Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1626 (2003) 102–105
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Short sequence paper

A heme-deficient strain of *Escherichia coli* has a three-base pair deletion in a "hotspot" in *hemA*

Juncheng Li[a], Berthony Deslouches[b,1], Sharon D. Cosloy[a], Charlotte Sananes Russell[b,c,*]

[a] *Department of Biology, City College, City University of New York, New York, NY 10031, USA*
[b] *Department of Biochemistry, City College, City University of New York, New York, NY 10031, USA*
[c] *Department of Chemistry, City College, City University of New York, New York, NY 10031, USA*

Received 30 July 2002; received in revised form 31 January 2003; accepted 7 February 2003

This work is dedicated to the memory of Sharon D. Cosloy, deceased September 2, 2001.

Abstract

The key regulatory step in heme biosynthesis in *Escherichia coli* is at the level of glutamyl-tRNA reductase (GTR), an enzyme which is encoded by *hemA*. A strain, HU227, with a spontaneous in-frame mutation in *hemA* has no GTR activity. The mutation is shown to be a three-base deletion at a "hotspot" in the gene. The amino acid sequence in this region is highly conserved.

Keywords: heme biosynthesis; *hemA*; *Escherichia coli*; glutamyl-tRNA reductase

The biosynthesis of the tetrapyrroles, such as hemes, chlorophylls, siroheme and corrin, begins with the synthesis of δ-aminolevulinic acid (ALA). The pathway is highly conserved except for the synthesis of ALA which is derived from glycine and succinyl CoA in some bacteria, in fungi and in eukaryotes except for green plants. In most bacteria, archaea and green plants, ALA is derived from glutamate in three steps: glutamate (glu) is converted to glu-tRNA by glu-tRNA synthetase; glu-tRNA is reduced by glu-tRNA reductase (GTR), the gene product of *hemA*, to form glu-1-semialdehyde (GSA); GSA is converted to ALA by GSA-aminotransferase [1].

Escherichia coli synthesizes heme and siroheme. When *E. coli* is grown in medium supplemented with ALA, it accumulates the intermediates of tetrapyrrole biosynthesis [2]. This suggests that regulation of the biosynthesis of ALA regulates the whole pathway. Because glu-tRNA is an intermediate in protein biosynthesis as well, and because some GSA is converted to ALA spontaneously, the GTR step is the putative regulatory step for ALA synthesis. *hemA* from *E. coli* has been cloned and sequenced and the amino acid sequence of GTR has been derived from its nucleotide sequence [3]. The amino acid sequence of GTR has several highly conserved areas (Fig. 1, Ref. [4] and references therein). Mutations in *hemA* and their effects on GTR activity have been reported from various sources.

Nakayashiki et al. [5] reported that a mutation in *E. coli* at nt 314 in *hemA*, which substitutes N for S_{105}, caused lower activity in the expressed GTR. A *B. subtilis* strain with a mutation in *hemA* caused by a G-to-A transition, which caused the substitution of C_{105} by Y, had no GTR activity [6]. In *E. coli* and *Chlorobium vibrioforme*, S is at that position, and barley, cucumber, *Arabidopsis* and *Synechocystis* have A at that position. The amino acid at position 105 is adjacent to a highly conserved area (Fig. 1, Ref. [4] and references therein). When each C in GTR from *Methanopyrus kandleri* was replaced with S, only C_{48} proved to be essential for activity, while replacement of H_{84} with A or Q reduced enzyme activity significantly [7].

For *E. coli*, Drolet et al. [8] reported that deletion of the N terminal 29 codons of *hemA* eliminated GTR activity. Wang et al. [9] showed that in *Salmonella typhimurium* replacement of $L_{3,4}$ with K stabilized GTR to proteolysis. Deletions of codons for amino acids 3–7 resulted in no GTR activity.

* Corresponding author. Department of Chemistry, City College, City University of New York, Convent Avenue at 138th Street, New York, NY 10031, USA. Tel.: +1-212-650-6681; fax: +1-212-650-6657.
E-mail address: chrcc@sci.ccny.cuny.edu (C.S. Russell).
[1] Department of Molecular Genetics and Biochemistry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA.

