AraPerox. A Database of Putative Arabidopsis Proteins from Plant Peroxisomes

Sigrun Reumann*, Changle Ma², Steffen Lemke²,³, and Lavanya Babujee
Georg-August-University of Goettingen, Albrecht-von-Haller-Institute for Plant Sciences, Department for Plant Biochemistry, D–37077 Goettingen, Germany (S.R., C.M., L.B.); and Molecular Biology Graduate Program, Georg-August-University of Goettingen, GZMB, D–37077 Goettingen, Germany (S.L.)

To identify unknown proteins from plant peroxisomes, the Arabidopsis genome was screened for proteins with putative major or minor peroxisome targeting signals type 1 or 2 (PTS1 or PTS2), as defined previously (Reumann S [2004] Plant Physiol 135: 783–800). About 220 and 60 proteins were identified that carry a putative PTS1 or PTS2, respectively. To further support postulated targeting to peroxisomes, several prediction programs were applied and the putative targeting domains analyzed for properties conserved in peroxisomal proteins and for PTS conservation in homologous plant expressed sequence tags. The majority of proteins with a major PTS and medium to high overall probability of peroxisomal targeting represent novel nonhypothetical proteins and include several enzymes involved in β-oxidation of unsaturated fatty acids and branched amino acids, and 2-hydroxy acid oxidases with a predicted function in fatty acid α-oxidation, as well as NADP-dependent dehydrogenases and reductases. In addition, large protein families with many putative peroxisomal isoforms were recognized, including acyl-activating enzymes, GDSL lipases, and small thioesterases. Several proteins are homologous to prokaryotic enzymes of a novel aerobic hybrid degradation pathway for aromatic compounds and proposed to be involved in peroxisomal biosynthesis of plant hormones like jasmonic acid, auxin, and salicylic acid. Putative regulatory proteins of plant peroxisomes include protein kinases, small heat shock proteins, and proteases. The information on subcellular targeting prediction, homology, and in silico expression analysis for these Arabidopsis proteins has been compiled in the public database AraPerox to accelerate discovery and experimental investigation of novel metabolic and regulatory pathways of plant peroxisomes.

Peroxisomes constitute ubiquitous eukaryotic cell organelles in which a large variety of oxidative metabolic reactions are compartmentalized. Plant peroxisomes deserve increasing attention because the size of the peroxisomal proteome from plants seems to be much larger than that of yeast, trypanosomes, insects, and mammals (Emanuelsson et al., 2003) and because plants possess an exceptionally large number of metabolically specialized microbodies. Apart from their well-known function in photorespiration and lipid mobilization plant peroxisomes also play a significant role in nitrogen metabolism (Verma, 2002), degradation of branched amino acids (Zolman et al., 2001), and biosynthesis of plant hormones including jasmonic acid (JA) and auxin (Stintzi and Browse, 2000; Zolman et al., 2001; Feussner and Wasternack, 2002), as well as in the production of the compatible osmo-solute Gly betaine (Nakamura et al., 1997).

Flavin-containing oxidases like glycolate oxidase, acyl-CoA oxidase, urate oxidase, and amine oxidase, which transfer electrons directly to molecular oxygen and generate the highly toxic by-product hydrogen peroxide (H₂O₂), play a central role in most peroxisomal pathways. The high concentration of catalase (CAT) of up to 10% to 25% of the organelle’s protein guarantees rapid detoxification of H₂O₂ at the site of production and prevents, under normal conditions, leakage of the membrane-permeable molecule out of peroxisomes and oxidation of extraperoxisomal proteins. Apart from a strict compartmentalization of intraperoxisomal H₂O₂, peroxisomes are discussed to play an important role in detoxification of reactive oxygen species (ROS) generated outside of peroxisomes and to be a source of ROS acting as signal molecules (Willekens et al., 1997; Corpas et al., 2001; Titorenko and Rachubinski, 2004). The significance of peroxisomes in tolerance of oxidative stress is further supported by a pronounced proliferation of plant peroxisomes upon application of oxidative stress (Lopez-Huertas et al., 2000). Further antioxidative enzymes, such as superoxide dismutase, membrane-bound ascorbate peroxidase, monodehydroascorbate reductase, and glutathione peroxidase have been
partly cloned or localized by biochemical means to plant peroxisomes and are thought to play an auxiliary role in detoxification of ROS (Bunkelmann and Trelease, 1996; Kliebenstein et al., 1998; del Río et al., 2002). Peroxisomal Ser-glyoxylate aminotransferase from *Cucumis melo* has only recently been shown to confer enzymatic resistance against downy mildew caused by the pathogen *Pseudoperonospora cubensis* possibly via enhanced production of photorespiratory *H₂O₂* (Taler et al., 2004). Apart from classical enzymes that catalyze metabolic reactions, our knowledge on peroxisomal matrix proteins is rather limited due to difficulties in identifying low-abundance and inducible proteins by biochemical approaches. Evidence for the existence of regulatory proteins in peroxisomes, such as heat shock proteins, kinases, and phosphatases, is just emerging.

Most known peroxisomal matrix proteins are targeted to the peroxisomal matrix by the peroxisomal targeting signal type 1 (PTS1), the C-terminal so-called SKL-motif, whereas other matrix proteins carry a PTS2, which is a conserved nonapeptide of the prototype RLx₃HL that is embedded in the N-terminal domain (Gould et al., 1989; Swinkels et al., 1991; see also references in Reumann, 2004). Plant-specific PTS motifs have been deduced from experimental targeting studies in vivo and provided valuable information. They tend, however, to yield partially contradictory results, to deduce motifs of low specificity, and to neglect the proven role of accessory elements located in close proximity to the PTS and displaying an auxiliary targeting function (Hayashi et al., 1997; Mullen et al., 1997a; Kragler et al., 1998; Kato et al., 1998; Flynn et al., 1998). From an in silico study, in which protein and expressed sequence tag (EST) databases were searched for a maximum number of sequences that are homologous to PTS1- and PTS2-targeted plant peroxisomal proteins, it has been concluded that only a small number of nine major PTS1 and two major PTS2 peptides are widespread in plant proteins of different orthologous groups and represent strong indicators for peroxisomal targeting (Reumann, 2004).

To identify novel proteins located in plant peroxisomes, we screened the Arabidopsis genome (The Arabidopsis Genome Initiative, 2000) for genes encoding proteins with putative PTSs. As deduced from the results of various targeting prediction programs, examination of the targeting domain for properties conserved in plant peroxisomal proteins, and analysis of PTS conservation in homologous ESTs, a large number of novel proteins are localized in peroxisomes with high targeting probability. The presence of some of these enzymes in plant peroxisomes has been demonstrated earlier by biochemical means and can now be supported by genomic evidence. Other proteins have not been predicted in plant peroxisomes yet, and their postulated localization in plant peroxisomes yields surprising insights into the complexity of catalytic and biosynthetic pathways of plant peroxisomes.

**RESULTS**

**Extraction of Arabidopsis Proteins Carrying a Putative PTS**

Novel proteins of a particular cell compartment can possibly be detected by screening eukaryotic genomes for genes encoding proteins that carry the respective targeting signals. In contrast to targeting signals of peroxisomal membrane proteins, which have only been identified in a few cases and for which consensus sequences remain to be deduced (Jones et al., 2001), the targeting motifs of soluble matrix proteins have been specified by experimental and in silico studies. Nine major PTS1 tripeptides have been defined previously and are considered to indicate peroxisomal targeting with high probability ([SA][RK][LM]> without AKM> plus SRI> and PRL>; Reumann, 2004). In our initial screen for PTS1-targeted proteins, these major PTS1 peptides were supplemented by 11 minor PTS1 tripeptides, which comprised 5 canonical PTS1 tripeptides (SKI>, PRM>, PKL>, [C][RK][L]> of the relatively restrictive Hayashi motif ([PSAP][KR][LMI]>; Hayashi et al., 1997) and 6 minor noncanonical PTS1 tripeptides (SRV>, [SA][NL]>, SML>, SNM>, SSM>; Reumann, 2004). The tripeptides SHL> and AHL> were added as well because they have been shown to represent efficient targeting peptides in yeast, to direct a passenger protein to plant peroxisomes, or to interact with tobacco Pex5 in the yeast-two-hybrid system (Elgersma et al., 1996; Mullen et al., 1997a; Kragler et al., 1998; Lametschwandtner et al., 1998; Fig. 1A). This strategy allowed the detection of a maximum number of true peroxisomal proteins while reducing the number of false positives to a minimum, and was combined with a subsequent detailed analysis of targeting prediction for identification and exclusion of nonperoxisomal proteins.

About 220 different open reading frames (ORFs) were extracted out of the Arabidopsis genome (alternative splice or translation variants not counted) that carried one of these 22 PTS1 tripeptides. About 130 proteins (59%) of these carried a major PTS1. The largest number of proteins contained the PTS1 SKL> (43; 0.15% of all predicted Arabidopsis proteins) or SRL> (31 proteins), followed by those with S[RK][A][RK][L]> or PKL> (each 9–16 proteins), and then S[RK][M]> and PRL> (each 6 proteins; Fig. 1A). About 90 proteins carried a minor PTS1 or the tripeptides [SA][HL]>, of which a considerable number of 50 proteins contained a noncanonical PTS1 tripeptide (SRV>, SNL>, and SSM>, each 10–13 proteins; SML> and ANL>, each 6–7 proteins; Fig. 1A). If random distribution of amino acids is assumed, on average 3.6 of the 28,800 predicted Arabidopsis proteins are expected to contain one of 8,000 possible C-terminal tripeptides. Most of the PTS1 tripeptides that are present in a large number of Arabidopsis proteins well above average have recently been defined as major tripeptides and are considered strong indicators
for peroxisomal targeting. This rough correlation most likely reflects the late evolutionary stage of optimizing the efficiency of peroxisomal targeting peptides and provides further support for a specific targeting function of these peptides. Vice versa, most canonical tripeptides of the Hayashi motif (e.g. AKM, PKI, CR\[MI], CKM) that are present in a low number of Arabidopsis proteins (2 proteins; data not shown), were also rare in plant homologs of PTS1-targeted peroxisomal proteins and are not considered to target proteins efficiently to plant peroxisomes (Reumann, 2004).

