

Dual specificities of the glyoxysomal/peroxisomal processing protease Deg15 in higher plants

Michael Helm[†], Carsten Lück[‡], Jakob Prestele[†], Georg Hierl[†], Pitter F. Huesgen[§], Thomas Fröhlich[¶], Georg J. Arnold[¶], Iwona Adamska[§], Angelika Görg[‡], Friedrich Lottspeich^{||}, and Christine Gietl^{†**}

[†]Lehrstuhl für Botanik, [‡]Fachgebiet Proteomik, Technische Universität München, Wissenschaftszentrum Weihenstephan, D-85350 Freising, Germany; [§]Lehrstuhl für Physiologie und Biochemie der Pflanzen, Fachbereich Biologie, Universität Konstanz, D-78457 Konstanz, Germany; [¶]Laboratorium für funktionale Genomanalyse, Ludwig-Maximilian-Universität München, D-81377 Munich, Germany; and ^{||}Max-Planck-Institut für Biochemie, Proteinanalytik, D-82152 Martinsried, Germany

Communicated by Diter von Wettstein, Washington State University, Pullman, WA, May 21, 2007 (received for review April 11, 2007)

Glyoxysomes are a subclass of peroxisomes involved in lipid mobilization. Two distinct peroxisomal targeting signals (PTSs), the C-terminal PTS1 and the N-terminal PTS2, are defined. Processing of the PTS2 on protein import is conserved in higher eukaryotes. The cleavage site typically contains a Cys at P1 or P2. We purified the glyoxysomal processing protease (GPP) from the fat-storing cotyledons of watermelon (*Citrullus vulgaris*) by column chromatography, preparative native isoelectric focusing, and 2D PAGE. The GPP appears in two forms, a 72-kDa monomer and a 144-kDa dimer, which are in equilibrium with one another. The equilibrium is shifted on Ca²⁺ removal toward the monomer and on Ca²⁺ addition toward the dimer. The monomer is a general degrading protease and is activated by denatured proteins. The dimer constitutes the processing protease because the substrate specificity proven for the monomer (Φ -Arg/Lys \downarrow) is different from the processing substrate specificity (Cys-Xxx \downarrow /Xxx-Cys \downarrow) found with the mixture of monomer and dimer. The *Arabidopsis* genome analysis disclosed three proteases predicted to be in peroxisomes, a Deg-protease, a pitrilysin-like metallopeptidase, and a Lon-protease. Specific antibodies against the peroxisomal Deg-protease from *Arabidopsis* (Deg15) identify the watermelon GPP as a Deg15. A knockout mutation in the *DEG15* gene of *Arabidopsis* (At1g28320) prevents processing of the glyoxysomal malate dehydrogenase precursor to the mature form. Thus, the GPP/Deg15 belongs to a group of trypsin-like serine proteases with *Escherichia coli* DegP as a prototype. Nevertheless, the GPP/Deg15 possesses specific characteristics and is therefore a new subgroup within the Deg proteases.

Arabidopsis thaliana | Ca²⁺ signal | *Citrullus vulgaris* | monomer/dimer equilibrium

Numerous matrix enzymes have to be imported from the cytosol into peroxisomes, in plants especially into glyoxysomes for seed storage oil mobilization or into leaf peroxisomes for photorespiration. The majority of these enzymes are imported in their mature form and targeted by a C-terminal SKL designated peroxisomal targeting signal 1 (PTS1) (1). In a few matrix enzymes, the peroxisomal targeting signal 2 (PTS2) with the consensus RL-X5-HL is located in the N-terminal 30 to 50 amino acids of the protein (2). In plants, these are four enzymes of the glyoxylate cycle and β -oxidation of fatty acids: glyoxysomal malate dehydrogenase (gMDH), glyoxysomal citrate synthase (gCS), acyl-CoA oxidase, and thiolase. In mammals, three enzymes with a PTS2 have been identified: thiolase, alkyl-DHAP synthase, and phytanoyl-CoA hydroxylase. In higher eukaryotes, such as plants and mammals, the PTS2 is removed on import; a Cys is consistently found near the cleavage site [supporting information (SI) Table 2]. In lower eukaryotes, such as yeasts, a PTS2 is present in the N terminus of the mature subunit of thiolase and amine oxidase (2) (SI Table 2). The Cys in position P2 is required for processing the presequences of gCS and gMDH in pumpkin; deletion of the Cys prevents cleavage,

and its exchange to Gly or Phe diminishes the processing efficiency to 20–30% of wild type (3, 4). Import or enzyme activity is not impeded by the absence of presequence removal.

In this paper, we present a procedure for purification and characterization of the glyoxysomal processing protease (GPP) from cotyledons of 3-day-old dark-grown watermelon seedlings (*Citrullus vulgaris*). The protein was identified with a fluorogenic substrate covering the cleavage sites of gMDH, gCS, and thiolase, and with an antibody raised against the *Arabidopsis* peroxisomal AtDeg15 protease. A T-DNA insertion mutant in the *AtDEG15* gene of *Arabidopsis* (At1g28320) prevents processing of the higher molecular mass precursor of gMDH. The GPP from *C. vulgaris* appears in two forms, as a 72-kDa monomer and a 144-kDa dimer under reducing and denaturing conditions. They migrate with 55 kDa and 110 kDa under conditions preserving a native and folded protein conformation. They are in equilibrium with one another and could be detected with gelatine-activity gels. The equilibrium is shifted on Ca²⁺ removal toward the monomer and on Ca²⁺ addition toward the dimer. The dimeric GPP/Deg15 protease cleaves specifically substrates with Cys in the P1 or P2 position and constitutes the peroxisomal processing peptidase. The monomer cleaves denatured peroxisomal matrix proteins into small peptides; it also exhibits an unusual temperature optimum for eukaryotic proteases at 45°C and a pH optimum in the basic range.