0167-4781/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0167-4781(03)00041-1

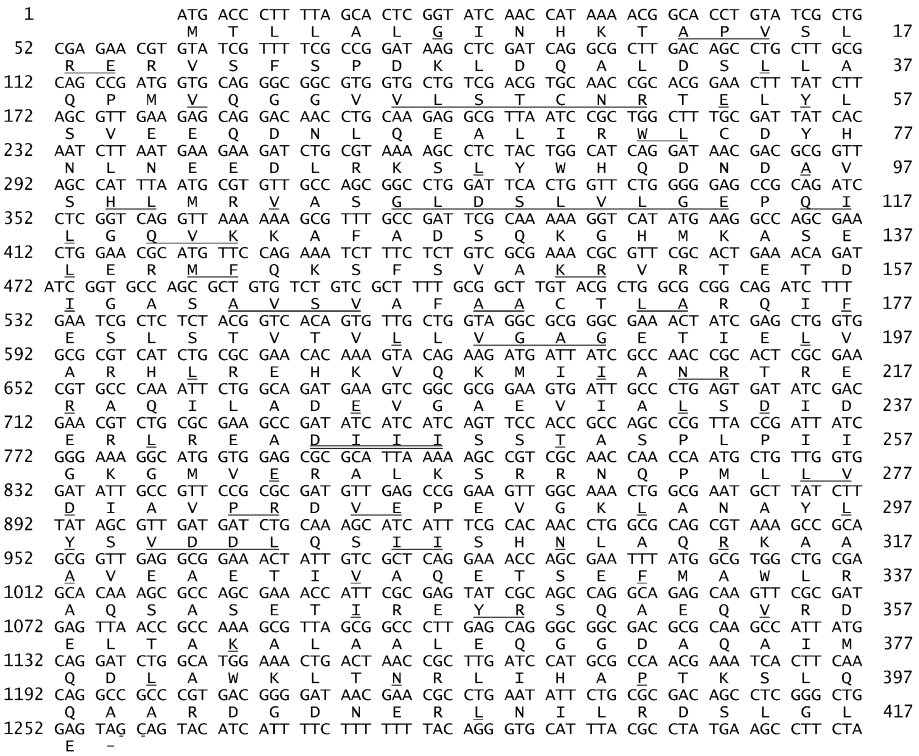

Fig. 1. Nucleotide sequence of *hemA* and amino acid sequence of GTR in Hfr Cavalli. Underlined sequences are conserved regions [4] and the double underline sequence is the conserved sequence concerned in the mutation in *hemA* of HU227. The numbers on the left refer to nucleotide positions and the numbers on the right refer to amino acid positions.

A. Fragments of *hemA* amplified by PCR for sequencing:

+1_____***hemA***_____**1256bp**

-56 (**I**)_____424(**VI**)

353(**II**)_____(733-742 ATCATCATCA)_____844 (**V**)

783(**III**)_____1293 (**IV**)

B. Sequences of the primers used for PCR:

Forward primer I (-77 -> -56bp) 5'CTCTTTTATTGATCTTAGGC

Reverse primer VI (424 - 443bp) 5'ACAGAGAAAGATTTCTGGAA

Forward primer II (334 - 353bp) HU227
5'CTGGGGGAGCCGCAATCCT
II' (383 – 413bp) Hfr Cavalli
5'GCAAAAAGGTCATATGAAGGCCAGCGAACT

Reverse primer V (844 - 863bp) HU227
5'TCCGGCTCAACATCGCGCGG
V' (864 - 893bp) Hfr Cavalli
5'TAAAGATAAGCATTCGCCAGTTTGCCAACT

Forward primer III (764 - 783bp) 5'CGATTATCGGGAAAGGCATG

Reverse primer IV (1293 - 1312bp) 5'GATAGAAGGCTTCATAGGCG

Fig. 2. *hemA* of Hfr Cavalli and of the *hemA* strain, HU227, were sequenced by PCR of segments of the respective genes on their chromosomes. (A) The fragments of *hemA* from HU227 amplified by PCR for sequencing are shown with their boundaries within and outside the structural gene. The relative position of the repetitive sequences is shown as inside the center fragment. The scale is not accurate. (B) The sequences of the primers for PCR. A different set of primers, II′ and V′, was used for PCR of the internal fragment of the wild-type gene from Hfr Cavalli.

