Siroheme Biosynthesis in Higher Plants

ANALYSIS OF AN S-ADENOSYL-L-METHIONINE-DEPENDENT UROPORPHYRIN GEN III METHYLMTRANSFERASE FROM ARABIDOPSIS THALIANA*

(Received for publication, May 22, 1996, and in revised form, August 29, 1996)

Thomas Leustek‡, Michael Smith‡, Michael Murillo‡, Davinder Pal Singh†, Alison G. Smith‡, Sarah C. Woodcock**, Sarah J. Awan**, and Martin J. Warren**

From the ‡Center for Agricultural Molecular Biology and Plant Science Department, Rutgers University New Brunswick, New Jersey 08903-0231, *Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom, and **Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, United Kingdom

Siroheme, the prosthetic group for both nitrite and sulfite reductases, is a methylated, iron-containing modified tetrapyrrole. Here we report the first molecular characterization of the branch point enzyme in higher plants, which directs intermediates toward siroheme synthesis. A cDNA was cloned from Arabidopsis thaliana (UPM1) that functionally complements an Escherichia coli cysG mutant, a strain that is unable to catalyze the conversion of uroporphyrinogen III (Uro'gen-III) to siroheme. UPM1 is 1484 base pairs and encodes a 369-amino acid, 39.9-kDa protein. The UPM1 product contains two regions that are identical to consensus sequences found in bacterial Uro'gen-III and precorrin methyltransferases. Recombinant UPM1 protein was found to catalyze S-adenosyl-l-methionine-dependent transmethylation by UPM1 in a multistep process involving the formation of a covalently linked complex with S-adenosyl-l-methionine. The UPM1 product has a sequence at the amino terminus that resembles a transit peptide for localization to mitochondria or plastids. The protein produced by in vitro expression is able to enter isolated intact chloroplasts but not mitochondria. Genomic blot analysis showed that UPM1 is encoded in the A. thaliana genome. The genomic DNA corresponding to UPM1 was cloned and sequenced and found to contain at least five introns.

The biosynthesis of siroheme was elucidated recently but only in the enteric bacterium Escherichia coli (1, 2). This methylated, iron-containing isobacteriochlorin functions as the prosthetic group of both sulfite and nitrite reductases (3) and shares part of its biosynthetic pathway with other modified tetrapyrroles including chlorophylls, bilins, hemes, and cobalamins (vitamin B₁₂) (4). The tetrapyrrole pathway begins with the synthesis of 5-aminolevulinate (ALA)¹ either by the condensation of succinyl CoA and glycine, as in animals, yeast, and the α-gene of the proteobacteria (5); or by the transformation of glutamate, as in higher and lower plants and most other bacteria. Eight molecules of ALA are converted into uroporphyrinogen III (Uro'gen-III) by the concerted action of three enzymes, namely ALA-dehydratase, porphobilinogen deaminase, and Uro'gen-III synthase (Fig. 1). Uro'gen-III serves as the central template from which all biologically functional modified tetrapyrroles are derived by a number of enzyme-mediated modifications that include peripheral alterations to the acetate and propionate side chains, methylations, oxidation state of the macrocycle, ring size, and the central metal ion. The siroheme biosynthetic branch deviates from the heme and chlorophyll pathway by methylation of Uro'gen-III at positions 2 and 7 to form precorrin-2 (PC-2); siroheme synthesis is completed after dehydrogenation and ferrochelation. PC-2, known as dihydrodiosirohydrochlorin, is also the precursor for cobalamins (vitamin B₁₂). To date, the biosynthesis of corrinoids appears to be confined to certain microorganisms (6).

A gene involved specifically in siroheme synthesis in E. coli, named cysG, was first identified by isolation of cysteine auxotrophs (7). This mutant is also unable to utilize nitrate as a source of nitrogen, and the cysG gene is physically located within the same transcriptional unit as nirB encoding nitrite reductase (8). Recent functional analysis of the cysG product revealed that it is a trifunctional enzyme catalyzing the three reactions that convert Uro'gen-III to siroheme (1, 2). The complete enzyme, 457 amino acids, consists of a carboxyl domain from amino acids 202 to 457, termed CysG₃, that catalyzes S-adenosyl-l-methionine (SAM)-dependent methylation of Uro'gen-III; and an amino-terminal domain encompassing the first 201 amino acids, termed CysG₄, that catalyzes NAD⁺-dependent dehydrogenation of PC-2 and ferrochelation. Expression in E. coli of the CysG₃ domain results in accumulation of PC-2 and a trimethylated product termed a trimethylporphyrin (Fig. 1), a compound with no known biological function that is strongly fluorescent when illuminated with 300 nm of light. Indeed, this property has been adapted as a method for the identification of recombinant plasmids in E. coli (9).

