PHENYLALANINE BIOSYNTHESIS IN Arabidopsis thaliana: IDENTIFICATION AND CHARACTERIZATION OF AROGENATE DEHYDRATASES*
Man-Ho Cho‡, Oliver R. A. Corea§, Hong Yang‡, Diana L. Bedgar‡, Dhrubojoyti D. Laskar‡, Aldwin M. Anterola§, Frances Anne Moog-Anterola‡, Rebecca L. Hood§, Susanne E. Kohalmi§, Mark A. Bernards§, ChulHee Kang¶, Laurence B. Davin‡ and Norman G. Lewis‡

From ‡the Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, USA, §the Department of Biology, University of Western Ontario, London, Ontario, Canada, N6A 5B7 and ¶the School of Molecular Biosciences, Washington State University, Pullman, WA 99164-4660, USA

Running Title: Phenylalanine biosynthesis in Arabidopsis thaliana

Address correspondence to: Norman G. Lewis, Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, Tel. 509-335-8382; Fax. 509-335-8206; Email: lewissn@wsu.edu

There is much uncertainty as to whether plants use arogenate, phenylpyruvate, or both as obligatory intermediates in Phe biosynthesis, an essential dietary amino acid for humans. This is because both prephenate and arogenate have been reported to undergo decarboxylative dehydration in plants via the action of either arogenate (ADT) or prephenate (PDT) dehydratases; however, neither enzyme(s) nor encoding gene(s) have been isolated and/or functionally characterized. An in silico data mining approach was thus undertaken to attempt to identify the dehydratase(s) involved in Phe formation in Arabidopsis, based on sequence similarity of PDT-like and ACT-like domains in bacteria. This data mining approach suggested that there are six PDT-like homologues in Arabidopsis, whose phylogenetic analyses separated them into three distinct subgroups. All six genes were cloned and subsequently established to be expressed in all tissues examined. Each was then expressed as a Nus-fusion recombinant protein in E. coli, with their substrate specificities measured in vitro. Three of the resulting recombinant proteins, encoded by ADT1 (At1g11790), ADT2 (At3g07630) and ADT6 (At1g08250), more efficiently utilized arogenate than prephenate, whereas the remaining three, ADT3 (At2g27820), ADT4 (At3g44720) and ADT5 (At5g22630) essentially only employed arogenate than prephenate, whereas the remaining three, ADT3 (At2g27820), ADT4 (At3g44720) and ADT5 (At5g22630) essentially only employed arogenate. ADT1, ADT2 and ADT6 had $k_{cat}/K_{M}$ values of 1050, 7650 and 1560 M⁻¹ s⁻¹ for arogenate, versus 38, 240 and 16 M⁻¹ s⁻¹ for prephenate, respectively. By contrast, the remaining three, ADT3, ADT4 and ADT5, had $k_{cat}/K_{M}$ values of 1140, 490 and 620 M⁻¹ s⁻¹, with prephenate not serving as a substrate unless excess recombinant protein (>150 μg/assay) was used. All six genes, and their corresponding proteins, are thus provisionally classified as arogenate dehydratases and designated ADT1 through ADT6.

The aromatic amino acid Phe, formed in both plants and microorganisms, serves as a building block for proteins, and as a pathway intermediate to a wide range of aromatic compounds. For mammals, Phe is also an essential dietary requirement. In vascular plants, some of the physiologically important secondary metabolites derived from Phe include pigments and defense molecules (e.g. flavonoids, proanthocyanidins, oligomeric lignans, cyanidins, etc.) (1-3), phytoalexins (e.g. isoflavones) (4), UV protectants (5), as well as the structural lignin and suberin biopolymers (6-8). Taken together, the metabolism of Phe in planta can account for up to 30-40 % of organic carbon depending upon the species (9).

In principle, prephenate [from the shikimate-chorismate pathway (10-12)] can be converted into Phe via either phenylpyruvate with subsequent transamination, and/or via arogenate by transamination of prephenate followed by dehydration/decarboxylation (Fig. 1). With microbes, two different enzymes, prephenate dehydratase (PDT) and arogenate dehydratase (ADT), are able to catalyze the dehydration/decarboxylation reactions of prephenate and arogenate to afford phenylpyruvate and Phe, respectively. Moreover, while most known microorganisms contain PDTs (13-15), an arogenate-specific pathway has also been described in Pseudomonas diminuta and P. vesicularis (16). A third class of enzyme, cyclohexadienyl dehydratase (CDT), common in...
enteric bacteria, can also either utilize prephenate or arogenate as a substrate (17,18). In addition, some microorganisms contain dual-function “P-proteins” with a N-terminal chorismate mutase (CM) domain followed by a PDT domain (19).

Microbial PDTs are often allosterically regulated by binding of pathway end products (19), this being envisaged to occur through a C-terminal ACT domain (derived from aspartokinase, chorismate mutase and TyrA), whereas CDTs are not (18). The “P-proteins” also contain a C-terminal ACT domain that has been shown to bind Phe (19). As a result of these different enzymes types, some bacteria, such as Pseudomonas aeruginosa, can use both routes to Phe (20) whereas others, such as Escherichia coli, only utilize phenylpyruvate (21).