Plant GTR's "have 31–34 amino acids at the N terminus which are not present in bacterial enzymes" [10], and which are not part of the transit peptide required for transport of GTR into the chloroplast. Deletion of the 30 N-terminal amino acids of barley GTR does not change GTR enzymatic activity but the cell becomes resistant to feedback inhibition by heme. A truncated enzyme lacking the 19 N-terminal amino acids can rescue an *E. coli hemA* mutant strain, but the entire intact protein cannot. The following mutations were made in barley *hemA* which had been truncated by deletion of the first N-terminal 30 codons, and fused to glutathione *S*-transferase (GST): a mutant with four amino acid substitutions in GTR, $M_{122} \rightarrow K$, $K_{154} \rightarrow N$, $F_{371} \rightarrow L$, $E_{400} \rightarrow K$, gave about 0.05 of wild-type GTR activity. K_{154}, F_{371} and E_{400} are conserved residues. A mutant with two amino acid substitutions, $L_{302} \rightarrow S$ and $L_{387} \rightarrow H$, where L_{302} and L_{387} are conserved in higher plants and cyanobacterial enzymes, gave about 0.1 activity of wild type. A mutant with three amino acid substitutions, $I_{318} \rightarrow L$, $R_{322} \rightarrow G$ and $N_{454} \rightarrow D$, where the R_{322} is highly conserved and I_{318} is highly conserved (except for *Chlorobium* in which leu is in this position), and N_{454} is conserved in plant but not bacterial sequences, the activity was 0.05 of wild type. A mutant with one amino acid substitution, L_{464} to P, had the lowest activity, about 0.01 of wild type. L_{464} is highly conserved except for *B. subtilis* where it is A.

This paper reports the identification of a non-lethal spontaneous mutation in a "hotspot" in *E. coli* which causes ALA auxotrophy.

A *hemA* strain, SASX41B (supplied by Barbara Bachman, Coli Genetic Stock Center [CSGC]), was first isolated by Sasarman et al. [11] as a spontaneous mutant of Hfr Cavalli, by neomycin selection. It was mutagenized to hemin-permeability [12]. The resultant strain, HU227, has no GTR activity, but it can grow on ALA or hemin. *hemA* is part of an operon which includes *prfA*, encoding the essential release factor 1 (RF1) [3]. However, the mutation in *hemA* in HU227 is not lethal and the strain can grow fermentatively on glucose.

The mutant *hemA* gene from HU227 was copied as three overlapping pieces by PCR of chromosomal DNA prepared from HU227 (Fig. 2). The wild-type gene in the parent strain of HU227, Hfr Cavalli [11], was copied in the same way. The PCR products were cloned into plasmid pUC18 (Invitrogen) and the inserts were sequenced. The same procedure was followed for the wild-type parent strain, Hfr Cavalli, except that a different set of primers was used for the central overlapping piece.

The coding sequence of the mutant strain differed from that of the parent strain in only one feature: a three-base pair deletion from nucleotide positions 734–742, which have a run of three ATC triplets in the coding strand. This deletion