The siroheme pathway has not been directly studied in other bacteria, although enzymes involved in cobalamin biosynthesis that catalyze SAM-dependent methylation of Uro'gen-III to form PC-2, and further methylations of PC-2, have been iden-

¹ The abbreviations used are: ALA, 5-aminolevulinate; Uro'gen-II, uroporphyrinogen III; PC-2, precorrin-2; SAM, S-adenosyl-l-methionine; PAGE, polyacrylamide gel electrophoresis; IPTG, isopropylthiogalactoside; bp, base pair(s); PCR, polymerase chain reaction.
tified in several bacterial species including *Pseudomonas denitrificans*, *Bacillus megaterium*, *Methanobacterium ivanovii*, and *Propionibacterium freudenreichii* (6, 10–13). The genes encoding these enzymes, including *cobA*, *corA*, *cobI*, and *cobM* show identity with CysG A but not CysG B. The similarity of the enzymes is particularly striking in two regions that constitute consensus amino acid sequences for this class of methyltransferases (14). The enzymes are only able to catalyze SAM-dependent transmethylation and not dehydrogenation or metal chelation reactions, as is observed with CysG. Organisms that make both cobalamin and siroheme complete the synthesis of siroheme from PC-2 in two different ways. In *B. megaterium* a separate oxidase/chelatase exists (12), whereas a dual function enzyme exists in *Salmonella typhimurium* that is closely related to *E. coli* CysG but functions both in siroheme and cobalamin synthesis (15).

Taken together, the significant differences in the siroheme pathway among bacterial species and the use of common intermediates for both siroheme and cobalamin synthesis were of interest to determine how siroheme is made in higher plants. This important prosthetic group is undoubtedly made by plants because it has been detected in association with nitrite and sulfite reductase (3, 16). In contrast, cobalamin has not been detected in plants (17), and plant enzymes are not known to require this coenzyme, although there is a single reported exception (18). The tissue specialization, cellular compartmentation, photosynthetic capacity, and large number of pyrrolic compounds in higher plants provides an added dimension of complexity to a hypothetical siroheme biosynthetic pathway. Nitrite and sulfite reductase are localized predominantly in plastids, as are the porphyrin biosynthetic pathway enzymes (19, 20). The enzymes for porphyrin biosynthesis in plastids are regulated developmentally and by light as required for chlorophyll production (4, 21), yet the development of nitrate and sulfate assimilation capacity is not strictly coordinated with the development of photosynthetic capacity (22, 23). For example, both nitrate and sulfate assimilation are coordinate regulated by the availability of these mineral nutrients (24, 25). In this regard siroheme, constituting the link between these assimilatory pathways, could play a role in cross-pathway regulation (26).

Recently, many genes encoding enzymes involved in sulfate assimilation have been cloned by functional complementation of microorganism mutant strains (27). The availability of stable *E. coli* cysG mutants provided the opportunity to exploit a bacterial genetic system to learn about siroheme biosynthesis in higher plants. Here we report on the cloning and characterization of a higher plant cDNA that is able to functionally complement a specific mutant allele of cysG. The results demonstrate that in the higher plant *Arabidopsis thaliana*, the first step in siroheme biosynthesis may be carried out by a Uro’gen-III, SAM-dependent methyltransferase, more similar to the CobA and CorA methyltransferases found associated with the cobalamin pathway of bacteria than the multifunctional CysG enzyme. The implication of this finding is that the dehydrogenation and ferrochelation steps in siroheme synthesis in plants may be performed by a separate enzyme(s) that has yet to be identified.

**EXPERIMENTAL PROCEDURES**

**General Methods**—All bacteriological media and general nucleic acid protocols such as plasmid mini-preparations, nucleic acid quantitation, buffers, etc., were carried out or prepared as described by Sambrook et al. (28). Denaturing protein gel electrophoresis, referred to in the text as SDS-PAGE, was carried out according to Laemmli (29). Protein was quantitated according to the Bradford (30) procedure.

**Cloning of a cysG Homolog from *A. thaliana***—A plasmid-based cDNA library from *A. thaliana* (Columbia), constructed in *AYES* (31), was used to screen for clones that can complement the Cys requirement of *E. coli*.
Carried out in 50°C, primer A (AGAGTTAAAG-MTAAAACAAGG) and primer B; and primer D (TCACGCTCGCAATTCTCTCGT) and primer B; primer E (GTGTGTGTGTGTTGTTGAATG). Biosystems, Inc., 373 DNA sequencer.

