

# Analysis of the Plastidic *phosphate translocator* Gene Family in *Arabidopsis* and Identification of New *phosphate translocator*-Homologous Transporters, Classified by Their Putative Substrate-Binding Site<sup>1</sup>

Silke Knappe, Ulf-Ingo Flügge, and Karsten Fischer\*

Botanisches Institut der Universität zu Köln, Lehrstuhl II, Gyrhofstrasse 15, D-50931 Cologne, Germany

Analysis of the *Arabidopsis* genome revealed the complete set of *plastidic phosphate translocator* (*pPT*) genes. The *Arabidopsis* genome contains 16 *pPT* genes: single copies of genes coding for the triose phosphate/phosphate translocator and the xylulose phosphate/phosphate translocator, and two genes coding for each the phosphoenolpyruvate/phosphate translocator and the glucose-6-phosphate/phosphate translocator. A relatively high number of truncated *phosphoenolpyruvate/phosphate translocator* genes (six) and *glucose-6-phosphate/phosphate translocator* genes (four) could be detected with almost conserved intron/exon structures as compared with the functional genes. In addition, a variety of *PT*-homologous (*PTh*) genes could be identified in *Arabidopsis* and other organisms. They all belong to the drug/metabolite transporter superfamily showing significant similarities to nucleotide sugar transporters (NSTs). The *pPT*, *PTh*, and NST proteins all possess six to eight transmembrane helices. According to the analysis of conserved motifs in these proteins, the *PTh* proteins can be divided into (a) the lysine (Lys)/arginine group comprising only non-plant proteins, (b) the Lys-valine/alanine/glycine group of *Arabidopsis* proteins, (c) the Lys/asparagine group of *Arabidopsis* proteins, and (d) the Lys/threonine group of plant and non-plant proteins. None of these proteins have been characterized so far. The analysis of the putative substrate-binding sites of the *pPT*, *PTh*, and NST proteins led to the suggestion that all these proteins share common substrate-binding sites on either side of the membrane each of which contain a conserved Lys residue.

Plastids, typical plant organelles, arose by endosymbiosis of a cyanobacterial-like prokaryotic cell (Schimper, 1883; McFadden, 1999). Engulfment of the cyanobacterial cell generated a plastid that is surrounded by two membranes, the outer and inner envelope membranes. During evolution, more than 95% of the cyanobacterial genes were subsequently lost or transferred to the nucleus of the host cell (Martin and Herrmann, 1998; Martin et al., 1998). These nuclear-encoded plastidic proteins acquired an N-terminal extension (the transit peptide) that directs the attached protein to the plastids. One of the first processes to establish endosymbiosis was, besides the development of the protein import apparatus, the insertion of transport proteins into the envelope membranes to tap photosynthates, e.g. phosphorylated sugars and amino acids (Cavalier-Smith, 2000) and to connect the metabolism of the endosymbiont and the host cell.

Only a small number of envelope transporters have been characterized at the molecular level so far. These include two dicarboxylate translocators that are involved in ammonia assimilation (Weber and Flügge, 2002); a putative hexose transporter that

exports hexoses, the product of hydrolytic starch degradation (Weber et al., 2000); an ADP/ATP translocator that supplies plastids with energy for biosynthesis of starch, fatty acids, and other compounds (Neuhaus et al., 1997); and a H<sup>+</sup>/P<sub>i</sub> symporter (Pht2;1) that affects P<sub>i</sub> allocation within the plant (Daram et al., 1999; Versaw and Harrison, 2002).

The triose phosphate/phosphate translocator (TPT) was the first plastidic transporter that has been characterized at the molecular level (Flügge et al., 1989). It belongs to a group of plastidic phosphate translocators (*pPT*) that function as antiport systems using inorganic phosphate and phosphorylated C<sub>3</sub>, C<sub>5</sub>, or C<sub>6</sub> compounds as counter substrates (Flügge, 1999). Under physiological conditions, the substrates are transported via a strict 1:1 counter exchange. Transport proceeds via a ping-pong type of reaction mechanism, i.e. the first substrate is transported across the membrane before the second substrate can be bound and transported (Flügge, 1992). In its functional form, *pPT* proteins are dimers composed of two identical subunits (Wagner et al., 1989). In this respect, the *pPT*s differ from other transporters of the plastid envelope membrane, which function as monomers that contain 12 transmembrane helices (Neuhaus et al., 1997; Weber et al., 2000; Weber and Flügge, 2002). However, the *pPT* structures resemble the mitochondrial transporter superfamily without any significant similarity on the DNA or protein level (Walker and Runswick, 1993).

<sup>1</sup> This work was supported by the Deutsche Forschungsgemeinschaft.

\* Corresponding author; e-mail karsten.fischer@uni-koeln.de; fax 49-221-470-5039.

Article, publication date, and citation information can be found at [www.plantphysiol.org/cgi/doi/10.1104/pp.016519](http://www.plantphysiol.org/cgi/doi/10.1104/pp.016519).

The pPT proteins can be classified into four different subfamilies based on their substrate specificities and their sequence similarities. The TPT mediates the export of fixed carbon in the form of triose phosphates and 3-phosphoglycerate (3-PGA) from chloroplasts to the cytosol (Fliege et al., 1978; Flügge et al., 1989; Flügge, 1999) and thus represents the day path of carbon. In the cytosol, triose phosphates are used for the biosynthesis of Suc and other metabolites. Analysis of transgenic plants with a reduced activity of the TPT showed that the lack of triose phosphate export for cytosolic Suc biosynthesis can be compensated for by an accelerated starch turnover and export of neutral sugars from the stroma (Häusler et al., 1998; Schneider et al., 2002).

The phosphoenolpyruvate (PEP)/phosphate translocator (PPT) accepts PEP and 2-PGA as substrates, i.e. C3 compounds that are phosphorylated at C-atom 2, whereas triose phosphates and 3-PGA are not transported (Fischer et al., 1997). The physiological function of the PPT in C<sub>3</sub> plants is to supply plastids with PEP for fatty acid synthesis and, more importantly, the shikimic acid pathway (Fischer et al., 1997), which leads to the synthesis of aromatic amino acids and a large number of secondary metabolites (Herrmann and Weaver, 1999). A *PPT1* knock-out mutant shows a reticulate leaf phenotype and is unable to produce anthocyanins as a product of secondary plant metabolism (Streatfield et al., 1999).

The Glc-6-phosphate (Glc-6-P)/phosphate translocator (GPT), representing the third subfamily of pPTs, shows the broadest substrate specificity accepting phosphorylated C3 (triose phosphates, 3-PGA), C5 (xylulose-5-phosphate, Xul-5-P), and C6 compounds (Glc-6-P; Kammerer et al., 1998; Eicks et al., 2002). In heterotrophic tissues, the GPT mediates the uptake of carbon in the form of Glc-6-P into plastids, where it serves as substrate for starch synthesis, fatty acid synthesis, or the oxidative pentose phosphate pathway (Borchert et al., 1989; Bowsher et al., 1992; Flügge, 1999). Analysis of starchless mutant lines that are deficient in the plastidic phosphoglucomutase (catalyzing the interconversion of Glc-6-P and Glc-1-P), led to the conclusion that in most plants, Glc-6-P is the sole precursor for starch synthesis (Harrison et al., 2000; Kofler et al., 2000).

The Xul-5-P/phosphate translocator (XPT) represents the fourth subfamily of pPTs. The XPT shows a similar substrate specificity as the GPT but does not transport Glc-6-P (Eicks et al., 2002). The proposed function of the XPT is to provide the plastidic pentose phosphate pathways with cytosolic carbon skeletons in the form of Xyl-5-P, especially under conditions of high demand for intermediates of the cycles.

In the last decade, cDNAs encoding pPT proteins of all four subfamilies have been isolated and sequenced (Flügge et al., 1989; Fischer et al., 1994, 1997; Kammerer et al., 1998; Eicks et al., 2002). Members of a distinct subfamily share a high degree of identical

amino acids with each other (>80%), whereas identities between the members of the subfamilies are only approximately 35%, with the exception of XPTs and GPTs that show a higher degree of identity (50%) with each other (Kammerer et al., 1998; Eicks et al., 2002).

Little is known about the number of *pPT* genes in Arabidopsis and other plants, the structure of these genes, and their evolution. The complete sequence of the Arabidopsis genome (Arabidopsis Genome Initiative, 2000) opened the way to address these questions. Here, we present the complete set of *pPT* genes and *PT*-homologous (*PTh*) genes of Arabidopsis and their structures and phylogenetic relationship in comparison with *pPT* genes of other plants and organisms.

## RESULTS

### Structure and Expression of *pPT* Genes in Arabidopsis

First, we addressed the question how many *pPT* genes are present in the Arabidopsis genome. Sequences of cDNAs coding for members of the four pPT subfamilies were used to conduct BLAST searches (BLASTP and TBLASTN) against the genome sequence of Arabidopsis (Altschul et al., 1990). A total of 16 genes encoding pPT proteins were found, 10 of which were probably pseudogenes. Mapping of all Arabidopsis *pPT* genes showed that these sequences are scattered throughout all five chromosomes (Table I). Although both the AtTPT (at5g46110) and the AtXPT (at5g17630) are encoded by single genes, small gene subfamilies were identified for AtPPTs and AtGPTs.

