



HAL
open science

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

Fui-Ching Tan, Qi Cheng, Kaushik Saha, Ilka U Heinemann, Martina Jahn, Dieter Jahn, Alison G Smith

► To cite this version:

Fui-Ching Tan, Qi Cheng, Kaushik Saha, Ilka U Heinemann, Martina Jahn, et al.. Identification and characterization of the Arabidopsis gene encoding the tetrapyrrole biosynthesis enzyme uroporphyrinogen III synthase. *Biochemical Journal*, 2008, 410 (2), pp.291-299. 10.1042/BJ20070770 . hal-00478823

HAL Id: hal-00478823

<https://hal.science/hal-00478823>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Fui-Ching TAN¹, Qi CHENG¹, Kaushik SAHA¹, Ilka U. HEINEMANN², Martina JAHN², Dieter JAHN² and Alison G. SMITH^{1,3}

¹Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

²Institute of Microbiology, Technical University Braunschweig, Spielmannstr. 7, 38106 Braunschweig, Germany

³Author for correspondence: Prof. Alison G. Smith
Tel: +44 1223-333952
Fax: +44 1223-333953
Email: as25@cam.ac.uk

Short title: Uroporphyrinogen III synthase gene from *Arabidopsis*

Abbreviations: ALA, 5-aminolaevulinic acid; AtUROS, *Arabidopsis thaliana* UROS; AtUROS_{Δ81}, AtUROS missing first 81 residues and with N-terminal His₆-tag; CAT, catalase; GFP, green fluorescent protein; HMB, 1-hydroxymethylbilane; ORF, open reading frame; PCR, polymerase chain reaction; UROS, uroporphyrinogen III synthase;

Summary

Uroporphyrinogen III synthase (UROS; 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 has 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 of 321 amino acids which, when subcloned into an *E. coli* expression vector, was able to complement an *E. coli hemD* mutant defective in UROS. Final proof that the ORF encodes UROS came from the fact that the recombinant protein expressed with an N-terminal His-tag was found to have UROS activity. Comparison of the sequence of AtUROS 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 to 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 confirmed that AtUROS was targeted exclusively to chloroplasts *in vivo*.

Key words: deletion mutant, functional complementation, chloroplast import *in vitro*, 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 vitamin B₁₂, or oxidatively decarboxylated in four steps to form protoporphyrin IX, the last common intermediate of haem and chlorophyll synthesis [1].

Uroporphyrinogen III is made in three enzymic steps from a five-carbon compound, 5-aminolaevulinic acid (ALA). Two molecules of ALA are condensed to form the pyrrole porphobilinogen (PBG) by a metalloenzyme, porphobilinogen 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, 1-hydroxymethylbilane (HMB) or preuroporphyrinogen III. Finally, uroporphyrinogen III synthase (UROS; EC 4.2.1.75) catalyzes the cyclization of HMB with a concomitant inversion of the fourth ring of the porphyrin 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 metabolised past the next step in the pathway. Congenital erythropoietic porphyria is a human disease caused by deficiency in UROS. This results in the accumulation of the oxidised derivatives, uroporphyrin I and coproporphyrin I, in plasma, tissues, and red 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 *E. coli* [5], with those from human [6], *Bacillus subtilis* [7], *Pseudomonas aeruginosa* [8], *Anacystis nidulans* R2 [9], mouse [10], and budding yeast *Saccharomyces cerevisiae* [11], isolated over the next few years. A comparison between UROS sequences found that there are 7 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 10 of the invariant or conserved residues that are surface-exposed. However, the overall sequence similarity between UROSs from different organisms is low, for example the *E. coli* and human sequences have less than 20% identity. This is in contrast to 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. 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 hemD* mutant deficient in UROS, to identify the corresponding gene from the cyanobacterium *A. nidulans* [9]. In this paper, 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, bacto-tryptone, bacto-peptone and bacto-agar came from Difco Laboratories Inc. (Detroit, USA), and yeast extract was obtained from Oxoid Ltd. (Basingstoke, UK). Deoxyribonucleoside triphosphates were purchased from Amersham Pharmacia Biotech (UK), Expand™ High Fidelity PCR Enzyme Mix from Roche (UK), BIOTAQ™ DNA polymerase from Bioline (UK), custom synthetic oligonucleotides and G418 (geneticin) were from Gibco BRL (Paisley, UK), and restriction enzymes were purchased from Roche (UK) and New England BioLabs (Hitchin, UK). Thermolysin and haemin were purchased from Sigma. Uroporphyrin III was from Porphyrin Products, Inc. (Logan, USA). PRO-Mix™ L-[³⁵S-methionine/cysteine] (>1000 Ci/mmol) was from Amersham Pharmacia. The Riboprobe® System, the T7 RNA polymerase, the rabbit reticulocyte lysate as well as the RNasin were supplied by Promega Corporation (Madison, USA).

Yeast strains and growth conditions

S. cerevisiae strain S150-2B (*MAT α ura3-52 trp1-289 leu2-3 leu2-112 his3 Δ 1*) was grown at 30°C in either rich glucose medium [YPD; 1% (w/v) yeast extract, 2% (w/v) bacto-peptone, 2% (w/v) glucose], or minimal glucose medium [YNB D; 0.67% (w/v) bacto-yeast nitrogen without amino acids, 2% (w/v) glucose] supplemented with 20 μ g/ml L-histidine, 20 μ g/ml L-tryptophane 20 μ g/ml uracil and 30 μ g/ml L-leucine. Strain S150-2B Δ HEM4 (*MAT α ura3-52 trp1-289 leu2-3 leu2-112 his3 Δ 1 Δ hem4::kan'*), constructed as described below, was maintained in YPD but with a supplement of 15 μ g/ml haemin and 200 μ g/ml geneticin. Transformants of S150-2B Δ HEM4 harbouring a pFL61 plasmid [14] were cultured in minimal glucose medium but without uracil, and with 200 μ g/ml geneticin. For phenotypic analysis, the yeast strains were cultured in the appropriate liquid media for one to two days at 30°C before the cultures were adjusted to the same OD₆₀₀ and serially diluted with sterile distilled water (1:10 dilutions). The diluted cells were spotted in 5- μ l droplets onto YPD, YPD + 15 μ g/ml haemin, and YPG [1% (w/v) yeast extract, 2% (w/v) bacto-peptone, 3% (v/v) glycerol] agar media, and incubated for 3 to 5 days at 30°C.

