Identification and characterization of the Arabidopsis gene encoding the tetrapyrole biosynthesis enzyme uroporphyrinogen III synthase

Fu-Ching TAN*, Qi CHENG*, Kaushik SAHA*, Ilka U. HEINEMANN†, Martina JAHN†, Dieter JAHN† and Alison G. SMITH*†

*Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, U.K., and †Institute of Microbiology, Technical University Braunschweig, Spielmannstr. 7, 38106 Braunschweig, Germany

UROS (uroporphyrinogen III synthase; EC 4.2.1.75) is the enzyme responsible for the formation of uroporphyrinogen III, the precursor of all cellular tetrapyrroles including haem, chlorophyll and bilins. Although UROS genes have been cloned from many organisms, the level of sequence conservation between them is low, making sequence similarity searches difficult. As an alternative approach to identify the UROS gene from plants, we used functional complementation, since this does not require conservation of primary sequence. A mutant of Saccharomyces cerevisiae was constructed in which the HEM4 gene encoding UROS was deleted. This mutant was transformed with an Arabidopsis thaliana cDNA library in a yeast expression vector and two colonies were obtained that could grow in the absence of haem. The rescuing plasmids encoded an ORF (open reading frame) of 321 amino acids which, when subcloned into an Escherichia coli expression vector, was able to complement an E. coli hemD mutant defective in UROS. Final proof that the ORF encoded UROS came from the fact that the recombinant protein expressed with an N-terminal histidine-tag was found to have UROS activity. Comparison of the sequence of AtUROS (A. thaliana UROS) with the human enzyme found that the seven invariant residues previously identified were conserved, including three shown to be important for enzyme activity. Furthermore, a structure-based homology search of the protein database with AtUROS identified the human crystal structure. AtUROS has an N-terminal extension compared with orthologues from other organisms, suggesting that this might act as a targeting sequence. The precursor protein of 34 kDa translated in vitro was imported into isolated chloroplasts and processed to the mature size of 29 kDa. Confocal microscopy of plant cells transiently expressing a fusion protein of AtUROS with GFP (green fluorescent protein) confirmed that AtUROS was targeted exclusively to chloroplasts in vivo.

Key words: chloroplast import in vitro, deletion mutant, functional complementation, green fluorescent protein (GFP), plastid location.

INTRODUCTION

Tetrapyrroles such as chlorophyll, haem, sirohaem and bilins are essential cofactors for many fundamental biological processes, including photosynthesis, oxygen transport and electron transfer. In all organisms, tetrapyrroles are derived from a common macrocyclic precursor, uroporphyrinogen III. This is methylated as the first step in the pathway to sirohaem and corrins such as uroporphyrinogen I, the last common intermediate of haem and chlorophyll synthesis.[1].

Uroporphyrinogen III is made in three enzymatic steps from a five-carbon compound, ALA (5-aminolaevulinic acid). Two molecules of ALA are condensed to form the pyrrole PBG (porphobilinogen) by a metalloenzyme, PBG synthase (EC 4.2.1.24). The following enzyme, PBG deaminase (EC 4.3.1.8), then mediates a stepwise linkage of four molecules of PBG to yield a linear tetrapyrrole. HMB (1-hydroxymethylbilane) or preuroporphyrinogen III. Finally, UROS (uroporphyrinogen III synthase; EC 4.2.1.75) catalyses the cyclization of HMB with a concomitant inversion of the fourth ring of the porphin macrocycle, giving rise to uroporphyrinogen III [2]. In the absence of UROS, HMB cyclizes non-enzymatically to form uroporphyrinogen I without any rearrangement of the fourth pyrrole ring. This is not a precursor to biological tetrapyrroles, and cannot be metabolized past the next step in the pathway. Congenital erythropoietic porphyria is a human disease caused by a deficiency in UROS. This results in the accumulation of the oxidized derivatives, uroporphyrin I and coproporphyrin I, in plasma, tissues and red blood cells, leading to severe photosensitivity with skin fragility, hypertrichosis and lesions on light-exposed areas [3,4].