Accordingly vascular plants might be anticipated to employ either one or both of these pathways for Phe biosynthesis. Indeed, in the mid-1980s, two research groups reported ADT activities using partially purified enzyme preparations from Nicotiana silvestris cell cultures and spinach (Spinacia oleracea) chloroplasts (22), as well as from etiolated seedlings of Sorghum bicolor (23). Since then arogenate has been provisionally considered as an intermediate in Phe biosynthesis in plants (7,9,22-27). In agreement with this, a prephenate aminotransferase activity was described in some plant species, such as N. silvestris (28) and etiolated shoots of S. bicolor (29). It has not, however, been possible to detect arogenate in the study of various extracts from a number of vascular plants (26) under the conditions employed, indicating that it must be present (at best) at very low levels. This, however, is quite frequently observed in primary metabolism.

Attempts to establish the presence of PDT activity in higher plants, by comparison, have either been inconclusive (30) (in pea), or unsuccessful (22) (using both N. silvestris and spinach). Yet, a recent report indicated that Phe, nevertheless, results from the activity of PDTs in Arabidopsis (31). Thus it remains unclear in the literature whether Phe results from ADT and/or PDT conversions in plants, and it has become essential to biochemically characterize the corresponding dehydratases involved.

In the TAIR database, there are six Arabidopsis genes, (At1g11790, At3g07630, At2g27820, At3g44720, At5g22630 and At1g08250) putatively annotated as PDT homologues. This annotation is based solely on sequence comparison (and thus a reliance on existing annotations) of bacterial sequences. However, since there is no existing functional linkage between any ADT enzyme activity and the corresponding gene sequence, none are annotated as being arogenate dehydratase-like. Yet recently, Ehlting et al (27) assigned ADT labels (ADT 1-6, respectively) to all six Arabidopsis PDT-like genes, but also without any functional characterization in support of this nomenclature. To investigate whether any of these genes encode functional ADTs, their coding sequences were cloned with the corresponding functional recombinant proteins expressed in E. coli. While all six displayed ADT activity, only ADT’s 1, 2 and 6 showed any level of PDT activities, albeit at much lower levels related to ADT. This study describes the biochemical parameters obtained for each recombinant protein and provides the first functional characterization of ADTs, the phenylalanine-forming machinery, in Arabidopsis.

**Experimental Procedures**

**Instrumentation**—HPLC analyses employed an Alliance 2695 HPLC system (Waters, Milford, MA), equipped with either a NovaPak C18 column (150 × 3.9 mm, Waters), or a Pico-Tag column (300 × 3.9 mm, Waters). Protein purification was carried out on a fast-protein liquid chromatography (FPLC, Amersham Pharmacia Biotech) system.

**Materials**—Phenylmethyl sulfonylfluoride (PMSF) and reagents for making buffers were bought from Sigma; phenylisothiocyanate (PITC) was from Pierce. PD-10 and DEAE-Sepharose were purchased from GE Healthcare, with POROS MC 20 obtained from Applied Biosystems, and Dowex 21K Cl− and AG1-X8 Cl− from Bio-Rad. The expression vector, pET43.1 Ek/LIC, the Rosetta™ (DE3) E. coli cells and BugBuster Protein Extraction Reagent were purchased from Novagen. SuperScript™ II reverse transcriptase, BL21 Star™ (DE3) E. coli cells and TRIzol reagent were procured from Invitrogen, whereas rDNAse1 (DNA-free) was from Ambion and QBioGene Fast RNA ProGreen Kit from QBioGene. PicoTag reversed-phase column (150 × 3.9 mm i.d) and PicoTag sample diluent were...
purchased from Waters (Milford, MA). Prephenate (circa 78% purity) was obtained from Sigma and purified (> 95% purity) using a Chromabond C18 cartridge (Macherey-Nagel) eluted with HPLC grade water. Crude arogenate was isolated from Neurospora crassa (75001/5212/C-167) cultures as described in (32), then purified (33), lyophilized and stored at –80°C until needed.

Cloning of Arabidopsis ADT/PDT-Clones U16103 (for At1g11790) and U11052 (for Atg07630) were obtained from the Arabidopsis Biological Research Center (ABRC), with the open reading frames (ORFs) of each gene individually amplified and subcloned into the expression vector, pET43.1 Ek/LIC, which was designed for cloning and high-level expression of peptide sequences fused with the Nus•Tag™ protein. Primers used for subcloning were 5′-GACGACGACAAGATGGCTCTGCTGAGGTGT-3′ and 5′-GAGGAGAAGCCCGGTATTATCTGACTAGATC-3′ for At1g11790 and 5′-GACGACGACAAGATGGCAATGCACACTGTTCGA-3′ and 5′-GAGGAGAAGCCCGGTATTAGAGCATTGTAGTGTC-3′ for Atg07630, respectively. The remaining ADTs/PDTs (At2g27820, At3g44720, At5g22630 and At1g08250) were individually amplified from an Arabidopsis cDNA library and subcloned into the expression vector, pET44 Ek/LIC. Primers used for subcloning were 5′-GACGACGACAAGATGCGTGTAGCTTATCAAGG-3′ and 5′-GAGGAGAAGCCCGGTATTAGAGCATTGTAGTGTC-3′ for At2g27820, 5′-GACGACGACAAGATGCGTGTAGCTTATCAAGG-3′ and 5′-GAGGAGAAGCCCGGTATTAGAGCATTGTAGTGTC-3′ for At3g44720, 5′-GACGACGACAAGATGCGTGTAGCTTATCAAGG-3′ and 5′-GAGGAGAAGCCCGGTATTAGAGCATTGTAGTGTC-3′ for At5g22630 and 5′-GACGACGACAAGATGCGTGTAGCTTATCAAGG-3′ and 5′-GAGGAGAAGCCCGGTATTAGAGCATTGTAGTGTC-3′ for At1g08250, respectively. Recombinant plasmids, named pADT1 through pADT6, bearing ORFs of At1g11790, Atg07630, At2g27820, At3g44720, At5g22630 and At1g08250, respectively, were transformed into Gigasingles competent cells, according to the manufacturer’s instructions (Novagen). After sequence confirmation, pADT1-pADT6 were individually transformed into either BL21 Star™ (DE3) (pADT1) or Rosetta™ (DE3) (pADT2-pADT6) E. coli cells for expression of recombinant proteins.

