Order Number 9009756

Molecular biological study of early steps of heme biosynthesis in *Escherichia coli*

Li, Jianming, Ph.D.

City University of New York, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

MOLECULAR BIOLOGICAL STUDY OF EARLY STEPS OF HEME BIOSYNTHESIS IN *ESCHERICHIA COLI*

by

JIANMING LI

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1989

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Sept 8, 1989
Date

Charlotte S Russell
Chair of Examining Committee

Sept. 11, 1989
Date

Executive Officer

Supervisory Committee

The City University of New York

ABSTRACT

MOLECULAR BIOLOGICAL STUDY OF EARLY STEPS OF HEME BIOSYNTHESIS IN *ESCHERICHIA COLI*

by

JIANMING LI

Adviser: Professor Charlotte Russell and Professor Sharon Cosloy

The heme molecule is the prosthetic group of hemoglobins, cytochromes, catalases and peroxidases, and modified tetrapyrroles are the active moieties of the chlorophylls and vitamin B_{12}. The pathway of heme biosynthesis is highly conserved in living organisms. It is almost the same in every organism, from bacteria to man, except for the formation of 5-aminolevulinic acid (ALA), the first common intermediate. ALA can be synthesized in two ways: the C_4 pathway, which uses succinyl CoA and glycine as the substrates and the C_5 pathway, which uses glutamic acid as the precursor and involves three enzymatic steps, aminoacylation of $tRNA^{glu}$, dehydrogenation of glutamyl-tRNA and aminotransference (Figure 5). The C_4 pathway operates in animals, fungi, and facultative aerobic and photosynthetic bacteria. Plants, algae, strict anaerobic bacteria and some oxygenic prokaryotes synthesize ALA using the C_5 pathway.

In *E. coli*, the heme biosynthetic pathway consists of more than eight enzymatic steps. The genes encoding these enzymes are widely scattered on the chromosome. This work focuses on the molecular events of the early steps of the pathway in *E. coli*.

An *E. coli* heme-requiring, hemin-permeable mutant had no detectable 5-aminolevulinic acid dehydratase (ALA D) or porphobilinogen deaminase (PBG D) activities The gene which complemented this mutation was cloned to a high copy-number plasmid, and PBG D activity was restored to normal levels, but the activity of ALA D was 20-30 fold higher than normal. A maxicell procedure confirmed that the cloned gene was *hemB*. The insert of pJL2, which contained the *hemB* gene, was double-digested with restriction enzymes and subcloned into pTZ18U, pTZ19U and M13mp18. The *hemB* gene was sequenced using the dideoxynucleotide chain termination procedure with [^{35}S]dATP. Computer software was used to identify coding regions and the correct reading frame. Two promoter regions, two Shine-Dalgarno sequences and two possible initiation sites were identified. Extensive homologies with yeast (36%), human liver (40%) and rat liver (40%) amino acid sequences were observed, especially in the sixteen-amino acid Zn-binding region (75%) and the four amino acids surrounding the essential lysine at the active site (100% for rat and human proteins). Analysis of promoter strength and two independent analyses of codon usage indicated that the *hemB* gene is moderately-expressed.

An *E. coli* gene which complemented two independent *hemA* mutants of *E. coli* was cloned onto a multi-copy plasmid and both its strands have been sequenced. Both complemented mutants produced 5-ALA and displayed fluorescence after 24 h, indicating that porphyrins were accumulating. The cloned sequence appears to encode a 46 kD protein. When a short insert which complemented the mutation was expressed in the maxicell procedure a 41 kD protein was produced as determined on SDS gels. The amino acid sequence of the cloned gene

product showed no significant homologies with any cloned ALA synthase (ALA S) nor with any protein in two databanks. A second cloned gene fragment, which has its coding region 41 bp away from the coding region of the gene that complemented *hemA*, has been identified as part of protein release factor 1 (RF1), thus confirming the location of *hemA* at min 26.7 and mapping it precisely near RF1. Two ORF's oriented in the opposite direction are located upstream from *hemA*.

A *hemA* mutant of *E. coli* containing a multi-copy plasmid which complemented the mutation, excreted ALA into the medium. [1-^{14}C] Glutamic acid was substantially incorporated into ALA by this strain whereas [2-^{14}C] glycine was not. Periodate degradation of labeled ALA showed that C-5 of ALA was derived from C-1 of glutamic acid. The synthesis of ALA by two sonicate fractions which had been processed by gel filtration and dialysis respectively, was dependent on glutamic acid, ATP, NADPH, $tRNA^{glu}$ and pyridoxal phosphate. $tRNA^{glu}$ stimulated ALA synthesis in a concentration-dependent manner. Pretreatment with RNase lowered this stimulation. These results suggest that *E. coli* synthesizes ALA by the C_5 pathway from the intact five-carbon chain of glutamate.

ACKNOWLEDGEMENT

This thesis is dedicated to my wife whose love, patience and encouragement have help to bring this thesis to its successful completion, to my mother who made sure that my education came before all other things, to my daughters who gave me the inspiration to go on.

I would like to thank the members of my thesis committe, Dr. Donald Sloan, Dr. Steven Meshnick and Dr. Rivka Rudner for their valuable advice and cooperation.

I would also like to thank Dr. Horst Schulz and Mrs. Sophie Weintraub for their advice, suggestions and support over the past several years and Dr. Nick Hoffman for his showing me how to use the Macintosh computer software.

I want to especially thank my colleagues, Woody Umanoff, Ormand Brathwaite, Pi-Shiang Lai, and Lillian Fugler, not only for their scientific discussion and advice but for their friendship.

No matter what I say, it is not enough to express my appreciation to my mentors, Dr. Charlotte Russell and Dr. Sharon Cosloy, for their years of guidance, advice, encouragement and support throughout every aspect. I am very grateful for having spent my critical training years with them, and I will never forget that for the rest of my life.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

aa, amino acid;

ALA, 5-aminolevulinic acid;

amp, ampicillin;

arg, arginine;

ATP, adenosine triphosphate;

bp, base pair;

BSA, bovine serum albumin;

dATP, deoxyadenosine triphosphate;

dCTP, deoxycytidine triphosphate;

dGTP, deoxyguanosine triphosphate;

dITP, deoxyinosine triphosphate;

DMAB, dimethylaminobenzaldehyde;

DTT, dithiothreitol;

dTTP, deoxythymidine triphosphate;

E. coli, Escherichia coli;

EDTA, ethylenediaminetetraacetic acid;

gly, glycine;

glu, glutamic acid;

GSA, glutamic-1-semialdehyde;

h, hour;

his, histidine;

IPTG, isopropylthiogalactopyranoside;

kb, kilobase;

kDa, kilodalton;

leu, leucine;

met, methionine;

NADPH, nicotinamide-adenine dinucleotide phosphate reduced;

nt, nucleotide;

ORF, open reading frame;

PBG, porphobilinogen;

PEG, polyethylene glycol;

PLP, pyridoxal phosphate;

pro, proline;

SDS, sodium dodecyl sulfate;

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis;

sec, second;

ser, serine;

TCA, trichloroacetic acid;

TBE, Tris-Borate-EDTA;

TE, 10 mM Tris-Cl,1 mM EDTA;

TEMED, N,N,N',N'-tetramethyl-ethylenediamine;

thi, thiamine;

thy, thymine;

thr, threonine;

Tris, tris(hydroxymethyl)aminomethane;

trp, tryptophan;

tyr, tyrosine;

ura, uracil;

X-gal, 5-bromo-4-chloro-3-indolyl--galactopyranoside;

APPENDICES

INTRODUCTION

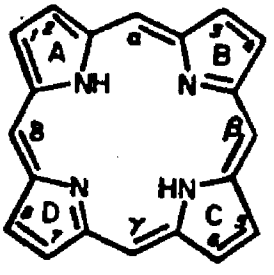
1. THE HEME MOLECULE

1.1. SKELETON AND NOMENCLATURE OF PORPHYRINS

Porphyrins are compounds which have a tetrapyrrolic ring structure in common. The fundamental macrocyclic tetrapyrrolic ring system is shown in Figure 1. It consists of four pyrrole rings joined by four methene bridges to give a macrocycle. This cyclic tetrapyrrole structure was first suggested by Kuster (1912). The nomenclature most generally used has been that of Fischer (Fischer and Orth, 1937; Fischer and Stern, 1940). In this classical system of nomenclature, the eight peripheral positions are numbered from 1 to 8 and the interpyrrolic methene positions, usually termed '*meso* ', are designated α, β, γ, and δ. The rings are usually lettered A, B, C, and D . This nomenclature system is based on a very large number of trivial names and a numeration scheme, shown in Figure 1, which is simple but incomplete. A more systematic nomenclature is recommended by the IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (Moss, 1988). The new 1-24 numbering scheme is adopted as shown in Figure 1. The fundamental macrocyclic tetrapyrrolic ring structure labeled in accordance with the recommendations of IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) is shown in Figure 1. The peripheral positions are now numbered 2, 3, 7, 8, 12, 13, 17, and 18. There are two hydrogen atoms which may be associated with any two of the four nitrogen atoms. However, for nomenclature purpose, the name 'porphyrin' implies that the saturated nitrogen atoms are at positions 21 and 23.

Most porphyrin compounds existing in nature have substituted side chains. Although substituted porphyrins can be named systematically on

A



B

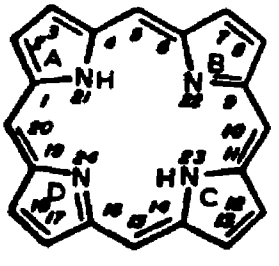


Figure 1. Structure of the porphyrin nucleus and the numbering systems. The structure of the porphyrin nucleus is shown in conventioal (Fischer) numeration in A and IUPAC-IUB numeration in B.

the basis of the fundamental porphyrin ring nuclei, some well-established trivial names are retained for naming certain substituted porphyrin structures containing the porphyrin nucleus. Table 1 shows some porphyrins related to this study.

Chlorins are 17,18-dihydroporphyrins in which the saturation is traditionally shown in ring D; the various chlorophylls are magnesium complexes of chlorins. Bacteriochlorophyll *a* is a tetrahydroporphyrin. The most important hexahydroporphyrins are porphyrinogens, which are colorless and contain four pyrrole rings linked in a macrocycle by four methylene groups; the importance of these materials lies in the fact that they are the immediate biosynthetic precursors of the porphyrins and plant pigments, as well as vitamin B_{12}.

The porphyrin macrocycle is highly conjugated and a number of resonance forms can be written. There are 22 π–electrons, but only 18 of these are included in any one delocalization pathway; this conforms with Huckel's 4n+2 rule for aromaticity. Porphyrins are highly colored, their most intense visible absorption band (called the 'Soret' band) has a very high extinction coefficient around 400nm, and this is characteristic of the macrocyclic conjugation. In addition to the Soret band, each different porphyrin has its own characteristic visible spectrum, depending on its substitutents, ionization state, and the solvents used.

The porphyrin nucleus is stable, even in concentrated sulfuric acid and neat trifluoroacetic acid. However, solutions of porphyrins with substituted side chains are very light-sensitive. They can be photo-oxidized, or photo-reduced in the presence of compounds such as ascorbic acid, glutathione, or tertiary amines, to give phlorins.

Porphyrins can combine with many metal ions. Biological cyclic

Table 1. Some substituted porphyrins related to this study with their trivial names

Trivial name	Substituents							
	2	3	7	8	12	13	17	18
Uroporphyrin I	Cm	Cet	Cm	Cet	cm	Cet	Cm	Cet
Uroporphyrin III	Cm	Cet	Cm	Cet	cm	Cet	Cet	Cm
Coproporphyrin I	Me	Cet	Me	Cet	Me	Cet	Me	Cet
Coproporphyrin III	Me	Cet	Me	Cet	Me	Cet	Cet	Me
Protoporphyrin IX	Me	Vn	Me	Vn	Me	Cet	Cet	Me

Cm, $-CH_2CO_2H$; Cet, $-CH_2CH_2CO_2H$; Me, CH_3; Et, $-CH_2CH_3$; Vn, $-CH=CH_2$. The use of the symbols A and P for $-CH_2CO_2H$ and $-CH_2CH_2CO_2H$ is often seen in the field of porphyrin chemistry, too.

metallotetrapyrrole compounds are known which contain Fe (heme and siroheme), Mg (chlorophylls), and Co (vitamin B_{12}). Other derivatives that can be readily formed include those with Zn, Ni, Cu, and Ag. The common structural pattern for the metal coordination complex consists of a metal ion coordinated to a roughly tetrapyrrole, acting as a tetradentate dianionic ligand, and, possibly, to one or two axial ligands. The tetrapyrrolic ligand is oriented with the numbering clockwise when seen from above. The axial ligands above and below the plane are then designated by β and α respectively.

1.2. THE STRUCTURE OF THE HEME MOLECULE

Heme is an iron porphyrin. Its structure is shown in Figure 2. The porphyrin, protoporphyrin type III, also called protoporphyrin IX, contains four pyrrole-like rings linked by four methene (=CH- group) bridges in an alternating double-bond ring system. The numbering system of the ring is shown according the recommendations of IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (Moss, 1988). The porphyrin contains methyl groups on C_2, C_7, C_{12} and C_{18}, vinyl groups on C_3 and C_8, and propionic acid groups on C_{13} and C_{17}. This structure is one of fifteen possible isomers.

In the center of the heme molecule an iron atom replaces the dissociable hydrogen atoms of two pyrrole rings and is simultaneously bound by coordination valences to the tertiary nitrogen atoms of the other two pyrrole rings. Because of resonance no distinction is made between these valences. The iron can form two additional bonds, one on either side of the heme plane. When the iron atom is in the ferrous (Fe^{2+}) state the coordination complex is called ferroprotoporphyrin, or heme; if it is in

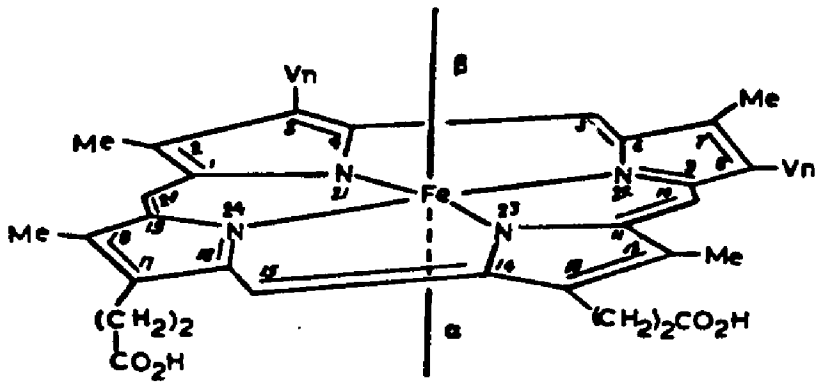


Figure 2. The structure of the heme molecule. Me, methyl group; Vn, vinyl group;

the Fe^{3+} state the complex is called ferriprotoporphyrin, or hemin. The heme molecule exists in the square -planar form while hemin is essentially square pyramidal with the extra ligand perpendicular to the porphyrin plane. Free heme is unstable and rapidly oxidized to hemin.

1.3. THE HEME BIOSYNTHETIC PATHWAY

The pathway of heme biosynthesis (Figure 3) is highly conserved. It is almost the same in every organism, from bacteria to man, from plant to animal, except for the formation of 5-aminolevulinic acid (ALA), the first common intermediate. Two molecules of ALA then condense to generate porphobilinogen (PBG) in a reaction catalyzed by ALA dehydratase (ALA D). Four molecules of PBG are utilized by PBG deaminase (PBG D) (also called uroporphyrinogen synthase, and hydroxymethylbilane synthase) to form a linear tetrapyrrole, which probably remains bound to the enzyme. An ammonium ion is released for each methylene bridge formed. This linear molecule, hydroxymethylbilane, can cyclize non-enzymatically to produce uroporphyrinogen I, which has a symmetric arrangement of side chains. Uroporphyrinogen I may undergo non-enzymatic oxidation to form one of the urinary pigments, uroporphyrin I, or it may be decarboxylated to yield coproporphyrinogen I , which in turn can be oxidized to coproporphyrin I. These porphyrins do not normally accumulate. When both PBG D and uroporphyrinogen III cosynthase are present, the bilane will ring-close with rearrangement of ring D to uroporphyrinogen III. Uroporphyrinogen III is decarboxylated by uroporphyrinogen decarboxylase to yield coproporphyrinogen III. Coproporphyrinogen III is modified by coproporphyrinogen oxidase to form protoporphyrinogen IX, which is oxidized by protoporphyrinogen

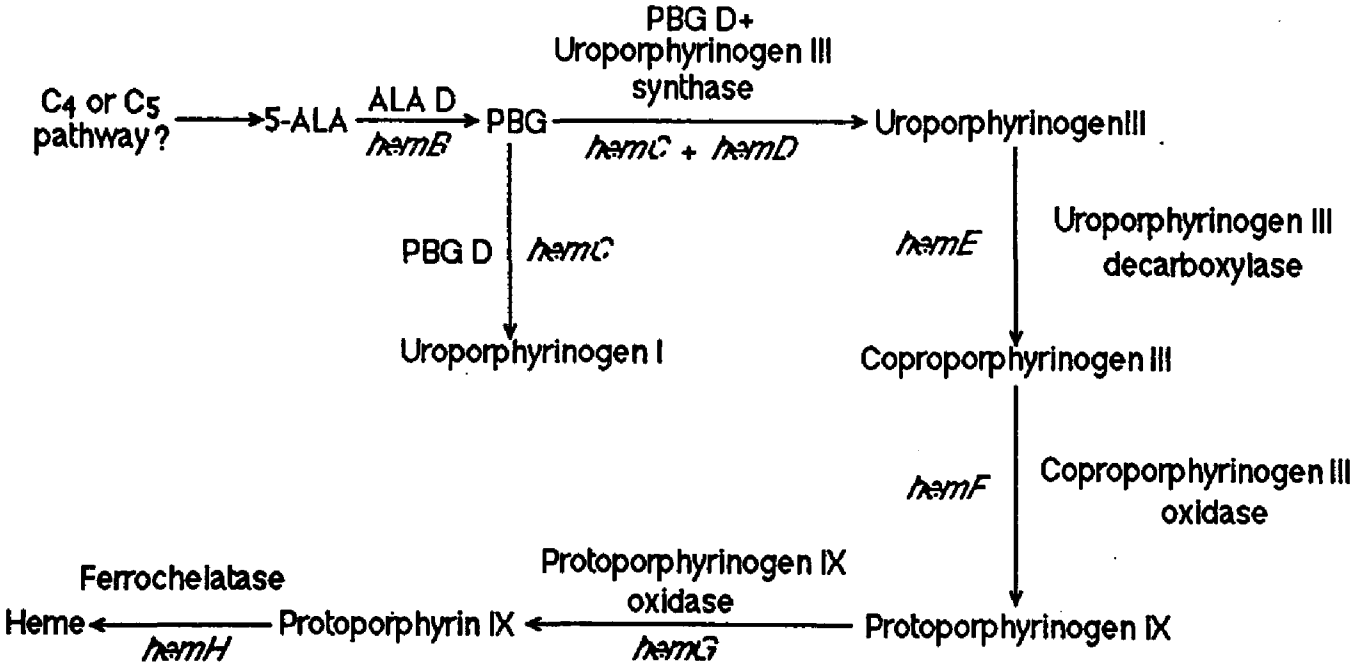


Figure 3. Heme biosynthetic pathway in *E. coli*

oxidase to protoporphyrin IX. Ferrochelatase incorporates an iron atom (Fe^{2+}) into protoporphyrin IX, resulting in the formation of heme.

In eukaryotes, ALA D, PBG D, uroporphyrinogen III cosynthase and uroporphyrinogen decarboxylase are located in the cytoplasm, while ALA S, coproporphyrinogen oxidase, protoporphyrinogen oxidase and ferrochelatase are found in the mitochondria.

1.4. THE BIOLOGICAL FUNCTIONS OF HEME

The heme molecule is involved in several vital biological process: O2 transport, energy production in aerobic respiration, cell detoxification and synthesis of methionine and cysteine.

Oxygen transport The transition from anaerobic to aerobic life was a major step in evolution because it uncovered a rich reservoir of energy. In the presence of oxygen eighteen times as much energy is extracted from glucose. The oxygen needed for aerobic respiration in vertebrates is transported by the oxygen carrier molecules, hemoglobin and myoglobin. Both proteins contain the heme molecule as the oxygen-carrying group. Hemoglobin, which is contained in red blood cells, serves as the oxygen carrier in blood and also plays a vital role in the transport of carbon dioxide and hydrogen ion. Myoglobin, which is located in muscle, serves as a reserve supply of oxygen and facilitates the movement of oxygen within muscle. The heme molecule is located in a crevice in the myoglobin molecule and is surrounded by nonpolar amino acid residues except for two histidines. The iron atom of the heme is bound to one of the two histidines on one side and to O_2 on the other side by coordination linkage. Hemoglobin has four polypeptide chains, two of one kind and

two of another, generally $\alpha_2\beta_2$. Each contains a heme group and a single-oxygen binding site. Only ferrohemoglobin can bind to oxygen. The binding of oxygen in hemoglobin is cooperative. Binding at one heme promotes the binding of oxygen at the other hemes on the same hemoglobin molecule. Conversely, the unloading of oxygen at one heme facilitates the unloading at the others. These allosteric effects greatly enhance the efficiency of action of hemoglobin.

Energy production In aerobic respiration, electrons are passed through the electron transport chain in order to generate energy. Cytochromes are important electron-transfer components of the electron transport chain. The prosthetic group of cytochromes is heme, as in myoglobin and hemoglobin, or a derivative of heme. In cytochrome *c*, the heme group is surrounded by tightly packed hydrophobic chains and is covalently bound to the apoprotein by a vinyl side chain as a thioether. The iron atom is bound to the sulfur atom of methionine residue 80 and to the nitrogen atom of histidine residue 18. The iron atom alternates between a reduced ferrous (+2) state and an oxidized ferric (+3) state during electron transport. Cytochrome reductase and cytochrome oxidase also contain different cytochromes with heme or heme A (a heme derivative) as a prosthetic group.

Cytochrome P450 Cytochrome *P450* is a heme-containing molecule. Hydroxylation in the synthesis of steroid hormones and bile salts requires the activation of oxygen. The activation is accomplished by cytochrome *P450*. Cytochrome *P450* is also important for the

detoxification of a variety of compounds including drugs in the liver, adrenal cortex, lung, and placenta.

Cell detoxification Other heme enzymes include catalase and peroxidase. Catalase catalyzes the decomposition of hydrogen peroxide formed by aerobic oxidation of reduced flavoproteins and O_2^-. This activity is present in nearly all animal cells and organs, all plant materials studied and almost all microorganisms other than obligate anaerobes. Peroxidase catalyzes oxidation of various organic substances by peroxides. The iron in the heme group of catalases and peroxidases is in the ferric (+3) state.

Chlorophylls, vitamin B_{12} and siroheme Among other biologically important porphyrin derivatives are chlorophylls, vitamin B_{12} and siroheme. The absorption of light by chlorophylls is the fundamental event in photosynthesis. Chlorophyll *a*, present in the chloroplast of all green plant cells, contains four substituted pyrrole rings, one of which (ring D) is reduced. Chlorophyll *a* also has a fifth ring , which is not pyrrole. The four central nitrogen atoms are coordinated with Mg^{2+}. Vitamin B_{12} is an essential nutrient, not made by either plants or animals and can be synthesized by only a few species of microorganisms. In the center of vitamin B_{12} is a cobalt atom which coordinates with the four nitrogen atoms of the ring system. The complex corrin ring system is chemically related to the porphyrin ring system of heme. Its biosynthesis branches from the heme pathway after the formation of uroporphyrinogen III. Vitamin B_{12} is known to serve as the cofactor for four enzymes in *S. typhimurium*; these include a homocysteine methyltransferase (*metH*) that

can function in methionine synthesis as an alternative to a B_{12}-independent (*metE*) enzyme (Cauthen et al., 1966), and the catabolic enzyme ethanolamine ammonia lysae (Chang and Chang, 1975; Jeter et al., 1987). Siroheme is a derivative of heme and is the prosthetic group of sulfite reductase and nitrite reductase (Siegel et al., 1973) Sulfite reductase is required for the synthesis of cysteine from inorganic sulfate (Kredich, 1983).

Inhibitory effect of heme on its synthesis In the liver, ALA synthase is the rate-limiting enzyme of the heme biosynthetic pathway. Heme exerts feedback inhibition on activity and repression of synthesis of ALA synthase (Sassa and Kappas, 1981). Heme also inhibits the translocation of ALA synthase precursor into mitochondria. Heme and oxygen have been shown to exert positive control over the synthesis of many (heme)proteins having respiratory functions or involved in oxygen metabolism in the yeast *Saccharomyces cerevisiae*. This regulation operates at the level of transcription in the cases of iso-1-cytochrome *c*, catalases A and T, Mn superoxide dismutase, and probably also for subunits V and VII of cytochrome *c* oxidase (Zagorec and Labbe-Bois, 1986). Heme has also been shown to be a negative effecter of anaerobic genes, such as the *ANB1* gene, which is produced only in the absence of oxygen, and a positive effecter of aerobic genes, e.g. *tr-I* and *CYC1* of yeast. The coordinated regulation is accomplished through the action of a common trans-acting factor, the *ROX1* gene product (Lowry and Lieber, 1986).