The first gene encoding UROS was isolated from Escherichia coli [5], with those from human [6], Bacillus subtilis [7], Pseudomonas aeruginosa [8], Anacystis nidulans R2 (now reclassified as Synechococcus PCC 7942) [9], mouse [10] and budding yeast Saccharomyces cerevisiae [11] being isolated over the next few years. A comparison between UROS sequences found that there are seven invariant residues and a further 15 positions have conservative substitutions. The crystal structure of the human enzyme revealed that the enzyme has two α/β domains linked by a β-ladder [12]. The active site is between the two domains, and is lined by ten of the invariant or conserved residues that are surface-exposed. However, the overall sequence similarity between UROS enzymes from different organisms is low; for example the E. coli and human sequences have less than 20% identity. This is in contrast with other tetrapyrrole enzymes, such as PBG deaminase and coproporphyrinogen oxidase that are 55–60% identical. Primary sequence conservation is a necessary prerequisite to identify putative orthologues by sequence database mining.

Abbreviations used: ALA, 5-aminolaevulinic acid; AtUROS, Arabidopsis thaliana UROS; CAT, catalase; EST, expressed sequence tag; GFP, green fluorescent protein; HMB, 1-hydroxymethylbilane; LB, Luria–Bertani; IPTG, isopropyl β-D-thiogalactoside; Ni-NTA, Ni²⁺-nitrilotriacetate; ORF, open reading frame; PBG, porphobilinogen; UROS, uroporphyrinogen III synthase; YNB D medium, 0.67% (w/v) bacto-peptone and 2% (w/v) glucose; YPD medium, 1% (w/v) yeast extract, 2% (w/v) bactopeptone and 2% (w/v) glucose; YPG medium, 1% (w/v) yeast extract, 2% (w/v) bactopeptone, 3% (v/v) glycerol.

1 To whom correspondence should be addressed (email as25@cam.ac.uk).
An alternative approach is to use functional complementation, which requires conservation of function only, not of nucleotide or amino acid sequence, so it can be used to identify genes from heterologous sources. Complementation of bacterial and yeast mutants has been used to great effect to identify plant cDNAs for a range of different proteins, including cell-cycle components, membrane transporters and transcription factors, as well as metabolic enzymes [13]. Indeed, we used the E. coli hemB mutant deficient in UROS, to identify the corresponding gene from the cyanobacterium Aspergillus nidulans [9]. In the present study, we describe the use of a mutant of S. cerevisiae, in which the HEM4 gene encoding UROS was deleted, for the isolation of an Arabidopsis cDNA for UROS. This provided the means to establish the subcellular location of the enzyme.

EXPERIMENTAL

Materials

Bacto-yeast nitrogen base without amino acids, bactotryptone, bactopeptone and bacto-agar came from Difco Laboratories and yeast extract was obtained from Oxoid. Deoxynucleoside triphosphates were purchased from Amersham Pharmacia Biotech, Expand™ High Fidelity PCR Enzyme Mix was from Roche, BIOTAQ™ DNA polymerase was from Bioline, custom synthetic oligonucleotides and G418 (geneticin) were from Gibco, and restriction enzymes were purchased from Roche and New England BioLabs. Thermolysin and haematin were purchased from Sigma. Uroporphyrin III was from Porphyrin Products. PRO-Mix™ [L-[35S]methionine/cysteine (> 1000 Ci/mmol) was from Amersham Pharmacia. The Ribobrobe® System, T7 RNA polymerase, rabbit reticulocyte lysate as well as RNasin were supplied by Promega.

Yeast strains and growth conditions

S. cerevisiae strain S150-2B (MATα ura3-52 trp1-289 leu2-3 leu2-112 his3Δ1) was grown at 30°C in either rich glucose medium [YPD; 1% (w/v) yeast extract, 2% (w/v) bactopeptone and 2% (w/v) glucose], or minimal glucose medium [YNB D; 0.67% (w/v) bacto-yeast nitrogen without amino acids and 2% (w/v) glucose] supplemented with 20 µg/ml L-histidine, 20 µg/ml L-tryptophan, 20 µg/ml uracil and 30 µg/ml L-leucine. Strain S150-2B∆HEM4 (MATα ura3-52 trp1-289 leu2-3 leu2-112 his3Δ1 Δhem4::kan′), constructed as described below, was maintained in YPD supplemented with 15 µg/ml haemin and 200 µg/ml geneticin. Transformants of S150-2B∆HEM4 harbouring a pFL61 plasmid [14] were cultured in minimal glucose medium but without uracil, and with 200 µg/ml geneticin. For phenotypic analysis, the yeast strains were cultured in the appropriate liquid medium for 1–2 days at 30°C before the cultures were adjusted to the same D600 and serially diluted with sterile distilled water (1:10 dilution). The diluted cells were spotted in 5-µl droplets on to YPD, YPD + 15 µg/ml haemin and YPG [1% (w/v) yeast extract, 2% (w/v) bactopeptone, 3% (v/v) glycerol] agar medium, and incubated for 3–5 days at 30°C.

