Volume 14 Number 15 1986 Nucleic Acids Research

Nucleotide sequence of the *hemC* locus encoding porphobilinogen deaminase of *Escherichia coli* K12

Steven D.Thomas and Peter M.Jordan*

Department of Biochemistry, University of Southampton, Bassett Crescent East, Southampton SO9 3TU, UK

Received 10 June 1986; Revised and Accepted 14 July 1986

ABSTRACT

Porphobilinogen deaminase, the product of the hemC locus in Escherichia coli K12, catalyses the tetrapolymerisation of porphobilinogen (PBG) into the hydroxymethylbilane, preuroporphyrinogen. The hemC locus has been subcloned from the Clarke and Carbon plasmid pLC41-4. The sequence of the hemC structural gene and flanking DNA was determined by the dideoxy chain termination method of Sanger. The structural gene for hemC is located within a 942bp sequence encoding the monomeric PBG deaminase, molecular weight 33,857. The extent of the coding region was confirmed by sequencing the N-terminus of the purified enzyme and by determination of the molecular weight. The hemC locus is closely linked to the cyaA locus, the genes being transcribed in a divergent manner. Upstream of the hemC coding region, a possible promoter and three repeated GGATG sequences were identified. This is the first report of a complete DNA sequence for a structural gene specifying an enzyme of the heme biosynthetic pathway in prokaryotes.

INTRODUCTION

Porphobilinogen (PBG) deaminase (EC.4.3.1.8.) catalyses the tetrapolymerisation of the monopyrrole, PBG, into the hydroxymethylbilane, preuroporphyrinogen (1,2). Preuroporphyrinogen subsequently acts as the substrate for uroporphyrinogen III cosynthase (3) resulting in the formation of uroporphyrinogen III, the key cyclic tetrapyrrole which acts as the precursor for all hemes, chlorophylls, corrins and related structures. PBG deaminases have been purified and characterised from a wide variety of prokaryotic and eukaryotic sources (4,5,6,7) and all appear to exist as a monomer with molecular weights ranging from 35,000 - 44,000.

PBG deaminase has been shown to catalyse the tetrapolymerisation reaction in several discrete steps and stable enzyme substrate intermediates have been isolated and characterised from both Rhodopseudomonas spheroides (8) and human (6) deaminases. The order in which the four molecules of the substrate PBG are assembled into the tetrapyrrole structure has also been elucidated (9) (10).

The E.coli PBG deaminase is the product of the hemC gene. The hemC gene has been mapped at approximately 84 minutes on the E.coli K12 chromosome (11) and more recently has been located adjacent to the cyaA gene locus (12).

The levels of the enzymes of the tetrapyrrole biosynthetic pathway are extremely low in E.coli, the PBG deaminase being less than 0.01% of the E.coli total soluble protein. In order to understand fully the structural features of the deaminase and the mechanism by which the enzyme catalyses the polymerisation of four substrate molecules into the hydroxymethylbiline, preuroporphyrinogen, our strategy was to clone and sequence the hemC gene and to incorporate the gene into a high expression plasmid vector so that larger amounts of the protein could be produced for structural studies. In this paper, we describe the subcloning and sequencing of the entire hemC structural gene and its flanking DNA, the identification of a putative promoter, and the location of a possible regulatory region containing a repeated pentanucleotide sequence.

MATERIALS AND METHODS

Strains and Media

E.coli K12 strain HB101 (13) and JM103 (14) were grown on L-broth medium supplemented with 0.1% w/v glucose. Solid media contained 1.5% w/v Difco agar. Transformed strains used for selection were grown on media supplemented with either tetracycline (15μg/ml) or ampicillin (50μg/ml), or both.

Plasmids

Plasmid pLC41-4 from the Clarke and Carbon library (15), previously identified as containing the hemC gene (12), was obtained from the Coli Genetic Stock Center, Yale University, New Haven, CT. The plasmid pBR322 (16) was employed for the subcloning of restriction fragments (17).