Generation of the yeast S150-2 Δ 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 chimeric gene fusion of the *E. coli* transposon *Tn903* (*kan^r* gene; [16]) coding sequence, and the promoter as well as the terminator of the *Ashbya gossypii* translation elongation factor 1 α [17], flanked at both ends by 45 bp nucleotide sequences homologous to the *HEM4* ORF, was generated by PCR. The pFA6-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 μ M forward primer (ScHEM4-KAN.for: 5'-

AGGATAAGGAAACAGAAAGGTAAAATAGACCTTGCTCGAGAGATGCTGCAGGTC
GACGGATCC-3') and 0.25 μ M reverse primer (ScHEM4-KAN.rev: 5'-
AAGTAAATAAATATAAATAGAGAGAAATATGACGTATCAATATTAATCGATGAA
TTCGAGCTC-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 sec, 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^r* gene on a YPD agar medium supplemented with 15 μ g/ml haemin and 200 μ g/ml geneticin, incubated at 30°C for 5 days. Normal-sized colonies were restreaked onto geneticin-supplemented YPD medium plus or minus haemin. Out of 16 randomly chosen transformants, one (termed S150-2 Δ 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 S150-2 Δ HEM4 by PCR

The correct replacement of the *HEM4* ORF in S150-2 Δ HEM4 by the *kan^r* disruption cassette was confirmed by PCR. Genomic DNA was extracted from S150-2B and S150-2 Δ 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'-ATAGGATCCGCTGTAGTCAGCTAAGGCGC-3'), ScUROS-rev (5'-TATGAATTCCATCGCATTCTTTCATGCCG-3') and KANMX4.iprev (5'-ACTGAATCCGGTGAGAATGGC-3'), as described in Results under the following conditions: 1 cycle of 95°C for 3 min; 30 cycles of 95°C for 30 sec, 52°C for 30 sec, and 72°C for 90 sec; 1 cycle of 72°C for 5 min.

Functional complementation of the yeast S150-2BAHEM4 mutant

The S150-2BAHEM4 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 approximately 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-tryptophane and 30 µg/ml L-leucine.

Cloning and characterisation 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'-CTTCTCCTTCCCCAATTCG-3'), AtUROS.ipF2 (5'-CCTTCTGCAGTTCGCGCC-3'), AtUROS.ipF3 (5'-GTAAGATATCTCAGATAGC-3'), AtUROS.ipR1 (5'-GATATCTTACAAGGGCTTC-3'), CAT.ipF1 (5'-ATCCAAGAGTACTGGAGG-3'), and CAT.ipF2 (5'-CTTCCAGTCAATGCTCCC-3'). Sequencing was carried out by DNA sequencing facilities in the Department of Biochemistry, University of Cambridge, UK. DNA and protein sequences were analysed using software packages of the Genetics Computer Group (GCG), University of Wisconsin, USA. Comparison of multiple sequences was conducted using the CLUSTAL W version 1.81 programme [20]. Growth and functional complementation of *E. coli* strain SASZ31 (*hemD*) [21], was 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 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_{Δ81}F (5'-GA AcatatgGCTTTGGAGAAAAATGGC-3') and AtUROS_{Δ81}R

(5'-CTTgaattcTCAATTCCTGCTGCTAGG-3'), followed by cloning the fragment into pET28b (Novagen) between *Nde*I and *Eco*R1 to form pAtUROS_{Δ81}.ET28b. This allowed synthesis of a chimeric protein with an N-terminal His₆-tag.

Recombinant production and purification of recombinant AtUROS

Recombinant AtUROS was produced in *E. coli* BL21(DE3) RIL (Stratagene, Heidelberg, Germany) containing pAtUROS_{Δ81}.ET28b. Cells were grown in LB medium at 37°C under vigorous aeration. When the cultures reached an absorbance at λ_{578} of 0.7 protein production was induced by the addition of 100 μ M IPTG. Further cultivation followed overnight at 25°C and 150 rpm. Cells were harvested, washed with buffer A (20 mM Hepes pH 7.5, 5 mM MgCl₂, 0.01% (v/v) Triton X-100) and resuspended in a minimal volume of buffer A. Bacteria were disrupted via sonication (Bandelin HD 2070, 0.5 sec sound, 0.5 sec paused, MS73 tip, 70% amplitude) and the cell free extract cleared by centrifugation at 150,000 g for 45 min. Protein integrity was verified via Western blot analysis. Recombinant UROS was purified by Ni-NTA affinity chromatography, His₆-tagged AtUROS_{Δ81} was eluted with 300 mM imidazole 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_{Δ81} were combined, concentrated (Centricon-10 filter, MWCO 10 kDa; Amicon Corp. Lexington, MA) and purified to apparent homogeneity by gel permeation chromatography using a 30 mL Superdex 200 HR 10/30 (General Electric Company, Freiburg, Germany), at a flow rate of 0.5 mL/min in buffer A. For the purposes of calibration, bovine carbonic anhydrase ($M_r = 29,000$), bovine serum albumin ($M_r = 66,000$), yeast alcohol dehydrogenase ($M_r = 150,000$), and amylase ($M_r = 200,000$) were used as marker proteins and chromatographed under identical conditions.

Determination of UROS enzymatic activity

UROS activity was determined using a coupled enzyme assay. Recombinant *P. aeruginosa* porphobilinogen synthase and *B. megaterium* porphobilinogen deaminase were purified as described before [22]. The standard assay mixture contained 25 μ g of purified recombinant AtUROS_{Δ81}, 0.2 mM aminolevulinic acid, 10 μ g porphobilinogen synthase and 10 μ g porphobilinogen deaminase in a total volume of 800 μ l 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 μ L KI/I₂ (0.5% w/v and 1% w/v in H₂O) to oxidise any uroporphyrinogen formed to uroporphyrin. Na₂S₂O₅ solution (1% w/v in water) was added to oxidise residual I₂. Proteins were precipitated by the addition of 100 μ L 50% TCA and harvested by subsequent centrifugation (13,400 rpm, 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 spectrometer (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 a 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 cyclisation of HMB, and which has identical absorbance and fluorescence properties, control experiments were performed, using either no AtUROS _{Δ 81} or heat-inactivated enzyme [24]. In neither case was any uroporphyrin detected. Thus all the uroporphyrin formed came from the activity of the recombinant 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 16 h day photoperiod. The shoots of 7 to 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 translation *in vitro* in the rabbit reticulocyte system in the presence of [³⁵S]methionine/cysteine. Chloroplast isolation and import of radiolabelled precursor protein was 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 [26].

Targeting of AtUROS-GFP fusion protein in tobacco leaves *in vivo*

The full-length protein coding sequence of AtUROS was amplified from pU6.FL61 by PCR with *Bam*HI sites at either end for in-frame fusion to the 5' end of the coding sequence for green fluorescent protein (GFP) in psmRSGFP [27] to form pAtUROS-GFP. Leaves from 5-week old tobacco (*Nicotiana tabacum* cv. Xanthi) were excised and used for biolistic transformation with pAtUROS-GFP, psmRSGFP and recA-GFP [28], followed by confocal microscopy as described by Cleary and coworkers [25].

RESULTS & DISCUSSION

Complementation of a Yeast *HEM4* mutant with *Arabidopsis* cDNAs

The Short-Homology-Flanking PCR technique [15] was used to construct yeast strain 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 supplemented with haemin (Figure 2A). However, the mutant was unable to grow on nonfermentable 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 the required nutrients plus haemin (not shown), thus providing a distinctive phenotype for selection of functionally complemented cells.

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 (data not shown).

Characterization of the complementing *Arabidopsis* cDNAs

The two complementing plasmids named pU2.FL61 and pU6.FL61 were digested with *Not* I 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 open reading frame (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 (www.ncbi.nlm.nih.gov/blast/Blast.cgi), and an *Arabidopsis* genomic BAC clone, T9J22 (GenBank accession: AC002505), was identified that was identical to 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 is common, and it is estimated that only about 20% of the originally annotated genes are structurally correct.