Tissue Preparation and RNA Isolation—Leaf, stem, floral and whole silique tissues were harvested from at least four week-old (flowering) wild type A. thaliana plants (TAIR accession number CS3879, ecotype Columbia) grown in sterile soil under 24 h fluorescent light (150 μE/m²/s) at 20°C. For root tissues, plants were initially grown on 0.5 × Murashige and Skoog basal medium with Gamborg’s vitamins (MS; Sigma) and then transferred into 1 × MS liquid medium and grown to maturity. Media were supplemented with PPM (Plant Cell Technology). Upon harvest, all tissues were immediately submerged into liquid nitrogen. To isolate total RNA, 100 mg of each plant tissue was lysed in 1 ml TRIzol (Invitrogen) containing a large lysing matrix bead and shaken in a FastPrep machine (BIO101/QBioGene) once (flower and root) or twice (leaf, stem, silique) for 45 sec (setting 4). RNA was isolated using the QBioGene Fast RNA ProGreen Kit according to the manufacturer’s instructions.

Reverse Transcriptase (RT)-PCR analysis—RNA (10 μg) was treated with rDNase1 (2 units) to remove single- and double-stranded DNA, and 2 μg was reverse transcribed using SuperScript™ II reverse transcriptase (200 units). PCR amplification was performed using primers pairs specific to the 3′UTR region of each ADT (ADT1-F: 5′-AATTGATTTTTAGATGACAG-3′; ADT1-R: 5′-TCTCAACATGAAGTGTGCA-3′; ADT2-F: 5′- TCTAATCCATTTATCCATCC-3′; ADT2-R: 5′-
AAACTAAACAGAATTGGTAACAAGAGC-3'; ADT3-F: 5'-CAACGTTGAAAGCTCAATTGTG-3'; ADT3-R: 5'-CGATCAAAGCAAACCTCAAAG-3'; ADT4-F: 5'-TCATTATCAGTGTTAGG-3'; ADT4-R: 5'-TTTGCTTTCTGGTTAAAAATCG-3'; ADT5-F: 5'-TCGTTTCTGATGATGTTGG-3'; ADT5-R: 5'-TCATCGAAGCAAACTACAGAACC-3'; ADT6-F: 5'-TGTTAACATTTGGAAGAGATAACAAAG-3'; ADT6-R: 5'-TTTACCAATTGGTTATGTATACGG-3'). To ensure that PCR products originated solely from mRNA, each template was PCR amplified using a primer set that specifically recognizes a genomic DNA sequence which is not transcribed (GEN-F: 5'GCTTTCATGTTTTAGCAATGGCG-3'; GEN-R: 5'ATTAATCTTCTGGGATGCGCCG-3'). The absence of amplification of a 600 bp product in DNAse treated versus untreated samples demonstrates that experimental PCR amplifications are solely based on RNA templates. PCR products were sequenced to ensure that the proper ADT was amplified. As a control, PCR was also performed for all RNA extracts using an RNA-specific primer set (TUB2-F: 5'-ATGCTAAGGGGT TTCAGTGG-3'; TUB2-R: 5'-GCACGAGGA CGAATCTACC-3') that amplifies the ubiquitously expressed gene TUBULIN2 (TUB2) (34,35). All PCR amplifications were performed according to the following protocol: 2 min at 94°C; 35 cycles of: 15 sec at 94°C, 1 min at 55°C, 2 min at 68°C; and a final extension for 10 min at 68°C. PCR products were size-separated in a 6% polyacrylamide gel and stained with ethidium bromide.

Expression in E. coli and Purification of Recombinant Nus-Proteins—The E. coli transformants, individually harboring pADT1 – pADT6, were grown at 37°C in LB media supplemented with carbenicillin (0.1 mg ml⁻¹) until an OD₆₀₀ ~0.9 was reached at which time isopropyl thio-β-D-galactoside (IPTG, 0.3 mM) was added. Cells were grown at 16°C for 20 – 24 h, individually harvested by centrifugation (3,000 × g for 20 min) and stored at ~20°C until required. The cell pellets were individually resuspended in BugBuster Protein Extraction Reagent (Novagen) according to the manufacturer’s instructions, with the cell debris removed by centrifugation (10,000 × g, 20 min). These crude protein preparations were individually applied to a DEAE-Sepharose column (1.6 × 3 cm) equilibrated in buffer A (50 mM potassium phosphate, pH 7.5), and eluted with buffer A containing 1 M NaCl. Fractions from the flow-through were combined and further subjected to POROS MC 20 column (1 × 9 cm) chromatography with Cu²⁺ (CuSO₄) as metal ions (36). After washing with buffer B (20 mM Tris·HCl, pH, 7.9, 500 mM NaCl, 50 ml) containing 20 mM imidazole, recombinant proteins were individually eluted with an imidazole gradient (20 – 500 mM, 100 ml) in buffer B. Fractions containing each recombinant protein (i.e. eluting between 50 and 130 mM imidazole) were combined and buffer-exchanged to Buffer A and concentrated to a small volume (~0.5 ml) using a Centricon Plus-20 (Amicon). Due to the instability of the ADT proteins, the purification steps described above (from cell harvest to affinity chromatography) were carried out over a 24 h duration, with purified proteins immediately used for enzyme assays.