Results

Partial Purification of the GPP. Purified glyoxysomes with an isocitrate lyase activity of 0.14 units/ml were sonicated and filter-sterilized, and the membranes were removed by centrifugation. The resulting glyoxysomal extract was loaded onto a DEAE-anion exchange column. Protease activity of the starting material and eluate were analyzed with the fluorogenic peptide *N*-benzoyl-carbonyl-Phe-Arg-7-amino-4-methylcoumarin (Cbz-FR-AMC), yielding an eluate fraction containing 15 mg of protein, a recovery of 97% of the activity applied, and a purification factor of 12.3. A further 301-fold purification was achieved by Superdex75 gel filtration, hydroxyapatite chromatography, and MonoQ-anion exchange chromatography, but the losses resulted in a yield of no more than 5% (SI Table 3).

Author contributions: M.H., C.L., J.P., G.H., G.J.A., I.A., A.G., F.L., and C.G. designed research; M.H., C.L., J.P., G.H., P.F.H., T.F., F.L., and C.G. performed research; M.H., C.L., J.P., G.H., P.F.H., T.F., G.J.A., I.A., A.G., F.L., and C.G. analyzed data; and M.H. and C.G. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: Cbz-FR-AMC, *N*-benzoyl-carbonyl-Phe-Arg-7-amino-4-methylcoumarin; gCS, glyoxysomal citrate synthase; gMDH, glyoxysomal malate dehydrogenase; GPP, glyoxysomal processing protease; IEF, isoelectric focusing; PTS1, peroxisomal targeting signal 1; PTS2, peroxisomal targeting signal 2; RFU, relative fluorescence unit.

**To whom correspondence should be addressed. E-mail: christine.gietl@wzw.tum.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0704733104/DC1.

© 2007 by The National Academy of Sciences of the USA

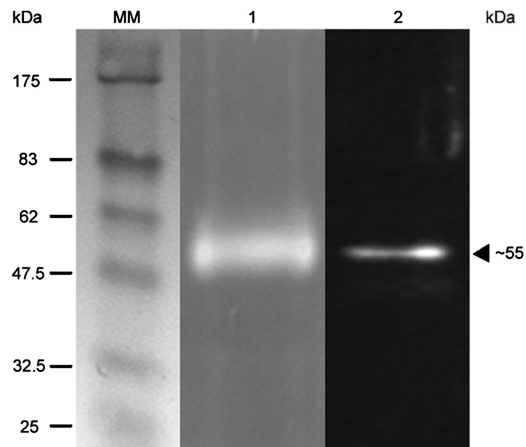


Fig. 1. The watermelon GPP is immunologically identical to the AtDeg15 protease. The glyoxysome extract was purified by DEAE-anion exchange and *p*-aminobenzamidine affinity chromatography. The partially purified GPP was analyzed by native PAGE and incubation with the fluorogenic peptide Cbz-FR-AMC (lane 1), followed by Western blot analysis of the same gel and decoration with the α -Deg15 peptide antibody (lane 2).

A synthetic peptide of 15 amino acids (ETTAAVHPGGSG-GAV) close to the catalytic Ser (SI Fig. 7) of the AtDeg15 protease was used to prepare polyclonal antibodies. These antibodies labeled a glyoxysomal protease band of 55 kDa visualized through activity gel assay with Cbz-FR-AMC in a subsequent Western blot (Fig. 1, lanes 1 and 2). The 55-kDa band was visible in the protease fraction purified by DEAE anion exchange and benzamidine affinity chromatography and separated on a native 8% polyacrylamide gel [360,000 relative fluorescence units (RFUs)]. The same 55-kDa activity band was obtained by assaying the hydroxyapatite pool (data not shown). A Western blot of the same gel revealed the GPP to be immunologically identical to the AtDeg15 protease.

Purification of GPP by DEAE-Anion Exchange Chromatography, Preparative Native Sephadex-Isoelectric Focusing (IEF), and Narrow pH Range 2D PAGE. The partially purified GPP from the DEAE column was mixed with the Sephadex matrix, the immobilines pH range 3–10, and the colored pI markers and pipetted into the flat-bed. Separation in the electric field yielded a single-activity peak with Cbz-FR-AMC as a substrate with a pI of 4.7–5.2 (Fig. 2). For additional proof of identity of GPP with AtDeg15 and more precise determination of the molecular mass of the protein under reducing and denaturing conditions, 2D PAGE was performed (Fig. 3). The protease-containing fraction from the native Sephadex-IEF was resolved by IEF in an immobilized pH gradient (IPG) of 4–7 (IPG-IEF). For separation of the proteins according to molecular mass, duplicate IPG strips were mounted onto 13% SDS polyacrylamide gels and subjected to electrophoresis in the second dimension. One gel was stained with silver nitrate (Fig. 3A), and the other was decorated with α -AtDeg15 antibodies (Fig. 3B). The single spot identified by the antibody corresponded to a single spot in the 2D gel as shown by superimposition (Fig. 3C). The 2D PAGE revealed a pI of 5.25 for the GPP/AtDeg15; this pI is in accordance with the pI of 4.7–5.2 as estimated by the preparative native IEF, assuming that complete unfolding of the protease in the 2D PAGE might lead to a slight pI shift. Watermelon GPP had an apparent molecular mass of 72 kDa under reducing and denaturing conditions, whereas the GPP migrated at 55 kDa under conditions preserving a native and folded protein conformation (Fig. 1). This molecular mass of the watermelon GPP is in agreement with the