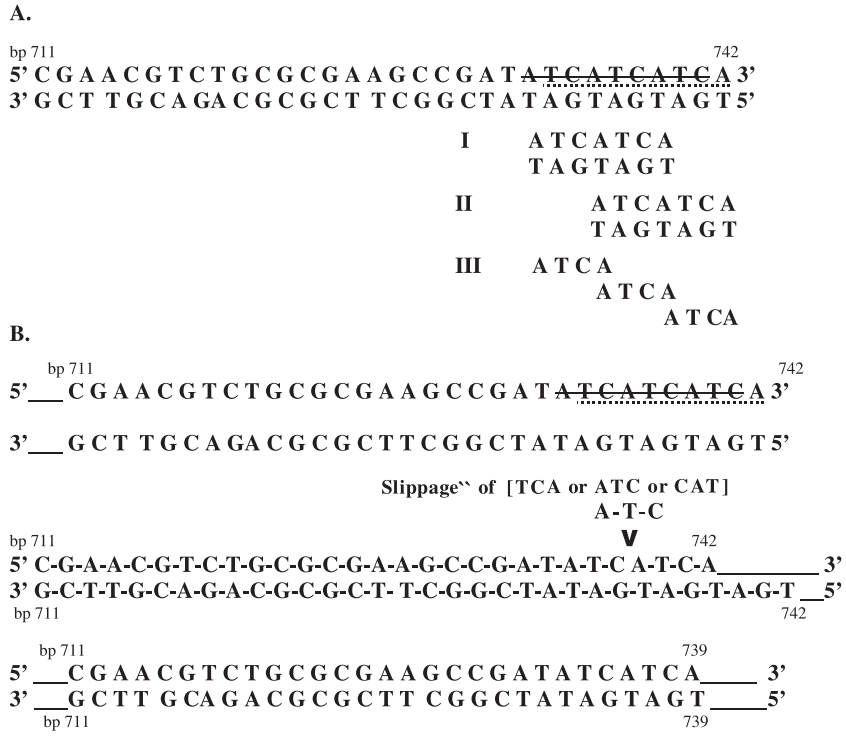

Fig. 3. (A) The nucleotide sequence of *hemA* from 711–742 bp. Features of the "hotspot". Repetitive sequences in the region (711–742 bp) around the site of the three-base pair deletion mutation in HU227. There are three TCA repeats (dotted underline) and three ATC repeats (strikethrough), two overlapping seven-base pair repeats and three overlapping four-base pair repeats. (B) The wild-type sequence is shown, followed by the intermediate in which any one of three TCA's or three ATC's or two CAT's is unpaired as a result of misalignment. This structure could lead to addition or deletion of the triplet. The third structure shows the two strands after deletion of a triplet from both strands.

results in the absence of one of a triplet of isoleucines from a conserved region of the gene at amino acid residues 245–247 (I–I–I) (Fig. 1, double underline). The amino acid sequence at positions 244–247, D–I–I–I, is conserved in the amino acid sequence of GTR in *E. coli*, *S. typhimurium* and *C. vibrioforme*. In *Arabidopsis* it is D–V–I–F, in *Synechocystis* it is D–I–V–F, in *Coxellia burnetii* it is D–I–V–I, in *B. subtilis* it is D–I–L–I, and in *Methanobacterium thermoautotrophicum* and *Clostridium josui*, it is D–V–V–I. Thus, an aspartic residue with three hydrophobic amino acid residues appears to be a highly conserved sequence.

It is clear from the base and amino acid sequence below and Fig. 1 that removal of any one of three TCA's or ATC's or any one of two CAT's from the coding strand sequence would cause only the deletion of an I and the reading frame would not be disrupted.

$$\underset{D}{GAT}\ \underset{I}{ATC}\ \underset{I}{ATC}\ \underset{I}{ATC}\ \underset{S}{AGT} \rightarrow \underset{D}{GAT}\ \underset{I}{ATC}\ \underset{I}{ATC}\ \underset{S}{AGT}$$

In an analysis of the populations of spontaneous mutations in *lacI* in *E. coli*, it was found that most of the mutations occurred in a "hotspot" due to structural features around a triple repeated tetramer, which resulted in addition or deletion of one tetramer [13,14]. This was attributed to misalignment of one copy of the repeat on the complement of another. In this case, the result was a frame shift. Structural features such as inter- and intra-strand base pairing in the upstream sequences were proposed to support the misaligned intermediates.

The region around the mutation in *hemA* (FIG. 3A) shows two overlapping seven-base pair repeats (ATCATCA), three overlapping four-base pair repeats (ATCA), and three three-base pair repeats each of ATC and TCA, and two three-base pair repeats of CAT. These structural features explain why loss of anyone of the multiple triplets, ATC, TCA, or CAT, would produce the same base pair-stabilized misaligned intermediate which would result in the same mutation (Fig. 3B).