PDT Assays—PDT activities were determined by measuring phenylpyruvate formation. Each assay consisted of prephenate (1 mM), recombinant protein (15 μg for ADT1 and ADT6, 2 μg for ADT2, and >150 μg for ADT3 – ADT5) and buffer A in a total volume of 50 μl. The enzymatic reaction was initiated by individual addition of recombinant protein and, after incubation for 30 min at 37°C, MeOH (50 μl) was added to stop the reaction. An aliquot (50 μl) of each assay mixture was then subjected to reversed-phased HPLC (NovaPak C₁₈, Waters) analysis. Phenylpyruvate and prephenate were separated using a linear MeOH gradient (0 to 18%) in eluant A (20 mM sodium phosphate, pH 6.9) at a flow rate of 1 ml min⁻¹, with detection at 210 nm. The amount of phenylpyruvate was determined using an external calibration curve.

ADT Assays—ADT activities were measured by determining the amounts of phenylthiocarbamyl derivatives of Phe (26). Each assay mixture consisted of arogenate (1 mM), the recombinant protein (100 ng for ADT1, ADT3 – ADT6 and 50 ng for ADT2) and potassium phosphate buffer (200 mM, pH 7.5), in a total volume of 50 μl. The enzymatic reaction was initiated by individual addition of recombinant
protein. After incubation at 37°C for 30 min, reaction mixtures were immediately placed on ice to stop the reaction and freeze-dried. Lyophilized reaction mixtures were subsequently individually derivatized with phenylisothiocyanate (PITC) as previously described (26). After drying in vacuo, the reaction mixtures were reconstituted in 100 µl of PicoTag Sample Diluent. An aliquot (50 µl) was subjected to HPLC analysis on a PicoTag reversed-phase column. The phenylthiocarbamyl (PITC) derivatives were eluted with a linear gradient (0 to 100%) of eluant B (CH₃CN-MeOH-H₂O, 45:15:40) to eluant C (0.07 M sodium acetate, titrated to pH 6.5 with glacial acetic acid-CH₃CN, 97.5:2.5) over 20 min at 1 ml min⁻¹. Detection was at 254 nm. Amounts of Phe formed were determined using an external calibration curve.

Control assays were conducted for both arogenate and prephenate activities as described above but in absence of recombinant proteins.

**Results and Discussion**

In silico Analysis of Putative PDT and ACT Domains in Bacterial and Plant Genes and Proteins—While ADT activity has been biochemically detected in bacteria, plants and green algae (16,22,23,37,38), there are no reports of molecular studies of any putative ADT genes in any organisms, only that of PTDs and CDTs. Our in silico screening, however, resulted in the provisional identification of six putative PDT domain-containing genes in Arabidopsis (At1g08250, At1g11790, At2g27820, At3g07630, At3g44720 and At5g22630) as mentioned above. Additionally, 35 possible ADT/PDT sequences were identified for other plant species, i.e. in Allium cepa, Glycerine max, Gossypium hirsutum, Hordeum vulgare, Lotus japonicus, Lycopersicon esculentum, Medicago truncatula, Nicotiana benthamiana, Oryza sativa, Pinus taeda, Saccharum officinarum, Solanum tuberosum, Triticum aestivum, Vitis vinifera and Zea mays, in available databases (NCBI, TIGR), based on sequence similarity to the six Arabidopsis genes (nucleotide and translated amino acid sequences). All plant putative ADTs/PDTs contained three distinct (albeit putative) domains at the amino acid level: a N-terminal transit peptide, followed by a PDT-like domain and a C-terminal ACT or ACT-like domain (Fig. 2A). By contrast, all bacterial PDT sequences lack the N-terminal signal peptide extension, while P-proteins have an N-terminal CM domain. While previously only two of the ADT/PDTs were annotated as possessing a partial (At1g11790) or full (At3g07630) ACT domain, our analysis suggests that the remaining four also do as well. All share many of the conserved amino acids (Fig. 2B) considered important for feedback regulation in bacteria (see below). However, whether these enzymes utilize prephenate, arogenate, or both, as a substrate could not be predicted by sequence analysis alone.

Alignments also indicate that there are many conserved amino acids throughout the putative ADT/PDT domains in the selected plant and bacterial sequences examined (Fig. 2B). However, the rooted phylogenetic tree, constructed from amino acid sequences encompassing ADT/PDT and ACT domains from all species included, suggests that the plant and bacterial sequences are evolutionary quite distinct from one another (Fig. 3 and Supplementary Fig. 1). The plant sequences seem to form three distinct subgroups that are conserved across monocots and dicots. Interestingly, the Arabidopsis genes that segregated into subgroup I (At1g11790) and II (At3g07630) contain introns, while the remaining four (At2g27820, At3g44720, At5g22630 and At1g08250), which cluster within a subgroup III, do not (data not shown). Further distinction between Arabidopsis ADTs/PDTs is evident in the amino acid sequence similarities within their catalytic (ADT/PDT) domains. That is to say, all four Arabidopsis ADTs/PDTs from subgroup III show a higher degree of ADT/PDT domain amino acid similarity to each other (81-98%) than to either of the other two ADTs (61-72%) (Table 1). The conservation of these subgroups may denote unique roles (e.g. in distinct metabolic branches) for the members of each group (discussed later).