Generation of the yeast S150-2B∆HEM4 mutant

Deletion of the yeast HEM4 gene was conducted according to the short-flanking-homology PCR strategy described by Wach et al. [15]. A disruption cassette, comprising a chimaeric gene fusion of the E. coli transposon Tn905 (kan′ gene; [16]) coding sequence and the promoter, as well as the terminator of the Ashbya gossypii translation elongation factor 1α [17], flanked at both ends by 45 bp nucleotide sequences homologous to the HEM4 ORF (open reading frame), was generated by PCR. The pFA6-A-KANMX4 plasmid [15] was amplified in the presence of 1 unit of BIOTAQ™ DNA polymerase, 0.5 mM dNTPs, 2.5 mM MgCl2, 0.25 µM forward primer (ScHEM4-KAN.rev: 5′-AGGATAAGGAACAGAAAGTAAATAGACCTTGCTCG-3′) and 0.25 µM reverse primer (ScHEM4-KAN.rev: 5′-AAGTAAATAATTATATAATAGAGAAGAATATGCTCAATATATATCAGTGAATTTCGAGC-3′); the underlined regions of both primers correspond to the sequence of the target ORF. The reaction was carried out at 30 cycles of 94°C for 30 s, 50°C for 1 min and 72°C for 2 min. The PCR product was gel-purified, and then used to transform the S150-2B strain using the method of Gietz and Woods [18]. Transformed cells were selected for incorporation of the kan′ gene on a YPD agar medium supplemented with 15 µg/ml haemin and 200 µg/ml geneticin, incubated at 30°C for 5 days. Normal-sized colonies were restreaked on to geneticin-supplemented YPD medium plus or minus haemin. Out of 16 randomly chosen transformants, one (termed S150-2B∆HEM4) was identified as a bona fide deletion mutant based on its inability to grow normally on YPD in the absence of haemin.

Confirmation of the deletion of the HEM4 gene in S150-2B∆HEM4 by PCR

The correct replacement of the HEM4 ORF in S150-2B∆HEM4 by the kan′ disruption cassette was confirmed by PCR. Genomic DNA was extracted from S150-2B and S150-2B∆HEM4 as reported by Rose et al. [19] except that the concentration of lyticase was 0.9 mg/ml. The isolated DNA was used as a template using different combinations of primers (ScUROS.for, 5′-ATAAGATCCGCTGTAAGTCATACGTAAGGCGC-3′; ScUROS.rev, 5′-TATGAATTCCATCGCATTCTTTATCATGGC-3′; and KANMX4.iprev, 5′-ACTGAATCCGCTGTAAGGCGC-3′) as described below under the following conditions: 1 cycle of 95°C for 3 min, 30 cycles of 95°C for 30 s, 52°C for 30 s and 72°C for 90 s, followed by 1 cycle of 72°C for 5 min.

Functional complementation of the yeast S150-2B∆HEM4 mutant

The S150-2B∆HEM4 strain was transformed with an Arabidopsis cDNA library constructed in a yeast expression vector, pFL61 [14], according to the protocol described by Gietz and Woods [18] with some minor modifications. The transformation was scaled up to 20 times of a standard reaction, using approx. 9 µg of plasmid library DNA. Haem prototrophs were directly selected on YNB D agar medium supplemented with 200 µg/ml geneticin, 20 µg/ml L-histidine, 20 µg/ml L-tryptophan and 30 µg/ml L-leucine.