Diseases caused by disturbances of heme biosynthetic pathway

Since the heme molecule is the prosthetic group of hemoglobins, cytochromes, catalases and peroxidases, and modified tetrapyrroles are the active moieties of the chlorophylls and vitamin B_{12}, any disruption of the fundamental pathway has profound effects on the organism. In mammals, partial defects of the heme pathway lead to a group of diseases known generically as porphyrias, which are characterized by the accumulation of porphyrins and porphyrin precursors. Table 2 lists some porphyrias and their enzymatic abnormalities. These porphyrias can be divided into two main groups, hepatic and erythropoietic, depending on the primary site. Most of the porphyrias are, in fact, inherited as Mendelian dominant autosomal characters, although one, congenital erythropoietic porphyria, is inherited as a Mendelian recessive and another, cutaneous hepatic porphyria, exists in two forms, a hereditary disease and an acquired disease. In patients with congenital erythropoietic porphyria, there is a deficiency of uroporphyrinogen III cosynthase, the isomerase that yields the asymmetric isomer on cyclization of the linear tetrapyrrole. Thus uroporphyrin I and coproporphyrin I, derivatives devoid of a physiologic role, accumulate. Erythrocytes are prematurely destroyed in this disease, which is transmitted as autosomal recessive. The urine of patients having this disease is red because of the excretion of large amounts of uroporphyrin I. Their teeth fluoresce strongly under ultroviolet light and their skin shows abnormal sensitivity to sunlight. Acute intermittent porphyria occurs mostly in the liver and is transmitted as autosomal dominant. This disease presents with intermittent attacks of systemic illness characterized by abdominal pain, peripheral neuropathy, and mental dysfunction. ALA

Table 2. Observed and suggested enzymatic defects in porphyrias and lead intoxication

Condition	Primary defects	Secondary regulatory consequences	Genetic and manifestation
PLP	ALA D		Probably autosomal recessive
AIP	PBG D ↓	ALA S ↑↑	Autosomal dominant hereditary
	Δ^4-5α-Steroid reductase↓	Uroporphyrinogen cosynthase↑	Manifestation by drugs, sexual hormones and fasting
HCP	Coprogen oxidase↓	ALA S ↑↑	Autosomal dominant hereditary
PV	Ferrochelatase↓	ALA S ↑↑	Autosomal dominant hereditary
PCT	Urogen decarboxylase↓↓	Urogen-1-synthase ↑	Uroporphyrinogen decarboxylase
CHP	Urogen cosynthase↓↓	ALA S ↑	Defect inherited or acquired; manifestation by liver disease, alcohol, estrogen and certain chemicals (HCB and chlorinated aromatic compounds)
EPP	Ferrochelatase↓	ALA S ↑	Autosomal dominant hereditary
Pb-I	ALA D↓↓	ALA S ↑↑	Lead directly inhibits ALA D and ferrochelatase
	Ferrochelatase↓↓		
	Coprogen oxidase ↓		

PLP, Plumboporphyria; AIP, Acute intermittent porphyria; HCP, hereditary coproporphyria; PV, porphyria variegate; PCT, porphyria cutanea tarta; CHP, chronic hepatic porphyria; EPP, erythropoietic protoporphyria; Pb-I, lead intoxication;

and PBG are excreted in high amounts into urine in patients with this disease. The urine darkens on long exposure in sunlight to a deep-red wine color. George III, King of England at the time of the American war for independence, is believed to have suffered from this type of porphyria. Medical historians have suggested that the behavioral symptoms of this disease may have been responsible for the unreasonable insistence of George III on excessive taxation and punishment of the American colonies.

1.5. GENETICS OF *HEM* GENES OF *E.COLI*

The selection of heme-deficient mutants of *E. coli* usually involves several steps, mutagenesis, followed by selection for respiratory mutants which are devoid of catalase activity, characterization of the mutants by genetic mapping and assays of enzymes in the heme biosynthetic pathway before and after complementation. As *E. coli* is naturally impermeable to heme, it is difficult to assay mutants for the heme requirement (Sasarman, et al., 1968). In 1957, Beljanski and Beljanski described the isolation of the first hemin-requiring mutant of *E. coli* M by means of streptomycin selection. The mutant failed to synthesize catalase and cytochromes, displayed respiratory deficiency, and grew poorly on the usual culture media. Under special circumstances (anaerobiosis, or intensely aerated broth culture) hemin increased growth or catalase synthesis. The first ALA-requiring mutant of *E. coli* was isolated by mutagenesis with N-methyl-N'-nitro-N-nitrosoguanidine and selection for ALA auxotrophy (Wulff, 1967). This marker was shown to lie in the *pro-thr-leu* region of the *E. coli* chromosome. Sasarman and Horodniceanu selected a group of respiratory mutants using neomycin, and found that

most of them are linked to the *lac* locus (1967). In 1968, Sasarman found a second group of ALA-requiring mutants, which were linked to the *trp* marker at different position from Wulff's mutant. Powell and coworkers established 5 groups of heme deficient mutants, *popA*, *popB*, *popE*, *popD*, and *popC* (Powell, et al, 1973). The first four accumulate protoporphyrin IX, coproporphyrin, porphobilinogen and ALA respectively; *popC* requires ALA for growth. With the development of technology and accumulation of data, more accurate positions of the *hem* genes on *E. coli* K-12 chromosome became available (Bachmann, 1983). To date nine *hem* genes have been discovered and located on the chromosome: *hemA* (27 min, Guest, 1969; Sasarman, et al, 1968b; Sasarman, et al, 1968a), *hemB* (8 min, Sasarman, et al, 1968a; McConville and Charles, 1979a), *hemC* (85 min, McConville and Charles, 1979b), *hemD* (85 Chartrand, et al, 1979), *hemE* (90 min, Sasarman, et al, 1975), *hemF* (17 min, Cox and Charles, 1973; Powell, et al, 1973),*hemG* (86 min, Sasarman, et al, 1979), *hemH* (11 min, Cox and Charles, 1973; Powell, et al, 1973), and *popC* (4 min, Powell, et al, 1973) respectively (Figure 4). *hemB, C, D, E, F, G,* and *H* encode ALA dehydratase (EC 4.2.1.24), PBG deaminase (also called uroporphyrinogen I synthase and hydroxymethylbilane synthase) (EC 4.3.1.8), uroporphyrinogen III cosynthase, uroporphyrinogen decarboxylase (EC 4.1.1.37), coproporphyrinogen III oxidase (EC 1.3.3.3), protoporphyrinogen oxidase, and ferrochelatase (EC 4.99.1.1) respectively. It was believed that *hemA* encoded for ALA synthase. However, our work in this study (Li, et al, 1988b; Li, et al, 1989b) shows that *hemA* encodes a protein in the heme biosynthetic pathway of *E. coli* K-12, involved in ALA synthesis, but it is not ALA synthase. This has been confirmed by others (Avissar and Beale,

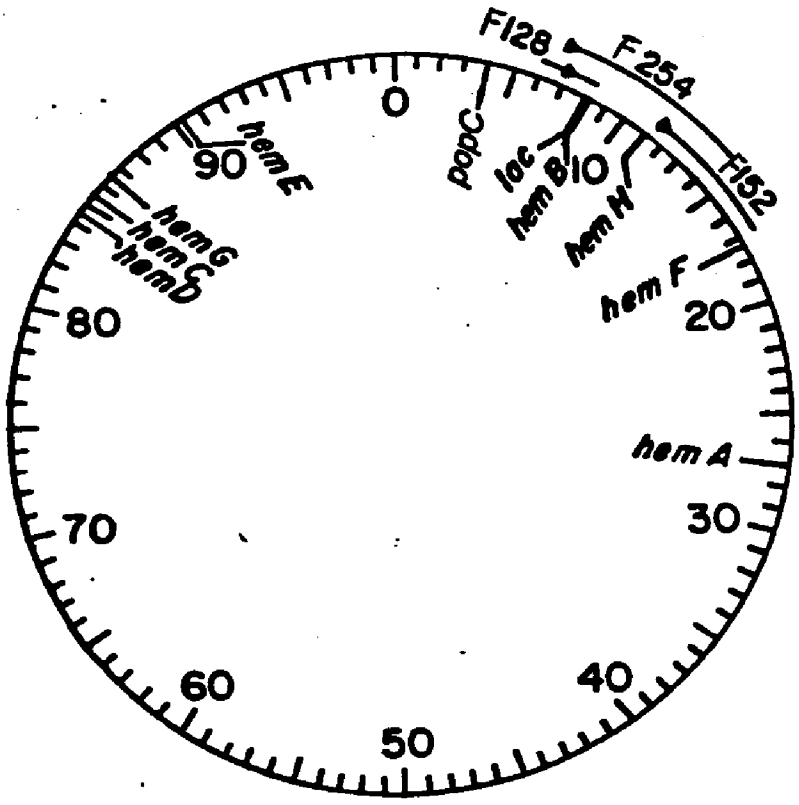


Figure 4. Locations of *E. coli hem* genes on the genetic map

1989; Hederstedt, private communication,1989). Although *popC* appears to be involved in ALA synthesis, its function is not known. *popA*, *popB* and *popE* are now called *hemH*, *hemF* and *hemC*, respectively.

In prokaryotic organisms, the genes of a metabolic pathway are usually arranged together to form a operon. All the genes are switched on or off simultaneously by its regulatory machinery. In contrast, the *hem* genes in *E.coli* are widely-scattered on the chromosome (Figure 4) except for *hemC*, *hemD* and possibly *hemG* which are linked and appear to form an operon (Jordan et al., 1987; Jordan et al., 1988; Sasarman et al., 1987). Thus the regulation and coordination of the expression of *hem* genes in *Escherichia coli* are more complicated and interesting. The elucidation of the regulatory mechanism of the *hem* genes may provide some information about the regulation of gene expression in other organisms, when the genes of a metabolic pathway are scattered on a chromosome or on different chromosomes.

A new, simple method for screening porphyrin secretion mutants was described recently (Javor and Kim, 1989). In this method, *E. coli* is grown in the present of 1-thioglycerol on a modified medium; DEAE-Sephadex beads are incorporated into solid medium. When colonies are viewed under long-wavelength ultraviolet light, wild type *E. coli* secretes porphyrin and forms a fluorescent halo, while mutants blocked at an early stage are haloless, or if blocked at a late step produce a very bright halo. This method appears to be quite simple and allows quick characterization of mutants.

2. THE FIRST STEP: SYNTHESIS OF ALA

2.1 C_4 PATHWAY

The synthesis of ALA is the first step in the biosynthesis of hemes, chlorophylls, corrins, and bile pigments. In the early of 1950s, it was demonstrated that glycine and succinyl-CoA were the precursors of ALA in birds (Shemin and Russell, 1953; Shemin, et al, 1954; Shemin, et al, 1955). The enzyme catalyzing the reaction is called ALA synthase (ALA S; succinyl-CoA:glycine C-succinyltransferase [decarboxylating]; EC 2.3.1.37). This reaction is called the C_4 pathway for ALA synthesis (Figure 5). In animals (Bottomly and Smithee, 1968; Gibson, et al, 1958; Marver, et al, 1966), fungi (Porra, et al, 1972), and facultative aerobic and photosynthetic bacteria (Kikuchi, et al, 1958; Lascelles, 1964; Leong, et al, 1982; Sasaki, et al, 1987; Tait, 1972; Tait, 1973; Warnick and Burnham, 1971; Wright, et al, 1987), ALA is synthesized by the C_4 pathway (Figure 5).

ALA synthase is the rate-limiting step in porphyrin biosynthesis and the major control point of heme production in liver (Kappas et al, 1983). The enzyme is located in the mitochondrial matrix and is derived from a larger cytoplasmic precursor (Yamauchi et al, 1980). The N-terminal (presequence) of the precursor of ALA synthase has to be cleaved in order for the enzyme to be translocated into mitochondria (Borthwick et al, 1985). However, Volland and Urban-Grimal showed that the presequence of ALA synthase is not required for targeting to mitochondria in yeast (1988). ALA synthase activity increases significantly in liver when large amounts of heme are required for microsomal detoxification systems and during the differentiation of erythroid cells where heme is needed for the synthesis of hemoglobin. In

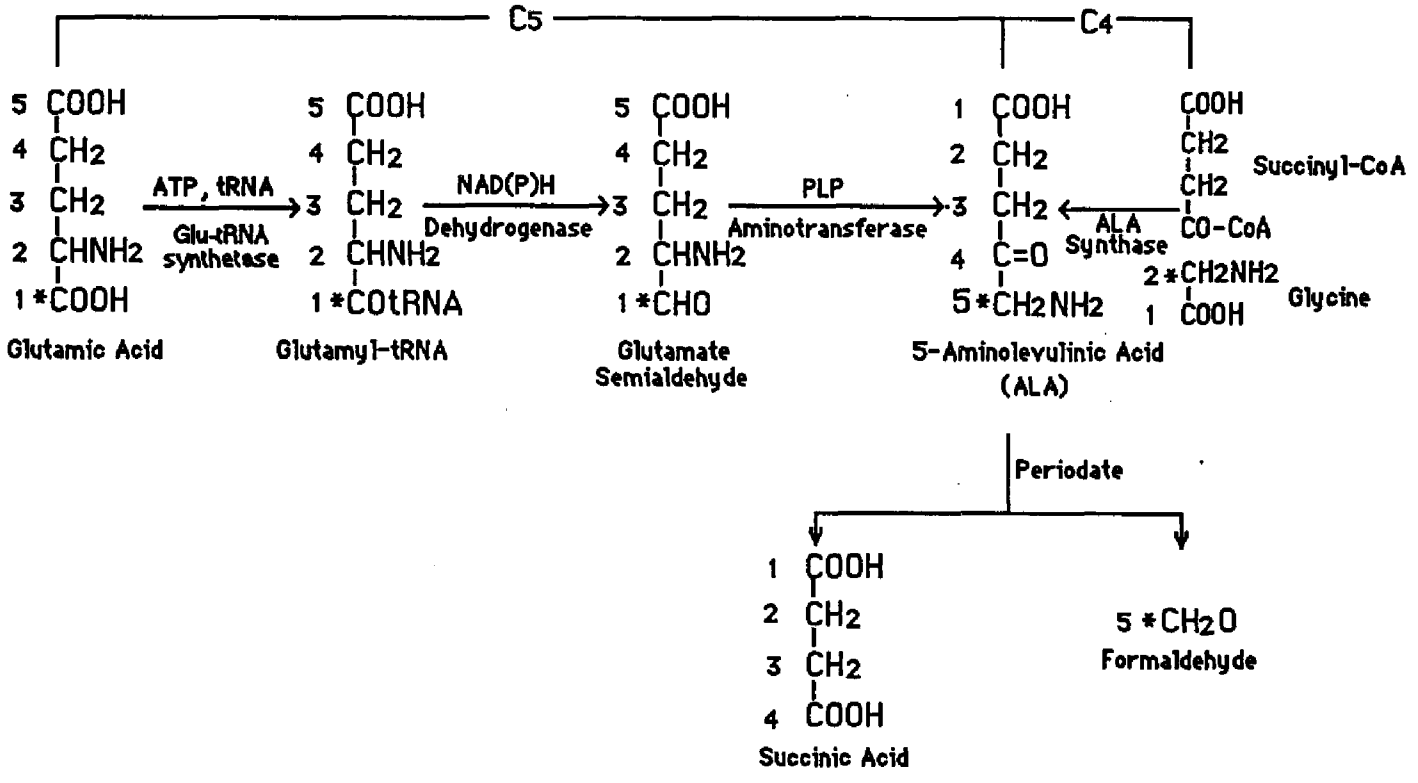


Figure 5. C_5 and C_4 pathways for ALA biosynthesis and periodate degradation scheme, with numbering systems for intermediates and fragmentation products. *, Origion of C-5 of ALA from the C_5 and C_4 pathways and fate of C-5 of ALA after periodate degradation. COtRNA, Ester linkage between tRNA and the carboxyl group of glutamic acid.

liver, ALA synthase is repressed by heme and increased by porphyrinogenic compounds. In contrast, heme increases or does not affect ALA synthase activity and little or no effect of porphyrinogenic drugs is observed in erythroid cells (Kappas et al, 1983). Thus, the expression of ALA synthase is regulated differently in these tissues.

The genes coding for ALA synthase from human (Bawden et al, 1987), chicken (Borthwick et al, 1984), mouse (Schoenhaut and Curtis, 1986),*Bradyrhizobium japonicum* (McClung et al, 1987), yeast (Keng et al, 1986), *Rhodobacter sphaeroides* (Tai et al, 1988), and *Rhizobium meliloti* (Leong et al, 1982) have been cloned. The complete sequences for the human (Bawden et al, 1987), chicken (Borthwick et al, 1985), mouse (Schoenhaut and Curtis, 1986), *Bradyrhizobium japonicum* (McClung et al, 1987) and partial sequences for the yeast (Keng et al., 1986), and *Rhizobium meliloti* (Leong et al., 1985) enzymes have been published.

2.2. C_5 PATHWAY

Attempts to detect ALA synthase activity in plants were largely unsuccessful although it was known that plants synthesize ALA. The administration of levulinic acid, a competitive inhibitor of ALA dehydratase, to the cells of the unicellular chlorophyte, *Chlorella vulgaris*, resulted in inhibition of chlorophyll a synthesis with concomitant accumulation of ALA (Beale, 1970; Beale, 1971). The reason for this observation was explained in an series of elegant studies. In greening cucumber cotyledons, radiolabeled glutamic acid and α-ketoglutarate were incorporated into ALA much more efficiently than were glycine and succinic acid (Beale and Castelfranco, 1974). When the radiolabeled

ALA was degraded with periodate in order to analyze the pattern of incorporation of the precursors, it was found that C-1 of glutamic acid was preferentially incorporated into C-5 of ALA and C-3 and C-4 of glutamic acid were preferentially incorporated into the C_1-C_4 fragment of ALA (Beale et al, 1975). The soluble protein fraction of isolated chloroplasts from greening barley leaves incorporated glutamic acid into ALA in a reaction dependent on ATP, NADPH, and $MgCl_2$ (Kannangara and Gough, 1977; Gough and Kannangara, 1977). The protein fraction also converted glutamic-1-semialdehyde (GSA) into ALA when NADPH was absent (Kannangara and Gough, 1978). The formation of ALA from the intact chain of glutamic acid is called the C_5 pathway. Recent evidence suggests that the C_5 pathway requires three steps: synthesis of glutamyl-tRNA (Weinstein and Beale, 1985; Huang and Huang, 1986; Schneegurt and Beale, 1986; Peterson et al, 1988; Kannangara et al, 1988; O'Neill et al, 1988), reduction with NADPH or NADH to yield GSA, and aminotransference to produce ALA (Mau et al,1987; Wang et al, 1981; Wang et al, 1984). These steps and the numbering system for the molecules involved are shown in Figure 5 . In plants (Beale and Castelfranco, 1974; Beale et al, 1975; Ford and Friedmann, 1979; Harel and Neeman, 1983), algae (Avissar, 1980; Huang and Huang, 1986; Oh-hama et al, 1986; Oh-hama et al, 1986; Troxler and Offner, 1979; Wang et al, 1984; Weinstein et al, 1987), strict anaerobic bacteria (Friedmann and Thauer, 1986; Gollis et al, 1983), and some oxygenic prokaryotes (Rieble and Beale, 1988), ALA is synthesized by the C_5 pathway from the intact five-carbon chain of glutamic acid. Both C_5 and C_4 pathways have been shown to operate in *Euglena gracilis* (Weinstein and Beale,1983) and

Scenedesmus obliquus (Klein and Senger, 1978a; Klein and Senger, 1978b).

2.3. ALA SYNTHESIS IN *E.COLI* K-12

Two different *E.coli* K-12 mutants, which require ALA for growth, have been described. One mutation maps at 27 min on the linkage map and is called *hemA* (Sasarman et al, 1968). Another is at 4 min and is called *popC* (Powell et al, 1973). It had been believed that these mutants were deficient in ALA synthase activity, since ALA synthase was assumed to be the enzyme which synthesizes ALA from succinyl CoA and glycine in *E. coli* K-12 and other bacteria. This assumption was supported by the observation that *hemA* mutants transformed by plasmids carrying the ALA synthase gene from other organisms grew on media without supplementation of ALA (Leong et al, 1982; Schoenhaut and Curtis, 1986; Tai et al, 1988). However, ALA synthase activity has never been detected in wild type *E. coli*. One explanation put forth for this observation was that *E.coli* has very low levels of ALA synthase activity, as is known for some tissues of animals, and that the assay method used was not sensitive enough to detect it.

Since ALA synthase catalyzes the first step of heme biosynthesis and is shown to be the rate-limiting step of the pathway in animal liver, one of the original objectives of this study was to clone and sequence the *hemA* gene for the study of its regulation and to characterize the enzyme.

2.4. THE ECONOMIC VALUE OF ALA

Recently, ALA, the precursor of chlorophylls, has been reported to show marked and selective herbicidal activity against dicotyledons, many

of which are weeds, but not against monocotyledons such as crops, with no harm to humans or other animals (Rebeiz et al., 1984). The mechanism of action of the compound depends on its conversion to tetrapyrroles within plants after application in darkness. The six tetrapyrroles especially important for this effect are magnesium mono- and divinylprotoporphyrin IX, their monoesters, and mono- and divinylprotochlorophyllide. At sunrise, the excess concentration of tetrapyrroles act as photosensitizers, converting normal oxygen to the potently oxidizing singlet form. They do this by absorbing photons and transferring them to free oxygen atoms in the cell, where they so excite electrons that the electrons reverse their direction of spin. This, in turn, oxidizes nearby atoms, altering, and even damaging, the molecules of which the atoms are a part. Singlet oxygen, one of the most potent oxidants known, can, in sufficient quantities, kill cells and tissues within hours.

Since all plants make tetrapyrrole from ALA, the interesting question is why ALA action has a different effect on different plants. Rebeiz suggests that plants green in different ways (1987). When some plants are treated with ALA, the amino acid arrives in the chloroplast at a time and in a way that coincide with the the organism's normal greening process; the resultant tetrapyrroles undergo conversion before they have time to photosensitize oxygen. In other plants, however, ALA treatment induces the plant to accumulate the wrong tetrapyrrole, one that, instead of being quickly metabolized, lingers in the plant's chloroplasts, generating singlet oxygen, which oxidizes the cells until the plant withers. From a survey of about a hundred species, they found that plants contain different kinds of tetrapyrroles at various stages in their formation of chlorophylls.

Based on whether they make use of monovinyl (MV) or divinyl tetrapyrroles (DV) and on the time of day (some plants switch pathways in the day and night), the plants can be divided into four groups, DV/DV, MV/DV, DV/MV, and MV/MV, the first referring to night and the second to day (Rebeiz, 1987).

ALA is also an insecticide. When it is sprayed, with a chemical modulator on or ingested by an insect, it dies within 10 seconds after exposure to light. ALA has effectively killed the corn earworm or the cabbage looper which are two of the most destructive agricultural pests.

However, ALA is very expensive and currently is used mainly as a research compound. Its use as a herbicide and insecticide will depend upon development of more economic ways to synthesize it. Some laboratories are trying to develop bacteria as the source of ALA (Rhee et al., 1987; Sasaki et al., 1987). During this study, two *E. coli* strains were seen to accumulate ALA, and one of them excretes the newly synthesized ALA into the medium. These strains may be a good source of ALA.

3. THE SECOND STEP: SYNTHESIS OF PORPHOBILINOGEN

The second step of the pathway of heme biosynthesis is the asymmetric condensation of two molecules of ALA to form the monopyrrole, porphobilinogen, which is catalyzed by ALA dehydratase (ALA D; porphobilinogen synthase; 5-aminolevulinate hydro-lyase). ALA D has been isolated and characterized from various sources (Gibson et al, 1955), including human (Anderson and Desnick, 1979), duck (Schmid and Shemin, 1955), and chicken erythrocytes (Granick, 1954), bovine liver (Cheh and Neilands, 1973; Gurba et al, 1972; Wilson et al,1972; Wu et al, 1974), mouse (Coleman, 1966; Doyle and Schimke, 1969), guinea pig

(Weissberg and Voytek, 1974), plants (Liedgens et al, 1983; Maralihalli et al, 1985; Nandi and Waygood, 1967; Shetty and Miller, 1969; Shibata and Ochiai, 1976; Stella and Batlle, 1978), and bacteria (*Rhodopseudomonas spheroides* [Burnham and Lascelles, 1963; Nandi and Shemin, 1968], *Rhodopseudomonas capsulata* [Nandi and Shemin, 1973], and *Mycobacterium phlei* [Yamasaki and Moriyama, 1971]). While the enzymes from all sources appear to be octamers of subunits of 31 to 35 kD, there are significant differences in susceptibility to chelating agents and activation by cations. A lysine-containing sequence, M-V-K-P-G-M, was identified at the active site of human ALA D (Gibbs and Jordan, 1986). The binding of a zinc atom to an essential SH group of the enzyme was shown by Barnard et al. (1977), Tsukamoto et al. (1979) and Gibbs and Jordan (1981).

ALA D is of considerable clinical interest. A sensitive diagnostic indicator of lead poisoning is the inhibition of erythrocyte ALA D activity. The marked inhibitory effect of lead (and other metals) is reversible *in vitro* and *in vivo* by zinc atoms. ALA accumulation is the cause of the neurologic involvement in hepatic porphyria resulting from the inherited deficiency of ALA D because ALA is a potent agonist of the γ-amino-butyric acid presynaptic receptor in the nervous system (Brennan and Cantrill, 1979; Kappas et al.,1983).

The genes for ALA D from human (Wetmur et al.,1986a; Wetmur et al., 1986b) , rat (Bishop et al.1986a; Bishop et al.1986b) , mouse (Bishop et al., 1989), *E. coli* K-12 (Li et al., 1988a; Echelard et al., 1988; Li et al., 1989a), and yeast (Myers et al., 1987) have been cloned and sequenced. The structural gene encoding human ALA D is assigned to the chromosomal region 9q34 (Elberg et al., 1983; Wang et al., 1985).

Interestingly, the product of ALA D, PBG, plays a regulatory role in the activity of the next enzyme of the heme biosynthetic pathway. In this study we found that a *hemB* mutant has no PBG D enzymatic activity. Although a *hemB* mutant complemented with a multi-copy plasmid harboring *hemB* showed a 25-fold increase over wild type in ALA D activity, only wild type level of PBG D activity was observed (Li et al., 1988a). This indicates that PBG is required for the appearance of PBG D activity. How PBG D activity is regulated by PBG has been studied by Umanoff et al in this laboratory and by Battersby and coworkers, Jordan and coworkers, Scott and coworkers. It was shown that when exogenous PBG is supplied to a *hemB* mutant at 40 μg/ml, PBG activity is normal (Umanoff et al., 1988). The reason for this observation is that a dipyrrylmethane (diPBG) is a covalently-bound cofactor of PBG D (Hart et al., 1987; Jordan and Warren, 1987; Miller et al.,1988; Jordan et al., 1988; Scott et al., 1988).