Characterization of the UPM1 Genomic Sequence—The following primer pairs were used to amplify flanking regions of the gene by polymerase chain reaction (PCR): primer A (AGAGTTAAAG-MTAAAACAAGG) and primer B; primer D (TCACGCTCGCAATTCTCTCGT) and primer B; primer E (GTGTGTGTGTGTTGTTGAATG) and primer B; primer D (TCACGCTCGCAATTCTCTCGT) and primer C (TCACGCTCGCAATTCTCTCGT) and primer B; primer E (TCACGCTCGCAATTCTCTCGT) and primer C; and primer D (CCCAAAGAAGACCACTCTCC) and primer C. The reactions were carried out in 50 µl containing 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 2 mM MgCl2, 200 µM of each deoxynucleotide, 0.4 µg of genomic DNA (isolated from axenically grown A. thaliana as described below), and 50 pmol of each primer. The conditions were 1 cycle at 94°C for 2 min, followed by 35 cycles at 94°C for 1 min, 60°C for 1 min, and 72°C for 1.5 min. The PCR products were directly sequenced as described above using the amplification primers and four additional primers; primer F (AATCCCAAGAATCATACATAC), primer G (TTTGGCGAAGAGGTTAAGG), primer H (CATCAACACACACACACACACACACAC), and primer I (GTGGTTGTTGTTGTTGTAAG).

Nucleic Acid Blot Analysis—Total genomic DNA was isolated from axenic root cultures of A. thaliana (Columbia). Approximately 100 seeds were soaked in water for 15 min. They were then surface sterilized by rinsing for 5 min in 70% (v/v) ethanol, followed by 15 min in 2% (w/v) sodium hypochlorite and finally three times each for 5 min in sterilized water. The seeds were germinated and grown for 14 days in liquid M-S medium (34) lacking phytohormones, with gentle shaking, in the dark at 21°C. Genomic DNA was purified by CsCl density gradient centrifugation (35). Approximately 10 µg of DNA were digested with restriction enzymes, and the fragments were separated by electrophoresis through a 0.8% (w/v) agarose gel prepared in TAE (40 mM Tris-acetate, 1 mM EDTA). The DNA fragments were nicked by exposing the gel for 8 min to short wavelength UV light (254 nm) and then denatured by soaking the gel for 8 min in a solution containing 0.4 N HNO3. The DNA fragments were transferred to nylon membrane (Zeta Probe; Bio-Rad, Inc.) using a pressure blotter. The nylon membrane was treated with 50 mM Tris-HCl, pH 7.2, for 10 min and then prehybridized in a solution containing 0.5 M Na2 (supplied in all hybridization and washing buffers as NaH2PO4, pH to 7.2 with H3PO4), 1 mM EDTA, and 7% (w/v) SDS at 65°C for 4 h. Hybridization was carried out in the same buffer, containing the radioactive labeled probe, at 65°C for 16 h. Membranes were washed at maximum stringency in 0.1 M Na2, 1 mM EDTA, and 5% (w/v) SDS at 65°C. The entire UPM1 cDNA was used as a DNA probe and was labeled by the random primer method (36) using [α-32P]dCTP (3000 Ci/mmol).

Uro'gen-III Activity Assays—A rapid binding assay was used to measure interaction of UPM1p with SAM. Purified protein (0.1 mg-1.0 mg) was incubated with 0.5 µCi of [methyl-3H]SAM (50 Ci/mmol) in a final volume of 200 µl of 50 mM Tris-HCl buffer, pH 7.8, and 1 mM 2-mercaptoethanol at 37°C for 15 min. The reaction mixture was applied to a column of Sephadex G25 (1 × 15 cm), and fractions (1 ml) were collected. Radioactivity in the individual fractions was identified with a Pharmacia Wallac 1410 scintillation counter, whereas protein-containing fractions were identified by protein assay.

Expression of UPM1 in E. coli—UPM1 was subcloned into the E. coli expression vector pMAL-c2 (New England BioLabs, Inc.) in two steps. A 394-bp fragment was removed from pUPM1, corresponding to the 5′ end of the cDNA up to the HindIII site, to give pUPM1-H. Then, the remaining DNA fragment from pUPM1-H was subcloned as a 1070-bp SalI-PstI fragment into pMAL-c2 to give pMAL-UPM1. Reombinant plasmids were identified by their ability to complement CBK103 on minimal medium supplemented with 10 µM IPTG, followed by confirmation by restriction enzyme analysis. Colonies carrying pMAL-UPM1 showed bright fluorescence when illuminated with UV light.