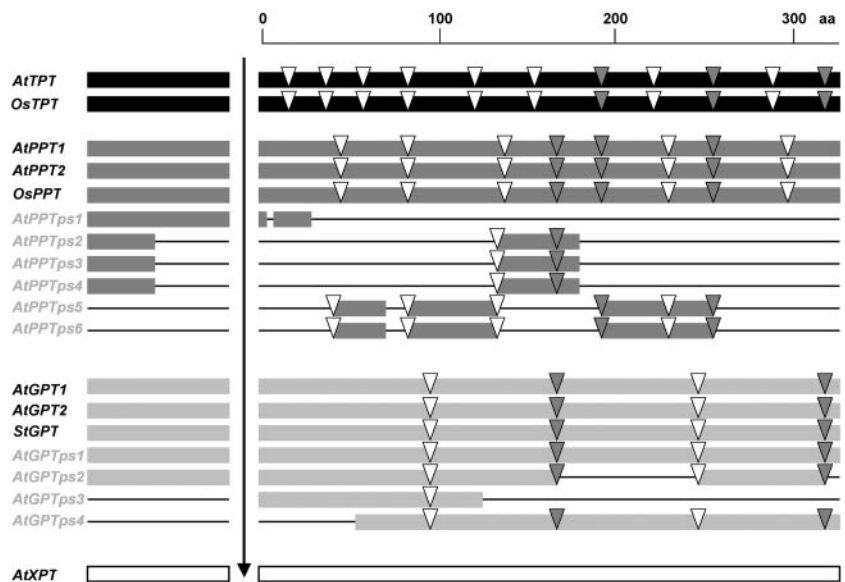
The coding region of the *AtTPT* gene is interrupted by 11 introns (Fig. 1). The structure of the *TPT* gene of Arabidopsis is identical to the ortholog of rice

**Table I.** The *PT* genes and pseudogenes (*PTps*)

The corresponding clones, the chromosomal localization, the AGI genome codes, and the accession numbers are listed.

| Gene            | Clone  | Chromosome | AGI Genome Codes | GenBank Accession No. |
|-----------------|--------|------------|------------------|-----------------------|
| <i>AtTPT</i>    | MCL19  | 5          | At5g46110        | AAC83815              |
| <i>AtGPT1</i>   | MGB8   | 5          | At5g54800        | BAB08759              |
| <i>AtGPT2</i>   | F8K4   | 1          | At1g61800        | AAC28500              |
| <i>AtPPT1</i>   | F19N2  | 5          | At5g33320        | BAB10483              |
| <i>AtPPT2</i>   | F4P13  | 3          | At3g01550        | AAF01540              |
| <i>AtXPT</i>    | K10A8  | 5          | At5g17630        | CAC01907              |
| <i>AtGPTps1</i> | T24M8  | 4          | –                | –                     |
| <i>AtGPTps2</i> | F26C17 | 5          | –                | –                     |
| <i>AtGPTps3</i> | MMG15  | 3          | –                | –                     |
| <i>AtGPTps4</i> | T23J18 | 1          | –                | –                     |
| <i>AtPPTps1</i> | F25G13 | 4          | –                | –                     |
| <i>AtPPTps2</i> | T32A11 | 3          | –                | –                     |
| <i>AtPPTps3</i> | K12G2  | 5          | –                | –                     |
| <i>AtPPTps4</i> | F1121  | 1          | –                | –                     |
| <i>AtPPTps5</i> | MJ16   | 3          | –                | –                     |
| <i>AtPPTps6</i> | T1D16  | 2          | –                | –                     |

**Figure 1.** Exon/intron structure of *pPT* genes and pseudogenes of Arabidopsis, rice, and potato. Alignment of deduced amino acid (aa) sequences of the genes depicting their structure within and between the *pPT* families. Homologous regions are indicated by bars (*TPTs*, black bars; *PPTs*, dark gray bars; *GPTs*, light gray bars; and *XPT*, white bar). Interrupted bars in the pseudogenes indicate lack of amino acid homology in comparison with the particular functional gene. The arrow marks the processing sites. Intron positions are indicated by triangles. Gray triangles indicate conserved intron positions. Accession numbers for genes from rice and potato are: *OsTPT*, OSJNBa0010K01; *OsPPT*, P0583G08; and *StGPT*, AY163867).



(*Oryza sativa*; BAC clone OSJNBa0010K01, chromosome 1) with all but one intron localized at exactly the same position in the DNA sequence. Only intron 2 has undergone a shift of 2 bp in rice compared with Arabidopsis.

The *PPT* gene subfamily consists of eight genes, only two of them representing full-length genes (*AtPPT1*, at5g33320; and *AtPPT2*, at3g01550) and six being truncated genes (*AtPPTps1–6*). The truncated *AtPPTps* genes split into three different classes according to their structure (Fig. 1). All truncated genes show high identities to *AtPPT1* but low identities to *AtPPT2* (Table II). However, the similarity to *AtPPT1* is restricted to the exons, whereas the introns show no sequence identities. In contrast, the high sequence identity between the truncated genes of one structural class covers both exons and introns. Three genomic *PPT* clones from tobacco (*Nicotiana tabacum*; A. Weber, unpublished data) and a *PPT* gene from rice (P0583G08) show completely conserved intron-exon structures compared with the Arabidopsis *PPT* genes and are highly similar to *AtPPT1*.

The *AtGPT* subfamily has six members, two of them representing functional genes (*AtGPT1*, at5g54800;

and *AtGPT2*, at1g61800) and four that are probably pseudogenes (*AtGPTsp1–4*) that all share about 60% to 90% identical amino acids (Table III; Fig. 1). The two *AtGPT* genes contain four introns at the same positions. The structures of two *GPT* genes from potato (*Solanum tuberosum*) that have been sequenced (Frank Ludwig, personal communication) are identical to the Arabidopsis orthologs.

The other four *GTP* genes contain several mutations leading to premature stop codons and to frame shifts, both preventing the synthesis of a functional protein. These genes were therefore considered to be nonfunctional pseudogenes (*AtGPT-pseudo1–4*).

Taken together, the structures of the different *pPT* genes turned out to be quite different between the four *pPT* subfamilies. The intron positions are surprisingly quite different between the three intron-containing subfamilies even if intron “slippage” of up to 12 bp is considered, whereas the *AtXPT* gene lacks any intron. Only three pairs of introns have the same position in different subfamilies, and no intron position is conserved between all three subfamilies (Fig. 1). However, the exon-intron structure of each

**Table II.** Amino acid identity of derived amino acid sequences of *pPT* genes and pseudogenes in the *PPT* subfamily

The values indicate the percentage of amino acid identity between protein sequences. The lengths of the putative proteins (no. of amino acid residues) are given in parentheses.

|                 | <i>AtPPT1</i><br>(410) | <i>AtPPT2</i><br>(380) | <i>AtPPTps1</i><br>(106) | <i>AtPPTps2</i><br>(89) | <i>AtPPTps3</i><br>(89) | <i>AtPPTps4</i><br>(90) | <i>AtPPTps5</i><br>(135) | <i>AtPPTps6</i><br>(169) |
|-----------------|------------------------|------------------------|--------------------------|-------------------------|-------------------------|-------------------------|--------------------------|--------------------------|
| <i>AtPPT1</i>   |                        | 52                     | 88                       | 89                      | 90                      | 88                      | 89                       | 90                       |
| <i>AtPPT2</i>   |                        |                        | 35                       | 29                      | 30                      | 30                      | 55                       | 55                       |
| <i>AtPPTps1</i> |                        |                        |                          | 76                      | 76                      | 71                      | –                        | –                        |
| <i>AtPPTps2</i> |                        |                        |                          |                         | 99                      | 96                      | –                        | –                        |
| <i>AtPPTps3</i> |                        |                        |                          |                         |                         | 97                      | –                        | –                        |
| <i>AtPPTps4</i> |                        |                        |                          |                         |                         |                         | –                        | –                        |
| <i>AtPPTps5</i> |                        |                        |                          |                         |                         |                         |                          | 86                       |

**Table III.** Amino acid identity of derived amino acid sequences of *pPT* genes and pseudogenes in the *GPT* subfamily

The values indicate the percentage of amino acid identity between protein sequences. The lengths of the putative proteins (number of amino acid residues) are given in parentheses.

|          | AtGPT1<br>(388) | AtGPT2<br>(388) | AtGPTps1<br>(387) | AtGPTps2<br>(309) | AtGPTps3<br>(317) | AtGPTps4<br>(265) |
|----------|-----------------|-----------------|-------------------|-------------------|-------------------|-------------------|
| AtGPT1   |                 | 74              | 65                | 59                | 67                | 79                |
| AtGPT2   |                 |                 | 70                | 66                | 72                | 80                |
| AtGPTps1 |                 |                 |                   | 67                | 74                | 92                |
| AtGPTps2 |                 |                 |                   |                   | 67                | 77                |
| AtGPTps3 |                 |                 |                   |                   |                   | 84                |

subfamily is conserved in different plants, e.g. Arabidopsis, potato, tobacco, and rice.

To answer the question of whether the truncated genes are expressed, reverse transcriptase-PCR assays were performed with RNA from whole Arabidopsis plants to analyze the transcript levels of all *pPT* genes. Figure 2 shows that only two putative pseudogenes (*AtGPTps2* and *AtGPTps3*) are expressed, whereas no signals were detected for the other pseudogenes.