4. THE THIRD STEP: FORMATION OF UROPORPHYRINOGEN

Porphobilinogen deaminase (PBG D) catalyzes the deamination of four molecules of the monopyrrole, PBG, to form the linear tetrapyrrole, hydroxymethylbilane, which is nonenzymatically cyclized to uroporphyrinogen I (Battersby et al., 1980). In the presence of uroporphyrinogen III cosynthase, hydroxymethylbilane is rapidly converted to uroporphyrinogen III (Battersby et al., 1978; Battersby et al., 1980; Jordan and Berry, 1980). Deaminases from a wide variety of sources have been purified to homogeneity and all appear to be monomeric enzymes with a molecular weight ranging from 34,000 to 50,000.

In humans a deficiency in PBG D activity results in acute intermittent porphyria, a dominantly-inherited disease (Strand et al.,1970; Meyer et al., 1972). During acute attacks of this disease, hepatic δ-aminolevulinate synthase is induced; δ-aminolevulinate and porphobilinogen accumulate because of low PBG D activity; and the amount of heme produced is too low to retain the negative feedback control on ALA S. However, PBG D activity is increased in erythrocytes of patients in the pre-leukemic state (Pasanen et al., 1981) and in patients with malignant lymphoproliferative disorders (Epstein et al., 1983). The increase in PBG D is caused by the increased growth rate of these cells (Schoenfeld et al., 1988).

The structural gene for PBG D in humans has been assigned to the distal part of the long arm of chromosome 11 (11q23->11qter) (Wang et al., 1981). The PBG D genes from human (Raich et al., 1986), *E. coli* (Thomas and Jordan, 1986) yeast (Gellerfors et al., 1986), and rat (Stubnicer et al., 1984) have been cloned and the DNA sequences of human (Raich et al., 1986), rat (Stubnicer et al., 1988), and *E. coli* (Thomas and Jordan, 1986; Alefounder et al., 1988) have been determined. Unlike other *hem* genes, *hemC* , *hemD* , probably *hemG* and an additional unknown gene, cluster together and form an operon (Sasarman et al., 1987; Jordan et al., 1988; Alefounder et al., 1988). Due to alternative transcription of the single PBG D gene in human, two distinct mRNAs are formed, one is housekeeping and the other is erythropoietic-specific (Chretien et al., 1988; Grandchamp et al., 1987). These two mRNA species differ solely in their 5' termini and are translated into two isoforms of PBG D. An additional peptide of 17 amino acid residues at the NH_2 terminus of the non-erythropoietic isoform of PBG D accounts for its

higher molecular mass. Two tissue-specific factors bind to the erythroid promoter of the human porphobilinogen deaminase gene: one is NF-E1, which also binds to the promoter and the enhancer of the human β-globin gene, the other is called NF-E2 (Mignotte et al., 1989).

MATERIALS

1. CHEMICALS AND ENZYMES

Restriction enzymes were purchased from Boehringer Mannheim (Indianapolis, IN), International Biotechnologies (New Haven, CT), and New England Biolabs (Beverly, MA). T4 ligase, L-[^{35}S] methionine, [α-^{35}S]dATP, L-[^{14}C(U)]-glutamic acid, L-[1-^{14}C]glutamic acid, [2-^{14}C]-glycine were from NEN (Boston, MA). Media were prepared from Difco products (Detroit, MI). 5-ALA, PBG, protoporphyrin IX and hemin were purchased from Sigma (St. Louis, MO) and Porphyrin Products (Logan, UT). Levulinic acid, Dowex 50W8 (200-400 mesh), tRNA^{glu}, RNase (type I-AS), ATP, periodic acid, *p*-dimethylaminobenzaldehyde, PEG 8000, kanamycin, neomycin, actinomycin and glutathione were obtained from Sigma (St. Louis, MO). Gabaculin (3-amino-2,3-dihydro-benzoic acid hydrochloride) was purchased from Fluka Chemie AG (CH-9470 Buchs, Switzerland). Sephadex G-25 and NADPH were from Pharmacia Inc. (Piscataway, NJ). Molecular biology grade agarose, ampicillin, phenol, chloroform, acrylamide, X-gal, urea, and IPTG were from International Biotechnologies Inc. (IBI)(New Haven, CT), and acetylacetone, ethyl acetoacetate, N-methyl-N'-nitro-N-nitrosoguanidine were from Aldrich Chemical Co. Inc. (Milwaukee,WI). All reagents for oligonucleotide synthesis were purchased from American Bionetics, Inc. (Hayward, CA). The Sequenase and GeneScribe-Z kits were bought from United States Biochemical Corp. (Cleveland, OH). The Cyclone system was obtained from IBI. Other chemicals were from standard suppliers and were reagent grade or better.

2. STRAINS AND MEDIA

The strains used in the cloning of the *hemA* and *hemB* genes are listed in Tables 4 and 5, respectively. An *E. coli* W3110 library in λSE6 and plasmid pSE103 were obtained from American Type Culture Collection (Rockville, MD). The F' kit (Table 5), C600, and SASX41B were obtained from Dr. Barbara Bachmann, the Coli Genetic Stock Center (Yale University, New Haven, CT). SHSP19 and TBI/pU1553, which harbors the structural gene for ALA synthase from *Rhodobacter spheroides*, and CSR603 were kindly supplied by Mark Moore, University of Illinois at Urbana-Champaign, and Dr. David Calhoun, City College of New York, respectively. JM101 came with the GeneScribe-Z kit of U. S. Biochemical Corp..

Some strains and plasmids were used in the cloning of both the *hemA and hemB* genes for their advantages in cloning procedures. JM101 has a deletion in *lac-proAB* , thus allowing the easy selection of transformants carrying plasmids with foreign inserts in the *lacZ* gene on the plasmid by growth on X-gal supplemented medium. Transformants without the insert form blue colonies while those with inserts form white colonies. This strain also has the F pilus, which is required for infection by M13 phages which may be useful for sequencing procedure.

pSE103 contains the λ represser gene *c* I857 with a temperature-sensitive mutation. In the presence of pSE103, λSE6 acts as a single-copy plasmid at permissive-temperature and as a lytic phage at non-permissive temperature (Elledge and Walker, 1985). Thus, λSE6 is also called a phasmid. A toxic gene from other organisms can be cloned at the permissive-temperature because a single copy of the foreign gene has

Table 3. Strains, plasmids, and phages used in the cloning and sequencing of the *hemB* gene

	Relevant genotype and/or phenotype[a]	Source (reference)[b]
***E. coli* strains**		
C600		CGSC
RP522	C600, *hemB*	Li et al. (1988a)
RP523	RP522, heme permeable	Li et al. (1988a)
HU1000	RP523/pSE103	Li et al. (1988a)
HU1016	HU1000(phHU16)	Li et al. (1988a)
CSR603	*phr-1 recA uvrA6*	Sancar et al. (1978)
JL1002	RP523/pJL2	Li et al. (1988a)
JL1102	C600/pJL2	Li et al. (1988a)
JM101	Δ(*lac-proAB*) (F' *traD36 proAB* + *lacI*^{q} *lacZΔM15*)	U. S. Biochemical
JC7623	*recB recC sbcB*	Kushner et al. (1971)
Plasmids or phages		
λSE6	*E. coli* genomic library	Elledge and Walker (1985)
pSE103	λ cI857 kan^r	Elledge and Walker (1985)
pTZ18U		U. S. Biochemical
phHU16	Hem$^+$ (from λSE6)	Li et al. (1988a)
pJL1	Hem$^+$ *amp* (from pTZ18U)	Li et al. (1988a)
pJL2	Hem$^+$ *amp* (from pJL1)	
F' kit		CGSC

[a] Other genetic markers are: C600, *thr1 leuB6 thi1 lacY1 tonA21 supE44* λ$^-$,F$^-$; CSR603, *thr1, leuB6, proA2, argE3, thi1, ara14, lacY1, galK2, xyl5, mtl1, rpl31, tsx33, sup44,* λ$^-$,F$^-$; JM101, *supE, thi1* ; JC7623, *his4, ara14, thr1, thi1, leuB6, proA2, argE3, rpl31, lacY1, galK2, xyl5, tsx33, sup37* amber, λ$^-$,F$^-$.

[b] CGSC, Coli Genetic Stock Center, Yale University, New Haven, CT; ATCC, American Type Culture Collection, Rockville, MD; U. S. Biochemical, U. S. Biochemical Corp., Cleveland, OH.

Table 4. Strains, plasmids, and phages used in the cloning and sequencing of the *hemA* gene

	Relevant Genotype and/or Phenotype[a]	Source
***E. coli* strains**		
SASX41B	*hemA*	CGSC [b]
HU227	SASX41B, hemin-permeable	Umanoff et al. 1988
SHSP19	*hemA8*	Sasarman et al. (1968)
HU2000	HU227/pSE103	Umanoff et al. (1988)
HU2001	HU2000/phHU201	Umanoff et al. (1988)
JL1268	HU227/pJL68	Li et al. (1989c)
JM101	Δ(*lac-proAB*) [F'*traD36, proAB* + *lacI*^{q} *lacZDM15*]	US Biochemical
JC7623	*recB recC sbcB*	Kushner et al. (1971)
CRS603	*phr-1 recA uvrA6*	Sancar et al. (1978)
Plasmids and phages		
λSE6	*E. coli* genomic library	Elledge and Walker (1985)
pSE103	λcI857, kan^r	Elledge and Walker (1985)
phHU201	Hem$^+$ (from λSE6)	Umanoff et al. (1988)
pTZ19U		US Biochemical
pJL68	Hem$^+$ (from pTZ19U)	Li et al. (1989c)
pJL69	Hem$^+$ (from pJL68)	Li et al. (1989c)

[a] Other genetic markers are: SASX41B, *met1*, HfrCavalli; SHSP19, *met, lac, str*, F$^-$; JM101, *supE, thi1*; JC7623: *his4, ara14, thr1, thi1, leuB6, proA2, argE3, rpl31, lacY1, galK2, xyl5, tsx33, sup37* amber, λ$^-$,F$^-$ and CSR603, *thr1, leuB6, proA2, argE3, thi1, ara14, lacY1, galK2, xyl5, mtl1, rpl31, tsx33, sup44*, λ$^-$,F$^-$

[b] Coli Genetic Stock Center, Yale University, New Haven, CT.

Table 5. F' strains used for genetic mapping

CGSC #	Episome	Region carried by F'	Supplements
4251	F104	*thr argF*	arg, his, thi
4288	F128	*proA,B lac*	lactose, thi
6350	F254	*lac lip*	lactose, ser, thi
4287	F152	*fep gal*	galactose,thi,ura,trp,his,tyr
4253	F126	*nadA recE*	galactose, thi, his
4256	F123	*galU recE*	thi, ura, his
5760	F506	*man lpp aroD pps*	arg, thi
5505	F500	*relB man his*	mannose, met
4326	F150	*eda his*	leu, met, arg
4280	F129	*his dsdA*	leu, arg met
4279	F142	*tyr ptsI supN*	thi, his, trp, ura
4291	F143	*tyrA lysA*	thi, his, ura, trp
4254	F116	*fuc metC*	arg, his, leu, met
4248	F141	*argG asd*	his, leu, met
4289	F140	*rpoD mtl pyrE*	met, his, leu
4258	F111	*pyrE malB*	his, leu, arg
4265	F133	*ilv argH*	arg, his, leu
4260	F112	*metB pyrB uxuAB*	arg, his, leu
4255	F117	*melA pyrB*	thi, his, leu, pro, thr, thy

lower toxicity than multi-copies of same gene. It also permits the study of the regulation of a gene when a mutant is complemented with normal gene dosage in λSE6. If the cloned gene is on a multi-copy plasmid the expression of the gene is probably deregulated. JC7623 was chosen for the amplification of phasmid DNA.

CSR603 was used for the identification of the cloned gene products. When irradiated with ultraviolet light, the chromosomal DNA of CSR603 is damaged and eventually destroyed because CSR603 cannot repair the damage. Plasmid DNA survives the irradiation because of its small size and multiple copy number. The plasmid takes over the cellular transcriptional and translational machinery and directs the synthesis of proteins encoded by the plasmid. These proteins can be radiolabeled and identified by gel electrophoresis and autoradiography.

Other strains were constructed in this laboratory by standard phage infection, transfection and transformation procedures.

Plasmids pTZ18U, pTZ18R, pTZ19U, and pTZ19R in the GeneScribe-Z kit were obtained from U. S. Biochemical (Cleveland, OH). These plasmids were chosen as the subcloning vectors because they contained several important features: (1) multiple cloning sites within the *lac* Z' gene which allowed screening on X-gal plates, (2) the replication origin of f1 for rapid production of single-stranded plasmid DNA, (3) all useful orientations for cloning and sequencing available, (4) high copy-number origin for ultra-high plasmid yields, (5) a T_7 promoter 2-4 bases downstream from the cloning site.

LB (Luria-Bertani), TB, YT, media were prepared according to the recipes of Schleif and Wensink (1981). LB medium contained 10 g of tryptone, 5 g of yeast extract, and 10 g of NaCl per liter, YT medium

contained 8 g of tryptone, 5 g of yeast extract, and 5 g NaCl per liter, 2XYT medium contained 16 g of tryptone, 10 g of yeast extract, and 5 g NaCl per liter, TB medium contained 10 g of tryptone, and 10 NaCl per liter. M9, SM, and NZCYM were made according to the recipes of Maniatis et al. (1982). One liter of M9 medium contained 6 g of Na_2HPO_4, 3 g of KH_2PO_4, 0.5 g of NaCl and 1 g of NH_4Cl, 2 ml of 1 M $MgSO_4$, 10 ml of 20% glucose and 0.1 ml of 1 M $CaCl_2$. The last three ingredients were individually sterilized and added after sterilizing the first four components. For SM, 5.8 g of NaCl, 2 g of $MgSO_4.7H_2O$, 50 ml of 1 M Tris-Cl, pH 7.5), 5 ml of 2% gelatin and water were mixed and autoclaved to make 1 liter of medium. NZCYM contained 10 g of NZ amine (hydrolysate of casein), 5 g of NaCl, 1 g of casamino acids, 5 g of yeast extract, and 2 g of $MgSO_4.7H_2O$, per liter. The pH was adjusted to 7.5 with NaOH. Solid medium was made by adding 15 g of agar to 1 liter of medium. 7 g of agar per liter of medium was used for top agar.

The hemin solution (4 mg/ml) was made by mixing 0.1 g of hemin and 2.5 ml of Tween 80 in 22.5 ml of sterile water with 4 drops of 10 N NaOH. This was stored at 4° C. The final concentration of hemin in medium is 10 µg per ml.

The final concentration of amino acids and thiamine supplements in media were usually 50 µg and 5 µg per ml, respectively. Antibiotics ampicillin and kanamycin were added at a final concentration of 50 µg/ml.

EXPERIMENTAL

1. CLONING AND SEQUENCING OF THE *hemB* GENE OF *E. COLI*

1.1. CHARACTERIZATION OF THE *hemB* MUTANT

A heme-requiring, hemin-permeable mutant, RP523, isolated in this laboratory was provided for this work. This mutant was further characterized as follows.

1.1.1. Detection of the intermediates of the heme biosynthetic pathway accumulated in *hem* mutant

Porphyrins The method of McConville and Charles (1979b) was followed and modified to detect any porphyrin accumulated in the culture medium. Strain RP523 was inoculated into two 2 liter flasks each containing 400 ml of glucose minimal medium (M9) supplemented with leu, thr (50 μg/ml each), ALA (50 μg/ml), vitamin B_1 (5 μg/ml) $Na_2S_2O_3$ (10 μg/ml); one flask was also supplemented with hemin (10 μg/ml). The cells were grown with vigorous shaking at 37 ^{0}C for 5 days. Each day, an aliquot of the cell culture was observed under long-wavelength UV to look for porphyrin fluorescence.

PBG To detect PBG, 100 ml of cells grown to saturation (48 h) in M9 were centrifuged. The cell pellet was washed with sodium phosphate buffer (50 mM, pH 5.8) and resuspended in 1 ml of the same buffer. The cells were disrupted by sonication (10 sec pulses with 10 sec rests for a total of 90 sec of sonication). Both the cell extract and the supernatant of the cell culture were assayed for PBG. 600 μl of supernatant or 20 μl of cell extract and 580 μl of water, was mixed with 600

µl of modified Ehrlich's reagent. The modified Ehrlich's reagent was prepared as follows: 0.7 g of $HgCl_2$ were dissolved in 168 ml of glacial acetic acid while the mixture was gently warmed, followed by addition of 40 ml of 70% perchloric acid to the cooled mixture. The resulting solution was diluted to 220 ml with glacial acetic acid. 1.0 g of *p* - dimethylaminobenzaldehyde (DMAB) were dissolved in 55 ml of the above solution prior to use (Sassa, 1982). After 10 min incubation of material with modified Ehrlich's reagent, the spectrum from 500 to 650 nm was recorded. In this assay the PBG-DMAB compound has a molar absorption coefficient of $6.1X10^4$ at 553 nm and has another absorbance peak at 525 nm with a molar absorption coefficient of $5.0X10^4$ (Mauzerall and Granick, 1956).

ALA Cells were grown overnight in LB medium supplemented with hemin (10 µg/ml). 50 ml of cells were centrifuged, washed with 10 mM Tris-buffer (pH 7.5), resuspended in 1 ml of the same buffer, and disrupted by sonication as described above. The supernatant of the cell culture and the cell extract were assayed for ALA accumulation using Burnham's method (1970). 1 ml of the supernatant, or 50 µl of the cell extract plus 0.95 ml of the buffer, was mixed with 0.5 ml of 10% TCA. After centrifugation in a microfuge for 5 min, the supernatant was transferred to a tube containing 2 ml of 1 M sodium acetate buffer (pH 4.7). 50 µl of acetylacetone was added and the tubes were placed in a boiling water bath for 15 min. Upon cooling, an equal volume of modified Ehrlich's reagent was added and after incubation for 10 min, the spectrum from 450-650 nm was measured. The condensation product of ALA with acetylacetone (ALA pyrrole) gives a pink color with modified Ehrlich's

reagent with maxima at 553 nm, and a shoulder at about 525 nm. The ratio of the latter to the former is 0.69. The molar absorption coefficient at 553 nm is $6.8X10^4$ (Mauzerall and Granick, 1956).

1.1.2. Genetic mapping of the mutated *hem* gene

Strain RP523 was cross-streaked against the F' donor strains (Table 5) on minimal medium supplemented with thr and leu. The plates were incubated at 37° C for 24 h and complementation to heme prototrophy was assessed.

1.1.3. Assays of ALA D, PBG D, and ferrochelatase

RP523 and its derivative strains, HU1016, JL1002 and JL1102 were assayed for ALA D, PBG D and ferrochelatase activity. Cells were grown overnight in LB medium. Strains harboring plasmids were selected by being grown in medium supplemented with 50 µg/ml of amp.

ALA D assay The procedure of Sassa (1982) was scaled up four times and modified slightly. 50 ml of cells were washed with 50 mM sodium phosphate buffer (pH 6.4), resuspended in 1 ml of the same buffer, and sonicated for 90 sec as described previously. 20 µl of the cell extract was mixed with 80 µl of the buffer and 200 µl of assay mix (50 mM sodium phosphate, pH 6.4, 20 mM DTT, 8 mM ALA) and incubated for 1.5 h. 0.6 ml of 6% TCA-0.1 M $HgCl_2$ was added to stop the reaction. After centrifugation in an Eppendorf microfuge (5415) for 5 min at full speed, 0.6 ml of supernatant was mixed with an equal volume of modified Ehrlich's reagent. After 10 min incubation, the absorbance was determined by taking a difference spectrum between the sample which

had been incubated with complete assay mix and the control which had been incubated with the mix devoid of ALA over the wavelength range of 500-650 nm (PERKIN-ELMER, Lambda 3B). Because the absorption of the Ehrlich-PBG compound is zero at 650 nm, the absorbance was set at zero at this wavelength. Net absorbance at 553 nm represents the enzymatic formation of PBG from ALA. ALA D activity was reported as A_{553} per milligram of protein. The Ehrlich-PBG compound has a molar coefficient of 6.1×10^4 at 553 nm.

PBG D assay 50 ml of cells were washed with 0.1 M Tris-Cl (pH 8.2), resuspended in 1 ml of the same buffer, and sonicated as above. 20 µl of the extract, 10 µl of PBG (1 mg/ml in 0.1 M Tris-Cl, pH 8.2), and 70 µl of 0.1 M Tris-Cl (pH 8.2) were incubated at 37° C for 2 h in the dark; 1 ml of 1 M HCl was then added, and the precipitate was removed by centrifugation. The spectrum of the mixture from 350 to 450 nm was recorded against the control in which the extract had been incubated without PBG. The Soret band for uroporphyrin in HCl is at 405 nm with a molar absorption coefficient of 5.18×10^5. Activity was reported as A_{405} per milligram of protein.

Ferrochelatase assay The procedure of Porra and Jones (1963) was followed for the ferrochelatase assay. 100 ml of cells were washed with 47.6 µM potassium phosphate buffer (pH 7.8), resuspended in 2 ml of the same buffer and sonicated as above. 1.1 ml of the extract, 2.1 ml of 95.2 µM K phosphate, pH 7.8, 95.2 µM protoporphyrin IX and 0.6 ml of water were mixed in a Thunberg tube. In the side arm was 0.4 ml of 95.2 µM reduced glutathione and 952 µM $FeSO_4$. The tube was

evacuated and flushed with N_2 gas repeatedly over 30 min at 4° C. The reaction was started by mixing both parts and incubating at 37° C. After 1.5 h, the reaction was stopped by adding 1 ml of pyridine, 0.5 ml of 1 N NaOH, and 1 ml of water sequentially. The resulting solution was divided into two 3 ml aliquots. 10 mg sodium dithionite ($Na_2S_2O_4$) was added to one tube and 50 µl of 3 mM $K_2Fe(CN)_6$ was added to the other, respectively. After mixing , the difference spectrum of the reduced and the oxidized pyridine haemochromogen was recorded from 500-650 nm. Activity was reported as nanomoles of heme produced per mg of protein (nanomoles per mg). The formula used is as follows:

$$\text{heme (nanomoles)} = \frac{10^3 v.\Delta E}{\Delta \varepsilon_{mM}}$$

where v is the volume (ml) of alkaline pyridine solution and ΔE represents the difference in extinction between the maximum of the α-band and the minimum occuring between the α- and β-bands. The values of $\Delta \varepsilon_{mM}$ for pyridine protohaemochromogen is as follows:

$$\varepsilon_{mM}^{557} - \varepsilon_{mM}^{541} = 20.7$$

Protein assay Protein concentrations of all cell extracts were determined by the Bio-Rad method (Bio-Rad Laboratories, Richmond CA).

1.2. CLONING PROCEDURES

1.2.1. Amplification of phasmid phage and isolation of phasmid DNA

Strains were constructed which were either *hemA* or *hemB* mutants, and which were complemented by an *E. coli* genomic library in λSE6. The *hemA* and *hemB* genes were carried on single copy phasmids which could be induced to form λ lysates. HU2001 is a strain

carrying the *hemA* gene on an insert in a phasmid and HU1016 is a strain which carries *hemB* gene.

Isolation of phasmid DNA from transductant Phasmid DNAs were isolated from 18 of the 30 transductants from HU1000 by using a mini-prep procedure (de Bruijin and Ausubel, 1981). The DNAs were used to transfect JC7623 (Davis et al., 1980). Each transductant was grown in 10 ml of LB medium supplemented with 50 μg per ml of kanamycin overnight at 30° C. Cells were pelleted by centrifuging at 6000 rpm at 4° C for 5 min. The pellet was resuspended in 50 μl of 25% sucrose by vigorous vortexing. To the cell suspension 300 μl of M-STET (5% Triton X-100, 50 mM Tris, pH 8.0, 5% sucrose) was added and mixed. The mixture was transferred to an Eppendorf tube and 25 μl of freshly prepared lysozyme (10 mg/ml) in H2O was added. The tube was heated in boiling water for 1 min. After 15 min centrifugation in an Eppendorf microfuge the supernatant was transferred to a fresh tube and DNA was precipitated by adding an equal volume of isopropanol, freezing at -80° C for 30 min and recentrifuging for 15 min in the Eppendorf microfuge. The pellet was dried and resuspended in 100 μl of H_2O, stored at -20° C.

Transfection 1 ml of an overnight culture of JC7623 in TB medium supplemented with 0.2% maltose, was inoculated into 50 ml of LB with maltose (0.2%). The cells were grown to mid-log phase (Klett 60) at 37° C, pelleted at 6000 rpm in a Sorvall SS34 rotor for 6 min at 4° C, resuspended in 25 ml of 50 mM $CaCl_2$ and incubated on ice for 15 min. The cells were pelleted again, resuspended in 2.5 ml of 50 mM $CaCl_2$, and incubated for 15 min. 0.1 ml of $CaCl_2$-treated cells, 40 μl of phasmid DNA, and 60 μl of TE buffer (10 mM Tris-Cl,1 mM EDTA) (pH 7.4) were

mixed and the resulting mixture was incubated on ice for 45 min, then heat-shocked for 90 sec at 42° C and put back on ice immediately. It was then mixed with 0.2 ml of indicator bacteria (mid-log phase or overnight JC7623, in this case) and 3 ml of melted top agar (55° C), and the mixture was poured on a prewarmed TB plate. The plate was incubated at 37° C overnight to get plaques.

Transduction The phage obtained as above was amplified once by transduction. Single plaques were picked from the incubated plates with an inoculating loop and suspended in 0.5 ml of TB. After vortexing 0.2 ml of overnight culture of JC7623 and 3 ml of top agar were added. The entire contents of the tube were poured on a TB plate.

After overnight incubation, 3 ml of TB was added to the top of the plate. The top agar overlay was scraped off and transferred to a test tube, and soaked for 30 min. 0.5 ml of chloroform was added and the tube was vortexed vigorously and centrifuged at 6000 rpm for 5 min. The supernatant was transferred to a fresh tube and a few drops of chloroform were added. The resulting mix is called phage stock.

