Gene dissection demonstrates that the *Escherichia coli* *cysG* gene encodes a multifunctional protein

Martin J. WARREN,*†, Edward L. BOLT,* Charles A. ROESSNER,† A. Ian SCOTT,† Jonathan B. SPENCER† and Sarah C. WOODCOCK*

*School of Biological Sciences, Queen Mary and Westfield College, Mile End Road, London E1 4NS, U.K.

and †Centre for Biological NMR, Department of Chemistry, Texas A & M University, College Station, TX 77843, U.S.A.

The C-terminus of the *Escherichia coli* CysG protein, consisting of amino acids 202–457, was expressed as a recombinant protein using gene dissection methodology. Analysis of the activity of this truncated protein, termed CysG^A, revealed that it was able to methylate uroporphyrinogen III in the same *S*-adenosyl-L-methionine (SAM)-dependent manner as the complete CysG protein. However, this truncated protein was not able to complement *E. coli* *cysG* cells, thereby suggesting that the first 201 amino acids of the CysG protein had an enzymic activity associated with the conversion of dihydrosirohydrochlorin into sirohaem. Analysis of the N-terminus of the CysG protein revealed the presence of a putative pyridine dinucleotide binding site. When the purified

CysG protein was incubated with NADP⁺, uroporphyrinogen III and SAM the enzyme was found to catalyse a coenzyme-mediated dehydrogenation to form sirohydrochlorin. The CysG^A protein on the other hand showed no such coenzyme-dependent activity. Analysis of the porphyrinoid material isolated from strains harbouring plasmids containing the complete and truncated *cysG* genes suggested that the CysG protein was also involved in ferrochelation. The evidence presented in this paper suggests that the CysG protein is a multifunctional protein involved in SAM-dependent methylation, pyridine dinucleotide dependent dehydrogenation and ferrochelation.

INTRODUCTION

Sirohaem, the prosthetic group found in the enzymes sulphite and nitrite reductases, is a modified tetrapyrrole and thus belongs to the same family of prosthetic groups as haem, chlorophyll and vitamin B₁₂ (Warren and Scott, 1990). These chromophoric compounds differ in their patterns of peripherally modified side chains, methylation, oxidation state and centrally chelated metal ion. An understanding of the basic route to the construction of these compounds will give not only a perception of the control and evolution of the biosynthetic pathways involved but also an insight into the intricate chemistry used to make these complex biological products.

All tetrapyrroles are synthesized along a common pathway up to the assembly of the first macrocyclic intermediate, uroporphyrinogen III (Warren and Scott, 1990; Jordan 1991). At this point there is a divergence in the biosynthetic routes. Modification of the side chains of uroporphyrinogen III by decarboxylation directs the tetrapyrrole along the route to haem and chlorophyll construction, whereas methylation of the macrocycle directs the molecule towards the synthesis of sirohaem and vitamin B₁₂ (Figure 1).

Uroporphyrinogen methylase is the enzyme responsible for the two *S*-adenosyl-L-methionine SAM-dependent methylations of uroporphyrinogen III at positions 2 and 7, which yields dihydrosirohydrochlorin, a key intermediate in the biosynthesis of both sirohaem and vitamin B₁₂ (Figure 1). The gene encoding the enzyme in *Escherichia coli* has been isolated and sequenced (Peakman et al., 1990) and has been used in expression systems to produce large quantities of the enzyme (Warren et al., 1990a). Mutations and deletions in this gene produce *E. coli* cysteine auxotrophs due to the involvement of sirohaem in sulphur metabolism with the enzyme sulphite reductase (Cole et al.,

1980). The gene has been termed *cysG* and has been shown to contain 1374 bases encoding 457 amino acids, which gives the CysG protein a predicted *M_r* of 49928 (Peakman et al., 1990). Experiments performed on the isolated purified protein have demonstrated that the enzyme converts uroporphyrinogen III into dihydrosirohydrochlorin (precorrin-2) in a reaction that requires SAM as the methyl donor (Warren et al., 1990b). If excess CysG protein is added to the incubation then the enzyme performs a third, SAM-dependent, methylation at position 12 of the macrocycle, producing a trimethylpyrrocorphin which has no known physiological role. In *E. coli*, dihydrosirohydrochlorin is destined solely for the biosynthesis of sirohaem, as *E. coli*, unlike *Salmonella*, is unable to synthesize vitamin B₁₂ *de novo*. To complete the construction of sirohaem, dihydrosirohydrochlorin has to be oxidized to produce sirohydrochlorin, which then has to be chelated with iron to produce the final product. No enzyme has been identified from *E. coli* or any other organism which is capable of performing these latter two functions.