Purification of Recombinant UPM1 (UPM1p) and Other Enzymes—Two liters of CBK103 transformed with pMAL-UPM1 were grown at 37°C in LB media to an absorbance (600 nm) of 0.5. IPTG was then added to 0.25 mM, and the cells were grown for an additional 3 h at 37°C. The cells were harvested by centrifugation at 10,000 × g and were resuspended in 25 ml of ice-cold lysis buffer, Tris-HCl, pH 7.5, containing 200 mM NaCl and 1 mM EDTA. The resuspended cells were placed in an ice-water bath and sonicated four times for 1 min each time in an MSE Ultrasonicator at an amplitude of between 12–14 µm with 2 min cooling between sonications. The cell lysate was centrifuged at 30,000 × g, and the supernatant was diluted with an additional 80 ml of lysis buffer and was applied directly to a column of amylose resin (2.5-ml bed volume × 10 cm). The column was washed with 200 ml of lysis buffer before the fusion protein was eluted with buffer containing 10 mM maltose. Fractions (5 ml) were collected from the column and analyzed. SDS-PAGE analysis of the fractions containing the fusion protein were pooled, concentrated in an Amicon ultrafiltration cell fitted with a PM10 membrane, and either stored on ice for immediate use or frozen at −70°C. The fusion protein could be cleaved to release the UPM1 product by cleavage with factor Xa according to the pMAL fusion vector using the manufacturer’s instructions (New England BioLabs). Purification of recombinant E. coli PBG-deaminase, Uro'gen-III synthase, and CysG were carried out as described previously (39, 40).

Uro'gen-III Activity Assays—A rapid binding assay was used to measure interaction of UPM1p with SAM. Purified protein (0.1 mg-1.0 mg) was incubated with 0.5 µCi of [methyl-3H]SAM (50 Ci/mmol) in a final volume of 200 µl of 50 mM Tris-HCl buffer, pH 7.8, and 1 mM 2-mercaptoethanol at 37°C for 15 min. The reaction mixture was applied to a column of Sephadex G25 (1 × 15 cm), and fractions (1 ml) were collected. Radioactivity in the individual fractions was identified with a Pharmacia Wallac 1410 scintillation counter, whereas protein-containing fractions were identified by protein assay.

Labeling of UPM1p with [methyl-3H]SAM was also measured. Protein (20 µg) was mixed with [methyl-3H]SAM (0.01 µCi) for 15 min at 37°C prior to boiling for 2 min in sample buffer. The protein was subsequently loaded onto an SDS-12% (w/v) polyacrylamide gel, and after electrophoresis, the protein was stained with Coomassie Blue and destained in 7% (v/v) acetic acid. The gel was soaked in Autorfluor for 1 h before drying and exposure to x-ray film.

Enzyme assays were performed in a final volume of 10 ml of thoroughly degassed 0.1 mM Tris-HCl, pH 7.8, containing 250 µg of porphobilinogen deaminase, 20 µg of Uro'gen-III synthase, and 0.1 µmol of PBG. After 10 min at 37°C to generate Uro'gen-III (approximate concentration, 2 µM), 1–5 mg of the transmethylase protein were added, and the synthesis of FC-2 was started by the addition of SAM (final concentration, 100 µM). Dehydrogenase activity was monitored after the addition of NAD+ (50 µM). The reaction was followed by recording of the UV-visible spectra (280–650 nm) of the reaction in a Hewlett Packard photodiode array spectrophotometer.

Alternatively, SAM-dependent methyltransferase activity was measured as described by Blanche et al. (41). In this case, enzyme activity was measured by the ability of the enzyme to transfer the methyl group from [methyl-3H]SAM to Uro'gen-III. Enzyme activity is expressed as nanomoles of methyl group transferred per hour.
RESULTS