Cloning and characterization of U2 and U6 cDNAs

The U2 and U6 inserts were digested with NotI from pU2.FL61 and pU6.FL61 respectively, and then subcloned into pBluescript II KS+ to form pU2.KS and pU6.KS, before being sequenced on both strands with T7 and T3 primers, and the following specific primers: AtUROS.ipF1 5′-CTTCTTCTCCATCCAAATTCCG-3′, AtUROS.ipF2 5′-CTTCTTGCATTCCACCAGGC-3′, AtUR-OS.ipF3 5′-GTAAGATATCCTAGATAGC-3′, AtUROS.ipR1 5′-GATACTTCTACAGGCTGTC-3′, CAT.ipF1 5′-ATCCAAAGGTAATGGGAC-3′, and CAT.ipF2 5′-CTTCAGGTCAATGTCTC-3′. Sequencing was carried out by DNA sequencing facilities in the Department of Biochemistry, University of Cambridge, U.K. DNA and protein sequences were analysed using software packages of the GCG (Genetics Computer Group), University...
of Wisconsin, Madison, WI, U.S.A. Comparison of multiple se-
quences was conducted using the ClustalW version 1.81 program
[20]. Growth and functional complementation of E. coli strain
SASZ31 (hemD<sup>−</sup>) [21], were as described in Jones et al. [9].

**Overexpression of AtUROS E. coli**

For overexpression studies, the insert from pU6.KS was subcloned
into vector pET24a (Novagen) such that the cDNA was
under the control of the T7 promoter. The resulting plasmid
pU6.ET24a was introduced into E. coli BL21(DE3) cells and
the protein was induced by addition of IPTG (isopropyl-
β-D-thiogalactoside) overnight. Total cell protein was released
from a cell pellet by sonication, and analysed by SDS/PAGE
followed by staining with Coomassie Blue. Because this did not
yield soluble protein, another construct was generated in which
the first 81 amino acids had been removed. PCR was carried
out using the following primers AtUROS<sup>′</sup>F, 5′-GAAcatatG-GCTTTGGAGAAAAATGGC-3′ and AtUROS<sup>′</sup>R, 5′-CTTgaattc-
TCAATCTCTGCTGTCAGG-3′ (lower case letters indicate the
NdeI and EcoR1 restriction sites), followed by cloning the frag-
ment into pET28b (Novagen) between NdeI and EcoR1 to
form pAtUROS<sup>′</sup>ET28b. This allowed synthesis of a chimaeric
protein with an N-terminal His<sub>6</sub>-tag.

**Recombinant production and purification of AtUROS**

Recombinant AtUROS was produced in E. coli BL21(DE3)
RIL (Stratagene) containing pAtUROS<sup>′</sup>ET28b. Cells were
grown in LB (Luria–Bertani) medium at 37°C under vigorous
aeration. When the cultures reached an OD<sub>750</sub> of 0.7, protein
production was induced by the addition of 100 µM IPTG.
Further cultivation followed overnight at 25°C and 150 rev./min.
Cells were harvested, washed with buffer A [20 mM Hepes
(pH 7.5), 5 mM MgCl<sub>2</sub>, 0.01% (v/v) Triton X-100] and res-
pended in a minimal volume of buffer A. Bacteria were disrupted
via sonication (Bandelin HD 2070, 0.5 s sound, 0.5 s paused,
MS73 tip, 70% amplitude) and the cell-free extract was cleared by
centrifugation at 150 000 g for 45 min. Protein integrity was
verified via Western blot analysis. Recombinant UROS was puri-
fied by Ni-NTA (Ni<sup>2+</sup>-nitrilotriacetate) affinity chromatog-
raphy, His<sub>6</sub>-tagged AtUROS<sup>′</sup> was eluted with 300 mM im-
dazol in buffer A. Fractions that contained recombinant UROS
were identified by SDS/PAGE and UROS activity (see below);
the two correlated closely. The fractions were combined, and applied
to a DEAE-Sepharose column at a concentration of 0.5 mg/ml
column volume, followed by elution with 200 mM NaCl in buffer
A. Fractions containing recombinant AtUROS<sup>′</sup> were combined,
concentrated [Centricon-10 filter, MWCO (molecular-mass cut-
off) 10 kDa; Amicon] and purified to apparent homogeneity by
gel-permeation chromatography using a 30 ml Superdex
200 HR 10/30 column (General Electric Company), at a flow rate of 0.5 ml/min in buffer A. For the purposes of
calibration, bovine carbonic anhydrase (Mr = 29 000), BSA (Mr =
66 000), yeast alcohol dehydrogenase (Mr = 150 000) and amylase
(Mr = 200 000) were used as marker proteins and chromatog-
graphed under identical conditions.