In a similar way, about 60 proteins with one of two major (R\[LI\]x5HL) or nine minor PTS2 nonapeptides (R[QTMAV]x5HL, RLx5H(IF), R[IA]x5HI) located in an N-terminal domain of defined size were extracted out of the Arabidopsis genome. The two major PTS2 nonapeptides, RLx5HL and RLx5HL, were most frequent in Arabidopsis proteins (17 and 14 proteins, respectively), whereas a lower number carried one of the nine minor PTS2 peptides (1–5 proteins; Fig. 1B). Considering the number of 160,000 different tetrapeptides (i.e. nonapeptides with four conserved residues at position 2–38) the number of Arabidopsis proteins is compared with the frequency at which specific PTS peptides occurred in plant homologs of PTS-targeted proteins (Reumann, 2004).
positions 1, 2, 8, and 9) and the fact that these tetrapeptides can occur at 29 different positions if the first residue of the PTS2 is located between position 2 and 30, each tetrapeptide is expected to be present on average in 5.1 Arabidopsis proteins if random distribution is assumed. Thus, also the two major but none of the minor PTS2 nonapeptides are present in a number of Arabidopsis proteins two to three times above average.

As summarized for both PTSs, 282 Arabidopsis proteins were identified that contained either a major (162 proteins, 57%), a minor PTS (109 proteins, 39%), or [SA]HL> (11 proteins), of which most carried a putative PTS1 (222, 79%). Of the proteins with a major PTS about 24 proteins (15%; Table I) can be classified as known proteins from plant peroxisomes either because their genes have been directly cloned from Arabidopsis and protein targeting to peroxisomes has been demonstrated experimentally, or because they most likely represent the ortholog of peroxisomal proteins from different plant species. The majority of the extracted proteins, however, represent unknown nonhypothetical proteins, plant homologs of which have not yet been cloned and the function of which remains to be studied (about 100 proteins with a major PTS; Table I). The unknown nonhypothetical proteins with a major PTS comprise on the one hand partly annotated proteins that share some sequence similarity with mammalian or prokaryotic proteins (81 proteins; Table I), and on the other hand “expressed proteins” without any detectable homology to functionally analyzed proteins in the database (22 proteins). For many unknown proteins, corresponding ESTs are restricted to collections of plants subjected to abiotic stress conditions like dehydration, supporting the idea that analysis of the entire proteome of plant peroxisomes is hardly possible by an experimental proteome approach and needs to be supplemented by such a bioinformatics strategy. Of the extracted proteins with a major PTS about 35 ORFs (22%) still lack evidence for expression and encode hypothetical proteins (Table I). The only known peroxisomal matrix proteins not identified by this search are the three isoforms of CAT with an unusual PTS1 and xanthine dehydrogenase/oxidase with an unknown PTS (Mullen et al., 1997b; Kamigaki et al., 2003).

Validation of Predicted Peroxisomal Localization by Bioinformatics Tools

The PTS1 reveals a lower hierarchy as compared to N-terminal signals (Neuberger et al., 2003a), and

Table I. Homology and targeting analysis of Arabidopsis proteins with a major PTS

<table>
<thead>
<tr>
<th>Major PTS</th>
<th>Proteins</th>
<th>Number of Proteins</th>
<th>Overall Probability of Targeting to Peroxisomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>High</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>162</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Known peroxisome proteins</td>
<td>24</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Unknown nonhypothetical proteins</td>
<td>103</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Hypothetical proteins</td>
<td>35</td>
<td>63</td>
</tr>
<tr>
<td>PTS1</td>
<td></td>
<td>131</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Known peroxisome proteins</td>
<td>16</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Unknown nonhypothetical proteins</td>
<td>90</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Hypothetical proteins</td>
<td>25</td>
<td>88</td>
</tr>
<tr>
<td>PTS2</td>
<td></td>
<td>31</td>
<td>n.d.</td>
</tr>
<tr>
<td></td>
<td>Known peroxisome proteins</td>
<td>8</td>
<td>n.d.</td>
</tr>
<tr>
<td></td>
<td>Unknown nonhypothetical proteins</td>
<td>13</td>
<td>n.d.</td>
</tr>
<tr>
<td></td>
<td>Hypothetical proteins</td>
<td>10</td>
<td>n.d.</td>
</tr>
</tbody>
</table>
PTS1-like C-terminal tripeptides are found in nonperoxisomal proteins in which the tripeptide is not properly exposed on the surface of the folded protein (see also ‘Discussion’). Thus, even though the Arabidopsis proteins with a major PTS are expected to include many true peroxisomal proteins, a moderate number of false positives need to be anticipated within this group as well. The proteins with a minor PTS are localized in peroxisomes by definition with lower probability and are expected to include a considerable number of nonperoxisomal proteins because peroxisomal targeting of minor PTS is suspected to depend on targeting enhancing elements located next to the PTS (Reumann, 2004). Discrimination between true positives and nonperoxisomal proteins by bioinformatics tools is a challenging follow-up task of this genome screen. Some false positives are obvious and can be identified by their annotation. For example, the cytoplasmic 205 proteasome subunits PAE1 and PAE2 (PAE1, At1g53850; ARL->; PAE2, At3g14290; SRL->), the putative plastid division protein FtsZ2-2 (At3g52750, PRL->), which is a paralog of plastidic FtsZ2-1, a closely related Arabidopsis homolog of the GIC-6-P translocator (At5g54800, AKL->), and several putative transcription factors with a minor PTS are probably nonperoxisomal proteins (data not shown).

To identify more subtle nonperoxisomal proteins and provide further support for peroxisomal targeting of true proteins of peroxisomes, various subcellular prediction programs were applied. Predicted targeting to peroxisomes by a PTS1 can be supported by a few subcellular prediction programs, including TargetP, PSORT, and PSORTII. Only recently, new efforts have been undertaken to optimize the prediction of PTS1-targeted proteins (PeroxIP, Emanuelsson et al., 2003; PTS1 predictor, Neuberger et al., 2003a, 2003b). Plant-specific properties of the PTS, however, have still not been elaborated because of the predominance of CAT isoforms with atypical PTSs (Neuberger et al., 2003a, 2003b). Thus, negative prediction results for plant PTS1 proteins need to be handled with care. Algorithms for predicting PTS2-targeted peroxisomal proteins lack in all programs to date.

By contrast, plastidic transit peptides, mitochondrial presequences, and ER signal peptides can sometimes be predicted with high accuracy. Dual subcellular targeting due to alternative splicing, an alternative use of two translational start codons or depending on environmental conditions has been reported for some peroxisomal proteins, such as plant isoforms of HSP70 and Asp aminotransferase (Wimmer et al., 1997; Gebhardt et al., 1998). These proteins, however, are expected to represent an exception from the general rule of an exclusive localization of peroxisomal proteins in microbodies. The presence of nonperoxisomal targeting signals is therefore considered to decrease the probability that the same protein is localized in peroxisomes and necessitates a more detailed targeting analysis. Targeting prediction was considered significant if two independent programs predicted subcellular targeting to the same subcellular compartment with high probability. High overall probability for targeting to peroxisomes was defined if targeting to peroxisomes was predicted with high probability at least by one program and if no nonperoxisomal targeting signal was predicted with high accuracy by two independent programs. Medium and low overall targeting probability was deduced if one or two of these criteria were not fulfilled, respectively. Due to a lack of prediction software for PTS2-targeted proteins, estimation of targeting probability was performed similarly but restricted to the categories of medium and low probability. Regarding the 90 unknown nonhypothetical proteins with a major PTS1 (Table I), most are targeted to plant peroxisomes with high (56%) or medium probability (38%). By contrast, a considerable percentage of the unknown nonhypothetical proteins with a major PTS2 are targeted to plant peroxisomes with low probability (46%; Table I).

Accessory elements that are located in close proximity of PTS peptides have been reported to play an auxiliary role in targeting proteins to peroxisomes (Mullen et al., 1997a, 1997b; Lametschwandtner et al., 1998). The targeting domain of both PTS1- and PTS2-targeted plant peroxisomal proteins was found to be characterized by a high content of basic and Pro residues and by a pronounced increase in pl (Reumann, 2004). Especially for proteins with a minor PTS peptide and in case of an ambiguous prediction of subcellular localization, the presence of these conserved structural elements can provide further support for peroxisomal targeting.

According to our experience, the most solid support for targeting of unknown proteins to peroxisomes, however, can be provided by the identification of homologous genes and ESTs that encode the same compartment-specific ortholog and carry a PTS as well, embedded in a moderately conserved targeting domain. Prerequisites for a successful detection of PTS conservation in homologous genes/ESTs supporting peroxisomal targeting, however, are: (1) the absence of paralogous nonperoxisomal Arabidopsis proteins of high sequence similarity shared with the Arabidopsis protein of interest; (2) expression of the gene of interest in a common plant organ, from which many plant EST collections have been generated; and (3) a certain degree of sequence variation within the targeting domain. Especially for proteins with a low-abundance PTS1, such as CKL->, SML->, SNL->, or ANL->, many of which have previously only been detected in one orthologous group (i.e. sulfate oxidase), or for proteins with a putative PTS2, detection of a majority of homologous ESTs with major PTS peptides strongly supports protein targeting to peroxisomes (Supplemental Fig. 1, A–C, available at www.plantphysiol.org). On the other hand, if several presumably orthologous ESTs from different plant species do not carry a PTS1-like C terminal tripeptide, targeting of these proteins to peroxisomes is questionable (Supplemental Fig. 1D). These Arabidopsis proteins with
a putative PTS, as well as the results of the analysis of overall targeting prediction, homology, and in silico expression by Digital Northern (Mekhedov et al., 2000), have been compiled in the database AraPerox (www.araperox.uni-goettingen.de; Fig. 2).

Proteins with overall low probability of targeting to peroxisomes were excluded from further analyses presented in this study. Focusing on proteins that are targeted to plant peroxisomes with medium or high probability, many currently unknown proteins and even protein families with numerous putative peroxisomal isoforms were identified. These proteins include large protein families of acyl-CoA activating enzymes and thioesterase- and GDSL-lipase related proteins, many enzymes probably involved in β- and α-oxidation of unsaturated and branched-chain fatty acids, NADP-dependent dehydrogenases and reductases, and enzymes involved in N metabolism as well as regulatory proteins. The focus of the subsequent presentation will be on enzymes, the peroxisomal localization of which is strongly supported by previous biochemical data, PTS conservation in homologous ESTs, and/or by their homology to peroxisomal proteins from yeast and mammals.