Cloning and Sequence Analysis of the UPM1 cDNA—As a first step toward characterizing the siroheme biosynthetic pathway of higher plants, a cDNA, UPM1, was isolated from A. thaliana that is capable of complementing the cysG mutation in E. coli strain CBK103. This E. coli strain carries a Tn5 transposon insertion, but the insertion site has not been mapped (32). Because CysG was reported to be a multifunctional enzyme, a second E. coli strain, 302D, carrying a cysG deletion (2), was used to genetically characterize UPM1. The result of a growth assay with CBK103 and 302D is shown in Fig. 2. Although all the strains are able to grow on minimal medium containing Cys, only CBK103 carrying UPM1 is able to grow on minimal medium lacking Cys. Growth of CBK103 on Cys-free medium is strictly dependent on expression of UPM1 as evidenced by growth dependence on IPTG. It should be noted that the colonies expressing UPM1 showed a bright fluorescence when illuminated with UV light, a property associated with accumulation of the trimethylated derivative of Uro'gen-III, trimethylpyrrocorphin, resulting from overexpression of a Uro'gen-III methyltransferase (6, 40). These results indicate that UPM1 is able to complement only one of the functions ascribed to the CysG protein, most likely the CysGA, methyltransferase domain that is localized to the carboxyl end of the enzyme. The results also indicate that CBK103 probably carries a transposon insertion within the 3′ end of cysG disrupting only the CysGα domain.

UPM1 was completely sequenced and found to be 1484 bp in length with an open reading frame predicted to encode a 369-amino acid, Mr 39,900 protein. The nucleotide sequence and deduced amino acid sequence of UPM1 are shown in Fig. 3. It contains a 114-bp 5′ and a 170-bp 3′ untranslated region, including a 35-bp poly(A) sequence, and a polyadenylation signal located 39 bp from the end of the cDNA. The codon bias of UPM1 is similar to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43). The UPM1 nucleotide sequence is 86% identical to that of other protein coding genes from A. thaliana (42). The first Met codon lies within an octanucleotide stretch that fits the consensus sequence around the translation initiation site of plant genes (43).

Fig. 2. Growth assay of E. coli cysG mutants expressing UPM1. E. coli strains CBK103 and 302Δ were transformed with either the parent vector, λYES, or with pUPM1. Following growth on LB medium containing 100 μg/ml Amp, each strain was tested for growth on M-9 minimal medium containing 18 amino acids, excluding Cys and Met (No Cys), or medium with all 20 amino acids (Plus Cys). IPTG at 0.1 mM was added to both media, and amino acids were added at a concentration of 25 μg/ml. To prepare the photographs shown in this figure, the cultures were grown at 30 °C for 48 h.

Fig. 3. Nucleotide sequence and deduced amino acid sequence of UPM1. The nucleotides and amino acids are numbered, the termination codon is indicated with an asterisk, and a possible polyadenylation signal is underlined. The position of introns is indicated by inverted triangles, and the size of the intron in bp is indicated above the triangle. Three nucleotide substitutions at positions 602, 638, and 641 were found in the gene sequence of UPM1. The substitution bases are indicated above the cDNA sequence. The forward slash indicates the HindIII site used for constructing the protein expression vector, pMAL-UPM1. The GenBank accession number for the UPM1 cDNA sequence is L47479, and the genomic sequence is U63734.

Errors in EST sequences.

The deduced translation product of UPM1, UPM1p, showed significant homology with several proteins in the GenBank data base. Fig. 4 shows an amino acid alignment of UPM1p with B. megaterium CobA, E. coli CysG, and a translation product of a yeast gene identified by the European yeast chro- mosome XI sequencing project (Max-Planck-Institut fuer Biochemie). There is a high degree of identity (46%) of UPM1p
with the entire length of CobA, a protein that is solely a Uro'gen-III methylase. UPM1p is homologous with only the carboxyl, methylase domain of CysG (43% identity) referred to as CysG^A. Homology of the putative yeast protein with UPM1p is less striking, with 31% identity. All of the sequences show very high homology in two regions perceived as signature sequences for SAM-dependent Uro'gen-III and PC-2 methyltransferases (14). The first is located between amino acids 120 and 134 of UPM1p; and the second is located between amino acids 195 and 228. The amino-terminal 114 amino acids of the UPM1p is not similar to any part of the SAM-dependent Uro'gen-III and PC-2 methyltransferases. Homology with the amino-terminal sequence of CysG, referred to as the CysG^B domain, responsible for dehydrogenation and ferrochelation of PC-2 is low, with only 19% identity. Indeed, a forced comparison using the Genetics Computer Group, Inc. program BestFit.
and detergent treated.

lane 2, total import reaction; lane 3, protease treated; lane 4, protease and detergent treated. P, precursor protein; M, mature protein.

