28237224

This is to certify that the

dissertation entitled

Isolation and Characterization
of the Escherichia coli K12 hemA Gene

presented by

Debra Elizabeth Verkamp

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Microbiology and
Public Health

Major professor

Date 8/10/89

ISOLATION AND CHARACTERIZATION OF THE *ESCHERICHIA COLI* K12 *HEMA* GENE

By

Debra Elizabeth Verkamp

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology and Public Health

1989

ABSTRACT

ISOLATION AND CHARACTERIZATION OF THE ESCHERICHIA COLI K12 HEMA GENE

By

Debra Elizabeth Verkamp

The initial step in heme formation is the biosynthesis of 5-aminolevulinic acid (ALA). Mutations which block ALA synthesis in Escherichia coli have been mapped to two loci, hemA and popC. The original characterization of the hemA mutation did not allow an unambiguous assignment for the role of its gene product in ALA formation, nor was there biochemical evidence that defined the ALA biosynthetic pathway that operates in E. coli. Biosynthesis of ALA occurs in biological systems by either a C4 route (via the ALA synthase reaction, with succinyl CoA and glycine as starting substrates) or a C5 route (a multi-enzyme reaction sequence that transforms glutamate into ALA). It was generally assumed that the ALA synthase pathway operates in bacteria, including E. coli. In this dissertation, I describe the isolation and characterization of the hemA gene from E. coli. Genetic analysis, gene replacement mutagenesis, and maxicell experiments identified the cloned

gene as *hemA*. The nucleotide and predicted amino acid sequences of *hemA* gene bore no resemblance to ALA synthase sequences cloned from other organisms and no ALA synthase activity was detected in *E. coli* extracts, including extracts prepared from strains carrying the cloned *hemA* gene on high copy number plasmids. These results suggested that the C4 route does not normally function in *E. coli* for the synthesis of ALA and that a re-evaluation of the role of the *hemA* gene product in ALA formation was required. Included in this dissertation is a discussion of my results in relation to biochemical studies on ALA biosynthesis in *E. coli* performed by other investigators. The combined studies support the conclusion that ALA synthesis normally occurs in *E. coli* by the C5 route and that the *hemA* gene encodes one of the components of this reaction sequence. Results of S1 nuclease protection experiments showed that the *hemA* mRNA appeared to have two different 5' ends and that a nonoverlapping divergent transcript was present upstream of the putative distal *hemA* transcriptional start site.

to my parents

ACKNOWLEDGMENTS

I would like to acknowledge Barry Chelm for providing support, instruction, and encouragement. I feel very fortunate to have worked with Barry and to have been a member of the Chelm lab.

Special thanks go to Larry Snyder for guidance and advisement during the latter stages of my research. I would also like to recognize the other members of my guidance committee, Mike Thomashow, Shelagh Ferguson-Miller, and Wendy Champness for advice and assistance. Thanks are also extended to Hans Kende and the PRL for providing financial support and to Tom Elliott for sharing results prior to publication.

The various members of the lab that I worked with over several years are duly recognized for making the years of my graduate career always stimulating and occasionally enjoyable. John Scott-Craig, Mary Lou Guerinot, Todd Cotter, Bill Holben and Prudy Hall deserve special thanks for sharing ideas and expertise that helped me to complete various aspects of my research project and writing endeavors.

TABLE OF CONTENTS

	PAGE
List of Tables	viii
List of Figures	ix
CHAPTER 1 Introduction	1
CHAPTER 2 Isolation and Characterization of the *hemA* Gene from *Escherichia coli* K12	
Introduction	15
Materials and Methods	16
Bacterial Strains	16
Culture Media and Growth Conditions	16
DNA Methods	18
Gene Replacement Mutagenesis	18
Cosmid Libraries	19
*Bal*31 Deletions and DNA Sequencing	20
Maxicell Experiments	21
Enzyme Assays	22
Results	22
Isolation of the *hemA* Gene	22
DNA Sequence Analysis	26
Maxicell Experiments	31
Construction of *hemA* Mutant by Gene Replacement	36
P1 Transductions	42

PAGE

ALA Synthase Assays 42

Discussion 43

CHAPTER 3 Fine Structure Analysis of the *Escherichia coli hemA* Transcriptional Unit

Introduction 49

Materials and Methods 51

E. coli Strains and Growth Conditions 51

RNA Isolation 51

S1 Nuclease Mapping 51

Results 53

Discussion 63

CHAPTER 4 Summary and Conclusions 69

REFERENCES 77

LIST OF TABLES

TABLE		PAGE
1	*E. coli* strains	17

LIST OF FIGURES

CHAPTER	FIGURE		PAGE
1	1	The biosynthetic pathway from ALA to protoporphyrin IX	2
1	2	Postulated biosynthetic pathway of ALA formation from glutamate and ALA synthesis via the ALA synthase reaction	11
2	1	Restriction endonuclease map of plasmid subclones of the isolated hemA region	25
2	2	Southern hybridization analysis of cosmids pMRC5, pMRC10, and genomic DNA	28
2	3	Partial nucleotide and predicted amino acid sequence of ORFC	29
2	4	Nucleotide and predicted amino acid sequence of ORFD	30
2	5	Nucleotide sequence of the hemA gene and the predicted amino acid sequence of the hemA polypeptide	33
2	6	Autoradiogram of plasmid polypeptides labeled in the maxicell procedure and schematic depiction of pMR81 and pMR81ΔBgl	35
2	7	Inserts of plasmids used for site-directed mutagenesis	39
2	8	Southern hybridization analysis of EV61 genomic DNA	41

CHAPTER	FIGURE		PAGE
3	1	Nucleotide sequence of the 5' end and the flanking upstream region of the *E. coli* *hemA* gene	55
3	2	S1 nuclease protection analysis	57
3	3	S1 nuclease protection mapping of the ORFD transcriptional start site	60
3	4	High resolution S1 nuclease mapping of the *hemA* transcript	62
3	5	Promoter sequence comparisons	64
3	6	Nucleotide sequence of the *hemA*-*prfA* intergenic region	67

CHAPTER 1

Introduction

Tetrapyrrole derivatives are distributed throughout nature and play a central role in the biology of both prokaryotic and eukaryotic organisms. With rare exceptions, members of this diverse group of compounds are present in all organisms and perform a variety of vital physiological functions. A schematic depiction of the tetrapyrrole biosynthetic pathway by which 5-aminolevulinic acid (ALA) is converted to protoporphyrin IX is presented in Figure 1. Excluding the formation of ALA, which is the first committed intermediate unique to the tetrapyrrole biosynthetic pathway, the enzymatic reactions that comprise this pathway appear to be conserved in biological systems. (For a review of tetrapyrrole synthesis and regulation, see Granick and Beale, 1978.)

Most biologically-active tetrapyrroles exist as metal complexes and can be classified according to the metal which is associated with the porphyrin nucleus. Iron is inserted into the protoporphyrin IX ring to form heme and the insertion of magnesium produces the biochemical precursor of chlorophyll and bacteriochlorophyll in another branch of the pathway. Corrins (including vitamin B_{12}) constitute a class of tetrapyrroles that is complexed to

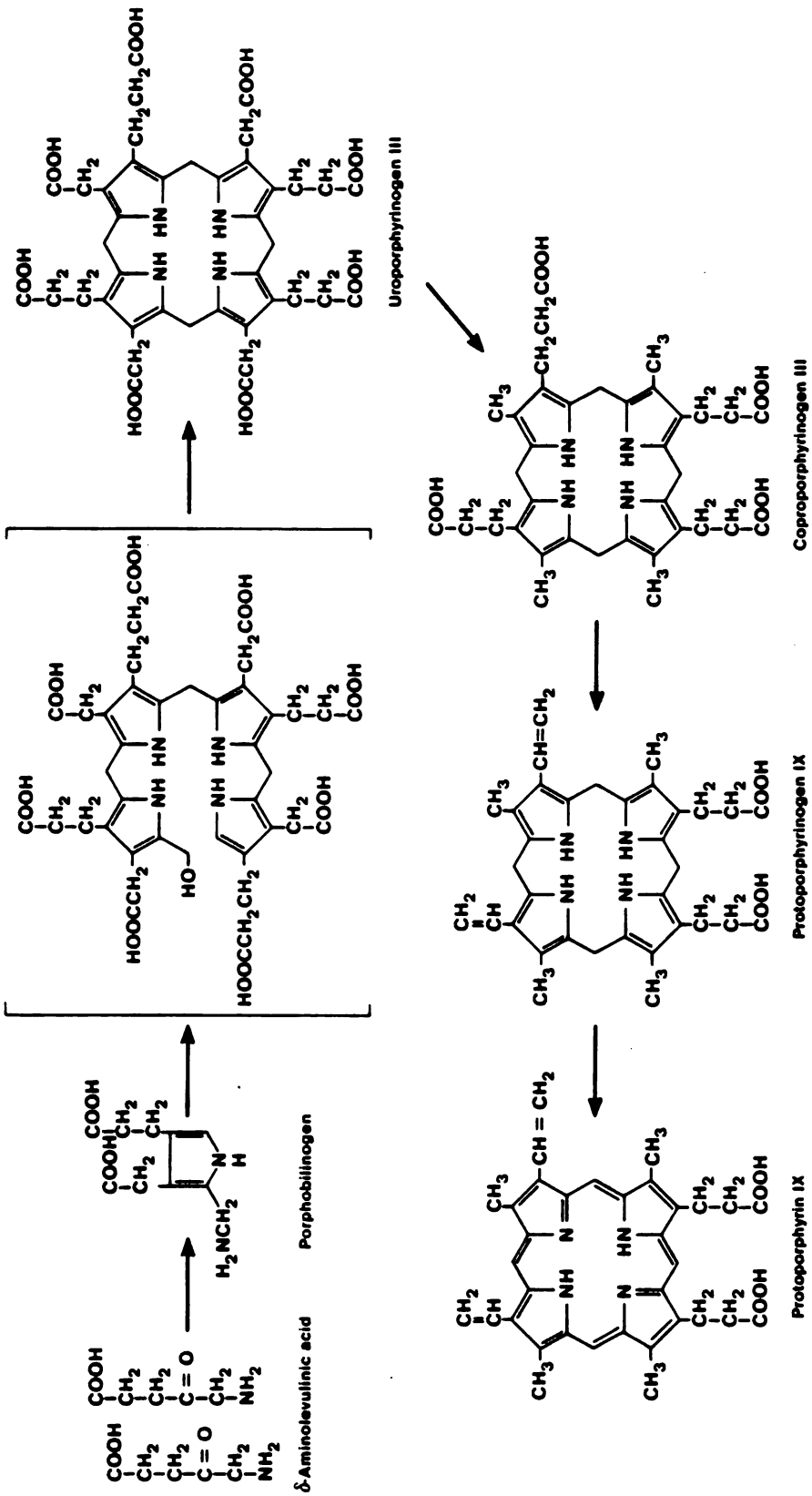


Figure 1. The biosynthetic pathway from ALA to protoporphyrin IX.

cobalt and diverges from the common pathway at the point of uroporphyrinogen III synthesis.

The regulatory aspects of this multi-branched pathway pose interesting biological questions. Requirements for tetrapyrroles vary among organisms and can be distinctly different between tissues of a single organism, depending on the physiological roles that they fulfill in a particular cell. Information about this regulation is rudimentary at present, even in the well-studied bacterium Escherichia coli. Results of investigations into porphyrin synthesis and its regulation in E. coli are discussed below.

Tetrapyrroles function in E. coli in the cellular processes of respiration and oxygen defense. As the prosthetic groups of cytochromes, hemes are essential for the operation of the electron transport chains required for cells to derive energy from non-fermentable carbohydrates (Poole and Ingledew, 1987). Heme also forms the prosthetic group of catalase, an enzyme that converts the toxic aerobic metabolic by-product hydrogen peroxide into oxygen and water (Gottschalk, 1986). By this mechanism, catalase protects cells against the deleterious effects of reduced oxygen intermediates. Siroheme is an iron tetrapyrrole that is derived from uroporhyrinogen III (Jacobs, 1974) and forms the prosthetic groups of both sulfite reductase, an enzyme required for assimilatory sulfate reduction (Kredich, 1987) and nitrite reductase, which reduces

nitrite to ammonia during anaerobic growth (Lin and Kuritzkes, 1987).

Research into the genetics and biochemistry of heme biosynthesis in E. coli dates back to the 1960's. Heme-deficient strains of E. coli were isolated initially by investigators studying small-colony-forming mutants selected on the basis of their resistance to low levels of aminoglycoside antibiotics (Sasarman and Horodniceanu, 1967; Sasarman et al. 1968a). These cells are resistant to aminoglycosides such as streptomycin presumably because they are unable to synthesize a functional respiratory chain and are unable to generate the transmembrane electron potential required for the uptake and accumulation of this class of antibiotics (Chopra and Ball, 1982). Upon further characterization, these small colony, respiratory-deficient mutants were determined to lack catalase and cytochromes because of their inability to synthesize heme. Advances in this area of investigation were somewhat hindered because of the impermeability of E. coli cells to all compounds in the heme biosynthetic pathway beyond ALA (Sasarman et al., 1968a). Further progress was facilitated after McConville and Charles (1979c) isolated heme-permeable mutants that could be manipulated more easily for genetic analysis.

Additional mutants deficient in heme biosynthesis were identified because the cells accumulated large quantities of the heme precursors synthesized prior to the block in the pathway (Cox and Charles, 1973; Powell et al., 1973;

Sasarman et al., 1975; McConville and Charles, 1979a; McConville and Charles, 1979b). The mutations which blocked heme synthesis were characterized based on the pattern of precursor accumulation, the absence of enzyme activities, and the ability of intermediate compounds to rescue the mutant phenotype. Data compiled from these studies provided the genetic scheme for heme synthesis in E. coli summarized below.

At least eight enzymatic steps are required for heme biosynthesis in E. coli. The associated genetic loci have been designated hemA through hemH and correspond to the following enzyme activities: ALA synthase (EC 2.3.1.37) (Evidence provided by this dissertation and other reports argues strongly against this assignment), ALA dehydratase (EC 4.2.1.24), porphobilinogen deaminase (EC 4.3.1.8) (formerly uroporphyrinogen I synthase), uroporphyrinogen III cosynthase (EC 4.2.1.75), uroporphyrinogen decarboxylase (EC4.1.1.37), coproporphyrinogen III oxidase (EC 1.3.3.3), protoporphyrinogen oxidase, and ferrochelatase (EC 4.99.1.1) (see references in Bachmann, 1983). Genes involved in heme biosynthesis map to several scattered regions of the chromosome, and, with the exception of one apparent heme biosynthetic operon (see below), are not physically linked. Two other loci, popC and popA, also appear to be involved in porphyrin synthesis, but have been less well characterized.

A systematic investigation into the regulation of heme biosynthesis has not yet been conducted, but it appears from several reports that this metabolic pathway may be controlled at multiple levels. It has been shown that wild-type strains of E. coli accumulate protoporphyrin and coproporphyrin when grown with excess ALA (Sasarman et al., 1975), suggesting that the formation of ALA may be a rate-limiting step in the pathway. However, since these cells did not accumulate heme, there may be additional control points later in the pathway, possibly at the ferrochelatase step at which iron is incorporated into the tetrapyrrole ring. In a study by Ishida and Hino (1972), the quantity of cytochromes and protoheme was shown to be several-fold higher in aerobically grown cells of E. coli as compared to anaerobically-grown cells, suggesting that some aspect of general oxygen control may contribute to the level of heme and hemoproteins in E. coli. In another study, Poulson et al. (1976) reported that protoporphyrinogen oxidase levels are subject to catabolite repression.

In summary, while genes involved in heme biosynthesis have been mapped in E. coli and some investigations have been conducted into the regulation of heme formation, knowledge about this pathway has lagged behind that of other biosynthetic pathways. This is somewhat surprising in light of the importance of heme to cellular metabolism. In recent years, however, renewed interest in the genetics,

biochemistry, and regulation of heme biosynthesis in E. coli has resulted in substantial progress toward an understanding of this fundamental process.

The second step in the heme pathway is catalyzed by ALA dehydratase (ALAD), coded for by the hemB gene. The reaction involves the condensation of two molecules of ALA to form porphobilinogen (PBG), with the removal of two molecules of water. The hemB gene has been isolated and sequenced (Echelard et al., 1988; Li et al., 1988; Li et al., 1989a) and the predicted amino acid sequence of the polypeptide displays extensive homologies with that of ALAD cloned from yeast (36%), human liver (40%), and rat liver (40%) (Li et al., 1989a). These studies also revealed a relationship between the activity of ALAD and the activity of the enzyme which catalyzes the third step in the pathway, porphobilinogen deaminase (PBG D), the product of the hemC gene. Umanoff et al. (1988) discovered that hemB mutants (lacking ALAD activity) also have extremely low PBG D activity, and, furthermore, that hemA mutants are also deficient in PBG D activity. The activity of PBG D is restored to the hemB mutant by introducing the cloned hemB gene into the cells on a single-copy plasmid or by growing the mutant on media supplemented with PBG. Likewise, PBG D activity is restored to a hemA mutant if it is grown with exogenous supplements of ALA or PBG. Further experiments demonstrated that the availability of PBG probably controls the appearance of PBG D activity at some

posttranscriptional level, possibly by binding to the protein and acting as a prosthetic group to activate the enzyme or by protecting the enzyme from degradation. This is the first demonstration that the product of one gene in the pathway, ALAD, is required to synthesize a biochemical intermediate (PBG) that is necessary for the activity of a second enzyme in the pathway, PBG D. The results imply that a complex and intricate array of regulation may exist to control the activities of enzymes in this pathway.

