Cloning, sequencing and heterologous expression of the gene for lupanine hydroxylase, a quinocytochrome *c* from a *Pseudomonas* sp.

David J. HOPPER^{*1}, Mustak A. KADERBHAI^{*}, Shirley A. MARRIOTT^{*}, Michael YOUNG^{*} and Jerzy ROGOZINSKI[†]

^{*}Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, U.K., and [†]Department of Biochemistry, Institute of Plant Biology, Academy of Agriculture, ul. Rakowiecka 26/30, Warsaw 02-528, Poland

The gene encoding the enzyme lupanine hydroxylase was isolated by PCR using chromosomal DNA from a lupanine-utilizing *Pseudomonas* sp. as template and primers based on the sequences of the N- and C-termini of the purified protein. The derived sequence for the mature gene product gave a protein with an M_r of 72256, in good agreement with the value found by SDS/PAGE of the pure enzyme, and contained the sequences of several peptides obtained after endoproteinase Lys-C digestion of the pure enzyme. The gene, under the transcriptional control of a *phoA* promoter and with the *Escherichia coli* alkaline phosphatase signal sequence, was expressed in *E. coli* containing a plasmid expressing the genes for cytochrome *c* maturation proteins constitutively. Haem-containing inactive protein in inclusion bodies was renatured and reactivated with pyrroloquinoline quinone (PQQ) and Ca^{2+} to give active enzyme. The lupanine hydroxylase (*luh*) gene coded for a protein with a cleavable 26-residue signal sequence at its N-terminus, required

for the transport of the enzyme to its periplasmic location. Analysis of the protein sequence showed that it contains two domains, a large PQQ-binding N-terminal domain and a smaller cytochrome *c* C-terminal domain. Comparison of the derived sequence with those of other proteins showed considerable similarity with other quino(haemo)proteins, including alcohol dehydrogenases from a variety of bacteria. The PQQ-binding domain sequence contains W motifs, characteristic of the eight-bladed 'propeller' structure of methanol dehydrogenase, but lacks the unusual disulphide ring structure formed from two adjacent cysteines seen in this enzyme. The C-terminus shares some similarity with bacterial cytochrome *c* and includes the haem-binding consensus sequence CXXCH.

Key words: cytochrome *c*, pyrroloquinoline quinone (PQQ), quinohaemoprotein, renaturation, signal sequence.

INTRODUCTION

Lupanine is a quinolizidine alkaloid produced by plants of the genus *Lupinus*. It was shown to be metabolized by a *Pseudomonas* sp. with conversion into 17-hydroxylupanine as the first step (Scheme 1) [1]. The enzyme involved is induced by growth of the organism on lupanine. It was partially purified by Rogozinski [2] and later purified and characterized by Hopper et al. [3]. It is located in the periplasm of the organism [4] and is a monomeric protein of M_r 72 000 containing one molecule of pyrroloquinoline quinone (PQQ) and one covalently bound haem as cofactors. It catalyses the dehydrogenation of lupanine with the PQQ acting as electron acceptor and the electrons are transferred to the haem before being passed to an external acceptor. This acceptor has not yet been identified but the enzyme will rapidly reduce horse

heart cytochrome *c*, which is used as electron acceptor in the enzyme assay. Addition of water to the dehydrogenated lupanine gives the hydroxylated product (Scheme 1).

In terms of its size and cofactor content lupanine hydroxylase is similar to the quinohaemoprotein type II alcohol dehydrogenases such as that from *Comamonas testosteroni*. This enzyme has an M_r of 73 200, contains one PQQ and one covalently bound haem and is also found in the periplasm [5]. The gene for this alcohol dehydrogenase has been cloned and sequenced [6] and recently the crystal structure has been determined [7] at a resolution of 1.44 Å. Both the derived amino acid sequence and the tertiary structure of the N-terminal PQQ-binding domain show considerable similarity with the α subunit of the PQQ-containing methanol dehydrogenases. The methanol dehydrogenases have been extensively studied and the three-dimensional structures of those from *Methylobacterium extorquens* and *Methylophilus methylotrophus* W3A1 have been elucidated using X-ray crystallography [8,9]. The haem-binding C-terminal domain of the alcohol dehydrogenase from *C. testosteroni* shows similarity to some *c*-type cytochromes.

Most PQQ-containing enzymes that have been described are dehydrogenases acting on alcohols or glucose, but lupanine hydroxylase, while catalysing a dehydrogenation, is unusual in acting on a tertiary amine. To study the relationship of lupanine hydroxylase with other quinohaemoproteins and to help us move towards an understanding of the reaction mechanism, we have now cloned and sequenced the gene for this enzyme. Previous studies of the enzyme have been hampered by the need to extract and purify lupanine from plant material for growth of the

Scheme 1 The reaction catalysed by lupanine hydroxylase

Abbreviations used: lys-C, endoproteinase Lys-C; PQQ, pyrroloquinoline quinone.

¹ To whom correspondence should be addressed (e-mail dvh@aber.ac.uk).

The nucleotide sequence of *luh* has been deposited in the GenBank®/EMBL/DDBJ Nucleotide Sequence Databases under accession number AJ318095.

bacterium and we also describe here the expression of the gene in *Escherichia coli* and reactivation of the product, giving a source of active enzyme without this requirement for lupanine.