(46) showed that the amino-terminal 114 amino acids of UPM1p is equally similar to either the CysG\(^{\Lambda}\) (amino acids 202–457) or CysG\(^{B}\) (amino acids 1–201) domains. Of the residues showing identity between UPM1p, CysG, and the yeast protein, none match the proposed dinucleotide binding site of the enzyme and are typically removed through proteolytic cleavage by a specific plastid-localized processing peptidase. The first 114 amino acids of UPM1p contain a number of potential processing sites; therefore, it is not possible to guess exactly its length, except that it is common for chloroplast transit peptides to be 50 to 100 amino acids (47).

An examination of the amino acid composition of the UPM1p amino terminus suggests that it may function as a signal sequence for localization to chloroplasts (47). The features of plastid transit peptides include a high concentration of hydroxyl amino acids, hydrophobic amino acids with small side chains, and a preponderance of basic over acidic amino acids, resulting in a net positive charge. Another feature of transit peptides is that they are not necessary for catalytic function of the enzyme and are typically removed through proteolytic cleavage by a specific plastid-localized processing peptidase. The first 114 amino acids of UPM1p contain a number of potential processing sites (48); therefore, it is not possible to guess exactly its length, except that it is common for chloroplast transit peptides to be 50 to 100 amino acids (47).

**Analysis of UPM1 Protein Localization**—To investigate the potential localization of the UPM1p further, in vitro import assays were performed. Transcription and translation in vitro of the precursor encoded by UPM1 gave a major protein band of approximately 41 kDa on SDS-PAGE (Fig. 5, lane 1). A band that is several kDa lower in mass is also visible. Incubation of the translation products with isolated intact chloroplasts resulted in the loss of the precursor band, but the lower mass band remained (lane 2), indicating that this is the mature protein that has been processed. The presence of the mature protein in the translation products is a common observation and is probably due to chloroplast-processing peptidase in the wheat germ system. The mature polypeptide is retained after protease treatment of the chloroplasts (lane 3) but is completely digested if the chloroplasts were treated with protease in the presence of the detergent Triton X-100 (lane 4). This indicates that the mature protein is protected by a membrane and confirms that the UPM1 protein has been imported into chloroplasts. A mitochondrial import assay was also performed, but no import was observed (data not shown).

**Analysis of the UPM1 Genomic Sequence**—Experiments were conducted to confirm that the UPM1 cDNA is derived from an A. thaliana gene. A genomic DNA blot was carried out to examine the number of homologous sequences in A. thaliana. Blotting of genomic DNA isolated from axenically grown A. thaliana plants showed that there are at least three strongly hybridizing DNA fragments and several weakly hybridizing fragments (Fig. 6). This indicates that there may be three or more genes with some identity to UPM1, or that there are a smaller number of genes containing introns that are cleaved with the enzymes tested in this experiment. Cross-hybridization to unrelated genes is also a possibility. The larger number of hybridizing bands in lane 1 can be partially explained by the existence of at least two HindIII sites within the UPM1 gene.

To further characterize the UPM1 gene, four overlapping segments were amplified by PCR, and the amplification products were directly sequenced. The results of this analysis are shown in Fig. 7. It should be noted that although the genomic DNA used as the PCR template was obtained from axenically grown plants, under low annealing temperatures a number of amplification products were observed with each primer set used. Thus, it was necessary to carry out the reactions at a high annealing temperature (60 °C) to obtain a single amplification product (data not shown). Comparison of the UPM1 genomic and cDNA sequences revealed that this gene contains at least five introns, indicated by the open boxes in Fig. 7. The exact positions and sizes of the introns are shown in Fig. 3. The genomic coding sequence was found to be nearly identical to the UPM1 cDNA with the exception of three nucleotide substitutions indicated in Fig. 3. The sequencing results obtained from the cDNA and genomic clones in this region were not ambiguous, indicating that the nucleotide substitutions are likely authentic. It is interesting to note that the changes are conservative with respect to the amino acid sequence. Four of the introns, ranging in size from 85 to 127 bp, are positioned within, and interrupt the coding sequence. A large intron of 430 bp is positioned 5′ of the initiation codon in a transcribed but untranslated region of the gene. Analysis of all open reading frames within this large 5′ intron did not reveal any homology with the amino-terminal domain of CysG or to the amino-terminal domain of the putative yeast methyltransferase.