DNA isolation

The rapid isolation of plasmid DNA was carried out by an alkaline lysis procedure (18), using 1.5ml E.coli K12 cultures. Larger scale preparations of plasmid DNA were carried out by the SDS cleared lysate method (19) and the DNA was further purified by centrifugation to equilbrium in CsCl gradients containing ethidium bromide (1mg/ml). Restriction endonuclease digests of plasmid DNA were carried out by published methods (17), or according to the supplier's instructions. After electrophoresis, DNA fragments were isolated from the agarose gels by a freeze-thawing procedure (20).

DNA sequencing

DNA sequences were determined by the dideoxy chain termination method of Sanger et al (21). The complete DNA sequence of a 1.69kbp DNA fragment containing the intact hemC gene was determined on both strands from overlapping DNA sequences. A detailed restriction map of the hemC locus and the sequencing strategy are shown in the Results. The restriction fragments indicated were ligated into the appropriate cloning sites within the

polylinker region of M13mp8 and M13mp9 (22) and transformed into JM103 (23). Colourless plaques were selected and propagated for the preparation of phage DNA (24). Commercially available universal primer, d(5'TCCCAGTCACGACGT3') and the following synthetic primers Syn1-8 were employed for the sequencing procedure.

Syn1	d(5'ACGTCGCCAGTGCCAA3')	Syn5	d(5'CTGGCGACGTAAAC3')
Syn2	d(5'GAGACGGGCGGATGAC3')	Syn6	d(5'TTTCCGCCTACTTTGC3')
Syn3	d(5'GATTCACGCACTCGC3')	Syn7	d(5'GCAGGCACACTATGTC3')
Syn4	d(5'GACGGGTATTCATGGC3')	Syn8	d(5'CCAGGCCACCTGCCCC3')

The synthetic primers were synthesised according to Caruthers (25,26) using phosphite chemistry. The DNA sequences were compiled manually and analysed finally by computer.

Purification of PBG deaminase from E.coli

PBG deaminase was isolated from the genetically engineered E.coli strain ST1047 (see below) by a method similar to that developed for the PBG deaminase from R.spheroides (4) except that the ultracentrifugation step was omitted and a final f.p.l.c. stage was employed using a Pharmacia Mono Q ion exchange column. The enzyme was adsorbed on to the column in 50mM Hepes pH 7.0 and eluted with a linear NaCl gradient (0-0.5M NaCl) in the same buffer, (total volume 20ml). The enzyme was assayed according to Jordan and Shemin (4). The PBG deaminase was pure as judged by polyacrylamide gel electrophoresis in the presence of SDS (27). The molecular weight of the monomeric enzyme was determined by the method of Weber and Osborn (27) or by Sephacryl S200 chromatography.

N-Terminal analysis of PBG deaminase

Purified enzyme (400µg) was carboxymethylated (28) and linked to PITC-control pore glass at pH 8.5 in 0.2M ethylmorpholine buffer (29). The excess PITC groups were masked using propylamine and the sequence of the N-terminal amino acids were determined using an LKB sequenator according to the manufacturer's instructions. The PTH amino acids were identified by h.p.l.c. according to Bloxham et al (30).

Enzymes and Chemicals

[α-^{35}S]-thio-dATP was obtained from Amersham International p.l.c., T4 DNA ligase, DNA polymerase I (Klenow fragment) and restriction endonucleases were obtained from Bethesda Research Laboratories (UK) Ltd. Deoxy- and dideoxy nucleoside triphosphates and M13 sequencing primer were purchased from Pharmacia-PL Biochemicals. All other chemicals were obtained from Sigma (London) Ltd., or British Drug Houses, Poole, Dorset, UK.