The protein sequences of the second ORF of U2, when searched against the *Arabidopsis* genome, matched imperfectly an annotated putative catalase-1 (CAT1) protein from the BAC F5M15 clone (GenBank accession: AC027665). The putative *CAT1* gene was predicted to encode a 1,013 amino acid protein, almost twice the size of the second ORF of U2 (Figure 3D). The extra sequences were predominantly found at the carboxyl-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 the amino acid sequences of the isolated cDNA matched a previously reported *CAT3* gene from chromosome I (GenBank accession: 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 bottom). The apparently longer amino-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 carboxyl-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 acids 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 media (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, and it 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 His₆-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 tracks 1 & 2). The identity of the protein was confirmed by N-terminal sequencing (data not shown). However, the majority of protein was in inclusion bodies, with very little in the soluble fraction (Figure 4B, track 3). AtUROS appears to have an N-terminal extension compared to UROS proteins from other organisms, most likely 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_{Δ81}-ET28b, was made in which the UROS protein started at residue 82, corresponding to the predicted start of the mature protein (see below). This time there was no effect on cell viability, and after induction with IPTG a protein of about 30 kDa was observable in *E. coli* extracts (Figure 4B, tracks 4 & 5). This would correspond to 240 residues from AtUROS with an extra 21 amino acids at the N-terminus from the vector sequence (predicted mass 27,981). Moreover, it was also present in the soluble phase (Figure 4B, track 6), providing the opportunity to carry out enzymatic analysis on the protein.

Accordingly we purified the recombinant AtUROS_{Δ81} using the N-terminal His-tag. Figure 4C shows an SDS-polyacrylamide gel of the different steps of the UROS purification. Two mg of AtUROS/liter of cell culture were recovered. A coupled assay was employed to test UROS activity, using purified recombinant *P. aeruginosa* porphobilinogen synthase and *B. megaterium* porphobilinogen deaminase to generate HMB (the substrate for UROS) enzymatically from 5-aminolaevulinic acid. Recombinant AtUROS_{Δ81} was able to convert HMB into uroporphyrinogen III as demonstrated by both fluorimetric (Figure 5) and spectroscopic detection of the oxidised product uroporphyrin III. The possibility that this was

due to oxidised uroporphyrinogen I, formed non-enzymatically, was ruled out by the fact that heat-inactivated enzyme [24] produced no measurable change in fluorescence, nor did the assay to which no AtUROS was added (data not shown). The rate of product formation was linear over 120 minutes, and also proportional to the amount of purified AtUROS added up to 50 $\mu\text{g}/\text{mL}$ (data not shown). The calculated specific activity was $76 \pm 3.8 \mu\text{mol}/\text{h}/\text{mg}$ protein, although this may not be the true rate of UROS, since this might be limited by substrate availability from the coupling enzymes.

During purification of AtUROS_{Δ81} a gel permeation chromatography step was included. By reference to the elution of molecular mass standards, a relative molecular weight of $62,000 \pm 5,000$ for native AtUROS was deduced, compared to 27,981 Da predicted for the monomeric protein with an N-terminal His₆-tag. This suggested that AtUROS is a homodimeric protein. However, purification of UROS from *E. coli* [24], rat liver [33], *Euglena gracilis* [34], and human erythrocytes [35] found in all cases that the purified enzyme was a monomer with a molecular mass of about 28-30 kDa, and it is unlikely that the plant enzyme should behave differently. Rather, the unusual asymmetric shape of the protein, as seen in the crystal structure of the human enzyme [12], may interfere with the gel filtration process.

Comparison of the 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 (about 50% identity) suggesting that these have diverged from the UROSs 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 UROSs from other organisms was made using the CLUSTAL W version 1.81 programme [20] (Figure 6). Mathews et al. [12] compared the human enzyme with those of *Drosophila*, yeasts 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 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 likely cleavage site after 81 residues (indicated by a diamond in Figure 5). We investigated this experimentally using 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 Experimental, and incubated with radiolabelled AtUROS precursor. Upon incubation with purified chloroplasts, the 34 kDa precursor was processed to a smaller mature protein of about 29 kDa (Figure 7A). The size difference would correspond to a loss of about 40 residues from the N-terminus, rather than the predicted 81 amino acids. The reason for this discrepancy is unknown, but the most likely 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. 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 (data not shown), suggesting that the protein is targeted only 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 green fluorescent protein (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 guard cell, the GFP fluorescence was detected in the chloroplasts, but not in any other compartments (Figure 7B, panel G). This was confirmed by the co-localization of GFP and chlorophyll autofluorescence (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 porphobilinogen deaminase is confined to the plastid [37], and a similar location would be likely so as to avoid the possibility of nonenzymic cyclisation of HMB to the non-metabolisable type I isomer. A later enzyme coproporphyrinogen oxidase is also found only in plastids [38, 39] whereas activity of the last two enzymes of haem synthesis, protoporphyrinogen oxidase and ferrochelatase, can be detected in plant mitochondria [38, 40] where they presumably contribute to haem biosynthesis in that organelle.

CONCLUSION

The plethora of genome sequences provides 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* spp., the initial tetrapyrrole precursor ALA is synthesised 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 synthesis 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 up the way to study the role of this enzyme in the plant pathway, positioned as it is at an important branchpoint [1].

ACKNOWLEDGEMENTS

We would like to thank the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom (CVCP) 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 psmRSGFP plasmid, Prof James A. H. Murray for his gift of the pAF6-KANMX4 plasmid, and Dr Ian Small for donating the recA-GFP plasmid.

Stage 2(a) POST-PRINT

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20070770

REFERENCES

- 1 Warren, M. J. and Scott, A. I. (1990) Tetrapyrrole assembly and modification into the ligands of biologically functional cofactors. *Trends Biochem. Sci.* **15**, 486-491
- 2 Battersby, A. R., Fookes, C.J.R., Gustafson-Potter, K.E., Matcham, G.W.J. and McDonald, E. (1979) *J. Chem. Soc. Chem. Commun.* 316-319
- 3 Desnick, R. J., Glass, I. A., Xu, W., Solis, C. and Astrin, K. H. (1998) Molecular genetics of congenital erythropoietic porphyria. *Semin. Liver Dis.* **18**, 77-84
- 4 Kappas, A., Sassa, S., Galbraith, R. A. and Nordmann, Y. (1995) *The Porphyrrias*. McGraw-Hill, New York
- 5 Sasarman, A., Nepveu, A., Echelard, Y., Dymetryszyn, J., Drolet, M. and Goyer, C. (1987) Molecular cloning and sequencing of the *hemD* gene of *Escherichia coli* K-12 and preliminary data on the Uro operon. *J. Bacteriol.* **169**, 4257-4262
- 6 Tsai, S. F., Bishop, D. F. and Desnick, R. J. (1988) Human uroporphyrinogen III synthase: molecular cloning, nucleotide sequence, and expression of a full-length cDNA. *Proc. Natl. Acad. Sci. USA* **85**, 7049-7053
- 7 Hansson, M., Rutberg, L., Schroder, I. and Hederstedt, L. (1991) The *Bacillus subtilis* *hemAXCDBL* gene cluster, which encodes enzymes of the biosynthetic pathway from glutamate to uroporphyrinogen III. *J. Bacteriol.* **173**, 2590-2599
- 8 Mohr, C. D., Sonstebly, S. K. and Deretic, V. (1994) The *Pseudomonas aeruginosa* homologs of *hemC* and *hemD* are linked to the gene encoding the regulator of mucoidy AlgR. *Mol. Gen. Genet.* **242**, 177-184
- 9 Jones, M. C., Jenkins, J. M., Smith, A. G. and Howe, C. J. (1994) Cloning and characterisation of genes for tetrapyrrole biosynthesis from the cyanobacterium *Anacystis nidulans* R2. *Plant Mol. Biol.* **24**, 435-448
- 10 Bensidhoum, M., Ged, C. M., Poirier, C., Guenet, J. L. and de Verneuil, H. (1994) The cDNA sequence of mouse uroporphyrinogen III synthase and assignment to mouse chromosome 7. *Mamm. Genome* **5**, 728-730
- 11 Amillet, J. M. and Labbe-Bois, R. (1995) Isolation of the gene HEM4 encoding uroporphyrinogen III synthase in *Saccharomyces cerevisiae*. *Yeast* **11**, 419-424
- 12 Mathews, M. A., Schubert, H. L., Whitby, F. G., Alexander, K. J., Schadick, K., Bergonia, H. A., Phillips, J. D. and Hill, C. P. (2001) Crystal structure of human uroporphyrinogen III synthase. *EMBO J.* **20**, 5832-5839