The hemC and hemD genes, encoding the enzymes PBG D and uroporphyrinogen III synthase, respectively, have been isolated (Thomas and Jordan, 1986; Jordan et al., 1987; Sasarman et al., 1987; Jordan et al., 1988) and are constituents of the Uro operon. This operon is located at approximately 85 min on the E. coli linkage map, adjacent to and transcribed divergently with respect to the cyaA gene, which encodes adenylate cyclase (Sasarman et al., 1987). Nucleotide sequence analysis revealed that the hemC gene is located promoter-proximally and that the start codon of the hemD gene overlaps the terminal nucleotide of the hemC coding sequence. Expression of hemD is dependent upon the hemC promoter, but studies on transcriptional regulation of the hemC hemD operon have not yet been published.

At the third step of the heme pathway, PBG D catalyzes the head-to-tail polymerization of four molecules of the monopyrrole, PBG, into the linear tetrapyrrole

intermediate, preuroporphyrinogen. Uroporphyrinogen III synthase rapidly converts this highly unstable intermediate to uroporphyrinogen III (Jordan et al., 1986), the first cyclic tetrapyrrole formed in the pathway to heme and all other biologically-active tetrapyrrole compounds. The result of the sequential and coordinate actions of these two enzymes is the formation of uroporphyrinogen III. In the absence of uroporphyrinogen III synthase or in the presence of excess PBG D, the nonphysiologic isomer uroporphyrinogen I is formed in a non-enzymatic reaction following the polymerization reaction catalyzed by PBG D (Jordan et al., 1988). The concerted expression of the hemC and hemD gene products, suggested by the physical arrangement of their coding sequences, may exist in order to ensure that their catalytic functions are coordinated and collaborative, analogous to a multi-enzyme complex.

Two additional open reading frames, denoted hemX and hemY have been detected downstream of the hemD coding region (Alefounder et al., 1988; Sasarman et al., 1988). These open reading frames direct the synthesis of polypeptides (Aldea et al., 1988) that may be involved in heme biosynthesis. The hemG locus, which encodes protoporphyrinogen oxidase, has been mapped to this region of the E. coli chromosome (Bachmann, 1983) and is a possible candidate for one of these genes (Jordan et al., 1988). At this time, there have been no further studies on the hemE, hemF and hemH genes.

The primary focus of this dissertation is the molecular characterization of the hemA gene in E. coli, which is required for the synthesis of ALA, the universal precursor of the common tetrapyrrole biosynthetic pathway. The synthesis of ALA occurs in biological systems via at least two different pathways.

In one pathway, the condensation of succinyl-coenzyme A and glycine to form ALA is catalyzed by ALA synthase (Figure 2). This enzyme has been studied extensively in Rhodobacter sphaeroides and other photosynthetic bacteria (e.g., Burnham and Lascelles, 1963; Lascelles, 1964; Burnham, 1970) and in avian and mammalian tissues (e.g., Marver et al, 1966; Bottomly and Smithee, 1968). This pathway has also been demonstrated in other eukaryotic and prokaryotic organisms and the genes which encode ALA synthase have been isolated from the following sources: Rhizobium meliloti (Leong et al., 1982); Bradyrhizobium japonicum (Guerinot and Chelm, 1986); yeast (Keng et al., 1986); Rhodobacter sphaeroides (Tai et al., 1988); Rhodobacter capsulatus (Biel et al., 1988); chicken (Borthwick et al., 1985; Yamamoto et al., 1985; Maguire et al., 1986); mouse (Schoenhaut and Curtis, 1986); and human (Bawden et al., 1987). The enzyme displays significant evolutionary conservation across a broad spectrum of organisms as illustrated by the 48.8% amino acid identity observed between the predicted amino acid sequence of ALA synthase from the symbiotic nitrogen-fixing soil bacterium

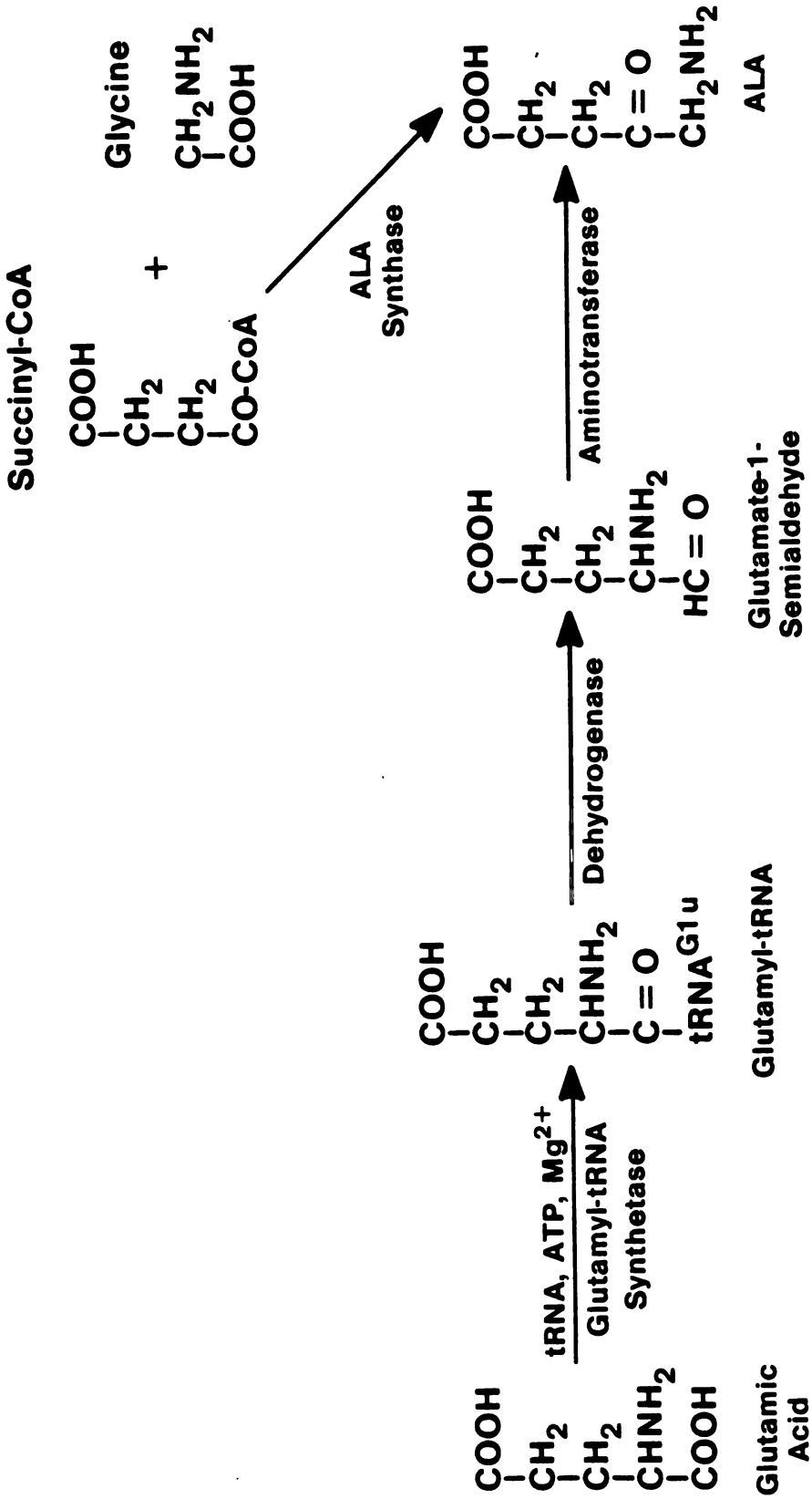


Figure 2. Postulated biosynthetic pathway of ALA formation from glutamate and ALA synthesis via the ALA synthase reaction.

Bradyrhizobium japonicum and from chicken embryonic liver (McClung et al., 1987).

In a second, multi-enzymatic pathway, the intact 5-carbon skeleton of glutamate is converted into ALA. This route has been demonstrated in plant chloroplasts (Beale and Castelfranco, 1974; Beale et al., 1975; Beale, 1976), Chlamydomonas (Wang et al., 1984; Huang and Wang, 1986), cyanobacteria (O'Neill et al., 1988), algae (Weinstein and Beale, 1985; Weinstein et al., 1987), an anaerobic archaebacterium, Methanobacterium thermoautotrophicum, (Gilles et al., 1983) and Euglena (Weinstein and Beale, 1983), which also has ALA synthase activity (Beale et al., 1981). The postulated biosynthetic sequence of catalytic reactions that transforms glutamate into ALA has been elucidated by several laboratories (Wang et al., 1981; Kannangara et al., 1984; Wang et al., 1984; Huang and Wang, 1986; Kannangara et al., 1988; Schneegurt and Beale, 1988) and is depicted schematically in Figure 2. The glutamic acid is activated by the ligation of a $tRNA^{Glu}$ in a reaction that is thought to be identical or analogous to the aminoacylation of tRNA catalyzed by glutamyl tRNA synthetase during protein synthesis. Following this activation at the α-carboxyl, a NADPH-dependent dehydrogenase catalyzes the reduction of this activated glutamate to glutamate- 1-semialdehyde (GSA). In the final step of this pathway, glutamate-1-semialdehyde aminotransferase transfers the amino group to the terminal

carbon to yield ALA. This pathway has been studied extensively in barley chloroplasts, where the component enzymes of this ALA-synthesizing system are soluble and located in the stroma of plastids (Gough and Kannangara, 1977). The glutamic acid tRNA ligase and the glutamate-1-semialdehyde aminotransferase have been purified (Bruyant and Kannangara, 1987; Kannangara et al. 1988), but the dehydrogenase has not yet been purified to homogeneity. The $tRNA^{Glu}$ which participates in ALA synthesis has been sequenced from barley and is encoded by chloroplast DNA (Schon et al., 1986). Isolation of the genes which encode the component enzymes of this pathway has not yet been reported.

At the time the series of experiments described in this dissertation was initiated, it was generally accepted that E. coli synthesized ALA via the ALA synthase reaction. However, reports of ALA synthase activity in E. coli have been inconsistent. While some investigators have failed completely to detect ALA synthase activity in in vitro enzyme assays (McConville and Charles, 1979b), others report values which range from 0.025 nmol/mg protein/hr (Schoenhaut and Curtis, 1986) to 65 nmol/mg protein/hr (Tai et al., 1988). Because of this disparity, ALA synthase activity in E. coli has remained controversial and a clarification of the role of the hemA gene product in ALA synthesis was warranted.

Mutations in E. coli that cause defects in ALA synthesis map to two different loci on the genome; the hemA (Sasarman et al. 1968b) and popC (Powell, et al. 1973) loci at approximately 27 min and 4 min, respectively. HemA mutants display a heme-deficient phenotype that can be rescued by exogenous ALA. However, the exact nature of the defect was not characterized and there was no definitive evidence that the hemA gene encoded ALA synthase. The role of the popC gene product in ALA synthesis has likewise not yet been defined.

The primary focus of this thesis project has been the isolation and characterization of the hemA gene in E. coli, with the aim of defining the role of its gene product in ALA synthesis. Experiments described herein were also designed to characterize the transcriptional unit of the hemA gene and to identify elements which might participate in the control of its expression. I also present the nucleotide sequences of additional open reading frames which have been identified in the region of the genome surrounding the hemA gene.

CHAPTER 2

Isolation and Characterization of the *hemA* gene from *Escherichia coli* K12

Introduction

Heme serves two major functions in the cellular metabolism of *E. coli*: respiration and protection from toxic oxygen metabolites. The initial step in heme biosynthesis is the formation of ALA, the first precursor unique to this pathway. Although mutations which result in ALA auxotrophy have been isolated (*hemA*, *popC*), the roles played by the gene products of the associated loci in the genetics and biochemistry of ALA production lack definition. The objective of experiments recorded in this chapter was the isolation and characterization of the *hemA* gene from *E. coli* with the aim of clarifying its function in ALA biosynthesis. The cloning strategies and molecular genetic techniques used to isolate the *hemA* gene from *E. coli* cosmid libraries are described herein along with the results of experiments that identified the cloned gene as *hemA*. The nucleotide and amino acid sequences of the cloned *E. coli* *hemA* gene display no significant similarity to ALA synthase genes which have been isolated from other organisms.

Materials and Methods

Bacterial Strains. E. coli strains used in this study are listed in Table 1.

Culture Media and Growth Conditions. E. coli strains were grown routinely on LB medium or M9 minimal medium (Davis, et al., 1980). Carbon sources were added as indicated at the following concentrations: glucose .010 M, acetate .020 M, succinate .016 M. Minimal medium was supplemented as required with L-amino acids at 40 ug/ml and thiamine at 1 ug/ml. Strains SASX41B and EV61 (Table 1) were grown routinely on medium supplemented with 50 ug/ml ALA. Antibiotics were incorporated into LB medium when indicated at the following concentrations: ampicillin 80 ug/ml, kanamycin 50 ug/ml and tetracycline 5 ug/ml.

Anaerobic conditions were achieved by sparging the cultures with N_2 in closed, stoppered bottles for 30 min. These cultures were then incubated at 37° in anaerobic GasPak jars (BBL Microbiology Systems). Aerobic cultures were grown in 2 liter flasks at 37°C with vigorous shaking.

For ALA synthase assays, the appropriate E. coli strains were inoculated into 400 ml of M9 glucose, succinate or acetate medium with 4 ml of a mid-log culture grown in the same medium and then incubated aerobically or

Table 1. *E. coli* strains

Strain	Genotype	Source of Reference
CR63	*supD60 lamB63*	Bachmann (1987)
HB101	*hsdS20* (r_B^-,m_B^-) *recA13 leu-6 thi-1 ara-14 proA2 lacY1 galK2 rpsL20* (Sm^r) *xyl-5 mtl-1 supE44*	Boyer and Roullane-Cussoix (1969)
DH5	*endA1 recA1 hsdR17* ($r_K^-m_K^+$) *supE44 thi-1 gyrA96 relA1*?	Hanahan (1985)
JC7623	*thr-1 ara-14 leuB6* Δ(*gpt-proA*)*62 lacY1 tsx-33 supE44 galK2* rac⁻ *hisG4 rfbD1 mgl-51 rpsL31 kdgK51 xyl-5 mtl-1 argE3 thi-1 recB21 recC22 sbcB15 sbcC201*	Bachmann (1987)
JM103	*hsdR4* Δ(*lac-pro*)/F'*traD36 proAB* $lacI^qZ$ *M15 thi strA supE endA sbcB*	Messing (1983)
ORN125	*zcg*::Tn*10 thr-1 leuB thi-1* Δ(*argF-lac*)*U169 malA1 xyl-7 ara-13 mtl-2 gal-6 rpsL fhuA2 supE44*	Spears *et al.* (1986)
SASX41B	HfrPO2A *hemA41 metB1 relA1*	(B.Bachmann)
EV61	Same as JC7623, except *hemA*	This study
MC1024	*araD139* Δ(*ara-leu*)*7697* Δ(*lacZ*)*M15 galU galK strA recA56 srl*::Tn*10*	Casadaban and Cohen (1980)
EV10	Same as SASX41B, except *recA56 srl*::Tn*10*	This study
JK268	*trpE trpA dadR1 purB* (same as JK266)	Hadar *et al.* (1976)

anaerobically. Cells were harvested at the mid-log phase of growth.

P1 transductions were performed according to methods described by Miller (1972).

DNA Methods. Small scale plasmid isolation was performed by the method of Holmes and Quigley (1981) and large scale isolation was by CsCl ethidium bromide equilibrium centrifugation (Clewell and Helinski, 1972). Genomic DNA was extracted and purified using the method of Marmur and Doty (1962). Restriction enzyme digests were performed according to the suppliers' specifications. Ligations and gel electrophoresis were done according to standard techniques (Maniatis, 1982). Restriction enzyme-cleaved DNA was size-fractionated by electrophoresis on 1% agarose gels, transferred to cellulose nitrate filters by the method of Southern (1975), and hybridized to ^{32}P-labeled probes as described (Adams et al., 1984).

Gene Replacement Mutagenesis. A method described by Winans et al. (1985) was used to construct a defined mutation in the hemA gene. The entire hemA gene and flanking sequences were present on a 6.0-kilobase pair (kbp) HindIII fragment that was cloned into pUC19 to create pMR57 (Figure 1). Plasmid pMR61 is a derivative of pMR57 in which a 278-base pair (bp) BglII fragment within the coding region of the hemA gene has been deleted and

replaced by a 1.3-kbp BamHI kanamycin resistance cassette derived from plasmid pRL161 (designated kanamycin cassette C.K1 in Elhai and Wolk, 1988). The nptII gene is oriented in the same direction as that of the putative hemA open reading frame (ORFA in Figure 1). Approximately 2 ug of this plasmid was digested with HindIII and transformed into E. coli strain JC7623 (Table 1). This strain can be transformed readily with linear fragments of DNA because of the loss of RecBCD nuclease activity, but is proficient for recombination due to a compensatory mutation in exonuclease I (sbcB). A marker for drug resistance flanked by E. coli DNA sequences transformed into JC7623 on a linear DNA fragment is maintained only after a double recombination event between the E. coli sequences introduced on the fragment and the homologous chromosomal sequences. Transformants were selected on LB containing kanamycin and ALA, tested for the absence of the ampicillin resistance marker present on vector sequences, and assessed for ALA auxotrophy. Southern hybridization analysis of total genomic DNA digests was performed in order to verify that the deleted copy of hemA with the nptII cassette insert was integrated at only one location in the genome, replacing the chromosomal copy of the hemA gene.

The construction of other plasmids used in gene replacement experiments (see Figure 7) made use of the Ωfragment, a 2.0-kbp antibiotic resistance gene (streptomycin resistance / spectinomycin resistance) that

is flanked by transcription and translation termination signals (Prentki and Krisch, 1984). The cassette was isolated as a SmaI DNA fragment and ligated to pMR19 DNA which had been digested with various restriction endonucleases; EcoRV for pMR19Δ, HpaI (partial) for pMR54 and pMR55, and AvaI (filled in with the Klenow fragment of DNA polymerase I) for pMR19Ω.