Titering of λ phage A serial dilution of the phage stock was made from 10^{-2} to 10^{-7}. 0.1 ml of each dilution was mixed with 0.2 ml of an overnight culture of JC7623 and 3 ml of melted top agar, and the mixture was poured onto TB plates. After incubation at 37° C overnight , plaques were counted and titer was determined (pfu/ml). The phage, which yield the highest titer (4.5×10^{10}), was called phHU16 and chosen for further study.

Amplification of phasmid phage and full-scale preparation of phasmid DNA Phage phHU16 was amplified according to the procedure of Maniatis et al. (1982) and the phasmid DNA was isolated using the method of Kaslow (1986) with slight modifications. JC7623 was grown overnight in NZCYM at 37° C with vigorous shaking. 10^{10} cells were withdrawn (1 OD_{600}=$8x10^8$ cells/ml), pelleted, and resuspended in 3 ml of sterile SM. After mixing with $5x10^8$ phages prepared as above, the cells were incubated for 20 min with intermittent shaking. The mixture was then inoculated into 500 ml of prewarmed NZCYM medium and grown with vigorous shaking at 37° C for 14 h. 10 ml of chloroform was added and the incubation was continued for 30 min. After the lysate was cooled to room temperature, DNase (pancreatic) and RNase A were added to a final concentration of 1 μg/ml each. 30 min later, NaCl (29.2 g) was added to a final concentration of 1 M and the mixture was kept on ice for 1 h . Cell debris was removed by centrifugation (11,000 g at 4° C for 10 min). PEG 8000 was added to the supernatant to a final concentration of 10%. λ phages were collected by centrifugation (same as above) after 1 h chilling on ice, and resuspended in 3 ml of SM. DNase and RNase A were added to a final concentration of 5 μg/ml and 100 μg/ml, respectively. After 30 min incubation at 37° C, 150 μl of 10% SDS, 120 μl of 0.5 M EDTA (pH 7.5) and 15 μl of proteinase K (20 mg/ml) were added. The mixture was heated at 68° C for 30 min. The mixture was extracted with phenol as described by Maniatis (1982), phenol/chloroform (1:1, v/v), and chloroform, and precipitated by adding 0.7 ml of 10.5 M NH_4OAC and 2 volumes of ethanol and incubating at -70° C for 30 min. DNA was recovered by centrifugation (same as above). To the DNA pellet, 1.7 ml of water, 0.32 ml of 5 M NaCl, and 2.0 ml of 13% PEG 8000 were added and

mixed. After 1 h incubation at 4° C, DNA was obtained by centrifugation, dried under N_2 gas, and dissolved in 1 ml of sterile water. DNA concentration and purity were estimated by measuring the absorbance at 260 nm and 280 nm, and determining the ratio of A_{260}/A_{280}, respectively. The purity and concentration, and the sizes of restriction-digested DNA were checked by agarose gel electrophoresis (for agarose gel electrophoresis, see below). Sometimes the procedure was repeated, starting from the proteinase K digestion.

The phasmid DNA and the phage from JC7623 were tested by transfection and transformation of HU1000 to heme prototrophy.

1.2.2. Subcloning and screening of transformants

Restriction digestion Phasmid DNA was partially-digested with Sau3A at 37° C for 2 min to get a broad range of molecular sizes. The vector plasmids, pTZ18U and pTZ18R, were digested with BamHI at 37° C for 2 h. A typical restriction digestion was as follows:

(1) In an Eppendorf tube the followings were mixed

A 20 µl reaction mixture contained 2 µl of 10X buffer (usually supplied with enzyme), 1-4 units of enzyme, about 1µg of DNA and H_2O to make up the total 20 µl.

(2) The mixture was incubated at 37° C for 2-4 h for most digestions, otherwise supplier's instructions were followed. Sau3A partial digestion was performed for different periods of time. In this case , 2 min digestion resulted in the correct size distribution of fragments which were from hundreds to thousands of bp. Longer digestion generated fragments which were too small for cloning.

(3) The reaction was stopped by heating at 70° C for 5 min to inactivate the enzyme, and chilled on ice to prevent annealing.

Agarose gel electrophoresis The sizes of restriction fragments and the quality of the digestion were determined by agarose gel electrophoresis. The concentration of agarose (usually 0.7-1%) used depended on the size of the fragment to be separated. 10 cm long minigels were run in 0.5XTBE at 90 V for 2 h. Stock solution of 10XTBE was made by mixing 108 g of Tris base, 55 g of boric acid, and 40 ml of EDTA, pH 8.0. Regular 20 cm long gels were run in 1XTBE at 35 V overnight. Hind III-EcoR I digested DNA fragments were used as the standards.

Ligation The digestion mixture was extracted with phenol, phenol/chloroform, and chloroform as described earlier. The DNA was precipitated by adding 0.2 volumes 10.5 M NH_4OAC and 2 volumes of ethanol and freezing at -70° C for 30 min. After centrifugation in the Eppendorf microfuge for 15 min at 4° C, the pellet was dried and resuspended in 10 µl of ligation buffer (50 mM Tris-Cl, pH 7.6 10 mM DTT, 1 mM ATP, 50 µg/ml BSA, 10 mM $MgCl_2$). The Sau3A-digested phasmid DNA and the BamHI digested vector DNA's were ligated with T4 ligase at 15° C for 2 h. In 20 µl of ligation reaction, 0.5-1 unit ligase was used for 1 µg of total DNA.

Transformation of JM101 JM101 was transformed with the ligation mix and transformants were screened on X-gal supplemented medium according to the procedure of the supplier, U. S. Biochemical.

0.5 ml of JM101 overnight culture was inoculated into 200 ml of 2XYT. The cells were grown to an absorbance of 60 units on a Klett-Summersen colorimeter, pelleted by centrifugation at 6000 rpm for 6 min at 4° C, and resuspended in 50 ml of ice-cold, sterile 0.1 M $MgCl_2$. After 20 min of incubation on ice, the cells were pelleted again and resuspended in 5 ml of 0.1 M $CaCl_2$, kept on ice for 1 h, then divided into 0.1 ml of aliquots. The ligated DNA (4-100 ng) was added to each 0.1 ml of cells, incubated for 45 min, then heat-shocked for 90 sec at 42° C and chilled on ice. 0.5 ml of 2XYT was added to each tube and incubated for 1 h at 37° C. 0.1 ml of each mixture was spread on solid medium containing IPTG (final concentration 0.1 mM), amp and X-gal (final concentration 50 µg/ml each). The plates were incubated at 37° C overnight.

White colonies were purified on fresh medium. Plasmid DNA was isolated from 1.5 ml of overnight culture of each transformant using the mini-prep procedure of U. S. Biochemical. The extraction with phenol and chloroform was not performed because it was not necessary for transformation.

Screening of plasmids RP523 was transformed with the plasmid DNA from each white colony to heme prototrophy. The procedure used is similar to the one above except RP523 was grown in LB with 10 µg/ml of hemin and the cells were treated with 50 mM $CaCl_2$ twice, instead of 0.1 M $MgCl_2$ once and 0.1 M $CaCl_2$ once. Transformed cells were spread on LB plates containing amp (50 µg/ml) to screen for heme prototrophy.

Subcloning Plasmid, pJL1 which has an insert of 7 kb and has the ability to complement the *hemB* mutation, was restriction-mapped

with EcoR I, Pst I, BamHI, Sal I, HindIII, Bgl II, Sac I, Xbal, Xho II, KpnI, and MboI. Subcloning was performed to look for the smallest fragment which still contained the gene activity.

1.3. IDENTIFICATION OF THE CLONED GENE PRODUCT

Maxicell method CSR603 was transformed with pTZ18U and pJL2 as above. The method of Sancar et al. (1979) was followed with the following modifications: (a) strain CSR603 and its derivatives were grown in M9 with 1% casamino acids and supplements of thr, leu, pro and arg at 50 µg/ml of each and thiamine at 5 µg/ml; (b) cells were irradiated at a density of $2x10^8$ cells/ml in a glass petri dish with rotary shaking for 6 sec. at a distance of 60 cm from a Westinghouse Sterillamp G15T8; (c) cells were grown at 37° C for 2 h. after irradiation and then D-cycloserine was added at a final concentration of 4% µg/ml (Perbal, 1984), with continued incubation for 14 h; (d) after pelleting and washing, cells were suspended in buffer depending on the enzyme assay to be performed and sonicated if enzyme assay and non-denaturing electrophoresis were to be performed.

0.5 ml of an overnight of strain culture of CSR603 or its derivatives was inoculated into 10 ml of M9 medium supplemented with leu, thr, pro, arg, and B1, and grown at 37° C with shaking. Cells were grown to an absorbance of 30 Klett units and were irradiated with UV (5 ml/each 10-cm glass petri dish), then transferred to a fresh flask, and its absorbance reading was recorded. The cells were then grown for 2 h, and 20 µl of D-cycloserine (1 mg/ml) was added. After overnight growth the cells were spun down, washed with 10 ml of M9 buffer, resuspended in 5 ml of M9 medium lacking $MgSO_4$, supplemented with the nutrients as above, and starved for 1 h. 1.5 µl of ^{35}S-methionine (18.5 µCi) was added to the

culture. One h later, the cells were centrifuged down and washed with M9 buffer. For SDS-PAGE the cells were resuspended in 0.2 ml of sample buffer (2% SDS, 5% mercaptoethanol, 10% glycerol, 0.0005% bromophenol blue, and 0.0625 M Tris-Cl, pH 6.8), and boiled for 2 min, applied to the gel, or stored at -20° C. If enzyme assays and non-denaturing gel electrophoresis were to be performed, the cells were resuspended in 50 mM sodium phosphate buffer (pH 6.4) for the ALA D assay, or 0.1 M Tris-Cl (pH 8.2) for the PBG D assay.

Polyacrylamide gel electrophoresis Discontinuous polyacrylamide gels were made using the modified procedure of Laemmli (1970). To make 50 ml of separating gel solution for two 14X10X0.3 cm gels, 16.65 ml of 30% acrylamide (2.7% bisacrylamide) , 12.5 ml of 1.5 M Tris-Cl, pH 8.8, 0.5 ml of SDS, and 20 ml of water were mixed and the resulting solution was deaerated under vacuum. 250 µl of 10% ammonium persulfate and 5 µl of TEMED were added and mixed gently, then the gel was poured. The final concentration of each component was 10% acrylamide, 0.375 M Tris-Cl, pH 8.8, 0.1% SDS, 0.1% ammonium persulfate (w/v), and 0.07% TEMED (v/v). Isobutanol was layered on the top of the gel gently. When the gel set (usually 30 min to 1 h) the isobutanol was removed and the gel surface was washed with deionized water. Excess water was removed by a piece of paper towel. To make 15 ml of stacking gel solution, 2 ml of 30% acrylamide, 3.75 ml of 0.5 M Tris-Cl, pH 6.8, 0.15 ml of 10% SDS, and 9.0 ml of water were mixed and deaerated. 75 µl of 10% ammonium persulfate and 6 µl of TEMED were added and mixed. The final concentration of each component was 4% acrylamide, 0.125 M Tris-Cl, pH 6.8, 0.1% SDS, 0.05% ammonium

persulfate (w/v), and 0.05% TEMED (v/v). The resulting solution was poured on the top of the separating gel. A comb was inserted. After polymerization took place, the comb was removed and the gel was pre-run for 30 min in electrophoresis buffer (0.025 M Tris-Cl, pH 8.3, 0.192 M glycine, 0.1% SDS). Samples were applied and the gels were run at 24 mA until the dye (bromophenol blue) was 1 cm from the bottom of the gel. The unit (Bio-Rad, Model 220) was kept cool by the circulation of cold tap water.

The gel was stained in 0.125% Coomassie blue R-250, 50% methanol, and 10% acetic acetic acid for 1 h, destained in 50% methanol, and 10% acetic acid for 1 h (several changes). The gel was then soaked in 7% acetic acid, 5% methanol overnight, dried under vacuum at 60° C for 1 h and autoradiographed on X-ray film (Kodak X-Omat 5). Sometimes the gel was fixed in 7% acetic acid for 30 min, soaked in Amplify solution (Amersham) for 30 min, and then dried and autoradiographed without staining.

Conditions for non-denaturing gel electrophoresis was similar except that SDS was not included and the samples were prepared differently. One of four duplicate gels was cut into 0.5 cm slices for the ALA D assay, the second for the PBG D assay, the third was for a heme binding test, and the last was used for autoradiography. Proteins were extracted from each gel slice by mashing and soaking the slices in 100 µl of appropriate buffer for 2 h at 4° C. After centrifugation the supernatant was used for the assays.

Hemin binding test The procedure of Fagard and London (1981) was used for the hemin binding test. The protein solution

extracted from the gel slices and a hemin solution (6 μM) were scanned respectively on the spectrophotometer (PERKIN-ELMER, Lambda 3B) against buffer (0.01 M Tris-Cl, pH 7.5, 0.1 M KCl, 0.5 mM $Mg(OAC)_2$ from 500 to 350 nm. Next, hemin was added to the protein extract and the protein-hemin mixture was scanned against the buffer. Afterwards, the same amount of hemin was added to the buffer and the protein-hemin mixture was scanned using the buffer-hemin mixture as a blank. Bovine globin (heme-depleted) was used as a positive control.

1.4. SEQUENCE OF THE CLONED GENE

1.4.1. Single-stranded plasmid sequencing

Isolation of plasmid High quality pJL2 plasmid DNA was isolated using a procedure from Promega Biotec (Madison, WI). 250 ml of JL1002 was grown overnight in LB with amp (50 μg/ml). Cells were centrifuged (6000 rpm for 5 min at 4° C in Sorvall SS-34 rotor), resuspended in 6 ml of freshly prepared buffer (25 mμ Tris-HCl, pH 7.5, 10 mM EDTA, 15% sucrose, 2 mg/ml lysozyme), and incubated in ice-water for 20 min. 12 ml of 0.2 M NaOH, and 1% SDS was added and mixed carefully and thoroughly by inversion, and incubated in ice-water for 10 min. 7.5 ml of 3 M sodium acetate, pH 4.6 was added, mixed, and incubated for 20 min in ice-water. After centrifugation at 15000 rpm for 15 min the supernatant was transferred to another tube, to which 50 μl of RNase A (1 mg/ml) was added. The mix was incubated in ice-water for 20 min, then extracted twice with an equal volume of phenol:chloroform. The DNA was precipitated with 2 volumes of ethanol. After centrifugation the DNA pellet was dissolved in 1.6 ml of water, mixed with 0.4 ml of 4 M NaCl and 2 ml of 13% PEG. The DNA mixture was incubated for 1 h in ice-water.

After centrifugation, the DNA precipitate was washed with 70% ethanol, dried, and dissolved in 20 µl of H_2O.

Subcloning for sequencing pJL2 DNA prepared as above was double digested with BamHI+Sal1, Sal1+HindIII and HindIII+EcoRI. Fragments from these digestions were ligated with plasmids pTZ18U and pTZ19U digested with the same enzymes, respectively. JM101 was transformed with the constructed plasmids. Plasmid DNAs from several colonies of each set were isolated and restriction-digested with appropriate enzymes. The digested DNAs were run on agarose gels to identify the clones which contained the correct inserts.

Isolation of single-strand plasmid Single stranded plasmid DNAs were isolated by the method of U. S. Biochemical. 1 ml of overnight culture of a JM101 transformant which carried a plasmid with an insert to be sequenced was inoculated into 50 ml of 2XYT with amp, and grown for 30 min at 37° C with vigorous shaking. The culture was infected with 133 µl of helper phage M13K07 (the titer of the phage stock was $2X10^{11}$ plaque forming units/ml). After 30 min growth, 70 µl of 50 mg/ml kanamycin was added and the culture was grown overnight. Cells were removed by centrifugation (17000 rpm, 15 min, 4° C). To the supernatant, 1/4 volume of 20% PEG 8000, 3.5 M ammonium acetate was added. The resulting mixture chilled on ice for 30 min and was then centrifuged (17000g, 15 min, 4° C). The pellet was resuspended in 0.2 ml of TE buffer (pH 7.4), extracted with an equal volume of phenol/chloroform six times, and finally with chloroform once. DNA was precipitated by adding 0.1 ml of 7.5 M ammonium acetate and 0.6 ml of ethanol and freezing at -70° C

for 30 min. After centrifugation the pellet was washed with 95% ethanol, dried, and resuspended in 10 µl of distilled water.

The inserts were sequenced using the Sequenase kit from U. S. Biochem. and following their instructions (see below for detail).

1.4.2. M13 sequencing

Subcloning into M13 vectors pJL2, M13mp18 and M13mp19 were digested with BamHI and EcoRI, respectively. The insert of pJL2 which contains *hemB* gene was subcloned into M13mp18 and M13mp19 by ligations. The ligated DNAs were used to transfect JM101. The transfection procedure was similar to JC7623 transfection except that the growth medium for JM101 was 2XYT, and the top agar contained X-gal and IPTG for easy screening of plaques containing inserts.

Isolation of single strand M13 phage DNA Single strand M13 phage DNA was prepared using the Cyclone procedure (IBI). An overnight culture of JM101 was diluted 50-fold and divided into 2 ml aliquots. Phages from single plaques were inoculated each into 2 ml of the culture and grown with vigorous shaking at 37° C for 6-7 h. Cells were removed by centrifuging in the Eppendorf microfuge (10 min at highest speed). 0.25 ml of 20% PEG 8000/3.5 M ammonium acetate was added to 1 ml supernatant of each lysate. After mixing by inversion, the mixture was incubated on ice for 1 h, then centrifuged to pellet the phages. All the supernatant was removed. Each pellet was resuspended in 0.1 ml of TE (pH 8.0) extracted with phenol, phenol/chloroform, chloroform, and the phage DNAs were precipitated as above (isolation of single strand plasmid DNA).

Construction of nested deletions A series of nested deletions of the insert in the phage vector was constructed with T4 DNA polymerase (Dale et al., 1985) using the Cyclone system (IBI). In this procedure, single-strand phage DNA was hybridized with a specific oligonucleotide to form a restriction site. Different oligomers were used for M13mp18 and M13mp19. In M13mp18, a HindIII site was formed; in M13mp19, an EcoRI site was formed. The annealed molecule was then digested with appropriate restriction enzyme for 1-2 h. The linearized molecule was selectively digested by utilizing the 3' to 5' exonuclease activity of T4 DNA polymerase. Aliquots were removed at specific time points and rapidly inactivated by heat so as to generate an overlapping set of molecules. The deleted molecules were then tailed with dATP (for M13mp18) or dGTP (for M13mp19) using terminal deoxynucleotide transferase. Specific primer was annealed to the deleted product joining the two ends of the molecule. T_4 DNA ligase was used to seal the remaining nick. The "bandaid" recombinants were used to transform JM101. The deleted phage DNA was isolated from each transformant and the size of the insert after deletion was determined by agarose gel electrophoresis. The appropriate number of the deletion mutants, which could cover the whole BamHI-EcoRI fragment of pJL2, was chosen for sequencing.

1.4.3. Sequencing reactions and electrophoresis

Dideoxynucleotide chain termination In sequencing experiments, Sequenase (U. S. Biochemical, OH), a modified T4 DNA polymerase which has high processivity, low 3' to 5' exonuclease activity, high speed, and the efficient use of nucleotide analogs important for DNA sequencing, was used as the polymerase. In most cases, an oligonucleotide, 5'-

GTAAAACGACGGCCAGT-3', was used as the primer. Purified single-stranded phage DNAs with M13mp18 and M13mp19 as the vectors were utilized as the templates. The chain terminating nucleotide analogs were 2',3'-dideoxynucleotide 5'-phosphates (ddNTPs) (Sanger et al., 1977). [α-^{35}S]dATP was used to label the newly synthesized DNA fragments. The DNA synthesis was carried out in two steps. The first was the labeling step, in which primer was extended using a limiting concentration of deoxynucleoside triphosphates, including radioactive labeled dATP. In the second step, the concentration of all the nucleoside triphosphates was increased and a didideoxynucleoside triphosphate was added. Processive DNA synthesis occurred until all growing chains were terminated by a dideoxynucleotide. In instances of ambiguity caused by compression, dITP was used to substitute for dGTP.

A typical sequencing reaction was conducted as follows:

(1) Annealing template and primer

1 µl of primer (0.5 pmol/µl), 2 µl of 5X sequencing buffer (0.2 M Tris-Cl, pH 7.5, 0.1 M $MgCl_2$, 0.25 mM NaCl), and 7 µl of DNA+H_2O were mixed to make a total volume of 10 µl. The mix was incubated at 65° C for 2 min, then cooled down slowly over a period of 30 min.

(2) Labeling reaction

The annealing mixture was mixed with 1 µl of 0.1 M DTT, 2 µl of labeling mix (1.5 µM of dGTP, dCTP and dTTP, obtained by 5 fold dilution of the original concentrated solution), 0.5 µl of [α-^{35}S]dATP (10 µCi/µl), and 2 µl (3 units) of Sequenase. The resulting mixture was incubated for 5 min at room temperature.

(3) Termination reaction

Each 3.5 µl of the mixture obtained above was mixed with 2.5 µl of

ddG, ddA, ddT, and ddC termination mix, respectively, which were prewarmed 1 min at 37° C. Each termination mix contained 80 µM dATP, 80 µM dGTP, 80 µM dCTP, 80 µM dTTP, 50 mM NaCl, and 8 µM of ddATP, or ddGTP, ddCTP, ddTTP). After 5 min incubation at 37° C, 4 µl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% Xylene Cyanol FF) was added to each tube. The mixtures were heated at 78° C for 2 min immediately before loading 3 µl per lane for electrophoresis.

Sequencing gel electrophoresis A standard thermoplate sequence gel electrophoresis unit (IBI, Model STS-45) was used to analyze the sequencing reactions. To make 100 ml of gel solution, 20 ml of 40% acrylamide (2% bisacrylamide), 10 ml of 10XTBE, 48 g of urea, and appropriate amount of water were mixed. After urea was dissolved at 37° C by stirring, water was added to bring the volume to 100 ml. The solution was filtered and 0.6 ml of 10% ammonium persulfate and 40 µl of TEMED were added and mixed gently. Immediately the gel was poured by capillary action using a pair of wedged spacers which allowed more bases to be read. The concentration of acrylamide varied according to the length of DNA fragments being sequenced. The gels were run, dried, and autoradiographed according to the suggestions of the suppliers (U. S. Biochemical, IBI)

1.5. ANALYSIS OF DNA SEQUENCES

Nucleotide sequences were read on an IBI Gel Reader (IBI). IBI DNA/Protein Sequence Analysis and DNA Inspector II+ software (Textco, NH) were used to analyze the sequences. Promoter searches were

carried out using the method of Mulligan and McClure (1986). Computerized databanks of nucleotide sequences (EMBL, Nucleic Acid Sequence Database: release 15.0 and NIH GenBank: version 55) were searched using the XFASTN, XFASTP and IFIND protocols available on Bionet (Intelligenetics, Inc., CA).

2. CLONING AND SEQUENCING OF *hemA* GENE OF *E.COLI*

A *hemA* mutant, SASX41B, and its derivatives were used for the purpose of cloning the gene by complementation, using the same strategy as the one used for the *hemB* gene. The mutation was located at 27 min on the *E. coli* genetic map. The gene products coded by the cloned insert were analyzed by the maxicell procedure and by SDS-PAGE gels. The insert was subcloned into M13mp18 and M13mp19 and sequenced using the same strategy as above. Oligonucleotides synthesized on the Cyclon™ Plus DNA Synthesizer (Milligen/Biosearch) were used as primers to fill two gaps left by the nested deletion method. The oligonucleotide primers were designed following the rules of Barnes (1987). (1) The following formula was used to determine the length of the primer: primer length = 18 + 1 extra nucleotide for each 2% deviation from 50% GC content of the DNA; (2) The sequences of the primers were checked with IBI DNA/Protein sequence analysis software for accidental base pairing with undesirable locations on the templates. The sequence was analyzed using the same computer software described above.

3. ALA SYNTHESIS

3.1. BACTERIAL GROWTH CONDITIONS

JL1268, which contains a multi-copy plasmid harboring a gene

which complements a *hemA* mutant allele (SASX41B), was grown as follows. (a) For the radiolabeling experiments, 1 ml of JL1268 overnight culture was inoculated into 100 ml glucose minimal medium supplemented with methionine, glutamic acid and amp (50 µg/ml each). After growth for 24 h at 37° C, cells ($5x10^8$ cells/ml) were centrifuged, washed with fresh medium, recentrifuged and the cell paste was resuspended in 20 ml of fresh medium, as above, but also containing levulinic acid, an inhibitor of ALA dehydratase, at 4 mM concentration. (b) For the preparation of cell-extracts of JL1268 and for growth of TBI/pU1553, a different *E. coli hemA* mutant strain which harbors a plasmid with the structural gene for ALA synthase from *Rhodobacter spheroides*, cells were grown overnight to saturation (> $2x10^9$ cells/ml) at 37° C in LB medium with amp (50 µg/ml).

3.2. PREPARATION OF BACTERIAL EXTRACTS

Ten 100 ml cultures of *E. coli* JL1268 were grown for 16 h at 37° C in LB medium with amp (50 µg/ml). Cells (>$2x10^9$ cells/ml) were harvested by centrifugation at 6000 rpm at 4° C, washed with 0.1M Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT, resuspended in 30 ml of the same buffer, and sonicated in 3 ml aliquots for 60 sec (10 sec on and 10 sec off, for a total of 120 sec). This sonicate was designated "crude extract," (protein concentration, 15.8 mg/ml). After centrifugation (10,000 rpm, 10 min, 4° C), the supernatant was designated S. The pellet was washed with 40 ml of buffer, resuspended in 6 ml of buffer and sonicated for 30 sec to suspend the pellet uniformly. This suspension was dialyzed against buffer for 3 h with two changes of buffer and designated "P," (protein concentration, 16.7 mg/ml). Supernatant S was passed through

a Sephadex G-25 column (100 ml bed volume, equilibrated with the same buffer) in two 13 ml portions. Fractions of 12 ml each were collected. Fractions 2 and 3 from both runs were combined and concentrated to 26 ml by ultrafiltration, and designated fraction S(1) (protein concentration, 8.3 mg/ml). This process was repeated to yield S(2) (protein concentration, 6.0 mg/ml) .