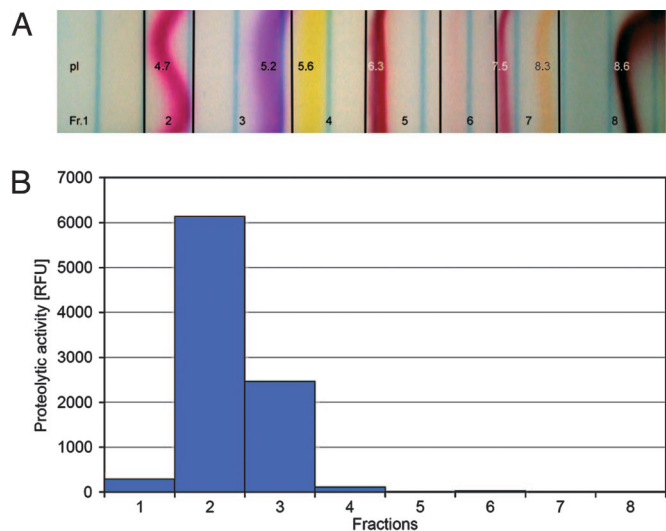


Fig. 2. Localization of the watermelon GPP activity in the preparative native Sephadex-IEF. (A) Separation of partially purified GPP in the electric field with colored pI markers and immobilines from pH 3–10. (B) A single peak for protease activity with the fluorogenic substrate Cbz-FR-AMC was obtained at pI 4.7–5.2.

glyoxysomal Deg15 protease in *Arabidopsis* (At1g28320), which has a calculated molecular mass of 76 kDa according to its amino acid sequence (5).

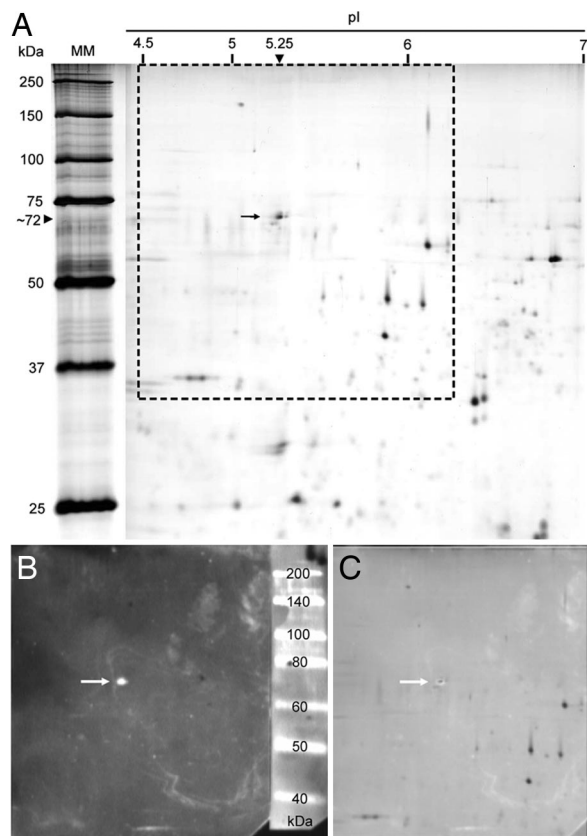


Fig. 3. Purification and identification of the watermelon GPP. (A) The glyoxysome extract purified by DEAE chromatography and native Sephadex-IEF was resolved by 2D PAGE with an immobilized pH gradient and SDS/PAGE and stained with silver nitrate. (B) A parallel gel was blotted (framed area in A) and decorated with α -Deg15 antibodies. (C) Superimposition of A (framed area) and B identify a single protein spot with a pI of 5.25 and an M_r of 72 kDa.

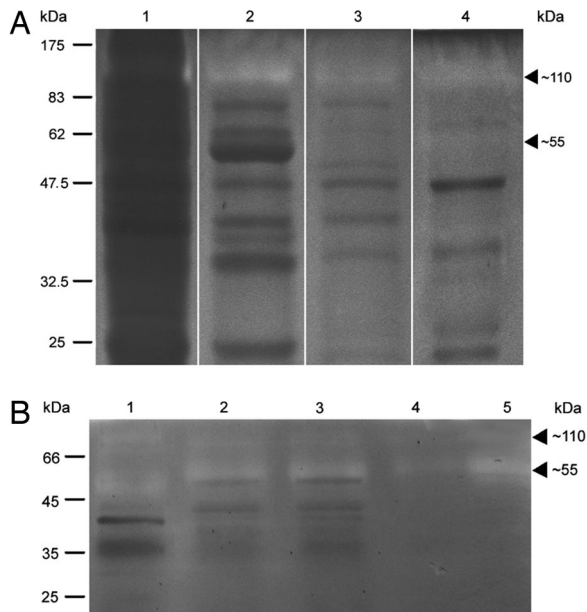


Fig. 4. The GPP/Deg15 exists as an equilibrium of monomers and dimers. The GPP was partially purified and proteolytic activity was localized by native PAGE in a gel with 0.05% gelatine after incubation at 30°C or 45°C. (A) At 30°C, analysis of the activity pools from DEAE (lane 1), Sephadex75 (lane 2), hydroxyapatite (lane 3), and MonoQ chromatography (lane 4) revealed a strong proteolytic activity at 110 kDa (lanes 1–4) and a barely visible activity at 55 kDa (lane 4). (B) At 45°C, analysis of the glyoxysome extract (lane 1), DEAE pool (lane 2), and Sephadex-IEF eluate (lane 5) revealed a strong activity at 55 kDa and a weaker activity at 110 kDa. Fractionation of the DEAE pool by ultrafiltration into proteins <100 kDa (lane 3) and >100 kDa (lane 4) likewise indicated activity at both positions. This result indicates interconversion among monomers and dimers.