Uroporphyrinogen methylase genes have also been isolated and sequenced from *Pseudomonas denitrificans* (*cobA* gene; Crouzet et al., 1990a), *Bacillus megaterium* (*cobA* gene; Robin et al., 1991), *Methanobacterium ivanovii* (*corA* gene; Blanche et al., 1991) and *Anacystis nidulans* R2 (*cobA* gene; Jones et al., 1993). The *corA* and *cobA* genes encode for proteins which have *M_r*s of between 26000 and 28000 and are therefore approximately half the size of the *E. coli* CysG protein. These uroporphyrinogen methylases show a strong sequence similarity with the C-terminus of the CysG protein. An analysis of the activity of the CorA and CobA proteins demonstrates that although they are able to convert uroporphyrinogen III into dihydrosirohydrochlorin they are unable to complement *E. coli* *cysG* auxotrophs. In this paper, the C-terminus of the CysG protein is shown to be responsible for the methylation of uroporphyrinogen III. Furthermore, a

Figure 1 Biosynthesis of some important tetrapyrrolic derived prosthetic groups from uroporphyrinogen III

Uroporphyrinogen III represents one of the most important branch points in the pathway. Thus, decarboxylation of the acetate side chains to methyl groups directs the intermediate towards haem and chlorophyll production. Methylation of uroporphyrinogen III at positions 2 and 7 forms dihydrosirohydrochlorin in a reaction that requires two molecules of SAM. The *E. coli* CysG protein will actually perform a third methylation at position 12 to yield the pyrrocorphin structure shown. This trimethylpyrrocorphin has no known physiological role. Methylation of dihydrosirohydrochlorin at position C-20 commits the intermediate to the biosynthesis of vitamin B₁₂. Oxidation of dihydrosirohydrochlorin yields sirohydrochlorin which is then chelated by iron to give sirohaem. Evidence is presented in this paper which demonstrates that these latter two processes are also performed by the *E. coli* CysG protein and that the oxidation process is an NAD⁺/NADP⁺-dependent reaction.

pyridine dinucleotide-dependent dehydrogenation that converts dihydrosirohydrochlorin into sirohydrochlorin, and also a ferro-chelation activity that converts sirohydrochlorin into sirohaem, are found to be associated with the N-terminus of the protein.

MATERIALS AND METHODS

Chemicals

Most chemicals, reagents and antibiotics were obtained from the Sigma Chemical Company. Sephracryl S-300, DEAE-Sephacel, restriction enzymes and sequencing kits were obtained from Pharmacia Fine Chemicals. Tryptone, yeast and agar were obtained from Difco Laboratories. Mimetic Green-1 affinity chromatography resin was purchased from Affinity Chromatography Ltd., Cambridge, U.K. A GeneAmp kit was purchased from Perkin-Elmer Cetus and was used for all PCR procedures.

Molecular biology techniques and the growth of organisms were carried out as described by Sambrook et al. (1989).

Plasmids and bacteria

All strains and plasmids used in this work are described in Table 1. *E. coli* strain CR252 was used as a source for the *cysG* gene and for the purification of the CysG protein (Warren et al., 1990a). *E. coli* strain 302Δa, *cysG*, was a gift from Dr. J. Cole, University of Birmingham, U.K. This strain has a deletion in the *cysG* gene and demonstrates no CysG activity. Dissection of the *cysG* gene to clone nucleotides 604–1374 was performed using a PCR method described by MacFerrin et al. (1990). Thus nucleotides 604–1374 were amplified from pCR252 using universal primer and a primer specific for the new N-terminus of the truncated protein. The N-terminus primer contained the sequence shown in Figure 2. The fragment was amplified using a program-

Table 1 Table of strains and plasmids used in this investigation

Strain or plasmid	Description/genotype/phenotype/properties	Reference
<i>E. coli</i> strains		
TB1	JM83 <i>hsdR</i> ($r_k^- m_k^-$)	Focus (1984)
302Δa	Nir ^s Lac ⁺ Cys ⁻ ; <i>cysG</i> ⁻	Griffiths and Cole (1987)
CR252	TB1 containing plasmid pCR252. Overexpresses CysG	Warren et al. (1990)
Plasmids		
pCR252	<i>cysG</i> gene cloned into pUC19. Expression regulated from lac promoter	Warren et al. (1990)
pEB1	<i>cysG</i> ^A gene cloned into pUC19. Expression regulated from lac promoter	This paper
pEB2	<i>cysG</i> ^B gene cloned into pUC19	This paper
pEB1-2	<i>cysG</i> ^A and <i>cysG</i> ^B cloned into pUC19	This paper
pSCW1	G224A mutation in CysG. A site directed mutation in the <i>cysG</i> gene changing glycine at position 224 to alanine. Cloned into pUC19	S. C. Woodcock and M. J. Warren, unpublished work
pSCW2	<i>cysG</i> ^A and G224A <i>cysG</i> in pUC19	S. C. Woodcock and M. J. Warren, unpublished work