RESULTS AND DISCUSSION

Cloning of hemC

Plasmid pLC41-4 from the Clarke and Carbon library (15) has been subjected to restriction endonuclease mapping (31) and more recently has been shown to carry the hemC gene locus (12). Our preliminary restriction map of pLC41-4 was essentially the same as that published and is shown in Fig.1. The plasmid pLC41-4 contains, in addition to the 6.2kbp ColE1 vector DNA, a 12.8kbp insert mapped at 85 minutes on the E.coli K12 chromosome, and shown to contain the cyaA locus in addition to hemC. As previously reported (12) the chromosomal DNA was subcloned after partial Sau3A digestion into the BamH1 site of pBR322 yielding pST46. After transformation into HB101, one of the resulting clones ST1046 exhibited PBG deaminase activity eighteen times that of the parental strain, when assayed for the presence of the enzyme. Restriction endonuclease analysis revealed a 2.89kbp BamH1-Sau3A insert within the pBR322 sequence (Fig.1). By means of a BamH1-Sal1 double digest the insert was further fragmented into a BamH1-Sal1 fragment of 1.69kbp. This fragment was cloned into pBR322, previously digested with the appropriate restriction endonucleases, giving the plasmid pST47 (Fig.1) which was transformed into HB101. The resulting strain (ST1047) exhibited levels of PBG deaminase fifty times that of the wild type. The BamH1-Sal1 fragment (1.69kbp) in pST47 was large enough to accommodate the DNA required (~1kbp) for a structural gene encoding a protein within the molecular weight range of other deaminases and this fragment provided the basis for all the DNA sequencing.

The increase in deaminase expression from eighteen to fifty times that of the wild type noted in strains ST1046 and ST1047 respectively, is almost certainly due to an increased plasmid copy number rather than to the removal

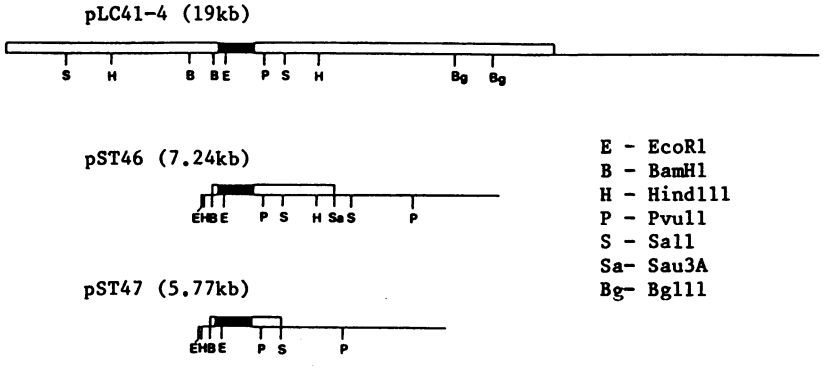

Figure 1: Restriction Map of Plasmids pLC41-4, pST46 and pST47. The thin line represents ColE1 DNA or pBR322 DNA (in pST46 and pST47). The position of hemC within the chromosomal DNA is shown by the shaded section. Only restriction sites pertinent to the hemC gene and to this study are shown.

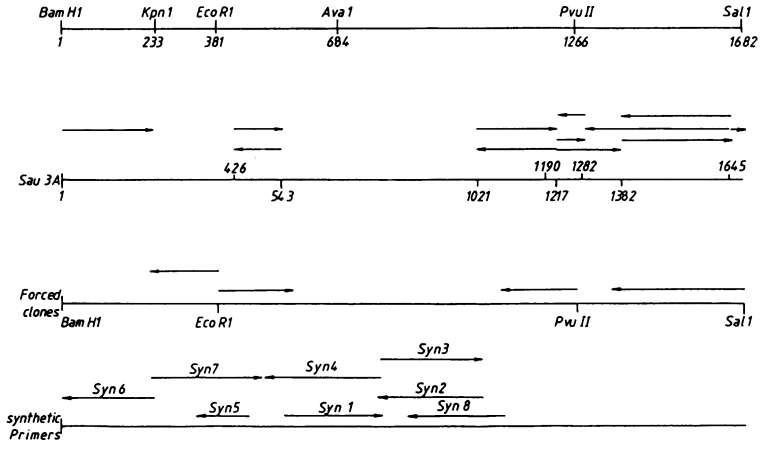

Figure 2: Sequencing Strategy for the hemC Locus. The arrows denote sequencing direction and length of the sequences obtained in individual experiments. Universal sequencing primer was used in all cases except where defined synthetic primers (Syn1-8) were used.

of any regulatory element. This is supported by the observation that strain ST1047 yields almost three times as much plasmid DNA as ST1046.