**Functional Analysis of UPM1 Protein**—A portion of the UPM1 cDNA was cloned in-frame into the pMAL fusion protein expression system. The net effect of the construction was to remove 93 amino acids from the amino terminus of UPM1p, corresponding to the entire postulated plastid transit peptide, and the addition of 13 amino acids encoded by plasmid polylinker sequence, up to the junction with the coding sequence for the maltose binding protein. The mass of the fusion protein is predicted to be 71 kDa, and the mass of the truncated UPM1p 31 kDa. Recombinant plasmids were identified by their ability to exhibit a bright fluorescence when illuminated under UV light, presumably resulting from accumulation of trimethylpyrocorphin, as did cells expressing full-length UPM1p. This compound has a characteristic orange/red fluorescence when
viewed under UV light (6, 40). The spectrum of the fluorescent material in cell extracts showed an absorption maximum at 354 nm (Fig. 8), identical to that reported previously for trimethylpyrrocorphin (40). The plasmid pMAL-UPM1 was also able to complement the Cys auxotrophy of CBK103. Taken together with the accumulation of a highly fluorescent intracellular product, this would suggest that the fusion protein is catalytically active. Furthermore, the analysis of whole-cell extracts by SDS-PAGE revealed the expression of a protein with a molecular mass of ~67 kDa. The recombinant protein was purified from cell extracts in a one-step procedure using an amylose affinity column. The purified protein is shown in Fig. 9, lane 2. The purified protein could be cleaved to release UPM1p from the fusion protein (lane 3), and UPM1p could subsequently be purified to homogeneity by passage through the amylose resin again (lane 4). However, because the fusion protein was catalytically active, the majority of the functional studies reported below were performed using the fusion protein.

The enzyme was assayed for its ability to transfer the methyl group of \([\text{methyl-}^{3}H]\text{SAM}\) to Uro'gen-III. The methylated Uro'gen-III was separated from the incubation mixture by binding to DEAE-cellulose, and the radioactivity associated with it was determined. The recombinant enzyme was found to have a specific activity of 550 nmol of methyl group h\(^{-1}\) mg\(^{-1}\) protein at 30 °C. The specific activity is in the same order as that reported for other Uro'gen methylases (11, 12, 41). Heat-denatured protein had negligible activity. This assay demonstrates that UPM1p is a functional Uro'gen III methyltransferase.

The activity of the enzyme was also monitored by UV-visible spectroscopy. Incubations containing Uro'gen III and SAM turned yellow in color and produced a UV-visible spectrum of a dipyrrorophyrin, as would be expected for a solution of PC-2 (Fig. 8). There was no apparent transformation of this solution into the orange/red color of trimethylpyrrocorphin as has been observed with \(E.\ coli\) CysG and CobA from \(P.\ freudenreichii\) (6). However, as we have stated earlier, the trimethylpyrrocorphin does accumulate in induced cells \(in\) \(vitro\) (Fig. 8). Trimethylpyrrocorphin may not be made \(in\) \(vitro\) because of the lower enzyme concentrations in the assay, or because the enzyme may be inhibited by the reaction product S-adenosyl-l-homocysteine, as has been observed for other Uro'gen-III methylases (12, 41). The addition of NAD\(^{+}\) or NADP\(^{+}\) to the incubation mixture did not lead to the oxidation of PC-2 to sirohydrochlorin (Fig. 8). Under the same conditions using NADP\(^{+}\), CysG is able to form sirohydrochlorin, as evidenced the production of material with absorption maxima at 376 and 600 nm.

Rapid SAM binding columns have been used in the past to identify the SAM-dependent transmethylases involved in cobalamin biosynthesis (13). To ascertain whether UPM1p shows a similar tight binding of SAM, the recombinant enzyme was incubated with stoichiometric quantities of \([\text{methyl-}^{3}H]\text{SAM}\), and the mixture was separated by gel filtration chromatography. The results, shown in Fig. 10, indicate that in the absence of protein, or if the \([\text{methyl-}^{3}H]\text{SAM}\) is mixed with bovine serum albumin, the label is found to elute in the small molecules fractions. If the \([\text{methyl-}^{3}H]\text{SAM}\) is mixed with UPM1p, the label is found to coelute with the protein-containing fractions.

Next we determined if UPM1p binds \([\text{methyl-}^{3}H]\text{SAM}\) covalently. The UPM1-SAM mixture was subjected to SDS-PAGE/fluorography and the results, shown in Fig. 11A, indi-
CysG has shown that the protein is able to bind both \textsuperscript{14}C\textsuperscript{-3}H\textsuperscript{3}SAM as well as \textsuperscript{methyl}\textsuperscript{3}H\textsuperscript{3}SAM, suggesting that the whole SAM molecule is bound to the enzyme (49). The exact nature of the interaction between SAM and the protein has still to be elucidated.