mable thermal cycler in a 0.5 ml Microfuge tube using the GeneAmp kit core reagents according to the manufacturer's instructions, with the following sequence: 1 min at 94 °C, 1 min at 45 °C and 2 min at 72 °C. This was repeated for a total of 30 cycles.

The amplified fragment was purified on a 1% agarose gel, extracted and cleaved with the restriction enzymes *Bam*HI and *Eco*RI and ligated into pUC19 which had been cut with the same restriction enzymes. This gave rise to plasmid pEB1. Expression of the C-terminus of the CysG protein was performed when this plasmid was transformed into *E. coli* TB1.

Cloning of the first 603 nucleotides was performed using the same technique except that the N-terminal primer contained nucleotides 1–30 together with a *Hind*III restriction site and a C-terminal primer was used containing a stop signal immediately after nucleotide 603 together with a *Bam*HI restriction site. The primers were based on the same design as the one shown above. The amplified product in this case was ligated into the *Hind*III–*Bam*HI site of pUC19 after appropriate digestion and gave rise to plasmid pEB2.

Complementation experiments

Complementation experiments were performed using the *cysG* auxotrophic mutant 302Δa. The auxotrophic strain was grown on minimal medium containing 2% glucose. Control plates contained cysteine at a concentration of 100 µg/ml.

Enzyme purification

Porphobilinogen deaminase and uroporphyrinogen III synthase were isolated as previously described (Warren et al., 1992). The

complete uroporphyrinogen methylase (CysG) was purified by a modification of the method described by Warren et al. (1990a) from strain CR252, except that the final stage on the f.p.l.c. MonoQ column was replaced by a Mimetic Green affinity column. Thus protein obtained after DEAE chromatography was applied to a column (2.5 cm × 10 cm) of Mimetic Green. The column was washed with 0.1 M phosphate buffer, pH 8.0, and the protein eluted using a 0–1.5 M NaCl gradient (total volume = 1 l). The C-terminus of the CysG protein was purified from *E. coli* TB1/pEB1 according to the following procedure. A 2 litre sample of the strain was grown overnight and harvested by centrifugation. The bacterial pellet was resuspended in 25 ml of 0.1 M potassium phosphate buffer, pH 8.0. The bacterial suspension was then sonicated for 1 min 4 times with 4 min cooling in a Soniprep ultrasonicator. The sonicated extract was centrifuged at 20000 g for 15 min to remove the cell debris, and the supernatant was made 30% (w/v) with respect to ammonium sulphate. The solution was then centrifuged again and the pellet collected. The pellet was resuspended in 10 ml of 0.1 M potassium phosphate buffer and was then dialysed against 5 l of the same buffer overnight at 4 °C. After this time the dialysed solution was applied to a column (2.5 cm × 15 cm) of DEAE-Sephacel. The column was washed with 20 ml of 0.1 M phosphate buffer, pH 8.0, and the truncated protein was eluted from the column by application of a 0–200 mM NaCl gradient. The truncated protein was detected by SDS/PAGE and fractions containing it were combined, concentrated and stored at –20 °C.

Enzyme assays

Enzyme assays were performed in a final volume of 10 ml of thoroughly degassed 0.1 M Tris buffer, pH 7.5, containing 1 mg of porphobilinogen deaminase, 20 µg of uroporphyrinogen III synthase and 0.1 µmol of porphobilinogen (PBG). This was allowed to incubate at 30 °C for 10 min in order to generate uroporphyrinogen III at a concentration of about 2 µM. At this point, 1–5 mg of the purified CysG protein or the truncated version of the protein was included and the reaction was started by the addition of SAM to a final concentration of 100 µM. If the enzyme was being tested for dehydrogenase activity then either NADP⁺ or NAD⁺ was added to a final concentration of 50 µM. The reaction was monitored in a Hewlett Packard photodiode array spectrophotometer. Alternatively, the reaction was stopped after 30 min by the application of the reaction mixture to a column (1 cm × 3 cm) of DEAE-Sephacel. After washing the column with 15 ml of distilled water the DEAE material was lyophilized and the bound tetrapyrroles were esterified in 95% methanol/sulphuric acid and analysed as described below.