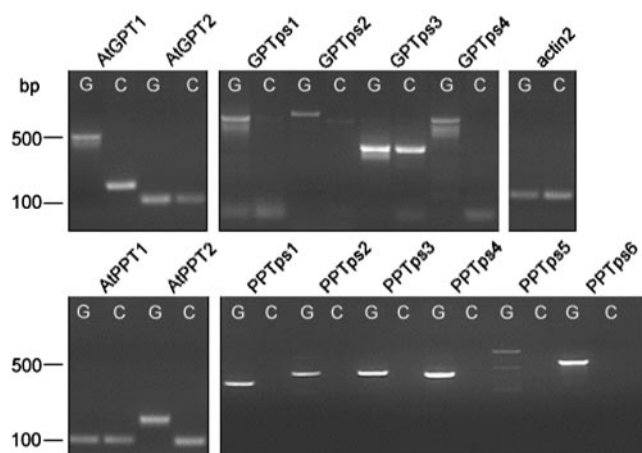
#### The *pPT* Genes Are Part of the Drug/Metabolite Transporter (DMT) Superfamily

BLAST searches against the Arabidopsis genome sequence and entries in GenBank using *pPT*-cDNAs revealed 28 homologous membrane proteins (PTh), which shared about 20% to 25% identical amino acids with the *pPT* proteins. They are all part of the DMT superfamily, which consists of 14 different families including the family of *pPT*s (Jack et al., 2001) and four families of nucleotide sugar transporters (NSTs). Because the *pPT*s are more related to the NSTs than to the other families of the DMT superfamily, Ward

(2001; <http://www.cbs.umn.edu/Arabidopsis>) combined these proteins in the TPT/NST-family.

All PTh protein sequences in Arabidopsis and some examples of characterized *pPT*- and NST- sequences from Arabidopsis and other organisms were used for the construction of a phylogenetic tree (Fig. 3). This comparison shows that these proteins can be split into different families of transport proteins. Only some of them have been functionally characterized, yet one family consists of *pPT*s, and the other families consist of NSTs of endoplasmic reticulum (ER) and Golgi membranes that transport UDP-glucuronic acid, GDP-Man, and other nucleotide sugars from the cytosol into the lumen of the ER and the Golgi apparatus (Baldwin et al., 2001; Lübke et al., 2001; Lühn et al., 2001). The family that is most similar to the *pPT* family consists of some uncharacterized proteins from animals and fungi (KR family). The PTh proteins from Arabidopsis obviously split into three new families, the KV/A/G, KT, and KD groups (see below).

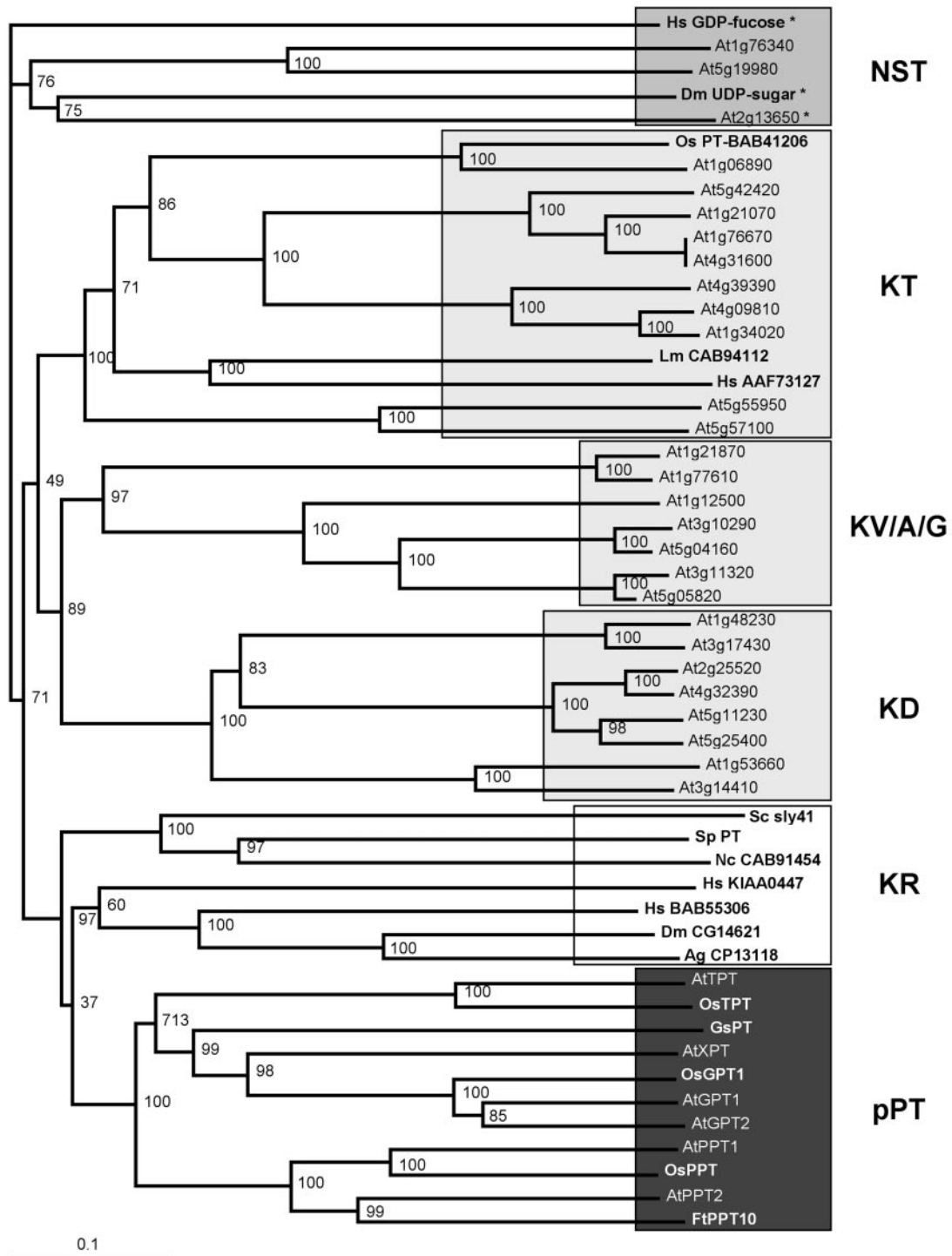
To assess the number of transmembrane spans in the *pPT* and PTh proteins, the *pPT* proteins from different plants and the PTh proteins from Arabidopsis were screened for membrane-spanning helices using six different algorithms (Schwacke et al., 2003; TmHMM 2.0, <http://www.cbs.dtu.dk/services/TmHMM/>; Hmmtop2.0, <http://www.enzim.hu/hmmtop/>; SosuiG1.1, <http://sosui.proteome.bio.tuat.ac.jp/>; TMPred, [http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html); TMap, <http://www.mbb.ki.se/tmap/>; TopPred2.0, <http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html>). Because the number of predicted membrane-spanning helices of a distinct protein differs significantly between the individual programs, the statistical median was calculated for each protein (Schwacke et al., 2003). The *pPT* proteins turned out to possess seven to eight membrane-spanning helices (Fig. 4), and the number of membrane spans in the PTh proteins was calculated to be six to eight.



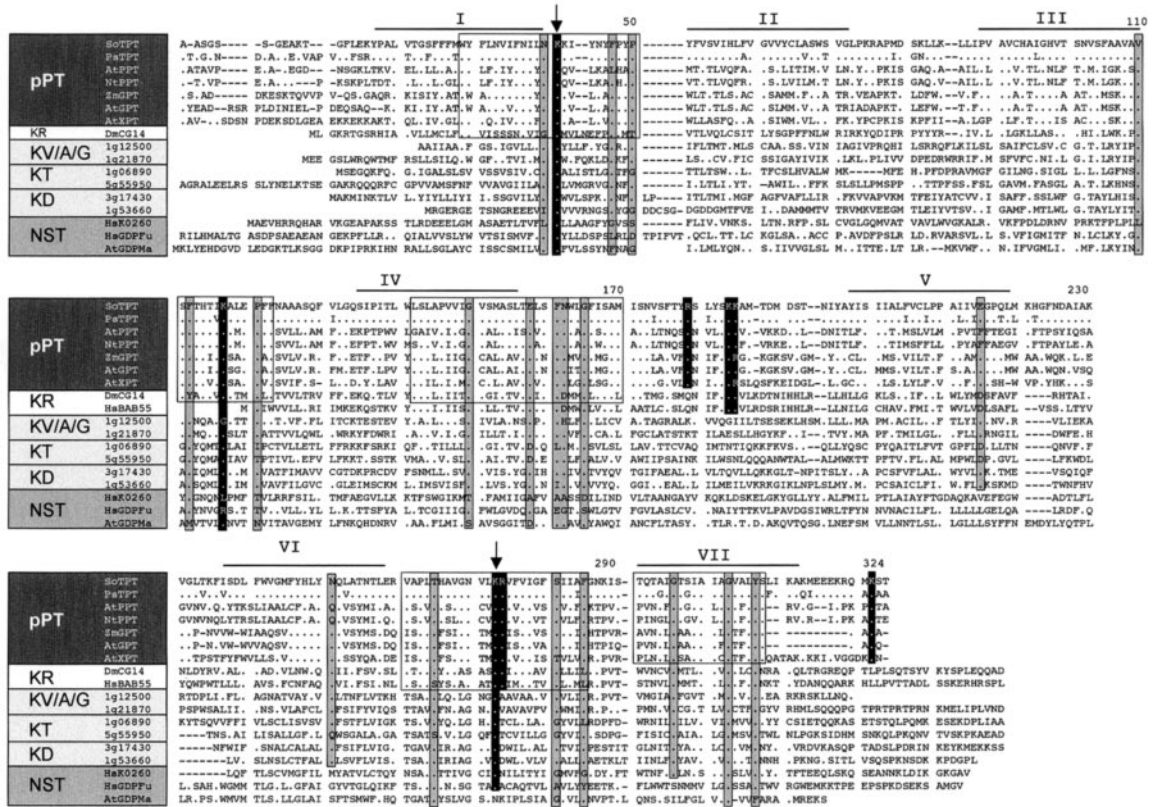
**Figure 2.** Expression of *PT* genes and pseudogenes. Specific primer combinations were used to amplify genomic DNA (G) and cDNA (C) from Arabidopsis. The numbers (500 and 1,000) indicate the length in base pairs. Differences in the length of G and C for one gene are due to intron sequences within the genomic DNA.