**DISCUSSION**

Here we describe the isolation of a cDNA clone from *A. thaliana* that encodes a SAM-dependent Uro'gen-III methyltransferase by functional complementation of a specific mutant allele of *cysG* in *E. coli* strain CBK103. Significantly, UPM1 cannot complement strain 302Δ carrying a complete deletion of *CysG*; and cells expressing UPM1p accumulate a fluorescent compound trimethylpyrrocorphin. One explanation for these results is that UPM1p is solely a methylase, lacking the ferrochelatase and dehydrogenase activities associated with the multifunctional CysG enzyme (2). (a) CBK103 carries a Tn5 transposon that disrupts only the methylase domain of CysG, maintaining intact the dehydrogenase and ferrochelatase functions of the enzyme; (b) trimethylpyrrocorphin is an overmethylated derivative of Uro'gen-III that is known to accumulate in cells that overexpress a truncated, recombinant version of CysG with only methylase activity (2); and (c) the UPM1p sequence shows high homology to the methylase domain of CysG, although homology to the ferrochelatase/dehydrogenase domain is very low. Consistent with this hypothesis, purified recombinant UPM1p was found to be active in methylation of Uro'gen-III to PC-2. The lack of ferrochelatase and dehydrogenase activity of this recombinant UPM1p is inconclusive, however, owing to the truncation of a portion from the amino terminus of the enzyme.

The presence of single-function Uro'gen-III methylases are common in bacterial species that synthesize cobalamin (6, 10–13). UPM1p is unlikely to be involved in coahalamin synthesis because higher plants are thought not to synthesize this vita-

![Fig. 10. Rapid SAM binding by UPM1p. The graphs show the elution profile of radioactivity from a rapid SAM binding assay performed on a Sephadex G-25 column. Bovine serum albumin was used as a control. Solid line, UPM1; dotted line, bovine serum albumin.](Image)

![Fig. 11. Fluorogram of CysG and UPM1p after incubation with [methyl-\textsuperscript{3}H]SAM. A: lane 1, UPM1p after labeling with [methyl-\textsuperscript{3}H]SAM; lane 2, CysG after labeling. B: lane 1, UPM1p after SAM labeling; lane 2, SAM-labeled UPM1p after addition of Uro'gen III for 15 min at 30° C prior to electrophoresis.](Image)

min. A more likely possibility is that siroheme biosynthesis in plants is carried out by separate enzymes, as is the case in *B. megaterium* (12). Thus, in *A. thaliana* UPM1p may carry out the first step in siroheme synthesis and at least one other enzyme, yet to be identified, catalyzes the oxidation of PC-2 into sirohydrochlorin and subsequently performs ferrochelation to yield siroheme. With the possibility that the steps in the siroheme pathway are catalyzed by separate enzymes, it will be possible in the future to design specific *E. coli* mutant strains that could be used to define all of the higher plant enzymes in this pathway.

UPM1p contains a 114-amino acid, amino-terminal region that resembles a transit peptide for transport into chloroplasts, and the protein is able to be transported into chloroplasts. This result is significant because other tetrapyrrole biosynthetic enzymes are known to be localized in plastids (19, 20). In addition, siroheme biosynthetic enzymes would be expected in plastids because both sulfite and nitrite reductase are thought to reside only in this organelle (24, 27).

UPM1p binds SAM in a manner similar to CysG, indicating that both belong to a class of enzymes that catalyze methyltransfer via a stable, covalently linked enzyme-SAM intermediate complex. Although the role of this mechanism for methylgroup transfer still remains to be answered, the identification of a eukaryotic enzyme with this property indicates that this mechanism may be a hallmark of all Uro'gen III methyltransferases.

**REFERENCES**

Siroheme Biosynthesis in Higher Plants: ANALYSIS OF AN S-ADENOSYL-L-METHIONINE-DEPENDENT UROPORPHYRINOGEN III METHYLTRANSFERASE FROM ARABIDOPSIS THALIANA
Thomas Leustek, Michael Smith, Michael Murillo, Davinder Pal Singh, Alison G. Smith, Sarah C. Woodcock, Sarah J. Awan and Martin J. Warren

doi: 10.1074/jbc.272.5.2744

Access the most updated version of this article at http://www.jbc.org/content/272/5/2744

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

This article cites 44 references, 23 of which can be accessed free at http://www.jbc.org/content/272/5/2744.full.html#ref-list-1