In vivo Expression of Arabidopsis ADTs/PDTs—It was next important to establish whether any of the ADT/PDT genes were differentially expressed in Arabidopsis tissues (e.g. roots, leaves, stems, flowers and siliques) of flowering Arabidopsis plants. However, since only two of the six ADT/PDT genomic sequences have introns we were unable to design primers that span intron-exon borders, as is normally done to create RNA-specific primers. Therefore, to ensure that RT-PCR amplifications solely derived from RNA
templates (and not contaminant genomic DNA in our preparations), each isolated RNA sample was treated with RNase1 prior to reverse transcription. The cDNAs from the DNase treated and untreated samples were then subsequently tested using a primer set complementary to a non-transcribing genomic sequence, which generates a 600 bp PCR product only from genomic DNA. Absence of this 600 bp PCR fragment in DNase1 treated samples confirmed that the samples were completely free from residual contaminating genomic DNA sequences. Only those RNA samples free from residual contaminating genomic DNA were used for the expression analyses. As a further control TUBULIN2, a constitutively expressed gene (34,35), was also amplified for each sample. From the data thus obtained, all six ADTs/PDTs appeared to be expressed in all tissues from which RNA was isolated, including roots, leaves, stems, flowers, and siliques (Figure 4). When comparing results from the different ADTs/PDTs, however, variations in band intensities cannot be interpreted as differences in ADT/PDT expression levels since different primers with potentially different amplification properties were used for each one. Since all ADT/PDT genes appear to be expressed in all tissues examined, it became essential to establish whether all members of the Arabidopsis ADT/PDT family possessed ADT or PDT enzymatic activity, or both, through recombinant protein expression and kinetic analyses.

Expression of Recombinant ADT/PDT Proteins—Each ADT was thus successfully expressed as a Nus-fused recombinant protein in E. coli in soluble His-tagged form using the pET 43.1 EK/Lic vector. This approach was necessary to overcome general inclusion body problems (39-42) that were initially encountered when many other attempts to express His-tagged proteins all failed. Each protein was then individually purified to apparent homogeneity by ammonium sulfate fractionation, DEAE-Sepharose column chromatography and Cu²⁺-affinity chromatography, respectively (data not shown); all manipulations including assays were carried out within 24 h. Following SDS PAGE analyses, the apparent molecular weight of each was estimated to be ~110 kDa, in good agreement with the deduced molecular weight of the Nus-fused ADT/PDTs (data not shown).

Arabidopsis ADT/PDTs Encode Functionally Competent ADTs—Individual enzyme assays performed using either arogenate or prephenate as potential substrates indicated that the six recombinant ADT/PDT homologues were catalytically active, i.e. all six gene products encoded a functional dehydratase. For detailed kinetic studies, enzyme reactions were carried out at pH 7.5, since arogenate and prephenate are unstable under acidic conditions and can be spontaneously converted into Phe and phenylpyruvate, respectively (43). The recombinant At1g11790, At3g07630 and At1g08250 proteins all had relatively similar affinities for both arogenate and prephenate (\(K_m\)'s ranging from 0.80 to 3.05 mM for arogenate, and 0.68 to 2.44 mM for prephenate, respectively; Table 2). These \(K_m\) values are only slightly greater than previously reported ADT activities in S. bicolor (23), E. herbicola (17), and P. aeruginosa (18), which ranged from 0.09 to 0.32 mM and PDT activities in Erwinia herbicola (17), P. aeruginosa (18), E. coli (19,44), Mycobacterium tuberculosis (15), and C. glutamicum (45) which ranged from 0.07 to 0.56 mM. The slightly higher \(K_m\) values for the ADTs may be due to the presence of the Nus-tag, as fusion tags have been shown to affect the \(K_m\) of other recombinant proteins (46). On the other hand, the calculated maximal velocities (\(V_{max}\)) of recombinant At1g11790, At3g07630 and At1g08250 were significantly higher for arogenate (31, 60.6 and 42.61 pkat\(\mu\)g\(^{-1}\) protein) than for prephenate (0.28, 1.6, and 0.4 pkat\(\mu\)g\(^{-1}\) protein) (Table 2). Moreover, the overall catalytic efficiencies (\(k_{cat}/K_m\)) for arogenate were ~28, 32, and 98 fold higher than that of prephenate (Table 2). These data thus indicate that these three gene products are more ADT- rather than PDT-like, and we thus provisionally designate At1g11790, At3g07630m and At1g08250 as ADTs (ADT1, ADT2 and ADT6, respectively, using the nomenclature of Ehtling et al. (27)).

The remaining three gene products (At2g27820, At3g44720 and At5g22630) also displayed a substrate preference for arogenate with \(K_m\) values ranging from 0.43 to 10.08 mM, and \(V_{max}\) values from 52.32 to 5.17 pkat \(\mu\)g\(^{-1}\) protein. The catalytic efficiency (\(k_{cat}/K_m\)) of At2g27820 (1140 M\(^{-1}\) s\(^{-1}\); Table 2) was similar to that of ADT1 and ADT6; the \(k_{cat}/K_m\) values of At3g44720...
and At5g22630 were lower (e.g. 490 and 620 M⁻¹ s⁻¹, respectively; Table 2). More striking, however, was the nearly complete lack of any detectable PDT activity of recombinant proteins derived from these three proteins (Table 2), with only minute levels of activity detected at very high protein concentration (>150 μg). Thus these three gene products are also provisionally designated as the functional ADTs, ADT3, ADT4 and ADT5 (27).