- 13 Murray, J.A.H. and Smith, A.G. (1996) Functional complementation in yeast and *E. coli*. In Plant Gene Isolation: Principles and Practice (Foster, G.D. and Twell, D., eds.), pp 177-211, J Wiley & Sons, Chichester
- 14 Minet, M., Dufour, M. E. and Lacroute, F. (1992) Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. *Plant J.* **2**, 417-422
- 15 Wach, A., Brachat, A., Pohlmann, R. and Philippsen, P. (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793-1808
- 16 Oka, A., Sugisaki, H. and Takanami, M. (1981) Nucleotide sequence of the kanamycin resistance transposon Tn903. *J. Mol. Biol.* **147**, 217-226
- 17 Steiner, S. and Philippsen, P. (1994) Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus *Ashbya gossypii*. *Mol. Gen. Genet.* **242**, 263-271
- 18 Gietz, R. D. and Woods, R. A. (1994) *Molecular Genetics of Yeast: A Practical Approach*. Oxford University Press, Oxford
- 19 Rose, M. D., Winston, F. and Hieter, P. (1990) *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbour Laboratory Press, New York
- 20 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680
- 21 Chartrand, P., Tardif, D. and Sasarman, A. (1979) Uroporphyrin- and coproporphyrin I-accumulating mutant of *Escherichia coli* K12. *J. Gen. Microbiol.* **110**, 61-66
- 22 Raux, E., Leech, H. K., Beck, R., Schubert, H. L., Santander, P. J., Roessner, C. A., Scott, A. I., Martens, J. H., Jahn, D., Thermes, C., Rambach, A. and Warren, M. J. (2003) Identification and functional analysis of enzymes required for precorrin-2 dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in *Bacillus megaterium*. *Biochem. J.* **370**, 505-516
- 23 Jordan, P. M. (1982) Uroporphyrinogen III cosynthetase: a direct assay method. *Enzyme* **28**, 158-169
- 24 Alwan, A. F., Mgbeje, B. I. and Jordan, P. M. (1989) Purification and properties of uroporphyrinogen III synthase (co-synthase) from an overproducing recombinant strain of *Escherichia coli* K-12. *Biochem. J.* **264**, 397-402

- 25 Cleary, S. P., Tan, F. C., Nakrieko, K. A., Thompson, S. J., Mullineaux, P. M., Creissen, G. P., von Stedingk, E., Glaser, E., Smith, A. G. and Robinson, C. (2002) Isolated plant mitochondria import chloroplast precursor proteins *in vitro* with the same efficiency as chloroplasts. *J. Biol. Chem.* **277**, 5562-5569
- 26 Suzuki, T., Masuda, T., Singh, D. P., Tan, F. C., Tsuchiya, T., Shimada, H., Ohta, H., Smith, A. G. and Takamiya, K. (2002) Two types of ferrochelatase in photosynthetic and nonphotosynthetic tissues of cucumber: their difference in phylogeny, gene expression, and localization. *J. Biol. Chem.* **277**, 4731-4737
- 27 Davis, S. J. and Vierstra, R. D. (1998) Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* **36**, 521-528
- 28 Akashi, K., Grandjean, O. and Small, I. (1998) Potential dual targeting of an *Arabidopsis* archaeobacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *FEBS Lett.* **431**, 39-44
- 29 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
- 30 AGI (2000) *Arabidopsis* genome initiative: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815
- 31 Frugoli, J. A., McPeck, M. A., Thomas, T. L. and McClung, C. R. (1998) Intron loss and gain during evolution of the catalase gene family in angiosperms. *Genetics* **149**, 355-365
- 32 Chow, K.-S., Singh, D. P., Walker, A. R. and Smith, A. G. (1998) Two different genes encode ferrochelatase in *Arabidopsis*: mapping, expression and subcellular targeting of the precursor proteins. *Plant J.* **15**: 531-541
- 33 Kohashi, M., Clement, R. P., Tse, J. and Piper, W. N. (1984) Rat hepatic uroporphyrinogen III co-synthase. Purification and evidence for a bound folate coenzyme participating in the biosynthesis of uroporphyrinogen III. *Biochem. J.* **220**, 755-765
- 34 Hart, G. J. and Battersby, A. R. (1985) Purification and properties of uroporphyrinogen III synthase (co-synthetase) from *Euglena gracilis*. *Biochem. J.* **232**, 151-160
- 35 Tsai, S. F., Bishop, D. F. and Desnick, R. J. (1987) Purification and properties of uroporphyrinogen III synthase from human erythrocytes. *J. Biol. Chem.* **262**, 1268-1273

- 36 Shi, J., Blundell, T.L. and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243-257
- 37 Witty, M., Jones, R. M., Robb, M. S., Jordan, P. M. and Smith, A. G. (1996) Subcellular location of the tetrapyrrole synthesis enzyme porphobilinogen deaminase in higher plants: an immunological investigation. *Planta* **199**, 557-564
- 38 Smith A. G., Marsh, O. and Elder, G. H. (1993) Investigation of the subcellular location of the tetrapyrrole-biosynthesis enzyme coproporphyrinogen oxidase in higher plants. *Biochem. J.* **292**, 503-508
- 39 Santana, M. A., Tan, F. C. and Smith, A. G. (2002) Molecular characterisation of coproporphyrinogen oxidase from *Glycine max* and *Arabidopsis thaliana*. *Plant Physiol. Biochem.* **40**, 289-298
- 40 Cornah, J. E., Roper, J. M., Singh, D. P. and Smith, A. G. (2002) Measurement of ferrochelatase activity using a novel assay suggests that plastids are the major site of haem biosynthesis in both photosynthetic and non-photosynthetic cells of pea (*Pisum sativum* L.). *Biochem. J.* **362**, 423-432
- 41 Jordan, P. M. (1991) The biosynthesis of 5-aminolaevulinic acid and its transformation into uroporphyrinogen III. In *Biosynthesis of Tetrapyrroles* (Jordan, P.M., ed.) pp. 1-35, Elsevier, Amsterdam
- 42 Schoenhaut, D. S., and P. J. Curtis (1986) Nucleotide sequence of mouse 5-aminolevulinic acid synthase cDNA and expression of its gene in hepatic and erythroid tissue. *Gene* **48**,55-63
- 43 Zavgorodnyaya A., Papenbrock J and Grimm B (1997) Yeast 5-aminolevulinic acid synthase provides additional chlorophyll precursor in transgenic tobacco *Plant J.* **12**, 169-178