Preparation of tetrapyrrole methyl esters and analysis of porphyrinoid material

Uroporphyrin, sirohydrochlorin, factor I and the 2,7,12-trimethylpyrrocorphin were detected as their octamethyl esters by the methods previously described (Warren et al., 1990b). The tetrapyrroles were separated by reverse-phase h.p.l.c. using an

5'CGCGCGGATCCAGGAGGAATTTAAATGGAAACGACCGAACAGTTAATCAACGAACCG3'

*Bam*HI Ribosome Start Nucleotides 604–633
binding site codon

Figure 2 Sequence of N-terminus primer used to amplify and clone the truncated *cysG*

85% (v/v) methanol/water solvent on a C_{18} reverse-phase column. The porphyrinoid components were monitored at 405 and 376 nm by a Hewlett Packard photodiode array spectrophotometer that had been adapted with a flow cell.

RESULTS AND DISCUSSION

Primary structure comparisons between the *E. coli* CysG protein and the CobA proteins from *P. denitrificans*, *B. megaterium* and *A. nidulans* as well as the CorA protein from *M. ivanovii* show a high degree of similarity, with an overall identity of about 40% (Crouzet et al., 1990a; Blanche et al., 1991; Warren et al., 1994). This similarity is restricted to the C-terminus of the CysG protein, between amino acids 202 and 457. In this respect the CysG protein can be viewed in terms of two distinct domains (Figure 3). The C-terminal portion of the protein from amino acids 202 to 457 has been termed CysG^A, as it shares the similarity with the CobA and CorA proteins. The N-terminus of the protein, consisting of the first 201 amino acids, is referred to as the CysG^B part of the protein.

To investigate the apparent bi-domain structure of the CysG protein it was decided to try and separate the two major domains of the CysG protein and to investigate their individual functions.

Figure 3 Comparison of the *E. coli* CysG protein with the CobA protein from *P. denitrificans*

The figure depicts the two major domains of the CysG protein. The CysG^A domain has a high degree of similarity with the CobA protein and contains the uroporphyrinogen methylase activity. The CysG^B domain shows no significant similarity with any other primary structure and contains the dehydrogenase/ferrochelatase activity.

Figure 4 Polyacrylamide gel of whole-cell lysates of *E. coli* TB1 harbouring pUC19, pCR252 and pEB1 and purified CysG protein as well as the purified CysG^A protein together with M_r mass markers

Lane 1, cell lysate of TB1/pUC19; lane 2, cell lysate of the strain harbouring the *cysG^A* gene (TB1/pEB1); lane 3, cell lysate of the strain harbouring the complete *cysG* gene (TB1/pCR252); lane 4, M_r standards; lane 5, purified CysG^A protein; lane 6, purified CysG protein; lane 7, M_r standards.

In this respect, the protein was dissected into the two putative domains, CysG^A and CysG^B, by genetic manipulation. The C-terminus of the CysG protein, CysG^A, was successfully expressed as a truncated protein by employing the technique of gene dissection described by MacFerrin et al. (1990). This technique relies on the use of appropriately designed primers to the part of the gene that is to be amplified. It was possible to use universal primer together with a specific primer which corresponded to nucleotides 604–633 and which contained a new ATG codon, an optimal ribosome binding site and a unique *Bam*HI restriction site. Thus amplification with these primers gives rise to an expression cassette which was directly cloned into pUC19, yielding plasmid pEB1. In this case the *cysG^A* gene is under the control of the lac promoter. Transformation of TB1 with this plasmid yielded highly fluorescent bacteria which could be detected by illumination with a u.v. light source. Analysis of the protein profile of this strain by SDS/PAGE clearly showed a new protein band with an M_r of around 29000, which corresponds to that expected for the CysG^A protein (Figure 4). Plasmid preparations from this strain showed that the plasmid contained the correctly sized insert, and sequencing of the plasmid revealed the expected sequence of the amplified fragment. It was therefore concluded that the latter portion of the *cysG* gene had not only been amplified but that it had been successfully cloned into pUC19. Furthermore, the presence of an abundant protein band at M_r 29000 and the accumulation of porphyrinoid material within the cell suggested that the truncated gene was expressing an active C-terminal truncated CysG protein, CysG^A, which was affecting flux through the tetrapyrrole biosynthetic pathway.