A Large Family of Peroxisomal Isoforms of Acyl-Activating Enzymes

Arabidopsis contains a large superfamily of 63 mostly unknown acyl-CoA activating enzymes (AAEs, or AMP-binding proteins; Fulda et al., 1997) that activate their acid substrates using ATP via an enzyme-bound adenylylate intermediate, and include mainly acyl-CoA synthetases and 4-coumarate-CoA ligases. As noticed earlier, a surprisingly large number of 17 enzymes of this family carry a putative PTS (Fulda et al., 2002; Shockey et al., 2002, 2003; Staswick et al., 2002; Supplemental Fig. 2). For long-chain acyl-CoA synthetases 6 and 7, targeting to peroxisomes by a PTS1/2 and an essential role in lipid mobilization has been demonstrated by Suc-dependent germination of a double T-DNA knockout (k.o.) mutant (Fulda et al., 2002, 2004). Clade V and the plant-specific clade VI in particular contain a large number of mostly unknown enzymes that probably activate specific fatty acids in peroxisomes for subsequent metabolism (clade V, seven PTS1 proteins out of eight; clade VI, six PTS1 proteins out of 14; Shockey et al., 2003; Supplemental Fig. 2). Gene arrangement next to each other suggests substrate specification by recent gene duplication. The enzymes of clade V are closely related to 4-coumarate-CoA ligases of the neighboring clade IV, which produce CoA thioesters of a variety of hydroxy- and methoxy-substituted cinnamic acids, which are precursors of several phenylpropanoid-derived compounds, including anthocyanins, flavonoids, lignin, and coumarins (see references in Shockey et al., 2002, 2003). The putative peroxisomal isoforms have been proposed to catalyze CoA activation of benzoic acid derivatives or of very long-chain fatty acids (Shockey et al., 2003). Combined homology data of this study support involvement of some AAEs in peroxisomal metabolism of aromatic compounds, possibly including some plant hormones (see Fig. 7).

Enzymes Involved in β-Oxidation of Unsaturated Fatty Acids

In addition to the basic β-oxidation machinery (Fig. 3), degradation of mono- and polyunsaturated fatty acids, which are predominantly in cis configuration, requires auxiliary enzymes for elimination or isomerization of double bonds. Degradation of unsaturated fatty acids with double bonds extending from even-numbered carbon atoms (e.g. linoleic acid) yields an intermediate, which is degraded either by an
NADP-dependent or an alternative pathway (Gerhardt, 1993; Graham and Eastmond, 2002; Hooks, 2002; Fig. 3). In the NADP-dependent pathway, 2,4-dienoyl-CoA (2-trans,4-cis-dienoyl-CoA) is reduced to 3-trans-enoyl-CoA by an NADP-dependent 2,4-dienoyl-CoA reductase (DECR) and further isomerized by the $\Delta^3, \Delta^2$-enoyl-CoA isomerase activity of multifunctional protein (MFP) to 2-trans-enoyl-CoA (Preisig-Müller et al., 1994), the common substrate of MFP. Even though 2,4-DECR has been detected biochemically in plant peroxisomes, except for NADPH of 2,4-DECR, other cofactors and small metabolites like CoASH, water, O$_2$, H$_2$O$_2$, FAD, and NADH are omitted due to space limitations. The identified Arabidopsis homologs of 2,4-DECR and $\Delta^3, \Delta^2$-DCI are shaded gray. For possible monofunctional peroxisomal isoforms of enoyl-CoA hydratase, $\Delta^3, \Delta^2$-enoyl-CoA isomerase, and acyl-CoA dehydrogenase, see Figures 4, 7, and Supplemental Figure 4. ACX, Acyl-CoA oxidase.

Figure 3. Alternative pathways of peroxisomal $\beta$-oxidation of unsaturated fatty acids in plants. Two Arabidopsis proteins (At3g12800, SKL$^>$; At5g43280, AKL$^>$) that are homologous to mammalian and yeast DECR and DCI, respectively, and carry a conserved PTS (see Supplemental Fig. 4, A and B), support the existence of an alternative NADP-dependent pathway for degradation of unsaturated fatty acids with double bonds extending from even-numbered (left, A) and odd-numbered C-atoms (right, B) in plant peroxisomes. Except for NADPH of 2,4-DECR, other cofactors and small metabolites like CoASH, water, O$_2$, H$_2$O$_2$, FAD, and NADH are omitted due to space limitations. The identified Arabidopsis homologs of 2,4-DECR and $\Delta^3, \Delta^2$-DCI are shaded gray. For possible monofunctional peroxisomal isoforms of enoyl-CoA hydratase, $\Delta^3, \Delta^2$-enoyl-CoA isomerase, and acyl-CoA dehydrogenase, see Figures 4, 7, and Supplemental Figure 4. ACX, Acyl-CoA oxidase.

An Arabidopsis protein of the family of weakly conserved short-chain dehydrogenases/reductases (At3g12800, 298 residues) carries the major PTS1 SKL$^>$ and is the closest Arabidopsis homolog of mammalian and yeast 2,4-DEC (about 40% identity over 250 residues, E value = $10^{-39}$; De Nys et al., 2001; Gurvitz et al., 2001; Supplemental Fig. 3). Because targeting of this Arabidopsis protein to peroxisomes is supported by PTS1 conservation (Supplemental Fig. 4A), this protein most likely represents the plant peroxisomal ortholog of NADP-dependent 2,4-DEC and supports the existence of the NADP-dependent pathway for degradation of unsaturated fatty acids in plants (Fig. 3). According to our analysis and in contrast to a previous suggestion (Kamada et al., 2003), two other Arabidopsis proteins of the family of short-chain dehydrogenases/reductases (At4g05530, 254 residues, SRL$^{-}$; At3g59710, 302 residues, SKL$^>$; De Nys et al., 2001). An Arabidopsis protein of the family of weakly conserved short-chain dehydrogenases/reductases (At3g12800, 298 residues) carries the major PTS1 SKL$^>$ and is the closest Arabidopsis homolog of mammalian and yeast 2,4-DEC (about 40% identity over 250 residues, E value = $10^{-39}$; De Nys et al., 2001; Gurvitz et al., 2001; Supplemental Fig. 3). Because targeting of this Arabidopsis protein to peroxisomes is supported by PTS1 conservation (Supplemental Fig. 4A), this protein most likely represents the plant peroxisomal ortholog of NADP-dependent 2,4-DEC and supports the existence of the NADP-dependent pathway for degradation of unsaturated fatty acids in plants (Fig. 3). According to our analysis and in contrast to a previous suggestion (Kamada et al., 2003), two other Arabidopsis proteins of the family of short-chain dehydrogenases/reductases (At4g05530, 254 residues, SRL$^{-}$; At3g59710, 302 residues, SKL$^>$;
E values > 10^{-13}, are more distantly related to yeast and mammalian 2,4-DECR even though they are targeted to plant peroxisomes with high probability as well, and are expected to differ in substrate specificity (Supplemental Fig. 3).

Degradation of unsaturated fatty acids with double bonds extending from odd-numbered carbon atoms (e.g. oleic acid) is thought to proceed to the intermediate 3-cis-enoyl-CoA and relies on the above-mentioned \( \Delta^{5,5} \), \( \Delta^{2,4} \)-enoyl-CoA isomerase activity of MFP for isomerization to 2-trans-enoyl-CoA (Gerhardt, 1993; Preissig-Müller et al., 1994). In mammals and yeast, an alternative NADP-dependent pathway is used for these fatty acids. In this pathway, the intermediate 3-trans,5-cis-dienoyl-CoA is isomerized by \( \Delta^{3,3} \)-\( \Delta^{2,4} \)-dienoyl-CoA isomerase (DCI) to 2,4-dienoyl-CoA (2-trans,4-trans-dienoyl-CoA), where both NADP-dependent degradation pathways of unsaturated fatty acids merge (see above, Henke et al., 1998). Rather surprisingly, an Arabidopsis protein that belongs to the family of enoyl-CoA hydratases/isomerases (At5g43280, 278 residues) shares about 40% identity over the entire length with human \( \Delta^{3,3} \)-, \( \Delta^{2,4} \)-DCI (Filippula et al., 1998) and is the only enzyme of about 10 Arabidopsis enoyl-CoA hydratases with a putative PTS that clusters together in clade II with mammalian \( \Delta^{3,3} \), \( \Delta^{2,4} \)-DCI (Fig. 4). The Arabidopsis protein carries the major PTS1 AKL>, which is conserved or changed to SKL> in all detectable homologous ESTs (Supplemental Fig. 4B) and thus most likely represents the plant peroxisomal ortholog of mammalian \( \Delta^{3,3} \), \( \Delta^{2,4} \)-DCI. Successful complementation of the yeast mutant with the Arabidopsis gene is required to conclusively prove the existence of this novel degradation pathway of unsaturated fatty acids in plant peroxisomes. Apart from the isoforms of MFP and CHY1 (clades I and IV, respectively; Richmond and Bleecker, 1999; Zolman et al., 2001), four other unknown Arabidopsis proteins with a putative PTS belong to different clades of the superfAMILY of enoyl-CoA hydratases/isomerases (e.g. clade V, At1g60550, RLx5HL; clade VI, At4g16210, SKL>; Fig. 4), two of which interestingly cluster with PECI from fungi and mammals (clade III, At1g65520, SKL>; At4g14430, PKL>; see also Hooks, 2002; Fig. 4).

**Provision of Reducing Equivalents for the Peroxisomal NADPH Pool**

The enzyme 2,4-DECR requires stoichiometric amounts of intraperoxisomal NADPH for \( \beta \)-oxidation of unsaturated fatty acids with double bonds at even and possibly also at uneven position (Fig. 3). Because the membrane permeability of plant and yeast peroxisomes seems to be restricted to diffusion of small carboxylates and to prevent passage of NAD(P)H (van Roermund et al., 1995; Reumann et al., 1997, 1998), reduced equivalents need to pass the membrane by a malate/Asp-oxaloacetate shuttle (Mettler and Beevers, 1980; Reumann et al., 1994). In yeast and mammals, peroxisomal isoforms of NADP-dependent isocitrate dehydrogenase (NADP-IDH) provide NADPH for degradation of unsaturated fatty acids (Henke et al., 1998; van Roermund et al., 1998). Plant peroxisomal isoforms of NADP-IDH have been characterized biochemically (Corpas et al., 1999; del Río et al., 2002) but not yet at the molecular level. The Arabidopsis family of NADP-IDH comprises three closely related genes (Supplemental Fig. 5), one isoform of which carries the C-terminal tripeptide SRL> (At1g54340, 416 residues). This enzyme shares about 85% and 75% identity over the entire length with the putative cytosolic (At1g65930, 410 residues) and plastidic isoforms (At5g14590, 485 residues), respectively, the latter of which is extended at the N terminus by a transit peptide about 70 residues (Supplemental Fig. 5). Expressed sequence tags that share highest sequence similarity with the PTS1-containing NADP-IDH and also carry a PTS1, are found in many plant species and support targeting of the isoform to plant peroxisomes (Supplemental Fig. 4C).