Cosmid Libraries. Methods used to prepare cosmid libraries were modified from those published (Adams et al., 1984). The first of two cosmid gene banks was prepared from genomic DNA purified from CR63 and partially digested with Sau3Al. Size-fractionated DNA fragments greater than approximately 20 kbp were ligated to a cosmid vector, pV35 (C. P. Wolk, unpublished), that had been digested with BglII and phosphatase-treated. This vector contains the origin of replication and the ampicillin resistance gene of pBR322. Products of ligation were packaged in vitro, and plated on E. coli strain SASX41B on LB agar containing ampicillin and ALA. Resulting colonies were screened for complementation of the hemA mutation by the resident cosmid. Colonies were transferred to media with and without added ALA, and cosmids were isolated from those which grew well without added ALA.

A second cosmid library was prepared by the method described above from genomic DNA extracted from E. coli strain ORN125. This strain possesses a transposon Tn10 insertion at 26.5 min that is approximately 82%

cotransducible with the hemA locus (Spears, et al., 1986). Purified DNA was partially digested with Sau3A and cloned into the unique BamHI site in pWH4, a cosmid vector that contains a lambda replicative origin (Herrero et al., 1984). Recombinant transformants of HB101 were selected on LB plus kanamycin and screened for tetracycline resistance and complementation of the hemA mutation.

Bal31 Deletions and DNA Sequencing. Bal31 exonuclease was used to generate a series of overlapping deletions spanning the DNA fragment to be sequenced (Poncz et al., 1982). The appropriate DNA fragments were then cloned into M13mp18 or M13mp19 (Yanisch-Perron et al., 1985) with the Bal31-deleted end inserted nearest the primer binding site in the lacZ coding sequence.

Single-stranded DNA template was prepared in JM103 (Messing, 1983) as described (Messing et al., 1981) and sequenced by the dideoxy chain termination method (Sanger et al., 1977). The nucleotide sequence of both strands was determined at least once. Materials and enzymes used for sequencing were purchased from either Bethesda Research Laboratories or United States Biochemical Corporation and used according to the supplier's recommendations.

End-labeled probes used in S1 protocols were sequenced by the method of Maxam and Gilbert (1980).

Maxicell Experiments. Plasmid-encoded polypeptides were identified using the maxicell technique of Sancar et al. (1979) except that E. coli strains HB101 and DH5 were used as the plasmid hosts. Plasmid pMR81 consists of a 2.5-kbp BglI-PvuI insert that includes ORFA plus 895 base pairs of the downstream prfA coding region and 336 bp of upstream sequence (See Figure 6) cloned as a blunt-ended fragment into the unique HincII site of pUC19. This plasmid was digested with BglII and re-ligated to yield pMR81ΔBgl, a plasmid with a deletion of the 278-bp BglII fragment within the ORFA coding sequence.

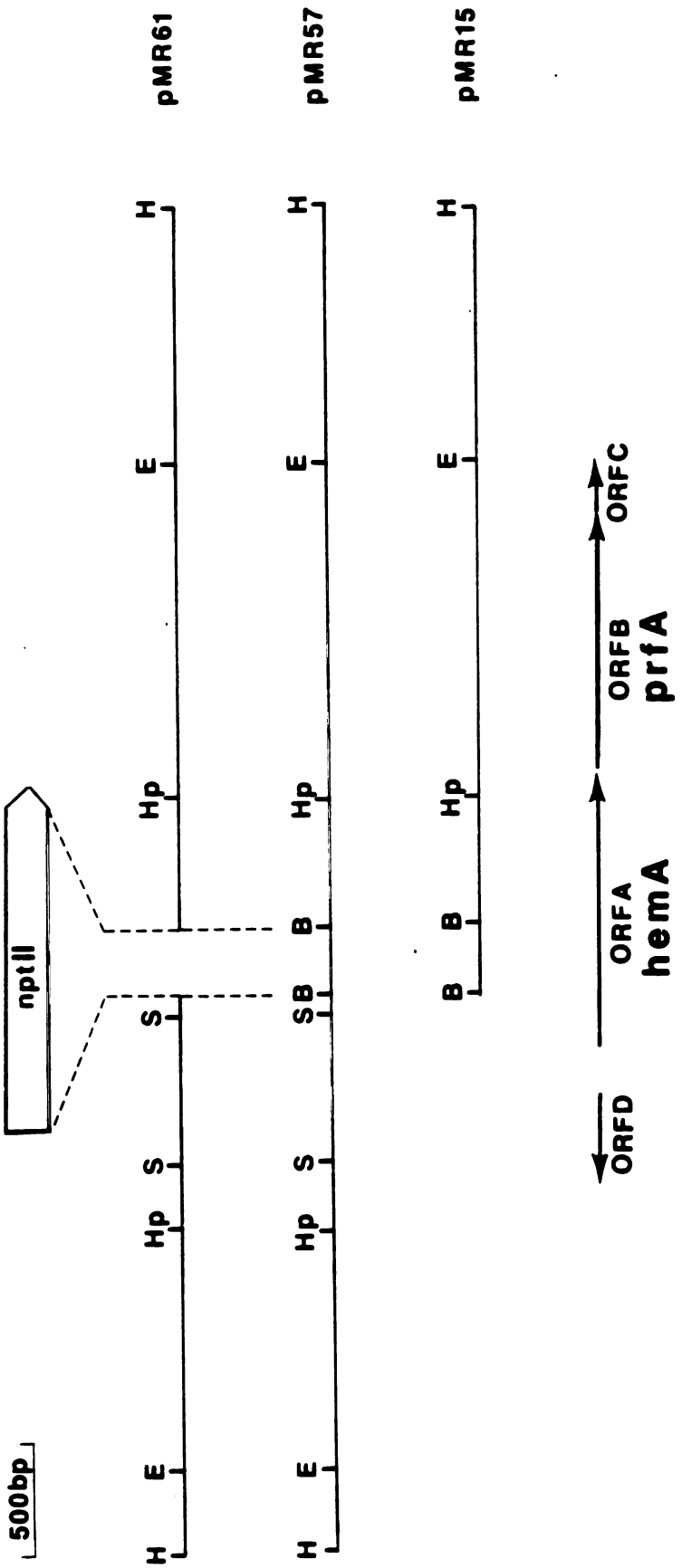
Enzyme Assays. ALA synthase activity was assayed in a 1 ml reaction containing 0.5-2.0 mg of protein for 2 hours at 37°C. Composition of the reaction mixture was that described by Guerinot and Chelm (1986). Cells were pelleted by centrifugation at 4,000 x g for 10 minutes at 4°C, washed one time with Hepes buffer (200 mM Hepes pH 8.0, 33 mM $MgCl_2$, 1 mM 2-mercaptoethanol, 10 mM pyridoxal 5-phosphate) and resuspended in 3-5 ml of Hepes buffer. Cells were disrupted at 4°C by two passages through a French pressure cell at 12,000 lb/in^2. Cell debris was removed by centrifugation at 17,000 x g for 20 min at 4°C. The supernatants were used for determination of protein concentration by the method of Lowry (1951) and for the measurement of enzyme activity. ALA synthase assays were

also performed by the method of Burnham (1970) as modified by Tai *et al*. (1988).

Results

Isolation of the hemA gene. Two schemes were devised to isolate the *hemA* gene on a cosmid vector. The first approach was the complementation of an *E. coli hemA* mutation with a cosmid gene bank prepared from *E. coli* strain CR63. The second strategy was to clone a Tn*10* tetracycline resistance marker that was inserted adjacent to *hemA* on the *E. coli* chromosome in strain ORN125, reasoning that some of the transformants which harbored Tn*10* on a recombinant cosmid would also contain *hemA*. The desired cosmids were isolated as described in Materials and Methods; one cosmid obtained by the first screen (pMRC2) and two cosmids isolated from the second (pMRC5 and pMRC10) were characterized in detail. The apparent complementation of SASX41B by the cosmid pMRC2 required that transformants were selected in the presence of exogenously-added ALA. Transformants could then grow in the absence of ALA at rates comparable to wild-type, suggesting that recombination was required for complementation. This was confirmed by showing that the apparent complementation did not occur in strain EV10, a derivative of SASX41B made RecA$^-$ by P1 transduction from MC1024 (Table 1). In contrast, pMRC5 and pMRC10 complemented SASX41B with no

Figure 1. Restriction endonuclease map of plasmid subclones of the isolated E. coli hemA region. The plasmid pMR15 is a subclone of cosmid pMRC2; the genomic DNA insert is carried on the cosmid vector pV35. The BglII site delineates one boundary of recombinant DNA present on the parent cosmid. The entire hemA and prfA genes and flanking sequences are carried on a 6.0-kbp HindIII DNA fragment that was cloned from pMRC5 into the unique HindIII site of pUC19 to yield plasmid pMR57. The location and direction of transcription of each ORF is indicated by a solid arrow; the DNA sequence was determined for the region delineated by these arrows. A 278-bp BglII fragment within the hemA coding region was deleted and replaced by a nptII cassette to create plasmid pMR61 that was used for gene replacement mutagenesis. Restriction enzymes: B=BglII, E=EcoRI, Hp=HpaI, H=HindIII, S=SalI.

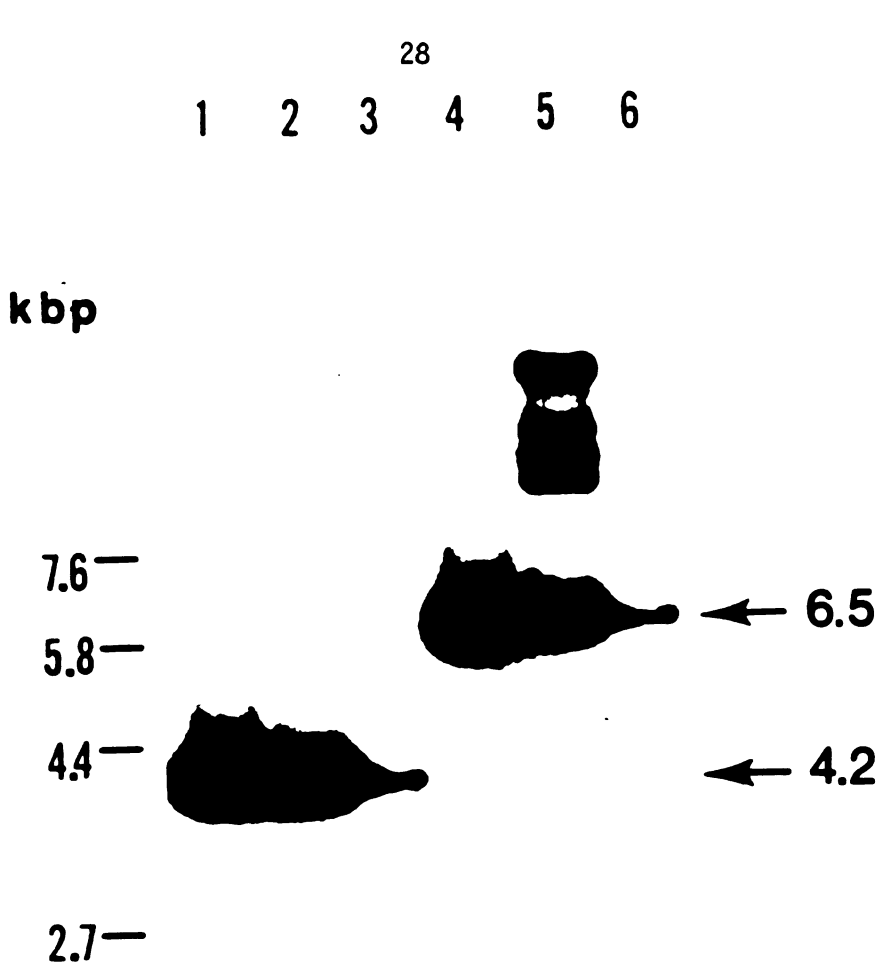

pMR61
pMR57
pMR15
H
E
Hp
B
SB
S
nptII
500bp
ORFA
hemA
ORFB
prfA
ORFC
ORFD

requirement for prior selection in the presence of ALA or recombination function. Plasmid subclones of pMRC2 (pMR15) and pMRC10 (pMR57) were obtained and these exhibited the complementation characteristics of the parent cosmids described above. A partial restriction enzyme map of each of these plasmids is presented in Figure 1. The BglII site at the far left end of pMR15 represents the endpoint of the recombinant DNA contained on its parent cosmid, pMRC2.

The presence of common restriction enzyme sites between the plasmid subclones suggested and Southern hybridization experiments verified that the DNA sequences isolated by the two different schemes were homologous. These data showed that pMR15, a plasmid subcloned from pMRC2, hybridized to the 6.0-kbp HindIII fragment present on cosmids pMRC5 and pMRC10 that was subcloned to create pMR57 (data not shown), as well as to a 4.2-kbp EcoRI fragment and a 6.5-kbp BamHI fragment common to pMRC5, pMRC10 and genomic DNA (Figure 2). The plasmid pMR57 therefore contained the recombinant DNA insert of pMR15, along with additional flanking DNA that appeared to be necessary for hemA activity.

DNA Sequence Analysis. The DNA sequence of the region delineated in Figure 1 was determined. An examination of the sequence data revealed two large ORFs, herein designated A and B. These two ORFs, along with their coding capacity and direction of transcription are shown schematically in Figure 1. If ORFA corresponds to the hemA

Figure 2. Southern hybridization analysis of cosmids pMRC5, pMRC10, and genomic DNA. EcoRI (lanes 1-3) and BamHI (lanes 4-6) digests of pMRC5 (lanes 1 and 4), pMRC10 (lanes 2 and 5), and genomic (lanes 3 and 6) DNA were transferred to cellulose nitrate and hybridized to ^{32}P-labeled pMR15 DNA.


1 2 3 4 5 6
kbp
7.6
5.8
4.4
2.7
6.5
4.2

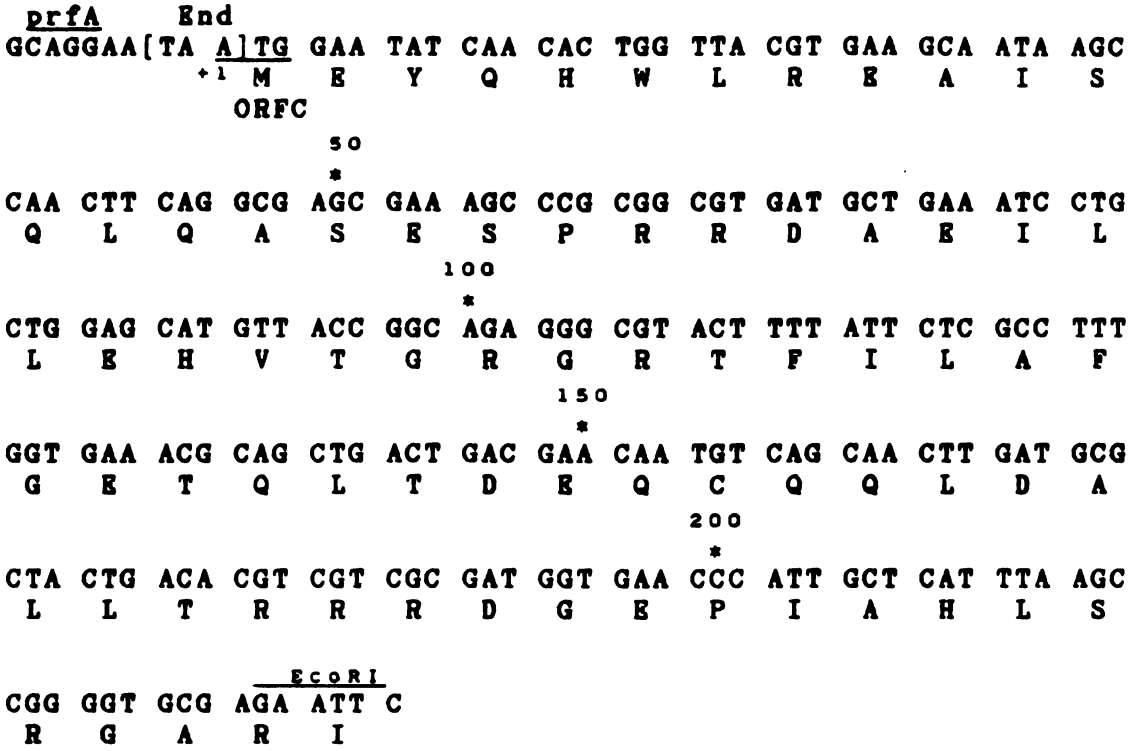


Figure 3. Partial nucleotide and predicted amino acid sequence of ORFC.

```
ACGGTAACGCTAGCATTAAGGGTTATAACTGCAACGTATCTCAAGGACTTGTCATCACT

+1 ORFD                                                      45
ATG CCC CTG CCC GAT TTT CGT CTT ATC CGC CTG CTA CCG CTG GCT
 M   P   L   P   D   F   R   L   I   R   L   L   P   L   A

                                                             90
GCT CTT GTG CTC ACT GCC TGT TCC GTT ACC ACG CCC AAA GGT CCT
 A   L   V   L   T   A   C   S   V   T   T   P   K   G   P

                                                            135
GGC AAA AGC CCG GAT TCG CCA CAA TGG CGT CAG CAT CAG CAA GAC
 G   K   S   P   D   S   P   Q   W   R   Q   H   Q   Q   D

                                                            180
GTG CGC AAT CTT AAT CAG TAT CAG ACT CGC GGC GCG TTC GCT TAT
 V   R   N   L   N   Q   Y   Q   T   R   G   A   F   A   Y

                                                            225
ATT TCT GAC CAA CAA AAA GTG TAC GCC CGC TTT TTC TGG CAG CAA
 I   S   D   Q   Q   K   V   Y   A   R   F   F   W   Q   Q

                                                            270
ACC GGC CAG GAT CGC TAC CGT CTG CTG CTC ACT AAC CCA TTG GGC
 T   G   Q   D   R   Y   R   L   L   L   T   N   P   L   G

                                                            315
AGC ACG GAA CTG GAG CTG AAT GCT CAA CCG GGT AAC GTC GAG TTA
 S   T   E   L   E   L   N   A   Q   P   G   N   V   E   L

                                                            360
GTC GAC AAT AAA GGT CAG CGT TAT ACC GCC GAT GAC GCC GAA GAG
 V   D   N   K   G   Q   R   Y   T   A   D   D   A   E   E

                                                            405
ATG ATT GGC AAA TTG ACC GGA ATG CCA ATT CCG CTC AAC AGC TTG
 M   I   G   K   L   T   G   M   P   I   P   L   N   S   L

                                                            450
CGC CAG TGG ATT TTA GGT TTA CCG GGT GAT GCA ACC GAC TAC AAA
 R   Q   W   I   L   G   L   P   G   D   A   T   D   Y   K

                                                            495
CTG GAC GAC CAG TAC CGC CTG AGC GAA ATT ACC TAC AGC CAG AAT
 L   D   D   Q   Y   R   L   S   E   I   T   Y   S   Q   N

                                                            540
GGC AAA AAC TGG AAG GTT GTT TAT GGT GGT TAT GAC ACC AAA ACG
 G   K   N   W   K   V   V   Y   G   G   Y   D   T   K   T

                                                            585
CAA CCT GCG ATG CCA GCC AAT ATG GAA CTC ACC GAC GGT GGT CAA
 Q   P   A   M   P   A   N   M   E   L   T   D   G   G   Q

CGC ATC AAG TTA AAA ATG GAT AAC TGG ATA GTG AAA TAA TGC GGA
 R   I   K   L   K   M   D   N   W   I   V   K  End

CAC AGT GGC CTC TCC GGC AAA ACT TAA TCT GTT TTT ATA CAT TGC
```

Figure 4. Nucleotide and predicted amino acid sequence of ORFD.

gene, it would explain why pMR15 does not display hemA activity, since it lacks a portion of this open reading frame. The experiments described below supported this interpretation and indicated that ORFA corresponded to the hemA gene. Two other ORFs of undetermined lengths (ORFC and ORFD in Figure 1) are also located in this region and their nucleotide sequences are presented in Figures 3 and 4. These sequences did not display significant similarity to any sequences contained in the GenBank data base and they have not been characterized further.