EXPERIMENTAL

Enzyme purification and reactivation

Native lupanine hydroxylase was purified from a *Pseudomonas* sp. grown on lupanine by the procedure described by Hopper et al. [3].

Enzyme produced by expression of the *luh* gene in *E. coli* was activated and partially purified as follows. Cells harvested from 12 l of culture were resuspended in 10 mM Tris/HCl buffer, pH 8, containing 1 mM EDTA, disrupted by sonication and centrifuged at 27000 *g* for 30 min at 4 °C. The pellet was resuspended in 50 ml of 8 M urea and left on ice for 30 min. After recentrifuging, the supernatant was quickly mixed with 1.5 l of buffer containing 0.13 mM PMSF and 0.2 mM 1,10-phenanthroline and incubated at 20 °C for 30 min. A slurry (30 ml) of DEAE-cellulose in buffer was added and, after gentle stirring for 30 min at 20 °C, this was removed by filtration and eluted with 300 ml of buffer containing 1 M NaCl. The eluate was dialysed at 4 °C against water and applied to a column (6 cm × 2.5 cm) of Q Sepharose FF equilibrated with 20 mM Tris/HCl buffer, pH 8.0. This was washed with 100 ml of buffer and then with a linear gradient from 0 to 0.5 M KCl in 500 ml of buffer. Pooled enzyme-containing fractions were concentrated by pressure filtration through a PM10 membrane and further purified by FPLC on a Superdex 75 16/60 column eluted with 20 mM Tris/HCl buffer, pH 8.0, at 4 °C. Enzyme was detected in column fractions by spectrophotometric assay after reactivation of 100 µl portions by incubation with 50 µM PQQ and 1 mM CaCl₂ for 30 min at 20 °C.

Enzyme assay

Lupanine hydroxylase activity was measured by the spectrophotometric assay with horse heart cytochrome *c* as an electron acceptor [3].

Protein sequencing

The N-terminal analysis of the enzyme and sequencing of peptides obtained by digestion with endoproteinase Lys-C (EC 3.4.21.50; lys-C) were performed by automated Edman degradation on an Applied Biosystems model 473A sequencer at the sequencing service of the University of Aberdeen. C-terminal analysis was performed at the Mayo Protein Core Facility (Mayo Clinic, Rochester, MN, U.S.A.) using an Applied Biosystems Procise C chemical C-terminal sequencer.

DNA manipulations

Chromosomal DNA was isolated from a *Pseudomonas* sp. by the cetyltrimethylammonium bromide (CTAB) method of Wilson [10]. The procedure of He et al. [11] or the Wizard Midiprep DNA purification system from Promega were used to purify plasmid DNA. The standard procedures described by Sambrook et al. [12] were used for DNA restriction, phosphorylation, agarose-gel electrophoresis, PCR and transformation of competent *E. coli* cells; competent cells were prepared as described by Hanahan [13]. PCR was performed on a Hybaid OmniGene thermal cycler using SuperTaq DNA polymerase (Kramel Biotechnology, Cramlington, Northumberland, U.K.). Primers were made by MWG-Biotech AG (Ebersberg, Germany). Their sequences were: 5'-AAC GAR CAR CAY TAC TTY-3' (1F); 5'-

TTC CAR TAR TCN GT-3' (1R); 5'-AAT GAA AAA GAC GGC TCT GCG GTG-3' (2F) and 5'-CTA YTG NGC NCC-3' (2R). The PCR reaction contained 10 mM Tris/HCl, pH 8.8, 1.5 mM MgCl₂, 50 mM KCl, 0.1 % (v/v) Triton X-100, 0.2 mM of each dNTP, 10 µg of template chromosomal DNA, 25 pmol of each primer and 2 units of SuperTaq in a final volume of 50 µl overlaid with 50 µl of paraffin oil. The mixture was subjected to 10 rounds of thermocycling involving denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min and extension at 72 °C using a Hybaid thermocycler. Amplification was continued for a further 20 cycles with the annealing temperature raised to 62 °C. The amplified products were separated by agarose gel electrophoresis and the regions of interest carefully excised. The DNA was recovered in solution by using either Pharmacia Sephaglas BandPrep kit or Promega Wizard DNA Clean-Up system. DNA sequencing was performed at the Sir John Dunn School of Pathology, University of Oxford, Oxford, U.K., using an automated fluorescence Applied Biosystem ABI DNA 377 sequencer.

RESULTS AND DISCUSSION

Peptide sequences

Lupanine hydroxylase was purified from a *Pseudomonas* sp. and the sequence of the first 31 amino acids from the N-terminus was determined as NEKDGSAVTSGNWSLLGGGNEQHYSAL-KDV (sequence 1). Several internal peptides that were produced by digestion with lys-C were also sequenced, giving the following results: AWTAYDTQQTLLK (sequence 2); MQAITGAPRVG-GGK (sequence 3); NLWYTYTPEVQ (sequence 4) and TWGT-DYWK (sequence 5).