Nucleotide sequence of the DNA fragment encoding hemC

The BamHl-Sall fragment was sequenced according to the strategy outlined in Fig.2. The nucleotide sequence of the hemC gene and the DNA extending at both 5' and 3' ends is shown in Fig.3. The DNA sequence contains an open reading frame extending from position 87 to 1058. Analysis of the entire BamHl-Sall nucleotide sequence using the frame-scan computer program (32) with a scan window of 21 nucleotides revealed a high probability of a coding sequence within this open reading frame which extended from bases 220 to 1040. Four methionine codons are present at the 5' end of this reading frame at positions 99, 105, 119 and 201, however, only those at 99 and 119 are preceded by putative Shine-Dalgarno ribosome binding sequences (AGGA and GACGG, respectively). The ATG at position 99 is unlikely to be the codon specifying the start of translation since the putative Shine-Dalgarno sequence upstream of this codon is within the -10 region of the putative hemC promoter (see below). The initiation codon is thus likely to be at position 119. This would generate a primary translation product containing 313 amino acids with a molecular weight of 33,857.

Derived amino acid sequence of the hemC gene product, PBG deaminase

The presence of an open reading frame of lower coding probability from bases 87 to 220, containing four in-frame ATG start codons, made it necessary to sequence the N-terminus of the PBG deaminase protein in order to define precisely the site for initiation of translation. The PBG deaminase,

```
1    GGATCCACTGCCAGACCTCATTTTACGGTTTGCGCAGGCGTCTACGTTTCACCACAACA

60   CTGACATCACTCTGGCAAGGATGTTAGGATGGACCACGGATGATAATGACGGTAACAAGC

      M  L  D  N  V  L  R  I  A  T  R  Q  S  P  L  A  L  W  Q  A
120  ATGTTAGACAATGTTTTAAGAATTGCCACACGCCAAAGCCCACTTGCACTCTGGCAGGCA

      H  Y  V  K  D  K  L  M  A  S  H  P  G  L  V  V  E  L  V  P
180  CACTATGTCAAAGACAAGTTGATGGCGAGCCATCCGGGCCTGGTCGTTGAACTGGTACCG

      M  V  T  R  G  D  V  I  L  D  T  P  L  A  K  V  G  G  K  G
240  ATGGTGACGCGCGGCGATGTGATTCTTGATACGCCGCTGGCGAAAGTAGGCGGAAAAGGC

      L  F  V  K  E  L  E  V  A  L  L  E  N  R  A  D  I  A  V  H
300  TTATTTGTAAAAGAGCTGGAAGTCGCGCTCCTCGAAAATCGCGCCCGATATCGCCGTACAC

      S  M  K  D  V  P  V  E  F  P  Q  G  L  G  L  V  T  I  C  E
360  TCAATGAAAGATGTGCCGGTTGAATTCCCGCAAGGTCTGGGACTGGTCACTATTTGTGAG

      R  E  D  P  R  D  A  F  V  S  N  N  Y  D  S  L  D  A  L  P
420  CGTGAAGATCCTCGCGATGCCTTTGTGTCCAATAACTATGACAGTCTGGATGCGTTACCG

      A  G  S  I  V  G  T  S  S  L  R  R  Q  C  Q  L  A  E  R  R
480  GCAGGCAGTATCGTCGGGACGTCCAGTTTACGTCGCCAGTGCCAACTGGCTGAACGCCGT

      P  D  L  I  I  R  S  L  R  G  N  V  G  T  R  L  S  K  L  D
540  CCGGATCTGATTATCCGCTCCCTGCGCGGCAACGTCGGCACTCGCCTGAGCAAACTGGAT

      N  G  E  Y  D  A  I  I  L  A  V  A  G  L  K  R  L  G  L  E
600  AACGGCGAATACGATGCCATCATTCTTGCCGTAGCCGGACTAAAACGTTTAGGTCTGGAG

      S  R  I  R  A  A  L  P  P  E  I  S  L  P  A  V  G  Q  G  A
660  TCACGTATTCGCGCCGCGTTGCCACCCGAGATTTCTCTTCCGGCGGTAGGACAAGGTGCG

      V  G  I  E  C  R  L  D  D  S  R  T  R  E  L  L  A  A  L  N
720  GTGGGTATTGAATGCCGCCTTGATGATTCACGCACTCGCGAGCTGCTTGCCGCGCTGAAT

      H  H  E  T  A  L  R  V  T  A  E  R  A  M  N  T  R  L  E  G
780  CACCACGAAACTGCACTGCGCGTTACCGCAGAACGCGCCATGAATACCCGTCTCGAAGGC

      G  C  Q  V  P  I  G  S  Y  A  E  L  I  D  G  E  I  W  L  R
840  GCATGTCAGGTGCCAATTGGTAGCTACGCCGAGCTTATTGATGGCGAAATCTGGCTGCGT

      A  L  V  G  A  P  D  G  S  Q  I  I  R  G  E  R  R  G  A  P
900  GGGCTGGTCGGCGCGCCGGACGGTTCGCAGATTATTCGCGGTGAACGCCGCGGTGCGCCG

      Q  D  A  E  Q  M  G  I  S  L  A  E  E  L  L  N  N  G  A  R
960  CAAGATGCCGAACAAATGGGGATTTCGCTGGCAGAAGAGCTACTGAATAACGGCGCGCGC

      E  I  L  A  E  V  Y  N  G  D  A  P  A  *
1020 GAGATCCTCGCTGAAGTCTATAACGGAGACGCCCCGGCATGAGTATCCTTGTCACCCGCC

1080 CGTCTCCCGCTGGAGAAGAGTTAGTGAGCCGTCTGCGCACACTGGGGACGGTGGCCTGGC

1140 ATTTTCCGCTGATTGAGTTTTCTCCGGGTCAACAATTACCGCAACTTGCTGATCAACTGG

1200 CAGCGCTGGGGGAGAGCGATCTGTTGTTTGCCCTCCTCGCAACACGCGGTTGCTTTTGCC

1260 CAATCACAGCTGCATTCAGCAAGATCGTAAATGGCCCCGACTACCTGATTATTTCGCCAT

1320 TGGACGCACCACCGCACTGGCACTACATACCGTAAGTCGGACAGAAGATTCTCTACCCGC

1380 AGGATCGGGAAATCAGCGAAGTCTTGCTACAATTACCTGAATTACAAAATATTGCGGGCA
```