FIGURE LEGENDS

Fig. 1. The deletion of the *HEM4* ORF in S150-2 Δ HEM4. The genomic DNA of S150-2B (WT) and S150-2 Δ HEM4 (h4) was individually amplified via PCR using three specific primers as follows: F1 (ScUROS.for; a forward primer homologous to a region upstream of the recombination site), R1 (ScUROS.rev; a reverse primer homologous to the coding sequence of *HEM4*), and R2 (KANMX4.iprev; a reverse primer homologous to the coding sequence of *kan^r*). 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, while the light grey and the hollow bars indicate the *HEM4* ORF and the *kan^r* gene respectively.

Fig. 2. Growth of the rescued S150-2 Δ HEM4 mutants on fermentable and non-fermentable carbon sources. The two rescuing strains, ScU2 and ScU6, together with S150-2B (WT) and S150-2 Δ HEM4 (Δ h4) were grown as suspension cultures, serially diluted in sterile distilled water, and then spotted onto **A.** glucose medium enriched with haemin (YPD+haemin), **B.** glucose medium alone (YPD), or **C.** glycerol medium alone (YPG). The plates were incubated at 30°C for 3 to 5 days.

Fig. 3. Structural features of cDNAs carried by Sch2 and Sch6. The complementing plasmids in Sch2 and Sch6 were isolated, and the cDNAs were subcloned into a pBluescript II KS vector for sequencing. **A.** The plasmids isolated from Sch2 and Sch6, pU2.FL61 and pU6.FL61 respectively, were digested with *NotI*, which released the inserts from the vector. Lane 1 is the digested pU2.FL61 plasmid, while Lane2 contains the restricted fragments of pU2.FL61. **B.** The 1.4 kb U6 insert encodes a single ORF, whereas the 3 kb U2 insert carries two adjoining cDNAs that are arranged in opposite direction, as determined by the presence of 5' and 3' UTRs, including poly A tails. The first ORF (black) in U2 is identical to the one in U6, whereas the second ORF (gray) in U2 completely unrelated. The dotted lines flanking both ends of the inserts are the vector sequences. The position of various sequencing primers are indicated by arrows. UTR, untranslated region. **C.** The exon-intron organisation of At2g26540, showing above the relationship to the hypothetical protein in the initial genome annotation [30], and below the U6 ORF, encoding AtUROS. The latter is organised into nine

exons (black solid boxes) separated by eight introns (thin solid bars). **D.** Genomic DNA from chromosome I showing the *CAT1* and *CAT3* genes encoding catalases are linked in tandem. Below are shown the exons of *CAT3*, encoded by the second ORF in U2, as light-grey boxes, while those of *CAT1* are coloured in dark-grey. The originally annotated *CAT1* (GenBank: AC027665) shown above is a merge of the *CAT1* and *CAT3* sequences.

Fig. 4. Expression of U6 cDNA encoding AtUROS in *E. coli*

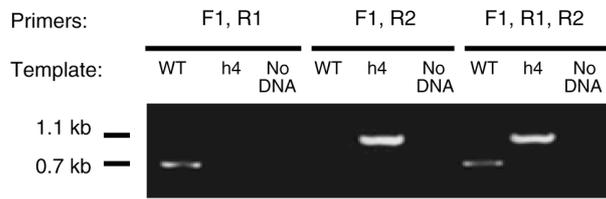
A. The *E. coli hemD* mutant SASZ31, defective in UROS, was transformed with the *Arabidopsis* U6 cDNA (pAtUROS), the *hemD* gene from the cyanobacterium *Anacystis nidulans* [9] (pAnUROS), or pBluescript SK alone (pSK), and plated onto minimal medium in the absence of haem. **B.** SDS-PAGE of extracts from *E. coli* BL21 cells containing the full-length cDNA pAtUROS (lanes 1-3) or pAtUROS_{Δ81}, in which the first 81 amino acids had been removed (lanes 4-6). Lane 1 was without induction, whereas lanes 2- 6 were after induction by 100 μM IPTG. Lanes 1, 2, 4 & 5 were total protein, and lanes 3 & 6 were the soluble fraction. The arrows indicate the overexpressed full-length and AtUROS_{Δ81} proteins respectively. **C.** SDS-PAGE illustrating purification of AtUROS_{Δ81}. Lane 1: molecular weight markers; lane 2: total cellular extract without induction; lane 3: total cellular extract after overnight induction with 100 μM IPTG; lane 4: eluate from the Ni-NTA column with imidazole; lane 5: eluate from DEAE sepharose column; lane 6: purified AtUROS_{Δ81} after gel filtration, showing a single band of 30 kDa.

Fig. 5. Formation of uroporphyrinogen III by recombinant AtUROS_{Δ81}. UROS catalyses the conversion of HMB into the first planar tertapyrrole uroporphyrinogen III. The substrate for UROS, HMB was formed in situ by the inclusion of porphobilinogen synthase, porphobilinogen deaminase and 5- aminolaevulinic acid. The amount of enzymatically formed uroporphyrinogen III produced by AtUROS_{Δ81} was determined via fluorimetric detection of its oxidised form, uroporphyrin III, with fluorescence maxima at 600 nm and 620 nm. Presented are the emission spectra from 540 to 700 nm with an excitation wavelength of 400 nm. The possibility that the changes seen here were due to nonenzymatic cyclisation of HMB to the type I isomer of uroporphyrin were ruled out by the fact that omission of AtUROS, or addition of heat-inactivated protein [24] resulted in no change in fluorescence over time (not shown).

Fig. 6. Alignment of UROS homologues from various organisms. The amino acid sequences of UROS from various organisms were aligned using the CLUSTALW 1.81 programme [20] with some manual adjustments. In addition to AtUROS (A.tha), the following sequences are from the SWISSPROT database: O.sat, *Oryza sativa* (rice; Q10QR9); S.cer, *Saccharomyces cerevisiae* (P06174); H.sap, *Homo sapiens* (P10746); A.nid, *Anacystis nidulans* R2 (P42452); E.col, *Escherichia coli* (P09126). C.alb, *Candida albicans* (CAA22001) and D.mel, *Drosophila melanogaster* (AAF46419), are from GenBank database entries. Bold residues are those that are completely conserved between all 8 sequences. Boxed residues are those identified as invariant (also asterisked) or conserved between animal, fungal and some bacterial enzymes by Mathews et al [12]. Three of the invariant residues they identified are not conserved in *E. coli* (marked with filled arrow head), although they are present in the two plant UROSs. Residue 82, indicated with an open diamond, is the predicted site of cleavage of the chloroplast transit peptide, and is where the enzyme was fused with an N-terminal His₆-tag to allow purification of AtUROS_{Δ81}.