Characteristics of the GPP/Deg15 Enzyme. A test of the proteolytic activity in the pH range from 4.5 to 9.0 revealed a linear increase with low activity at pH 4.5 to high activity at pH 9.0; in phosphate buffer, the enzyme activity increased sigmoidally from low to high pH, reaching its maximum at pH 8–9 (SI Fig. 8A and B). The temperature optimum of the enzyme for Cbz-FR-AMC was 45°C. Temperature dependence increased from 2,500 RFUs at 25°C to 6,500 RFUs at 45°C (SI Fig. 9). Several protease inhibitors were tested on the DEAE and Sephadex-IEF eluates (SI Fig. 10). GPP was inhibited by antipain, leupeptin, and *N*-*p*-tosyl-L-Lys chloromethylketone, which inhibit Cys proteases such as papain and trypsin-like Ser proteases. The lack of inhibition by the chymotrypsin inhibitors chymostatin and pepstatin A, as well as the serine proteases inhibitor PMSF, was noteworthy. An important characteristic was an increased proteolytic activity on incubation with EDTA and cystatin.

GPP Exists as a Monomer and a Dimer with Different Substrate Specificities. Separation of activity from the DEAE column, Sephadex75 gel filtration, hydroxyapatite, and MonoQ column by 8% native PAGE with 0.05% gelatine and incubation of the gel at 30°C identified protease activity at ≈110 kDa, with little or no activity at 55 kDa (Fig. 4A, lanes 1–4). Thus, a 110-kDa protease activity coeluted over four different chromatography columns with the 55-kDa protease that cleaves fluorogenic Cbz-FR-AMC, which was used for monitoring GPP activity during purification. This outcome raised the possibility that the 55-kDa and 110-kDa protease activity could result from a monomer and a dimer of the same protein. High concentrations of GPP activity separated at room temperature on a native polyacrylamide gel and elicited an extremely fluorescent signal at the 55-kDa position when incubated with Cbz-FR-AMC, whereas no cleavage of the fluorogenic peptide occurred by the 110-kDa form (data not shown). On incubation of the gelatine activity gel at 45°C, the glyoxysome extract, the DEAE eluate pool, and the native Sephadex-IEF eluate all exhibited protease activity at the 55-kDa level and only a weak activity at 110 kDa (Fig. 4B, lanes 1, 2, and 5). With Cbz-FR-AMC as a substrate, the optimal temperature for GPP *in vitro* was 45°C (SI Fig. 9).

The GPP dimer and monomer seem to be in equilibrium because the 110-kDa form can be generated from the 55-kDa form. During Superdex75 gel filtration, the dimeric 110-kDa form was in the void volume and did not react with Cbz-FR-AMC. This result can explain the loss of 40% of the loaded activity during Superdex75 gel filtration (SI Table 3). However, the proteolytically active fraction eluting at 55 kDa from the Superdex75 column, detectable with Cbz-FR-AMC, contained the gelatin-hydrolyzing 110-kDa form (Fig. 4A, lane 2). Furthermore, when the DEAE eluate was fractionated by ultrafiltration into <100 kDa and >100 kDa, both fractions contained monomeric as well as dimeric proteolytic activity, although the 55-kDa monomer should be absent from the >100-kDa fraction (Fig. 4B, lanes 3 and 4).

Specificity of the Sites Cleaved by Monomeric and Dimeric GPP. Four fluorogenic peptides covering the cleavage sites of gCS, gMDH, thiolase, and a negative control peptide, in which the Cys in P1 position of thiolase was substituted by Gly, were synthesized (SI Table 4). In all, 10 nmol of peptide was incubated for 3 h at 24°C with GPP (67,000 RFUs) that had been purified by DEAE chromatography, native Sephadex-IEF, and removal of peptides by 5-kDa ultrafiltration. The cleavage products were sequenced by N-terminal sequence analysis and mass spectrometry (Table 1). The gCS peptide was cleaved in the vicinity of Cys, but not at the authentic site with Cys at P2. The gMDH peptide was cleaved at four positions, including next to the predicted Arg at P1 with Cys at P2. The thiolase peptide was cleaved at two positions, one being the predicted one adjacent to Cys. The negative control peptide with the substituted Cys was resistant to cleavage. The furin substrate was cleaved after Arg (P1) and Lys

Table 1. Cleavage by GPP of fluorogenic peptides containing the cleavage sites of glyoxysomal higher molecular weight precursor proteins

Fluorogenic peptide	Sequence	Cleavage products
gCS	Abz-EAHCV [†] SAQ [†] TMY* [†] D	Abz-EAH . . . + TMY* [†] D
gMDH	Abz-R [†] ANCR [†] AK [†] G [†] GAY* [†] D	Abz-RANCR + ANCR + AKG + GGAY* [†] D + GAY* [†] D
Thiolase	Abz-SASVC [†] AA [†] GDSY* [†] D	Abz-SASV . . . + AAGDS . . . + GDSY* [†] D
Control	Abz-SASVG [†] AAGDSY* [†] D	None
Furin substrate	Abz-RVVKR [†] GLAY* [†] D	Abz-RVVKR + GLAY* [†] D

Y*, (NO₂)Y, m-nitro-tyrosine as quencher (acceptor); Abz, anthralinic acid as fluorogene (donor; excitation 320, emission 420 nm).

[†]Predicted cleavage site.

[†]Cleavage sites as identified by N-terminal sequence analysis and mass spectroscopy of the cleavage products.

(P2). The GPP protein was present as a mixture of a monomer and a dimer. The monomer, but not the dimer, cleaves the fluorogenic Cbz-FR-AMC peptide. We interpret the nonauthentic cleavages to be due to the activity of monomers and the authentic cleavages to be performed by dimers.