3.3. ASSAY FOR ALA SYNTHESIS

Succinate and glycine as substrates for ALA synthesis When succinate and glycine were used as substrates, a slightly modification of Burnham's method (1970) was used. 0.2 ml of crude extract was mixed with 0.15 ml of cofactor mixture (56 mM ATP, 2.46 mM CoA, 1.9 mM pyridoxal phosphate, and 33.3 mM levulinic acid), 0.35 ml of substrate mixture (286 mM glycine, 286 mM succinate, 28.6 mM $MgCl_2$, 0.071 mM succinyl CoA, and 143 mM Tris-Cl, pH 7.5), and 0.3 ml of water. The final assay contained: glycine, 100 mM; succinate, 100 mM; Tris-Cl, pH 7.5, 50 mM; CoA, 0.37 mM; pyridoxal phosphate, 0.27 mM; succinyl CoA, 25 µM, levulinic acid, 5 mM. The mixture was incubated for 10 min at 37° C. The blank was crude extract and buffer. ALA was determined as described in "Determination of ALA" (Section 3.4).

Glutamic acid as substrate for ALA synthesis To assess glutamic acid as substrate, the modified method of Oh-hama and co-workers (1988) was followed. 0.4 ml of cell extract, 0.4 ml of buffer (0.1 M Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT), and 0.2 ml of assay mix (15 mM glutamic acid, 25 mM ATP, 5 mM NADPH, 0.1 mM pyridoxal phosphate, 25 mM levulinic acid) were mixed and incubated for 90 min at 37° C. 50 µl of 70%

perchloric acid was added to stop the reaction. ALA was determined as described in "Determination of ALA" . The blank was crude extract and buffer.

Effects of cofactors and inhibitors To study the effects of cofactors (ATP, NADPH, PLP and tRNA), RNase, levulinic acid and gabaculin, the sonicate was processed as described above and assayed with selective exclusion of substrate, cofactors, levulinic acid or gabaculin from the incubation mixture. The reaction mixture contained 50 µl "P", 250 µl "S" run through Sephadex G-25 once or twice, 100 µl assay mix and buffer and additions (i. e. tRNA) to make up to 500 µl containing, for the complete system, 100 mM Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 5 mM ATP, 1 mM NADPH, 3 mM glutamic acid, 5 mM levulinic acid and 20 mM pyridoxal phosphate. The blank was "S" and "P" and buffer. After 2 h at 37° C, 50 µl 1M citric acid and 0.5 ml 10% SDS were added. The mixture was boiled for 2 min to ensure complete denaturation of proteins, cooled and ALA was determined as below.

3.4. DETERMINATION OF ALA

For crude extracts Burnham's method was used (1970) with modifications. After the reaction was finished, 50 µl of 70% perchloric acid was added to stop the reaction. The mixture was spun down and 1 ml of the supernatant was mixed with 2 ml of 1 M NaOAc, pH 4.7 and 50 µl acetylacetone. After boiling for 15 min and cooling, 3 ml modified Ehrlich's reagent was added to each sample, and the spectrum from 450 -630 nm was recorded after 10 min.

The method of Weinstein and Beale (1985b) was used to determine ALA for assays of purified extracts. The denatured mixture was applied to a Dowex 50X8-400 column (0.9 dia x 0.5 cm high). The column had been prewashed with 2 ml 1N NaOH and 2 ml 0.05 N Na citrate (pH 3.01). The incubation tube was rinsed with 2 ml water which was also applied to the column. The column was washed with 2 ml of 0.05 M Na citrate (pH 3.01) in 25% methanol (v/v), then 1 ml of water. ALA was eluted with 2 ml of 0.5 M sodium phosphate, pH 6.8, and collected in 1 ml fractions. Ethyl acetoacetate (40 µl) was added to each fraction and the mixtures were placed in a boiling water bath for 15 min. An equal amount of modified Ehrlich's reagent was added after cooling and the spectrum recorded from 450-650 nm. The ALA formed was quantitated with a standard curve obtained for ALA in the same way, in which the absorbances at 553 nm were plotted against ALA concentrations. The results were corrected for endogenous ALA by using as the control blank, enzyme extract to which no assay mix had been added.

3.5. EXPERIMENTS WITH LABELED PRECURSORS

Uptake of radiolabeled precursors 0.5 ml of an overnight culture of JL1268 was inoculated into 50 ml of M9 medium supplemented with met, glu, and amp (50µg/ml each). Another 0.5 ml of the overnight culture was inoculated into 50 ml of M9 medium supplemented with met, succinate, gly, and amp (50µg/ml each). The cultures were grown for 24 h at 37° C, centrifuged (6000 rpm, 6 min) and resuspended in 10 ml of the same fresh medium containing 4 mM levulinic acid. After 1 h growth at 37° C, 10 µCi of radiolabeled [^{14}C]glutamic acid or [^{14}C] glycine, respectively, was added to the cultures. At 2, 5, 10, 15, and 30 min, 0.1 ml

of each culture was removed and spotted on a piece of glass fiber filter. The filters were washed six times with 3 ml of 10% TCA each time, then dried and the radioactivity of each filter was counted in 5 ml of ScintiVerse (Fisher Scientific Co).

Radiolabeling, extraction and condensation of ALA 50 ml of an overnight culture of JL1268 in M9 medium containing met and amp (50 μg/ml each) supplemented with glu or succinate and gly (50 μg/ml), depending on the radiolabeled precursor to be added subsequently, was grown for 24 h ($5x10^8$ cells/ml). After centrifugation the cell paste was resuspended in 10 ml fresh medium containing levulinic acid at 4 mM. Radiolabeled precursor (10 μCi) was added after 1 h, at which time there was little appreciable accumulation of ALA in the medium (see Figure 19), and the mixture was incubated for one more hour. The cells were removed by centrifugation. ALA was extracted from the supernatant using the procedure of Troxler and Offner (1979). The supernatant was adjusted to pH 2.0 with HCl. 5 ml of the supernatant was applied to 5 ml of Dowex 50W-X8-400 in a column equilibrated against 0.2 N sodium citrate, pH 3.1. After washing with 60 ml of this buffer, ALA was eluted with 0.2 N sodium citrate, pH 5.1, in 5 ml fractions and an aliquot of each was assayed for ALA. Fractions which contained ALA were combined and adjusted to pH 2.5 with HCl, extracted with 4 ml ether three times and adjusted to pH 6.8 with 0.5 N Na_3PO_4. Of a total of 24 ml, 12 ml was condensed with 0.6 ml ethyl acetoacetate at 100° C for 10 min. The pH of the cooled mixture was adjusted to 2.5 with HCl and the pyrrole was extracted into ether by washing with 3 ml portions five times. The ether was back-washed with 2 ml water and then taken to dryness under N_2.

The residue was dissolved in 100 µl methanol.

Paper chromatography 12 µl of ALA pyrrole prepared as above was chromatographed on Whatman 3MM paper for 4 h using the upper layer of 1:1 butanol/1.5M ammonia as solvent. Amino acids (standards and any carry-over from experiments) were visualized with ninhydrin (0.2 g/100 ml acetone) and ALA pyrrole was visualized with Ehrlich's reagent. The paper was sliced into 15 sections and each was counted in 5 ml of ScintiVerse (Fisher).

Degradation of ALA Radiolabeled ALA which had been isolated by ion exchange chromatography was cleaved by periodate (Beale, 1976; Beale et al.1975; Troxler and Offner, 1979) after adding carrier ALA as described (Troxler and Offner, 1979), to determine the pattern of radiolabeling. 50 mg of carrier ALA was mixed with 6 ml of labeled ALA (10 nanomoles) off the Dowex 50W-X-8 column. 75 mg H_5IO_6 was added and the pH was adjusted to 8.5 with 1 N NaOH. The mixture was incubated for 1 h at room temperature. 90 mg of dimedon was added , and the pH was adjusted to 5.6 with 1 N HCl. The solution was incubated for 1 h at room temperature, cooled to 4° C, and methylene-bis-dimedon was collected on a pre-weighed Whatman filter by vacuum filtration. The precipitate on the filter was washed three times with 10 ml of distilled water, dried over P_2O_5 in a vacuum desiccator, weighed, and assayed for radioactivity in 5 ml of ScintiVerse.

The filtrate containing the succinic acid fragment of ALA (C_1-C_4) was adjusted to pH 6.0 with 1 N NaOH, extracted three times with diethyl ether to remove excess dimedon, and the solution adjusted to pH 3.0 with

1 N HCl. The succinic acid was extracted into diethyl ether (5X20 ml), the diethyl ether was dried over anhydrous sodium sulfate, transferred to pre-weighed vials, evaporated to dryness, weighed, and assayed for radioactivity in 5 ml of ScintiVerse.

RESULTS

1. CLONING AND SEQUENCING OF THE *hemB* GENE OF *E.COLI*

1.1. CHARACTERIZATION OF THE *hem* MUTANT

Strain RP523, a respiratory-deficient, catalase-negative, heme-permeable strain was constructed in this laboratory and provided for this work. Further characterization demonstrated that ALA supplementatiom did not support its growth. Neither PBG nor porphyrins could be detected in the cell extract or the growth medium, and the only intermediate in the pathway detectable in medium was ALA. Also strain RP523 had no ALA D or PBG D activities (Table 6).

In addition, the *hem* mutation was mapped using an F' kit. The *E. coli* map positions of the *hem* genes are shown in Figure 4. The mutation was complemented to yield heme ptototrophy with F254 (CGSC # 6350), but not with F128 (CGSC # 4288) or F152 (CGSC # 4287). F254 covers 8-15 minutes of the chromosome of E. coli, F128 covers 6-8 minutes, overlapping with F254 at minute 8. F152 contains 13-17 minutes. The portions of the *E. coli* map carried by these plasmids are also shown in Figure 4. Thus, the mutated gene is located in the region of 8-13 minutes. Both *hemB,* which encodes ALA D and *hemH* which encodes ferrochelatase, lie within the 8 to 11 map minutes (Bachmann, 1983). To further identify which gene was mutated, enzyme assays of ALA D, PBG D, and ferrochelatase of RP523 were carried out. As seen in Table 6, the mutant synthesized neither ALA D nor PBG D, but did synthesize a substantial amount of ferrochelatase.

Table 6. Activities of heme biosynthetic enzymes in wild-type strains and strains constructed in this study[a]

Strain[b]	Marker	ALA D	PBG D	Ferrochelatase
RP523 (H)	*hemB*	0.00	0.00	0.319
RP523 (H+A)	*hemB*	0.00	0.00	0.174
HU1016	HU1000/phHU16	0.41	0.21	0.630
C600	wild type	0.41	0.37	0.927
JL1002	RP523/pJL2	10.05	0.33	0.670
JL1002 (H)	RP523/pJL2	10.10	0.33	0.670
JL1102	C600/pJL2	8.30	0.32	0.606

[a], Activities are: ALA D, A_{553} per milligram of protein; PBG D, A_{405} per milligram of protein; ferrochelatase, heme (nanomoles) per milligram of protein.

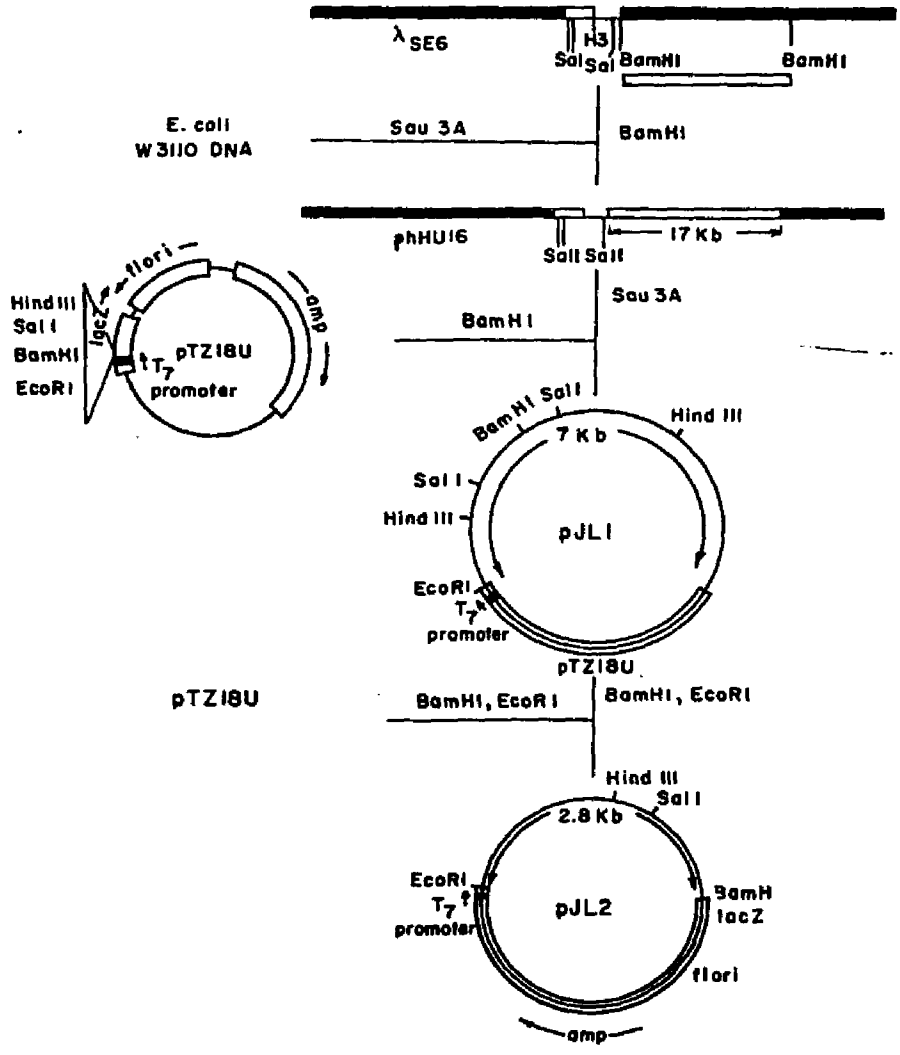
[b], H and A stand for hemin and ALA, respectively. Cells were grown in LB medium either unsupplemented or supplemented with hemin or ALA.

1.2. CLONING OF THE MUTATED GENE BY COMPLEMENTATION

In order to determine whether the isolated mutant carried a mutant allele for *hemB* or *hemC,* or for some heme pathway regulatory gene, the gene that complemented the mutation was cloned. The cloning strategy is presented in Figure 6. An *E. coli* genomic library in λSE6 (Elledge and Walker, 1985) was introduced by phage infection into HUI000, the *hem* mutant carrying a cloned temperature- sensitive allele of the gene encoding the λ phage repressor protein. Under these conditions, this phage behaves as a phasmid, a single copy plasmid. Hem+ colonies were selected on minimal medium plus nutritional requirements lacking hemin. Thirty such colonies were purified and phasmid DNA was isolated from 18 of them. The phasmid DNAs were amplified by transfection into JC7623 which allows the DNAs to be incorporated into phage particles, and subsequent growth of phage lysates. The phage lysates from many of the Hem+ transductants had low titers (<10^6 pfu/ml), but three of the transductants produced phages that gave relatively high titers upon growth (10^8 to 10^{10} pfu/ml). One of these transductants, HUI0I6, was chosen for recloning of its insert.

The phasmid (phHU16), isolated from HUI0I6 contained a 17 kb insert of *E. coli* genomic DNA. Since the *lac* operon maps close (0.25 min away) to *hemB* (Bachmann, 1983; Bachmann and Low, 1980; Hadley et al., 1983; Powell et al., 1973) HUI0I6 was tested for growth on medium containing lactose as a sole carbon source to determine whether the phasmid also carried the *lac* locus by complementing the *lacY* mutant allele of HUI0I6. The strain could not grow on medium containing lactose as a sole carbon source. A second phasmid carrying *hemB* was also tested for the *lac* locus. It too, did not complement the *lacY* mutation of

Figure 6. Strategy for cloning *hemB.* An *E. coli* genomic library in λSE6 was introduced by phage infection into HU1000. Hem^+ colonies were selected on minimal medium lacking hemin, and one of these, HU1016, was chosen for purification of its phasmid DNA. Phasmid DNA (phHU16) was partially digested with Sau3A at 37°C for 2 min. The vector plasmid, pTZ18U, was digested with BamHI at 37°C for 2 h. These DNA's were ligated with T_4 ligase at 15°C for 2 h in ligation buffer (50 mM Tris-HCl, pH 7.6; 10 mM DDT; 1 mM ATP; 50 mg/ml BSA; 10 mM $MgCl_2$). JM101 was transformed with the ligation mix and transformants were screened on X-gal supplemented medium according to the procedure of the supplier, U.S. Biochemical. Plasmid DNA was isolated from each transformant using the mini-prep procedure of U.S. Biochemical, except that the extraction with phenol and chloroform was not performed. HU227 was transformed with the DNA from each white colony to heme prototrophy. pJL1 had the ability to complement the mutation. The insert was subcloned to generate pJL2.


λ SE6
H3
Sal
Sal
BamHI
BamHI
E. coli
W3110 DNA
Sau 3A
BamHI
phHU16
17 Kb
Sall
Sall
Sau 3A
BamHI
Hind III
Sal I
BamHI
EcoRI
lacZ
flori
amp
T7
pTZ18U
promoter
BamHI
Sal I
Hind III
7 Kb
Sal I
Hind III
pJL1
EcoRI
T7
promoter
pTZ18U
pTZ18U
BamHI, EcoRI
BamHI, EcoRI
Hind III
Sal I
2.8 Kb
EcoRI
T7
promoter
pJL2
BamH
lacZ
flori
amp

HUI0I6. Thus, the inserts do not carry the *lac* operon genes, indicating that *lac* genes may lie further away than the 0.25 min as previously reported.

The Hem$^+$ transductant strain was assessed for its ability to synthesize heme pathway enzymes. As seen in Table 6, the transductant, which carries the single copy phasmid DNA, produced wild type levels of ALA D, PBG D and ferrochelatase. Thus the insert carried by the phasmid restores the activities of both enzymes lost by the mutant.

The phasmid DNA was partially digested with Sau3A and the DNA was ligated with plasmid pTZ18U at its BamHI site which is located in a *lacZ* gene . Resulting plasmids were transformed into JMI0I and white colonies were selected on medium with X-gal. DNA was prepared from 172 purified white colonies and was used to transform RP523 to heme prototrophy.

One such plasmid, carrying the gene for heme prototrophy, was found to have a 7 kb insert (pJL1). This plasmid was partially restriction mapped (Figure 6). It was double-digested with BamHI and EcoRI and the fragment was recloned into pTZ18U (pJL2). A 2.85 Kb fragment contained the gene. Digestion of this fragment with Sal1 or HindIII abolished gene activity. The fragment was not cut by the following enzymes: EcoRI, Bgl II, BamHI, SacI, XbaI, KpnI and MboI. A partial restriction map of the 2.85 kb fragment is shown in Figure 6. Strains carrying the plasmid produce about 20 to 30 times wild type levels of ALA D and normal levels of PBG D and ferrochelatase. The results also showed that growth on hemin had no effect on the activity of ALA D, PBG D and ferrochelatase when JL1002 was grown on hemin (Table 6).

1.3. IDENTIFICATION OF THE CLONED GENE PRODUCT

A maxicell protocol (Sancar et al., 1979) was used to determine which genes were cloned on pJL2 . The plasmid was transformed into CSR603, a strain which has a number of mutations in UV repair genes. After UV irradiation and subsequent incubation of the cells, UV damage cannot be repaired. Chromosomal DNA is more sensitive to UV irradiation because of its large size. Cells extracts were prepared and assayed for enzymatic activity and analyzed on gels. The extracts had no PBG D activity, but did have l00 times the normal ALA D activity of CSR603/pTZ18U (data not shown). pJL2 directed the production of two proteins (Figure 7) which have subunit sizes of 28 kDa and 38 KDa (Figure 8), as determined when resolved on SDS-PAGE gels. The control plasmid, pTZ18U only directed the synthesis of a 28 KDa protein, which is the ß-lactamase synthesized by the *amp* gene on the plasmid vector. Therefore, the insert codes for only one protein. Nondenaturing gels revealed three protein bands in addition to ß-lactamase (Figure 9). These three bands had ALA D activity and probably correspond to different aggregations of subunits. PBG D activity and heme-binding activity were not associated with the protein on the gel. Thus, the *hemB* gene, and not some regulatory gene was cloned, and the sole protein generated from the insert as resolved on SDS-PAGE gel is the ALA D enzyme.

1.4. SEQUENCE ANALYSIS OF THE *hemB* GENE

The strategy for sequencing *hemB* is diagrammed in Figure 10. The complete nucleotide and amino acid sequences are shown in Figure 11. A search of two DNA databases showed strong homologies to human

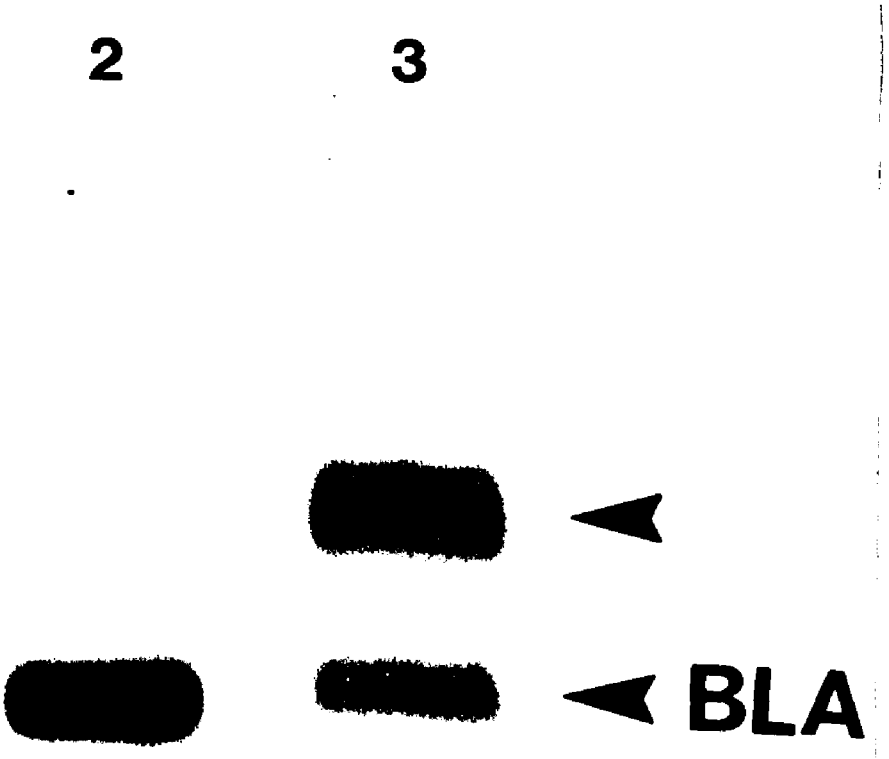


Figure 7. Autoradiogram of SDS-PAGE to show proteins produced by pJL2 in the presence of [^{35}S]-methionine in the maxicell procedure. The proteins were analyzed on sodium dodecyl sulfate-10% acrylamide disc gels. Lanes: 1, CSR603; 2, CSR603/pTZ18U; 3, CSR603/pJL2. In lane 3, bottom band is β-lactamase (BLA), top band is the product of *hemB.* gene.

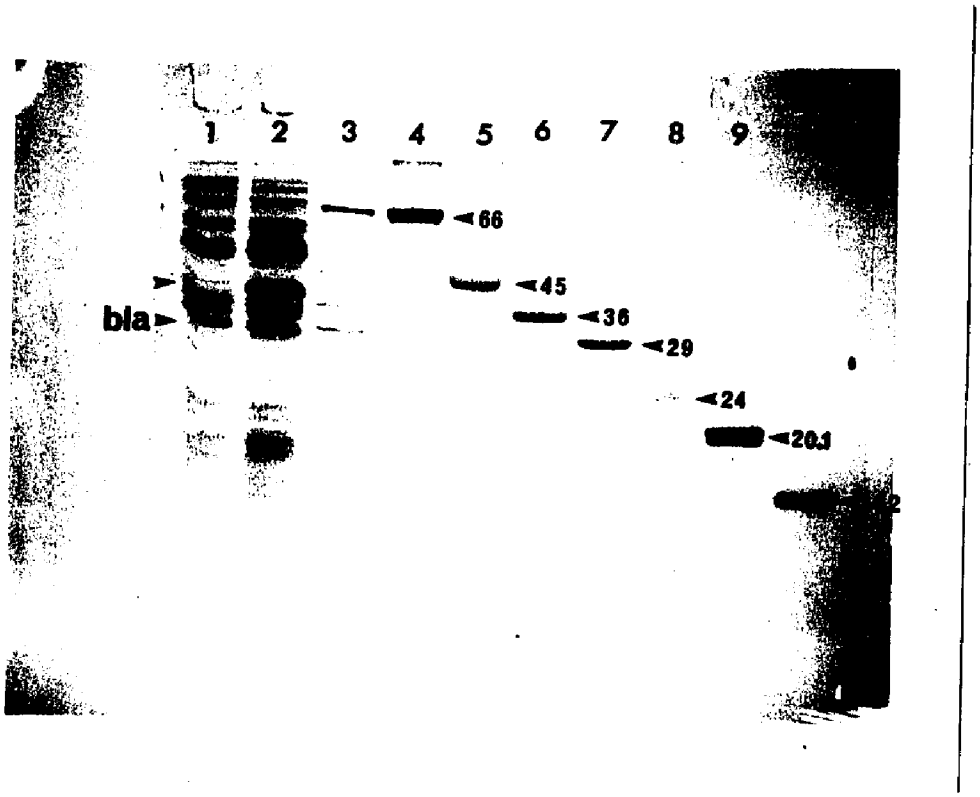


Figure 8. Coomassie blue staining of SDS-PAGE gel to determine the molecular weight of *E. coli* ALA D. Lanes: 1, CSR603/pTZ18U; 2, CSR603/pJL2; 3, protein markers (Dalton Mark VII-L, SDS-7, Sigma), a mixture of 7 proteins in lane 4-10; 4, bovine albumin; 5, egg albumin; 6, glyceraldehyde-3-phosphate dehydrogenase; 7, carbonic anhydrase; 8, trypsinogen; 9, trypsin inhibitor; 10. α-lactalbumin.