Fig. 7. Subcellular localization of Arabidopsis UROS. **A.** Isolated pea chloroplasts were incubated with ³⁵S-labelled precursor protein, and then fractionated into stroma, thylakoid and envelope. Each fraction was analyzed on 10% PAGE. TP, translation product; wC, washed chloroplasts; pC, protease-treated chloroplasts; S, stroma; T, thylakoid; pT, protease-treated thylakoid; E, envelope. **B.** GFP fusion cassettes were introduced into tobacco leaf guard cells via particular bombardment. GFP on its own displays cytoplasmic and nucleoplasmic localization: (A), (B) and (C); recA-GFP localizes exclusively to the chloroplasts: (D), (E) and (F); and AtUROS-GFP is confined to the chloroplasts: (G), (H) and (I). The scale bar at the bottom left corner in each picture represents 10 μm.

Figure 1



S150-2B (WT)

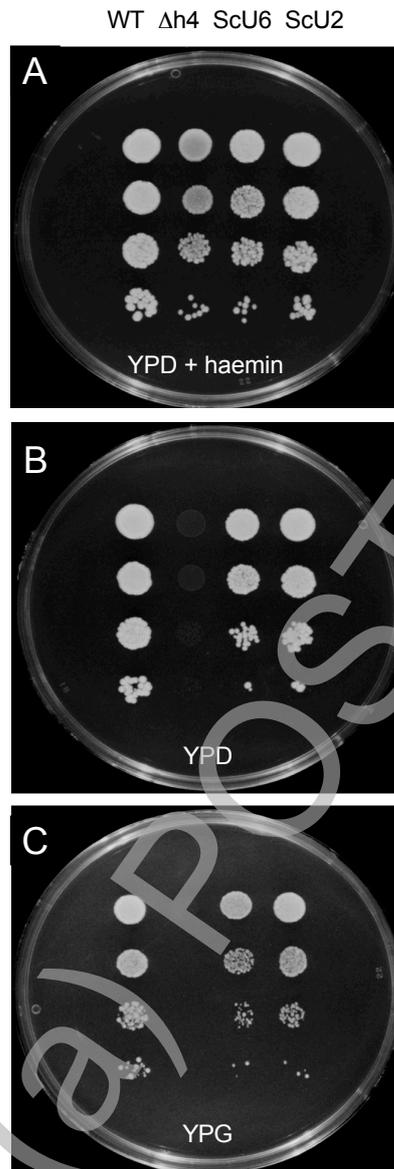


S150-2B Δ HEM4 (h4)