The predicted localization of NADP-IDH and 2,4-DECR in plant peroxisomes indicates that NADPH is an important cofactor of peroxisomal metabolism next to NADH (Fig. 5). Peroxisomal 2-oxophytodienoic acid (OPDA) reductase isoform 3 (OPR3) is a second NADPH-dependent reductase of plant peroxisomes (Fig. 5). Besides NADP-IDH, the oxidative pentose-phosphate pathway (OPPP) has been discussed as an alternative mechanism for intraperoxisomal generation of NADPH, because all three enzymes of the oxidative part have been characterized biochemically in leaf peroxisomes from *Pisum sativum* (Corpas et al., 1998; Fig. 5). Two proteins paralogous to enzymes of the OPPP with a putative PTS1 can be detected in the Arabidopsis genome. The second enzyme of the OPPP, 6-phosphogluconolactonase (6PGL), has not been cloned from plants yet but the Arabidopsis genome encodes five putative 6PGL, which share about 40% identity (over 250 residues) with mammalian 6PGL and yeast homologs (Collard et al., 1999). One enzyme (At5g24400, 325 residues), which surpasses the length of the others by a predicted transit peptide of about 70 residues, carries additionally the putative PTS1 SKL>, suggesting dual targeting to plastids and peroxisomes (Supplemental Fig. 5). Indeed, several homologous ESTs carry a putative PTS1 as well (Supplemental Fig. 4D). Experimental support is required to verify the hypothesis of the existence of peroxisomal isoforms and their translation from the second Met, which is conserved in all homologs with a predicted transit peptide (data not shown).

Of three Arabidopsis homologs of 6-phosphogluconate dehydrogenase (6PGDH), one isoform (At3g02360, 486 residues) carries the minor PTS1 SKL> and represents a peroxisomal candidate for oxidation of 6-P-glucuronate to ribulose-5-P, NADPH, and CO\(_2\) (Fig. 5; Supplemental Fig. 5). The enzyme shares about 85% sequence identity with other full-length PTS1-containing homologs, many of which are annotated as cytosolic proteins (e.g. Medicago, SRI-->...
Figure 4. Phylogenetic analysis of Arabidopsis enoyl-CoA hydratases/isomerases with putative PTSs. Five unknown Arabidopsis proteins with a putative PTS (At5g43280, AKL; At4g14430, PKL; At1g65520, SKL; At1g60550, RLx; and At4g16210, SKL) that belong to the superfamily of enoyl-CoA hydratases/isomerases (ECH), were aligned with MFP and CHY1 homologs from Arabidopsis and Oryza sativa as well as prokaryotic and mammalian homologs of the family of enoyl-CoA hydratases/isomerases, and a phylogenetic tree was calculated using ClustalX (1,000 bootstraps). An unrooted phylogenetic tree was graphically constructed by Treeview. For eukaryotic proteins, recognizable PTS are indicated, and the five unknown Arabidopsis enoyl-CoA hydratases/isomerases are labeled by a gray box. Clade I of enoyl-CoA hydratases/isomerases comprises MFP homologs from plants (At_MFP2, At3g06860, SRL; At_AIM1, At4g29010, SKL; and Cs_MFP, Cucumis sativus, Q39659, PRM), bifunctional proteins (BFP) from mammals (Hs_BFP, Homo sapiens, NP_001957, SKL; and Rn_BFP, Rattus norvegicus, NP_598290, SKL), and a prokaryotic homolog (Po_FadB, Pseudomonas oleovorans, AAA83058). Clade II of enoyl-CoA hydratases/isomerases comprises D3,5, D2,4-DCI from mammals (Hs_DCI, AAH11792, SKL; Rn_DCI, NP_072116, SKL; and Mm_DCI, Mus musculus, NP_058052, SKL), two putative DCI (Ce_putDCI, Cenorhabditis elegans, NP_494954, SKL; and Nc_putECH, Neurospora crassa, XP_326933, SKL), and two plant homologs (At5g43280, AKL; and Os_putECH, Oryza sativa, NP_914858, SKL). Clade III of enoyl-CoA hydratases/isomerases comprises PECI from mammals (Hs_PECI, NP_006108, SKL; Mm_PECI, NP_035998, PKL; and Rn_PECI, XP_214464, PKL) and fungi (Sc_PECI, Saccharomyces cerevisiae, NP_013386) as well as three Arabidopsis homologs (At4g14430, PKL; At1g65520, SKL; and At4g14440). Clade IV of enoyl-CoA hydratases/isomerases comprises hydroxysubstituted-CoA hydratases (HIBCH), such as the Arabidopsis protein At_CHY1 (At5g63940, AKL), two closely related homologs of CHY1 (At_putHIBCH_1, At2g30660, AKL; and At_putHIBCH_2, At2g30650, AKL), two enoyl-CoA hydratases from other plant species (Am_putECH, Avicennia maritima, AAF01467, AKL; and Pa_putECH, Prunus armeniaca, AAB88874, AKL), three Arabidopsis homologs without recognizable PTS (At1g60550, At4g31810, and At3g60510), and mammalian homologs from human (Hs_HIBCH_1, NP_055177, and RTx; and Hs_HIBCH_2, NP_566220, and RAXx). Clade V of enoyl-CoA hydratases/isomerases comprises dihydroxyalcohol synthases (DHNS) from prokaryotes (Hsom_DHNS, Haemophilus somnus, ZP_00132565; Pm_MenB, Pasteurella multocida, NP_246013; and Ec_DHNS, E. coli, AAA33682) and two plant homologs (At1g60550, RLx; and Os_putECH, NP_917386, RLx). Clade VI of enoyl-CoA hydratases/isomerases comprises two unknown enoyl-CoA hydratases from plants (At4g16210, SKL; and Ca_putECH, Cicer arietinum, CAB61740, SKL). Due to space limitations, bootstrap values are omitted in the center of the tree.
Figure 5. Peroxisomal reactions involving the cofactor NADPH. According to this study, NADP-IDH and two enzymes of the OPPP, namely 6-PGL and 6PGDH, are targeted to plant peroxisomes with high probability (see Supplemental Fig. 4, C-E) and can provide NADPH for reductive matrix enzymes. Two NADPH-dependent reductases of plant peroxisomes carry a PTS1, i.e. OPR3 of JA biosynthesis, which converts OPDA to 3-oxo-2-(cis-2'-pentenyl)-cyclo-pentane-1-octanoic acid (OPC:8:0) and DECR of the NADP-dependent degradation pathway of unsaturated fatty acids (see Fig. 3). Oxidation of NADPH in the course of the ascorbate-glutathione cycle has also been reported (Jimenez et al., 1997).

Spinacia, SKI>; Oryza, AKM>; Supplemental Fig. 4E), and putative PTS1 tripeptides are detected in several homologous ESTs as well (Supplemental Fig. 4E). The Arabidopsis protein shares about 75% identity over the entire length with the two other Arabidopsis isoforms (At5g41670, At1g64190), none of which carries rather surprisingly an N-terminal extension resembling a transit peptide (Supplemental Fig. 5). By contrast, a possible candidate for a peroxisomal homolog of Glc-6-P dehydrogenase has not been detected yet in this study and may either be targeted to peroxisomes by an alternative mechanism or carry an unusual PTS. Alternatively, 6-P-glucuronolacton may enter the peroxisomal matrix directly from the cytoplasmic pool.

Catabolism of Branched Amino Acids

Catabolism of branched amino acids poses a considerable challenge because 3-methyl and 2-oxo or 2-hydroxyl groups represent a barrier for the core enzymes of fatty acid β-oxidation. Therefore, the postulated catabolic mechanism is a complex multi-step peroxisomal pathway that seems to be initiated in mitochondria (Gerhardt and Gerhardt, 1989; Gerhardt, 1993; Zolman et al., 2001; Graham and Eastmond, 2002). Many biochemically characterized enzymes of these pathways remain to be cloned and localized at a subcellular level.

Regarding Leu catabolism, the 2-oxo acid has been proposed to undergo first oxidative decarboxylation and to be further degraded via a free methyl-branched fatty acid to 2-methyl-propanoyl-CoA, which can re-enter the core complex of β-oxidation (Gerhardt, 1993). In mammals, an alternative catabolic pathway for Leu catabolism has first been described in mitochondria, in which basically 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) is cleaved by a lyase to acetyl-CoA and acetoacetate. Only recently, a putative PTS1 has been recognized at the C terminus of mitochondrial HMG-CoA lyase, and targeting of the unprocessed form to mammalian peroxisomes has been demonstrated (Tuinstra et al., 2002). A single Arabidopsis gene is homologous to mammalian HMG-CoA lyase and apparently encodes by use of two alternative translation start Met and possibly by alternative splicing of the 6th exon either a longer protein with a predicted mitochondrial presequence (At2g26800, 468 residues) or a shorter protein (433 residues) that is suspected to be targeted to peroxisomes by the putative PTS1 SKI-.

Homologous C-terminal ESTs from a large number of plant species, most of which also carry a putative PTS1, support targeting of the shorter isoform to peroxisomes and its role in an alternative pathway of peroxisomal Leu catabolism (Supplemental Fig. 4F). Involvement of HMG-CoA lyase in peroxisomal catabolism of aromatic compounds is also possible, whereas a role in isoprenoid metabolism like in mammals is not expected due to a plastidic localization of this pathway in plants.