The parent of HU227, SASX41B, was selected by growth on neomycin and a requirement for ALA for normal growth. Therefore, the mutation in *hemA* in HU227 was selected not to be lethal. A reading frame change in *hemA* would result in the elimination of expression of the downstream gene coding for RF1, an essential protein. The reading frame must not have changed and RF1 must be expressed [15]. A mutation caused by a base-pair substitution might, or a three base pair deletion would, preserve the reading frame, as long as early termination does not take place. Based on the results presented here, we believe that a three-base pair deletion in a 'hotspot" in *hemA* results in the deletion of one I in a conserved region of the gene product, GTR, rendering the enzyme inactive.

The sequence, L_{226}ARSDVVVSAT$_{236}$ in GTR from *M. kandleri*, has the following amino acid sequence L-DVVVS-T which compares with L_{240}READIIISAT$_{250}$ in *E. coli*. L, D and T are highly conserved as are three hydrophobic amino acids after D. The X-ray structure for GTR from *M. kandleri* [16] shows that this sequence is in Domain II, the NADPH binding region and specifically in the diphosphate recognition area [17]. It appears, therefore, that the loss of an I from this sequence is affecting the ability of cofactor to bind to the enzyme or to undergo the conformational changes required to bring it close enough to Domain I to react with glutamyl-tRNA [16].

Acknowledgements

This work was supported by NIH(MBRS/SCORE) GM08168, NIH-RCMI RR03060, and PSC-CUNY #669178.

References

[1] S.I. Beale, Biosynthesis of hemes, in: F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed., American Society for Microbiology, Washington, DC, 1996, pp. 731–748.
[2] W.K. Philipp-Dormston, M. Doss, Z. Naturforsch. 30c (1975) 425–426.
[3] J.-M. Li, C.S. Russell, S.D. Cosloy, Gene 75 (1989) 177–184.
[4] B. Pontoppidan, C.G. Kannangara, Eur. J. Biochem. 225 (1994) 529–537.
[5] T. Nakayashiki, K. Nishimura, R. Tanaka, H. Inokuchi, MGG, Mol. Gen. Genet. 249 (1995) 139–146.
[6] M. Petricek, L. Rutberg, I. Schroder, L. Hederstedt, J. Bacteriol. 172 (1990) 2250–2258.
[7] J. Moser, S. Lorenz, C. Huberschwerlen, A. Rompf, D. Jahn, J. Biol. Chem. 274 (1999) 30679–30685.
[8] M. Drolet, L. Peloquin, Y. Echelard, L. Cousineau, A. Sasarman, MGG, Mol. Gen. Genet. 216 (1989) 347–352.
[9] L.Y. Wang, S. Wilson, T. Elliott, J. Bacteriol. 181 (1999) 6033–6041.
[10] U.C. Vothknecht, C.G. Kannangara, D. v. Wettstein, Phytochemistry 47 (1998) 513–519.
[11] A. Sasarman, M. Suredeanu, T. Horodniceanu, J. Bacteriol. 96 (1968) 1882–1884.
[12] H. Umanoff, C.S. Russell, S.D. Cosloy, J. Bacteriol. 170 (1988) 4969–4971.
[13] P.A. Farabaugh, U. Schmeissner, M. Hofer, J.H. Miller, J. Mol. Biol. 126 (1978) 847–863.
[14] R.M. Schaaper, B.N. Danforth, B.W. Glickman, J. Mol. Biol. 189 (1986) 273–284.
[15] T. Elliott, X. Wang, J. Bacteriol. 173 (1991) 4144–4154.
[16] J. Moser, W.-D. Schubert, V. Beier, I. Bringemeier, D. Jahn, D.W. Heinz, EMBO J. 20 (2001) 6583–6590.
[17] O. Carugo, P. Argos, Proteins, Struct. Funct. Genet. 28 (1997) 10–28.