Partial gene sequence and restriction map

Part of the N-terminal sequence and sequence 5 were used to design forward 18 nt and reverse 14 nt primers (1F and 1R). Using these primers in a PCR, with chromosomal DNA from the *Pseudomonas* sp. as a template, resulted in the selective amplification of a 600 bp fragment. This was ligated into a pGEM-A/T cloning vector and transformed into *E. coli* DH5α using standard procedures [12]. The sequence of the insert in plasmid DNA isolated from one of the several white recombinant colonies had an open reading frame that corresponded to the amino acid sequences from which the primers were designed. It also encompassed sequences 3 and 4 from the lys-C digestion products of lupanine hydroxylase indicating that the PCR product was part of the *luh* gene encoding this enzyme. The *Bam*HI-excised 600 bp DNA insert was used to construct a restriction map of the area around the *luh* gene (Figure 1). Southern blotting indicated that the complete gene was located on a 6.6 kb *Eco*RI fragment. To obtain some additional sequence information upstream from the 5' end of the coding sequence of the mature protein this part of the gene was amplified by ligation-mediated PCR. Chromosomal DNA was digested with *Hind*III and ligated with the vector pMTL20 [14]. A PCR using this as the template with a forward vector-based primer and a reverse primer based on a previously sequenced fragment gave an expected fragment of 500 bp. This was recovered and ligated with the vector pXCM with the transformants screened for ampicillin resistance and kanamycin sensitivity. The sequence of the cloned fragment confirmed the N-terminal sequence of the mature protein and extended it to include the potential secretory signal sequence region.

Sequence of *luh* gene

To obtain the complete gene sequence, the C-terminus of the purified lupanine hydroxylase was sequenced and gave the

Figure 1 A restriction map of the segment of chromosomal DNA including the lupanine hydroxylase gene

The boxed region represents the PCR product obtained using primers 1F and 2R.

tripeptide GAQ. A reverse primer (2R) based on this tripeptide plus a stop codon was designed and used in a PCR in conjunction with a forward primer (2F) based on the determined sequence of the N-terminus of lupanine hydroxylase. This resulted in production of an ≈ 1800 bp fragment that contained the complete sequence for the mature enzyme. The sequence of the original 600 bp PCR product was precisely reproduced in the 1800 bp fragment. The gene sequence and the derived protein sequence for lupanine hydroxylase are shown in Figure 2. All the peptide sequences obtained from the purified protein are seen in the derived protein sequence of the *luh* gene product (underlined in Figure 2) and the M_r of the derived mature protein, 72256, is close to that of 72000 found by SDS/PAGE for the purified protein [3]. This supports the proposal that this gene does encode lupanine hydroxylase and this is further reinforced by the presence of a typical cytochrome *c* haem-binding site (CXXCH) and the many similarities to the sequences of other quinoproteins (see below).

Signal sequence

Lupanine hydroxylase is a periplasmic protein [4] and, as the region coding for the N-terminal sequence of the mature protein starts at base 78 from the start codon of the open reading frame, it is reasonable to propose that the first 26 amino acids represent the signal sequence needed for translocation of the enzyme. Further evidence for this came from analysis of the N-terminus by the SignalP V2.0 computer program [15]. This gave a signal peptide probability of 1.000 and predicted the signal peptidase cleavage site to be between amino acids 26 and 27 with a probability of 0.986. It appears to be a Sec signal sequence and, although some periplasmic redox proteins use an export pathway signalled by a double-arginine signal peptide, those binding redox cofactors such as iron porphyrins and PQQ are generally associated with standard Sec signal sequences [16].

Expression of *luh* gene in *E. coli*

Final evidence that the cloned gene codes for lupanine hydroxylase required heterologous expression to give an active enzyme. Initial attempts to express the gene in *E. coli* were unsuccessful for a number of possible reasons including incompatibility of the signal sequence and the requirement for haem insertion. We have addressed these problems by replacing the signal sequence with the 21 residue *E. coli* alkaline phosphatase signal sequence and by including plasmid pEC86 which expresses constitutively the genes for cytochrome *c* maturation protein factors [17]. The region of the *luh* gene encoding the mature protein was amplified by PCR using 5'-*Nsi*I/3'-*Eco*RI-adapted primers. The DNA fragment (≈ 2000 bp) was cleaved with *Nsi*I and *Eco*RI and ligated into *Pst*I- and *Eco*RI-cleaved plasmid pLiQ [18] to give the plasmid pEV-LH32. Plasmid pEV-LH32 was introduced into *E. coli* TB1 host harbouring pEC86. The organism was grown

for 6 h aerobically and then for 12 h under conditions of low aeration in medium containing limited phosphate (0.1 mM) to induce expression. As controls, the organism was grown with a phosphate concentration of 25 mM, enough to prevent induction, and *E. coli* containing only the pEC86 plasmid was also grown under both inducing and non-inducing conditions.

Cells were harvested by centrifugation and only those containing both plasmids grown under inducing conditions gave a pink pellet, indicative of the presence of cytochromes. SDS/PAGE of total protein showed that only these cells contained an extra haem-staining protein of the same M_r as lupanine hydroxylase (Figure 3). On breakage of the cells and centrifugation the pink protein was present as inclusion bodies in the same fraction as the cell membranes. Inactive enzyme was prepared by solubilizing and incubating this fraction in 8 M urea followed by rapid dilution of the soluble fraction. After further incubation of this preparation with PQQ and CaCl_2 , lupanine hydroxylase activity could be detected in the spectrophotometric assay, showing that holoenzyme had been produced. Both compounds were needed for enzyme activation and this is the first demonstration that, in common with other PQQ-containing enzymes, lupanine hydroxylase requires calcium ions. Further evidence that this expression system had achieved the incorporation of the haem group into lupanine hydroxylase came from the partial purification of enzyme from 12 l of culture by ion-exchange chromatography and gel filtration. The results of SDS/PAGE at each stage are included in Figure 3. The spectrum of the protein was that of a *c*-type cytochrome in the oxidized state (Figure 4), as expected for lupanine hydroxylase [3]. Addition of PQQ and Ca^{2+} gave only minor changes in the spectrum but when lupanine was also included the haem group was rapidly reduced (Figure 4), evidence that active lupanine hydroxylase had been produced. Thus the isolated gene has been shown to code for lupanine hydroxylase and a comparison of the derived protein sequence with those of other proteins can now be made.