#### The PTh Proteins Can Be Divided into Different Groups

As mentioned above, the members of a distinct subfamily of *pPT*s share a high degree of identity



**Figure 3.** Phylogenetic tree of pPT, NST, and PTh families constructed by the neighbor-joining method. Amino acid sequences were aligned and the tree was created based on corrected distances using the programs ClustalX (v1.81) and TreeView v1.1.6. The numbers at the branches of the tree are bootstrap values (percentage; 1,000 repeats were performed). The first two letters of each sequence represent the organism: Ag, *Anopheles gambiae*; At, Arabidopsis; Dm, fruitfly (*Drosophila melanogaster*); Hs, human (*Homo sapiens*); Lm, *Leishmania major*; Ft, *Flaveria trinervia*; Gs, *Galdieria sulfuraria*; Nc, *Neurospora crassa*; Os, rice; Sc, Brewer's yeast (*Saccharomyces cerevisiae*); Sp, Fission yeast (*Schizosaccharomyces pombe*). At numbers are indicated except for the pPTs from Arabidopsis (see Table I). Accession numbers not mentioned within the tree are the following: HsGDP-Fuc transporter (Q96A29), Dm-UDP sugar transporter (Q95Y15), Sc-sly41 (CAA38144), SpPT (CAB36873), OsTPT (Bac clone OSJNBa0010K01), OsPPT (AAK51561), and GsPT (phosphate translocator of unknown substrate specificity; A. Weber, unpublished data). Proteins from organisms other than Arabidopsis are marked in bold/italics. Already characterized proteins are marked by an asterisk.



**Figure 4.** Alignment of pPT, PTh, and NST amino acid sequences. The TPT sequence from spinach (*Spinacia oleracea*; SoTPT, accession no. CAA32016) was aligned with sequences of TPT from pea (*Pisum sativum*; PsTPT, accession no. CAA48210), PPT1 from Arabidopsis (AtPPT1, At5g33320) and tobacco (NtPPT, accession no. AAB40648), GPT from maize (*Zea mays*; ZmGPT, accession no. AF020813) and Arabidopsis (AtGPT, accession no. At5g54800), XPT from Arabidopsis (AtXPT, accession no. At5g17630), two proteins from fruitfly (*DmCG14*, accession no. AAF50956) and human (HSBAB55, accession no. BAB55306) representing the KR family, six proteins from Arabidopsis (At1g12500, At1g21870, At1g06890, At5g5950, At3g17430, and At1g53660) that are members of the KV/A/G, KT, and KD families, the human UDP-glucuronic acid transporter (HsK0260, accession no. BAB18586), the human GDP-Fuc transporter (HsGDPFu, accession no. Q96A29), and the GDP-Man transporter from Arabidopsis (AtGDPMa, accession no. At2g13650). Identities of amino acid residues are indicated by dots. The numbers refer to the amino acid position in the SoTPT sequence. The locations of seven transmembrane helices are indicated by solid lines (I-VII), whereas an eighth potential membrane-spanning region is indicated by a dashed line. Five regions of high similarities between the pPT proteins were boxed (white boxes). Lys and Arg residues that are conserved in all pPT proteins are marked by black boxes, whereas K41 and K273, which are probably involved in substrate binding, are marked by an arrow. Other residues that are conserved in most families of the TPT/NST superfamily are marked by gray boxes.

with each other but only 35% to 40% identical amino acids with members of the other three subfamilies. However, all pPT proteins contain five regions of remarkable high similarity (Kammerer et al., 1998; Eicks et al., 2002; Fig. 4, white boxes). Because the substrates transported by the pPT proteins are bound as divalent anions (Fliege et al., 1978), positively charged amino acid residues, i.e. Lys and/or Arg residues, should be involved in binding and transport. A comparison of sequences of the pPTs revealed that six Lys residues and two Arg residues are conserved in all pPT proteins (Fig. 4, black boxes). Three of the Lys residues (K41, K117, and K273); the numbers refer to the position in the SoTPT sequence) and one Arg residue (R274) are located in regions of high similarity.

Extending these considerations to the PTh proteins, it is shown that two of the Lys residues (K41 and K273) are conserved in all proteins (Fig. 4, black boxes marked by an arrow). K273 is part of the fourth similarity box. Two other residues are well conserved in this box, namely T265 and F285 (motif T[X]<sub>7</sub>K[X]<sub>11</sub>F). The second conserved Lys residue, K41, is located in the first similarity box and is part of the conserved sequence motif NK[X]<sub>7</sub>F. Both motifs reside in hydrophilic loops of the protein. In contrast, the two Arg residues are not conserved in the PTh proteins. However, amino acid position 274 also seems to play a significant role in transport because particular residues can be found at this position that can be used to divide the PTh proteins into different groups (Figs. 3 and 4): (a) In a group of PTh proteins

from animals and fungi, an Arg residue is located at position 274, as is the case for the pPTs (KR-group). (b) In one group of PTh proteins from Arabidopsis, the amino acid downstream of K273 can be a Val, an Ala, or a Gly residue (KV/A/G-group, seven sequences). (c) In two additional groups, this position is occupied by either an Asp residue (KD-group, eight sequences from Arabidopsis) or (d) by a Thr residue (KT-group, nine sequences including one sequence from humans). Nothing is known about the physiological functions of the proteins of the KR, KV/A/G, KD, and KT groups so far.

In most of the NST proteins, K273 is also conserved (Fig. 4). However, in the GDP-Man transporter family, this position is occupied by an Asn residue, whereas a Lys residue is located at position 274.

Several other amino acid residues are conserved in all or most of the four protein groups and also in some NST families (Fig. 4, gray boxes). Especially the motive G[X]<sub>6</sub>G[X]<sub>3</sub>Y in the fifth region of high similarity, located in the last membrane-spanning region at the C terminus, is found in almost all families. Interestingly, most of the well-conserved residues are located in the five similarity boxes, suggesting an important role of these boxes in transport function.

#### Subcellular Localization of the Arabidopsis PTh Proteins

The pPT proteins are nuclear-encoded proteins that are synthesized in the cytosol with a N-terminal presequence. The proteins are posttranscriptionally inserted into the plastidic inner envelope membrane (Flügge et al., 1989; Flügge, 1999). All PTh proteins from Arabidopsis were examined by eight algorithms for N-terminal peptides targeting the proteins to plastids (and mitochondria; TargetP 1.0, <http://www.cbs.dtu.dk/services/TargetP/>; ChloroP 1.1, <http://www.cbs.dtu.dk/services/ChloroP/>; predotar 0.5, <http://www.inra.fr/predotar/>; iPSORT, <http://www.HypothesisCreator.net/iPSORT/>; SignalP\_NN\_v2 and SignalP\_HMM\_v2, <http://www.cbs.dtu.dk/services/SignalP-2.0/>; MitoProt\_v2, <http://www.mips.biochem.mpg.de/cgi-bin/proj/medgen/mitofilter/>; PCLR\_v0.9, <http://apicoplast.cis.upenn.edu/pclr>). Most of the PTh proteins do not possess any plastidic or mitochondrial targeting presequences, but several of these proteins contain cleavable N-terminal signal peptides that direct the proteins to the ER and the secretory pathway (Table IV). Only two proteins, the products of genes at1g12500 and at3g21090 have a plastid targeting sequence. Both are members of the KV/A/G family of PTh proteins.

## DISCUSSION

### Organization of the pPT Gene Family

The pPTs are a family of transport proteins that can be classified into four subfamilies based on their

substrate specificities, sequence similarities, and gene structures. In Arabidopsis, two subfamilies, the *TPT* and the *XPT*, consist of only one gene, whereas the *PPT* and *GPT* subfamilies have eight and six members, respectively. Remarkably, six *PPT* and four *GPT* genes seem to be pseudogenes (*AtPPTps1–6* and *AtGPTps1–4*). Nine of these genes, all six *PPT* and three *GPT* genes, are truncated versions lacking 20% to 80% of the coding region. Most of the truncated genes and also the *AtGPTps1* gene that has the same length and structure as the functional *GPT* genes contain mutations that introduce stop codons and frame shifts both leading to a premature stop of translation. We therefore consider these genes to be pseudogenes. We could show that only two of these genes, *AtGPTps2* and *AtGPTps3*, are transcribed. Whether or not these pseudogenes have any physiological function is not known. The high number of pseudogenes (62%) within the *pPT* family is unusual because in other gene families in Arabidopsis, their frequency is only 0% to 10%, with an average of 3% (Arabidopsis Genome Initiative, 2000). For example, in Arabidopsis, only 21 of 249 genes encoding ribosomal proteins and 5 of 40 genes coding for major intrinsic proteins are nonfunctional (Bakarati et al., 2001; Johanson et al., 2001).