A computer search of GenBank sequences disclosed that ORFB corresponded to prfA, the gene that encodes peptide chain release factor 1 (RF1), which had been isolated and sequenced previously (Craigen et al., 1985). Since the locus for the RF1 gene has been mapped precisely to 26.7 min on the genetic linkage map (Ryden et al., 1986; Lee et al., 1988), and hemA maps to about 27 min (Bachmann, 1983), the cloned DNA sequences came from the hemA region of the chromosome.

The coding sequence of the putative hemA gene, ORFA, along with upstream flanking sequences are presented in Figure 5. Assuming that the coding region begins at the first ATG, ORFA could theoretically code for a polypeptide consisting of 418 amino acids with a molecular weight of 46,312. The 41-bp intergenic region separating the ORFA stop codon and the first codon of prfA are also presented in the figure.

Figure 5. Nucleotide sequence of the hemA gene and the predicted amino acid sequence of the hemA polypeptide. The sequence of each DNA strand is shown for the region that separates hemA from the putative translational start site of the divergent ORFD. As described in Chapter 3, the proposed transcriptional start sites of the two major hemA RNA species are indicated by asterisks and sequences which comprise possible promoter elements are underlined. The likely promoter sequences of the divergent ORFD transcript are underscored with dashed lines and arrows indicate the nucleotides which correspond to the 5' end of the ORFD transcript.

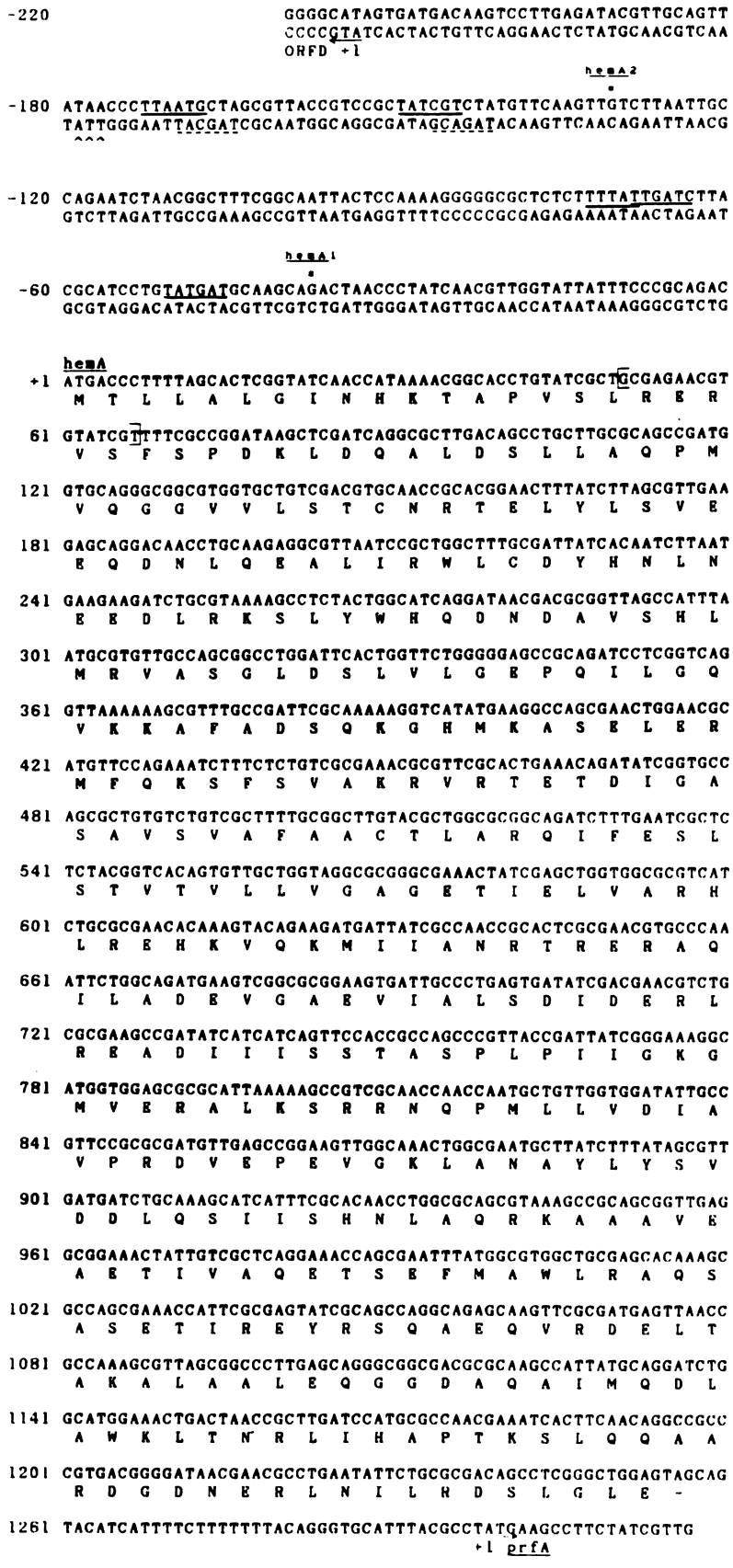
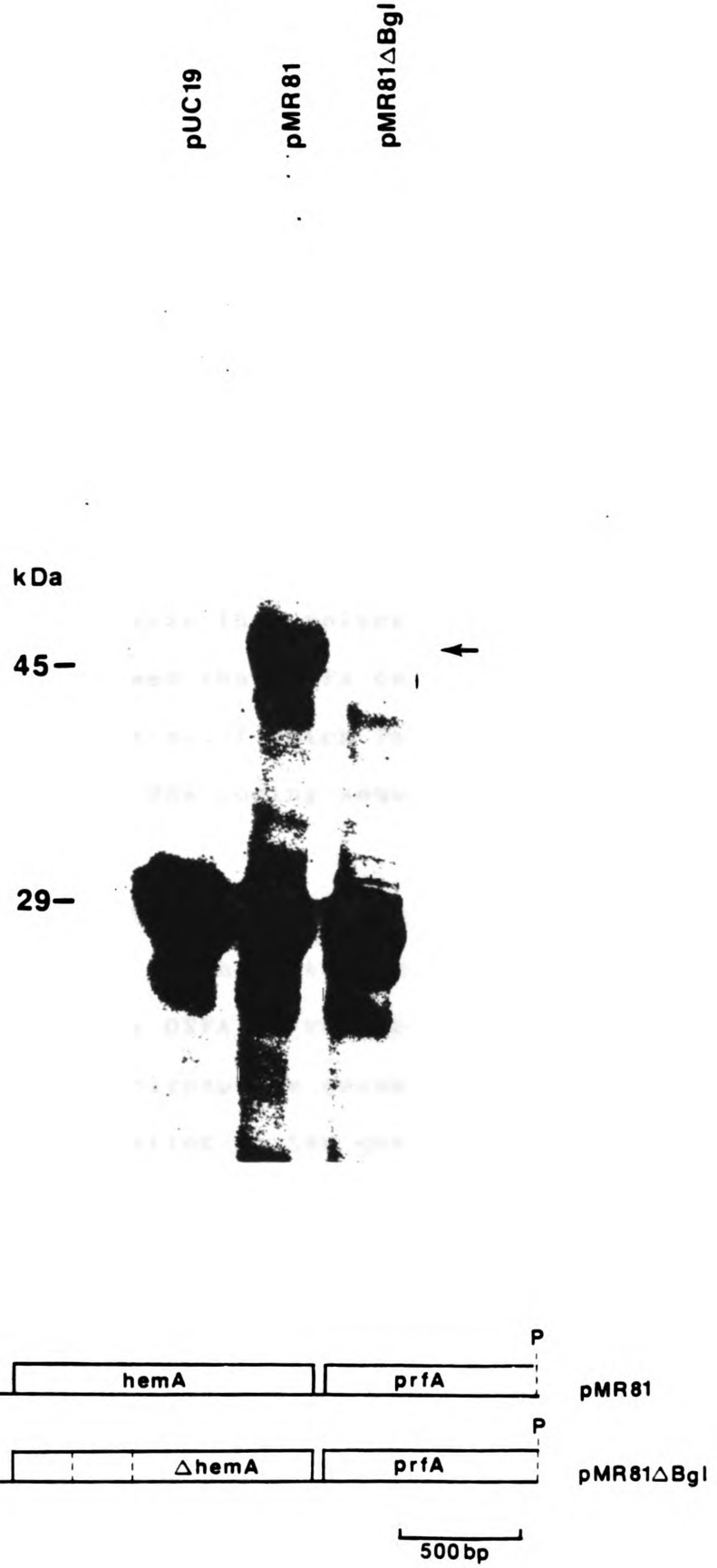

-220
GGGGCATAGTGATGACAAGTCCTTGAGATACGTTGCAGTT
CCCCGTATCACTACTGTTCAGGAACTCTATGCAACGTCAA
ORFD +1
hemA2
-180 ATAACCCTTAATGCTAGCGTTACCGTCCGCTATCGTCTATGTTCAAGTTGTCTTAATTGC
TATTGGGAATTACGATCGCAATGGCAGGCGATAGCAGATACAAGTTCAACAGAATTAACG
-120 CAGAATCTAACGGCTTTCGGCAATTACTCCAAAAGGGGGCGCTCTCTTTTATTGATCTTA
GTCTTAGATTGCCGAAAGCCGTTAATGAGGTTTTCCCCCGCGAGAGAAAATAACTAGAAT
hemA1
-60 CGCATCCTGTATGATGCAAGCAGACTAACCCTATCAACGTTGGTATTATTTCCCGCAGAC
GCGTAGGACATACTACGTTCGTCTGATTGGGATAGTTGCAACCATAATAAAGGGCGTCTG
hemA
+1 ATGACCCTTTTAGCACTCGGTATCAACCATAAAACGGCACCTGTATCGCTGCGAGAACGT
M T L L A L G I N H K T A P V S L R E R
61 GTATCGTTTTCGCCGGATAAGCTCGATCAGGCGCTTGACAGCCTGCTTGCGCAGCCGATG
V S F S P D K L D Q A L D S L L A Q P M
121 GTGCAGGGCGGCGTGGTGCTGTCGACGTGCAACCGCACGGAACTTTATCTTAGCGTTGAA
V Q G G V V L S T C N R T E L Y L S V E
181 GAGCAGGACAACCTGCAAGAGGCGTTAATCCGCTGGCTTTGCGATTATCACAATCTTAAT
E Q D N L Q E A L I R W L C D Y H N L N
241 GAAGAAGATCTGCGTAAAAGCCTCTACTGGCATCAGGATAACGACGCGGTTAGCCATTTA
E E D L R K S L Y W H Q D N D A V S H L
301 ATGCGTGTTGCCAGCGGCCTGGATTCACTGGTTCTGGGGGGAGCCGCAGATCCTCGGTCAG
M R V A S G L D S L V L G E P Q I L G Q
361 GTTAAAAAAGCGTTTGCCGATTCGCAAAAAGGTCATATGAAGGCCAGCGAACTGGAACGC
V K K A F A D S Q K G H M K A S E L E R
421 ATGTTCCAGAAATCTTTCTCTGTCGCGAAACGCGTTCGCACTGAAACAGATATCGGTGCC
M F Q K S F S V A K R V R T E T D I G A
481 AGCGCTGTGTCTGTCGCTTTTGCGGCTTGTACGCTGGCGCGGCAGATCTTTGAATCGCTC
S A V S V A F A A C T L A R Q I F E S L
541 TCTACGGTCACAGTGTTGCTGGTAGGCGCGGGCGAAACTATCGAGCTGGTGGCGCGTCAT
S T V T V L L V G A G E T I E L V A R H
601 CTGCGCGAACACAAAGTACAGAAGATGATTATCGCCAACCGCACTCGCGAACGTGCCCAA
L R E H K V Q K M I I A N R T R E R A Q
661 ATTCTGGCAGATGAAGTCGGCGCGGAAGTGATTGCCCTGAGTGATATCGACGAACGTCTG
I L A D E V G A E V I A L S D I D E R L
721 CGCGAAGCCGATATCATCATCAGTTCCACCGCCAGCCCGTTACCGATTATCGGGAAAGGC
R E A D I I I S S T A S P L P I I G K G
781 ATGGTGGAGCGCGCATTAAAAAGCCGTCGCAACCAACCAATGCTGTTGGTGGATATTGCC
M V E R A L K S R R N Q P M L L V D I A
841 GTTCCGCGCGATGTTGAGCCGGAAGTTGGCAAACTGGCGAATGCTTATCTTTATAGCGTT
V P R D V E P E V G K L A N A Y L Y S V
901 GATGATCTGCAAAGCATCATTTCGCACAACCTGGCGCAGCGTAAAGCCGCAGCGGTTGAG
D D L Q S I I S H N L A Q R K A A A V E
961 GCGGAAACTATTGTCGCTCAGGAAACCAGCGAATTTATGGCGTGGCTGCGAGCACAAAGC
A E T I V A Q E T S E F M A W L R A Q S
1021 GCCAGCGAAACCATTCGCGAGTATCGCAGCCAGGCAGAGCAAGTTCGCGATGAGTTAACC
A S E T I R E Y R S Q A E Q V R D E L T
1081 GCCAAAGCGTTAGCGGCCCTTGAGCAGGGCGGCGACGCGCAAGCCATTATGCAGGATCTG
A K A L A A L E Q G G D A Q A I M Q D L
1141 GCATGGAAACTGACTAACCGCTTGATCCATGCGCCAACGAAATCACTTCAACAGGCCGCC
A W K L T N R L I H A P T K S L Q Q A A
1201 CGTGACGGGGATAACGAACGCCTGAATATTCTGCGCGACAGCCTCGGGCTGGAGTAGCAG
R D G D N E R L N I L R D S L G L E -
1261 TACATCATTTTCTTTTTTTACAGGGTGCATTTACGCCTATGAAGCCTTCTATCGTTG
+1 prfA

Figure 6. Autoradiogram of plasmid polypeptides labeled in the maxicell procedure and schematic depiction of pMR81 and pMR81ΔBgl. The [^{35}S]-methionine-labeled extracts were analyzed on a 10% sodium dodecyl sulfate polyacrylamide gel. The plasmids harbored by the host E. coli strain DH5 are identified above the appropriate lanes and the chromosomal DNA inserts contained on the plasmids are illustrated schematically below. The dashed lines in the hemA coding sequence represent the region of the 278-bp deletion. The electrophoretic mobility of the protein molecular mass markers egg albumin (45,000) and carbonic anhydrase (29,000) are indicated. The arrow identifies the position of the 46-kilodalton polypeptide (middle lane) that is coded for by the hemA gene. A polypeptide of approximately 40 kilodaltons that probably corresponds to the truncated RF1 protein is synthesized by both plasmids. The intact RF1 protein migrates as a 48-kilodalton polypeptide in this gel system (my unpublished result). In addition, a band that migrates below the 28,000 dalton β-lactamase protein is unique to pMR81ΔBgl and probably represents a product of the deleted hemA gene. Restriction endonuclease sites are BglI (B) and PvuI (P).


pUC19
pMR81
pMR81ΔBgl
kDa
45—
29—
B
P
hemA
prfA
pMR81
B
P
ΔhemA
prfA
pMR81ΔBgl
500 bp

Maxicell Experiments. Maxicell labeling experiments were performed in order to determine whether a polypeptide of the anticipated size was produced by ORFA. A polypeptide of approximately 46 kilodaltons was synthesized in maxicell experiments in which ORFA, the presumptive hemA gene, was cloned into pUC19 on plasmid pMR81 (Figure 6). There are no other ORFs present on this plasmid which have a coding capacity sufficiently large to encode this polypeptide. Furthermore, an identical plasmid that was deleted for the 278-bp BglII fragment within ORFA (pMR81Δ Bgl) failed to synthesize this polypeptide (Figure 6). These experiments showed that ORFA can direct the synthesis of a polypeptide the size of which is in agreement with that predicted by the DNA coding sequence.