1440 AACGTGCGCTGATATTACGTGGCAATGGTGGTCGTGAGCTAATTGGGGATACCTGACGGC

1500 GCGCGGTGCTGAGGTCACTTTTTGTGAATGTTATCAACGATGCGCAATCCATTACGATGT

1560 GCAGAAGAAGCGATGCGCTGCAAGCCCGCGAGGTGACGATGGTCGTTGTTACCAGCGGTG

1620 AAATGTTGCAGCAACTCTGGTCACTGATCCCACAATGTATCGTGAGCACTGGTTACTACA

1680 CTGTCGAC

Figure 3: DNA Sequences and Translation of hemC. The sense strand is shown in the direction 5'→3'. The sequence is numbered from the first nucleotide of the BamHl recognition site and ends at the final nucleotide of the Sall recognition site. The sequences comprising the -10 and -35 regions of the putative hemC promoter are boxed. The proposed Shine-Dalgarno sequence is overlined and the repeated GGATG sequences are underlined. The location of a codon in the sequence is designated by the number of the first nucleotide of the codon.

therefore, was purified to homogeneity and the N-terminus was sequenced by a semi-automated Edman procedure. H.p.l.c. analysis of the resulting PTH amino acid derivatives revealed the following sequence:

$$NH_2\text{-M-L-D-N-V-L-}$$

This sequence is consistent with a translational start at the ATG codon in position 119 of the BamHl - Sall fragment, and not at the ATG in position 201 which was marginally more probable from the frame-scan computer analysis. The total coding region of hemC thus appears to occupy 942 nucleotides stretching from the ATG at 119 to the TGA at 1040. Further confirmation as to the extent of the reading frame was obtained by comparison of the predicted molecular weight of 33,857 derived from this gene sequence with those of 35,000 ± 2,000, obtained from the determination of the molecular weight of the PBG deaminase enzyme using SDS polyacrylamide gel electrophoresis (27) and 32,500 ± 3,000 obtained by gel filtration on Sephacryl S200.