The number of introns within a *pPT* subfamily is conserved but varies between different subfamilies. None of the intron positions is conserved in all subfamilies, but two identical positions between *TPT* and *PPT* genes and two positions between *PPT* and *GPT* genes have been found (Fig. 1). However, the structures of the *pPT* genes of different plant species are conserved within one subfamily, e.g. the structures of *TPT* genes from Arabidopsis and rice and of *GPT* genes from Arabidopsis and potato are identical. Similar results were reported in comparisons among four genes of unknown function. Their gene structure was largely conserved between genes from barley (*Hordeum vulgare*) and rice and to a lesser extent with homologous genes from Arabidopsis, with 42 of 53 intron positions conserved between Arabidopsis and the monocots (Dubcovsky et al., 2001).

The observation that pPT sequences from one subfamily are up to 85% identical at amino acid level between different plants whereas the similarity between members of different subfamilies within one plant is only about 35% indicates that the four pPT subfamilies already existed before angiosperms split into monocotyledons and dicotyledons. This also holds true for other gene families in Arabidopsis that have been analyzed recently. For example, the major intrinsic protein family can be divided into four different subfamilies based on their sequence similarities, gene structure, and substrate specificities. These four subfamilies and their specific exon-intron pattern evolved from a common ancestral gene before monocots and dicots diverged (Johanson et al., 2001). Because nothing is known about *pPT* genes from

**Table IV.** Predicted subcellular localization of pPT and PTh proteins

The putative localization was determined with different algorithms for all pPT and PT-homologous proteins. Pred. localization, Predicted localization; program, algorithms that predict the localization listed in pred. localization. A, TargetP\_v1; B, Predotar\_v0.5; C, iPSORT; D, ChloroP; E, SignalP\_NN\_v2; F, SignalP\_HMM\_v2; G, MitoProt\_v2; H, PCLR\_v0.9. mito, Mitochondrial targeting signal; sec, secretory pathway; Golgi, Golgi apparatus; plastid, plastid targeting signal. In cases where the predicted localization is validated by experimental data, the corresponding references are listed.

| Protein       | Pred. localization | Program | Reference                   |
|---------------|--------------------|---------|-----------------------------|
| <b>NST</b>    |                    |         |                             |
| Hs GDP-fucose | Mito               | A, B, G | Luhn et al. (2001)          |
| At1g76340     | No signal          |         |                             |
| At5g19980     | Sec                | A       |                             |
| Dm UDP-sugar  | Golgi              |         | Goto et al. (2001)          |
| At2g13650     | Golgi              |         | Baldwin et al. (2001)       |
| <b>KT</b>     |                    |         |                             |
| OsBAB41206    | Plastid or mito    | D/C     |                             |
| At1g06890     | Sec or plastid     | A, C/D  |                             |
| At5g42420     | Sec or plastid     | D/C     |                             |
| At1g21070     | Sec or plastid     | D/C     |                             |
| At1g76670     | Plastid or sec     | D/E     |                             |
| At4g31600     | No signal          |         |                             |
| At4g39390     | Plastid            | D       |                             |
| At4g09810     | Plastid or mito    | D/C     |                             |
| At1g34020     | No signal          |         |                             |
| LmCAB94112    | Mito               | G       |                             |
| HsAAF3127     | Mito               | B, G    |                             |
| At5g55950     | No signal          |         |                             |
| At5g57100     | Mito               | B, G    |                             |
| <b>KV/AG</b>  |                    |         |                             |
| At1g21870     | No signal          |         |                             |
| At1g77610     | Sec                | A       |                             |
| At1g12500     | Plastid            | D, H    |                             |
| At3g10290     | Plastid            | A, D, H |                             |
| At5g04160     | Mito or sec        | B, G/A  |                             |
| At3g11320     | Ambiguous          |         |                             |
| At5g05820     | Plastid or mito    | B/G     |                             |
| <b>KD</b>     |                    |         |                             |
| At1g48230     | Sec or mito        | A, E/G  |                             |
| At3g17430     | Sec or plastid     | A, C/D  |                             |
| At2g25520     | Sec or mito        | A, E/G  |                             |
| At4g32390     | Sec                | A, E    |                             |
| At5g11230     | Sec                | A, C, E |                             |
| At5g35400     | Sec                | A, E    |                             |
| At1g53660     | No signal          |         |                             |
| At3g14410     | Sec                | A, E    |                             |
| <b>KT</b>     |                    |         |                             |
| Sc sly41      | Mito               | B, C    |                             |
| Sp PT         | Mito or sec        | B, G/C  |                             |
| No CAB91454   | No signal          |         |                             |
| Hs KAA0447    | Mito               | C       |                             |
| Hs BAB55306   | Mito or sec        | B, G/C  |                             |
| Dm CG14621    | Mito               | B, G    |                             |
| Ag CP13118    | No signal          |         |                             |
| <b>pPT</b>    |                    |         |                             |
| AtTPT         | Plastid            |         | Flügge et al. (1989)        |
| OsTPT         | Plastid            |         | Flügge et al. (1989)        |
| GsPT          | Plastid            |         | A. Weber (unpublished data) |
| AtXPT         | Plastid            |         | Eicks et al. (2002)         |
| OsGPT1        | Plastid            |         | Kammerer et al. (1998)      |
| AtGPT1        | Plastid            |         | Kammerer et al. (1998)      |
| AtGPT2        | Plastid            |         | Kammerer et al. (1998)      |
| OsPPT         | Plastid            |         | Fischer et al. (1997)       |
| AtPPT2        | Plastid            |         | Fischer et al. (1997)       |
| AtPPT10       | Plastid            |         | Fischer et al. (1997)       |



gymnosperms, mosses, or algae, the structure of the ancestral *pPT* gene remains elusive.

### Duplications of *pPT* Genes

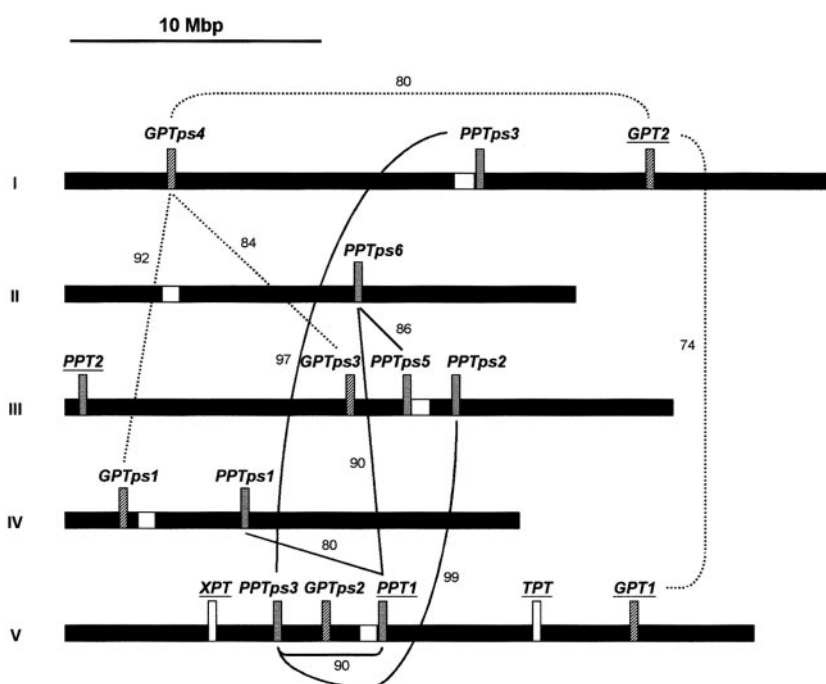
The Arabidopsis genome contains only 35% single-copy genes, whereas a high percentage of gene families with more than five members exist. Gene families arise from duplication of ancestral genes, followed by the divergence of both copies leading to a structural and functional specialization. Small-scale duplications, i.e. duplications of individual genes or groups of genes, often lead to genes arranged in tandem arrays. This kind of duplication plays a significant role in the Arabidopsis genome because 17% of all genes are arranged in tandem arrays (Arabidopsis Genome Initiative, 2000). However, roughly 60% of the Arabidopsis genome consists of large duplicated regions of 100 k $\text{B}$  or even larger (Arabidopsis Genome Initiative, 2000). Analysis of the Arabidopsis genome and comparative sequence analysis with segments of the genomes of other plants suggests that at least two rounds of large-scale duplication occurred in the lineage leading to Arabidopsis, most probably duplications of the whole genome (polyploidization; Blanc et al., 2000; Ku et al., 2000; Schmidt, 2002).