GPP Degrades Glyoxysomal Matrix Proteins at 45°C. The partially purified protein pool eluted from the DEAE column contains many glyoxysomal matrix proteins in addition to GPP. A 24-h digestion of the pool at 30°C displays many protein bands on SDS/PAGE. However, an incubation for 24 h at 45°C reduced the number of protein bands dramatically (SI Fig. 11 A and B). We interpret this result to show that the monomer degrades proteins by cleaving Φ -R/K↓ with Φ identifying hydrophobic and apolar residues. The limited endoproteolysis by GPP of higher molecular weight precursor proteins is carried out by the dimer at lower temperatures, implying two different proteolytic specificities for the two forms of the enzyme.

Ca²⁺ Ions Promote Dimerization of GPP. The DEAE eluate was concentrated by ultrafiltration. After incubating with 6.25 mM EDTA for 10 min, the proteolytic activity was increased by 35%. A similar result was obtained in the course of inhibitor studies with 10 mM EDTA final concentration, which increased the proteolytic activity 2.5-fold (SI Fig. 10). No significant change was observed by adding 25 mM Ca²⁺ ions to the DEAE eluate. A 10-min incubation with EDTA, followed by the addition of 25 mM Ca²⁺ for 10 min, diminished the proteolytic activity to 20%, as measured with the substrate for the monomer Cbz-FR-AMC. Apparently dimer formation requires Ca²⁺ ions. A pretreatment with EDTA might remove other divalent cations from the binding site.

A direct comparison of proteolytic activity of the DEAE eluate toward the Cbz-FR-AMC substrate in 20 mM Tris-HCl buffer (pH 8.5) with that in 50 mM Na-phosphate buffer (pH 8.5) revealed a 2-fold higher activity in phosphate than in Tris buffer (SI Fig. 12). The effect of phosphate buffer is probably because of the precipitation of Ca₃(PO₄)₂, thus pushing the equilibrium toward the monomer. Formation of the dimer on addition of Ca²⁺ ions might explain the 57% loss of activity against the Cbz-FR-AMC monomer substrate during hydroxyapatite chromatography (SI Table 3) because this ceramic material contains Ca²⁺, phosphate, and hydroxide ions.

A Knockout Mutation in the *DEG15* Gene of *Arabidopsis* Prevents Processing of pre-gMDH. SALK-line 007184 contains a T-DNA insertion in intron 4 of the *DEG15* gene (At1g28320; 13 exons). Genomic DNA from 10-day-old segregating plants was isolated. Two primers specific for the wild-type gene amplified a 1,000-bp product, whereas the gene-specific primer with the T-DNA primer gave a product of 600 bp. Genomic DNA of the heterozygous plants gave both amplification products, whereas genomic DNA of the homozygous plant yielded only the 600-bp fragment (Fig. 5A). These mutants had no obvious altered plant phenotype. A protein extract was isolated from homozygous mutant plants and separated by SDS/12.5% PAGE. Immunoblot with α -gMDH antibodies showed that processing of pre-gMDH to gMDH no longer occurred in the homozygous mutant plant, whereas the heterozygous and wild-type plants processed the pre-gMDH to the mature enzyme (Fig. 5B).

In addition to Deg15, analysis of the *Arabidopsis* genome disclosed two peroxisomal proteases with a PTS1 that may be possible candidates for a peroxisomal processing protease (5): a Zn-metalloprotease with the ptrylisin family typical inverted HXXEH-Zn binding motif, also called insulinase (At2g4170); and a Lon-protease, a serine protease with the catalytic diad Ser-Lys (At5g47040). Knockout mutants with T-DNA insertions have been identified for both candidates: for At2g4170 (26 exons), the SALK-

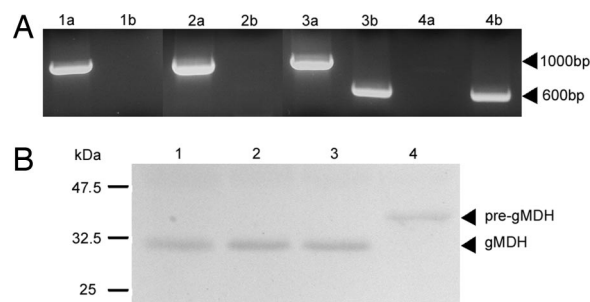


Fig. 5. Analysis of knockout T-DNA insertion mutation in the *DEG15* gene (At1g28320) of *Arabidopsis*. (A) PCR amplificates from wild-type and mutant plant DNA. Primers specific for the wild-type gene amplify a 1,000-bp fragment, whereas the gene-specific primer with the T-DNA gives a 600-bp product. PCR products are as follows: lanes 1 a and b, wild-type Columbia; lanes 2 a and b, wild type from segregating population; lanes 3 a and b, heterozygous mutant plant; lanes 4 a and b, homozygous mutant plant. (B) Western blots of protein extracts from wild-type and mutant plants decorated with antibodies specific for glyoxysomal malate dehydrogenase. Lane 1, wild-type control; lane 2, wild type; lane 3, heterozygous plant of segregating population; lane 4, homozygous mutant fails to process pre-gMDH to its mature form.

line 023917 with the T-DNA insertion in the 21 exon; and for At5g47040 (17 exons), the SALK-line 043857 with the T-DNA insertion in exon 17. Homozygous mutants were characterized in a segregating population. Protein extracts from both homozygous mutants processed the pre-gMDH to its mature form. The proteins were separated by SDS/12.5% PAGE, and decoration with α -gMDH antibodies consistently yielded the processed mature gMDH (data not shown). Thus, both the Zn-metalloprotease and the Lon-protease are not involved in the processing of glyoxysomal higher molecular weight precursor.