**Determination of UROS enzymatic activity**

UROS activity was determined using a coupled enzyme
assay. Recombinant P. aeruginosa PBG synthase and Bacillus
megaterium PBG deaminase were purified as described previously
[22]. The standard assay mixture contained 25 µg of purified
recombinant AtUROS<sup>′</sup>, 0.2 mM ALA, 10 µg of PBG synthase
and 10 µg of PBG deaminase in a total volume of 800 µl in
buffer A. The reaction mixture was incubated for up to 120 min
at 37°C in the dark. The reaction was stopped by addition of
300 µl KI/I<sub>2</sub> [0.5% (w/v) and 1% (w/v) in H<sub>2</sub>O] to oxidize
any uroporphyrinogen converted into uroporphyrin. Na<sub>2</sub>S<sub>2</sub>O<sub>5</sub>
solution [1% (w/v) in water] was added to oxidize residual I<sub>2</sub>.
Proteins were precipitated by the addition of 100 µl of 50% TCA
(trichloroacetic acid) and harvested by subsequent centrifugation
(10 000 g, 5 min, 4°C). The amount of uroporphyrin produced
was determined by both by absorbance at 408 nm [23] and by
fluorimetric detection using a PE LS50B luminescence spectro-
meter (PerkinElmer Instruments) with an excitation wavelength
of 400 nm, an emission scan range of 500–700 nm, a scan speed
of 200 nm/min and slit widths of 5 nm for emission and excitation.
To test whether this was enzymatically formed uroporphyrin III
isomer, rather than uroporphyrin I, which can arise by chemical
cyclization of HMB, and which has identical absorbance and
fluorescence properties, control experiments were performed,
using either no AtUROS<sup>′</sup> or heat-inactivated enzyme [24].
Uroporphyrin was not detected in either case. Thus all of
the uroporphyrin formed came from the activity of the recombi-
nant enzyme.

**Import of radiolabelled AtUROS precursor into pea chloroplasts**

Peas (Pisum sativum L. cv Feltham First) were grown at 25°C in
a greenhouse with a 16 h day photoperiod. The shoots of 7–8-day-
old peas were harvested for chloroplast isolation. Radiolabelling
of the full-length AtUROS precursor protein was prepared by
transcription in vitro of plasmid pU6.ET24a, followed by trans-
lation in vitro in the rabbit reticulocyte system in the presence
of [147M]methionine/cysteine. Chloroplast isolation and import
of radiolabelled precursor protein were carried out essentially as
described by Cleary et al. [25]. After import, protease-treated
chloroplasts were fractionated into stroma, thylakoid and envelope
fractions, and analysed by SDS/PAGE followed by fluorography
as described previously [26].

**Targeting of AtUROS–GFP (green fluorescent protein) fusion
protein in tobacco leaves in vivo**

The full-length protein coding sequence of AtUROS was ampli-
fied from pU6.ET24a by PCR with BamHI sites at either end
for in-frame fusion to the 5′-end of the coding sequence for GFP
in psmRSFGFP [27] to form pAtUROS–GFP. Leaves from 3-week-
old tobacco (Nicotiana tabacum cv. Xanthi) were excised and used
for biostatic transformation with pAtUROS–GFP, psmRSFGFP
and recA–GFP [28], followed by confocal microscopy as
described previously [25].

**RESULTS AND DISCUSSION**

**Complementation of a yeast UROS mutant with Arabidopsis cDNAs**

The short-homology-flanking PCR technique [15] was used to
custom grow S150-2BΔHEM4, in which the endogenous HEM4
gene encoding UROS was replaced with a kanamycin-
resistance gene via homologous recombination. This replacement
was confirmed by PCR using specific primers (Figure 1). Strain
S150-2BΔHEM4 could grow on YPD, as long as it was supple-
mented with haemin (Figure 2A). However, the mutant was unable
to grow on non-fermentable carbon sources such as glycerol, since
it lacked respiratory cytochromes. Interestingly, it was also unable
to grow on minimal glucose medium even in the presence of all
of the required nutrients plus haemin (results not shown), thus
providing a distinctive phenotype for selection of functionally
complemented cells.
The genomic DNA of S150-2B (WT) and S150-2BΔHEM4 (h4) was individually amplified via PCR using three specific primers as follows: F1 (ScUROS.for; a forward primer homologous with a region upstream of the recombination site), R1 (ScUROS.rev; a reverse primer homologous with the coding sequence of HEM4), and R2 (KANMX4.iprev; a reverse primer homologous with the coding sequence of kanr). Samples without DNA were used as negative controls. The region amplified from the corresponding templates was indicated with arrows. The solid black bars represent the upstream and the downstream regions of the HEM4 gene. The dark grey bars signify the site of recombination, whereas the light grey and the open bars indicate the HEM4 ORF and the kanr gene respectively.