The first 603 nucleotides of the *cysG* gene, which encode the CysG^B domain, were also dissected and cloned using the same procedure, giving rise to plasmid pEB2. Plasmid preparations and restriction analysis of pEB2 indicated that the gene fragment had been cloned correctly. However, there was no sign of protein over-expression when pEB2 was transformed into TB1 cells. The reason for the lack of over-expression is unclear.

Purification of truncated methylase, the CysG^A protein

The C-terminal truncated CysG protein, CysG^A, was purified using standard protein purification techniques described in the Materials and methods section. The migration of the purified protein corresponded to an M_r of 29000 (predicted gene derived M_r = 27992) co-migrating with the protein standard carbonic anhydrase (Figure 4). One litre of TB1 culture, harbouring the plasmid pEB1, yielded approximately 20 mg of purified truncated protein. When analysed by gel-filtration chromatography to determine the native molecular size of the protein, the enzyme eluted from the column at a position corresponding to an M_r of 60000. This suggests that the protein exists as a homodimer in its native state and in this respect it is similar to the other uroporphyrinogen methylases which also exist as homodimers. Interestingly, when analysed under the same conditions on the gel-filtration column, the complete CysG protein eluted at a point corresponding to an M_r of 100000. The complete CysG protein, which has a subunit M_r of 50000, would also appear to be a homodimer and not a monomer as previously reported (Warren et al., 1990a) (results not shown).

Analysis of the extracts from strains producing the CysG^A protein

The pigment that accumulated during growth of the *E. coli* strain expressing the CysG^A protein, and which caused the fluorescence, was analysed. The pigment was analysed from *E. coli* strains TB1 and 302Δa after both had been transformed with the plasmid pEB1. After extraction and esterification, the pigments were

Figure 5 N.m.r. spectra of dihydrosirohydrochlorin and 2,7,12-trimethylpyrrocorphin

(a) N.m.r. spectrum of dihydrosirohydrochlorin derived from $[3,5-^{13}\text{C}_2]\text{PBG}$. Incubation of the labelled PBG and SAM with the enzyme cocktail described in the Materials and methods section together with the CysG^A protein produced dihydrosirohydrochlorin with the spectrum shown. (b) N.m.r. spectrum of 2,7,12-trimethylpyrrocorphin synthesized from $[3,5-^{13}\text{C}_2]\text{PBG}$ by prolonged incubation with the CysG^A protein.

passed down a reverse-phase C_{18} column and the components analysed in terms of their retention time and u.v./visible spectra. In both cases the extracts were found to contain a mixture of sirohydrochlorin and overmethylated trimethylpyrrocorphin. This result indicates that, *in vivo*, the truncated protein performs the same methylation reaction as the complete CysG protein *in vitro*. However, unlike the *E. coli* strain harbouring the complete *cysG* gene, whose cellular content is dark green in appearance due to the overproduction of sirohaem, the C-terminal truncated protein does not accumulate sirohaem. In fact, the truncated protein does not make even trace amounts of sirohaem. This was concluded from complementation experiments with an *E. coli cysG* strain. Thus, whereas the complete *cysG* gene is able to complement the cysteine auxotroph 302Δa when it is grown on minimal medium, the C-terminal truncated gene, *cysG*^A, does not.

Taken together, the analysis of the porphyrinoid material within the strain harbouring pEB1 and the fact that this plasmid

is unable to complement the *cysG* strain, these results imply that the CysG^A protein, although capable of performing the methylation reaction, is deficient in ferrochelatase activity. The strain harbouring the CysG^A plasmid is able to accumulate sirohydrochlorin but not sirohaem. The ferrochelatase activity present in the complete *cysG* gene must therefore be mediated from the first 201 amino acids of the N-terminus. Whether the truncated protein actually makes dihydrosirohydrochlorin or sirohydrochlorin cannot be addressed from these results as dihydrosirohydrochlorin is very unstable and rapidly oxidizes to sirohydrochlorin. Therefore, although it can be deduced that the truncated protein is deficient in chelatase activity it cannot be concluded whether it is deficient in oxidase activity.