Discussion

In contrast to the majority of peroxisomal matrix enzymes targeted with the PTS1, a smaller number employ the PTS2 (2). Removal of the PTS2-containing presequence in higher eukaryotes by the peroxisomal processing protease takes place at a Cys at P1 or P2. In the latter case, P1 is occupied by Arg, Val, or Thr (SI Table 2). Processing of the PTS2 is not necessary for enzyme activity in higher eukaryotes (6, 7); it does not take place in lower eukaryotes such as yeasts (8, 9), and pre-gMDH from higher plants is enzymatically active in the heterologous host *H. polymorpha* (10). An *Arabidopsis* peroxisomal protease (GPP) of the Deg/HtrA family of ATP-independent serine endopeptidases was purified from watermelon. GPP occurs in monomeric and dimeric forms; dimerization requires Ca²⁺ ions. Evidence is presented that the dimer carries out the specific cleavages needed to remove the presequence, whereas the monomer attacks sites that can degrade the removed presequences and matrix enzymes that are misfolded or no longer needed during development. The protein extracted from a T-DNA insertion mutant in *Arabidopsis deg15* is not able to process the presequence. Of the three known peroxisomal *Arabidopsis* proteases (5), only the AtDeg15 protease can process presequences; the Zn-metalloprotease (11) and the Lon-protease^{††} are unable to remove the presequence.

Deg15 belongs to the Deg/HtrA proteases, which form trimeric and hexameric complexes (12). In addition to the trypsin-like domain, most of them contain one to three PDZ domains located toward the C terminus of the polypeptide chain. Crystallographic analyses of the trimers and hexamers suggest that

^{††}De Walque, S., Fransen, M., Mannaerts, G. P., Van Veldhoven, P. P., Poster presentation, International Symposium "Peroxisomal Disorders and Regulation of Genes," Sept. 23–25, 2002, Ghent, Belgium.

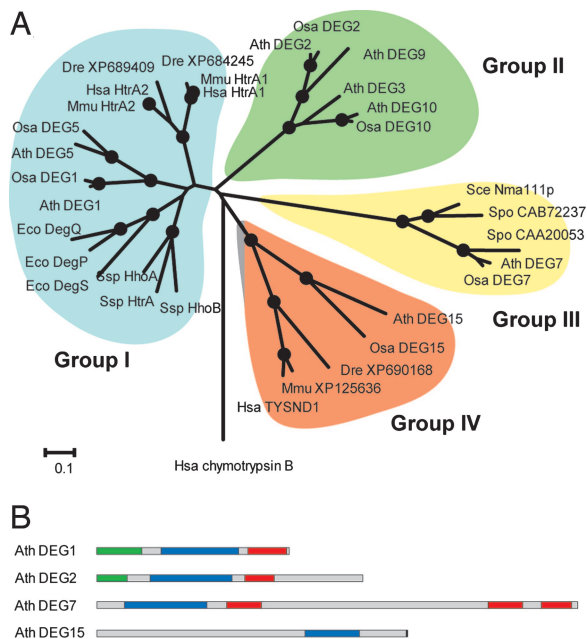


Fig. 6. Phylogenetic analysis and domain structure of Deg proteases. (A) The neighbor-joining phylogenetic tree of the Deg protease domains. Black dots indicate a support by 80% or more of 10,000 bootstrap replicates. National Center for Biotechnology Information protein accession numbers are given for unnamed proteins. (B) Domain structures of representative Deg/HtrA proteases from the four distinct groups. Blue, trypsin-like protease domains; red, PDZ domains; green, chloroplast transit peptides; black, SKL signal. The following abbreviations are used: Ath, *Arabidopsis thaliana*; Bsu, *Bacillus subtilis*; Dme, *Drosophila melanogaster*; Eco, *Escherichia coli*; Hsa, *Homo sapiens*; Mmu, *Mus musculus*; Osa, *Oryza sativa*; Sce, *Saccharomyces cerevisiae*; Spo, *Schizosaccharomyces pombe*; Ssp, *Synechocystis* sp. PCC 6803.

the PDZ domains play a role in substrate recognition and activation of the protease domain (13). Sixteen genes coding for putative Deg/HtrA proteases have been identified in the genome of *Arabidopsis*, of which six are predicted to target to mitochondria and seven to chloroplasts (14–16). *DEG15* (At1g28320) is the only gene coding for a peroxisomal protein. Paralogues of *DEG15* exist in various eukaryotic organisms (SI Table 5). These proteases contain the C-terminal PTS1 (17) and lack a PDZ domain (Fig. 6). The human and mouse homologues are annotated as trypsin domain-containing domain 1 (*TYSND1*) and are classified as members of the S1A subfamily of trypsin-like serine proteases (18). Similarity analysis of the protease domains of known and putative Deg/HtrA proteases shows that this family is divided into four distinct groups (Fig. 6A). This distribution of the Deg/HtrA proteases into four groups is reflected by their respective protein domain structures, rather than their distribution across different species (Fig. 6B). Group I includes the canonical Deg/HtrA proteases from bacteria and eukaryotes, which carry one (e.g., HsaHtrA2) or two (e.g., EcoDegP/HtrA) PDZ domains at the C terminus of the protease domain (12, 19). The only exception is the Deg5 protease from plants, which does not contain a PDZ domain. Analysis of the thylakoid lumen proteome of *Arabidopsis* found the AtDeg5 to be a truncated version of a Deg protease (20–22), which suggests that Deg/HtrA proteases can function without containing a PDZ domain. This model is supported by the fact that bacterial DegS protease is fully functional without its PDZ domain (23). Outer membrane porin peptide signals initiate the envelope stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. Group II includes mostly plant Deg proteases such as AtDeg2 (16, 24), which contain a putative PDZ domain and