Construction of hemA Mutant by Gene Replacement. The following experiments demonstrated that the 46-kilodalton polypeptide encoded by ORFA is required for ALA synthesis. If ORFA codes for a polypeptide necessary for ALA synthesis, then a deletion in the coding region of the chromosomal copy of this gene should result in ALA auxotrophy. Plasmid pMR61, (see Materials and Methods and Figure 1), unlike its parent plasmid, pMR57, could no longer complement the hemA mutation and did not produce the 46-kilodalton polypeptide in maxicells (data not shown), indicating that the gene had been interrupted. The gene replacement technique described in Materials and Methods

was used to introduce the inactivated copy of ORFA on plasmid pMR61 into the genome, replacing the chromosomal copy of ORFA. Fourteen of the 14 kanamycin resistant, ampicillin sensitive transformants selected by this procedure displayed the characteristics of ALA auxotrophs. These strains were respiratory-deficient, being unable to grow on a non-fermentable carbon substrate such as acetate in the absence of exogenous ALA. Aerobic growth in LB broth without ALA was undetectable even after prolonged incubation. When strains were grown on LB plates without ALA, they formed minute colonies that displayed no detectable catalase activity. These three phenotypes could be completely rescued by the exogenous addition of ALA to the growth media and thus were due to ALA auxotrophy resulting from the inactivation of ORFA. One of these strains, designated EV61, was characterized further.

In parallel experiments, it was not possible to introduce into the genome a hemA disruption in which the nptII cassette was inserted in the opposite orientation. Unsuccessful attempts were also made to introduce the deletion and insertions depicted in Figure 7 into the genome. Possible reasons why these gene replacements were not recovered are presented in the Discussion.

Total genomic DNA was extracted from EV61 and JC7623 and used in Southern hybridization studies in order to verify that the expected gene replacement event had occurred. The DNA samples were digested with EcoRI,

Figure 7. Inserts of plasmids used for site-directed mutagenesis. A restriction endonuclease map for the plasmid subclone (pMR19) used to construct plasmids for site-directed mutagenesis is shown along with a schematic depiction of genes contained on the plasmid. The dotted line denotes the segment of the *hemA* open reading frame not present on pMR19. The dashed lines represent either a region of *E. coli* DNA deleted and replaced by the Ωfragment or a site into which the Ωfragment was inserted. Restriction sites are *Bgl*II (B), *Eco*RV (V), *Hpa*I (Hp), *Ava*I (A), *Eco*RI (E).

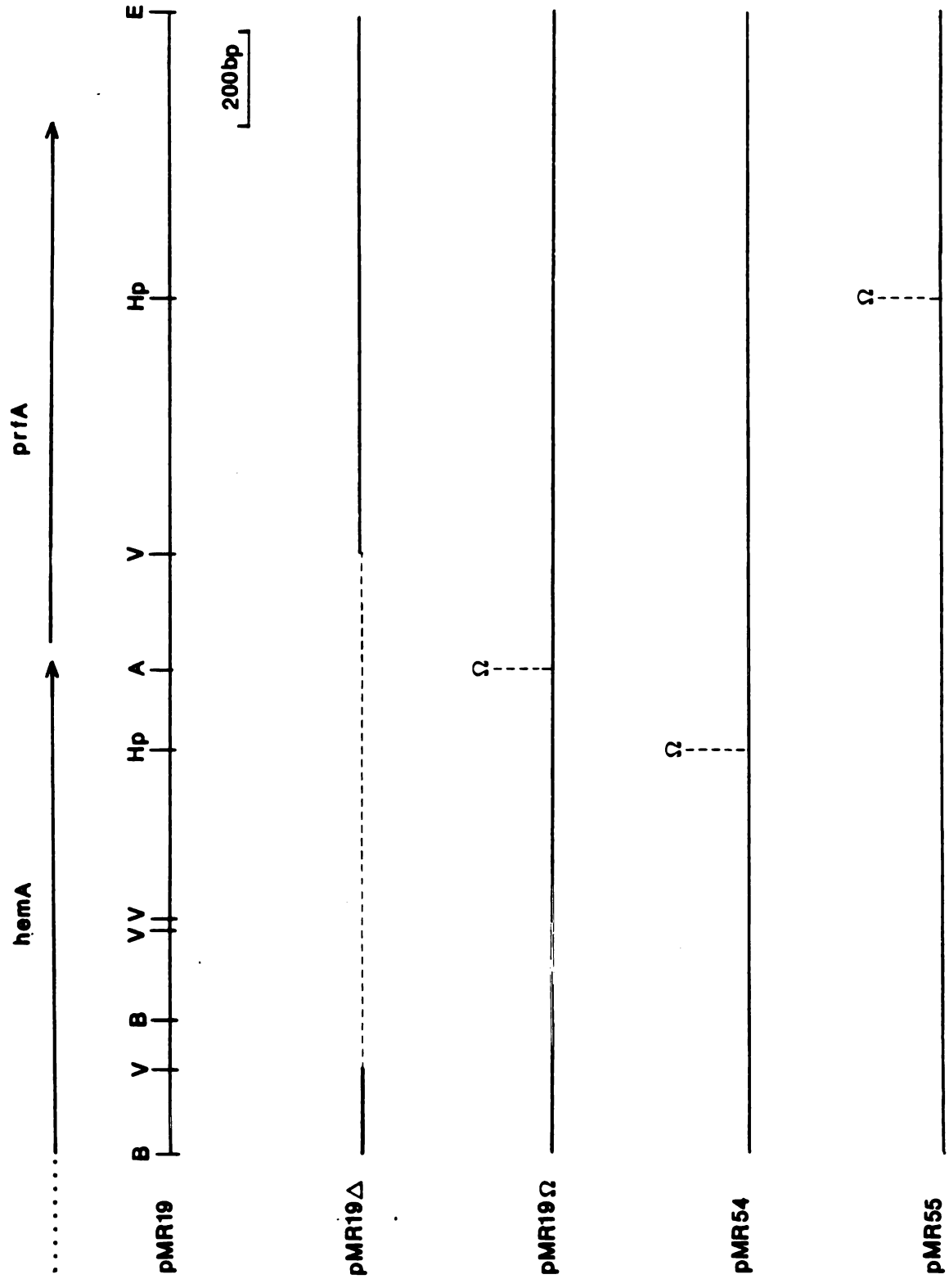

hemA
prfA
B
V
B
V V
Hp
A
Hp
V
Hp
E
200bp
pMR19
pMR19Δ
pMR19Ω
pMR54
pMR55
Ω
Ω
Ω

Figure 8. Southern hybridization analysis of EV61 genomic DNA. *EcoRI* digests of chromosomal DNA from strain JC7623 (lanes 1 and 3) and strain EV61 (lanes 2 and 4) were transferred to cellulose nitrate and hybridized to two radiolabeled probes. Lanes 1 and 2 were hybridized to the 1.4-kbp *nptII* gene and lanes 3 and 4 were hybridized to the 1.8-kbp *HpaI* fragment containing the *E. coli hemA* gene.


1 2
3 4
kbp
5.2—
4.2—

size-fractionated on 1% agarose gels, transferred to cellulose nitrate and hybridized to two different DNA probes: the 1.3-kbp nptII gene and a 1.9-kbp Hpa I fragment which contains the major portion of the ORFA gene along with 700 bp of upstream DNA sequences (Figure 1). The nptII probe hybridized to a 5.2-kpb Eco RI fragment of DNA derived from the mutant but did not hybridize to DNA isolated from the parent strain (Figure 8, lanes 1 and 2). The HpaI probe hybridized to a 4.2-kbp EcoRI fragment in DNA derived from strain JC7623, but, as expected, this fragment was shifted to 5.2 kbp in DNA derived from strain EV61 (Figure 8, lanes 3 and 4). This analysis showed clearly that the wild type chromosomal copy of the ORFA gene was absent in EV61 and had been replaced by the inactive copy.

P1 Transductions. If ORFA corresponds to the hemA gene, then the kanamycin resistance marker present in EV61 should map at approximately 26.7 min and should be co-transduced with trp at approximately 27.5 min (Bachmann, 1983). To test this, a P1 lysate was prepared from JK268 (trpA, trpE, hemA$^+$) and used to transduce strain EV61, selecting for HemA$^+$ (ability to form normal colonies on LB with no ALA supplement) and then testing for the Trp$^-$ phenotype. The hemA and trp markers were co-transduced at a frequency of 15.4% (10/65). All of the HemA$^+$ transductants of EV61 were sensitive to kanamycin, indicating that a wild-type ORFA

allele had replaced the inactive ORFA gene containing the nptII insertion. These genetic data corroborated the DNA sequence analysis that located ORFA to the 26.7 min region of the E. coli linkage map. For unknown reasons, we were unable to transduce the kanamycin resistance marker from EV61 into various recipient strains (see Discussion).

ALA Synthase Assays. Extracts from various strains of E. coli, including strains which harbored the putative hemA gene on high copy number plasmids, were tested for ALA synthase activity by the protocols described in Materials and Methods. Although strains which harbored the plasmid-bourne hemA gene appeared to accumulate tetrapyrroles in liquid culture and on solid media, I was unable to detect activity in any of these strains. However, ALA synthase activity could be readily demonstrated by these methods in an E. coli strain which contained the Bradyrhizobium japonicum hemA gene cloned into pUC9 (data not shown), indicating that the protocols were adequate for ALA synthase detection. Reports of ALA synthase activity in E. coli are inconsistent (Schoenhaur and Curtis, 1986; Tai et al., 1988) and recent experiments indicate that ALA is synthesized by the 5-carbon glutamate route in E. coli (see below). Additional experiments are required to resolve this question.

Discussion

This chapter describes the isolation and nucleotide sequence of the *hemA* gene from *E. coli*. The cloned gene was identified as *hemA* based on the following results.

(i) The gene complemented the *hemA* mutation of SASX41B. The apparent complementation of the mutant *hemA* allele by truncated copies of the cloned gene required RecA function, indicating that the cloned DNA corresponded to *hemA* and eliminating the possibility that the cloned gene suppressed the mutant phenotype of SASX41B by a mechanism involving a different locus.

(ii) The gene was localized to approximately 26.7 min on the *E. coli* linkage map, the site mapped previously for the *hemA* locus. Sequence analysis of the cloned DNA disclosed that the *hemA* gene lies directly upstream of and is transcribed in the same direction as *prfA*, which encodes RF1 and maps to 26.7 min. Transduction mapping using P1 also localized the cloned gene to the 27 min region of the chromosome, and, in addition, the restriction map of the DNA sequence surrounding the isolated *hemA* gene corresponds to the chromosomal restriction map at 26.7 min (Kohara *et al.*, 1987).

(iii) Inactivation of the chromosomal homolog of the cloned gene created an ALA auxotroph. Strain EV61 contained a defined insertion in the *hemA* gene that removed 278 nucleotides of the coding sequence. This strain

exhibited the characteristic phenotype of a hemA strain; it was respiratory-deficient and lacked catalase activity, but formed small colonies on agar plates if provided with a fermentable carbon source. The mutant phenotype was completely reversible when the media was supplemented with ALA. Thus, the hemA gene probably does not encode a polypeptide vital to the survival of E. coli, although the possibility that strain EV61 harbors a compensatory suppressor mutation or that some ALA is synthesized via an alternate pathway has not been ruled out by these experiments.

The inability to construct a gene replacement mutant in which the nptII cassette is transcribed in a direction opposite to that of hemA implies that this insertion may be polar on the downstream prfA gene. (In an extensive genetic analysis, Elliott and Roth [1989] failed to recover hemA insertion and deletion mutations in Salmonella typhimurium, although point mutations at this locus were readily isolated.) This polarity effect may be less severe or absent when the nptII gene is transcribed in the same direction as prfA. In this orientation, the nptII gene may provide a promoter that allows prfA transcription, or, alternatively, the insertion of opposite polarity may introduce a transcriptional terminator which prevents prfA from being expressed from an upstream promoter (hemA or nptII).

Attempts to recover gene replacement mutations in which *prfA* was disrupted by insertions containing transcriptional and translational terminators were unsuccessful, as were efforts to create chromosomal deletions that extended into the *prfA* coding region, suggesting that RF1 is essential for the viability of *E. coli*. This is not surprising, in light of its role in polypeptide chain termination. Termination of protein synthesis and release of the completed polypeptide chain require the activity of release factors which recognize specific termination codons, bind to the ribosome and facilitate peptidyl-tRNA hydrolysis. There are three release factors in *E. coli* which participate in this process: RF1 catalyzes termination at codons UAG and UAA; RF2 specifically recognizes termination codons UGA and UAA; and RF3 enhances the activities of RF1 and RF2 (both binding and dissociation) but does not recognize nonsense codons (Hershey, 1987). In the absence of RF1, amber (UAG) termination codons cannot be recognized, and its function would therefore be essential to survival. This hypothesis is supported by the fact that *prfB*, which encodes RF2 is indispensable for the growth of *E. coli* (Kawakami *et al.*, 1988). The above observations and considerations, in conjunction with the inability to transduce the *hemA* insertion from EV61 into different genetic backgrounds, suggest that some insertions in *hemA* may be polar on *prfA* and that strain EV61 may possess a

secondary mutation that suppresses the polarity effect on prfA expression.

The computer analysis program of Pustell (1984) was used to compare the DNA and amino acid sequences of the E. coli hemA gene and ALA synthase sequences which had been reported from several organisms. The E. coli hemA sequence exhibited no significant similarities with the ALA synthase sequences of B. japonicum, chicken, mouse, or human (data not shown). This observation was unexpected as substantial similarities exist between B. japonicum and chicken (McClung et al., 1987), mouse, and human (my unpublished results) and among the various ALA synthase sequences of eukaryotic origin (Bawden et al., 1987). However, the E. coli hemA gene exhibits extensive similarities with the DNA and amino acid sequences of the cloned S. typhimurium hemA gene (Elliott, 1989). The predicted amino acid sequences share 94.3% identical residues over the entire length of the polypeptides and the amino acids are 96.9% conserved (data not shown).

This lack of similarity to cloned ALA synthase sequences, coupled with the lack of detectable ALA synthase activity in E. coli cell extracts, suggested that the primary route of ALA synthesis in E. coli might not be via ALA synthase. Interestingly, the B. japonicum hemA gene cloned into pUC9 complemented the E. coli hemA strain SASX41B. This result implies that ALA synthase can function in E. coli but does not require that this is the

usual means of ALA biosynthesis in this organism. Two recent reports (Avissar and Beale, 1989; Li *et al.*, 1989b) present convincing evidence that the 5-carbon glutamate pathway operates in *E. coli*. By following the fate of labeled precursor compounds, both groups demonstrated that glutamate, but not glycine, is incorporated into ALA by *E. coli* cells. ALA formation in *in vitro* assays displayed characteristics typical of the 5-carbon pathway, requiring glutamate and reduced pyridine nucleotide and being sensitive to RNase and gabaculine, an inhibitor of the aminotransferase (Hoober *et al.*, 1988). Furthermore, Avissar and Beale (1989) have shown that the enzyme activity absent in the *hemA* strain SASX41B appears to be that of the dehydrogenase. These data are consistent with my results and suggest that the *hemA* gene codes for the dehydrogenase.

In summary, I have presented evidence that the *hemA* gene encodes a polypeptide which functions in ALA synthesis, but which shows no similarity to cloned ALAS sequences. I was unsuccessful in detecting ALAS activity in *E. coli* extracts. Both of these results agree with evidence discussed above which indicates that *E. coli* synthesizes ALA by the 5-carbon glutamate pathway. It is likely that the *hemA* gene codes for the dehydrogenase component of this reaction sequence, which converts the tRNA-activated glutamate to GSA.

CHAPTER 3

Fine Structure Analysis of the *Escherichia coli hemA* Transcriptional Unit

Introduction

This chapter describes experiments performed to characterize the transcriptional unit of the *hemA* gene with the eventual goal of determining whether its expression is regulated at the level of transcription initiation, and, if so, by what controlling elements and cellular signals. As noted in Chapter 1, the amount of heme present in *E. coli* cells appears to be regulated and can vary under different growth conditions (Ishida and Hino, 1972). Similarly, in *S. typhimurium*, the synthesis of cytochromes is reduced under conditions of anaerobic growth on glucose as compared to aerobic growth (Elliott and Roth, 1989). With the exception of the activation and/or stabilization of PBG D by the availability of its substrate, PBG, the molecular mechanisms that effect regulation of heme synthesis have not been characterized. *E. coli* cells grown in media with an excess of exogenous ALA are reported to accumulate and excrete porphyrin compounds (McConville and Charles, 1979a), suggesting that the first step in heme biosynthesis, ALA formation, may be a rate-limiting step for the overall pathway.

The biosynthesis of ALA from glutamate confers additional complexity to the question of regulation because at least three enzymes and a tRNA species are requisite components of this three-step reaction sequence. To date, regulation of the 5-carbon glutamate pathway has been investigated only in plants, where experiments with dark-grown seedlings suggest that ALA formation in leaf plastids controls tetrapyrrole biosynthesis (Kannangara et al., 1988). When supplied with radiolabeled ALA, such leaves convert the exogenous ALA into protochlorophyllide, indicating that the other enzymes in the pathway are present in the plastids and that ALA synthesis is the reaction inhibited in the dark. The molecular mechanism(s) that regulate the ALA synthesizing activity in plastids remains unresolved. The regulation of ALA synthesis in E. coli has not yet been investigated and may be very different from regulation of ALA formation in plants. It will be interesting to determine if the very different biological systems, i.e., a prokaryotic chemoheterotroph and a photosynthetic eukaryotic organelle, share common controlling elements.