As these enzymes are the first monofunctional ADTs to be individually expressed and assayed, it was also important to demonstrate that the conditions for the prephenate and arogenate assays did not favor ADT over PDT activity. Therefore, a previously characterized PDT from Methanocaldococcus jannaschii (MjPDT; courtesy of Dr. Peter Kast) was assayed using the same conditions, with 0.2 μg of enzyme in each reaction. The $K_m$ and $k_{cat}$ values obtained for prephenate were somewhat lower than that reported by Kleeb et al. (47); 370 vs. 22 μM and 8.9 vs. 12.3 s⁻¹, respectively. In our hands, the differences in $K_m$ and $k_{cat}$ were most likely due to the widely different assay conditions used. However, the arogenate activity of MjPDT was not previously tested. We found that ADT activity was virtually absent when using 0.2 μg of enzyme in each sample, and a minute conversion to Phe was only observed with addition of excess (5 to 10 μg) enzyme.

These data are in stark contrast to those of Warpeha et al. (31), who reported an indirect prephenate conversion into phenylpyruvate in crude protein extracts from WT Arabidopsis, but not PD1 mutants (deficient in At2g27820; ADT3). These authors did not, however, test arogenate conversion into Phe, precluding any comparison between PDT and ADT activities. Our ADT3 Nus-fusion protein, however, had essentially no PDT activity in vitro, while displaying substantial ADT activity. It is possible that there may be some residual PDT activity in the native protein. However, even this does not seem likely since ADT1, but not ADT3, was able to complement a PDT-deficient yeast mutant (pha2; (48)) (Corea, Bernards, Kohalmi, unpublished data).

Secondary Structure, Mode of Catalysis and Regulation Predictions—In spite of the low degree of identity noted earlier between bacterial and plant sequences, their corresponding ADT/PDT domains seem all to be composed of 5 or 6 α-helices as suggested by secondary structure predictions (data not shown). In addition, many of the specific catalytic sites identified in bacterial PDTs were conserved in the Arabidopsis homologues. Specifically, residues important for PDT catalysis in E. coli [N160, S208, Q215 and T274 (44) (Fig. 2B, ▲)] and Corynebacterium glutamicum [(E64, S99, T183, (45) (Fig. 2B, ◆)], were also present in Arabidopsis ADTs/PDTs). Moreover, the Thr identified in each of these is part of a TRF-motif that might be involved in either catalytic activity or substrate binding (44,45,49). Indeed, even though there are no crystal structures for ADTs/PDTs from any species, this well characterized TRF-motif is predicted to be in an α-helix. Furthermore, while CDTs have little apparent sequence similarity to ADTs/PDTs (see below), they nevertheless also have a Thr at this motif (18), perhaps indicative of its importance for enzyme function. The arginine in the TRF-motif (Fig. 2B, ◇) is also absolutely conserved across the 70 plant and bacterial species so far surveyed (i.e., the same as those in Supplementary Figure 1). Conservation of these amino acid residues in Arabidopsis ADTs thus also support the hypothesis that these enzymes potentially function as aromatic ring-forming dehydratases.

Sequence similarity within the ACT domain may suggest similar functions, since it is the 3D structure of this domain that is important for feedback regulation (for review, see (50)). Indeed, sequence alignments with Phe hydroxylase, together with mutational analysis and isothermal titration calorimetry of the E. coli P-protein, identified two candidate regulatory regions predicted to be involved in Phe binding and feedback inhibition, i.e. 309GALV312 and 329ESRP332, respectively (see Fig. 2B, △) (51,52). The ESRP-motif, which is the most hydrophilic region in the P protein regulatory domain, is predicted to be a part of a loop and is conserved across most ADT/PDT sequences examined, including all six Arabidopsis ADTs (see Fig. 2B). On the other hand, the GALV motif, which is the most hydrophobic region of the regulatory domain of the P protein is very likely part of a secondary structural element (α-helix). However, it is only partially conserved across these same sequences, with Gly and Leu having the highest degree of
conservation across bacterial and plant ADTs/PDTs. Meanwhile, another residue, W388 (not shown), which is also proposed to be involved in Phe binding (37), is not conserved; instead, it is Leu, Ile or Val in Arabidopsis ADTs. Furthermore, mutational analysis of _C. glutamicum_ PDT identified two more amino acids in the ACT domain, R202 and G224 (see Fig. 2B, C), that caused resistance to Phe-mediated feedback inhibition of PDT activity (53). These amino acids were conserved as Lys and Ala, respectively, in many other bacterial PDTs, as well as in the six Arabidopsis ADTs. In summary, the high degree of similarity within the ACT domain of Arabidopsis ADTs, and the high number of conserved residues with bacterial PDTs provisionally suggest that all six Arabidopsis ADT genes could possess functional ACT domains.

Finally, when the protein sequence of the cyclohexadienyl dehydratases (CDT) from _P. aeruginosa_ were compared to that of the ADTs/PDTs from both plants and bacteria, it clearly resided as an outlier (Fig. 3), suggesting little sequence similarity. This is in agreement with earlier reports that CDTs are distinct from bacterial PDTs (18), and confirms that they are even more distantly related to putative plant ADTs/PDTs. There were no homologues for CDT-like proteins found in the Arabidopsis genome.