Enzymic activity of the purified CysG^A protein, the C-terminal truncated CysG

The enzymic activity of the complete CysG protein has been

Binding site	Sequence
Consensus NADP ⁺	X X X X G X G X X A X X X A X X X X X G X X X X X X
Consensus NAD ⁺	X X X X G X G X X G X X X G X X X X X G X X X X X X
Consensus FAD	X X X X G X G X X G X X X A X X X X X G X X X X X X
CysG sequence (14–42)	D C L I V G G G D V A E P K A R L L L D A G A R L T V N
HemY sequence (10–38) Predicted	H V V I I G G G I T G L A A A F Y M E K E I K E K N L P
Secondary structure	T β β β β β T T T · α α α α α α α α α α β β β β β

Figure 6 Alignment of consensus dinucleotide binding sites with the putative dinucleotide binding sites from the CysG protein and the HemY protein

Letters in bold signify hydrophobic residues.

Figure 7 Spectra obtained from incubations of CysG (continuous line) and CysG^A (dotted line) with uroporphyrinogen III, SAM and NADP⁺

The CysG incubation produces a spectrum characteristic of sirohydrochlorin whereas the CysG^A incubation spectrum remains as dihydrosirohydrochlorin.

studied in the past by incubating the enzyme with uroporphyrinogen III under anaerobic conditions with SAM. In these experiments it was found that the enzyme catalysed the formation of dihydrosirohydrochlorin by methylation of uroporphyrinogen III at position 2 and 7. At higher enzyme concentrations the enzyme performed a third methylation at position 12 to give the overmethylated trimethylpyrrocorphin (Figure 1). The reaction of the CysG^A protein was investigated under similar conditions. Analysis of the products of this incubation, after oxidative esterification and extraction, revealed that at lower enzyme concentrations of the truncated protein (< 0.1 mg/ml) the extract contained a high relative proportion of factor I, the monomethylated product derived from methylation at position 2 of uroporphyrinogen III. Some sirohydrochlorin was also detected but in much lower amounts than was formed under similar conditions with the intact CysG protein. Using higher enzyme concentrations the yield of dimethylated material could be increased and the reaction turned the characteristic yellow colour associated with the formation of the dipyrrocorphin structure of dihydrosirohydrochlorin. Confirmation of this reaction was obtained by monitoring the products of the reaction by using ¹³C n.m.r. spectroscopy. Thus incubation with uroporphyrinogen III derived from [3,5-¹³C₂]PBG gave a spectrum identical to that obtained previously with the CysG protein,

which showed two sp³ carbons and six sp² carbons (Figure 5a). If this high concentration of enzyme was left for a prolonged period of time (over 2 h) in the incubation, a red colour developed. This red colour development was first noticed with the CysG protein and was shown to be due to overmethylation of the dihydrosirohydrochlorin by addition of a third SAM-derived methyl group at position 12. Analysis of the red pigment associated with the truncated protein by n.m.r. revealed that it too was due to an extra methylation at position C-12 of the substrate template with the spectrum indicating the presence of three sp³ carbons and five sp² carbons (Figure 5b). These results demonstrate that the truncated protein is capable of performing the same SAM-dependent methylation reactions as the complete CysG protein *in vitro* as well as *in vivo*. The major difference between the CysG protein and the CysG^A protein is in the amount of monomethylated material that is formed. In this respect the truncated protein is similar to the CobA protein of *P. denitrificans*, as this uroporphyrinogen methylase also makes a high proportion of monomethylated material (Blanche et al., 1989). There is little doubt that the C-terminus of the CysG protein is responsible for the methylation of uroporphyrinogen III. The methylation domain of the protein therefore consists of amino acids 202–457. This leaves the question of the function of the first 201 amino acids to be addressed.

The N-terminus of the CysG protein

Although the C-terminus of cysG protein revealed similarity to other uroporphyrinogen methylases there was no strong similarity to the first 201 N-terminal amino acid sequence of the protein to any structures in the databases. However, a closer examination of the N-terminal sequence for putative nucleotide or coenzyme binding sites, similar to that carried out on the C-terminal portion of the protein which identified the SAM binding site (Crouzet et al., 1990b; Haydock et al., 1991; Warren et al., 1994), revealed a sequence which constituted a putative dinucleotide binding fold. The consensus binding sites for NAD⁺, NADP⁺ and FAD together with the putative CysG dinucleotide binding site are shown in Figure 6 along with the HemY protein sequence from *B. subtilis*. The significance of the HemY sequence is dealt with in a later section. This dinucleotide binding site is located close to the N-terminus of the protein and shows strong similarity to the dinucleotide binding sites described by Hanukoglu and Gutfinger (1989) and Scrutton et al. (1990). The site on the CysG protein, between amino acids 14 and 42, displays the characteristics of an NADP⁺ binding site. These coenzyme binding sites adopt a characteristic βαβ fold, and in this respect the N-