As described in Chapter 2, the E. coli hemA gene is required for ALA synthesis and probably encodes the dehydrogenase component of the 5-carbon pathway. As an initial step toward the characterization of expression of this gene, I have localized the transcriptional initiation sites for the hemA message using S1 nuclease protection

procedures. An upstream transcript with opposite polarity to hemA was also revealed during the course of these experiments. Having established a detailed map of the region containing these initiation sites, it will be possible to use the cloned sequences as probes to determine whether the level of transcription of hemA is regulated by the cell and to correlate the level of its expression with the amount of ALA and heme synthesized by the cell in response to different growth environments.

Materials and Methods

E. coli Strains and Growth Conditions. The *E. coli* strains used are described in Chapter 1, Table 1. Routine plasmid construction and maintenance was done in HB101 and M13 manipulations were performed using JM103. The RNA samples used in S1 nuclease protection experiments were prepared from 500 ml cultures of CR63 grown under aerobic or anaerobic conditions (as described in Chapter 2) in M9 medium containing glucose or succinate and harvested at the mid-log phase of growth.

RNA Isolation. RNA was isolated by a CsCl gradient method as described (Adams and Chelm, 1984).

S1 Nuclease Mapping. The method of Berk and Sharp (1977) as modified by Adams and Chelm (1984) was used to

map the 5' transcriptional initiation sites of the hemA mRNA. A 447-bp SalI-BstNI fragment that includes 144 bp of the hemA coding sequence and 303 bp of upstream sequence was used in initial S1 nuclease protection experiments (Figure 1). This fragment was isolated from a polyacrylamide gel, treated with calf intestinal alkaline phosphatase, and then 5'-end-labeled with [γ^{32}P]ATP using T4 polynucleotide kinase. Single-stranded labeled fragments were separated on an 8% denaturing polyacrylamide gel, isolated, and hybridized to 20 ug of RNA according to procedures referenced above. To localize the 5' ends of the transcripts, each of the S1 nuclease-protected fragments was size-fractionated by electrophoresis adjacent to the DNA sequence ladder of its corresponding full-sized probe.

For better resolution of the 5' end of the hemA transcript, a hemA-specific single-stranded DNA probe was synthesized by a primer extension method described by Adams and Chelm (1988). A synthetic oligonucleotide, 5'-ACGATACACGTTCTCGC (Figure 1) was labeled at its 5' end and annealed to a single-stranded template prepared from a recombinant M13mp19 phage containing the 675-bp SalI fragment shown in Figure 1. The hybridized primer was extended, permitting the synthesis of an end-labeled DNA strand that was complementary to the hemA message. This labeled probe was purified and used in S1 nuclease protection experiments as described above. The same

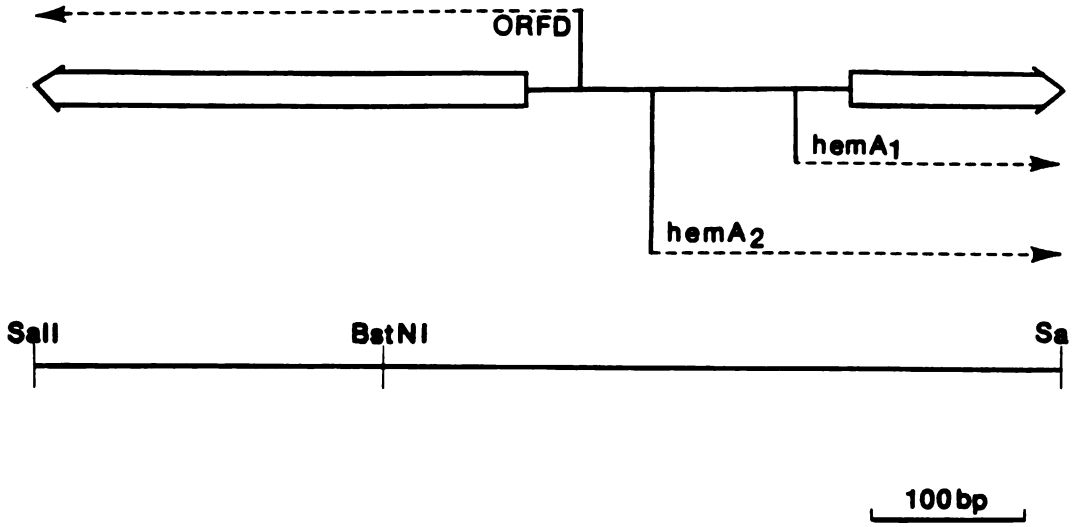
oligonucleotide primer and single-stranded DNA template were used in a dideoxy chain termination reaction to generate a sequence ladder.

Results

Determination of the Transcriptional Start Site of the hemA Gene. The initiation sites of transcripts which originated in the region upstream of the hemA gene were mapped by S1 nuclease protection methods using the SalI-BstNI probe described in Materials and Methods and depicted in Figure 1. Each end-labeled strand of probe was hybridized to RNA prepared from cultures of E. coli strain CR63 grown under several different conditions, selected to represent various energy-generating modes of metabolism. The results of one set of experiments are presented in Figure 2. Two RNA species of different sizes protected the probe corresponding to one of the DNA strands (lanes 6, 7, and 8) and one RNA species protected the probe corresponding to the opposite strand (lanes 3 and 4; a band in lane 2 is present, but faint). These results indicated that within the DNA encompassed by the SalI-BstNI fragment there were two divergent transcriptional units.

Electrophoresis of each of the S1 nuclease protected products adjacent to a Maxam Gilbert sequence ladder generated from its corresponding full-length probe enabled identification of the 5' end points of the presumptive

Figure 1. Nucleotide sequence of the 5' end and the flanking upstream region of the E. coli hemA gene. The open arrows on the schematic diagram of this region indicate the beginning of the hemA and ORFD coding region as predicted by the DNA sequence shown below. The SalI/BstNI fragment that spans the intergenic region was used in initial S1 nuclease protection experiments. The dashed lines and arrows below represent the RNA transcripts that initiate in this region. The nucleotide sequence with the reverse complement of the synthetic oligonucleotide used in high resolution S1 nuclease mapping of the hemA transcript is in brackets. The asterisks indicate the site of transcriptional initiation for hemA and the arrowheads indicate the approximate location of the 5' end of the ORFD transcript. Nucleotides that display homology to the E. coli promoter consensus sequences are underlined for the hemA transcripts and are indicated by lowercase letters for the ORFD transcript. The broken lines underscore the two different sequences with homology to the -35 consensus sequence for hemA1.



ORFD+1
gta
GTAGCAGGCG GATAAGACGA AAATCGGGCA GGGGCATAGT GATGACAAGT

vvv tacgat
CCTTGAGATA CGTTGCAGTT ATAACCCTTA ATGCTAGCGT TACCGTCCGC
-200

hemA2
gcagat *
TATCGTCTAT GTTCAAGTTG TCTTAATTGC CAGAATCTAA CGGCTTTCGG
-150

CAATTACTCC AAAAGGGGGC GCTCTCTTTT ATTGATCTTA CGCATCCTGT
-100

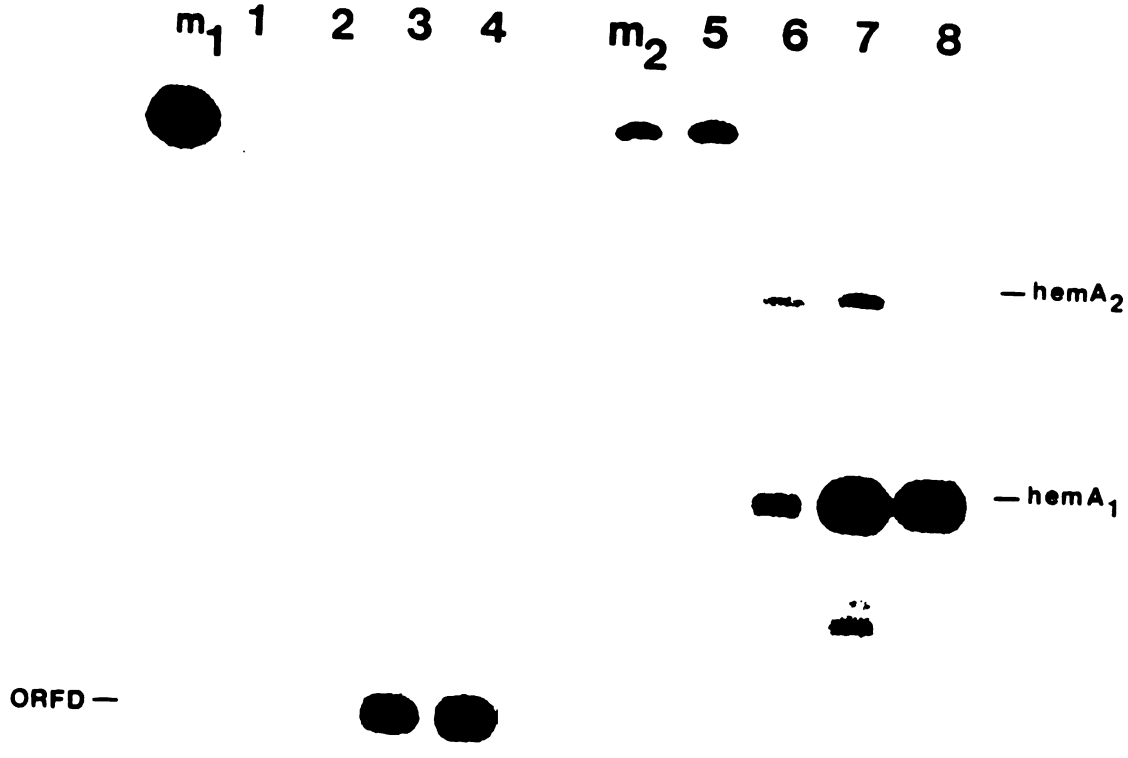
hemA1
*
ATGATGCAAG CAGACTAACC CTATCAACGT TGGTATTATT TCCCGCAGAC
-50

+1 hemA
ATGACCCTTT TAGCACTCGG TATCAACCAT AAAACGGCAC CTGTATCGCT

[GCGAGAACGT GTATCGT]TTT CGCCGGATAA GCTCGATCAG GCGCTTGACA

GCCTGCTTGC GCAGCCGATG GTGCAGGGCG GCGTGGTGCT GTCGACGTGC

Figure 2. S1 nuclease protection analysis. The probes used in S1 nuclease protection experiments were each of the 5' end-labeled strands of the 447-bp BstNI-SalI fragment shown in Figure 1. Each probe was hybridized to the following RNA types: aerobic M9-succinate (lanes 2 and 6); aerobic M9-glucose (lanes 3 and 7); anaerobic M9 glucose (lanes 4 and 8). Control lanes: lanes m_1 and m_2 contain the probes alone; lanes 1 and 5 contain the labeled probes hybridized with salmon sperm DNA.


m1 1 2 3 4 m2 5 6 7 8
— hemA2
— hemA1
ORFD —

transcriptional start sites (results for ORFD are shown in Figure 3). Because the nucleotides which corresponded to the initiation sites for the hemA message could not be determined unambiguously using this probe, a hemA-specific oligonucleotide probe was used for high resolution mapping of this transcript and these data are presented in Figure 4. The nucleotides which correspond to the 5' ends of the hemA message and the divergent ORFD transcript are indicated in Figure 1. The 5' ends of the hemA transcript are separated by 92 nucleotides and are evident in each of four different RNA preparations and in experiments using several different probes (Figure 2, 4, and data not shown). It is therefore unlikely that they are due to artifacts associated with the S1 nuclease digestion or RNA isolation procedure. These results suggested that there may be two functional transcriptional start sites for the hemA message or that a RNA processing (cleavage) event produces the two different RNA species.

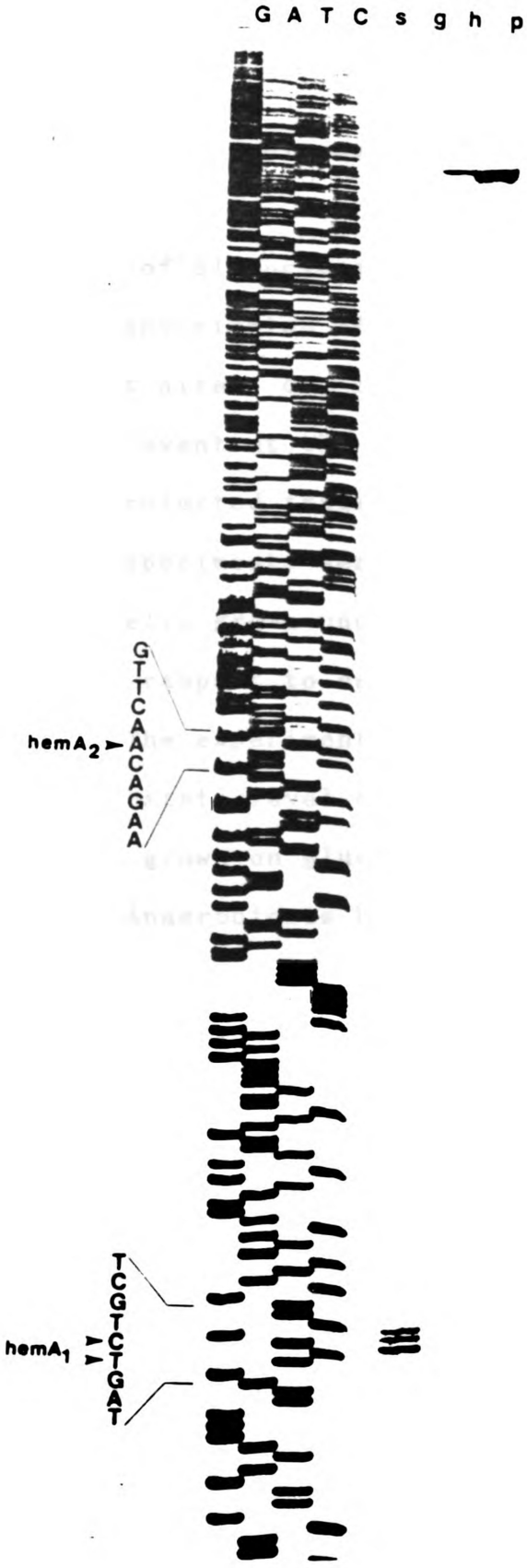
The apparent start sites of the distal hemA transcript and the ORFD transcript do not overlap, but are separated by approximately 45 nucleotides (Figure 1). Whether these sequences play a role in the expression of either transcript has not yet been addressed experimentally.

Nucleotides which comprise possible promoter sequences for each of the transcripts were identified upstream of each of the transcriptional start site. These are depicted

Figure 3. S1 nuclease protection mapping of the ORFD transcriptional start site. The product of an S1 nuclease protection experiment that was carried out using the 447-bp SalI/BstNI fragment that spans the hemA/ORFD intergenic region (Figure 1) was electrophoresed in lane 1, adjacent to the Maxam and Gilbert sequence reactions performed on the labeled ORFD coding strand. The nucleotide sequence depicted on the left is that of the ORFD coding strand and arrows indicate the approximate location of the ORFD transcriptional start site.


1
G
A>C
T+C
C
C
C
A
A
T
A
T
T
G
A
C
G
T
T

Figure 4. High resolution S1 nuclease mapping of the *hemA* transcript. A single-stranded DNA probe was generated by primer extension of the *hemA*-specific synthetic oligonucleotide (see Figure 1) and hybridized to 20 ug of RNA. The protected DNA fragments of S1 nuclease digestion were size-fractionated next to a DNA sequencing ladder which had been produced by primer extension of the same oligonucleotide. s = probe hybridized to RNA extracted from *E. coli* cultures grown aerobically on M9 medium plus succinate; g = probe hybridized to RNA extracted from *E.* cultures grown aerobically on M9 medium plus glucose; h = probe hybridized to an equal amount of heterologous denatured salmon sperm DNA; and p = probe not subjected to S1 nuclease digestion. Nucleotides depicted on the left correspond to the sequence of the *hemA* coding strand. Nucleotides which correspond to the most likely 5' ends of *hemA* transcripts are indicated by arrowheads.


G A T C s g h p
hemA2
G T T C A A C A G A A
hemA1
T C G T C T G A T

in Figure 5 in a comparison with the canonical *E. coli* σ^{70} consensus sequences (Hawley and McClure, 1983).

Discussion

The results of S1 nuclease protection experiments suggest that transcription of the *hemA* message may initiate at two different sites, or, alternatively, that there may be a processing event at the 5' non-translated end of the message. Two protected fragments of different sizes were apparent when experiments were performed with RNA samples isolated from cells grown under several different cultural conditions with respect to carbon substrate and anaerobicity. The experiments shown in Figure 2 suggest that the steady state level of the *hemA*$_1$ transcipt is higher in cells grown on glucose as compared to cells grown on succinate. Anaerobiosis had no apparent effect on the level of this transcript and the level of the *hemA*$_2$ transcript was approximately equal for all of the RNA samples tested. The ORFD transcript was present at levels comparable to those seen for *hemA*$_1$. It will now be possible to perform a thorough investigation into the question of whether *hemA* expression plays a role in the regulation of heme biosynthesis.