The predicted amino acid composition of the enzyme, compiled on the basis of the gene sequence, may be deduced from Table 1. An amino acid analysis of purified PBG deaminase is consistent with the predicted composition (data not shown). Further evidence for the extent of the structural gene was obtained by specific chemical cleavage of the PBG deaminase enzyme using formic acid. Cleavage of the single aspartic acid-proline link (between the amino acids at positions 103 and 104) yielded, as expected, two peptides of molecular weights 11,000 and 24,000 (data not shown).

Table 1: Codon Usage for PBG Deaminase.

Codon Usage											
F	TTT	2	S	TCT	1	Y	TAT	3	C	TGT	2
F	TTC	1	S	TCC	3	Y	TAC	2	C	TGC	2
L	TTA	6	S	TCA	3	-	TAA	0	-	TGA	1
L	TTG	2	S	TCG	2	-	TAG	0	W	TGG	2
L	CTT	7	P	CCT	1	H	CAT	1	R	CGT	7
L	CTC	5	P	CCC	1	H	CAC	4	R	CGC	19
L	CTA	2	P	CCA	3	Q	CAA	6	R	CGA	0
L	CTG	20	P	CCG	11	Q	CAG	4	R	CGG	0
I	ATT	13	T	ACT	4	N	AAT	6	S	AGT	3
I	ATC	6	T	ACC	2	N	AAC	5	S	AGC	4
I	ATA	0	T	ACA	1	K	AAA	7	R	AGA	1
M	ATG	6	T	ACG	3	K	AAG	1	R	AGG	0
V	GTT	4	A	GCT	2	D	GAT	14	G	GGT	8
V	GTC	8	A	GCC	13	D	GAC	5	G	GGC	12
V	GTA	6	A	GCA	7	E	GAA	16	G	GGA	6
V	GTG	6	A	GCG	12	E	GAG	8	G	GGG	2

Codon usage

The codon usage for the hemC structural gene is reported in Table 1. As expected, for a protein such as PBG deaminase which represents only 0.01% of the soluble protein in E.coli, the codon usage matches that of a weakly expressed gene (33). There are several other interesting features worth noting. The rare arginine codon AGA is used only once in the whole structural gene sequence, coding for the arginine residue at position 7 in the primary structure. The use of such a rare codon near the N-terminus coding region could well contribute to the poor expression of the deaminase since the level of the t-RNA which recognises this codon is very low in E.coli. The isoleucine codon ATA is not used.

Identification of the putative hemC promoter

A computer search for possible promoter sequences was made upstream of the initiation codon. Only one region which fulfilled the criteria for a RNA polymerase binding site was found, the location of which is indicated by the boxed nucleotides in Fig.3. The proposed -10 region (TAGGAT) differs from the concensus sequence, TATAAT (34), at two positions and the -35 region (GTGACA) differs only in a single position from the concensus sequence TTGACA. Importantly, these two regions are separated from one another by the near optimal 18 nucleotides. The nucleotide sequences within this putative promoter region are consistent with those of a weakly transcribed mRNA (33).

DNA sequences flanking the hemC structural gene

A computer search for regions of dyad symmetry revealed no obvious areas

of secondary structure. However, it is interesting to note part of the putative -35 sequence (ACA) is located in a region of dyad symmetry (ACATCACTCTGGCAAGGATGT). Another region of dyad symmetry (GCAACACGCGGTTGC) occurs downstream of the structural gene starting at position 1238 and is followed by TTTT. Such areas of dyad symmetry followed by either T rich or A rich regions have been postulated as signals for the termination of transcription (35).

More significant is a sequence straddling the putative -10 region which contains three repeated pentanucleotides within the space of 24 bases (GGATGTTAGGATGGACCACGGATG). The low probability of finding three such repeats within a relatively short length of DNA leads one to believe that this area of the DNA is playing an important functional role. Interestingly, the pattern of the repeat is conserved in the putative -10 region TAGGATG giving rise to a sequence not found to date in any other E.coli -10 region (36). Repeated sequences in the -35 regions of the P_I and P_{RE} promoters of bacteriophage λ have been shown to bind the cII protein activator and play a crucial role in the regulation of transcription by RNA polymerase (37). In the absence of any information about the mechanism of regulation of heme synthesis in E.coli, or of direct experimental evidence for the location of the hemC promoter, conclusions about the significance of the GGATG repeats can only be looked upon as tentative and will form the basis of further research.