Regarding Val catabolism, oxidative decarboxylation of the branched-chain 2-oxo acid is thought to yield 2-methyl-propanoyl-CoA, which can be degraded by the basic β-oxidation enzymes to 3-hydroxy-2-methyl-propanoyl-CoA. Instead of subsequent oxidation by MFP, 3-hydroxy-2-methylpropanoate is released from the CoA ester by CHY1, a 3-hydroxyisobutyryl-CoA hydrolase (At5g65940, 378 residues, AKL-), which belongs to the superfamily of enoyl-CoA hydratases/isomerases (Fig. 4). The enzyme has been identified in a screen of an Arabidopsis population mutagenized with ethyl methane sulfonate for plants that are resistant to the naturally occurring auxin indole-3-butyric acid but sensitive to indole-3-acetic acid (IAA; Zolman et al., 2001). Two closely related homologs of CHY1 (At1g30650, 410 residues; At2g30660, 376 residues, both AKL-), Zolman et al., 2001; Fig. 4) are candidate enzymes for hydrolysis of slightly different substrates, for instance 3-hydroxy-propionyl-CoA, an intermediate of propionate catabolism (Gerhardt 1993), but are still hypothetical proteins.

In the subsequent step of Val catabolism, 3-hydroxy-2-methyl-propionate (3-hydroxyisobutyrate) is oxidized to 3-oxo-2-methyl-propionate by a yet unknown alcohol dehydrogenase. Instead of an Arabidopsis protein similar to 3-hydroxyisobutyryl-CoA dehydrogenases (At5g15290, 294 residues, PRL-), which is expected to oxidize 3-hydroxy acids esterified to CoA (see Fig. 7), the two other short-chain reductases that are more
distantly related to 2,4-DECR and carry a putative PTS as well (Supplemental Fig. 3), may catalyze this reaction. A candidate for peroxisomal acyl-CoA dehydrogenase that mediates a subsequent step of Val catabolism (Zolman et al., 2001) and which has not been cloned from any organism yet, is the Arabidopsis protein At3g06810 (SKL>, 824 residues; data not shown). The enzyme is targeted to peroxisomes with high probability, contains the characteristic signature of acyl-CoA dehydrogenases and is related to several unknown homologs from mammals of the same size, all of which carry well-known putative PTS1 as well (e.g. ARM>, AKL>, SRL>.

A Novel Family of Small Peroxisomal Thioesterases

Acyl-CoA thioesterases catalyze the hydrolysis of fatty acyl-CoAs to free fatty acids and CoASH. Acyl-CoA thioesterases have been implicated in maintenance of adequate levels of free CoA within the peroxisomal matrix for continuous fatty acid β-oxidation by releasing CoA that is linked to non- or poorly-metabolizable acyl-CoA intermediates. Since the substrate specificity of enzymes involved in metabolism of complex fatty acids, such as ricinoleic acid and propionate, switches from CoA esters to free fatty acids (Gerhardt, 1993), several acyl-CoA thioesterases or hydrolases, such as the above-mentioned hydroxyisobutyryl-CoA hydrolase (CHY1; Zolman et al., 2001), may be required. Moreover, as β-oxidation of fatty acids includes not only catabolism but also biosynthesis of fatty acid derivatives, such as the important plant hormones JA, auxin, and possibly salicylic acid (SA; Stintzi and Browse, 2000; Zolman et al., 2001; Feussner and Wasternack, 2002), the precursor acyl-CoAs need to leave the β-oxidation complex at a certain point and the hormones must be released either by cytoplasmic or, more likely, by peroxisomal acyl-CoA thioesterases (see Fig. 7).

Five distinct families of acyl-CoA thioesterases have been defined for prokaryotes and eukaryotes, where they are found in different cell compartments. Two mammalian acyl-CoA thioesterases have been localized to peroxisomes and revealed a broad substrate specificity (Jones and Gould, 2000; Hunt et al., 2002). A corresponding plant homolog (ACH2, At1g01710, 427 residues, SKL>; homologous to At4g00520, 283 residues, AKL>), has been cloned only recently (Tilton et al., 2004). According to biochemical activity and expression data, the enzyme is unlikely to be linked to fatty acid oxidation as has been suggested for its eukaryotic homologs (Tilton et al., 2004). Our study revealed the existence of a large yet unknown Arabidopsis family of six small thioesterase-related proteins of only about 150 residues in size (At1g04290, At1g48320, At2g29590, At3g16175, At3g61200, and At5g48950). All six proteins carry a putative PTS1, such as SKL>, AKL>, and SNL> (Fig. 6), and their postulated function in plant peroxisomes is puzzling at first glance. The plant enzymes cluster into two different clades and are homologous to unknown small thioesterases from yeast, nematodes, insects, and mammals that strikingly do not carry a PTS1 or PTS2 (Fig. 6). The three small thioesterases from Arabidopsis of clade I are homologous to prokaryotic small thioesterases, such as Paal of Escherichia coli and Azorarcus evansii and FcbC of Arthrobacter sp., which belong to a novel catabolic pathway for the aromatic compounds phenylacetic acid and 4-Cl-benzoate, respectively (Benning et al., 1998; Ferrandez et al., 1998; Thoden et al., 2003; Fig. 6). This novel prokaryotic pathway has been termed aerobic hybrid pathway, because both free and CoA esterified fatty acids are pathway intermediates and substrates of the participating enzymes. Predicted targeting to peroxisomes of the homologs from Arabidopsis is supported by PTS1 conservation in homologs from Oryza (Fig. 6). The large number of small thioesterases in Arabidopsis and their presumable role in peroxisomal metabolism contrast the situation in fungi, nematodes, insects, and mammals (Fig. 6) and argue in favor of a pronounced plant-specific function of these enzymes, for instance in plant hormone biosynthesis (Fig. 7). Apart from this protein family, two other Arabidopsis proteins with a putative PTS1 belong to the family of esterases, lipases, and thioesterase and may also be involved in peroxisomal fatty acid metabolism (At3g28660, 348 residues, SSM>; At5g11910, 297 residues, SRI>.

Peroxisomal Enzymes Involved in Biosynthetic β-Oxidation of Plant Hormones

The peroxisomal enzyme OPR3 reduces OPDA to the intermediate 3-oxo-2-(cis-2-pentenyl)cyclopentane-1-octanoic acid, the octyl chain of which is shortened in three cycles of peroxisomal β-oxidation to an acetyl group to yield JA (Figs. 5 and 7). Similarly, indole-3-butyric acid is converted by one cycle of peroxisomal β-oxidation to IAA (Zolman et al., 2001). SA, a hormone involved in thermotolerance and pathogen resistance, is synthesized from trans-cinnamic acid by way of a side branch of the phenylpropanoid pathway. After one cycle of β-oxidation, which is assumed to take place in peroxisomes, benzoic acid is further hydroxylated in meta position to yield SA (Durner et al., 1997; Crozier et al., 2000; Fig. 7). The requirement for specific enzymes in prokaryotes strongly suggests that specific peroxisomal homologs of the enzymes of the basic β-oxidation machinery are necessary for biosynthesis of these aromatic or cyclic plant hormones. First, depending on whether these hormone precursors enter the peroxisomal matrix as free fatty acids or as CoA esters and depending on the substrate specificity of previous peroxisomal enzymes of the biosynthetic pathway (e.g. peroxisomal OPR3 acting on nonesterified substrates), the participation of peroxisomal AAEs needs to be postulated as an entrance point for peroxisomal β-oxidation. Second, considering the bulky size of the aromatic or cyclopentanone ring of IAA, SA, and JA...
and the fact that the substrate specificity of other enzymes, for instance that of the four peroxisomal isoforms of acyl-CoA oxidase, depends solely on the length of straight-chain acyl-CoA esters (Hayashi et al., 1999; Hooks et al., 1999; Eastmond et al., 2000; Froman et al., 2000), it is well possible that specific isoforms of enoyl-CoA hydratases including monofunctional D3, D2-enoyl-CoA isomerases (Fig. 4) and 3-hydroxyacyl dehydrogenases, catalyze the chain-shortening reactions of hormone precursors instead of MFP. Third and as outlined above, after thiolytic cleavage of the last C2 unit, the plant hormones are probably released from CoA esters by peroxisomal acyl-CoA thioesterases.

The gene encoding the small thioesterase FcbC of Arthrobacter sp.; phenylacetate, Paal of E. coli and Azorarcus evansi). For homologs carrying a putative PTS, the sequence of the PTS peptide is indicated. Homologous unknown small thioesterases were also detected in yeast, nematodes, insects, and mammals but lack a putative PTS, indicating a nonperoxisomal localization in all organisms except for higher plants. Abbreviations: Ae_ Paal, A. evansi, AAG28967; Ag_hom1, Anopheles gambiae, XP_308560; Asp_FcbC, Arthrobacter sp., C48956, Thoden et al., 2003; At, Arabidopsis; At_put_sT1 to At_put_sT6, At1g48320, At1g48320, At2g29590, and At3g16175; Ce_hom1 to Ce_hom3, C. elegans, NP_495115, NP_872068, and NP_498872; Dm_hom1 and Dm_hom2, Drosophila melanogaster, NP_647732 and NP_647730; Ec_ Paal, E. coli, P76084, Ferrandez et al., 1998; Hs_st, Homo sapiens, HTO12, PNAS-27, and NP_060943; Mus musculus, NP_080066; Nc_hom1 and Nc_hom2, Neurospora crassa, XP_325377 and XP_326708; Os_hom1 to Os_hom5, Oryza sativa, BAC79655, CAD40812, BAB90363, NP_913491, and NP_913490; Rn_st, Rattus norvegicus, XP_214475; Sp_hom1, Schizosaccharomyces pombe, NP_596564.

The three homologous genes of Pseudomonas sp. strain CBS-3 are arranged likewise (Benning et al., 1998). The closest Arabidopsis homologs of FcbA are the 4-coumarate-CoA ligase-like proteins with a putative PTS1 of clade V of AAEs (Supplemental Fig. 2, about 25% identity over the entire length) and other PTS1-containing AAEs. The enzyme 4-chlorobenzoyl-CoA dehalogenase (FcbB) from Arthrobacter sp. is a monofunctional enoyl-CoA hydratase, several Arabidopsis homologs of which carry a putative PTS (Fig. 4).