The S150-2BΔHEM4 mutant was transformed with an Arabidopsis cDNA library constructed in a yeast expression vector pFL61 [14]. Two independent transformants were obtained based on their ability to grow on a minimal glucose medium in the absence of uracil and haemin. Both could grow on rich glucose medium (YPD) in the absence of exogenous haemin (Figure 2B), and could utilize non-fermentable carbon sources such as glycerol, indicative of the restoration of normal respiratory function in the mitochondria of both clones (Figure 2C). To confirm that the complementation was due to the presence of the Arabidopsis cDNAs, rather than reversion, plasmids were isolated from the complemented strains, and transformed back into the S150-2BΔHEM4 mutant. As expected, both plasmids complemented the respiratory defect of the mutant (results not shown).

Characterization of the complementing Arabidopsis cDNAs

The two complementing plasmids named pU2.FL61 and pU6.FL61 were digested with NotI to excise the cDNAs from the vector, and found to be 3.0 and 1.4 kbp respectively (Figure 3A). The inserts were subcloned into pBluescript II KS, to generate pU2.KS and pU6.KS, and sequenced, whereupon the reason for the difference in size between the two clones was established (Figure 3B). Both shared an identical ORF of 321 amino acids, but U2 encoded a second ORF of 492 amino acids on the complementary strand downstream of the first ORF. The nucleotide sequence of the ORF common to both plasmids was used to query the Arabidopsis sequence database with BLAST [29] on the NCBI server (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi), and an Arabidopsis genomic BAC clone, T9J22 (GenBank® accession number AC002505), was identified that was identical with the cDNA, except at two positions, where single nucleotide polymorphisms occurred. This is probably due to microstructural differences between ecotype Landsberg erecta, the source of the cDNA library, and Columbia, the ecotype from which the genome was sequenced [30]. The BAC clone mapped to chromosome II, and no other region of the genome was identified that had sequence similarity to the cDNA. Comparison of the cDNA with the genomic sequence revealed that the gene comprised nine exons separated by eight introns (Figure 3C), all with consensus splice sites. In the original annotation of the Arabidopsis genome [30], part of the BAC sequence matching the U6 cDNA was incorrectly predicted to encode a 145-amino-acid hypothetical protein of unknown function (AGI reference At2g26540), starting from the middle of the fourth exon to the end of the gene, but omitting the sixth exon. The inaccuracy of initial annotation, particularly for genes without ESTs (expressed sequence tags) is common, and it is estimated that only approx. 20% of the originally annotated genes are structurally correct.

The authors acknowledge the Biochemical Society for the support of this work.
The protein sequences of the second ORF of U2, when searched against the Arabidopsis genome, matched imperfectly an annotated putative CAT1 (catalase-1) protein from the BAC F5M15 clone (GenBank\textsuperscript{®} accession number AC027665). The putative CAT1 gene was predicted to encode a 1013-amino-acid protein, almost twice the size of the second ORF of U2 (Figure 3D). The extra sequences were predominantly found at the C-terminus of the protein. At first glance, the apparent differences could be due to the isolated cDNA being a truncated clone. However, this was unlikely considering the fact that the coding sequences of the cDNA were flanked by untranslated regions at both ends. Moreover, the nucleotide and amino acid sequences of the isolated cDNA matched a previously reported CAT3 gene from chromosome I (GenBank\textsuperscript{®} accession number U43147; [31]). Subsequent analysis at the nucleotide level revealed some mistakes in the prediction of the coding sequences of the annotated gene which accounted for the discrepancy seen at the amino acid level (Figure 3D, middle and lower panels). The apparently longer N-terminus of the annotated protein was due to a falsely predicted first exon from a region that corresponds to the first intron of the CAT3 gene. The main factor contributing to the extra sequences at the C-terminus of the annotated protein came from the six additional exons predicted after the CAT3 termination codon, which entirely overlapped the downstream CAT1 gene. Hence, the annotated sequence is a fusion of the CAT3- and the CAT1-coding sequences, which explains the longer than expected translated polypeptide. This discovery also indicates that the second ORF of U2 in fact codes for a CAT3 enzyme of 492 amino acid residues whose gene resides on chromosome I of Arabidopsis. The chimeric U2 clone was probably an artefact generated during the construction of the pFL61 cDNA library, a not uncommon occurrence [14,32].