Sequence comparison with other proteins

A BLAST search [19] for proteins with sequences similar to that of the derived amino acid sequence for lupanine hydroxylase revealed similarities to many other quinoproteins. The greatest similarity was with the quinohaemoprotein ethanol dehydrogenase of *C. testosteroni*, which had 36% identity and 54% positives (i.e. identical and conserved residues) shown over the full length of the gene. Other quinohaemoproteins including the type III membrane-bound alcohol dehydrogenases of *Acetobacter* and *Gluconobacter* also showed similarity to the deduced *luh* gene product.

Lupanine hydroxylase from the *Pseudomonas* sp. and alcohol dehydrogenase from *C. testosteroni* are both monomers with M_r values of about 70000, containing one molecule of PQQ and a single haem *c*. Both show sequence similarity to the α subunit of

```

ATG TCG GCT AAT AAG AAT ATA TGG ATC ATA AGG CTT GGC GTG GCA TTT GTT TGT GTC GCT ATC GGA GCT GCC CAG
M S A N K N I W I I R L G V A F V C V A I G A A Q
|2F primer → → → → → → → →|
GCA AAT GAA AAA GAC GGC TCT GCG GTG ACC AGC GGT AAT TGG AGC CTG CTA GGA GGC GGA AAT GAG CAG CAC TAT
A N E K D G S A V T S G N W S L L G G G N E Q H Y
→ →|
TTT AGT GCG CTA AAG GAT GTG AAT AAG TCA AAC GTC AAA AAC CTC GGT CTA AGC TGG TTT ACG GAT ATG GAG GCC
F S A L K D V N K S N V K N L G L S W F T D M E A

GGT GAT GGA TTG GTA GGA AAC CCG TTA GTT GCT GAT GGA GTG ATT TAT CAA GGT GGG CCC CCA GGA AAG ATT TAC
G D G L V G N P L V A D G V I Y Q G G P P G K I Y

GCT AAC GAC CTA AAG ACC GGT AAA AAT CTT TGG ACC TAT ACC CCA GAA GTG CAG TAC GAC AAA GAC ACA TCT TGG
A N D L K T G K N L W T Y T P E V Q Y D K D T S W

ACT GGC TTT TGG GGG ACG CAT GTT AAC CGG GGG CTA GCC GTA GAC GAC GAT AAT GTC TAT ATC GGC TCC TAC TGT
T G F W G T H V N R G L A V D D D N V Y I G S Y C

AAG CTT CTC GCT GTG AGC AGG ACA ACT CAC AAA CTG ACC TGG TCG TCC CAA TCG TGC GAT CCA AAA AAA ATG CAG
K L L A V S R T T H K L T W S S Q S C D P K K M Q

GCG ATT ACC GGC GCG CCT AGG GTC GGC GGT GGC AAG GTC TTC ATA GGT AAC GCC TCG GGT GAT TTC GGC GGC GAT
A I T C G A P R V G G G K V F I G N A S G D F G G G D

CGT GGA CAT CTA GAT GCG TTC GAC GCC AAA ACG GGT AAA CAT CTG TGG CGC TTC TAC ACT ATG CCA GGC GAT CCG
R G H L D A F D A K T G K H L W R F Y T M P G D P

TCC AAG CCT TTT GAA AAC GAT CTC TTG GCT AAG GCT TCA AAA ACA TGG GGT ACC GAT TAC TGG AAA TAT ACC AAA
S K P F E N D L L A K A S K T W G T D Y W K Y T K
|← ← ← 1R primer|

GGT GGT GTC AGC CCC TGG GAC GCC ATA ACT TAC GAT GAA GCA AGC GAC ACT CTC TAT TTT GGC ACT GAT GGG CCA
G G V S P W D A I T Y D E A S D T L Y F G T D G P

TCG CCC TGG TCA CCA GCT CAA CGG GCC CCC GAT GCG GGA GAT GAG TTG TTT TCG CAC TCG ATC ATT GCG GTA GAT
S P W S P A Q R A P D A G D E L F S H S I I A V D

GCA AGC ACG GGA GCT TAC AAG TGG CAT TTT CAG ACC GTG CAA AAT GAC GGG AGC AAT ATG AGT GCC ACT ATG CAC
A S T G A Y K W H F Q T V Q N D G S N M S A T M H

ATC ATG CTT GCT GAT TTG CCC GTT GAA GGA GTT TCA AAG CGT GTG GTC ATG ACC GCC CCG AAA AAT GGT TAT TTC
I M L A D L P V E G V S K R V V M T A P K N G Y F

TAT GTG CTT GAC GCG AGT ACT GGT AAA TTT ATC AGT GCG GAT CAT TAT GTT CCT GTA AAT TGG ACG AAG GGT TTA
Y V L D A S T G K F I S A D H Y V P V N W T K G L

GAT CCC AAA ACG GGG CGT CCG ATC CCT TCA AAC GAA GCG AAT TAT TGG GAG CGT CCT GGT GAG ATG ACC ATT CCT
D P K T G R P I P S N E A N Y W E R P G E M T I P

CTT CCT GGT GAT GTG GGT GGA CAT AAC TGG GAG GCG ATG GCT TAT AAT CCT GAA TTG CGT ACG GTT TAT ATT CCT
L P G D V G G H N W E A M A Y N P E L R T V Y I P

TCT ACC TTA GTT CCT GTT ACC GTG GTG GCT AGC AAG GAT ACT GGC GAA TTA GAT TTA GAT TAT TAC TAT GGG ATG
S T L V P V T V V A S K D T G E L D L D Y Y Y G M

CGG CCA GAC GCG ACC ATC AAG ACG CAA GGT GAC CTT GTT GCG TGG GAT CCA TTG CTG CAG AAA GAA AAG TGG CGC
R P D A T I K T Q G D L V A W D P L L Q K E K W R

GCT AAA CGA AGT TTG CCA GTA AAC GGG GGA GTT CTT GCC ACT GCT GGA GGA TTG GTA TTT CAG GGA ACT GGT GAC
A K R S L P V N G G V L A T A G G L V F Q G T G D

GGC CAC TTT GAG GCG TTT GAT GCT AAT ACT GGC GAA AAG CTT TGG TCG TTC CAT GTC GGT GGT TCC ATC TTG GCT
G H F E A F D A N T G E K L W S F H V G G S I L A

GCG CCG ACG ACT GTC GAA GTA GAT GGT GAT CAA TAT CTC ATC GTA GCG AGC GGC AAC GGT GGC GCA TCC GGT ATG
A P T T V E V D G D Q Y L I V A S G N G G A S G M

CGA GGT ATA CCT CGC CTC ATG AAC AAT TTG CAA AGT CAA GGT CCT GCT CGC TTG TTG GCG TTT CGA TTA GGA GGA
R G I P R L M N N L Q S Q G P A R L L A F R L G G

AAG ACG GAG TTG CCT ATC ACT TCA ACG CCA GAT TTT CCT AAA CCA CAG TAT CCG AAG CCC ACG AGC GCC ATG GCT
K T E L P I T S T P D F P K P Q Y P P K P A M A

GAG AGT GGA CGA CAT ATA TTT AAT GCC AAC GCG TGT GGT GCT TGC CAC GGT TTC AAT GCG GAG GGC TCC ACG CCA
E S G R H I F N A N A C G A C H G F N A E G S T P

GGA TTA CCT GAT CTG CGT CGT TCG GAC AAA TTA GAC CTA GCA GTA ATG AAA TCT ATT GTC ATT GAT GGT GCG TTT
G L P D L R R S D K L D L A V M K S I V I D G A F

AAG CCT CTT GGG ATG CCG GGG CAT CCT CAC ATC TCG GAT GCG GAT TTG CAG GCG CTA CAG GCG TTC ATT CTG CAG
K P L G M P G H P H I S D A D L Q A L Q A F I L Q

AAA GCG TGG ACC GCT TAT GAT ACT CAA CAA ACC CTA AAG ACT AGC GAT ACG GGC GCG CAA TAG
K A W T A Y D T Q Q T L K T S D T G A Q Stop
|← ← ← 2R primer|