Stage 2(a) POST-PRINT

Figure 2



Stage 2 (e) POST-PRINT

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20070770

Figure 3

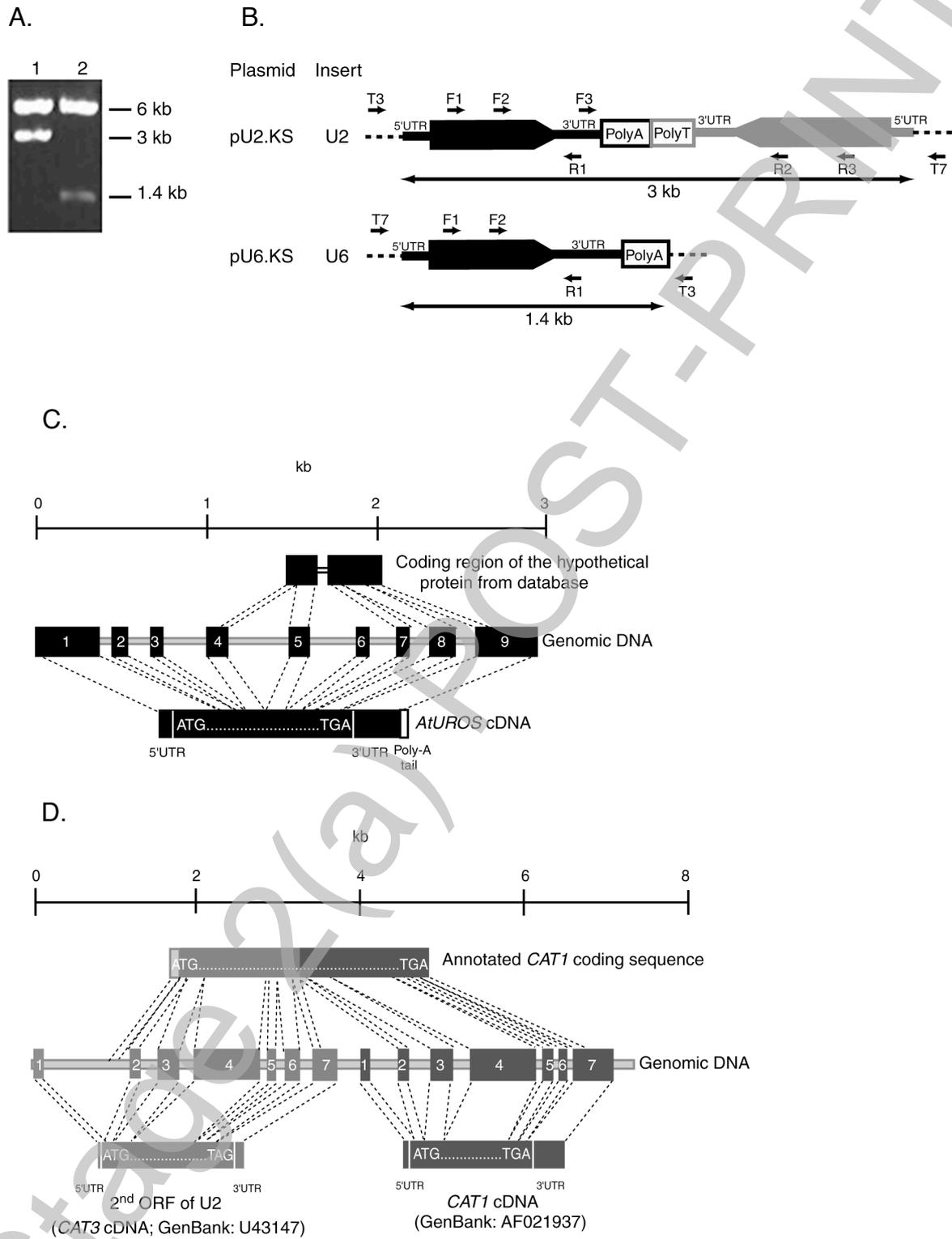


Figure 4

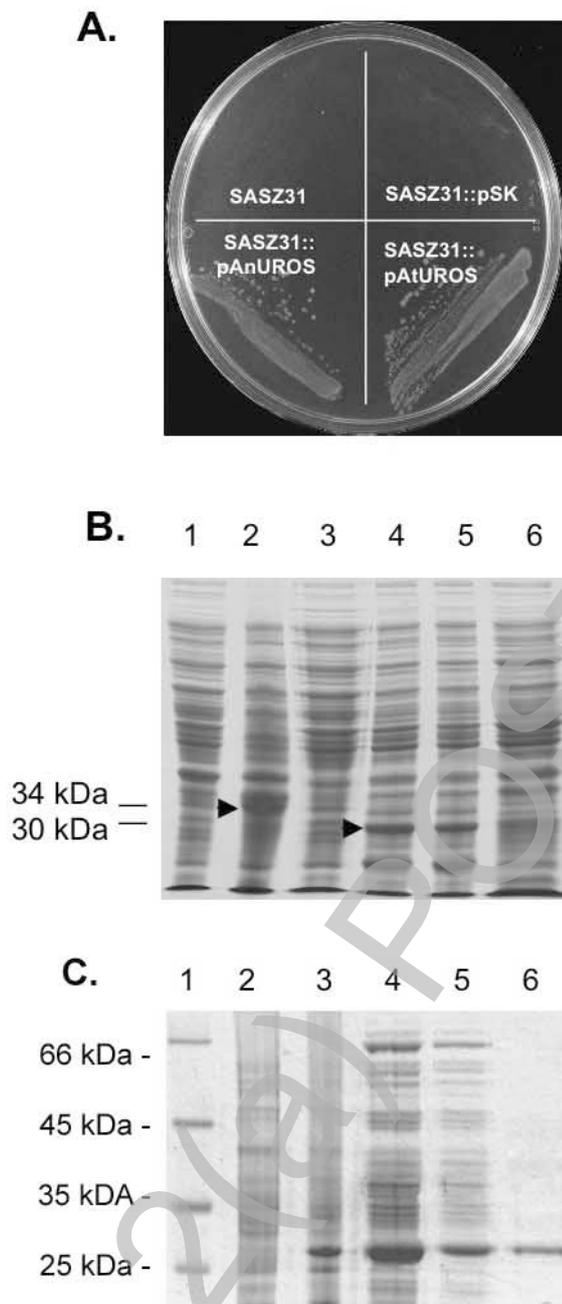


Figure 5

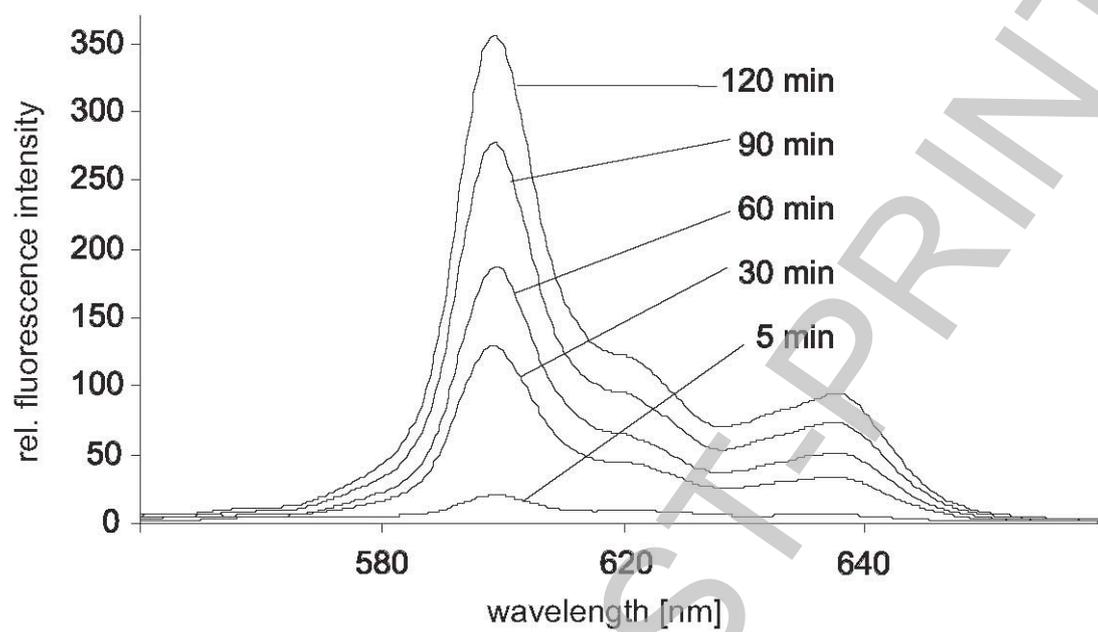


Figure 6

A.tha MALLLLSHCSILSFQPPLSSSSSFHSSHVQSLSKPVFASPSPIRNSISSSVSSSS
O.sat MALSSSSH-LLPFSRPPAT----FPRARHAGGGRG-----RAGATGR-----

◇

A.tha SSVSSSNSIPQVVVTRER GK---NNQIIKALEKNG-ISSLELPLIQHARGPDFDRLASVL
O.sat FMACSSPPPPDVVVTRER GK---NAKLIAALEKHN-VQSLELPLIKHVEGPDTRLSAVL
A.nid -MAEQPLIGKTILTTRAAGQ---SSPFAAQLRAAG-AAVIEMPTLEIG---PPSSWLPLD
E.col -----MSILVTRPSPA---GEELVSRLRTLQ-QVAWHFPLIEFS---PGQQLPQLA
H.sap -----MKVLLKDAKEDDCGQDPYIRELGLYG-LEATLIPVLSFEFLSLPSFSEKLS
S.cer ---MSSRKVRVLLKKNKTVP---IDKYELECR-SKAFEPFVPLIKHTHV-IQDFRNVLN
D.mel ---MTSRQRTVIFKSESES---SDVYAETLEKHD-FNPVVFVPTLSFGFKNLDELRAKLO
C.alb ---MTN-----VILLKNASIP---CDPYDIKFSNSNKYKPNFVPLLTHRHK-DKSQTL SFL

A.tha NDKSF D---WIIITTSPEAGSVFLEAW-----KTASSPEVOIGVVGAGTARVF
O.sat RDEKFD---WIIITTSPEAAAVFLEGW-----KAAGNPKVRIAVVGAGTERVF
A.nid EAIAAIADFDWIIILASANA VEAVOORLAAQ-----QKSWSDVPCAIAVVGKTTAQLV
E.col DQLAALGESDLIFALSOHAVAFASOQLHQO-----DRKWPRLPDYFAIGRTTALAL
H.sap HPEDYGG---IIFTS PRAVEAAELCLEQNNKTEVWERSLKEKWNKSVYVVGNTASLV
S.cer TIPNYLNTINYIIITTSQRTVESLNEAIIPTLTS-----EQKAALLSKTVYTVGPATTANFI
D.mel NPDKYAG---IIFTS PRCVEAVAESLNLG-----ELPGGWKMLHNYAVGVEVTHNLA
C.alb ISDEF LNPIFIITTSQRAVEMFKECIE-ELNH-----DIRORIYQKIGYTVGPATYKIL

* * *

A.tha EEA-MKSADGLLHVAFTPSKATGKVLASELPEKVGKRS SVIY--PASLKAGNDIVEGLSK
O.sat DEV-IQYNDGSLEVAFSPSKAMGKFLASELPRTTETTCKVLY--PASAKAGHEIQNGLSN
A.nid AAQ-GGKADYIPP-EFIAESLMEHFPQPVAG---QR---LLFPRVETGGREQITQALQS
E.col HTVSGQKILYPQDREISEVLLQLPELQNIAG-----KRALILRGNGG-RELIGDTLLTA
H.sap SKI-G--LDTEGETCGNAEKLA EYIC SRE---SSALP--LLFPCGNLK-REILPKALKD
S.cer RRS-GF INVKGGEDAGNGSILADIIIDDLSTD IKACPPSELLFLVGEIR-RDIIPKLLHS
D.mel LST-LDQLFTHGKQTGNARALGDYIVDTFDG--SRALP--LLLPCGNLA-TDTLLSKLAE
C.alb KSV-GFKDVRGGDEAGNGSKLADLIKQDTIG--RENIP--MVFFTGVIR-KDIIPRKLID