Figure 9. Autoradiogram of proteins produced by pJL2 in the maxicell procedure in the presence of [^{35}S]-methionine. The proteins were analyzed on a nondenaturing 10% polyacrylamide disc gel. Lanes: 1, CSR603/pTZ18U; 2, CSR603/pJL2. The enzyme assay showed that the bands indicated by arrows had ALA D activity.

liver and rat liver cDNA for ALA D (Wetmur et al., 1986; Bishop et al., 1986; Bishop et al., 1987), the only ALA D genes in the databases (May, 1988, Bionet), and poor homologies to all other genes in the databases.

The protein coding region was determined by two methods. A search for open reading frames (ORF) was made using DNA Inspector II+ software. Only one ORF was found which is from nt 483 to 1491(Appendix A). Additionally, the protein coding region was located with *E. coli* codon bias tables generated using the IBI codon bias table generator program (Appendix B). Both methods resulted in compatible protein coding regions.

The codon usage for the *hemB* gene is very similar to codon usage in 407 other *E. coli* genes (Aota et al., 1988). However, comparisons with codon usage for the rat liver (Bishop et al., 1987), human liver (Wetmur et al., 1986) and yeast (Myers et al., 1987) ALA D genes showed that the human and rat codon usages are virtually identical, whereas yeast and *E. coli* usages each differ from them, but by only a few codons (Table 7). The underlined codons in Table 7 are used much less frequently in human and rat ALA D genes. Where the table shows 0% usage in the *E. coli hemB* gene, appreciable usage is observed in the human and rat genes.

The P2' index is a function of codon usage which can be used to assess the level of expression of a gene (Sharp and Li, 1986). Among 166 *E. coli* genes assessed, the mean P2' value for the 27 very highly-expressed genes is 0.76 (range 0.62-0.92), and for the 15 highly-expressed genes 0.65 (range 0.52-0.82). P2' values for the other 115 genes range between 0.30 and 0.76. The lower half 58 genes with P2' <0.49 are considered to show low gene expression. For the *E. coli hemB*, the P2' index is 0.51, indicating that it is a moderately-expressed gene.

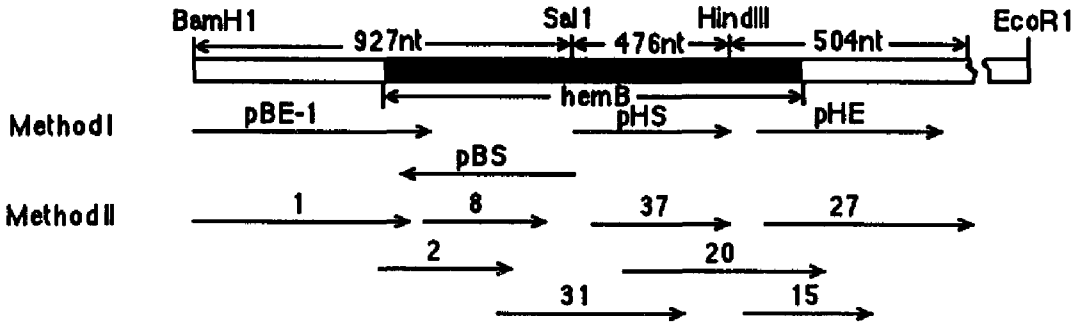


Figure 10. Sequencing strategy for *hemB* gene. In Method I, a series of double digestions the BamHI-EcoRI fragment of pJL2 DNA was performed. Fragments from BamHI + SalI (pBS), SalI+ HindIII (pSH), and HindIII + EcoRI (pHE) double digestions were subcloned into plasmids pTZ18U and pTZ19U to allow for sequencing in two directions. The ssDNA was prepared from each subclone utilizing the vectors' F1 origin of replication. Of the six possible sequences (three fragments in two directions), four were determined, as illustrated above. A second method was used to fill in gaps and to confirm the results of the above (Method II). pJL2 was double-digested with BamHI+ EcoRI and subcloned into M13mp18, and phage was prepared. A series of nested deletions of the insert in the phage vector was constructed with T_4 DNA polymerase using the Cyclone system. Nine overlapping fragments which spanned the 1.9 kb portion of the insert containing the structural gene were chosen and sequenced with the Sequenase kit. The numbers in Method II refer to the designations of the phages carrying the different inserts that were chosen for sequencing.

```
   1   TG CAT GCC TGC AGG TCG ACT CTA GAG GAT CCA TCA ATT ATC AAT GCA TAG AAA ACA TAT
  60   TAT GTT TCG CGC TAA CGA TTA TGC TGC GTT ACA GAC GAA TGG TTT CAA ATT ATT CTT TCC
 120   CTA AGT CCG AAT GTG GAC ATA TAT AAT TTA TGG AGA AAA ACA CGC ATC AAA TTC ATA AAA
 180   GCA CAG TAT TTC AAC ATG TTG CAG ATA TTT TGT TTT TGA AGC ATA CGC ATA ATA AAT ATC
 240   TCA AAC ACA AAC CAG ATA ATA TCC TAT GAA ATG CAA CAA AGT TTT AAT TAT TCA ATT TAT
 300   AGC AGC GCA ACC TTT CTC CAA AAA ATA ATA ACC AAC AGC TAT AAA TAA AGA TTA AAA ATA
 360   TTT TCA GTG AAT ATA AAA TTA TTG CTG TTC TAT TAA AAA AGA TGT CTG GCC TCT CTT CCA
 420   TTC TCT TCT TGT CAA CCA TCG CGA CAC TTT CGT AAA ACA TCC CTA CCC TGC TTC AGG TAT
 480   ACT ATG CCC CTC GAT TCC ACA AAC ATC AGG CAG ACC ATG ACA GAC TTA ATC CAA CGC CCT
           M   P   L   D   S   T   N   I   R   Q   T   M   T   D   L   I   Q   R   P
 540   CGT CGC CTG CGC AAA TCT CCT GCG CTG CCG CGT ATG TTT GAA GAG ACA ACA CTT AGC CTT
       R   R   L   R   K   S   P   A   L   P   R   M   F   E   E   T   T   L   S   L
 600   AAC GAC CTG GTG TTG CCG ATC TTT GTT GAA GAA GAA ATT GAC GAC TAC AAA GCC GTT GAA
       N   D   L   V   L   P   I   F   V   E   E   E   I   D   D   Y   K   A   V   E
 660   GCC ATG CCA GGC GTG ATG CGC ATT CCA GAG AAA CAT CTG GCA CGC GAA ATT GAA CGC ATC
       A   M   P   G   V   M   R   I   P   E   K   H   L   A   R   E   I   E   R   I
 720   GCC AAC GCC GGT ATT CGT TCC GTG ATG ACT TTT GGC ATC TCT CAC CAT ACC GAT GAA ACC
       A   N   A   G   I   R   S   V   M   T   F   G   I   S   H   H   T   D   E   T
 780   GGC GAG CGA GCC TGG CGG GAA GAT GGA CTG GTG GCG CGT ATG TCG CCG ATC TGC AAG CAG
       G   S   D   A   W   R   E   D   G   L   V   A   R   M   S   P   I   C   K   Q
 840   ACC GTG CCA GAA ATG ATC GTT ATG TCA GAC ACC TGC TTC TGT GAA TAC ACT TCT CAC GGT
       T   V   P   E   M   I   V   M   S   D   T   C   F   C   E   Y   T   S   H   G
 900   CAC TGC GGT GTG CTG TGC GAG CAT GGC GTC GAC AAC GAC GCG ACT CTG GAA AAT TTA GGC
       H   C   G   V   L   C   E   H   G   V   D   N   D   A   T   L   E   N   L   G
 960   AAG CAA GCC GTG GTT GCA GCT GCT GCA GGT GCA GAC TTC ATC GCC CCT TCC GCC GCG ATG
       K   Q   A   V   V   A   A   A   A   G   A   D   F   I   A   P   S   A   A   M
1020   GAC GGC CAG GTA CAG GCG ATT CGT CAG GCG CTG GAC GCT GCG GGA TTT AAA GAT ACG GCG
       D   G   Q   V   Q   A   I   R   Q   A   L   D   A   A   G   F   K   D   T   A
1080   ATT ATG TCG TAT TCG ACC AAG TTC GCC TCC TCC TTT TAT GGC CCG TTC CGT GAA GCT GCC
       I   M   S   Y   S   T   K   F   A   S   S   F   Y   G   P   F   R   E   A   A
1140   GGA AGC GCA TTA AAA GGC GAC CGC AAA AGC TAT CAG ATG AAC CCA ATG AAC CGT GCT GAG
       G   S   A   L   K   G   D   R   K   S   Y   Q   M   N   P   M   N   R   A   E
1200   GGC ATT GCT GAA TAC CTG CTG GAT GAA GCC CAG GGG CCA GAC TGC CTG ATG GTT AAA CCT
       G   I   A   E   Y   L   L   D   E   A   Q   G   P   D   C   L   M   V   K   P
1260   GCT GGA GCG TAC CTC AAC ATC GTG CGT GAG CTG CGT GAA CGT ACT GAA TTG CCG ATT GGC
       A   G   A   Y   L   N   I   V   R   E   L   R   E   R   T   E   L   P   I   G
1320   GCG TAT CAG GTG AGC GGT GAG TAT GCG ATG ATT AAG TTC GCC GCG CTG GCG GGT GCT ATA
       A   Y   Q   V   S   G   E   Y   A   M   I   K   F   A   A   L   A   G   A   I
1380   GAT GAA GAG AAA GTC GTG CTC GAA AGC TTA GGT TCG ATT AAG CGT GCG GGT GCG GAT CTG
       D   E   E   K   V   V   L   E   S   L   G   S   I   K   R   A   G   A   D   L
1440   ATT TTC AGC TAC TTT GCG CTG GAT TTG GCT GAG AAG AAG ATT CTG CGT TAA TTT TGT TTC
       I   F   S   Y   F   A   L   D   L   A   E   K   K   I   L   R   -
1500   CTC TCA GAA AGG GGG AGT TAA CCT GTT GAA AGT ACG GTG AAC GGG CCC TCA CCC TAA CCC
1560   TCT CCC CAA GGG GCG AGG GGA CCG TCC ACT CTC GTA TTA CCC CGC CCG ATA AAA CGG TTT
1620   ATC CCC AAG GAT CGC CCG ATG CAT TAT CCG TCG CTG TGG CAG GTA ATC GGC ATT GGC ATA
1680   GTG CTG GGT CAC GCC CTT ATC CAA ATC GCA ATA TAT TTG GTT GCC AGC GCC AGC CAC CTG
1740   AAA CTC CGG TTT GGT GAT ATG GGC AAA CAA AAA ACT TAA CAA GGC TTC GCT CTC TTT TCT
1800   CGC TCA CAT CAA CAA TTC GCG TAG GTA AAG CCT TCA TTC ACA AAC AGC GCC TGT TTA CCG
1860   CTC ACC GGA TGC GTT CGC ACC ACC GGA TGT AGC AAC GGT GGT TTT GGA
```

Figure 11.Nucleotide sequence of the sense strand of the *hemB* gene. Possible promoter regions are shown by single underlines, start codons are designated by double underlines and Shine-Dalgarno sequences are shown by single overlines. The 7-bp direct repeat associated with one promoter sequence is indicated by a double overline.

Table 7. Comparison of codon usage of ALA Ds from *E. coli*, yeast, rat and human genes[a,b]

Codon[b]	aa	E.coli	Yeast	Rat	Human	Codon[b]	aa	E.coli	Yeast	Rat	Human
TTT	F	6(1.8)	6(1.8)	7(2.1)	6(1.8)	TCT	S	3(0.9)	6(1.8)	2(0.6)	1(0.3)
TTC	F	6(1.8)	8(2.3)	6(1.8)	6(1.8)	TCC	S	5(1.5)	3(0.9)	7(2.1)	7(2.1)
TTA	L	4(1.2)	10(2.9)	1(0.3)	0(0.0)	TCA	S	1(0.3)	2(0.6)	3(0.9)	1(0.3)
TTG	L	3(0.9)	10(2.9)	3(0.9)	4(1.2)	TCG	S	4(1.2)	2(0.6)	0(0.0)	2(0.6)
CTT	L	2(0.6)	0(0.0)	3(0.9)	2(0.6)	CCT	P	4(1.2)	10(2.9)	6(1.8)	6(1.8)
CTC	L	3(0.9)	2(0.6)	9(2.7)	10(3.0)	CCC	P	1(0.3)	1(0.3)	10(3.0)	9(2.7)
CTA	L	0(0.0)	7(2.0)	4(1.2)	1(0.3)	CCA	P	5(1.5)	8(2.3)	4(1.2)	4(1.2)
CTG	L	16(4.8)	5(1.5)	17(5.2)	17(5.2)	CCG	P	5(1.5)	1(0.3)	2(0.6)	3(0.9)
ATT	I	12(3.6)	13(3.8)	4(1.2)	1(0.3)	ACT	T	4(1.2)	6(1.8)	2(0.6)	1(0.3)
ATC	I	9(2.7)	11(3.2)	6(1.8)	10(3.0)	ACC	T	6(1.8)	2(0.6)	8(2.4)	7(2.1)
ATA	I	1(0.3)	2(0.6)	1(0.3)	1(0.3)	ACA	T	4(1.2)	5(1.5)	0(0.0)	2(0.6)
ATG	M	15(4.5)	6(1.8)	9(2.7)	11(3.3)	ACG	T	1(0.3)	0(0.0)	2(0.6)	1(0.3)
GTT	V	5(1.5)	6(1.8)	5(1.5)	3(0.9)	GCT	A	9(2.7)	14(4.1)	10(3.0)	12(3.6)
GTC	V	2(0.6)	4(1.2)	8(2.4)	5(1.5)	GCC	A	12(3.6)	5(1.5)	21(6.4)	17(5.2)
GTA	V	1(0.3)	1(0.3)	5(1.5)	5(1.5)	GCA	A	5(1.5)	11(3.2)	9(2.7)	10(3.0)
GTG	V	10(3.0)	8(2.3)	9(2.7)	12(3.6)	GCG	A	16(4.8)	4(1.2)	0(0.0)	1(0.3)
TAT	Y	5(1.5)	7(2.0)	6(1.8)	3(0.9)	TGT	C	1(0.3)	6(1.8)	6(1.8)	5(1.5)
TAC	Y	5(1.5)	7(2.0)	5(1.5)	9(2.7)	TGC	C	5(1.5)	2(0.6)	2(0.6)	3(0.9)
TAA	-	1(--)	1(--)	0(--)	0(--)	TGA	-	0(--)	0(--)	1(--)	1(--)
TAG	-	0(--)	0(--)	0(--)	0(--)	TGG	W	1(0.3)	2(0.60	3(0.9)	3(0.9)
CAT	H	3(0.9)	5(1.5)	2(0.6)	4(1.2)	CGT	R	12(3.6)	7(2.0)	2(0.6)	0(0.0)
CAC	H	3(0.9)	4(1.2)	7(2.1)	5(1.5)	CGC	R	7(2.1)	0(0.0)	3(0.9)	6(1.8)
CAA	Q	2(0.6)	5(1.5)	2(0.6)	0(0.0)	CGA	R	0(0.0)	1(0.3)	4(1.2)	2(0.6)
CAG	Q	8(2.4)	1(0.3)	11(3.3)	10(3.0)	CGG	R	1(0.3)	0(0.0)	4(1.2)	9(2.7)
AAT	N	1(0.3)	9(2.6)	1(0.3)	1(0.3)	AGT	S	0(0.0)	5(1.5)	4(1.2)	2(0.6)
AAC	N	7(2.1)	4(1.2)	3(0.9)	4(1.2)	AGC	S	7(2.1)	1(0.3)	5(1.5)	6(1.8)
AAA	K	8(2.4)	8(2.3)	2(0.6)	2(0.6)	AGA	R	0(0.0)	7(2.0)	3(0.9)	2(0.6)
AAG	K	7(2.1)	7(2.0)	10(3.0)	10(3.0)	AGG	R	1(0.3)	7(2.0)	4(1.2)	4(1.2)
GAT	D	9(2.7)	17(5.0)	9(2.7)	7(2.1)	GGT	G	8(2.4)	19(5.6)	3(0.9)	4(1.2)
GAC	D	12(3.6)	8(2.3)	8(2.4)	11(3.3)	GGC	G	10(3.0)	3(0.9)	9(2.7)	6(1.8)
GAA	E	19(5.7)	16(4.7)	6(1.8)	6(1.8)	GGA	G	4(1.2)	0(0.0)	10(3.0)	10(3.0)
GAG	E	8(2.4)	2(0.6)	12(3.6)	14(4.2)	GGG	G	1(0.3)	3(0.9)	1(0.3)	3(0.9)

[a] Derived in this laboratory from nucleotide sequences using the IBI codon bias table generator program.

[b] Underlined codons are used frequently in *E. coli* and infrequently or not at all in rat and human genes.

Another criterion for determining the level of expression of a gene is to calculate the percentage of modulatory codons (Grosjean and Fiers, 1982) which correspond to the tRNA's of lowest abundance *in vivo*. Low usage of these codons occurs in highly expressed genes. According to its use of modulatory codons (2.4%), *hemB* is expressed at an intermediate level. A third method to analyze the level of expression is to calculate the Q/D ratios (quartet/duet for amino acids with six codons, these are, for arginine, leucine, and serine) which correlate with levels of expression, and are higher in highly-expressed genes (61/1, 63/6, and 48/11, respectively) than in genes with low levels of expression (57/13, 71/24, and 46/22, respectively) (Grantham et al., 1981). According to the Q/D ratio of the hemB gene, it is expressed at moderate level. Thus the results of the P2' index, modulatory codon usage and Q/D ratio classify the *hemB* gene as a moderately-expressed gene. Codon usages, P2' indices, modulatory codon usages and Q/D indices for *hemA, hemB, hemC,* and *hemD* are summarized in Table 8.

The nucleotide sequence of *hemB* shown in Figure 11, contains two possible translation start codons at positions 483 and 517 which would encode proteins of 324 and 335 amino acids, respectively, with molecular weights of 35, 506 and 36,763 Da, respectively. Each start codon is associated with a possible Shine-Dalgarno sequence, CAGG, six bp upstream from it.

Additionally two possible promoter sequences were identified (Mulligan and McClure, 1986) as shown in Figure 11. One is associated with a seven bp direct repeat and is tentatively favored as the initiation of transcription site, although both have the same score on the Mulligan and

Table 8. Codon usage, P2' index, usage of modulatory codons, and Q/D index of cloned *E.coli hem* genes[a]

Codon	aa[b]	*hemA*	*hemB*	*hemC*	*hemD*	Codon	aa[b]	*hemA*	*hemB*	*hemC*	*hemD*
TTT	F	5(1.2)	6(1.8)	2(0.6)	5(2.0)	TCT	S	4(1.0)	3(0.9)	1(0.3)	2(0.8)
TTC	F	2(0.5)	6(1.8)	1(0.3)	1(0.4)	TCC	S	1(0.2)	5(1.5)	3(1.0)	0(0.0)
TTA	L	7(1.7)	4(1.2)	6(1.9)	8(3.3)	TCA	S	2(0.5)	1(0.3)	3(1.0)	1(0.4)
TTG	L	3(0.7)	3(0.9)	2(0.6)	5(2.0)	TCG	S	7(1.7)	4(1.2)	2(0.6)	2(0.8)
CTT	L	10(2.4)	2(0.6)	7(2.2)	3(1.2)	CCT	P	1(0.2)	4(1.2)	1(0.3)	2(0.8)
CTC	L	6(1.4)	3(0.9)	5(1.6)	4(1.6)	CCC	P	0(0.0)	1(0.3)	1(0.3)	2(0.8)
CTA	L	0(0.0)	0(0.0)	2(0.6)	7(2.8)	CCA	P	2(0.5)	5(1.5)	3(1.0)	1(0.4)
CTG	L	26(6.2)	16(4.8)	20(6.4)	11(4.5)	CCG	P	7(1.7)	5(1.5)	11(3.5)	5(2.0)
ATT	I	10(2.4)	12(3.6)	13(4.2)	6(2.4)	ACT	T	5(1.2)	4(1.2)	4(1.3)	1(0.4)
ATC	I	14(3.3)	9(2.7)	6(1.9)	4(1.6)	ACC	T	5(1.2)	6(1.8)	2(0.6)	6(2.4)
ATA	I	0(0.0)	1(0.3)	0(0.0)	1(0.4)	ACA	T	2(0.5)	4(1.2)	1(0.3)	1(0.4)
ATG	M	10(2.4)	5(4.5)	6(1.9)	4(1.6)	ACG	T	6(1.4)	1(0.3)	3(1.0)	1(0.4)
GTT	V	12(2.9)	5(1.5)	4(1.3)	3(1.2)	GCT	A	5(1.2)	9(2.7)	2(0.6)	5(2.0)
GTC	V	5(1.2)	2(0.6)	8(2.6)	7(2.8)	GCC	A	17(4.1)	12(3.6)	14(4.5)	7(2.8)
GTA	V	4(1.0)	1(0.3)	6(1.9)	1(0.4)	GCA	A	8(1.9)	5(1.5)	8(2.6)	6(2.4)
GTG	V	9(2.2)	10(3.0)	6(1.9)	3(1.2)	GCG	A	20(4.8)	16(4.8)	11(3.5)	9(3.7)
TAT	Y	5(1.2)	5(1.5)	3(1.0)	3(1.2)	TGT	C	1(0.2)	1(0.3)	2(0.6)	3(1.2)
TAC	Y	1(0.2)	5(1.5)	2(0.6)	2(0.8)	TGC	C	2(0.5)	5(1.5)	2(0.6)	1(0.4)
TAA	-	0(--)	1(--)	0(--)	1(--)	TGA	-	0(--)	0(--)	1(--)	0(--)
TAG	-	1(--)	0(--)	0(--)	0(--)	TGG	W	4(1.0)	1(0.3)	2(0.6)	7(2.8)
CAT	H	6(1.4)	3(0.9)	1(0.3)	4(1.6)	CGT	R	9(2.2)	12(3.6)	7(2.2)	7(2.8)
CAC	H	3(0.7)	3(0.9)	4(1.3)	3(1.2)	CGC	R	18(4.3)	7(2.1)	19(6.1)	6(2.4)
CAA	Q	9(2.2)	2(0.6)	6(1.9)	15(6.1)	CGA	R	2(0.5)	0(0.0)	0(0.0)	3(1.2)
CAG	Q	16(3.8)	8(2.4)	5(1.6)	6(2.4)	CGG	R	1(0.2)	1(0.3)	0(0.0)	3(1.2)
AAT	N	4(1.0)	1(0.3)	6(1.9)	2(0.8)	AGT	S	2(0.5)	0(0.0)	3(1.0)	3(1.2)
AAC	N	9(2.2)	7(2.1)	5(1.6)	2(0.8)	AGC	S	16(3.8)	7(2.1)	4(1.3)	4(1.6)
AAA	K	15(3.6)	8(2.4)	7(2.2)	3(1.2)	AGA	R	0(0.0)	0(0.0)	1(0.3)	0(0.0)
AAG	K	3(0.7)	7(2.1)	1(0.3)	2(0.8)	AGG	R	0(0.0)	1(0.3)	0(0.0)	0(0.0)
GAT	D	18(4.3)	9(2.7)	14(4.5)	9(3.7)	GGT	G	4(1.0)	8(2.4)	8(2.6)	6(2.4)
GAC	D	7(1.7)	12(3.6)	5(1.6)	2(0.8)	GGC	G	10(2.4)	10(3.0)	11(3.5)	3(1.2)
GAA	E	22(5.3)	19(5.7)	16(5.1)	9(3.7)	GGA	G	0(0.0)	4(1.2)	5(1.6)	3(1.2)
GAG	E	12(2.9)	18(2.4)	7(2.2)	8(3.3)	GGG	G	4(1.0)	1(0.3)	3(1.0)	3(1.2)
P2'[c]		0.39	0.51	0.31	0.46						
Mod[d]		1.67%	2.39%	2.88%	5.35%						
Q/D[e]	R	30/0	20/1	26/1	19/0						
	L	42/10	21/7	34/8	25/13						
	S	14/18	13/7	9/7	5/7						

[a], Codon usages were derived from the deduced amino acid sequences of cloned *hem* genes. The references are: Li et al., 1989c for *hemA* ; Li et al., 1989a for *hemB* ; Thomas and Jordan, 1986 for *hemC* ; Sasarman et al., 1987; Jordan et al., 1988 for *hemD* .

[b], 1-letter abbreviation for amino acids is used.

[c], P2' was calculated as follows:

$(n_{TTC}+n_{TAC}+n_{CGT}+n_{ATC}+n_{AAC}+n_{GCT}+n_{GGT})/(n_{TTY}+n_{TAY}+n_{CGY}+n_{ATY}+n_{AAY}+n_{GCY}+n_{GGY})$. Where n_{TTC} was the number of occurrences of the codon pair n_{TTT} and n_{TTC}, etc (Sharp and Li, 1986). For example, the value for the *hemA* gene is

(2+1+9+14+9+5+4)/(2+5+1+5+9+18+14+10+9+4+5+17+4+10)=0.39.

[d], Mod was obtained by dividing total codons of a gene with the usage of modulatory codons (CTA, ATA, CGA, CGG, AGA, AGG, GGA and GGG) (Grosjean and Fiers, 1982). For example, the *hemA* gene has 8 modulatory codons and 418 total codons (8/418=1.91%).

[e], Q/D value (quartet/duet) is calculated for amino acids (arg, leu, ser) with six codons (Grantham et al., 1981). For example, the *hemA* gene uses 30 arg codons (9+18+2+1) from the quartet (CTG, CGC, CGA, and CGG, respectively) and none (0+0) from the duet (AGA and AGG, respectively).