an elongated C terminus lacking clear similarity to any known protein. All Deg proteases with a chloroplast or mitochondrial targeting sequence are located in group I or II. Group III includes all Deg/HtrA proteases found in fungi, such as the nuclear-localized Nma111p protease of *S. cerevisiae* (25), and some plant proteases. The putative proteases of this group are about twice as long as Deg proteases of groups I and II and possess three PDZ domains, one adjacent to the C terminus of the protease domain and two at the C terminus of the protein. Because the PTS2 in yeast enzymes investigated is not cleaved off on import, it is interesting to note that no *DEG15* paralogue has been identified in fungi. Group IV is formed by *AtDEG15* and related enzymes from other organisms. The proteins of this group carry the protease domain more toward the C terminus than other Deg proteases and, as mentioned, do not contain recognizable PDZ domains. The diverging domain structure of group IV Deg proteases, their placement in the phylogenetic tree, and their capacity to remove the PTS2-containing presequence suggest that this group might be a subfamily of the Deg/HtrA proteases.

The 709-amino acid-long protein sequence of AtDeg15 has a distinct primary structure. The protease domain with 240 amino acids is located at positions 375–614. The N- and C-terminal regions have no obvious homology to other proteins. The primary structure is, however, conserved in the Deg15 protein of rice (SI Fig. 13). The protease domains of AtDeg15 and Os-Deg15 share 45% identical amino acids, compared with 40% identity in the complete sequence. In the two plant Deg15 proteases (SI Fig. 7), 67 amino acids are inserted after the catalytic His, which is located in the short helix B in DegP (13). Because the loop is located in the vicinity of the catalytic pocket, it may determine access to the active center. The loop may be involved in electrostatic interactions due to the presence of many polar and charged amino acids.

It is not unusual for Deg proteases to alternate in two distinct forms. *E. coli* DegP alternates between proteolytic and chaperone functions (12), whereas *E. coli* DegS changes between a catalytically active and inactive form (26). Deg15 may exist as monomers and dimers exhibiting different substrate specificities.

Recently, Tysnd1 was identified as the mammalian peroxisomal processing protease responsible for the removal of the PTS2 peptide and was classified as a cysteine endopeptidase based on inhibitor studies (27). Sequence analysis, however, shows that the highly conserved catalytic triad His-Asp-Ser in the serine protease domain is present in Tysnd1 (SI Fig. 7), whereas no conserved Cys is detectable. Mutation analysis of recombinant *Arabidopsis* Deg15 further confirmed that predicted catalytic His, Asp, and Ser are essential for catalysis (28). The Tysnd1 processing activity using pre-thiolase as a substrate was completely abolished with the cysteine protease inhibitor *N*-ethylmaleimide (27). In the light of our data, we suggest that *N*-ethylmaleimide inhibited the dimer formation of Tysnd1, thus abolishing processing of the specific substrate pre-thiolase. The observed partial inhibition of Tysnd1 by metal chelators removing Ca^{2+} ions supports this interpretation. In both cases, the presumed dimer formation of Tysnd1 would be inhibited.

Structural investigations are required to determine which alternative protein interaction domains GPP/Deg15 uses. Further studies are necessary to understand how the two functions of the enzyme (i.e., limited proteolysis and proteolysis of denatured proteins) are accomplished.

In conclusion, a peroxisomal protease that cleaves PTS2-carrying presequences has been purified to a single protein that immunologically interacted with specific antibodies recognizing the *Arabidopsis* peroxisomal Deg15 endoprotease. Knockout of the Deg15 protein eliminates cleavage of the presequences of the peroxisomal matrix enzymes. This Deg protein lacks PDZ

domains and forms dimers in the presence of Ca²⁺ ions. Monomers and dimers have different substrate specificities.

Methods

Isolation of Glyoxysomes on a Sucrose (Suc) Gradient for Purification of the GPP. Watermelon (*C. vulgaris*) was grown in the dark at 30°C for 3 days, and glyoxysomes were isolated. For details, see *SI Methods*.

GPP Activity Assay. Preliminary experiments using processing of ³⁵S-labeled pre-gMDH as an assay to monitor GPP activity found activity in the 120 mM NaCl eluate of a DEAE-anion exchange column. To simplify the GPP assay, fluorogenic peptides covering the cleavage sites of the higher molecular weight precursor proteins of gMDH, gCS, and thiolase were synthesized; the thiolase peptide with the Cys in P1 exchanged to Gly was used as a control (*SI Table 4*). The fluorogenic peptide anthranilic acid-RVKRGLA-*m*-nitro-tyrosine-D, a substrate for the eukaryotic Ca²⁺-dependent proprotein convertase, was also used for characterization. Because of the low yield and stability of the fluorescent signal of anthranilic acid upon cleavage, these peptides were not suitable for monitoring the GPP purification by screening diluted protease-containing fractions. However, the peptides were highly suitable for functional analysis of the purified and concentrated GPP. The peptide Cbz-FR-AMC, which is a substrate for many serine and cysteine proteases, was used to follow the purification. For fluorogenic peptide synthesis and assay details, see *SI Methods*.

Purification of GPP. Pure glyoxysomes obtained from 90 Suc gradients with an isocitrate lyase activity of 0.14 unit/ml were diluted 1:10 with DEAE-buffer A, sonicated, centrifuged, and filtrated. The resulting glyoxysome extract was purified on a DEAE-anion exchange column followed by a Superdex75 column, a hydroxyapatite column, and a MonoQ column. In a second approach, the glyoxysome extract was purified on a DEAE column, followed by a *p*-aminobenzamidine affinity column. Peak activity fractions were pooled and used for activity gels and Western blot analysis. In a final approach, the purification of GPP was achieved by DEAE chromatography and

sample prefractionation with preparative Sephadex isoelectric focusing, followed by narrow pH range 2D PAGE (29). The gels were either silver stained or used for Western blot analysis with α -AtDeg15 antibodies. The DEAE eluate pool was also subdivided by ultrafiltration in fractions <100 kDa and >100 kDa for analysis by gelatin activity polyacrylamide gels. For details, see *SI Methods*.