We therefore asked whether the *pPT*s arose by the mechanisms of small- or large-scale duplications. In the case of the *XPT* gene, a duplication mechanism could be proposed. Because the *XPT* gene is highly similar to the *GPT* genes, it may be derived from the duplication of one of the *GPT* genes in Arabidopsis. However, it lacks all introns occurring in the *GPT* genes. This “sudden” disappearance of multiple in-

trons might be explained by a retrotranscription from the *GPT* mRNA, followed by genome insertion. Several cases of such intronless genes within families of intron-containing genes have been reported in animals and plants, e.g. in the family of SET domain proteins, the glycosyl transferase family, the expansin group  $\beta$ 2, and the catalase family from Arabidopsis (Frugoli et al., 1998; Tavares et al., 2000; Baumbusch et al., 2001; Li et al., 2002). In plants, several sequences, e.g. copia- and gypsy-like retroelement sequences, were found, which can be the source of a reverse transcriptase activity (Wessler et al., 1995).

None of the other Arabidopsis *pPT* genes and pseudogenes are located in the duplicated regions identified so far ([http://mips.gsf.de/proj/thal/db/gv/rv/rv\\_frame.html](http://mips.gsf.de/proj/thal/db/gv/rv/rv_frame.html)) or are arranged in tandem arrays (Fig. 5). These data suggest that other mechanisms could be responsible for the high number of genes and pseudogenes. The structure of the *pPT* pseudogenes in the Arabidopsis genome resembles those of the functional genes, with the positions of the introns being the same as in their functional counterparts. In cases of the truncated pseudogenes, the preserved regions often correspond to exons of the functional genes (Fig. 1). In addition, the pseudogenes show a high degree of identity at the DNA sequence level with the functional *pPT* genes (66%–90%) only within the exons. However, the identities between the pseudogenes cover both exons and introns. This proposes a probable explanation for the occurrence of the pseudogenes. The duplication of the *PPT1* gene (from which all *PPT* pseudogenes derived) and one of the *GPT* genes led to two functional copies of both genes. The exons of both copies were preserved, whereas the introns diverged through the introduction of mu-

**Figure 5.** Localization of *pPT* genes and pseudogenes on the five chromosomes of Arabidopsis. A map of Arabidopsis chromosomes is shown. Centromeres are indicated by white boxes. *pPT* genes and clones containing pseudogenes are indicated by bars (gray bars, *PPT* gene family; striped bars, *GPT* gene family; and white bars, other members of the *pPT* gene family). Functional genes are underlined. Black lines connect *PPT* genes and pseudogenes of highest homology, whereas dotted lines connect the most closely related *GPT* genes and pseudogenes. Numbers at the lines represent the percentage of identity at amino acid levels.



tations. One of the copies was then disrupted by transposition of parts of the gene to a different chromosomal location followed by one or two duplications of the pseudogenes leading to the different pseudogene classes (Figs. 1 and 5). Both gene duplication and translocation of the *PPT* and *GPT* genes were more recent events.

The mechanism of duplication/translocation of the *PPT* and *GPT* pseudogenes are not clear. Transposons account for at least 10% of the genome of Arabidopsis. Class I elements (retrotransposons) primarily occupy the centromere and the pericentromeric regions (Copenhaver et al., 1999), whereas functional genes are thought to be relatively rare in centromeric regions of higher eukaryotes. Expressed genes in the centromere region of chromosome 5 (CEN5) include those encoding galactinol synthase and *PPT1*. The *PPT1* gene defines one CEN5 border (Fig. 5) and is surrounded by several retrotransposon sequences like some LINE or Athila elements. Interestingly, three of six *PPT* pseudogenes are located in centromeric or pericentromeric regions, too (Fig. 5). It has been shown that retrotransposons contribute to the evolution of genomes by a transfer of genomic sequences like exons or promoters to new positions (Moran et al., 1999). However, in most cases, the translocated genes lack introns (Jin and Bennetzen, 1994; Kumar and Bennetzen, 1999; Elrouby and Bureau, 2001; Witte et al., 2001). Thus, it is unlikely that the *PPT* and *GPT* pseudogenes arose by a retrotranscriptional mechanism.

The similarity of the *GPT* pseudogenes is about 65% to 80% to *GPT1* and *GPT2*, showing that the duplications occurred before the proliferation of the *PPT* pseudogenes. The chromosomal localization is remarkably different from the *PPT* pseudogenes because none of the *GPT* pseudogenes is localized within or near centromeric regions (Fig. 5), probably indicating a different duplication and transposition mechanism.

### PTh Sequences

Through database searches, we identified about 40 genes, mainly from Arabidopsis but also from other organisms, which encode proteins sharing significant similarities with the pPTs (PTh proteins). According to specific dipeptides at position 273/274, the PTh proteins can be divided into different groups, the KR, KT, KV/A/G, and the KD groups, comprising more than 20 proteins from Arabidopsis, animals, and fungi, all with unknown functions. The pPTs remarkably are most closely related to the KR group of uncharacterized proteins from fungi and animals lacking any presequences. Because the pPTs and the PTh proteins cluster together with the NST, Ward (2001) suggested that the pPTs and NSTs could be combined to the TPT-NST gene superfamily, which also includes functionally uncharacterized members

found in yeast, animals, and Arabidopsis that belong to the new groups described here. The TPT/NST superfamily can be found in all eukaryotic cells but not in prokaryotes (Jack et al., 2001). Because no proteins related to pPTs could be detected in cyanobacteria, the pPTs (and the families of PTh proteins) might trace back to the genome of the ancestral host cell. It is tempting to speculate that the pPTs evolved from proteins of the ER and Golgi membranes, probably members of the KR-group, by acquisition of a plastidic targeting signal sequence.

Besides sequence homology, further evidence corroborates the notion of a close relationship between pPTs and NSTs: (a) Both NSTs and pPTs function as homodimers (Wagner et al., 1989; Gao and Dean, 2000). In contrast, all other plastidic transporters characterized so far are monomers of 12 to 14 transmembrane helices (Weber and Flügge, 2002). (b) Both types of transporters function as antiporters. NSTs exchange nucleotide sugars with nucleoside monophosphates (Capasso and Hirschberg, 1984), whereas pPTs mediate the transport of phosphorylated C3 and C6 compounds in exchange with inorganic phosphate (Flügge, 1999). (c) Both the size of the NST and pPT proteins and the number of membrane-spanning regions are similar. The NSTs consist of 320 to 340 amino acids and have six to 10 transmembrane regions (Abeijon et al., 1997; Kawakita et al., 1998); the length of mature pPT proteins is about 330 amino acid residues. The proteins of the KV/A/G, KG, and KD groups have similar properties. Thus, it is reasonable to assume that these proteins are also homodimers that mediate an antiport transport of so far unknown substrates. (d) 4,4'-diisothiocyanostilbene-2,2'-disulfonic acid (DIDS), an inhibitor of the pPT activity (Flügge and Heldt, 1986), also leads to a reduction of the transport activity of NST proteins (Norambuena et al., 2002).

### Members of the TPT/NST Superfamily Might Share a Conserved Substrate-Binding Site

The conserved structure of the proteins of the TPT/NST superfamily suggests that all proteins might share a common substrate-binding site or at least conserved amino acid residues that are involved in substrate binding. In the case of the GDP-Man transporter from yeast and other fungi, a conserved motif has been identified that is required for binding of the nucleotide sugar (Gao et al., 2001). This GALNK consensus motif that resides in a cytosolic loop is also found in a GDP-Man transporter from Arabidopsis (Baldwin et al., 2001) but not in other NSTs. Sequence comparison of the GDP-Man transporters with other NSTs and the pPTs shows that this motif is located in a region that is highly conserved between the pPTs. The Lys residue at the end of this motif (K273) is remarkably conserved in all proteins of the TPT/NST superfamily including the KV/A/G, KD, and KT

families, albeit K273 is located one residue further downstream in case of the GDP-Man transporters. Another residue that is found in almost all TPT/NST sequences is a Thr/Ser residue at position 265, which is also part of the fourth similarity box of all pPTs and PThs. Thus, we propose the common motif T[X]<sub>7(8)</sub>K that is similar to motifs identified in GDP-Man and UDP-sugar transporters (Gao et al., 2001). In the case of the GDP-Fuc transporter from human, a mutation in that region leads to a defect in GDP-Fuc transport into the lumen of the Golgi (Lübke et al., 2001).

All substrates transported by the pPTs and NST are negatively charged molecules. Therefore, positively charged amino acids, e.g. Lys residues are expected to be involved in binding of these substrates. Reagents that react selectively with the  $\epsilon$ -amino group of Lys residues like pyridoxal-5'-phosphate (PLP), 2,4,6-trinitrobenzene sulfonate, or DIDS are strong inhibitors of phosphate translocator activities (Flügge and Heldt, 1986; Rumpho et al., 1988; Gross et al., 1990). DIDS also inhibits the activity of the UDP-Gal transporter from *Arabidopsis* (Norambuena et al., 2002). Both PLP and DIDS were shown to react with the same Lys group of the TPT and PPT (Gross et al., 1990). Substrates of the TPT, like phosphate or 3-PGA, have been shown to prevent binding of the inhibitors to the protein, suggesting that a Lys residue is located at the active site of the translocator (Flügge and Heldt, 1977). Exchange of K273 to Gln remarkably leads to a total loss of transport activity of the spinach TPT (B. Kammerer, K. Fischer, and U.I. Flügge, unpublished data).