The intergenic region upstream of *hemA* is likely to contain controlling elements that affect the expression of ORFD as well as *hemA*. Nucleotides upstream of each of the

E. coli consensus	...TTGACA....17 bp....TATAAT..5-9 bp...*
hemA$_1$	...TTtAtt....16 bp....TATgAT...7 bp....g
hemA$_1$	...TTGAtc....12 bp....TATgAT...7 bp....g
hemA$_2$	...TTaAtg....17 bp....TATcgT...12 bp...g
ORFD	...TaGACg....17 bp....TAgcAT..7-9 bp...t or a

Figure 5. Promoter sequence comparisons. Uppercase letters denote homology between the *E. coli* consensus sequences and the proposed promoter sequences for *hemA* and ORFD. The initiation of transcription is indicated by an asterisk.

transcriptional start sites were identified that display similarity to the canonical -10 and -35 sequences of E. coli σ^{70} consensus promoter elements (Figure 5). For hemA1, two different hexanucleotides are denoted in Figure 5 that show similarity to the -35 promoter element. It should be noted that although the sequence proximal to hemA1 is better matched to the consensus, it is not spaced appropriately with respect to the -10 element, and thus, may not function with the designated -10 element as a RNA polymerase binding site. Whether these or other sequences comprise functional promoters has not been addressed in this study. Regions of divergent transcription are common in E. coli, and in many cases, the product of one transcript plays a regulatory role in the level of expression of transcripts that originate in the region (Beck and Warren, 1988). The possibility that ORFD could play a role in the level of hemA mRNA or vice versa therefore warrants investigation, particularly since the putative RNA polymerase binding sites for ORFD and hemA2 overlap.

A 41-bp region, detailed in Figure 6, separates the stop codon of the hemA open reading frame from the start codon of the prfA gene. Because the evidence presented in Chapter 2 suggests that prfA expression is dependent upon hemA, it is reasonable to hypothesize that this region of nucleotide sequence may contain control signals that affect prfA expression. The stop codon of the hemA gene is an

amber codon, which is recognized specifically by RF1 (Craigen et al., 1985). Whether this observation is of any functional significance has not been determined experimentally, but a model for autoregulation of RF1 has been suggested (Elliott, 1989).

A phenomenon termed translational readthrough has been documented in E. coli, whereby, instead of terminating protein synthesis, an amino acid is inserted for a termination codon by the translational machinery and protein synthesis proceeds until another in-frame termination codon is reached (Hershey, 1987). If protein synthesis continued through the hemA amber termination codon (facilitated by low abundance of RF1), it could continue to an in-frame RF2-specific UGA termination codon, the last two nucleotides of which overlap the AUG initiation codon of prfA (Figure 6). Because there is not a nucleotide sequence with homology to the Shine-Dalgarno consensus sequence upstream of prfA, the translation of prfA may be dependent upon reinitiation allowed by translational readthrough of the hemA gene. According to this hypothesis, under conditions of sufficient or excess RF1 protein in the cell, translation would be terminated at the amber codon, preventing prfA translation.

A putative promoter sequence for the prfA gene was identified by Craigen et al. (1985) (see Figure 6) in what has now been identified as the COOH-terminal region of the hemA coding sequence. There has been no demonstration that

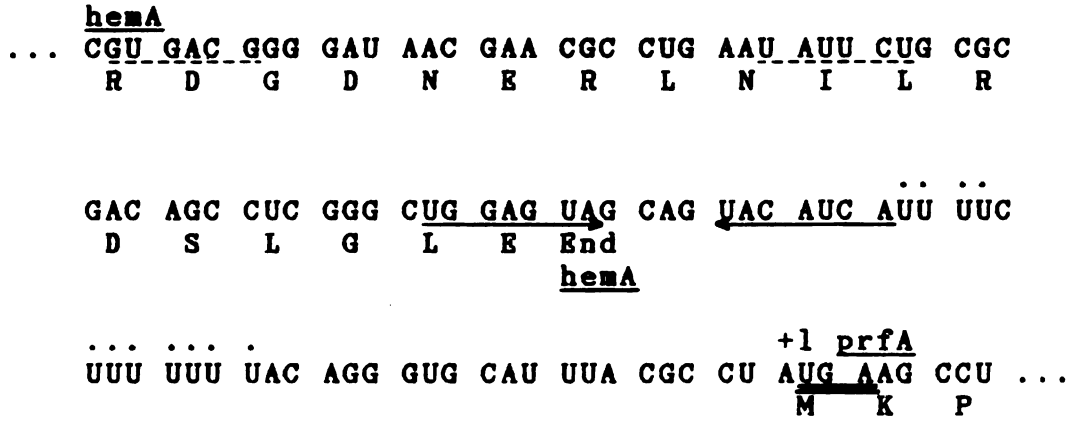


Figure 6. Nucleotide sequence of the hemA-prfA intergenic region. The potential stem-loop structure in the RNA transcript is indicated with arrows and the run of uridine residues is denoted by dots. The promoter sequences proposed for the prfA gene are identified by dashed lines and the downstream hemA in-frame UGA termination codon discussed in the text is underlined twice.

the sequence functions as a promoter for transcriptional initiation, but S1 nuclease protection experiments should clarify this point. Another noteworthy feature of this intergenic region is the presence of sequences which resemble a rho-independent terminator, having a possible hairpin structure followed by a series of U's (Figure 6) (Yager and von Hippel, 1987). Whether this sequence terminates hemA transcription has not been determined.

With these cloned genes available, it will be possible to test the prfA translational reinitiation hypothesis and also to perform further experiments to study the transcriptional control of this putative operon. The proposed AUG codon of the open reading frame downstream of prfA overlaps the termination codon of prfA (Chapter 2, Figure 3), which makes it a likely candidate for a third member of this genetic unit.

The 5-carbon glutamate pathway for ALA biosynthesis appears to share features with protein biosynthesis, with the involvement of a tRNA and a glutamyl tRNA synthetase enzyme. Results of this study indicate that the hemA gene probably encodes a second enzyme in the pathway, the dehydrogenase, which is genetically linked to an enzyme that is a component of the translational machinery of the cell. Further studies into the possible ties between these two fundamental cellular processes could provide informative data on overall metabolic regulation in E. coli.

Chapter 4

Summary and Conclusions

At the outset of this thesis project, it had been shown that the the E. coli hemA gene is required for ALA synthesis, but the initial genetic characterization of this locus did not allow a definitive assignment for its gene product. It was generally assumed that the ALA auxotrophy associated with the hemA mutation was due to the absence of ALA synthase (Bachmann, 1983), the enzyme thought to catalyze ALA synthesis in bacteria. Data on ALA synthase activity in E. coli have been contradictory; some investigators have been unable to detect enzyme activity (McConville and Charles, 1979b; my results) while other laboratories have reported a disparate range of enzyme activities (Ishida and Hino, 1972; Schoenhart et al., 1986; Tai et al., 1988). These ambiguities could be explained if there were inherent difficulties in assaying the enzyme under the conditions of the in vitro assay, possibly due to its instability, requirement for activation, or the presence of inhibitors in bacterial extracts (Jacobs, 1978).

Although it is rare for a fundamental metabolic precursor such as ALA to be synthesized via two independent biosynthetic pathways, Beale and Castelfranco (1974) discovered that a second enzymatic route exists in higher

plants that uses glutamate as the starting substrate for ALA formation. A scheme for the multi-enzymatic reaction sequence has since been proposed by several laboratories (Huang et al., 1984; Huang and Wang, 1986; Bruyant and Kannagara, 1987; Weinstein et al., 1987). This route was originally thought to be limited to plants and algae, but recent studies indicate that it also operates in some eubacteria (Oh-hama et al., 1986; Oh-hama et al., 1988; Rieble and Beale, 1988; Avissar et al., 1989) and a species of archaebacteria (Friedman and Thauer, 1986).

One of the specific goals of this thesis project was to evaluate the role of the hemA gene product in ALA synthesis in E. coli. Chapter 2 details the molecular genetic techniques employed for the isolation, characterization, and DNA sequencing of the hemA gene from E. coli. If the hemA gene encoded ALA synthase, I expected that its nucleotide sequence would display significant similarity to other cloned ALA synthase genes, given that these sequences showed similarity across broad evolutionary boundaries. Unexpectedly, the E. coli hemA gene bore no resemblance to the ALA synthase genes from other organisms. Furthermore, ALA synthase activity was not detectable in E. coli extracts from cells grown under a variety of culture conditions. Activity was also absent in cells which appeared to overproduce porphyrin compounds after being transformed with recombinant multicopy plasmids carrying the cloned hemA gene. A number of explanations could be

put forward to account for these results, the most satisfactory being that an alternate pathway for ALA synthesis normally operates in E. coli.

Two laboratories have recently presented evidence that E. coli uses glutamate as the precursor for ALA formation via the 5-carbon pathway, rather than using glycine and succinyl-CoA as substrates for ALA synthesis via the ALA synthase reaction (Avissar and Beale, 1989; Li et al, 1989b). Avissar and Beale (1989) extended these studies to show that the enzyme activity absent in extracts of the hemA strain SASX41B is that of the dehydrogenase, which catalyzes the second reaction in the 5-carbon pathway, i.e., the reduction of activated glutamate to GSA. Extracts prepared from the hemA strain aminoacylated $tRNA^{Glu}$ that was capable of acting as substrate for ALA formation and converted GSA to ALA by the aminotransferase reaction, demonstrating that the other enzyme activities in the reaction sequence were present. Contrary to previous assumptions, this biochemical characterization of the hemA mutation and the results presented in Chapter 2 argue strongly that the E. coli hemA gene encodes the dehydrogenase component of the 5-carbon pathway, rather than ALA synthase. However, the possibility that the hemA gene product participates in the activation of the dehydrogenase or the regulation of its expression has not been excluded by myself or others.

The discovery of the 5-carbon pathway for ALA synthesis in E. coli has resolved the nature of the hemA mutation, first isolated twenty years ago, and has also opened many avenues for biochemical and genetic investigation. A thorough genetic analysis aimed toward the isolation and characterization of additional ALA auxotrophs is warranted because of the likelihood that there are other loci on the E. coli genome that affect this multi-enzymatic process. The popC mutation which results in ALA auxotrophy, but maps to the 4 min region of the chromosome, requires further investigation in order to resolve the basis for the associated block in ALA synthesis. A mutation analogous to the popC mutation has been isolated in S. typhimurium (Elliott and Roth, 1989) but, as in E. coli, the nature of the lesion has not yet been characterized.

The identification of genes that affect the other components of the 5-carbon pathway might be complicated by the possibility that their actions may not be restricted to the ALA biosynthetic process. For example, if the same glutamyl-tRNA synthetase and $tRNA^{Glu}$ participate in both ALA synthesis and protein translation, then mutations in these genes will not be specific for ALA synthesis. Whether the glutamyl-tRNA synthetase and the $tRNA^{Glu}$ species that activates glutamate for its conversion to ALA are identical to those which participate in protein synthesis needs to be determined. There are four loci in E. coli (gltT, 90 min; gltU, 84.5 min; gltV, 90.5 min; and

gltW, 57 min) that encode glutamyl-tRNA, all of which have the UUC anticodon (Fournier and Ozeki, 1985), and three loci (*gltE*, 81 min; *gltM*, 43 min; *gltX*, 52 min) which affect the activity of glutamyl-tRNA synthetase (see references in Bachmann, 1983).

Characterization of the $tRNA^{Glu}$ required for ALA synthesis has been investigated in other systems. In a study of the tRNA that participates in ALA formation in barley, where there is only one chloroplast-encoded gene for $tRNA^{Glu}$, it was shown that the tRNA which supports ALA formation contains a highly modified UUC glutamate anticodon (Schon *et al.*, 1986). Other glutamate-accepting tRNA species were identified but were shown to have glutamine anticodons (Schon *et al.*, 1988). The glutamate on these mis-charged tRNA species is subsequently converted to glutamine through an amidotransferase reaction, with glutamine or asparagine as the amido donor. These glutamate-accepting tRNA species are not capable of participating in ALA synthesis, suggesting that the anticodon may be an important structural component of tRNA species that support ALA formation. Although the UUC anticodon is required, it is apparently not sufficient for activity, as demonstrated by the results of heterologous reconstitution experiments in which *E coli* $tRNA^{Glu(UUC)}$ was capable of supporting ALA synthesis in isolated enzyme fractions prepared from *Chlamydomonas*, but could not support ALA formation in extracts prepared from

Synechocystis, Chlorella or Euglena, even though these latter extracts could charge the tRNA with glutamate (Kannangara et al., 1988; O'Neill et al., 1988). Furthermore, studies using aplastidic mutants of Euglena indicate that plastid $tRNA^{Glu}$, but not other cellular tRNA species, is effective in supporting ALA biosynthesis (Mayer et al., 1987). There must therefore be other constraints which determine whether the $tRNA^{Glu}$ species can function in ALA synthesis, perhaps dictated by the ability of the dehydrogenase to recognize the glutamyl tRNA. In the blue-green alga Synechocystis sp 6803, two fractions of $tRNA^{Glu(UUC)}$ have been resolved by HPLC that differ in their effectiveness to support ALA formation. Both were also capable of participating in protein synthesis, as determined by in vitro translation experiments (Schneegurt et al., 1988). This raises the question of how the charged $tRNA^{Glu}$ is divided between the two processes of protein synthesis and tetrapyrrole synthesis and whether diversion of available glutamyl-tRNA into ALA biosynthesis is regulated. If the tRNA which supports ALA formation were processed post-transcriptionally by specific modifying enzymes, then the regulation of these enzyme activities could be involved in the control of overall tetrapyrrole synthesis. Further biochemical and genetic experiments are require before these question can be addressed in E. coli.

The glutamyl-tRNA^{Glu} formed in the first step of the reaction sequence is the substrate for the dehydrogenase, which catalyzes the reduction of the activated carboxyl group to yield GSA. The evidence that the hemA gene in E. coli encodes this enzyme was discussed above. If the level of activity of this enzyme controlled the overall pathway, then an overproduction of this enzyme would be predicted to increase the formation of ALA in a cell. Li, et al. (1989b) have shown that an E. coli strain transformed by a multicopy plasmid containing the cloned hemA gene excretes ALA into the medium. I have observed an accumulation of porphyrin compounds in similarly-transformed strains. In light of these results, the possibility that the regulation of dehydrogenase activity could be a control point for ALA and overall tetrapyrrole synthesis warrants investigation.

This control could be exerted at several different levels, including activity of the enzyme (possibly by feedback inhibition by the end product or intermediate products in the pathway), translational regulation of the hemA gene product, or control of transcription of the hemA gene. As an initial step in characterizing hemA expression, a detailed analysis of transcripts which originate in the region upstream of the hemA gene was performed and results of these studies were presented in Chapter 3. It appears that the hemA mRNA has two different 5' ends and that an ORF of opposite polarity is transcribed upstream of the distal hemA transcriptional start site.

With this information available, it will be possible to perform a thorough quantitative analysis of hemA transcription utilizing S1 nuclease or gene fusion techniques in order to determine if expression of hemA is controlled by transcriptional regulation.

One intriguing aspect of the characterization of the hemA gene is its location directly upstream of prfA and the possibility that there is coordinate regulation of the two genes. This genetic link between tetrapyrrole biosynthesis and protein synthesis raises the question of whether regulation of these two fundamental cellular processes share controlling elements. Two other components of the translational machinery of the cell, $tRNA^{Glu}$ and glutamyl-tRNA synthetase also participate in ALA biosynthesis. Future investigation in this area may supply additional insight into the global regulation of cellular metabolism.

LIST OF REFERENCES

LIST OF REFERENCES

Adams, T.H., and B.K. Chelm. 1988. Effects of oxygen levels on the transcription of nif and gln genes in Bradyrhizobium japonicum. J. Gen. Microbiol. 134:611-618.

Adams, T.H., and B.K. Chelm. 1984. The nifH and nifDK promoter regions from Rhizobium japonicum share structural homologies with each other and with nitrogen-regulated promoters from other organisms. J. Mol. Appl. Genet. 2:392-405.

Adams, T.H., C.R. McClung, and B.K. Chelm. 1984. Physical organization of the Bradyrhizobium japonicum nitrogenase gene region. J. Bacteriol. 159:857-962.

Aldea, M., V.F. Maples, and S.R. Kushner. 1988. Generation of a detailed physical and genetic map of the ilv-metE-udp region of the Escherichia coli chromosome. J. Mol. Biol. 200:427-438.

Alefounder, P.R., C. Abell, and A.R. Battersby. 1988. The sequence of hemC, hemD and two additional E. coli genes. Nucleic Acids Res. 16:9871.

Avissar, Y.J., and S.I. Beale. 1989. Identification of the enzymatic basis for 5-aminolevulinic acid auxotrophy in a hemA mutant of Escherichia coli. J. Bacteriol. 171:2919-2924.

Avissar, Y.J., J.G. Ormerod, and S.I. Beale. 1989. Distribution of 5-aminolevulinic acid biosynthetic pathways among phototrophic bacterial groups. Arch Microbiol. 151:513-519.

Bachmann, B.J. 1983. Linkage map of Escherichia coli K-12, Edition 7. Microbiol. Rev. 47:180-230.

Bachmann, B.J. 1987. Derivations and genotypes of some mutant derivatives of Escherichia coli K-12, p.1190-1224. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.

Bawden, M.J., I.A. Borthwick, H.M. Healy, C.P. Morris, B.K. May, and W.H. Elliott. 1987. Sequence of human 5-aminolevulinate synthase cDNA. Nucleic Acids Res. 15:8563.

Beale, S.I. 1976. The biosynthesis of 5-aminolevulinic acid in plants. Phil. Trans. R. Soc. London B 273:99-108.

Beale, S.I., and P. Castelfranco. 1974. The biosynthesis of 5-aminolevulinic acid in higher plants. Plant Physiol. 53:291-296.

Beale, S.I., S.P. Gough, and S. Granick. 1975. Biosynthesis of 5-aminolevulinic acid from the intact carbon skeleton of glutamic acid in greening barley. Proc. Natl. Acad. Sci. USA 73:2719-2723.

Beale, S.I., T. Foley, and V. Dzelzkalns. 1981. Aminolevulinic acid synthase from Euglena gracilis. Proc. Natl. Acad. Sci. USA 78:1666-1669.

Beck, C.F., and R.A.J. Warren. 1988. Divergent promoters, a common form of gene organization. Microbiol. Rev. 52:318-326.

Berk, A.J., and P.A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12:721-732.