**Concluding Remarks**

Although it has been proposed that vascular plants preferentially form Phe biosynthesis via arogenate (22,23), no direct evidence, e.g. purification of an ADT, cloning of an ADT-encoding gene and/or expression of a functionally competent recombinant protein, has been reported. In this study, we cloned all six putative Arabidopsis ADTs/PDTs predicted as possibly involved in Phe biosynthesis. Interestingly, all six appeared to be expressed throughout the plant, preventing strict assignment of individual isoforms to specific cellular processes at this time. Future work will, therefore, determine to what extent the ADTs are differentially expressed in the various cell types of each tissue at various points of growth/development, using, for example, quantitative real time PCR.

All six genes were subsequently expressed in active recombinant form in _E. coli_, and each was shown to preferentially catalyze the decarboxylative dehydration of arogenate to afford Phe (Table 2). Three of these, ADT1, ADT2 and ADT6, were also able to catalyze the decarboxylation/dehydration of prephenate directly, while the remaining three had an apparent strict requirement for arogenate. Our biochemical studies also indicated that the ADT activities were >28 times higher than that of PDT (Table 2), where PDT activity was even detected, providing support that Arabidopsis preferentially utilizes arogenate in Phe biosynthesis. Phylogenetic analysis places the six Arabidopsis ADT genes into three distinct subgroups (Fig. 3, Supplementary Fig. 1). The three ADTs studied herein that showed both ADT and PDT activity were distributed across all three subgroups, whereas the remaining three were restricted to subgroup III. In future, more extensive studies will be needed both in vitro (e.g. to investigate feedback analysis of all six Arabidopsis ADTs, and site-directed mutagenesis targeting ADT and ACT domain residues), as well as in vivo (e.g. gene knockout, expression) studies. These are necessary in order to fully understand the last steps of Phe biosynthesis, and the involvement of each isoform (e.g. in the various possible metabolic pathways derived from Phe). Other detailed studies will also be directed towards establishing the catalytic mechanism of ADTs and their 3D structures.

**Acknowledgments.** The authors thank Dr. Peter Kast for the _Methanocaldococcus jannaschii_ (MjPDT) PDT.
References

FOOTNOTES

*This research was supported in part by the National Science Foundation (MCB-0417291), the United States Department of Energy (DE-FG-0397ER20259), the National Institutes of Health (GM66173), the G. Thomas and Anita Hargrove Center for Plant Genomic Research and the Natural Sciences and Engineering Research Council of Canada (MAB, SEK). The cost of publication of this article was defrayed in part by the payment of page charges. This article must therefore be hereby-marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1The abbreviations used are: ADT, arogenate dehydratase; CM, chorismate mutase; ACT, domain whose name derives from aspartokinase, chorismate mutase and TyrA (prephenate dehydrogenase); CDT, cyclohexadienyl dehydratase; IPTG, isopropyl thio-β-D-galactoside; PDT, prephenate dehydratase; PITC, phenylisothiocyanate; PMSF, phenylmethyl sulfonylfluoride; PPM, plant preservative medium.

FIGURE LEGENDS

FIG. 1. Proposed biosynthetic pathway from prephenate and arogenate to Phe in plants and microorganisms.

FIG. 2. Domain structure and amino acid sequence alignment of putative plant ADTs/PDTs, bacterial PDTs and bacterial P-proteins. (A) Schematic representation of domain arrangements. Putative domains are: ADT/PDT, arogenate/prephenate dehydratase; PDT, prephenate dehydratase; ACT, regulatory domain; TP, (putative) N-terminal transit peptide and CM, chorismate mutase. (B) Partial sequence alignment using a highly conserved C-terminal portion of the putative ADT/PDT (underlined in blue) and the N-terminal portion of the putative ACT domains (underlined in green). The boundaries between the two domains are as described for E. coli P-protein by Zhang et al. (19). Amino acids with similar properties are shown in the same color. Amino acid residues that are either identical or similar in ≥66% of sequences are shaded in green or gray, respectively. Additional symbols below the alignment indicate residues identified by mutagenesis to be involved in catalytic activity in either E. coli (▲ and △), or C. glutamicum (● and ○) (43,44,48,50,51), as well as the absolutely conserved arginine of the TRF motif (◊). Sequences marked with an asterisk are annotated as having an ACT domain. Numbers on the right hand side are in reference to the N-terminal end of each protein, i.e. the beginning of the transit peptide (Arabidopsis), chorismate mutase domain (E. coli, Haemophilus influenzae, and P. aeruginosa), or PDT domain (all other species). BjPDT = Bradyrhizobium japonicum, NP_768061; CgPDT = Corynebacterium glutamicum, AAA23304; CrPDT = Chlorobium tepidum, AMM72891; EcPDT = Escherichia coli, ZP_00708880; EfPDT = Enterococcus faecium, EAN09096; HiPDT = Haemophilus influenzae, P43900; LlPDT = Lactococcus lactis, AAK05840; MlPDT = Mesorhizobium loti, BAB51940; PaPDT = Pseudomonas aeruginosa, AAG06554; and SpPDT = Streptococcus pneumoniae, AAK75467.

FIG. 3. Phylogenetic relationship between selected plant and bacterial ADTs/PDTs. The rooted phylogenetic tree, generated with DNAMAN using a bootstrap of 1000, compares amino acid sequences containing both ADT/PDT and ACT domains. Numbers at the branch points give the bootstrapping values. Horizontal scale indicates sequence divergence. The clustering of bacterial sequences into Gram positive and Gram negative groups, as well as plant ADT/PDT subgroups are delineated with dashed lines. The CDT sequence for P. aeruginosa is included as a representative of this group. Bacterial sequences used are as in Figure 2 legend as well as PaCDT = Pseudomonas aeruginosa, AAC08596 and Spo = Schizosaccharomyces pombe, CAB10811. One O. sativa sequence (Os07g32774), which does not appear to contain an ACT domain, was omitted from the analysis. A more complete phylogenetic tree, which includes an additional ~ 40 plant and bacterial sequences, is presented in the supplementary material that accompanies this article.

