e.g., substrate specificity. For instance, apart from two glycolate oxidase homologs, two *Arabidopsis* ORFs are only distantly

related to glycolate oxidase (about 60% sequence identity), but share high sequence similarity and conserved active site residues with medium- and long-chain hydroxyacid oxidases that have been cloned only recently from mammals (Jones et al., 2000), indicating that they probably represent homologs of the mammalian enzymes. Group IV contains annotated ORFs that are not yet known to be localized in peroxisomes. For instance, betaine aldehyde dehydrogenase was previously thought to be localized in dicots in chloroplasts. A second BADH isoform, however, contains a perfect SKL motif. Group V finally contains proteins that are largely unknown and share only little sequence similarity with other proteins. The function of these proteins needs to be investigated by overexpression studies or analyses of knock-out mutants. For all proteins the subcellular localization needs to be verified experimentally.

Proteomic studies of plant peroxisomes

Proteomics is fast emerging as a major research focus as we attempt to understand the functions of several genes. The establishment of the proteome of leaf peroxisomes can unravel hitherto unknown functions of this essential cell organelle. The fully completed genome sequence of *Arabidopsis thaliana* offers an exciting possibility to explore the peroxisomal proteome. However, preparations of Arabidopsis leaf peroxisomes with reasonable purity as dictated by proteome analysis are yet difficult to achieve. On the other hand, considerable amounts of highly enriched leaf peroxisomes can be obtained with relative ease from *Spinacia oleracea* L. Therefore, we have chosen this plant for the identification of novel leaf peroxisomal proteins.

A complete analysis of the proteome of leaf peroxisomes requires efficient protein solubilization and highly reproducible separation of proteins by 2-D gel electrophoresis. Solubility of proteins is an important factor that contributes to the quality of 2-D gels and determines the number of proteins to be identified. We have been using a buffer containing high concentration of urea together with a zwitter-ionic detergent for the solubilization of leaf peroxisomal proteins. An SDS-PAGE analysis of the insoluble fraction did not result in the detection of additional protein bands as compared to those detected in the soluble fraction (data not shown), indicating efficient solubilization of the large majority of proteins. A very high reproducibility with respect to pI and mol wt for individual spots is essential for the establishment of 2-D maps in order to identify novel proteins. Although our preliminary results suggest that the 2-D gel patterns of soluble proteins from leaf peroxisomes are highly reproducible, it will be worthwhile to readdress this question across several independent peroxisomal preparations and subsequent 2-D gel analysis of proteins.

Following separation of fifty micrograms of protein by 2-D electrophoresis and silver staining, as many as 150 polypeptide spots can be resolved (Fig. 2). Most of the proteins are mildly acidic as indicated by their pIs (4 to 6) although several neutral, basic and highly acidic proteins are also detected when a gradient with a nonlinear pH range of 3 to 10 is used in the first dimension IEF. A large majority of the proteins have a high mol wt between 70 and 30 and few outside this range. "Trains" of spots, indicative probably of post-translational modifications, are characteristic patterns on 2-D gels (see Fig. 2). It will be interesting to find out the nature of changes that create such differences in pI. Higher

sample loads and Coomassie blue staining will facilitate identification of novel and possibly low-abundance polypeptides by mass spectrometry.

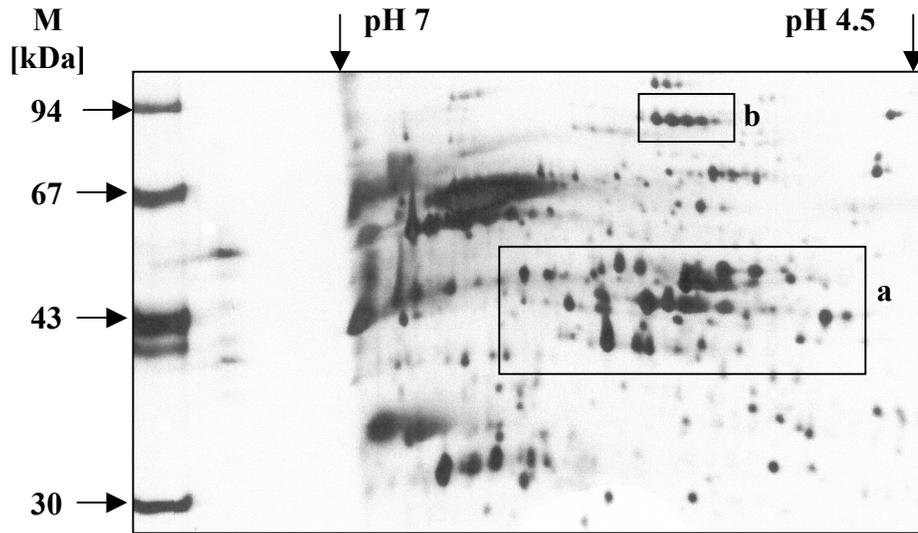


Fig. 2: 2-D gel electrophoresis of matrix proteins from spinach leaf peroxisomes.

Fifty μg protein were separated using a 13 cm linear immobilized pH gradient of 4-7 followed by denaturing PAGE (12.5% acrylamide) and stained with silver. Frame “a” encloses dominant discrete spots to be analyzed by mass spectrometry. Frame “b” shows a “train” of protein spots indicative probably of post-translational modifications.

Despite the high purity of peroxisomal fractions, minor contaminations with chloroplasts and mitochondria will still remain an inevitable reality that prevent the identification of peroxisomal proteins represented by minor spots. Subtractive gels that involve the comparison of 2-D patterns of highly purified proteins from all three cell organelles are in progress to help identify proteins that are truly characteristic of leaf peroxisomes.

The high purity of spinach leaf peroxisomes and the large number of protein spots on 2-D gels of matrix proteins compared to the low number of known proteins of leaf peroxisomes suggest that many of these proteins are currently unknown. We expect to be able to obtain partial sequence information of the proteins after tryptic digest and MALDI-TOF post source decay measurements or ESI-MS/MS of the peptides. These data are expected to allow for many spinach proteins the identification of homologs in the databases.

Acknowledgment

We are grateful to Prof. H.W. Heldt for all his support. This work is funded by the Deutsche Forschungsgemeinschaft.

References

- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) *J Mol Biol* **300**: 1005-1016.
- Gould SJ, Keller GA, Hosken N, Wilkinson J, Subramani S (1989) *J Cell Biol* **108**, 1657-1664
- Hayashi M, Aoki M, Kondo M, Nishimura M (1997) *Plant Cell Physiol* **38**, 759-768

Jones JM, Morrell JC, Gould SJ (2000) *J Biol Chem* **275**, 12590-12597
Nakai K, Kanehisa M (1992) *Genomics* **14**, 897-911.
Reumann S, Maier E, Benz R, Heldt HW (1995) *J Biol Chem* **270**, 17599-17565.