Biel, S.W., M.S. Wright, and A.J. Biel. 1988. Cloning of the Rhodobacter capsulatus hemA gene. J. Bacteriol. 170:4382-4384.

Borthwick, I.A., G. Srivastava, A.R. Day, B.A. Pirola, M.A. Snoswell, B.K. May, and W.H. Elliott. 1985. Complete nucleotide sequence of hepatic 5-aminolevulinate synthase precursor. Eur. J. Biochem. 150:481-484.

Bottomley, S.S., and G.A. Smithee. 1968. Characterization and measurement of 5-aminolevulinate synthetase in bone marrow cell mitochondria. Biochim. Biophys. Acta 159:27-37.

Boyer, H.W., and D. Roulland-Dussoix. 1969. A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. 41:459-472.

Bruyant, P., and C.G. Kannangara. 1987. Biosynthesis of 5-aminolevulinate in greening barley leaves. VIII. Purification and characterization of the glutamate-tRNA ligase. Carlsberg Res. Commun. 52:99-109.

Burnham, B.F. 1970. 5-Aminolevulinic acid synthase. Methods Enzymol. 17:195-204.

Burnham, B.F., and J. Lascelles. 1963. Control of porphyrin biosynthesis through a negative-feedback mechanism. Biochem. J. 87:462-472.

Casadaban, M.J., and S.N. Cohen. 1980. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. J. Mol. Biol. 138:179-207.

Chopra, I. and P. Ball. Transport of antibiotics into bacteria. 1982. Advances in microbial physiology. 23:183-217.

Clewell, D.B., and D.R. Helinski. 1972. Effect of growth conditions on the formation of the relaxed complex of supercoiled ColE1 deoxyribonucleic acid and protein in Escherichia coli. J. Bacteriol. 110:1135-1146.

Cox, R. and H.P. Charles. 1973. Porphyrin-accumulating mutants of Escherichia coli. J. Bacteriol. 113:122-132.

Craigen, W.J., R.G. Cook, W.P. Tate, and C.T. Caskey. 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc. Natl. Acad. Sci. USA 82:3616-3620.

Davis, R.W., D. Botstein, and J.R. Roth. 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Echelard, Y., J. Dymetryszyn, M. Drolet, and A. Sasarman. 1988. Nucleotide sequence of the hemB gene of Escherichia coli K12. Mol. Gen. Genet. 214:503-508.

Elhai, J., and C.P. Wolk. 1988. A versatile class of positive-selection vectors based on the nonviability of palindrome-containing plasmids that allows cloning into long polylinkers. Gene 68:119-138.

Elliott, T. 1989. Cloning, genetic characterization, and nucleotide sequence of the hemA-prfA operon of Salmonella typhimurium. J. Bacteriol. 171:3948-3960.

Elliott, T. and J.R. Roth. 1989. Heme-deficient mutants of Salmonella typhimurium: Two genes required for ALA synthesis. Mol. Gen. Genet. 216:303-314.

Fournier, M.J., and H. Ozeki. 1985. Structure and organization of the transfer ribonucleic acid genes of Escherichia coli K-12. Microbiol. Rev. 49:379-397.

Friedmann, H.C., and R.K. Thauer. 1986. Ribonuclease-sensitive 5-aminolevulinic acid formation from glutamate in cell extracts of Methanobacterium thermoautotrophicum. FEBS Lett. 207:84-88.

Gillis, H., R. Jaenchen, and R.K. Thauer. 1983. Biosynthesis of 5-aminolevulinic acid in Methanobacterium thermoautotrophicum. Arch. Microbiol. 135:237-240.

Gottschalk, G. 1986. Bacterial metabolism, second edition. Springer-Verlag, New York, New York.

Gough, S.P., and C.G. Kannangara. 1977. Synthesis of 5-aminolevulinate by a chloroplast stroma preparation from greening barley leaves. Carlsberg Res. Commun. 42:459-464.

Granick, S. and S.I. Beale. 1978. Hemes, chlorophylls, and related compounds: biosynthesis and metabolic regulation. Adv. Enzymol. 46:33-203.

Guerinot, M.L., and B.K. Chelm. 1986. Bacterial 5-aminolevulinic acid synthase activity is not essential for leghemoglobin formation in the soybean/Bradyrhizobium japonicum symbiosis. Proc. Natl. Acad. Sci. USA 83:1837-1841.

Hadar, R., A. Slonim, and J. Kuhn. 1976. Role of D-tryptophan oxidase in D-tryptophan utilization by Escherichia coli. J. Bacteriol. 125:1096-1104.

Hanahan, D. 1985. Techniques for transformation of E. coli, p.109-135. In D.M. Glover (ed.), DNA Cloning, Vol.I, IRL Press, Oxford.

Hawley, D.K., and W.R. McClure. 1983. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. 11:2237-2255.

Herrero, A., J. Elhai, B. Hohn, and C.P. Wolk. 1984. Infrequent cleavage of cloned Anabaena variabilis DNA by restriction endonuclease from A. variabilis. J. Bacteriol. 160:781-784.

Hershey, J.W.B. 1987. Protein synthesis, p.613-647. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.

Holmes, D.S., and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114:193-197.

Hoober, J.K., A. Kahn, D.E. Ash, S. Gough, and C.G. Kannangara. 1988. Biosynthesis of Δaminolevulinate in greening barley leaves. IX. Structure of the substrate, mode of gabaculine inhibition, and the catalytic mechanism of glutamate-1-semialdehyde aminotransferase. Carlsberg Res. Commun. 53:11-25.

Huang, D., and W. Wang. 1986. Chlorophyll biosynthesis in Chlamydomonas starts with the formation of glutamyl-tRNA. J. Biol. Chem. 261:13451-13455.

Ishida, A., and S. Hino. 1972. Effect of oxygen on cytochrome pattern and heme synthesis in Escherichia coli. J. Gen. Appl. Microbiol. Tokyo 18:225-237.

Jacobs, N.J. 1974. Biosynthesis of heme, p. 125-148. In J.B. Nielands (ed.), Microbial iron metabolism. A comprehensive treatise. Academic Press, Inc., New York.

Jordan, P.M., B.I.A. Mgbeje, A.F. Alwan, and S.D. Thomas. 1987. Nucleotide sequence of hemD, the second gene in the hem operon of Escherichia coli K-12. Nucleic Acids Res. 15:10583.

Jordan, P.M., B.I.A. Mgbeje, S.D. Thomas, and A.F. Alwan. 1988. Nucleotide sequence for the hemD gene of Escherichia coli encoding uroporphyrinogen III synthase and initial evidence for a hem operon. Biochem. J. 249:613-616.

Jordan, P.M., L. Marshall, and S. Thomas. 1986. The enzymes of uroporphyrinogen III biosynthesis, p. 83-93. In Y. Nordmann (ed.), Porphyrins and porphyrias. Colloque INSERM/John Libbey Eurotext Ltd.

Kannangara, C.G., S.P. Gough, P. Bruyant, J.K. Hoober, A. Kahn, and D. von Wettstein. 1988. $tRNA^{Glu}$ as a cofactor in 5-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis. TIBS 13:139-143.

Kannangara, C.G., S.P. Gough, R.P. Oliver, and S.K. Rasmussen. 1984. Biosynthesis of 5-aminolevulinate in greening barley leaves. VI. Activation of glutamate by ligation to RNA. Carlsberg Res. Commun. 49:417-437.

Kawakami, K., I. Toshifumi, and Y. Nakamura. 1988. Conditionally lethal and recessive UGA-suppressor mutations in the prfB gene encoding peptide chain release factor 2 of Escherichia coli. J. Bacteriol. 179:5378-5381.

Keng, T., E. Alani, and L. Guarante. 1986. The nine amino-terminal residues of 5-aminolevulinate synthase direct B-galactosidase into the mitochondrial matrix. Mol. Cell. Biol. 6:355-364.

Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50:495-508.

Kredich. N.M. 1987. Biosynthesis of cysteine, p. 412-418. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.

Lascelles, J. 1964. Tetrapyrrole biosynthesis and regulation. W.A. Benjamin Inc., New York.

Lee, C.C., Y. Kohara, K. Akiyama, C.L. Smith, W.J. Craigen, and C.T. Caskey. 1988. Rapid and precise mapping of the Escherichia coli release factor genes by two physical approaches. J. Bacteriol. 170:4537-4541.

Leong, S.A., G.S. Ditta, and D.R. Helinski. 1982. Heme biosynthesis in Rhizobium. Identification of a cloned gene coding for 5-aminolevulinic acid synthetase from Rhizobium meliloti. J. Biol. Chem. 257:8724-8730.

Li, J., C.S. Russell, and S.D. Cosloy. 1989a. The structure of the Escherichia coli hemB gene. Gene 75:177-184.

Li, J., H. Umanoff, R. Proenca, C.S. Russell, and S.D. Cosloy. 1988. Cloning of the Escherichia coli K-12 hemB gene. J. Bacteriol. 170:1021-1025.

Li, J., O. Brathwaite, S.D. Cosloy, and C.S. Russell. 1989b. 5-Aminolevulinic acid synthesis in Escherichia coli. J. Bacteriol. 171:2547-2552.

Lin, E.C.C., and D.R. Kuritzkes. 1987. Pathways for anaerobic electron transport, p. 201-221. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.

Lowry, O.H., N.J. Rosebrough, A.L. Farr, and R.J. Randall. 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193:265-275.

Maguire, D.J., A.R. Day, I.A. Borthwick, G. Srivastava, P.L. Wigley, B.K. May and W.J. Elliott. 1986. Nucleotide sequence of the chicken 5-aminolevulinic acid synthase gene. Nucleic Acids Res. 14:1379-1391.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.

Marmur, J., and P. Doty. 1962. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. Mol. Biol. 5:109-118.

Marver, H.S., A. Collins, D.P. Tschudy, and J. Rechcigl. 1966. 5-Aminolevulinic acid synthetase. II. Induction in rat liver. J. Biol. Chem. 241:4323-4329.

Maxam, A.M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499-560.

Mayer, S.M., J.D. Weinstein, and S.I. Beale. 1987. Enzymatic conversion of glutamate to 5-aminolevulinate in soluble extracts of Euglena gracilis. J. Biol. Chem. 262:12541-12549.

McClung, C.R., J.E. Somerville, M.L. Guerinot, and B.K. Chelm. 1987. Structure of the Bradyrhizobium japonicum gene hemA encoding 5-aminolevulinic acid synthase. Gene 54:133-139.

McConville, M.L., and H.P. Charles. 1979a. Isolation of haemin-requiring mutants of Escherichia coli K12. J. Gen. Microbiol. 113:155-164.

McConville, M.L., and H.P. Charles. 1979b. Mutants of Escherichia coli K12 accumulating porphobilinogen: a new locus, hemC. J. Gen. Microbiol. 111:193-200.

McConville, M.L., and H.P. Charles. 1979c. Mutants of Escherichia coli K12 permeable to haemin. J. Gen. Microbiol. 113:165-168.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101:20-78.

Messing, J., R. Crea, and P.A. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309-321.

Miller, J.H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Oh-hama, T., H. Seto, and S. Miyachi. 1986. ^{13}C-NMR evidence of bacteriochlorophyll a formation by the C_5 pathway in Chromatium. Arch. Biochem. Biophys. 246:192-198.

Oh-hama, T., N.J. Stolowich, and A.I. Scott. 1988. 5-Aminolevulinic acid formation from glutamate via the C_5

pathway in Clostridium thermoaceticum. FEBS Lett. 228:89-93.

O'Neill, G.P., D.M. Peterson, A. Schon, M. Chen, and D. Soll. 1988. Formation of the chlorophyll precursor 5-aminolevulinic acid in cyanobacteria requires aminoacylation of a $tRNA^{Glu}$ species. J. Bacteriol. 170:3810-3816.

Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz, and S. Surrey. 1982. "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA 79:4298-4302.

Poole, R.K., and W.J. Ingledew. 1987. Pathways of electrons to oxygen, p. 170-200. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.

Poulson, R., K.J. Whitlow, and W.J. Polglase. 1976. Catabolite repression of protoporphyrin IX biosynthesis in Escherichia coli K-12. FEBS Lett. 62:351-353.

Powell, K.A., R. Cox, M. McConville, and H.P. Charles. 1973. Mutations affecting porphyrin biosynthesis in Escherichia coli. Enzyme 16:65-73.

Prentki, P., and H.M. Krisch. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene 29:303-313.

Pustell, J. and F.C. Kafatos. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res. 12:643-655.

Rieble, S., and S.I. Beale. 1988. Transformation of glutamate to 5-aminolevulinic acid by soluble extracts of Synechocystis sp. PCC6803 and other oxygenic prokaryotes. J. Biol. Chem. 263:8864-8871.

Ryden, M., J. Murphy, R. Martin, L. Isaksson, and J. Gallant. 1986. Mapping and complementation studies of the gene for release factor 1. J. Bacteriol. 168:1066-1069.

Sancar, A., A.M. Hack, and W.D. Rupp. 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. 137:692-693.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Sasarman, A., A. Nepveu, Y. Echelard, J. Dymetryszyn, M. Drolet, and C. Goyer. 1987. Molecular cloning and sequencing of the hemD gene of Excherichia coli K-12 and preliminary data on the Uro operon. 169:4257-4262.

Sasarman, A., M. Surdeanu, G. Szegli, T. Horodniceanu, V. Greceanu, and A. Dumitrescu. 1968a. Hemin-deficient mutants of Escherichia coli K-12. J. Bacteriol. 96:570-572.

Sasarman, A., M. Surdeanu, and T. Horodniceanu. 1968b. Locus determining the synthesis of 5-aminolevulinic acid in Escherichia coli K-12. J. Bacteriol. 96:1882-1884.

Sasarman, A., P. Chartrand, R. Proschek, M. Desrochers, D. Tardif, and C. Lapointe. 1975. Uroporphyrin-accumulating mutant of Escherichia coli K-12. J. Bacteriol. 124:1205-1212.

Sasarman, A., and T. Horodniceanu. 1967. Locus determining normal colony formation on the chromosome of Escherichia coli K-12. J. Bacteriol. 94:1268-1269.

Sasarman, A., Y. Echelard, J. Letowski, D. Tardif, and M. Drolet. 1988. Nucleotide sequence of the hemX gene, the third member of the Uro operon of Escherichia coli K12. Nucleic Acids Res. 16:11835.

Schneegurt, M.A., and S.I. Beale. 1988. Characterization of the RNA required for biosynthesis of 5-aminolevulinic acid from glutamate. Plant Physiol. 86:497-504.

Schneegurt, M.A., S. Rieble, and S.I. Beale. 1988. The tRNA required for in vitro 5-aminolevulinic acid formation from glutamate in Synechocystis extracts. Plant Physiol. 88:1358-1366.

Schoenhaut, D.S., and P.J. Curtis. 1986. Nucleotide sequence of mouse 5-aminolevulinic acid synthase cDNA and expression of its gene in hepatic and erythroid tissues. Gene 48:55-63.

Schon, A., C.G. Kannangara, S. Gough, and D. Soll. 1988. Protein biosynthesis in organelles requires misaminoacylation of tRNA. Nature. 331:187-190.

Schon, A., G. Krupp, S. Gough, S. Berry-Lowe, C.G. Kannangara, and D. Soll. 1986. The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. Nature (London) 322:281-284.

Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503.

Spears, P.A., D. Schauer, and P.E. Orndorff. 1986. Metastable regulation of type 1 piliation in Escherichia coli and isolation and characterization of a phenotypically stable mutant. J. Bacteriol. 168:179-185.

Tai, T., M.D. Moore, and S. Kaplan. 1988. Cloning and characterization of the 5-aminolevulinate synthase gene(s) from Rhodobacter sphaeroides. Gene 70:139-151.

Thomas, S.D., and P.M. Jordan. 1986. Nucleotide sequence of the hemC locus encoding porphobilinogen deaminase of Escherichia coli K12. Nucleic Acids Res. 14:6215.

Umanoff, H., C.S. Russell, and S.D. Cosloy. 1988. Availability of porphobilinogen controls appearance of porphobilinogen deaminase activity in Escherichia coli K-12. J. Bacteriol. 170:4969-4971.

Wang, W., D.D. Huang, D. Stachon, S.P. Gough, and C.G. Kannangara. 1984. Purification, characterization and fractionation of the 5-aminolevulinic acid-synthesizing enzymes from light-grown Chlamydomonas reinhardtii cells. Plant Physiol. 74:569-575.

Wang, W., S.P. Gough, and C.G. Kannangara. 1981. Biosynthesis of 5-aminolevulinate in greening barley leaves. IV. Isolation of three soluble enzymes required for the conversion of glutamate to 5-aminolevulinate. Carlsberg Res. Commun. 46:243-257.

Weinstein, J.D., and S.I. Beale. 1983. Enzymatic conversion of glutamate to 5-aminolevulinic acid in soluble extracts of Euglena gracilis. J. Biol. Chem. 258:6799-6807.

Weinstein, J.D., and S.I. Beale. 1985. Enzymatic conversion of glutamate to 5-aminolevulinate in soluble extracts of the unicellular green alga, Chlorella vulgaris. Arch. Biochem. Biophys. 237:454-464.

Weinstein, J.D., S.M. Mayer, and S.I. Beale. 1987. Formation of 5-aminolevulinic acid from glutamic acid in algal extracts. Plant Physiol. 84:244-250.

Winans, S.C., S.J. Elledge, J.H. Krueger, and G.C. Walker. 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in Escherichia coli. J. Bacteriol. 161:1219-1221.

Yager, T.D., and P.H. von Hippel. 1987. Transcript elongation and termination in Escherichia coli, p.1241-1275. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger, (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.

Yamamoto, M., N.S. Yew, M. Federspiel, J.B. Dodgson, N. Hayashi, and J.D. Engel. 1985. Isolation of recombinant cDNAs encoding chicken erythroid 5-aminolevulinate synthase. Proc. Natl. Acad. Sci. USA 82:3702-3706.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.