FIG. 4. RT-PCR analyses of Arabidopsis ADTs. mRNA for RT-PCR analysis was collected from mature flowering plants as described in Experimental Procedures. PCR products were size-separated in a 6%
polyacrylamide gel and stained with ethidium bromide. R: root L: leaf S: stem Si: silique F: flower. The amplification products for ADTs can be distinguished by size: $ADT1$: 146 bp; $ADT2$: 135 bp; $ADT3$: 145 bp; $ADT4$: 176 bp; $ADT5$: 117 bp; $ADT6$: 103 bp; $TUB2$: 204 bp. M: 100 bp ladder; the 200 bp and 100 bp markers are shown for all ADTs, while for $TUB2$ the 200 bp and 300 bp markers are visible.
Table 1. Amino acid sequence identity of the PDT domains of putative Arabidopsis ADTs/PDTs. Sequences were aligned using ClustalW, then identity matrices were created using BioEdit, version 7.0.0. Genes are named according to the nomenclature of Ehlting et al. (26).

<table>
<thead>
<tr>
<th>Gene ID (PDT)</th>
<th>At3g07630 (ADT2)</th>
<th>At2g27820 (ADT3)</th>
<th>At3g44720 (ADT4)</th>
<th>At5g22630 (ADT5)</th>
<th>At1g08250 (ADT6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g11790 (ADT1)</td>
<td>64.2</td>
<td>62.4</td>
<td>60.7</td>
<td>60.7</td>
<td>64.0</td>
</tr>
<tr>
<td>At3g07630 (ADT2)</td>
<td>72.3</td>
<td>67.4</td>
<td>68.5</td>
<td>73.4</td>
<td></td>
</tr>
<tr>
<td>At2g27820 (ADT3)</td>
<td>81.2</td>
<td>82.3</td>
<td>97.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At3g44720 (ADT4)</td>
<td>92.2</td>
<td>81.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5g22630 (ADT5)</td>
<td></td>
<td></td>
<td>82.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Kinetic parameters for the arogenate- and prephenate dehydratase activities of recombinant protein derived from six Arabidopsis ADT/PDT genes.

<table>
<thead>
<tr>
<th>Arabidopsis Gene ID</th>
<th>Protein designation</th>
<th>Arogenate</th>
<th>Substrate</th>
<th>Prephenate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_m$ (mM)</td>
<td>$V_{max}$ (pkat μg$^{-1}$)</td>
<td>$k_{cat}$ (s$^{-1}$)</td>
</tr>
<tr>
<td>At1g11790 (ADT1)</td>
<td>3.05</td>
<td>31.00</td>
<td>3.2</td>
<td>1050</td>
</tr>
<tr>
<td>At3g07630 (ADT2)</td>
<td>0.80</td>
<td>60.60</td>
<td>6.1</td>
<td>7650</td>
</tr>
<tr>
<td>At2g27820 (ADT3)</td>
<td>0.43</td>
<td>5.17</td>
<td>0.5</td>
<td>1140</td>
</tr>
<tr>
<td>At3g44720 (ADT4)</td>
<td>10.08</td>
<td>52.32</td>
<td>4.9</td>
<td>490</td>
</tr>
<tr>
<td>At5g22630 (ADT5)</td>
<td>1.72</td>
<td>11.26</td>
<td>1.1</td>
<td>620</td>
</tr>
<tr>
<td>At1g08250 (ADT6)</td>
<td>2.58</td>
<td>42.61</td>
<td>4.0</td>
<td>1560</td>
</tr>
</tbody>
</table>

$^1$Nomenclature from Ehtling et al. (27)

$^2$nd= not determined
Figure 1
**Figure 2**

(A) Schematic representation of the predicted ADT/PDT and P-protein domains in bacterial and plant proteins. The domain organization is indicated by different colors: red for ADT/PDT, blue for P-protein, and black for other domains. The relative positions of the domains are shown for selected species.

(B) Alignment of the amino acid sequence of bacterial ADT/PDT and P-protein domains with their plant counterparts. The sequences are aligned using ClustalW, and the conserved residues are highlighted in red. The alignment shows the conservation of key residues across the species, indicating their importance in the function of these domains.

The figure illustrates the conservation of key residues across various species, highlighting the importance of these domains in bacterial and plant metabolism. The alignment provides insights into the evolutionary conservation of these domains and their potential roles in specific biological processes.
Figure 3
Figure 4
Phenylalanine biosynthesis in Arabidopsis thaliana: Identification and characterization of arogenate dehydratases
Man-Ho Cho, Oliver R. A. Corea, Hong Yang, Diana L. Bedgar, Dhrubojyoti D. Laskar, Aldwin M. Anterola, Frances-Anne Moog-Anterola, Rebecca L. Hood, Susanne E. Kohalmi, Mark A. Bernards, ChulHee Kang, Laurence B. Davin and Norman G. Lewis

J. Biol. Chem. published online August 28, 2007

Access the most updated version of this article at doi: 10.1074/jbc.M702662200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2007/08/29/M702662200.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2007/08/28/jbc.M702662200.citation.full.html#ref-list-1