```

Figure 2 Nucleotide sequence of the gene for lupanine hydroxylase and the derived protein sequence

The signal sequence at the N-terminal is shown in bold italics. Underlined sequences are those found by sequencing the purified protein. Double underlining is used to denote the putative haem-binding sequence. The nucleotide sequences corresponding to the primers used in PCR are also indicated.

Figure 3 Demonstration by SDS/PAGE of *luh* gene expression in *E. coli*

The total protein fraction from non-induced (lane 3) and induced (lane 4) *E. coli* containing both pEV-LH32 (carrying the *luh* gene) and pEC86 (carrying cytochrome *c* maturation genes) plasmids was examined by SDS/PAGE and stained for protein with Coomassie Blue (A) or for haem (B). As controls, protein fractions from non-induced (lane 1) and induced (lane 2) *E. coli* containing only the pEC86 plasmid were also examined. Lane 5 is crude *luh* gene product after partial purification by ion-exchange chromatography and lane 6 is after further purification by gel filtration. For comparison, lane 7 shows a sample of purified lupanine hydroxylase from the *Pseudomonas* sp. and lane 8 contains molecular-mass markers.

Figure 4 Absorption spectrum of partially purified heterologously produced enzyme

The UV/visible spectrum of the partially purified lupanine hydroxylase produced by expression of the *luh* gene in *E. coli* was recorded (dashed line). The enzyme was activated by addition of PQQ (5 μM) and CaCl₂ (33 μM) to the cuvette. Further addition of lupanine (10 μM) resulted in reduction of the haem (solid line).

the PQQ-containing methanol dehydrogenases. The methanol dehydrogenases have an $\alpha_2\beta_2$ tetrameric structure and bind one PQQ molecule per α subunit, but do not contain haem. As might be expected, their similarity with lupanine hydroxylase extends only over the first 550 amino acids, the putative PQQ-binding

domain, and not into the haem-binding domain of the C-terminus.