It has been shown previously that phosphate translocator proteins are asymmetrically integrated into the inner envelope membrane (Flügge, 1992). The affinities toward the substrates phosphate and 3-PGA are different on either side of the membrane, the outward-facing (cytosolic) binding site showing 5-fold higher affinities than the inward-facing (stromal) site. In addition, PLP and DIDS bind only to the cytosolic site of the phosphate translocator protein (Flügge, 1992). These data suggest that two different substrate-binding sites exist facing to different sites of the membrane. One Lys residue that is involved in binding of substrates might be represented by K273 in the fourth homology box between the penultimate and the last membrane span. As shown above, a second Lys residue (K41), located in the first highly conserved region of the pPTs and between the first and the second membrane span, is conserved in all proteins of the pPT, NST, and PTh protein families. Both Lys residues are located in hydrophilic loops of the pPT proteins but face opposite sides of the membrane (Fig. 4). It is tempting to speculate that K41 and K273 are part of two binding sites that show different affinities toward the transported substrates.

## On the Function of the Unknown TPT/NST Proteins

We have identified several new proteins of unknown function in *Arabidopsis* that belong to the TPT/NST superfamily. Because of the homology to the pPTs and NSTs, we propose that these proteins also transport negatively charged substrates like phosphorylated compounds or nucleotide sugars.

What could be the physiological function of this high number of pPT/NST-related proteins in plants? Nucleotide sugars are synthesized in the cytosol. Thus, transport of nucleotide sugars by NSTs into the ER lumen is required to render the substrate available to sugar transferases that transfer the nucleotide sugars to endogenous acceptors, e.g. polysaccharides, glycoproteins, or glycolipids (Hirschberg et al., 1998). In plants, nucleotide sugars are mainly used in the synthesis of noncellulose polysaccharides and glycoproteins of the cell wall that takes place in the lumen of the Golgi cisternae (Dupree and Sherrier, 1998). Because the majority of NSTs are specific for only one nucleotide sugar, several transporters should be necessary for the transport of substrates for the syntheses of pectin and hemicellulose (Wulf et al., 2000).

Two of the proteins (at1g12500 and at3g10290, both members of the KV/A/G group) contain N-terminal plastidic targeting sequences. Nothing is known about the physiological function of these potential plastidic proteins. They are possibly involved in the biosynthesis of galactolipids (monogalactosyldiacylglycerol [MGDG] and digalactosyldiacylglycerol) in plastids, which are the major constituents of plastid membranes (Joyard et al., 1996). MGDG synthase transfers Gal from the donor (UDP-Gal) to a hydrophobic acceptor molecule, diacylglycerol, to synthesize MGDG and UDP, a reaction that occurs within the envelope membranes (Miège et al., 1999). Another lipid, sulfoquinovosyldiacylglycerol, is found in the photosynthetic membranes of plants and bacteria (Essigmann et al., 1998). The precursor of the synthesis of this sulfolipid is UDP-sulfoquinovose, which is synthesized from UDP-Glc and an intermediate of sulfate reduction in plastids (Tietje and Heinz, 1998). The origin of the nucleoside sugar is unknown, but it can be speculated that it is imported from the cytosol by a pPT/NST-related protein.

## MATERIALS AND METHODS

### Screening of Databases, Sequence Analysis, and Comparisons

Screening of databases was carried out with the BLAST algorithms (Altschul et al., 1990) at The *Arabidopsis* Information Resource (<http://www.Arabidopsis.org/blast/>) to detect PTh sequences. Both BLASTP and TBLASTN searches were done. Putative splicing sites and potential coding regions were predicted by the NETPLANTGENE software (Hebsgaard et al., 1996) and compared with previously sequenced cDNAs and expressed sequence tags. Multiple alignments were performed using ClustalX (Thompson et al., 1997). Unrooted trees were prepared by the neighbor-joining method using ClustalX v1.81 and TreeView v1.1.6, and 1,000 bootstrap replicates were performed.

## Plant Material

Plants of *Arabidopsis* ecotype Columbia were used in all experiments unless specified otherwise. Seedlings were grown at 20°C under 12-h:12-h light:dark regime with approximately 180  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ .

## Isolation of Genomic DNA and Gene Expression Studies

Genomic DNA was isolated from *Arabidopsis* (3 weeks old) leaves according to Ausubel et al. (1997). Total RNA was purified from whole *Arabidopsis* plants as previously described (Eggermont et al., 1996), with some small modifications. Reverse transcriptase (SuperscriptII, Invitrogen, Carlsbad, CA) was used to synthesize first strand cDNA from 2  $\mu\text{g}$  of total RNA (DNase treated) according to the manufacturer's instructions. Reverse transcriptase-PCR was done with 2  $\mu\text{L}$  of first-strand cDNA and gene-specific primers using 0.7 unit of *Taq* polymerase (Qiagen USA, Valencia, CA) in a total volume of 25  $\mu\text{L}$ . PCR conditions were 5 min at 95 followed by 40 cycles of 30 s of denaturation at 95°C, 30 s of annealing at 55°C, and extension for 1 min at 72°C.

Received October 22, 2002; returned for revision November 12, 2002; accepted November 12, 2002.

## LITERATURE CITED

- Abeijon C, Mandon EC, Hirschberg CB (1997) Transporters of nucleotide sugars, nucleotide sulfate and ATP in the Golgi apparatus. *Trends Biochem Sci* **22**: 203–207
- Altschul SF, Stephen F, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410
- Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–805
- Ausubel FM, Brent R, Kingston RE, Moore DD, Snuth FA, Seidman JG, Struhl K (1997) *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Intersciences, New York
- Bakarat A, Szick-Miranda K, Chang I-F, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J (2001) The organization of cytoplasmic ribosomal genes in the *Arabidopsis* genome. *Plant Physiol* **127**: 398–415
- Baldwin TC, Handford MG, Yuseff M-I, Orellana A, Dupree P (2001) Identification and characterization of GONST1, a Golgi-localized GDP-mannose transporter in *Arabidopsis*. *Plant Cell* **13**: 2283–2295
- Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionary conserved classes. *Nucleic Acids Res* **29**: 4319–4333
- Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**: 1093–1101
- Borchert S, Grosse H, Heldt HW (1989) Specific transport of inorganic phosphate, glucose 6-P, dihydroxyacetone phosphate and 3-phosphoglycerate into amyloplasts from pea roots. *FEBS Lett* **253**: 183–186
- Bowsher CG, Boulton EL, Rose J, Nayagam S, Emes MJ (1992) Reductant for glutamate synthase is generated by the oxidative pentose phosphate pathway in non-photosynthetic root plastids. *Plant J* **2**: 893–898
- Capasso JM, Hirschberg CB (1984) Mechanisms of glycosylation and sulfation in the Golgi apparatus: evidence for nucleotide sugar/nucleoside monophosphate and nucleotide sulfate/nucleoside monophosphate antiporters in the Golgi apparatus membrane. *Proc Natl Acad Sci USA* **81**: 7051–7055
- Cavalier-Smith T (2000) Membrane heredity and early chloroplasts evolution. *Trends Plant Sci* **5**: 174–182
- Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**: 2468–2474
- Daram P, Brunner S, Rausch C, Steiner C, Amrhein N, Bucher M (1999) Pht2;1 encodes a low-affinity phosphate transporter from *Arabidopsis*. *Plant Cell* **11**: 2153–2166
- Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001) Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. *Plant Physiol* **125**: 1342–1353
- Dupree P, Sherrier DJ (1998) The plant Golgi. *Biochim Biophys Acta* **1404**: 259–270
- Eicks M, Maurino V, Knappe S, Flügge UI, Fischer K (2002) The plastidic pentose phosphate translocator represents a link between the cytosolic and the plastidic pentose phosphate pathways in plants. *Plant Physiol* **128**: 512–522
- Eggermont K, Goderis IJ, Broekaert WF (1996) High-throughput RNA extraction from plant samples based on homogenisation by reciprocal shaking in the presence of a mixture of sand and glass beads. *Plant Mol Biol Rep* **14**: 273–279
- Elrouby N, Bureau TE (2001) A novel hybrid open reading frame formed by multiple cellular gene transduction by a plant long terminal repeat retroelement. *J Biol Chem* **276**: 41963–41968
- Essigmann B, Guler S, Narang RA, Linke D, Benning C (1998) Phosphate availability affects the thylakoid lipid composition and the expression of SQD1, a gene required for sulfolipid biosynthesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **95**: 1950–1955
- Fischer K, Arbinger B, Kammerer B, Busch C, Brink S, Wallmeier H, Sauer N, Eckerskorn C, Flügge UI (1994) Cloning and *in vivo* expression of functional triose phosphate/phosphate translocators from C<sub>3</sub>- and C<sub>4</sub>-plants: evidence for the putative participation of specific amino acid residues in the recognition of phosphoenolpyruvate. *Plant J* **5**: 215–226
- Fischer K, Kammerer B, Gutensohn M, Arbinger B, Weber A, Häusler R, Flügge UI (1997) A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. *Plant Cell* **9**: 453–462
- Fliege R, Flügge UI, Werdan K, Heldt HW (1978) Specific transport of inorganic phosphate, 3-phosphoglycerate and triosephosphates across the inner membrane of the envelope in spinach chloroplasts. *Biochim Biophys Acta* **502**: 232–247
- Flügge UI (1992) Reaction mechanism and asymmetric orientation of the reconstituted chloroplast phosphate translocator. *Biochim Biophys Acta* **1110**: 112–118
- Flügge UI (1999) Phosphate translocators in plastids. *Annu Rev Plant Physiol Plant Mol Biol* **50**: 27–45
- Flügge UI, Fischer K, Gross A, Sebald W, Lottspeich F, Eckerskorn C (1989) The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the *in vitro* synthesized precursor protein into chloroplasts. *EMBO J* **8**: 39–46
- Flügge UI, Heldt HW (1977) Specific labelling of a protein involved in phosphate transport of chloroplasts by pyridoxal-5'-phosphate. *FEBS Lett* **82**: 29–33
- Flügge UI, Heldt HW (1986) Chloroplast phosphate-triose phosphate-phosphoglycerate translocator: its identification, isolation and reconstitution. *Methods Enzymol* **125**: 716–730
- Frugoli JA, McPeck MA, Thomas TL, McClung CR (1998) Intron loss and gain during evolution of the catalase gene family in angiosperms. *Genetics* **149**: 355–365
- Gao XD, Dean N (2000) Distinct protein domains of the yeast Golgi GDP-mannose transporter mediate oligomer assembly and export from the endoplasmic reticulum. *J Biol Chem* **275**: 17718–17727
- Gao XD, Nishikawa A, Dean N (2001) Identification of a conserved motif in the yeast Golgi GDP-mannose transporter required for binding to nucleotide sugar. *J Biol Chem* **276**: 4424–4432
- Gross A, Brückner G, Heldt HW, Flügge UI (1990) Comparison of the kinetic properties, inhibition and labelling of the phosphate translocators from maize and spinach mesophyll chloroplasts. *Planta* **180**: 262–271
- Harrison CJ, Mould RM, Leech MJ, Johnson SA, Turner L, Schreck SL, Baird KM, Jack PL, Rawsthorne S, Hedley CL et al. (2000) The rug3 locus of pea encodes plastidial phosphoglucomutase. *Plant Physiol* **122**: 1187–1192
- Häusler RE, Schlieben NH, Schulz B, Flügge UI (1998) Compensation of decreased triose phosphate/phosphate translocator activity by accelerated starch turnover and glucose transport in transgenic tobacco. *Planta* **204**: 366–376
- Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996) Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* **24**: 3439–3452
- Herrmann KM, Weaver LM (1999) The shikimate pathway. *Annu Rev Plant Physiol Plant Mol Biol* **50**: 473–503