McClure scale (1986), i. e. 58%, designating them as moderate or Class 2 promoters.

A number of possible terminator sequences have been identified which are GC rich, contain six to seven bp in the stem and have loops of between four to six bases. These, however are not followed by a sequence of T's. Another possible termination sequence starts much further away from the end of the coding region, nt 1879, and consists of a six bp stem with a ten base loop, or a seven bp stem encompassing an AA loop, and a six base loop. The stem includes five or six GC pairs, respectively, and is followed by a sequence of four T's. This sequence occurs within another ORF. None of the sequences identified as possible terminator sequences is optimal, because either they are not followed by a series of T's, or the loop is more than 5 bases (Rosenberg and Court, 1979), or the site is too far from the end of the coding region.

A search of two protein sequence databases showed strong homology to rat liver ALA D, the only ALA D in the databases and poor homologies to all the other proteins in the databases.

The amino acid sequence of ALA D from *E. coli* is compared with the sequences of ALA D from human liver (Wetmur, et al., 1986), rat liver (Bishop et al., 1986) and yeast (Myers et al., 1987) (Figure 12). There is fairly extensive homology between this prokaryotic protein and the eukaryotic proteins (36% for yeast, 40% for human and 40% for rat). Local regions of strong homologies are observed (84.2% and 70%) including the zinc-binding region, residues 129-144, where twelve out of sixteen (75%) amino acids are identical for all the proteins. This region includes the consensus sequence for zinc binding sites (Berg, 1986). The active site lysine (Gibbs and Jordan, 1986 and Wetmur et al., 1986) at position

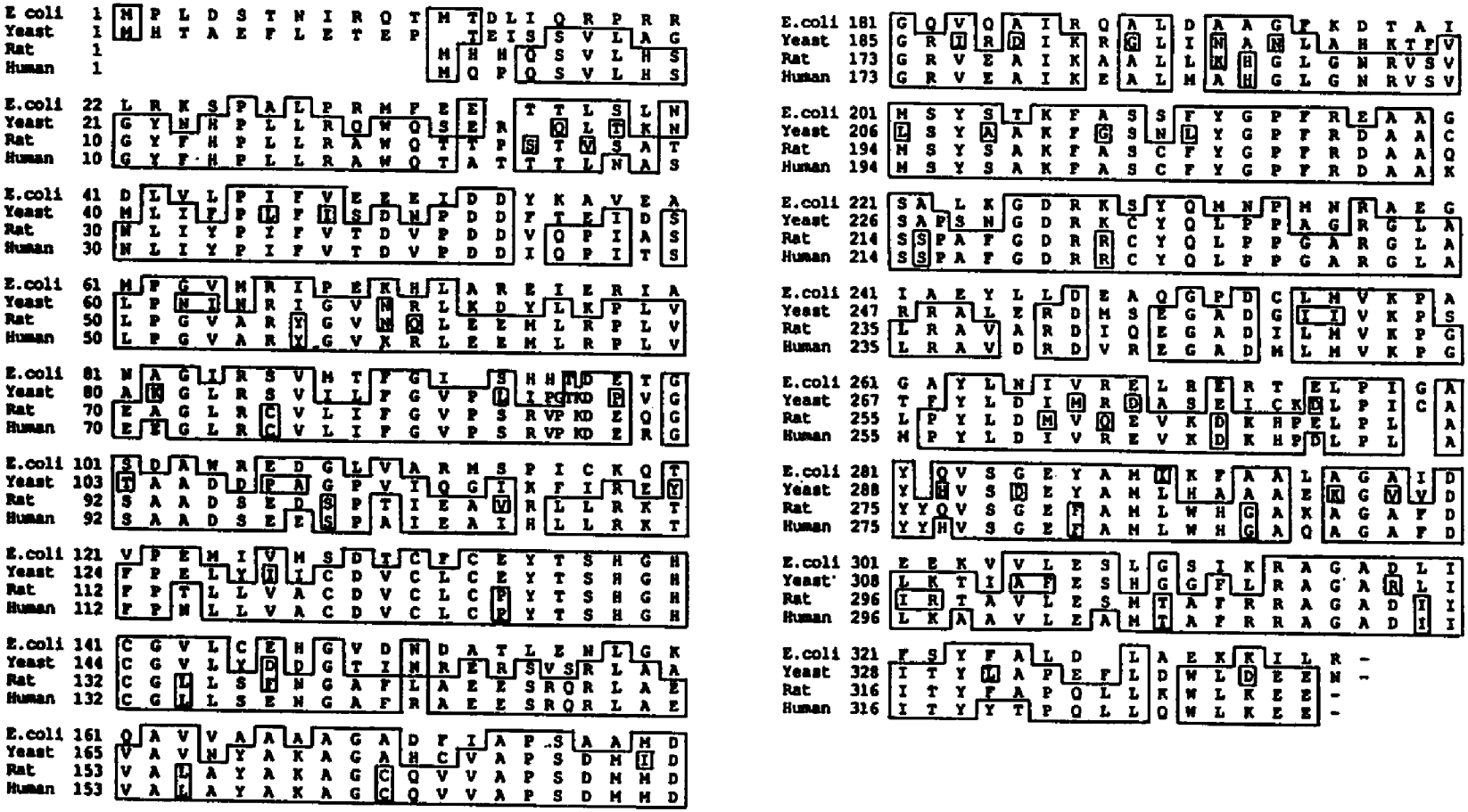


Figure 12. Comparison of amino acid sequences of E. coli, yeast, rat liver and human liver ALA D's deduced from the sequences of cloned genes. Areas of identity are enclosed by boxes.

258 is also in an area of homology, where the four amino acid residues surrounding lysine for the *E. coli*, rat and human proteins are identical. The apparent homology is also shown by an homology matrix search with *E. coli* ALA D against yeast, human liver and rat liver ALA Ds (Appendix C).

The hydropathy profiles for human liver, rat liver, yeast and *E. coli* ALA D's are shown in Figure 13. The distribution of hydrophilic and hydophobic areas confirms that *E. coli* ALA D, like the rat liver, human liver and yeast enzymes is a soluble enzyme. The N terminal regions for the rat and human enzymes are hydrophobic, while the N terminal regions for *E. coli* (for either initiation of translation site) and yeast are hydrophilic. Otherwise, the regions of hydrophilicity and hydrophobicity are almost superimposable. It is especially noteworthy that the Zn-binding region and the region with the essential lysine are both hydrophobic areas in all of the ALA D's.

The secondary structures of ALA D's from *E. coli*, rat liver and human liver also show great similarity. The Zn-binding region and its flanking sequence has a beta(about 10 aa)-turn(6 aa)-beta(5-6 aa)-turn(4-6 aa)-alpha(more than 10 aa) secondary structure. The structure around active lysine is also very similar (Appendix D).

A detailed restriction map (Appendix E), antigen plot (Appendix F), amino acid composition, molecular weight, pI, and cleavage map with CNBr and trypsin and the size of fragments (Appendix G) of *E. coli* ALA D were also obtained using computer software.

The results of the ALA D gene sequence study show that the gene for ALA D is indeed highly conserved, as is the gene for PBG D (Thomas and Jordan, 1986). This is significant because ALA D plays a part in the

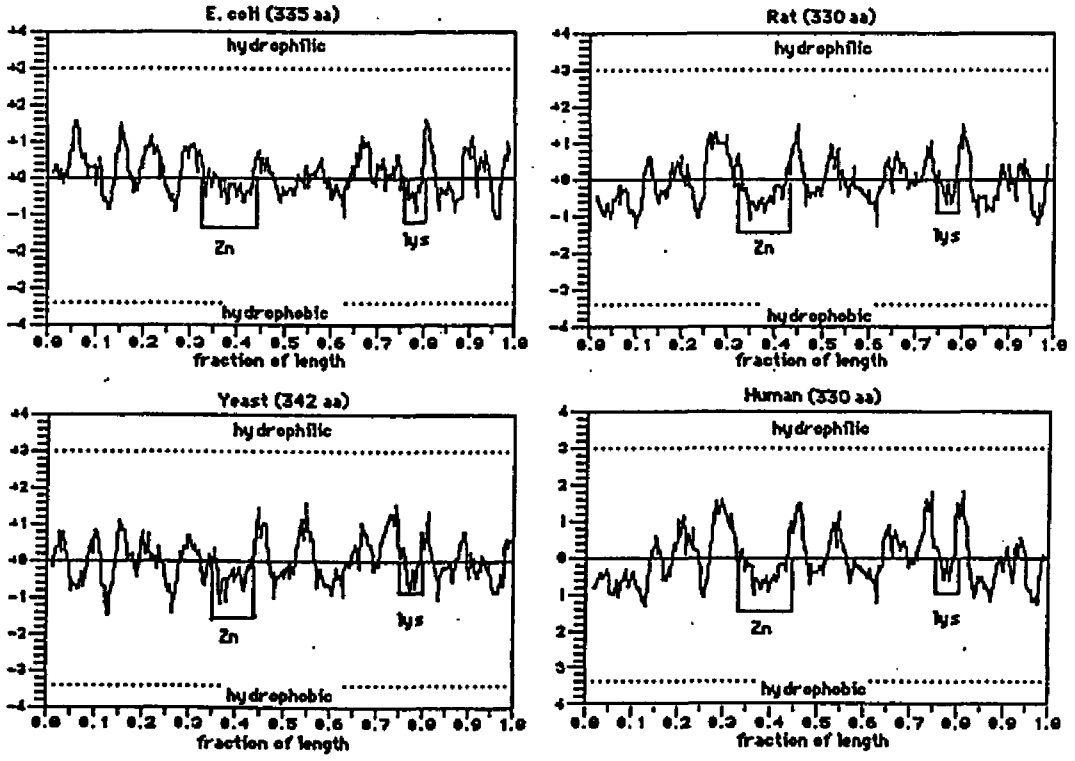


Figure 13. Hydrophathy profiles of E. coli, yeast, rat liver and human liver ALA D's amino acid sequences. These profiles were obtained from a Hopp-Woods hydrophilicity plot using DNA Inspector II+ software. The number of amino acids of each protein is shown in parentheses. The Zn-binding region and the active site lysine region are indicated by brackets.

regulation of an important enzymic step in the heme biosynthesis pathway. PBG availability controls the activity of PBG D in *E. coli* (Umanoff et al., 1988). This suggests that this regulation is ubiquitous because, Beaumont and coworkers (1986) showed that in dimethyl sulfoxide-mediated differentiation of Friend erythroleukemia cells, inhibition of ALA D, and therefore PBG production, lowered the half life of PBG D while synthesis of the enzyme protein was unaffected.

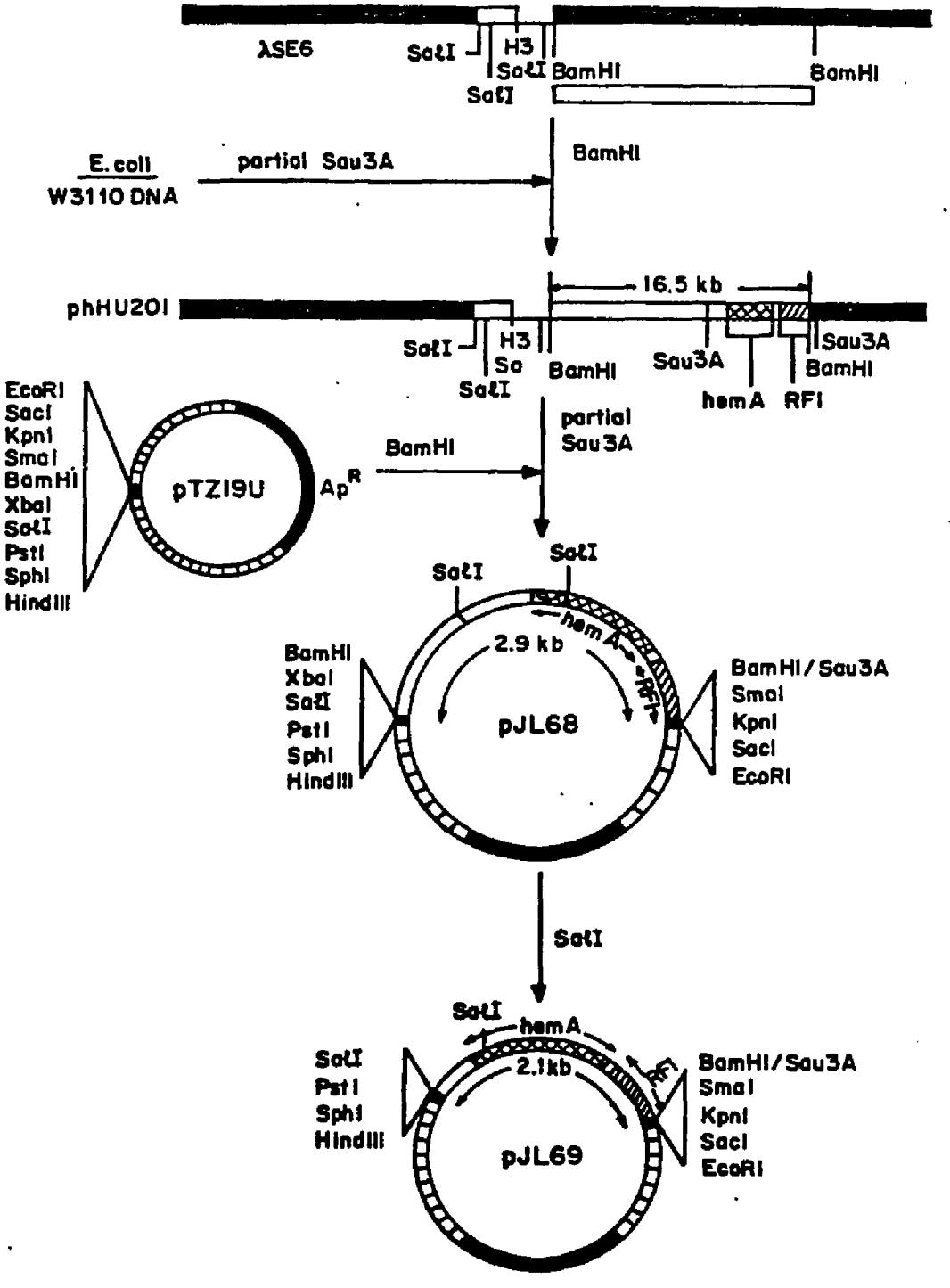
2. CLONING AND SEQUENCING OF THE *hemA* GENE OF *E. COLI*

2.1. CLONING OF THE *hemA* GENE

The cloning strategy for the *hemA* gene is presented in Figure 14. HU2001 was obtained as described previously (Umanoff et al, 1988). An *E. coli* genomic library in λSE6 was introduced by phage infection into HU2000, the hemin permeable *hemA* mutant which carried a cloned temperature-sensitive allele of the gene encoding the λ phage represser protein. Hem$^+$ colonies were selected on minimal medium lacking ALA, and one of these, HU2001, was chosen for purification of its phasmid DNA.

The phasmid (phHU201) isolated from HU2001 contained a 16.5 kb insert. The phasmid DNA was partially digested with Sau3A (5 min at 37°C) and the DNA ligated into the BamHI site of pTZ19U and pTZ19R, which is located in a *lacZ* gene. Resulting plasmids were transformed into JM101, and white colonies were selected on medium with X-gal and amp. DNA was prepared from each of the 172 purified white colonies and used to transform HU227 to heme prototrophy. One of the plasmids resulting from the ligation of pTZ19U and Sau3A digested phHU201 contained the gene for heme prototrophy, and was found to have a 2.9 kb insert (pJL68).

Figure 14. Strategy for cloning *hemA*. An *E. coli* genomic library in λSE6 was introduced by phage infection into HU2000. Hem$^+$ colonies were selected on minimal medium lacking ALA, and one of these, HU2001, was chosen for purification of its phasmid DNA. Phasmid DNA (phHU201) was partially digested with Sau3A at 37°C for 5 min. The vector plasmid, pTZ19U, was digested with BamHI at 37°C for 2 h. These DNA's were ligated with T_4 ligase at 15°C for 2 h in ligation buffer (50 mM Tris-HCl, pH 7.6; 10 mM DDT; 1 mM ATP; 50 mg/ml BSA; 10 mM $MgCl_2$). JM101 was transformed with the ligation mix and transformants were screened on X-gal supplemented medium according to the procedure of the supplier, U.S. Biochemical. Plasmid DNA was isolated from each transformant using the mini-prep procedure of U.S. Biochemical, except that the extraction with phenol and chloroform was not performed. HU227 was transformed with the DNA from each white colony to ALA independence. ALA was determined in culture supernatants or cell extracts as described previously (Li et al., 1989b).


λSE6
SalI
H3
SalI
BamHI
BamHI
SalI
E. coli
W3110 DNA
partial Sau3A
BamHI
16.5 kb
phHU201
SalI
H3
So
SalI
BamHI
Sau3A
Sau3A
hemA
RFI
BamHI
EcoRI
SacI
KpnI
SmaI
BamHI
XbaI
SalI
PstI
SphI
HindIII
pTZ19U
Ap^R
BamHI
partial
Sau3A
SalI
SalI
hemA
2.9 kb
RFI
pJL68
BamHI
XbaI
SalI
PstI
SphI
HindIII
BamHI/Sau3A
SmaI
KpnI
SacI
EcoRI
SalI
SalI
hemA
2.1kb
RFI
pJL69
SalI
PstI
SphI
HindIII
BamHI/Sau3A
SmaI
KpnI
SacI
EcoRI

This was cut partially with SalI and religated. A plasmid, pJL69, carrying a 2.1kb insert with the ability to transform HU227 to heme prototrophy was selected.

HU227 harboring pJL68 (JL1268) produced ALA and excreted it into the medium. pJL68 transformed a second *E. coli hemA* mutant, SHSP19, to heme prototrophy. When both JL1268 and SHSP19 carrying pJL68, were incubated for more than 24 h on solid medium, they fluoresced, indicating that they were accumulating porphyrins which are synthesized from ALA. However, enzyme assays for ALA S activity using glycine and succinyl CoA as substrates were unsuccessful even though JL1268 contains a high-copy number plasmid which enables the strain to overproduce ALA. Subsequently, it was shown, as discussed below, that JL1268 does not use succinyl CoA and glycine as substrates to produce ALA. Instead, it incorporates the intact five carbon chain of glutamic acid into ALA and [^{14}C]-1-glutamic acid specifically labels C-5 of ALA (Li et al., 1989b).

2.2. SEQUENCE OF THE CLONED INSERT

The sequencing strategy is shown in Figure 15. The 2.9 kb insert of pJL68 was subcloned into phages M13mp18 and M13mp19 and a series of nested deletions (Yanisch-Perron et al., 1983; Dale et al., 1985) in a single strand of the DNA of the insert was made for both strands. Thirty-two overlapping fragments which covered almost all of both strands of the entire insert were isolated from phages. The sequences of both strands were determined by the dideoxy method (Sanger et al., 1977). Two gaps were filled by fragments formed with synthesized oligomers as primers.

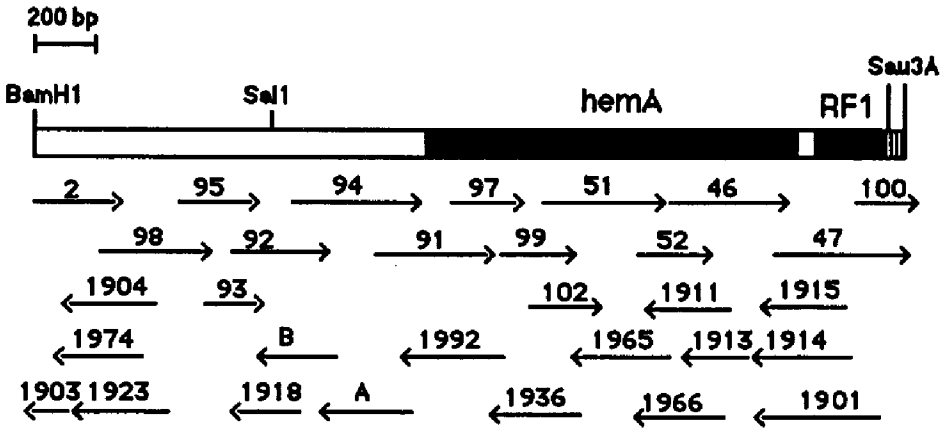


Figure 15. Sequencing strategy for the *hemA* gene. pJL68 was double-digested with BamHI and EcoRI and subcloned into M13mp18 and M13mp19. A series of nested deletions of the inserts in the phage vector were constructed with T_4 DNA polymerase (Dale et al., 1985)using the Cyclone system (IBI). Thirty-two overlapping fragments of the insert of pJL68, cloned into these phages were sequenced by the Sanger dideoxy method (Sanger et al., 1977) using the Sequenase kit (U. S. Biochemical) and IBI sequencing apparatus (STS45) with a wedged spacer. The black boxes are ORF's: hemA=ORF1 and RF1=part of ORF2. The striped box represents a vector sequence which is part of ORF2. The open box represents the 5' flanking region. Numbered fragments are from M13 phages. Fragments A and B were obtained using oligonucleotide primers synthesized in this work.