Biochemical Procedures. Protease activity was analyzed by non-denaturing (native) PAGE, with SDS only in the running buffer; gelatine was added at a final concentration of 0.05% to the resolving gel. The gels were run at 4°C overnight, equilibrated with buffer, and incubated at either 30°C or 45°C overnight, followed by Coomassie blue staining. Alternatively, native PAGE without gelatine was performed, followed by equilibration with buffer, incubation with the fluorogenic substrate Cbz-FR-AMC, and recorded with a Nikon Coolpix4500 (Tokyo, Japan). Protease inhibitor studies, SDS/PAGE, and immunoblotting using the polyclonal primary rabbit α -Deg15 peptide or the α -gMDH antibody were performed. Anti-Deg15 antibodies were raised against the synthetic peptide ETAAVHPGGSG-GAV comprising 15 amino acids close to the catalytic Ser of the AtDeg15 protease. Anti-gMDH antibodies were raised against the overexpressed protein. Immunoreactive proteins were detected either with horseradish peroxidase-conjugated α -rabbit IgG, visualized by using chemiluminescent substrate, or with alkaline phosphate-conjugated α -rabbit IgG using nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate color reaction. For details, see *SI Methods*.

Bioinformatics. Similarity searches against public protein sequence databases were performed as implemented at the web pages of the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/blast) and the Institute for Genomic Research (www.tigr.org) using default parameters. For details, see *SI Methods*.

This work was supported by Deutsche Forschungsgemeinschaft Grants Gi154/9-4, Gi154/9-5, (to C.G.), and Ad92/8-2 (to I.A.).

1. Keller GA, Krisans S, Gould SJ, Sommer JM, Wang CC, Schliebs W, Kunau W, Brody S, Subramani S (1991) *J Cell Biol* 114:893–904.
2. Flynn CR, Mullen RT, Trelease RN (1998) *Plant J* 16:709–720.
3. Kato A, Hayashi M, Kondo M, Nishimura M (1996) *Plant Cell* 8:1601–1611.
4. Kato A, Takeda-Yoshikawa Y, Hayashi M, Kondo M, Hara-Nishimura I, Nishimura M (1998) *Plant Cell Physiol* 39:186–195.
5. Reumann S, Ma C, Lemke S, Babujee L (2004) *Plant Physiol* 136:2587–2608.
6. Gietl C, Seidel C, Svendsen I (1996) *Biochim Biophys Acta* 1274:48–58.
7. Cox B, Chit MM, Weaver T, Gietl C, Bailey J, Bell E, Banaszak L (2005) *FEBS J* 272:643–654.
8. Mathieu M, Zeelen J, Pauptit RA, Erdmann R, Kunau WH, Wierenga RK (1994) *Structure (London)* 2:797–808.
9. Faber KN, Haima P, Gietl C, Harder W, Ab G, Veenhuis M (1994) *Proc Natl Acad Sci USA* 91:12985–12989.
10. Gietl C, Faber KN, Van der Klei IJ, Veenhuis M (1994) *Proc Natl Acad Sci USA* 91:3151–3155.
11. Authier F, Bergeron JJM, Ou WJ, Rachubinski RA, Posner BI, Walton PA (1995) *Proc Natl Acad Sci USA* 92:3859–3863.
12. Clausen T, Southan C, Ehrmann M (2002) *Mol Cell* 10:443–455.
13. Krojer T, Garrido-Franco M, Huber R, Ehrmann M, Clausen T (2002) *Nature* 416:455–459.
14. Adam Z, Adamska I, Nakabayashi K, Ostersetzter O, Haussuhl K, Manuell A, Zheng B, Vallon O, Rodermel SR, Shinozaki K, Clarke AK (2001) *Plant Physiol* 125:1912–1918.
15. Sokolenko A, Pojidaeva E, Zinchenko V, Glaser VM, Herrmann RG, Shestakov SV (2002) *Curr Genet* 41:291–310.
16. Huesgen PF, Schuhmann H, Adamska I (2005) *Physiol Plant* 123:413–420.
17. Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003) *J Mol Biol* 328:581–592.
18. Rawlings ND, Tolle DP, Barrett AJ (2004) *Nucleic Acid Res* 32:D160–D164.
19. Pallen MJ, Wren BW (1997) *Mol Microbiol* 26:209–221.
20. Adam Z, Clarke AK (2002) *Trends Plants Sci* 7:451–456.
21. Peltier JB, Emanuelson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, van Heijne G, van Wijk KJ (2002) *Plant Cell* 14:211–236.
22. Schubert M, Petersson UA, Haas BJ, Funk C, Schroder WP, Kieselbach T (2002) *J Bio Chem* 277:8354–8365.
23. Walsh NP, Alba BM, Bose B, Gross CA, Sauer RT (2003) *Cell* 113:61–71.
24. Haußühl K, Andersson B, Adamska I (2001) *EMBO J* 20:713–722.
25. Fahrenkrog B, Sauder U, Aebi U (2004) *J Cell Sci* 117:115–126.
26. Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, Clausen T (2004) *Cell* 117:483–494.
27. Kurochkin IV, Mizuno Y, Konagaya A, Sakaki Y, Schönbach C, Okazaki Y (2007) *EMBO J* 26:835–845.
28. Schuhmann H, Huesgen PF, Gietl C, Adamska I (2007) *J Biol Chem*, in press.
29. Görg A, Boguth G, Köpf A, Reil G, Parlar H, Weiss W (2002) *Proteomics* 2:1652–1657.