terminal sequence of the CysG protein is compatible with forming such a motif. Such a coenzyme could be involved in the conversion of dihydrosirohydrochlorin to sirohydrochlorin by the coenzyme-dependent dehydrogenation of dihydrosirohydrochlorin. With this in mind, the CysG protein was incubated with uroporphyrinogen III under strict anaerobic conditions in the presence and absence of NADP⁺. The incubation containing the NADP⁺ did not turn the usual yellow colour but instead turned a mauve colour similar to that of sirohydrochlorin. Indeed a u.v./visible spectrum of the incubation with NADP⁺ revealed that the enzyme had produced sirohydrochlorin as opposed to the dihydro compound, with the spectrum showing a new absorption maximum at 376 nm. Experiments on the C-terminal truncated CysG protein, which, when incubated under the same conditions with NADP⁺, did not catalyse the formation of sirohydrochlorin (Figure 7). These experiments, together with the identification of a putative pyridine dinucleotide binding site on the N-terminus of the protein, provide strong evidence that the N-terminus of the CysG protein is involved in the dehydrogenation of sirohydrochlorin. More recent work reported by Spencer et al. (1993) would appear to indicate that the CysG protein is performing an NAD⁺-rather than an NADP⁺-dependent dehydrogenation, as the enzyme would appear to work more efficiently with NAD⁺.

It was not possible to detect any dehydrogenase activity in extracts from strains harbouring the *cysG^B* gene. This was not altogether surprising as it was not possible to detect any CysG^B protein in these strains by SDS/PAGE analysis. Complementation experiments with the plasmid pEB2 showed that it did not complement the *cysG* auxotroph. Even a plasmid containing both the *cysG^A* and *cysG^B* genes, pEB1-2, was not able to complement the *cysG* auxotroph. This suggested that there was no CysG^B protein expression. In fact the only way possible so far to obtain active CysG^B activity, with no CysG^A activity, is by site-directed mutagenesis of key amino acids in the CysG^A domain of the complete CysG protein. Thus, the mutant CysG protein, G224A, where a glycine at position 224 has been substituted for an alanine, is unable to methylate uroporphyrinogen III, and the plasmid harbouring the mutant *cysG* gene, pSCW1, is unable to complement the *E. coli cysG* strain. However, if this mutant *cysG* gene is placed on the same plasmid as the *cysG^A* gene, pSCW2, then functional complementation of the *E. coli cysG* mutant is obtained (S. C. Woodcock and M. J. Warren, unpublished work).

Parallels with haem biosynthesis

The dehydrogenase activity described in this report, whereby dihydrosirohydrochlorin is oxidized by removal of two protons and two electrons, obviously draw a parallel with the oxidation of protoporphyrinogen IX to protoporphyrin IX in haem and chlorophyll biosynthesis. In this latter case the oxidation requires the removal of six electrons and six protons in a reaction catalysed by the enzyme protoporphyrinogen oxidase. It has been reported that this enzyme contains a tightly bound flavin dinucleotide (Siepker et al., 1987). Although genetic or protein primary structures have been attributed to protoporphyrinogen oxidase, it is interesting to note that the *hemY* gene from *Bacillus subtilis* (Hansson and Hederstedt, 1992) encodes a protein which contains a putative flavin dinucleotide binding site. This site and surrounding sequence shows good similarity to human monoamine oxidase and is shown in Figure 5. Considering the fact that the *hemY* gene is located in a hem operon on the *B. subtilis* genome and that mutations in this gene prevent the later steps in

haem biosynthesis (Hansson and Hederstedt, 1992), it can be hypothesized that the *hemY* gene encodes for an FAD-dependent protoporphyrinogen oxidase.

In this paper, evidence has been provided which demonstrates that the CysG protein is a multifunctional enzyme responsible for the complete transformation of uroporphyrinogen III into sirohaem. This conversion requires the enzyme to perform two SAM-dependent methylations at positions 2 and 7, an NAD⁺/NADP⁺-dependent dehydrogenation, and finally, ferrochelation. *In vitro* experiments demonstrating that the CysG protein is capable of performing this latter chelation activity have now also been performed, thereby confirming the conclusions reached here (Spencer et al., 1993). Thus it is possible to incubate uroporphyrinogen III in the presence of SAM, NAD⁺ and Fe²⁺ and obtain good quantitative yields of sirohaem. The CysG protein is therefore a trifunctional enzyme and should more aptly be named sirohaem synthase. The origin of the enzyme is of course debatable, but it is likely that the protein has arisen as a result of gene fusion. In this respect it has been reported that in *B. megaterium* there is a separate protein associated with oxidation and chelation (Robin et al., 1991). It would not seem unreasonable that the *E. coli cysG* gene has arisen by a gene fusion between a uroporphyrinogen methylase and a separate oxidase/chelate enzyme. The fact that the methylase can be dissected out of the CysG protein would also support this theory. Work is currently under way to investigate the biosynthesis of sirohaem in several other systems in order to identify the presence of separate independent oxidase/chelate enzymes.