The operon of E. coli involved in phenylacetate degradation comprises 14 ORFs (Fig. 7), of which Paak, Paaf/G, and Paal are homologous to FcbA, FcbB, and FcbC from Arthrobacter sp., respectively (Schmitz et al., 1992; Ferrandez et al., 1998). In addition, Paaz encoded in the same operon is homologous to betaine aldehyde dehydrogenases (BADH), one of two Arabidopsis isoforms is probably peroxisomal (At5g20700, SKL.; see below). The closest Arabidopsis
Figure 7. Model of $\beta$-oxidation of aromatic and cyclic plant hormones in plant peroxisomes. The final steps of the biosynthesis of the aromatic or cyclic plant hormones JA, auxin, and SA are thought to take place in peroxisomes and involve one to three cycles of $\beta$-oxidation for shortening of the side chain, but the enzymes involved in precursor activation, $\beta$-oxidation, and hormone release are currently unknown. Prokaryotes possess specific enzymes for degradation of aromatic compounds that participate in a novel aerobic hybrid pathway and are homologous to Arabidopsis proteins with a putative PTS, which accordingly are proposed to be involved in peroxisomal reactions of aromatic and cyclic compounds. The operon of *Arthrobacter* sp. involved in degradation of 4-Cl-benzoate comprises three enzymes, namely an AAE (FcbA; about 25% identical over the entire length with 4-coumarate-CoA ligase-like proteins of clade V and other PTS1-containing AAEs from Arabidopsis; see also Supplemental Fig. 2), an enoyl-CoA hydratase (4-chlorobenzoyl-CoA dehalogenase, FcbB; about 30% identical over 200 residues with Arabidopsis enoyl-CoA hydratases, e.g. At5g43280, AKL; At4g16210, SKL; see Fig. 4), and a small thioesterase (FcbC; about 20% identical with the small thioesterases from Arabidopsis of clade I; Fig. 6). The operon of *E. coli* involved in degradation of phenylacetate (Ferrandez et al., 1998) comprises 14 ORFs, three of which are homologous to FcbA (PaaK), FcbB (PaaF/G), and FcbC (PaaI) of *Arthrobacter* sp. In addition, an aldehyde dehydrogenase (PaaZ) is homologous to Arabidopsis BADH (At3g48170, SKL; 20% identical over 400 residues). A hydroxybutyryl-CoA dehydrogenase (PaaH) is homologous to a multifunctional hydroxybutyryl-CoA dehydrogenase from Arabidopsis (At3g15290, PRL; about 40% identical over the entire length; see Supplemental Fig. 4G). The order in which the enzymes of *E. coli* are involved in degradation of phenylacetate has not yet been analyzed in all detail and has been drawn according to available information.
homolog of PaaH is a monofunctional hydroxybutyryl-CoA dehydrogenase (At3g15290, 294 residues, PRL>, about 40% identical over the entire length) with high targeting probability to peroxisomes due to PTS1 conservation in homologous ESTs (Supplemental Fig. 4G). Distinct isoforms of six acyl-CoA oxidases and three thiolases in Arabidopsis may also be specific for aromatic compounds. Taken together, these homology data suggest that several Arabidopsis proteins with a putative PTS are involved in peroxisomal biosynthetic and possibly also catabolic reactions of aromatic and cyclic plant hormones and pave the way for straightforward experimental strategies to test these hypotheses.

**Novel Paralogs of Glycolate Oxidase with a Postulated Role in α-Oxidation**

Glycolate oxidase (GOX) is the well-known first peroxisomal enzyme of the glycolate pathway and catalyzes the oxidation of glycolate to glyoxylate coupled to the reduction of O$_2$ to H$_2$O$_2$ via the cofactor flavinmononucleotide (Volokita and Somerville, 1987). The genome of Arabidopsis contains an unusually large gene family of five GOX homologs (Reumann, 2002). Two proteins (GOX1, At3g14420; GOX2, At3g14415) are closely related to each other and to Spinacia oleracea GOX, and their role in photosphere is supported by a high number of ESTs derived from photosynthetic tissue (Fig. 8A). The third GOX paralog (GOX3, At4g18360, AKL>) seems to have the same substrate specificity but is specifically expressed in nongreen tissue (Kamada et al., 2003; Fig. 8A). Two more distantly related Arabidopsis paralogs (HAOX1, At3g14130; HAOX2, At3g14150, about 60% identical over the entire length) carry the minor PTS1 SML and are, according to their tissue expression profile, predicted to play a role in nonphotosynthetic metabolism (Fig. 8A). Prediction of different substrate specificity is based on the presence of two differing active site residues as compared to GOX (Y-24→F and W-108→M; Lindqvist and Braenden, 1989; Fig. 8B). Interestingly, three peroxisomal GOX homologs have been identified recently in mammals (hydroxy acid oxidase 1-3 [HAOX1-3]; Kohler et al., 1999; Jones et al., 2000; Williams et al., 2000). In contrast to one glycolate-oxidizing isoenzyme (HsHAOX1), HsHAOX2 and HsHAOX3 are specific for long-chain and medium-chain 2-hydroxy acids, respectively, and reveal exchanges of active site residues as compared to photosynthetic GOX that resemble those of the novel Arabidopsis paralogs (Fig. 8B). It may be predicted that the two novel GOX paralogs from Arabidopsis have a function closer to that of HsHAOX2/3 and metabolize a substrate of longer chain length than glycolate. Because HsHAOX2/3 are discussed to be involved in α-oxidation of fatty acids, a process that is only marginally understood in both mammals and plants, the substrate specificity is expected to yield important insights into the...
function of plant peroxisomes in α-oxidation of 3-methyl- and 2-hydroxy acids.

Other Enzymes Related to Fatty Acid Catabolism

Lipolytic enzymes constitute a large family of different classes, including not only true lipases, which display maximal activity toward water-insoluble long-chain triacylglycerides, but also carboxylesterases, which hydrolyze small ester-containing molecules at least partly soluble in water, and various types of phospholipases (Upton and Buckley, 1995; Arpigny and Jaeger, 1999). Class II carboxylesterases from prokaryotes contain the so-called GDSL lipases that carry the active site residue Ser of the catalytic triad (GDS(L) instead of GxSxG) and localized very close to the N terminus (Arpigny and Jaeger, 1999). Prokaryotic GDSL lipases are membrane-bound or secreted at the N terminus (Arpigny and Jaeger, 1999). Class II carboxylesterases from Arabidopsis contain eight proteins that are indeed targeted to plant peroxisomes, they are soluble enzymes. At least six of these Arabidopsis proteins are expressed (At1g33811, ANL; At2g38180, ARL; At3g04290, SKI; At4g18970, ARL; At4g28780, SRI; At5g08460, SRL). The GDSL motif of the Arabidopsis proteins contains an additional Val residue at pos. 5 (GDSLV) like the mammalian enzyme in contrast to prokaryotic enzymes (GDS(L); Arpigny and Jaeger, 1999; Wilhelm et al., 1999; Fig. 9). The two other active site residues, Asp and His, are conserved as well (FxDXxHPT, Fig. 9; Wilhelm et al., 1999). If these isoforms of GDSL lipases are targeted to plant peroxisomes, they are suspected to be involved in releasing fatty acids from complex conjugates, for instance Glc and amino acid conjugates of plant hormones.

Peroxisomal Enzymes Involved in N Metabolism

With the peroxisomal enzymes uricase and xanthine dehydrogenase/oxidase, plant peroxisomes are involved in N metabolism and may compartmentalize further biosynthetic or catabolic pathways of N-containing metabolites. Aldehyde oxidase 2 (At3g43600, 1321 residues, SNL→), which is distantly related to xanthine oxidase/dehydrogenase and contains the same co-factor binding sites (Sekimoto et al., 1998; Seo et al., 2000), is one example and may be targeted to peroxisomes by the minor PTS1 SNL→. Polyamines are crucial components for growth and cell proliferation of all organisms and stimulate many reactions by their polycationic nature. Polyamines comprise the primary amines putrescine and cadaverine as well as spermidine and spermine, of which the latter carry additional secondary amino groups (Walden et al., 1997; Pandey et al., 2000; Perez-Amador et al., 2002). During catabolism of polyamines in plants, the primary amines are subjected to oxidative deamination by copper-containing diamine oxidases producing H2O2, ammonia, and aminodehydes. The secondary amines present in spermine and spermidine can also be catabolized by a second class of enzymes, namely flavin-containing polyamine oxidases, which form H2O2, 1,3-diaminopropan and aminodehydes. The first polyamine oxidase has only recently been cloned from mammals, and postulated targeting to peroxisomes by the putative PTS1 PRL→ remains to be demonstrated (Vijucic et al., 2003; Wu et al., 2003). Preferred substrates were N1-acetylated polyamines, indicating a physiological role of the enzyme in polyamine back-conversion of spermine and spermidine to putrescine and spermidine, respectively (Vijucic et al., 2003; Jänne et al., 2004). Plant peroxisomes may also play a significant role in catabolism and/or back-conversion of polyamines as well, since three (poly) amine oxidases from Arabidopsis carry a putative PTS1 as noticed earlier (At1g65840, 497 residues, SRM→; At3g59050, 488 residues, SRM→; At2g43020, 490 residues, SRL→; Kamada et al., 2003).

Even though the enzyme BADH, which catalyzes the second and final biosynthetic step of the compatible osmosolutes Gly betaine, is a chloroplastic enzyme in Gly betaine-accumulating plant species, BADH

---

**Figure 9.** Sequence alignment of putative GDSL lipases from plant peroxisomes. Seven Arabidopsis GDSL-like lipases with a putative PTS1 were aligned. The numbers at the top correspond to the residues of the first protein (At1g33811). Due to space limitation, only the N-terminal GDSL domain, an internal domain with a putative active site residue (D174 for At1g33811), and the C-terminal end of the Arabidopsis proteins with two putative active site residues (D336 and H338 for At1g33811) and the PTS1 are shown. The GDSL motif and conserved residues that correspond to active site residues in prokaryotic homologs are indicated at the bottom. The putative PTS1 tripeptides are underlined.
homologs of *Oryza sativa* and *Hordeum vulgare* are peroxisomal enzymes and also seem to confer salt tolerance (Nakamura et al., 1997). Arabidopsis contains two closely related BADH isoforms, one of which terminates with the PTS1 SKL> (At5g48170, 503 residues). An analysis of ESTs suggests coexistence of a peroxisomal and a nonperoxisomal (chloroplastic) isoform in dicots and monocots (S. Lemke and S. Reumann, unpublished data). Because the presumed high substrate specificity of chloroplastic BADH has been questioned recently (Vojtechova et al., 1997; Sebela et al., 2000), the more general annotation as an aminoaidehyde dehydrogenase seems to be more appropriate (Livingstone et al., 2003). The kinetic parameter and subcellular localization of both BADH paralogs require reinvestigation to elucidate their physiological function, with special consideration of the production of aminoaidehyde derivatives in the course of polyamine catabolism.