The X-ray structure of the α subunit of methanol dehydrogenase is unusual and consists of a superbarrel structure made up of eight antiparallel β sheets (W motifs) arranged like the blades of a propeller. These are held together by novel tryptophan-docking motifs of 11 amino acids [20]. Similar motifs are seen in other quinoproteins including the dimeric, non-haem ethanol dehydrogenase from *Pseudomonas aeruginosa* and the quino-haemoprotein from *C. testosteroni*. The crystal structures of these proteins show that their PQQ-binding domains have a tertiary structure similar to that of the α subunit of methanol dehydrogenase [7,21]. Alignment [22] of the lupanine hydroxylase sequence with those of ethanol dehydrogenase of *C. testosteroni* and methanol dehydrogenase of *M. extorquens* also shows eight W motifs (Figure 5), again suggesting a similar tertiary structure for the PQQ-binding domain in this enzyme.

Some of the invariant residues seen in PQQ enzymes and shown in methanol dehydrogenase of *M. extorquens* and the haemoquinoprotein alcohol dehydrogenase of *C. testosteroni* to be involved in the active site and interacting with PQQ are also present in lupanine hydroxylase. These include Arg-109 (109), Thr-152 (159) and Asn-383 (394) with the number in brackets referring to the residue position in the methanol dehydrogenase. Arg-331 of methanol dehydrogenase is replaced by Lys-320 as is also seen in the haem-containing ethanol dehydrogenase. Also present is the invariant residue Trp-230, which in the methanol dehydrogenase (Trp-243) and the ethanol dehydrogenase (Trp-245) forms the floor of the active-site chamber. In both of these enzymes the PQQ is sandwiched between this tryptophan and a novel disulphide ring structure formed from two adjacent cysteines (Cys-103 and Cys-104 in methanol dehydrogenase and Cys-116 and Cys-117 in the ethanol dehydrogenase) but a notable difference from the sequence of lupanine hydroxylase is that this structure is missing. Such a structure is conserved in many of the PQQ enzymes and, because of its rarity in proteins in general, it is reasonable to suggest that it has an important biological role

Luphy	-----NEKDG-----SAVTSGNMSLLGGNEQHYFSAIKDVNKS NVKNGLSMF	44
Edh-ct	TGPAAQAAAAVQRVDGDFIRANAARTPDMPITIGVDYAETRYSRDQINAANVKDLGLAMS	60
Mdh-mex	-----NDKLVELSKSDDNVMPGKNYDSNNFSDLKQINKGNVKQIRPANT	45
Luphy	TDMEAGDGLVGNFLVADGVIIQGGPPGK-IYANDLK-TGKNLWITYTFEVQYDKDTSWTGF	102
Edh-ct	YNLESTRGVEATEVVDGIMYVSASWSV-VHAIDTR-TGNRIWITYDF--QIDRSTGFKG-	115
Mdh-mex	FSTGLLNGHEGAPLVVDGKMYIHTSFNNFTFALGLDDPGTILMQDKF---KQNPAAARAVA	102
Luphy	WGTHVNRGLAVDDN-----VYIGSYC-KLLAVSRTTHKLTWSSQSCDPKK-MQAITGA	154
Edh-ct	CCDVVNRGVALWKGK-----VYVGAWDGRLLALDAATGKEVWHQNTFEGQKGSLLTITGA	169
Mdh-mex	CCDILVNRGLAYWPGDGKTPALILKTQLDGNVAALNAETGETVMKVENS-DIKVGSTLTIA	161
Luphy	PRVGGGVFIIGNASGDFGGRGHLDFFAKTGKHLWFFYTMPGDPSPKPFENDLLAKASKT	214
Edh-ct	PRVFKGVIIIGKRGAEYG-VRGYITAYDAETGERKWRWFSVPGDPSPKPFEDSMKRAART	228
Mdh-mex	PYVVKDKVIIIGSSGAELG-VRGYLTAYDVKTGEQVWRAYATGPDKDLLASDFNIKPNHY	220
Luphy	WG----TDYWK---YTKGGVSPDAITYDEASDTLIEGTGDPSPWSPAQRAPDAGDELFS	267
Edh-ct	WDP----SGKWW---EAGGGGTMDSMTFDAELNTMIVGTGNGSPWSHKVRSPKGGENLYL	282
Mdh-mex	GQKGLGTGTWEGDAWKIGGGTNWGWAYDPGTNLIIEGTGNPAWNETMRP---GDNKWT	277
Luphy	HSIIAVFASSTGAYVWHFCTVQNGS NMSATMHIMADLP-VEGVSKRVVMTAPKNGYFYV	326
Edh-ct	ASIVALLPDTGKYVWHYCEETPGENWDYTSTQPMILADIK-IAGKPRKVLHAPKNGFFV	341
Mdh-mex	MTIFGRADTGEAKFGYCKTPHDEWDYAGVNVMMISEQKDKDKARKLLTHPDRNGIVYT	337
Luphy	LDASTGKFISADHYVP-VNWTGKLSPRTGRIIPSNEANYWERPGEMTIPLFGDVGSHNWE	385
Edh-ct	LDRTNKGFISAKNFVP-VNWSGYE-MHGKPIGIAAARDGSKPD---AVPGPYGAHNWH	396
Mdh-mex	LDRTDCAVLSANKLDDTVNVFKSVLNTGQPVDRDPEYG-TRMDHLAKDICPSAMGYHNQ	396
Luphy	AMAYNEELRTVYIPSTLVPVTVVASK-----DTGELDLDYYYGMRPDATIK	431
Edh-ct	PMSFNBQTGLVYLPQNVVPVNLMDKKWEFNQAGPGKPSGTGWNTAKFFNAEPYKS--K	454
Mdh-mex	HDSYDKKRELFFMGINHICMDWEPFMLPYRAG-----QFFVGSATLNMYPGPKGDRQNYE	450
Luphy	TQSDLVWDPDLLQKEKRAKRSLPVNGCVLATAGGLVFGGTGDCHEFAFDANTGKELWSE	491
Edh-ct	PFGRLLAWDPAQKAAMSVEHVSFWNGGTLTTAGNVVFGGTADGRLVAYHAATGKELWEA	514
Mdh-mex	GLSQIKRYNAITGDYKMEKMERFAVWGGTMTATAGDLVFGYGTLDGYLKARDSDTGDLWKE	510
Luphy	HVGGSILAAETTVEVDGDOYLIVASNGGASGMRI-----PRLMNNLQS	537
Edh-ct	PTGTGVVAAPSTYMVDGRQYVSVAVWGGVYGLAAR-----ATER	555
Mdh-mex	KIPSGAIGYEMTYTHKCTOYVAIYYVGGWPGVGLVFDLADPTAGLGAVGAFKKLANYT	570
Luphy	GPARLLAERLGGKTELP-ITSTPDFPKQYPKPTSAMAESGRHIFNANACGACHGFNAEG	596
Edh-ct	GPGTYTEVAVAGKARMEFVAQRTQQLLGVKYDPAKVEAGTMLYVAN-CVFCHGVPGVD	614
Mdh-mex	MGGGVVVFSLDGGKGPYDDPNVGEWKSAAK	599
Luphy	STPGLPDLRRSDKLDLAVMKSIVIDGAFKPLGMPGHP-HISDADLQALQAFILQKAWTAY	655
Edh-ct	RGGNIPNLGYMDASYIENLPNFVFKGPAMVRGMPDFTGKLSGDDVESLKAFIQGTADAIR	674
Luphy	DTQQLKTSDTGAQ	669
Edh-ct	PKP	677