**Confirmation that U6 cDNA encodes UROS**

To provide independent evidence of the function of the ORF common to the two complementing plasmids, plasmid pU6.KS was transformed into the E. coli hemD mutant SASZ31, which grows as microcolonies [21]. Normal-sized colonies were observed on LB medium (Figure 4A), demonstrating that the U6 cDNA was able to complement the defect, and to the same extent as the hemD gene from A. nidulans [9]. This indicates that the ORF encodes a functional UROS; this is referred to as AtUROS from now on. The complete cDNA from pU6.KS was subcloned into a pET expression vector to express the protein with an N-terminal His\textsubscript{6}-tag, and transformed into E. coli strain BL21. Analysis of total cell proteins by SDS/PAGE revealed that after induction with IPTG a strongly staining band of 34 kDa was visible, which was not seen in uninduced cells (Figure 4B, compare lane 1 with lane 2). The identity of the protein was confirmed by N-terminal sequencing (results not shown). However, the majority of protein was in inclusion bodies, with very little in the soluble fraction (Figure 4B, lane 3). AtUROS appears to have an N-terminal extension compared with UROS proteins from other organisms, most probably an organelle-targeting peptide. We made a construct in which the first 40 residues were removed but, on induction with IPTG, the cells died. The reason for the lethality of this construct is unknown, so in an attempt to avoid this problem, another construct, pAtUROS\textsubscript{Δ41}.ET28b, was made in which the UROS protein started at residue 82, corresponding
Comparison of AtUROS with homologues from other organisms

Sequence similarity searches with AtUROS using the BLAST algorithm [29] did not convincingly identify UROS from bacteria
or animals. However, BLAST searches of the EST databases for various crop plant species enabled us to identify clones for UROS from potato, tomato, soybean and wheat, and from rice genome. These plant proteins are all very similar to one another (approx. 50% identity) suggesting that these have diverged from the UROS enzymes found in other kingdoms early on in evolution. The plant enzymes share 26% identity with that from cyanobacteria.

A comparison of the Arabidopsis and rice UROS sequences with those for UROS enzymes from other organisms was made using the ClustalW version 1.81 program [20] (Figure 6). Mathews et al. [12] compared the human enzyme with those of Drosophila, yeast and some bacteria, and found seven invariant residues, of which mutation in just three, Thr103, Tyr168 and Thr228, affected enzyme activity. AtUROS, and the other plant sequences, contain all seven invariant residues (boxed and asterisked in Figure 6), and 12 of the 15 conserved residues (boxed). Interestingly, the E. coli UROS differs at three of the invariant positions (arrowed), including Thr228. When a structure-based search was carried out with AtUROS on the structures in the PDB (Protein Data Bank) using the program FUGUE [36], the top hit was that for human UROS, with a Z-score of 20.41. This demonstrates that, despite a lack of primary-sequence conservation, the structural elements of the protein have been conserved.

**Subcellular localization of AtUROS**

As mentioned above, the one striking difference between the plant enzymes and those from other species is an N-terminal extension (Figure 6), which is rich in hydroxylated residues, suggesting that it was a chloroplast transit peptide. Analysis of the Arabidopsis sequence by ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) strongly predicted a plastid location for the protein, with a probable cleavage site after 81 residues (indicated by a diamond in Figure 5). We investigated this experimentally using an import assay in vitro and GFP-fusion proteins in vivo. For the import assay in vitro, chloroplasts were isolated from 8-day-old pea shoots as described in the Experimental section, and incubated with radiolabelled AtUROS precursor. Upon membrane purification, the 34 kDa precursor was processed to a smaller mature protein of approx. 29 kDa (Figure 7A). The size difference would correspond to a loss of approx. 40 residues from the N-terminus, rather than the predicted 81 amino acids. The reason for this discrepancy is unknown, but the most probable explanation is that the protein runs anomalously on SDS/PAGE. As mentioned above, we attempted to overexpress a form of AtUROS in E. coli in which the first 40 amino acids were removed, but this proved to be toxic to the cells. When the chloroplasts were treated with thermolysin, the mature protein was protected from proteolytic degradation, indicating that it was enclosed within the chloroplasts (Figure 7A). The mature protein was found mainly in the stroma, although a small proportion was associated with the membrane fractions. Incubation of the radiolabelled precursor protein with isolated pea mitochondria did not result in any import or processing (results not shown), suggesting that the protein is only targeted to plastids.