**Regulatory Proteins of Peroxisomes**

Our knowledge on peroxisomal matrix proteins with a regulatory function is rather limited due to difficulties in identifying low-abundance and inducible proteins by biochemical approaches. Evidence for the existence of regulatory proteins, such as heat shock proteins, kinases, and phosphatases, in peroxisomes is just emerging. Regarding heat shock proteins expressed in response to increased temperature and other forms of abiotic stresses, a chloroplastic isoform of an HSP70 homolog from *Citrus vulgaris* is also targeted to peroxisomes by alternative use of two successive translation start codons (Wimmer et al., 1997). Small heat shock proteins (sHSPs) are a ubiquitous superfAMILY of HSPs characterized by the relatively small mass of the polypeptide chain (16–42 kD) and the presence of a conserved a-crystallin domain (Kim et al., 1998; Haslbeck, 2002). Plants house an exceptionally large family of sHSPs (Scharf et al., 2001). Experimental evidence for the presence of sHSPs in peroxisomes has not been provided yet for any organism. Two members of this family, however, carry a putative PTS1 (At5g37670, 137 residues, SKL>; At1g06460, 285 residues, PKL> and RLX5HF) and may be targeted to plant peroxisomes. The second sHSP is intriguing because of the presence of a putative PTS1 (PKL>) as well as a putative PTS2 (RLX5HF), a feature that has so far only been reported for LACS7 (Fulda et al., 2002).

Our largest gap of knowledge is probably that of the regulation of peroxisomal enzymes by posttranslational modifications and protein turnover. Some peroxisomal enzymes, in particular CAT, have been suggested to be modified posttranslationally by proteolytic cleavage (Kleff et al., 1997). Fragmentary signaling cascades involving protein kinases and phosphatases have not been unraveled in any case. A protein kinase with yet unknown target was recently found in glyoxysomes in a proteome study of Arabidopsis (At3g17420, AKI>; Fukao et al., 2003). A calcium-dependent protein kinase is anchored in the peroxisomal membrane by an acyl residue (Dammann et al., 2003). Calmodulin was found in a peroxisomal fraction and regulates CAT in the presence of Ca2+ (Yang and Poovaiah, 2002), and a light-responsive nucleoside diphosphate kinase is reported to interact with CAT (Fukamatsu et al., 2003). Our screen of the Arabidopsis genome revealed seven putative protein kinases that are suspected to mediate signal transduction across the peroxisomal membrane and to alter the activity of peroxisomal enzymes by reversible phosphorylation.

Turnover of peroxisomal enzymes may be regulated by a peroxisomal homolog of the mitochondrial Lon protease (At5g47040, 888 residues, SKL>), a homolog of which has also recently been localized in mammalian peroxisomes (Kikuchi et al., 2004). A glutathione transferase (At5g41120, 245 residues, SKL>), which may be a homolog of peroxisomal glutathione S-transferase from mammals (Morel et al., 2004), or two protease-related proteins (At1g28320, 709 residues, SKL>; At2g41790, 970 residues, PKL>) may be further regulators of peroxisomal protein degradation. A Ca2+-binding protein (At5g07490, 153 residues, SNL>) could be involved in signal transduction of peroxisomal CDPK1 and CAT-activation by calmodulin, but the targeting function of the putative PTS1 can currently not be supported PTS1 conservation in homologous ESTs (Supplemental Fig. 1D). One homolog of a purple acid phosphatase also carries a putative PTS1 (At2g01880, 328 residues, AHL>). By contrast, considering the large number of more than 550 F-box proteins in Arabidopsis (Kuroda et al., 2002) and the current lack of evidence for expression of most of the genes encoding about 8 F-box proteins with a putative PTS, most of these proteins are not expected to play a major role in plant peroxisomal metabolism.

**DISCUSSION**

In Silico Analysis of Protein Targeting to Peroxisomes

Well-defined conserved targeting signals that direct nuclear-encoded proteins to a particular cell compartment offer the possibility to screen eukaryotic genomes for unknown proteins that carry these signals and are probably localized in the organelle of interest. The number of false positives relies primarily on correct gene prediction and on the specificity of the applied targeting motifs. In the first years after publication of the Arabidopsis genome (The Arabidopsis Genome Initiative, 2000), incorrect gene prediction affected about 30% to 40% of our extracted proteins, for instance a ‘forever young protein’ now predicted with an internal SKL tripeptide (At4g27760; Kamada et al., 2003). Gene prediction, however, has improved considerably since and been supported by large-scale full-length cDNA projects (Yamada et al., 2003) with the result that only a minor portion of our extracted proteins encoded by predicted genes are not expected to play a major role in plant peroxisomal metabolism.
proteins are estimated to be subjected to incorrect PTS prediction at present. However, a considerable number of additional splice variants and alternative start Mets are still being detected and require a regular update of genome screens.

Even though PTSSs with their conservation of three to four residues suggest a high specificity of the deduced targeting motifs, a clear determination of allowed residue combinations and the definition of plant-specific PTS motifs remain a challenge. Prediction of peroxisomal targeting is made difficult because even the presence of a well-known PTS (e.g. SKL>, SRL>, or RLx2HL) is not a sufficient criterion for unambiguous targeting of a “natural” protein to the peroxisome matrix and contrasts experimental results obtained with artificial fusion proteins. Some proteins with a putative PTS are not targeted to peroxisomes because the tiny PTS1 tripeptide in particular is not properly exposed on the surface of the folded protein or because peroxisomal targeting of low-abundance “weaker” PTS depends on the presence of marginally defined accessory elements that substantially enhance the affinity to its receptor. Furthermore, the C-terminal PTS1 is probably of lower hierarchy as compared to N-terminal signals (Neuberger et al., 2003a) because of the temporary advantage of N-terminal signals for their differing rates of false positives, and focused in the subsequent in silico analyses primarily on those proteins with a major PTS. Proteins with a minor PTS can provide valuable hints on novel targeting properties, and a considerable number of false positives need to be anticipated among proteins with a minor PTS. According to our experience, distinction of true positives, however, may indeed be hidden among these Arabidopsis proteins with putative PTS peptides other than those defined as major and minor PTS. For instance, a homolog of monodehydroascorbate reductase carries the putative PTS1 AKI> (At5g52880; Kamada et al., 2003) and has been characterized biochemically in plant peroxisomes (Bunkelmann and Trelease, 1996). The analysis of PTS conservation in novel proteins in the course of this study and the detection of homologous ESTs with PTS peptides that had so far been restricted to single orthologous groups (e.g. SML>, unique ESTs (e.g. AKI>, CRI>, SSL>), or which had completely been absent (RLx4HV, SSI>, SNN>, AML>, or FKL>; see Supplemental Figs. 1 and 4; Reumann, 2004), indicate that additional peptides represent low-abundance but functional PTS1 tripeptides. These peptides will be defined as minor PTS peptides and added to our genome screen upon experimental localization of one of these novel orthologs to peroxisomes. Our limitation dealing with the localization of the PTS2 nonapeptide within the N-terminal 38 residues is also more restrictive (compared to Kamada et al., 2003) but supported by an analysis of more than 150 plant sequences (R of position 1 of the PTS2 at most at position 26; data not shown).

The Choice of the Screening Peptides

Our screen of the Arabidopsis genome for peroxisomal proteins is characterized by a restrictive application of targeting peptides, as we included only peptides that had previously been detected in a significant number of plant homologs of PTS-targeted matrix enzymes (Reumann, 2004). In contrast to a comparable genome analysis (Kamada et al., 2003), 10 peptides of the Hayashi motif (i.e. A[RI]K>,[PC][RI]>,[PC][K][MI]>; Hayashi et al., 1997) and 10 additional tripeptides with H at position 2 (SAH>[MI]>,[PC][H][LMI]>) were excluded in our study due to a current lack of experimental and in silico support for their peroxisomal targeting properties and in light of the subsequent difficulties in sorting out false positives by further bioinformatics analyses. Our in silico analysis strongly suggested that position-specific residues of functional PTS peptides cannot be combined freely and that two to three low-abundance residues are unlikely to constitute functional PTS1 tripeptides (Reumann, 2004). For the same reasons, the large number of additional PTS peptides included in more permissive experimentally determined PTS motifs were also omitted (e.g. PTS1, [SAPC]IG[KR]HNL[LMTI]Y>; Mullen et al., 1997a; Kragler et al., 1998; PTS2, [RK]x6[HQ][ALF]; Flynn et al., 1998). A low number of true positives, however, may indeed be hidden among these Arabidopsis proteins with putative PTS peptides other than those defined as major and

Bioinformatics Validation of Peroxisomal Targeting

Despite our restrictive application of PTS peptides, some proteins have been extracted from the Arabidopsis genome that are obviously nonperoxisomal proteins, and a considerable number of false positives need to be anticipated among proteins with a minor PTS. According to our experience, distinction of true and false positives is essential before drawing premature conclusions with respect to the identification of novel proteins from plant peroxisomes and before proceeding with experimental studies (Kamada et al., 2003). Several bioinformatic strategies have been applied in our study to increase the prediction accuracy of protein targeting to peroxisomes. These analyses comprised algorithm-based prediction of PTS1-targeted proteins as well as nonperoxisomal targeting signals, analysis of the targeting domain for properties conserved in plant peroxisomal proteins, and analysis of PTS conservation in homologous ESTs. Even though subcellular targeting prediction of different programs

Database of Arabidopsis Peroxisomal Proteins

www.plantphysiol.org
Novel Proteins from Plant Peroxisomes

Several unknown proteins with a putative PTS are localized in peroxisomes with high probability as indicated by overall targeting prediction, EST analysis, and functional evidence provided by homology analysis. The enzymes 2,4-DECR, Δ^5, Δ^24-DCI, and NADP-IDH have been localized to yeast and mammalian peroxisomes and are involved in β-oxidation of fatty acids with double bonds at even- and possibly also odd-numbered carbon atoms. Three Arabidopsis proteins are probably orthologous, catalyze the same reaction as their counterparts in yeast and mammals, and strongly suggest that degradation of unsaturated fatty acids by NADP-dependent enzymes plays a more pronounced role in plant peroxisomal metabolism than previously assumed. It will be interesting to gain insights into the physiological conditions under which these alternative pathways are active and how they are regulated. In addition, we have presented indications that two isoforms of the OPPP, namely 6PGL and 6PGDH, are targeted to peroxisomes and may provide NADPH for JA biosynthesis or DECR activity (Fig. 5). These molecular data support the previous biochemical characterization of the enzymes in peroxisomes (Corpas et al., 1998). Because minor amino acid exchanges can alter the substrate specificity, experimental verification of the postulated subcellular localization and substrate specificity of all novel proteins is mandatory.