Note added in proof (received 20 July 1994)

As predicted, the HemY protein has been shown to possess porphyrinogen oxidase activity (Dailey et al., 1994).

Financial support from S.E.R.C. is gratefully acknowledged.

REFERENCES

- Blanche, F., Debussche, L., Thibaut, D., Crouzet, J. and Cameron, B. (1989) *J. Bacteriol.* **171**, 4222–4231.
- Blanche, F., Robin, C., Couder, M., Faucher, D., Cauchois, L., Cameron, B. and Crouzet, J. (1991) *J. Bacteriol.* **173**, 4637–4645.
- Cole, J. A., Newman, B. M. and White, P. (1980) *J. Gen. Microbiol.* **120**, 475–483.
- Crouzet, J., Cauchois, L., Blanche, F., Debussche, L., Thibaut, D., Rouyez, M. C., Rigault, S., Mayaux, J.-F. and Cameron, B. (1990a) *J. Bacteriol.* **172**, 5968–5979.
- Crouzet, J., Cameron, B., Cauchois, L., Rigault, S., Rouyez, M. C., Blanche, F., Thibaut, D. and Debussche, L. (1990b) *J. Bacteriol.* **172**, 5980–5990.
- Dailey, T. A., Meissner, D. and Dailey, H. A. (1994) *J. Biol. Chem.* **269**, 813–815.
- Focus* (1984) **6**, 4–7.
- Griffiths, L. and Cole, J. A. (1987) *Arch. Microbiol.* **147**, 364–369.
- Hansson, M. and Hederstedt, L. (1992) *J. Bacteriol.* **174**, 8081–8093.
- Hanukoglu, I. and Gutfinger, T. (1989) *Eur. J. Biochem.* **180**, 479–484.
- Haydock, S. F., Dowson, J. A., Dhillon, N. Roberts, G. A., Cortes, J. and Leadlay, P. F. (1991) *Mol. Genet.* **230**, 120–128.
- Jones, M. C., Jenkins, J. M., Smith, A. G. and Howe, C. J. (1994) *Plant Mol. Biol.* **24**, 435–448.
- Jordan, P. M. (1991) in *Biosynthesis of Tetrapyrroles* (Jordan, P. M., ed.), pp. 1–59. Elsevier, Amsterdam.
- MacFerrin, K., Terranova, M., Schrieber, S. and Verdine, G. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1937–1941.
- Peakman, T., Crouzet, J., Mayaux, J.-F., Busby, S., Mohan, S., Harborne, N., Wootton, J., Nicholson, R. and Cole, J. (1990) *Eur. J. Biochem.* **191**, 315–323.
- Robin, C., Blanche, F., Cauchois, L., Cameron, B., Couder, M. and Crouzet, J. (1991) *J. Bacteriol.* **173**, 4893–4896.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Scrutton, N. S., Berry, A. and Perham, R. N. (1990) *Nature (London)* **343**, 38–43.
- Siepker, L. J., Ford, M., de Kock, R. and Kramer, S. (1987) *Biochim. Biophys. Acta* **913**, 349–358.
- Spencer, J. B., Roessner, C. A., Stolowich, N. J. and Scott, A. I. (1993) *FEBS Lett.* **335**, 57–60.

- Warren, M. J., Roessner, C. A., Santander, P. J. and Scott, A. I. (1990a) *Biochem. J.* **265**, 725–729
- Warren, M. J., Stelowich, N. J., Santander, P. J., Roessner, C. A., Sowa, B. A. and Scott, A. I. (1990b) *FEBS Lett.* **261**, 76–80

- Warren, M. J. and Scott, A. I. (1990) *Trends Biochem. Sci.* **15**, 486–491
- Warren, M. J., Roessner, C. A., Ozaki, S., Santander, P. J., Stelowich, N. J. and Scott, A. I. (1992) *Biochemistry* **31**, 603–609
- Warren, M. J., Bolt, E. and Woodcock, S. C. (1994) *CIBA Found. Symp.* **180**, 26–40

Received 2 December 1993/13 April 1994; accepted 20 April 1994