To verify that the targeting of AtUROS in vitro reflected the location in vivo, the entire coding sequence of AtUROS was fused in-frame to the 5'-end of the coding sequence for GFP [27], under the regulation of a cauliflower mosaic viral promoter (CaMV 35S). The expression cassette was introduced into tobacco leaves via biolistic bombardment. In a transformed cell expressing recA–GFP fusion proteins (Figure 7B, panel I). Furthermore, the distribution pattern of the GFP fluorescence resembled the discrete pattern displayed by a transformed cell expressing recA–GFP fusion proteins (Figure 7B, panel D). In contrast, GFP on its own accumulated mainly in the cytoplasm and the nucleoplasm, but was excluded from other organelles (Figure 7B, panel A), as reported previously [25]. The absence of UROS from other subcellular compartments is in line with the fact that the preceding enzyme PBG deaminase is confined to the plastid [37], and a similar location would be likely so as to avoid the possibility of non-enzymatic cyclization of HMB to the non-metabolizable type I isomer. A later enzyme coproporphyrinogen oxidase is also found only in plastids [38,39],...
bacteria (including *E. coli*) the enzyme ALA synthase, whereas in plants, algae and most plant pathways, positioned as it is at an important branchpoint [1].

Errors in initial genome annotation can also confound gene discovery between homologous proteins requires alternative experimental approaches to the use of sequence similarity searches. In contrast, functional complementation overcomes these difficulties, both in initial isolation, and in the verification of genes found from database searching. However, there are pitfalls to this approach as well, since complementation of a mutant phenotype can occur via a different route. As an example, the recombination protein. The identification of the plant gene for haem biosynthesis in that organelle.

Conclusions

The plethora of genome sequences provide a tremendous resource for gene identification, but in some cases limited sequence conservation between homologous proteins requires alternative experimental approaches to the use of sequence similarity searches. Errors in initial genome annotation can also confound gene discovery. In contrast, functional complementation overcomes these difficulties, both in initial isolation, and in the verification of genes found from database searching. However, there are pitfalls to this approach as well, since complementation of a mutant phenotype can occur via a different route. As an example, in mammals, fungi and the α-subgroup of proteobacteria such as *Rhodobacter* and *Rhizobium* ssp., the initial tetrapyrrole precursor ALA is synthesized from succinyl-CoA and glycine by the enzyme ALA synthase, whereas in plants, algae and most bacteria (including *E. coli*), ALA is derived from glutamate [41]. Nevertheless, mouse ALA synthase rescues the *E. coli* *hemA* mutant [42], and plants defective in ALA synthase were complemented by the ALA synthase gene from yeast [43]. For the UROS cDNA from *Arabidopsis* that we isolated by functional complementation, its identity was verified by enzyme activity of the recombinant protein. The identification of the plant gene for UROS opens the way to study the role of this enzyme in the plant pathway, positioned as it is at an important branchpoint [1].

We thank the CVCP (Committee of Vice-Chancellors and Principals) of the Universities of the United Kingdom for the Overseas Research Students Award to F.-C.T., and for funding from the Deutsche Forschungsgemeinschaft. We are grateful to the *Arabidopsis* Biological Resource Center for supplying the pmiGFP plasmid, Professor James A.H. Murray (Institute for Biotechnology, University of Cambridge, Cambridge, U.K.) for his gift of the pAF6-KANMX4 plasmid, and Dr Ian Small (ARC Centre of Excellence in Plant Energy Biology, University of Western Australia, Western Australia, Australia) for donating the *recA–GFP* plasmid.

REFERENCES


© The Authors. Journal compilation © 2008 Biochemical Society
Uroporphyrinogen III synthase gene from Arabidopsis


Received 7 June 2007; accepted 26 November 2007
Published as BJ Immediate Publication 28 November 2007, doi:10.1042/BJ200703770

© The Authors Journal compilation © 2008 Biochemical Society