A further interesting finding is the presence of an open reading frame having a high coding probability and extending from an ATG start codon at position 1058. This ATG is made up of two of the bases from the TGA termination codon of the hemC structural gene. In addition to this finding, a good Shine-Dalgarno (38) sequence (GGAG) is present ~10 bases upstream of this start codon suggesting that an additional protein could be produced under the control of the putative hemC promoter. The significance of this protein coding region is not known.

Location of hemC on the E.coli K12 Chromosome

Comparison of the nucleotide sequence of the BamHl-Sall fragment with the nucleotide sequence reported for the cyaA locus(39) revealed an overlap between the flanking DNA upstream of the hemC gene with that upstream of cyaA (Fig. 4) indicating that the two genes are linked and that they are expressed in a divergent manner. Since it is known that the cyaA gene maps at 84 minutes and is transcribed in the direction 100→0 minutes (40), hemC must be expressed in the direction 0→100 minutes on the E.coli K12 chromosome. Within the sequence of 386 nucleotides between the start codons of the two structural genes, several stop codons are present in all reading frames in both directions suggesting that this intervening DNA is not a protein coding region. It is also most significant that in other studies (41) evidence was

Figure 4: Intergenic Region between cyaA and hemC. Both strands are shown extending from the coding region of cyaA to the coding region of hemC. The hemC initiates with ATG and cyaA with TTG. The cyaA promoter and the putative hemC promoters are boxed. The sequence shown reads in the direction 0→100 minutes on the E.coli K12 chromosome.

provided for the existence of a promoter functioning in a divergent manner to that of cyaA and which was located in a region of the intervening DNA consistent with our predictions for the position of the hemC promoter.

Structure of the PBG deaminase

Since PBG deaminase is not known to belong to any class of proteins whose X-ray structure is known,it is not possible to make any meaningful conclusions about the tertiary enzyme structure from the gene-derived amino acid sequence. It is worth noting, however, the presence of two sequences which show nucleotide homology stretching from positions 657 to 671 and 774 to 758 and coding for amino acid sequences ESRIR and DSRTR, respectively. These sequences, GATTCACGCACTCGC and GAGTCACGTATTCGC, have 12 out of 15 nucleotides in common and could represent two structurally similar regions of the enzyme. In fact, many of the arginine residues are found in pairs or closely associated with one another as in the sequences above. Since the PBG deaminase catalyses the polymerisation of four molecules of porphobilinogen, each of which has two negatively charged carboxyl groups, it is expected that several of these arginines may be involved in substrate binding. It is significant that studies with PBG deaminase using butadione (42) have highlighted the importance of an active site arginine.

Although cDNA and gene sequences from the heml (hemA) gene from prokaryote and eukaryote sources have been published (43, 44, 45), this is

the first report of a DNA sequence for an entire gene coding for an enzyme of the heme biosynthetic pathway in bacteria.

ACKNOWLEDGEMENTS

This work was supported by the SERC and BTG. We are indebted to Dr. I.G. Giles for running the frame-scan computer programs and for printing out the nucleotide sequences and to Dr. M.G. Gore who carried out the N-terminal analysis.

* To whom all correspondence should be addressed.