Two novel paralogs of GOX have been detected that are expected to catalyze a non-photorespiratory reaction and to be involved in α-oxidation of 3-methyl- and 2-hydroxy fatty acids, such as the intermediate 2-hydroxy-3-methyl-butanoate of Leu catabolism and D-2-hydroxy-8:0 of the degradation of ricinoleic acid (Gerhardt, 1993). Analysis of postulated targeting of AtHAOX1/2 to peroxisomes by the rare PTS SML> and their substrate specificity may yield awaited insights at the molecular level into peroxisomal α-oxidation of fatty acids in plants.

Complete degradation of complex fatty acids has been postulated to be accomplished in a hybrid pathway, involving enzymes with specificity for free fatty acids as well as those converting CoA esters (Gerhardt, 1993). Depending on their substrate specificity, several AAEs and thioesterases may be involved. The number of 15 largely unknown AAEs as well as 6 small thioesterases all with putative PTSs supports the idea that complex fatty acids can be oxidized in peroxisomes by a sequential action of thioesterases, hydroxy acid oxidases, and AAEs. By contrast, our data according to which GDSL lipases play a role in peroxisomal lipid metabolism raises currently more questions than it provides answers. Much remains to be learned about unusual input and output metabolites of peroxisomal β-oxidation and their catabolic pathways.

Apart from metabolic enzymes and peroxisome biogenesis proteins (PEX proteins), our knowledge on the presence and function of further peroxisomal matrix proteins is rather limited. Evidence for the existence of heat shock proteins, kinases, and phosphatases in peroxisomes, however, is supported by the detection of homologs with a putative PTS in this study. In addition of an HSP70 and a DnaJ homolog (HSP40; Wimmer et al., 1997; Diefenbach and Kindl, 2000), two members of the large plant family of small HSPs (Scharf et al., 2001) may be targeted to plant peroxisomes and help refolding of heat-labile matrix proteins. A few protein kinases have recently been localized to plant peroxisomes (Yang and Poovaiah, 2002; Dammann et al., 2003; Fukamatsu et al., 2003; Fukao et al. 2003). Our screen of the Arabidopsis genome reveals seven additional putative protein kinases that are suspected to mediate signal transduction across the peroxisomal membrane and to alter the activity of peroxisomal enzymes by reversible phosphorylation. In line with this result many proteins from leaf peroxisomes seem to be posttranslationally modified, of which some modifications most likely represent reversible phosphorylations (L. Babujee and S. Reumann, unpublished data). Accumulating evidence also suggests that turnover of peroxisomal enzymes is highly regulated, for instance under transition of glyoxysomes to leaf peroxisomes and vice versa during senescence. Whether a homolog of the Lon protease, a zinc protease, or a glutathione S-transferase is active in peroxisomes and involved in orchestrated degradation of matrix proteins needs to be elucidated. The considerable number of partly annotated unknown proteins not mentioned in this article as well as the 22 completely unknown Arabidopsis proteins with a major PTS lacking significant
The Potential of a Genomics Approach

Even though proteome approaches, in which the powerful resolution of two-dimensional gel electrophoresis is combined with high-sensitivity mass spectrometry, allowed considerable advances in the area of “organelle proteomics” and led to the identification of many unknown proteins (Peltier et al., 2000; Fukao et al., 2002, 2003; Heazlewood et al., 2004), detection of most regulatory and inducible proteins of small, fragile, and low-abundance cell organelles is restricted to an in silico genomics approach. Indeed, for many of our Arabidopsis proteins with a putative PTS, corresponding ESTs are only detected in collections from plants subjected to specific stress conditions. About 15% of the extracted proteins are still hypothetical, some of which may be induced under currently unknown conditions or in highly distinct plant organs.

Plant homologs of peroxisomal proteins from fungi and mammals can often be detected by homology analysis using general protein databases. The main advantages of our genome screen and the compilation of all available information on putative peroxisomal matrix proteins in a user-friendly html-based database are: (1) identification of plant-specific peroxisomal proteins; (2) recognition of distantly related homologs (e.g. small thioesterases); (3) detection of proteins the function of which has not yet been connected to the organelle of interest (e.g. Lon protease, GDSL lipases); (4) coverage of the major portion of soluble matrix proteins from plant peroxisomes; and (5) assembly of, at first glance, apparently unrelated metabolic mosaic pieces to novel metabolic pathways. The latter is illustrated by the detection of Arabidopsis proteins with a putative PTS that are homologous to prokaryotic enzymes of a novel prokaryotic pathway for degradation of aromatic fatty acids (Benning et al., 1998; Ferrandez et al., 1998; Thoden et al., 2003). Close plant homologs of the three enzymes involved in degradation of 4-CI-benzoate, i.e. an acyl-activating enzyme, a monofunctional enoyl-CoA hydratase, and a small thioesterase, belong to Arabidopsis protein families with a considerable number of members with high peroxisome targeting probability. If prokaryotes require specific isoforms for β-oxidation-related reactions of these aromatic compounds, specific isoforms need to be postulated for higher plants as well. To which extend aromatic and cyclic compounds, such as aromatic amino acids, chinones, and the plant hormones auxin, SA, and JA are degraded in the peroxisomal compartment, is currently largely unknown and requires experimental investigation, considering the influence of their turnover rate on cytoplasmic hormone concentrations and the induction of short-term signaling cascades. Peroxisomal AAEs and small thioesterases, both specific for aromatic and cyclic substrates also need to be postulated for biosynthesis of auxin, SA, and JA and are probably PTS-carrying members of the protein families presented here. It is also possible that the two isoforms of MFP are not suitable for conversion of aromatic and cyclic intermediates and that peroxisomal isoforms of monoclonal enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase catalyze these reactions. Analysis of the corresponding Arabidopsis T-DNA insertion k.o. mutants may yield interesting insights into the role of peroxisomes in the regulation of plant hormone levels.

CONCLUSIONS

We have screened the Arabidopsis genome for novel proteins of plant peroxisomes carrying putative PTSs. These proteins have been compiled in the database AraPerox and have been supplemented by known and putative Arabidopsis PEX homologs of yeast and mammalian proteins involved in peroxisome biogenesis (Charton and López-Huertas, 2002) and by membrane-associated enzymes as well as membrane proteins involved in metabolite transport. Detailed information on the prediction of subcellular localization is provided to sort out peroxisomal from non-peroxisomal proteins and to design straight-forward strategies for experimental validation of postulated protein targeting. In addition, quick access to analytical in silico data on homology and gene expression as well as literature on experimentally studied homologs is provided and expected to facilitate the identification of new functions and complex biochemical pathways of plant peroxisomes.

MATERIALS AND METHODS

Identification of Arabidopsis Proteins with a Putative PTS1 or PTS2

Genes encoding Arabidopsis proteins with putative PTS1 or PTS2 were identified using the software PatternSearch at the Arabidopsis Information Resource (TAIR) server (www.arabidopsis.org). Major and minor PTS1 tripeptides and PTS2 nonapeptides have been defined in a previous publication (Reumann, 2004). The localization of the PTS2 nonapeptide (R of position 1 of the nonapeptide) was restricted to position 2 to 30 of the N-terminal domain based on the analysis of about 150 plant homologs of PTS-targeted proteins (Reumann, 2004). ORF prediction was verified by homology analysis against the nonredundant database at the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov/). Homologous plant ESTs were retrieved and analyzed as described (Reumann, 2004). Homologous sequences were aligned and bootstrap analysis performed using ClustalX. Treeview was used for graphical presentation of phylogenetic analyses. Molecular mass and pIs were either taken from the information file at the TAIR server or, in cases of obvious mistakes, calculated using the Expasy server (www.expasy.org/tools/pi_tool.html). Targeting, expression, and homology analyses were updated in January and February 2004 and gene prediction, annotation, and expression of hypothetical proteins in April 2004.

Targeting Analysis

For targeting prediction the following programs were applied: TargetP (www.cbs.dtu.dk/services/TargetP/; Emanuelsson et al., 2000), PSORT, PSORTII, iPSORT (www.psort.org/; Nakai and Kanehisa, 1991; Horton and Nakai, 1997; Bannai et al., 2002), Predotar (www.inra.fr/predotar/), Mitoprot (www.mips.biochem.mpg.de/cgi-bin/proj/medgen/mitofinder; Claro and...
Expression Analysis by Digital Northern

The nucleotide sequence of the protein of interest was retrieved using ENTREZ (www.ncbi.nlm.nih.gov/Entrez/). The EST database was searched for the presence of corresponding Arabidopsis ESTs using TBLASTN (www.ncbi.nlm.nih.gov/BLAST/). Expressed sequence tags with a minimum length of 150 bp and a sequence identity of \( \sim 98\% \) were considered matches to the protein of interest. For analysis of GOX expression, \( \sim 97\% \) sequence identity at the nucleotide level was chosen for A_GOX/2.

ACKNOWLEDGMENTS

We thank Katharina Pawlowski, Martin Fulda, Ivo Feussner, and Hans-Walter Heldt for stimulating discussions and suggestions on the manuscript. The support of our research project by Hans-Walter Heldt and Ivo Feussner is appreciated. The help of Carolin Farke and Caroline Mayer in setting up the database AraPerox, citation of this article is appreciated. The support of our research project by Hans-Walter Heldt and Ivo Feussner is acknowledged. If gene cloning and functional analysis of peroxisomal proteins is based in major parts on information provided by the database AraPerox, citation of this article is appreciated.

Received March 28, 2004; returned for revision June 14, 2004; accepted June 16, 2004.

LITERATURE CITED


Mullen RT, Lee MS, Flynn CR, Trelace RN (1997a) Diverse amino acid residues function within the type 1 peroxisomal targeting signal. Implications for the role of accessory residues upstream of the type 1 peroxisomal targeting signal. Plant Physiol 115: 881–889


van Roermund CW, Elgersma Y, Singh N, Wanders RJ, Tabak HF (1995) The membrane of peroxisomes in Saccharomyces cerevisiae is impermeable to NAD(H) and acetyl-CoA under in vivo conditions. EMBO J 14: 3480–3486