* * *

A.tha RGFE-VVRLNTYTT--VPVQSVDTVLLQOALSAPV-----LSVA SPSAVRAWLHLI
O.sat RGFE-VTRLNTYTT--VPVQVDVPLILK PALSAPV-----VAVASPSALRAWLNLA
A.nid QGAI-VVEVPAYESRCPSQIPDDALIALRQ AHLNL-----ISFTSSKTVRNFCQLM
E.col RGAE-VTFCECYQR-CAIHYDGAE EAMRWQAREVTM-----VVVTS GEMLQQLWSLI
H.sap KGIA-MESITVYQTVAHPGIQGNLNSYYSQQGPAS-----ITFFSPSGLTYSLKHI
S.cer KGIK-VREVVTYKT---EELSDGFKRFIHAM-KECDEDEVFSDVWVVFSPQGTKEITQYL
D.mel NGFS-VDACEVYETRCHPELGANVERALEIYGESIEF-----LAFFSPSGVNCAQQYF
C.alb SGYNNFOEFIIYQTDRLDIIDNFKKVIHGLDKDKDND DV---WIVFFSPQGTKEIVNYL

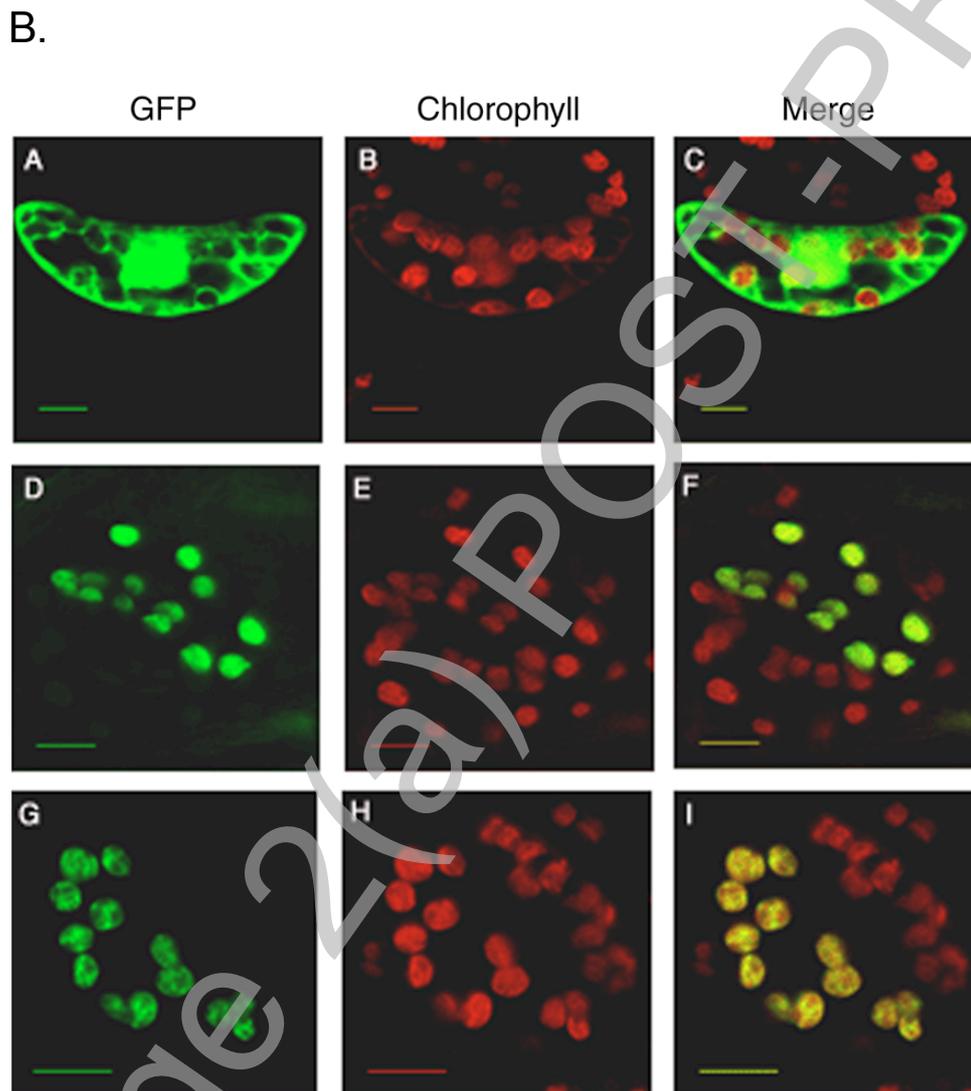
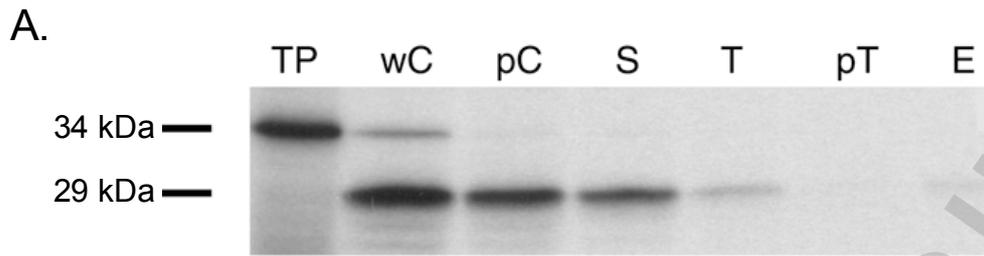
* * *

A.tha QN-----EEQWSNYVACIGETTASAARRLGLKNVYYPEKPGLEGWVESIMEALGAHAD
O.sat SQ-----VDNWGNALACIGETTASA AKKFLKSIYYPTTPGLDGWVESILEALRAHGQ
A.nid ASN LGVDWSARISGVAIASIGPOTSITCOELLGRVEVEAQEYTL DGLLLAIE-----
E.col PQ---WYREHWLLHCRLLVVSERLAKLARELGWODIKVADNADNDALLRALQ-----
H.sap QE---LS-GDNIDQIKFAAIGPTTARALAAOGLPV SCTAESPTPQALATGIR-----K
S.cer GD---SNRLPG-SHLRVASIGPTTKKYLDDNDVTS DVVSPKPDPKSLLDAIE-----
D.mel TS---RQ-LS-MDKWKLVAIGPSTRRALES LGQKVYCTAERPTVEHLVKVLLNPQDSRER
C.alb ID---RDGNSNQKNWKIASIGPTTRDYLLKDFDLSPHVIAPKPDPEALFETIF-----

* * *

A.tha SSNPSSRN-----
O.sat SKEVLSRPRGYFP
A.nid -QWARQTT-----
E.col -----
H.sap ALQPHGCC-----
S.cer LYQRHK-----
D.mel LLKERERMAAIENNN
C.alb QYDKTYAL-----

Figure 7



THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20070770