REFERENCES

1. Burton, G., Fagerness, P.E., Hosozawa, S., Jordan, P.M. and Scott, A.I. (1979) J.Chem.Soc. Chem.Commun. 202-204.
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. 1155-1158.
3. Jordan, P.M., Burton, G., Nordlöv, H., Schneider, M.M., Pryde, L.M. and Scott, A.I. (1979) J.Chem.Soc. Chem.Commun. 204-205.
4. Akhtar, M. and Jordan, P.M. (1979) Comprehensive Org.Chem. 5, 1121-1163, and references therein.
5. Jordan, P.M. and Shemin, D. (1973) J.Biol.Chem. 248, 1019-1024.
6. Anderson, P.M. and Desnick, R.J. (1980) J.Biol.Chem. 255, 1993-1999.
7. Williams, D.C., Morgan, G.S., McDonald, E. and Battersby, A.R. (1981) Biochem. J. 193, 301-310.
8. Jordan, P.M. and Berry, A. (1981) Biochem.J. 195, 177-181.
9. Seehra, J.S. and Jordan, P.M. (1980) J.Amer.Chem.Soc. 102, 6841-6846.
10. Battersby, A.R., Fookes, C.J.R., Matcham, G.W.J., McDonald, E. and Hollenstein. (1983) J.Chem.Soc., Perkin Trans.1, 3031-3040.
11. McConville, M.L. and Charles, H.P. (1979) J.Gen.Micro. 111, 193-200.
12. Jordan, P.M., Marshall, L.M. and Thomas, S.D. (1986) Colloque INSERM Vol.134 "Porphyrins and Porphyrias" Ed. Nordmann, Y. 83-93.
13. Boyer, H.W. and Roulland-Dussoix, D. (1969) J.Mol.Biol, 41, 459-472.
14. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl. Acids Res.9, 309-321.
15. Clarke, L. and Carbon, J. (1976) Cell 9, 91-99.
16. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.V., Heynecker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 75-93.
17. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.
19. Godson, G.N. and Vapnek, D. (1973) Biochim.Biophys.Acta. 299, 516-523.
20. Thuring, R.W.J., Sanders, J.P.M. and Borst, P. (1975) Anal.Biochem. 66, 213-220.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci. 74, 5463-5467.
22. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
23. Mandel, M. and Higa, A. (1970) J.Mol.Biol. 53, 154-160.
24. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J.Mol.Biol. 143, 161-178.
25. Matteucci, M.D. and Caruthers, M.H. (1981) J.Amer.Chem.Soc. 103, 3185-3191.
26. Itakura, K., Rossi, J.J. and Wallace R.B. (1984) Ann.Rev.Biochem. 53 323-356.

27. Weber, K. and Osborn, M. (1969) J.Biol.Chem. 244, 4406-4412.
28. Hirs, C.H.W. (1967) Methods in Enzymol. 11, 199-203.
29. Laursen, R.A., Horn, M.J. and Bonner, A.G. (1972) FEBS.Letts. 21, 67-71.
30. Bloxham, D.P., Parmelee, D.C., Kumar, S., Walsh, K.A. and Titani, K. (1982) Biochemistry 21, 2028-2036.
31. Roy, A. and Danchin, A. (1981) Biochemie 63, 719-722.
32. Staden, R. and McLachlan, A.D. (1982) Nucl.Acids Res. 10, 141-156.
33. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.
34. Rosenberg, M. and Court, D. (1979) Ann.Rev.Genet. 13, 319-353.
35. Holmes, W.M., Platt, T. and Rosenberg, M. (1983) Cell 32, 1029-1032.
36. Hawley, D.K. and McClure, W.R. (1983) Nucl. Acids Res. 11, 2237-2255.
37. Ho, Y-S., Wulff, D. and Rosenberg, M. (1986) Soc. Gen. Microbiol. Symp. 39, 79-103. Eds. Booth, I.R. and Higgins, C.F. Cambridge Univ. Press.
38. Shine, J. and Dalgarno, L. (1974) Proc.Natl.Acad.Sci. 71, 1342-1346.
39. Aiba, H., Mori, K., Tanaka, M., Ooi, T., Roy, A. and Danchin, A. (1984) Nucl.Acids.Res. 12, 9427-9440.
40. Bachmann, B.J. (1983) Microbiol.Rev. 47, 180-230.
41. Roy, A. and Danchin, A. (1982) Mol.Gen.Genet. 188, 465-471.
42. Russell, C.S., Pollack, S. and James, J. (1984) Fed.Proc., Fed.Amer.Soc. Exp.Biol. 43, 1545.
43. Urban-Grimal, D., Volland, C., Garnier, T., Dehoux, P. and Labbe-Bois, R. (1986) Eur. J. Biochem. 156, 511-519.
44. Leong, S.A., Williams, P.H. and Ditta, G.S. (1985) Nucl. Acids Res. 13, 5965-5976.
45. Borthwick, I.A., Srivastava, G., Day, A.R., Pirola, B.A., Snoswell, M.A., May, B.K. and Elliot, W.H. (1985) Eur. J. Biochem. 150, 481-484.