Figure 5 Alignment of the amino acid sequence of lupanine hydroxylase (Luphy) with those of the quinohaemoprotein ethanol dehydrogenase from *C. testosteroni* (Edh-ct) and the quinoprotein methanol dehydrogenase from *M. extorquens* (Mdh-mex)

Residues that are common to all three proteins are shaded in black and identical residues in the lupanine hydroxylase and one of the other proteins are shaded grey. The double underlined sequences are tryptophan-docking motifs of the eight-bladed propeller structure found in methanol dehydrogenase. The boxed sequences are the putative haem-binding regions in the cytochrome *c* domains of the two haemoproteins.

in the quinoproteins. However, there are other PQQ enzymes that also lack this disulphide ring, including the soluble form of glucose dehydrogenase from *Acinetobacter calcoaceticus* [23,24] and the membrane-bound form of glucose dehydrogenase from *E. coli* [25]. In the latter enzyme a histidine replaces the disulphide ring in maintaining the position of the PQQ in the active site and whether a similar replacement also occurs in lupanine hydroxylase must await further structural studies.

Examination of the C-terminal sequence of about 150 amino acids showed the presence of a putative cytochrome *c* haem-

binding sequence (CXXCH) of CGACH from residue 586 to 590 (Figures 2 and 5). This aligns with a similar sequence in the *C. testosteroni* ethanol dehydrogenase for which it has been shown that the haem is covalently bound to the cysteines with the histidine providing one of the ligands to the haem iron. The sixth ligand to the iron is provided by Met-647, which is also conserved in the lupanine hydroxylase sequence. A BLAST search of the databases using the C-terminal domain (residues 551–659) did not show any close matches but there were similarities to the C-terminal domains of other quinohaemoproteins, including the

alcohol dehydrogenases from *Acetobacter* and *Gluconobacter* (26% and 30% identities respectively). There were also some similarities with *c*-type cytochromes, with *cyt*₅₅₃ of *Cyanidium caldarium* and the *cyt*₅₅₂ precursor of *Nitrosomonas europaea* appearing the most similar (32% and 30% identities respectively). In a similar analysis of the sequence of the quinohaemoprotein from *C. testosteroni*, Stoorvogel et al. [6] suggested that the gene might have arisen by the fusion of genes for a quinoprotein alcohol dehydrogenase and a cytochrome *c*. This interesting proposal might also be applicable to the lupanine hydroxylase gene.

Lupanine is a larger molecule than many of the substrates for quinoprotein enzymes. Moreover, it is a tertiary amine that is dehydrogenated rather than the alcohol group used by most enzymes of this class. Despite this, analysis of the gene shows that lupanine hydroxylase has many similarities with other quinoproteins and suggests possible common overall structural features. However, there could be detailed differences around the active site, as illustrated by the lack of the disulphide bridge from adjacent cysteines. Further structural studies may reveal other differences. Successful expression of the *luh* gene in *E. coli* and development of a procedure for reactivation of the protein will now allow the production of sufficient enzyme for further structural and other studies without restrictions due to insufficient lupanine supply.

We thank B. Dunbar and B. Madden for help with protein sequencing. We are also grateful to Alice Taylor and Andrew Gane for assisting us with DNA sequencing and to Professor L. Thony-Meyer for the gift of the pEC86 plasmid. This work was supported by a grant from the Biotechnology and Biological Sciences Research Council.