- Hirschberg CB, Robbins PW, Abejón C (1998) Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* **67**: 49–69
- Jack DL, Yang NM, Saier MH (2001) The drug/metabolite transporter superfamily. *Eur J Biochem* **268**: 3620–3639
- Jin YK, Bennetzen JL (1994) Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bsl1 retroelement of maize. *Plant Cell* **8**: 1177–1186
- Johanson U, Karlsson M, Johanson I, Gustavsson S, Sjövall S, Frayse L, Weig AR, Kjellbom P (2001) The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol* **126**: 1358–1369
- Joyard J, Block MA, Douce R (1996) Molecular aspects of plastid envelope biochemistry. *Eur J Biochem* **199**: 489–509
- Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A, Flügge UI (1998) Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose-6-phosphate phosphate antiporter. *Plant Cell* **10**: 105–117
- Kawakita M, Ishida N, Miura N, Sun-Wada GH, Yoshioka S (1998) Nucleotide sugar transporters: elucidation of their molecular identity and its implication for future studies. *J Biochem* **123**: 777–785
- Kofler H, Häusler RE, Schulz B, Gröner F, Flügge UI, Weber A (2000) Molecular characterisation of a new mutant allele of the plastidic phosphoglucomutase and complementation of the mutant with the wild-type cDNA. *Mol Gen Genet* **263**: 978–986
- Ku HM, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc Natl Acad Sci USA* **97**: 9121–9126
- Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* **33**: 479–532
- Li Y, Darley CP, Ongaro V, Fleming A, Schipper O, Baldauf SL, McQueen-Mason SJ (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin. *Plant Physiol* **128**: 854–864
- Lübke T, Marquardt T, Etzioni A, Hartmann E, von Figura K, Körner C (2001) Complementation cloning identifies CDG-IIc, a new type of congenital disorders of glycosylation, as a GDP-fucose transporter deficiency. *Nat Genet* **28**: 73–76
- Lühn K, Wild MK, Eckhard M, Gerardy-Schahn R, Vestweber D (2001) The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. *Nat Genet* **28**: 69–72
- Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: How much, what happens and why? *Plant Physiol* **118**: 9–17
- Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik K (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**: 162–165
- McFadden GI (1999) Endosymbiosis and evolution of the plant cell. *Curr Opin Plant Biol* **2**: 513–519
- Miège C, Maréchal E, Shimojima M, Awai K, Block MA, Ohta H, Takamiya E, Douce R, Joyard J (1999) Biochemical and topological properties of type A MGDG synthase, a spinach chloroplast envelope enzyme catalyzing the synthesis of both prokaryotic and eukaryotic MGDG. *Eur J Biochem* **265**: 990–1001
- Moran JV, DeBerardinis RJ, Dazarian HH (1999) Exon shuffling by L1 retrotransposition. *Science* **283**: 1530–1534
- Neuhaus HE, Thom E, Möhlmann T, Steup M, Kampfenkel K (1997) Characterization of a novel ATP/ADP translocator located in the plastid envelope of *Arabidopsis thaliana*. *Plant J* **11**: 73–82
- Norambuena L, Marchant L, Berninsone P, Hirschberg CB, Silva H, Orellana A (2002) Transport of UDP-galactose in plants: identification and functional characterization of AtUTr1, an *Arabidopsis thaliana* UDP-galactose/UDP-glucose transporter. *J Biol Chem* **277**: 32923–32929
- Rumpho ME, Edwards GE, Yousif AE, Keegstra K (1988) Specific labelling of the phosphate translocator in C<sub>3</sub> and C<sub>4</sub> mesophyll chloroplasts by tritiated dihydro-DIDS (1,2-ditritio-1,2-[2,2'-disulfo-4,4'-diisothiocyano]-diphenylethane). *Plant Physiol* **86**: 1193–1198
- Schimper AFW (1883) Über die Entwicklung der Chlorophyllkörner und Farbkörper. *Bot Zeitung* **41**: 104–114
- Schmidt R (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. *Plant Mol Biol* **48**: 21–37
- Schneider A, Häusler RE, Kolkusaoglu Ü, Kunze R, van der Graaf E, Schwacke R, Catoni E, Desimone M, Flügge UI (2002) An *Arabidopsis thaliana* knock-out mutant of the chloroplast triose phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. *Plant J* **32**: 685–699
- Schwacke R, Schneider A, Van der Graaf E, Fischer K, Catoni E, Desimone M, Frommer WB, Flügge UI, Kunze R (2003) ARAMEMNON: a novel database for *Arabidopsis* integral membrane proteins. *Plant Physiol* **131**: 16–26
- Streatfield SJ, Weber A, Kinsman EA, Häusler RE, Li J, Post-Beittenmiller D, Kaiser WM, Pyke KA, Flügge UI, Chory J (1999) The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development and plastid-dependent nuclear gene expression. *Plant Cell* **11**: 1609–1621
- Tavares R, Aubourg S, Lecharny A, Kreis M (2000) Organization and structural evolution of four multigene families in *Arabidopsis thaliana*: AtLCAD, AtLGT, AtMYST and AtHD-GL2. *Plant Mol Biol* **42**: 703–717
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882
- Tietje C, Heinz E (1998) Uridine-diphospho-sulfoquinovose:diacylglycerol sulfoquinovosyltransferase activity is concentrated in the inner membrane of chloroplast envelopes. *Planta* **206**: 72–78
- Versaw WK, Harrison MK (2002) A chloroplast phosphate transporter, PHT2;1, influences allocation of phosphate within the plant and phosphate-starvation responses. *Plant Cell* **14**: 1751–1766
- Wagner R, Apley EC, Gross A, Flügge UI (1989) The rotational diffusion of chloroplast phosphate translocator and of lipid molecules in bilayer membranes. *Eur J Biochem* **182**: 165–173
- Walker JE, Runswick MJ (1993) The mitochondrial transport protein superfamily. *J Bioenerg Biomembr* **25**: 435–446
- Ward JM (2001) Identification of novel families of membrane proteins from the model plant *Arabidopsis thaliana*. *Bioinformatics* **17**: 560–563
- Weber A, Flügge UI (2002) Interaction of cytosolic and plastidic nitrogen metabolism in plants. *J Exp Bot* **53**: 865–874
- Weber A, Servaites JC, Geiger DR, Kofler H, Hille D, Gröner F, Hebbeker U, Flügge UI (2000) Identification, purification and molecular cloning of a plastidic glucose translocator. *Plant Cell* **12**: 787–801
- Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITES: important players in the evolution of plant genomes. *Curr Opin Genet Dev* **5**: 814–821
- Witte C-P, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc Natl Acad Sci USA* **98**: 13778–13783
- Wulf C, Norambuena L, Orellana A (2000) GDP-fucose uptake into the Golgi apparatus during xyloglycan biosynthesis requires the activity of a transporter-like other than the UDP-glucose transporter. *Plant Physiol* **122**: 1–11