REFERENCES

- Toczko, M., Brzeski, W. and Kakolewska-Baniuk, A. (1963) Microbial degradation of lupanine. V. Identification of 17-hydroxylupanine. *Bull. Acad. Pol. Sci., Cl. II*, **11**, 161–164
- Rogozinski, J. (1975) Molecular properties of the inducible lupanine hydroxylase from growing cultures of *Pseudomonas lupanine*. *Acta Biochim. Pol.* **32**, 57–66
- Hopper, D. J., Rogozinski, J. and Toczko, M. (1991) Lupanine hydroxylase, a quinocytochrome *c* from an alkaloid-degrading *Pseudomonas* sp. *Biochem. J.* **279**, 105–109
- Hopper, D. J. and Rogozinski, J. (1998) Redox potential of the haem *c* group in the quinocytochrome, lupanine hydroxylase, an enzyme located in the periplasm of a *Pseudomonas* sp. *Biochim. Biophys. Acta* **1383**, 160–164
- de Jong, G. A. H., Caldeira, J., Sun, J., Jongejan, J. A., de Vries, S., Loehr, T. M., Moura, I. and Duine, J. A. (1995) Characterization of the interaction between PQQ and heme *c* in the quinohaemoprotein ethanol dehydrogenase from *Comamonas testosteroni*. *Biochemistry* **34**, 9451–9458
- Stoorvogel, J., Kraayveld, D. E., van Sluis, C. A., Jongejan, J. A., de Vries, S. and Duine, J. A. (1996) Characterization of the gene encoding quinohaemoprotein ethanol dehydrogenase of *Comamonas testosteroni*. *Eur. J. Biochem.* **235**, 690–698
- Oubrie, A., Rozeboom, H. J., Kalk, K. H., Huizinga, E. G. and Dijkstra, B. W. (2002) Crystal structure of quinohaemoprotein alcohol dehydrogenase from *Comamonas testosteroni*. *J. Biol. Chem.* **277**, 3727–3732
- Ghosh, M., Anthony, C., Harlos, K., Goodwin, M. G. and Blake, C. C. F. (1995) The refined structure of the quinoprotein methanol dehydrogenase from *Methylobacterium extorquens* at 1.94 Å. *Structure* **3**, 177–187
- Xia, Z., Dai, W., Zhang, Y., White, S. A., Boyd, G. D. and Mathews, F. S. (1996) Determination of the gene sequence and the three-dimensional structure at a 2.4 Å resolution of methanol dehydrogenase from *Methylophilus W3A1*. *J. Mol. Biol.* **259**, 480–501
- Wilson, K. (1993) Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*, vol. 1 (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 1–5, Greene Publishing Associates & Wiley, New York
- He, M. Y., Wilde, A. and Kaderbhai, M. A. (1990) A simple single-step procedure for small-scale preparation of *Escherichia coli* plasmids. *Nucleic Acids Res.* **8**, 1660–1661
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning, a Laboratory Manual*, 2nd edn, Cold Spring Harbor Press, Cold Spring Harbor, NY
- Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580
- Chambers, S. P., Prior, S. E., Barstow, D. A. and Minton, N. P. (1988) The PMTL NIC-cloning vectors.1. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **68**, 139–149
- Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6
- Berks, B. C. (1996) A common export pathway for proteins binding complex redox cofactors. *Mol. Microbiol.* **22**, 393–404
- Thony-Meyer, L. (1997) Biogenesis of respiratory cytochromes in bacteria. *Microbiol. Mol. Biol. Rev.* **61**, 337–376
- Kaderbhai, M. A., Ugochukwu, C. C., Kelly, S. L. and Lamb, D. C. (2001) Export of cytochrome P450 105D1 to the periplasm of *Escherichia coli*. *Appl. Environ. Microbiol.* **67**, 2136–2138
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res.* **25**, 3389–3402
- Anthony, C. and Ghosh, M. (1998) The structure and function of the PQQ-containing quinoprotein dehydrogenases. *Progr. Biophys. Mol. Biol.* **69**, 1–21
- Keitel, T., Diehl, A., Knaute, T., Stezowski, J. J., Höhne, W. and Görsch, H. (2000) X-ray structure of the quinoprotein ethanol dehydrogenase from *Pseudomonas aeruginosa*: basis of substrate specificity. *J. Mol. Biol.* **297**, 961–974
- Morgenstern, B. (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**, 211–218
- Oubrie, A., Rozeboom, H. J., Kalk, K. H., Duine, J. A. and Dijkstra, B. W. (1999) The 1.7 Å crystal structure of the apo form of the quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus* reveals a novel internal conserved sequence repeat. *J. Mol. Biol.* **289**, 319–333
- Oubrie, A. and Dijkstra, B. W. (2000) Structural requirements of pyrroloquinoline quinone dependent enzymatic reactions. *Protein Sci.* **97**, 1265–1273
- Cozier, G. E. and Anthony, C. (1995) Structure of the quinoprotein glucose dehydrogenase of *Escherichia coli* modelled on that of methanol dehydrogenase from *Methylobacterium extorquens*. *Biochem. J.* **312**, 679–685

Received 8 May 2002/11 July 2002; accepted 16 July 2002

Published as BJ Immediate Publication 16 July 2002, DOI 10.1042/BJ20020729