Figure 16. The complete nucleotide sequence of the sense strand and the amino acid sequence of the *hemA* gene. The sequences were read on an IBI gel reader and analyzed with IBI DNA/protein sequence analysis software and DNA Inspector II+ (Textco, NH). DNA databases (NIH GenBank Version 54.0 Release, EMBL Nucleic Acid Sequence Database Release 14.0) and protein databanks (PIR Protein Sequence Databank Release 15.0, SWISS-PROT Release 6.0) were searched on Bionet April 1988 (Intelligenetics, Inc., CA), using XFASTN/P and IFIND protocols. The promoter search was performed using the program of Mulligan and McClure (1986).The protein coding region was determined by two methods. A search for ORF's of more than 100 amino acids was made using DNA Inspector II+. Additionally, protein coding regions were located with three *E. coli* codon bias tables generated in this lab using the IBI codon bias table generator program. A protein coding region for a low expression gene (Sharp and Li, 1986) obtained in this manner corresponded to the ORF obtained with DNA Inspector II+. The possible promoter region is shown by a single underline. The possible Shine-Dalgarno sequence is shown by a double underline. The initiation codon for RF1 is designated by an asterisk. The region of homology (nt 2272 - 2841) with the published sequence of RF1 and its upstream sequences is bracketed (Craigen et al., 1985). The corrections to the published sequence are shown in parentheses.

```
   1 GGA TCC ACC GGC GTT AGT ATT TCA CCA ACG CCT TCG GCA AAC GCG GCA TGC CCC CGA ACA
  61 AAG TCA GGA ACA TCT GCG CCC AGC GTC AGC CCC ATT TCC GCC AGC TCA TCC ATG CTT AGC
 121 CCG CAT TGC CAG AGA TGA TTT AAT GCC ACC AGG ACC GTG CGG CAT TGG ATG AAC CAC CGC
 181 CGA GAC CGC CGC CCA TCG GCA AAC GCT TGT CAA TGC TGA TAT TCG CAC CGC TTC CCG TCG
 241 GAA GAC GCC CGC TGT CTG CCG CAG TTT TCA TCA ACA ATC GCG CTG CGC GAA CGA TCA GGT
 301 TAT CTT CAT GTT CCA CGC CTT CAA CGG GCG TTA ACA GAC GAA TAT CCC CAT CGT CAC GAA
 361 GCT CAA TGC TGA TGG TGT CGC CGT AAT CAA GAA ACT GAA ACA GCG TTT GCA GCG TGT GGT
 421 AAC CAT CCG CAC GCT GAC CGG TAA TGT ATA AAA ACA GAT TAA GTT TTG CCG GAG AGG GCC
 481 ACT GTG TCC GCA TTA TTT CAC TAT CCA GTT ATC CAA TTT TTA ACT TGA TGC GTT GAT CCA
 541 CCG TCG GTG AGT TCC ATA TTG GCT GGC ATC GCA GGT TGC GTT TTG GTG TCA TAC CAC CAT
 601 AAA CAA CCT TCC AGT TTT TGC CAT TCT GGC TGT AGG TAA TTT CGC TCA GGC GGT ACT GGT
 661 CGT CCA GTT TGT AGT CGG TTG CAT CAC CGG TAA ACC TAA AAT CCA CTG GCG CAA CGT GTT
 721 GAG CGG AAT TGG CAT TCC GGT CAA TTT GCC AAT CAT CTC TTC GGC GTC ATC GGC GGT ATA
 781 ACG CTG ACC TTT ATT GTC GAC TAA CTG CAC GTT ACC CGG TTG AGC ATT CAG CTC CAG TTC
 841 CGT GCT GCC CAA TGG GTT AGT GAG CAG CAG ACG GTA GCG ATC CTG GCC GGT TTG CTG CCA
 901 GAA AAA GCG GGC GTA CAC TTT TTG TTG GTC AGA AAT ATA AGC GAA CGC GCC GCG AGT CTG
 961 ATA CTG ATT AAG ATT GCG CAC GTC TTG CTG ATG CTG ACG CCA TTG TGG CGA ATC CGG GCT
1021 TTT GCC AGG ACC TTT GGG CGT GGT AAC GGA ACA GGC AGT GAG CAC AAG AGC AGC CAG CGG
1081 TAG CAG GCG GAT AAG ACG AAA ATC GGG CAG GGG CAT AGT GAT GAC AAG TCC TTG AGA TAC
1141 GTT GCA GTT ATA ACC CTT AAT GCT AGC GTT ACC GTC CGC TAT CGT CTA TGT TCA AGT TGT
1201 CTT AAT TGC CAG AAT CTA ACG GCT TTC GGC AAT TAC TCC AAA AGG GGG CGC TCT CTT TTA
1261 TTG ATC TTA CGC ATC CTG TAT GAT GCA AGC AGA CTA ACC CTA TCA ACG TTG GTA TTA TTT
1321 CCC GCA GAC ATG ACC CTT TTA GCA CTC GGT ATC AAC CAT AAA ACG GCA CCT GTA TCG CTG
                M   T   L   L   A   L   G   I   N   H   K   T   A   P   V   S   L
1381 CGA GAA CGT GTA TCG TTT TCG CCG GAT AAG CTC GAT CAG GCG CTT GAC AGC CTG CTT GCG
     R   E   R   V   S   F   S   P   D   K   L   D   Q   A   L   D   S   L   L   A
1441 CAG CCG ATG GTG CAG GGC GGC GTG GTG CTG TCG ACG TGC AAC CGC ACG GAA CTT TAT CTT
     Q   P   M   V   Q   G   G   V   V   L   S   T   C   N   R   T   E   L   Y   L
1501 AGC GTT GAA GAG CAG GAC AAC CTG CAA GAG GCG TTA ATC CGC TGG CTT TGC GAT TAT CAC
     S   V   E   E   Q   D   N   L   Q   E   A   L   I   R   W   L   C   D   Y   H
1561 AAT CTT AAT GAA GAA GAT CTG CGT AAA AGC CTC TAC TGG CAT CAG GAT AAC GAC GCG GTT
     N   L   N   E   E   D   L   R   K   S   L   Y   W   H   Q   D   N   D   A   V
1621 AGC CAT TTA ATG CGT GTT GCC AGC GGC CTG GAT TCA CTG GTT CTG GGG GAG CCG CAG ATC
     S   H   L   M   R   V   A   S   G   L   D   S   L   V   L   G   E   P   Q   I
1681 CTC GGT CAG GTT AAA AAA GCG TTT GCC GAT TCG CAA AAA GGT CAT ATG AAG GCC AGC GAA
     L   G   Q   V   K   K   A   F   A   D   S   Q   K   G   H   M   K   A   S   E
1741 CTG GAA CGC ATG TTC CAG AAA TCT TTC TCT GTC GCG AAA CGC GTT CGC ACT GAA ACA GAT
     L   E   R   M   F   Q   K   S   F   S   V   A   K   R   V   R   T   E   T   D
1801 ATC GGT GCC AGC GCT GTG TCT GTC GCT TTT GCG GCT TGT ACG CTG GCG CGG CAG ATC TTT
     I   G   A   S   A   V   S   V   A   F   A   A   C   T   L   A   R   Q   I   F
1861 GAA TCG CTC TCT ACG GTC ACA GTG TTG CTG GTA GGC GCG GGC GAA ACT ATC GAG CTG GTG
     E   S   L   S   T   V   T   V   L   L   V   G   A   G   E   T   I   E   L   V
1921 GCG CGT CAT CTG CGC GAA CAC AAA GTA CAG AAG ATG ATT ATC GCC AAC CGC ACT CGC GAA
     A   R   H   L   R   E   H   K   V   Q   K   M   I   I   A   N   R   T   R   E
1981 CGT GCC CAA ATT CTG GCA GAT GAA GTC GGC GCG GAA GTG ATT GCC CTG AGT GAT ATC GAC
     R   A   Q   I   L   A   D   E   V   G   A   E   V   I   A   L   S   D   I   D
2041 GAA CGT CTG CGC GAA CGC GAT ATC ATC ATC AGT TCC ACC GCC AGC CCG TTA CCG ATT ATC
     E   R   L   R   E   R   D   I   I   I   S   S   T   A   S   P   L   P   I   I
2101 GGG AAA GGC ATG GTG GAG CGC GCA TTA AAA AGC CGT CGC AAC CAA CCA ATG CTG TTG GTG
     G   K   G   M   V   E   R   A   L   K   S   R   R   N   Q   P   M   L   L   V
2161 GAT ATT GCC GTT CCG CGC GAT GTT GAG CCG GAA GTT GGC AAA CTG GCG AAT GCT TAT CTT
     D   I   A   V   P   R   D   V   E   P   E   V   G   K   L   A   N   A   Y   L
2221 TAT AGC GTT GAT GAT CTG CAA AGC ATC ATT TCG CAC AAC CTG GCG CAG CGT AAA GCC GCA
     Y   S   V   D   D   L   Q   S   I   I   S   H   N   L   A   Q   R   K   A   A
2281 GCG GTT GAG GCG GAA ACT ATT GTC GCT CAG GAA ACC AGC GAA TTT ATG GCG TGG CTG CGA
     A   V   E   A   E   T   I   V   A   Q   E   T   S   E   F   M   A   W   L   R
                                                        (-)
2341 GCA CAA AGC GCC AGC GAA ACC ATT CGC GAG TAT CGC AGC CAG GCA GAG CAA GTT CGC GAT
     A   Q   S   A   S   E   T   I   R   E   Y   R   S   Q   A   E   Q   V   R   D
2401 GAG TTA ACC GCC AAA GCG TTA GCG GCC CTT GAG CAG GGC GGC GAC GCG CAA GCC ATT ATG
     E   L   T   A   K   A   L   A   A   L   E   Q   G   G   D   A   Q   A   I   M
                                            (C)
2461 CAG GAT CTG GCA TGG AAA CTG ACT AAC CGC TTG ATC CAT GCG CCA ACG AAA TCA CTT CAA
     Q   D   L   A   W   K   L   T   N   R   L   I   H   A   P   T   K   S   L   Q
                                                            (-)
2521 CAG GCC GCC CGT GAC GGG GAT AAC GAA CGC CTG AAT ATT CTG CGC GAC AGC CTC GGG CTG
     Q   A   A   R   D   G   D   N   E   R   L   N   I   L   R   D   S   L   G   L
2581 GAG TAG CAG TAC ATC ATT TTC TTT TTT TAC AGG GTG CAT TTA CGC CTA TGA AGC CTT CTA
     E   -                                                   *
2641 TCG TTG CCA AAC TGG AAG CCC TGC ATG AAC GCC ATG AAG AAG TTC AGG CGT TGC TGG GTG
2701 ACG CGC AAA CTA TCG CCG ACC AGG AAC GTT TTC GCG CAT TAT CAC GCG AAT ATG CGC AGT
2761 TAA GTG ATG TTT CGC GCT GTT TTA CCG ACT GGC AAC AGG TTC AGG AAG ATA TCG AAA CCG
2821 CAC AGA TGA TGC TCG ATG ATC TGC GCT ACC TGC TGT ACC TGC GGC TTT TCG TCG TAC TGT
2881 TCC GGC ATG TTG TTG GCG ATC CCC GGG TAC CGA GCT CGA ATT C
```

The complete nucleotide sequence of the sense strand and the amino acid sequence are presented in Figure 16.

Two long ORF's were found in the presented sequence by searching for ORF's with the computer software, Inspector II+ (Appendix H), and these two ORF's were confirmed by locating protein-coding regions with *E. coli* codon bias tables generated in this study (Appendix I). ORF1 is located at nt 1328-2583 and codes for a 418 amino acid protein with a MW of 46 kDa. ORF2 starts at nt 2628 and continues into part of the vector, but the part of the ORF2 derived from the 2.9 Kb insert ends at nt 2901. Thus nt 2628-2901 encodes a protein fragment of 91 aa. Two additional ORF's were seen in the upstream region of *hemA* in the opposite direction (ORF3 and ORF4). ORF4 is located in the region of nt 1116 to nt 646 and encodes a protein of 156 amino acids, which has a molecular weight of 17.6 kDa. The other one (ORF3) starts at nt 493 and stops at nt 140. ORF3 encodes a 117 aa protein which has a molecular weight of 12.8 kDa. These two ORF's were also confirmed by locating protein-coding regions with *E. coli* codon bias tables. Certain regions of the deduced amino acid sequences of these two ORF's showed some homologies to several proteins in the two protein databases. The functions of the two proteins has not been determined.

A possible promoter site for ORF1 was identified (Mulligan and McClure, 1986) and is indicated in Figure 16. The possible promoter site has a score of 50.3 on the Mulligan and McClure scale, indicating that it is a Class I or low expression promoter. A possible Shine-Dalgarno sequence was found before the initiation codon of ORF1 and is indicated in Figure 16. A stem-loop structure for the termination of transcription

which is located from nt 2579 to nt 2592 is found at the stop codon of the gene, and is followed immediately by a string of T's (Figure 16).

A homology search of the amino acid sequences deduced from all six reading frames of the nucleotide sequence was carried out against amino acid sequences of cloned ALA synthases. No homology was found between the deduced amino acid sequence of ORF1 and ALA synthases from human, chicken, *Bradyrhizobium japonicum,* yeast and part of *Rhizobium meliloti* , using a homology matrix generated using the analysis program of Pustell and Kafatos (1984) set as described by McClung et al. (1987), although these genes show strong homology among themselves. *B. japonicum* shows over 48% homology with chicken liver ALA synthase and 53% homology with a 76 aa sequence of the *R. meliloti* protein (McClung et al., 1987). The aa sequence of the yeast enzyme is 41% homologous to that of chicken liver ALA synthase, and almost identical to a portion of the *R. meliloti* protein (Urban-Grimal et al., 1986). In an analysis done in this study, using a computer generated matrix, the human and *B. japonicum* ALA synthases showed strong homologies (Appendix J). In addition, the amino acid sequence of ORF1 was not homologous to any other sequence in two protein databases. DNA databases did, however, show strong homology (99.5%) between the sequence of nt 2272 to 2841 (569 nt long) and the sequence of part of protein synthesis release factor 1 (RF1) and its upstream region which had been determined by Craigen et al. (1985). The difference is one mismatch and two unmatches. The three different nucleotides were seen on both strands in our experiments and are indicated in Figure 16. This sequence includes the end of ORF1 and the intervening 41 nt region and part of ORF2. Thus the beginning of ORF2, nt 2628-2841 encodes 71 aa of RF1. RF1 maps at

min 26.7 on the *E. coli* genome (Ryden et al., 1986; Lee et al. 1988), thus establishing that ORF1 encoding the *hemA* allele of SASX41B, like that of the *hemA* allele of SHSP19B (Sasarman et al., 1968), maps at 26.7 min. The sequence of nt 2837 to 2901(right terminal 64 nt of the insert) showed 100% homology with a fragment of λ phage DNA, indicating that it is derived from the right arm of the λ phage vector. Thus the 2.9 kb insert of pJL68 is located at the right end of the insert of phHU201. The nt 1-2841 came from *E. coli* chromosome and codes for *hemA* and part of RFI.

Table 8 shows the P2' value, usage of modulatory codons and Q/D (quartet/duet) ratios of *hemA* gene. The P2' index for the protein encoded by ORF1 is 0.31(Sharp and Li, 1986), indicating that it is expressed at a low level. According to its use of modulatory codons (1.67%) ORF1 is a moderately-expressed gene. A third method to analyze the expression of a gene is to calculate the Q/D (quartet/duet), as described earlier (Grantham et al., 1981). According to the ratios of Q/D, ORF1 is expressed at a low level. Combining the result of codon usage and the low strength of the promoter of ORF1, *hemA* should be a low-expressed gene.

Since the *hemA* gene showed no homology to that of ALA S from a variety of organisms, a different search was made to understand the function of the cloned gene. The sequences of DNA-binding regions of 10 known regulatory proteins were searched against the *hemA* amino acid sequence. These DNA-binding proteins interact with their target site through specific hydrogen bonds made by amino acid side chains present in characteristic helix-turn-helix motif (Pabo and Sauer, 1984). Many other regulatory proteins possess sequence homologies to this bihelical region (Gicquel-Sanzey and Cossart, 1982). No such sequence was observed.

The hydropathy profile (Figure 17) of the deduced protein sequence suggested that the protein is water-soluble. The secondary structure of the amino acid sequence deduced from ORFI was predicted by a Chou and Fasman algorithm using a computer program on Bionet (Chou and Fasman,1974) (Appendix K). A typical βαβ structure is seen in the region of aa 180 to aa 215 of *hemA* protein. This βαβ structure is known to be associated with nucleotide binding activity (Wierenga and Hol,1983; Taylor, 1987). Two glycine residues, which are highly conserved in the βαβ structure for the nucleotide binding, are seen. This is expected because the dehydrogenase of other organism use NADPH as a cofactor and we have shown that ALA synthesis in *E. coli* is NADPH dependent.

2.3. MAXICELL ENZYME SYNTHESIS OF THE CLONED GENE

A maxicell protocol (Sancar et al., 1979) was used to analyze the genes cloned on pJL69, the smaller of the two plasmids which complemented the *hemA* mutants. After UV irradiation and subsequent growth of the cells, extracts were prepared and proteins which had been synthesized in the presence of [^{35}S]-methionine were resolved on SDS-PAGE gels and visualized by autoradiography. As seen in Figure 18, pJL69 produced three proteins, which have subunit weights of 23 kDa, 28 kDa and 41 kDa. The control plasmid, pTZ19U, produced the 28 kDa protein, which is the β-lactamase encoded by the *amp* gene of the vector. The 41 kDa protein is probably produced by ORF1, because the sequence specifies a 46 kDa protein. The sequencing data suggest that a fusion protein of part of RF1 and a polypeptide fragment specified by the plasmid vector has a MW of 23 kDa. However, unlike the result with the *hemB* protein (ALA D) labeling in the maxicell procedure, Coomassie blue

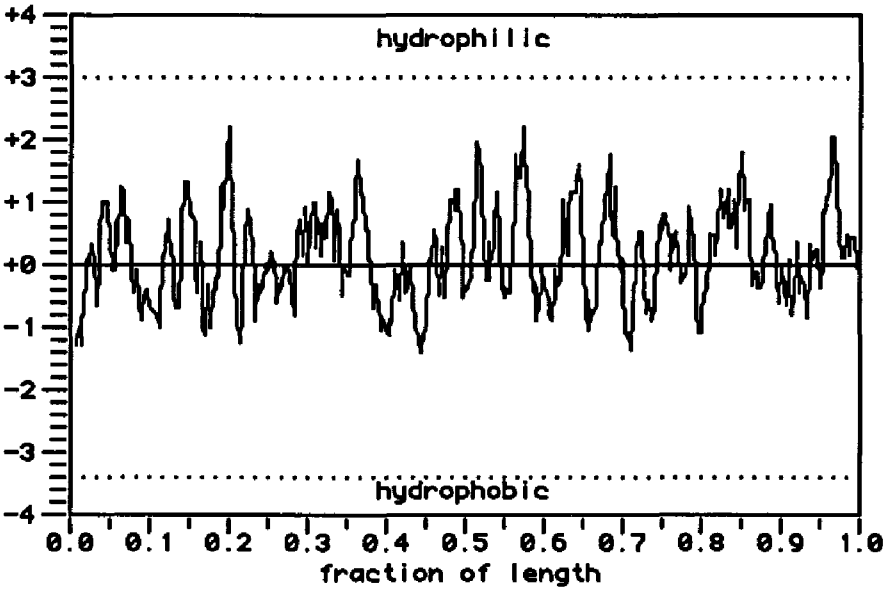


Figure 17. Hydropathy profile of the *hemA* protein obtained from a Hopp-Woods hydrophilicity plot using DNA Inspector II+ software.

Figure 18. Autoradiogram of proteins produced in the presence of L-[^{35}S]methionine in the maxicell procedure (Sancar et al., 1978, 1979). The proteins were analyzed on sodium dodecyl sulfate-10% polyacrylamide disc gels. Lanes: 1, CSR603; 2. CSR603/pTZ19U; 3. CSR603/pJL69, incubation with unlabeled methionine for 30 min after exposure to L-[^{35}S]methionine; 4. same as 3, but incubation with unlabeled methionine was for 60 min; 5. same as 3, but incubation with unlabeled methionine was for 90 min; 6. same as 3, but incubation with unlabeled methionine was for 120 min; 7. same as 3, but incubation with unlabeled methionine was for 150 min. Coomassie blue staining of these samples and of standards showed that the band in lane 2 had a molecular size of 28 kDa (β-lactamase) and the bands in lanes 3-7 had molecular sizes of 28 (β-lactamase), 41 (ORF1) and 23 (ORF2, part of RF1 and vector) kDa, respectively.

staining did not show overproduction of the 41 kDa protein. To investigate the stability of the protein, irradiated cells containing the plasmids were pulse-labeled with [^{35}S]-methionine for 5 min, followed by a chase with cold methionine in order to assess the stability of the proteins. Both the 41 kDa and 23 kDa bands became lighter with increased incubation in cold methionine, but the intensity of the 28 kDa band did not change. Thus, the proteins encoded by ORF1 and ORF2 appear to be less stable than β-lactamase. The 23 kDa band was lighter than the other two. The 23 kDa protein is not a native *E. coli* protein and may be degraded. This fusion protein contains 71 aa residues of RF1, 20 aa residues encoded by a fragment of λ phage from the original library, and 109 aa which result from translation through the insert into part of the vector (nt 295-625 of the pTZ19U map).

3. ALA SYNTHESIS

Since *hemA* did not encode ALA S, studies were undertaken to determine how ALA is synthesized in *E. coli*. As described in the introduction, ALA can be synthesized through either the C_4 or C_5 pathway. In the C_5 pathway, ALA is synthesized from glutamic acid via three enzymatically catalyzed steps: the activation of glutamic acid to form glu-tRNA, dehydrogenation to form GSA and aminotransference to form ALA. These steps require tRNA and other cofactors.

JL1268 excreted ALA into the medium (data not shown). When it was grown in the presence of levulinic acid, an inhibitor of ALA dehydratase, more ALA accumulated in the medium. (Figure 19). Crude extracts of JL1268 were able to synthesize ALA when incubated with glutamic acid and ATP (2.32 nanomoles ALA/mg protein in extract) or

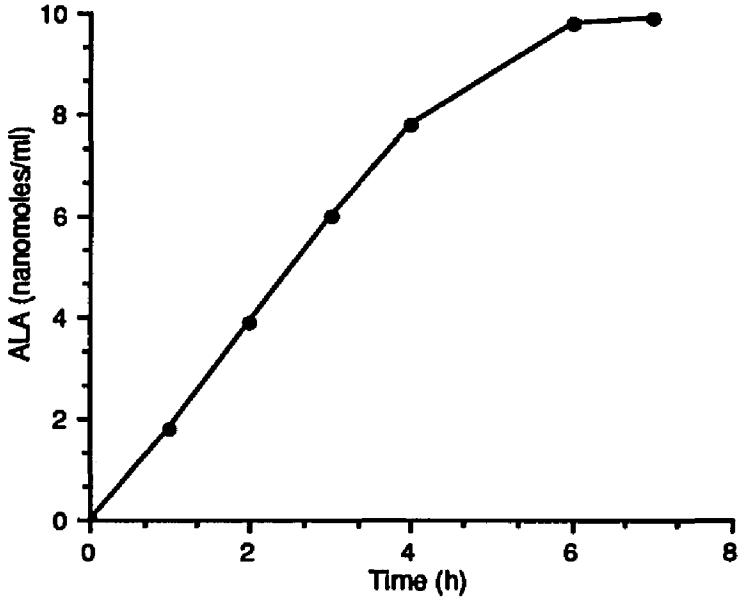


Figure 19. Time dependence of excretion of ALA into the medium by JL1268. A cell suspension (10 ml, $2x10^9$ cells per ml) in glucose minimal medium supplemented with met, glu and amp (50 μg/ml each) and also containing levulinic acid (4 mM) was incubated at 37^0C. A 1-ml aliquot was removed every hour. The cells were removed by centrifugation and ALA was determined in the supernatants by the method of Burnham (1970). ALA is expressed as nanomoles per milliliter of supernatant.

glycine, succinyl CoA and ATP (1.13 nanomoles ALA/mg protein), twice as much ALA being produced in the former condition (Table 9). The blank was an extract control with no additions. This blank gave higher results than a 0 time control, used by other investigators, and was used as a correction for calculations of net synthesis. The synthesis of ALA by JL1268 with glycine and succinyl CoA as substrates was probably due to stimulation by ATP in the assay mixture (see below). Significantly, under the same conditions, a *hemA* mutant harboring a plasmid containing the structural gene for ALA synthase from *Rhodobacter spheroides,* which uses the C_4 pathway [strain TBI/pUI553], used glycine, succinyl CoA and ATP to make ALA (1.77 nanomoles/mg protein) but produced no net ALA from glutamic acid and ATP (Table 9).

In order to determine which substrate *E. coli* uses, the following labeling studies were undertaken. [^{14}C]-Glutamic acid and [^{14}C]-glycine were both taken up by the cells of JL1268 (Table 10). A substantial amount of the radioactivity recovered from the chromatogram lane comigrated with ALA pyrrole when glutamic acid was the precursor while [2-^{14}C] glycine gave practically no counts above background in the same region. This demonstrated that [1-^{14}C]-glutamic acid was a precursor for ALA and [2-^{14}C]-glycine was not. However, the same amount of ALA was recovered from ion exchange chromatography regardless of the labeled precursor (Table 11). ALA, obtained by the incorporation of labeled precursors was degraded with periodate to obtain labeling patterns (Figure 5) which could be correlated with the C_4 or with the C_5 pathway. These results are shown in Table 12. The C-5/C-1-4 ratios of specific radioactivities is much higher for [1-^{14}C] glutamic acid as a precursor than for uniformly-labeled [^{14}C] glutamic acid or for [2-^{14}C] glycine. In fact, the

Table 9. ALA synthesis by crude extracts of *E. coli* strains in the presence of different substrates

Strain[a]	Substrate[b]	
	Glutamic acid	Succinyl CoA+Glycine
JL1268	2.32	1.13
TBI/pU1553	0	1.77

[a], Cells were grown overnight in LB medium with 50 μg/ml of amp and disrupted by sonication.

[b], When succinyl CoA and glycine were used as substrate,the reaction mixture contained: glycine, 100 mM; succinate, 100 mM; Tris-Cl, pH 7.5, 50 mM; CoA, 0.37 mM; pyridoxal phosphate, 0.27 mM; succinyl CoA, 25 μM, levulinic acid, 5 mM and 0.2 ml of the crude extract. When glutamic acid was the substrate, the reaction mixture contained 0.1 M Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT, 3 mM glutamic acid, 5 mM ATP, 1 mM NADPH, 20 μM pyridoxal phosphate, 5 mM levulinic acid and 0.4 ml of cell extract. The blank was crude extract and buffer. ALA was determined by Burnham's method (1970). Specific activity is defined as nanomoles ALA/mg protein.

Table 10. Uptake of ^{14}C-glutamic acid and ^{14}C-glycine by *E. coli* JL1268

Time (min)	[1-^{14}C]glutamic acid	[2-^{14}C]glycine
2	4851.00	1683.00
5	8125.00	2923.00
10	13705.0	5045.00
15	14543.0	6639.00
30	14215.0	7932.00

Cells were grown overnight in M9 medium with supplements, washed with fresh medium, concentrated to 1/5 of the original volume in the same medium but containing 4 mM levulinic acid. After 1 h incubation at 37° C, radiolabeled precursors were added into the individual cultures, respectively. 0.1 ml of the culture was withdrawn at each time interval and washed six times on a fiber glass filter with 10% TCA. The radioactivity of each filter was counted in 5 ml of ScintiVerse.

Table 11. Paper chromatography of ALA pyrrole derived from ALA excreted into the medium of *E. coli* JL1268

Paper strip no.	Rf	ALA pyrrole labeled (cpm) with precursor: [1-^{14}C]glu	[U-^{14}C]glu	[2-^{14}C]gly
1	0.032	841[a]	827[a]	93[a]
2	0.097	756	67	20
3	0.162	1385	1497	16
4	0.227	638	889	11
5	0.292	640	901	8
6[b]	0.354	2870[c]	2397[c]	20[c]
7[b]	0.422	2598[c]	1624[c]	16[c]
8[b]	0.487	682[c]	1052[c]	10[c]
9	0.552	402	494	3
10	0.617	336	526	2
11	0.682	303	521	16
12	0.747	259	482	47
13	0.812	200	424	3
14	0.877	111	180	6
15	0.942	38	50	9

[a] Ninhydrin positive

[b] Strips 6, 7, and 8 together contained 51% of the [1-C^{14}]glutamic acid, 40% of the [U-C^{14}]glutamic acid, and 16% of [2-C^{14}]glycine total counts.

[c] Ehrlich reagent positive

Table 12. Periodate degradation of ALA isolated from medium of *E. coli* JL1268 grown with ^{14}C-labeled precursors[a]

Precursor	Incorporation of label				Ratio
	C-5[b]		C-1-4[c]		
	cpm	cpm/μmol	cpm	cpm/μmol	C-5/C-1-4
[1-^{14}C]glutamic acid	25,907	104	964	1.83	57
[U-^{14}C]glutamic acid	4707	22	4989	11.03	2
[2-^{14}C] glycine	690	3	132	.35	9

[a] Numbers refer to the numbering shown in Figure 5. Each precursor was used at 10 μCi per culture.

[b] Dimedone derivative of formaldehyde.

[c] Succinic acid.

highest ratio obtained by Troxler and Offner (1979) for standard [5-^{14}C] ALA was an average of 66/1 which compares favorably with our ratio of 57/1. The labeling by glycine might be explained by carry-over of a small amount of glycine during the ion exchange chromatography isolation of ALA, since paper chromatography of the ALA fractions in the [2-^{14}C]glycine experiment contained a small amount of ninhydrin-positive material at the origin and the highest radioactivity in the ALA pyrrole chromatogram for this experiment was in this spot.

Extracts of JL1268 were assessed for the ability to use glutamic acid as substrate and their cofactor dependence in several assays. The soluble and pellet portions of sonicates of JL1268 each had very low ALA synthesis activity, compared with the whole sonicate, but, when they were mixed, activity was restored. The pellet preparation may contain active but insoluble aggregates or a membrane-bound species of the cloned protein which is essential for ALA synthesis. Dialysis of the washed and sonicated pellet did not affect its activity. As shown on Table 13 the maximum synthesis of ALA by the combination of S and P fractions was dependent on glutamic acid, NADPH, ATP and pyridoxal phosphate when S was processed through a Sephadex G-25 column twice to remove small molecules (Weinstein and Beale, 1985b). *E coli* $tRNA^{glu}$ (Sigma Type II) stimulated ALA synthesis in a concentration-dependent fashion (Figure 20), while non-specific *E. coli* tRNA was only marginally stimulatory. RNase only weakly inhibited ALA synthesis in extracts of JL1268 when it was added to a mixture of $tRNA^{glu}$ and the extract. However, preincubation of $tRNA^{glu}$ with RNase and then addition of extract, lowered synthesis of ALA (Figure 21). Gabaculin, an inhibitor of aminotransferase (PLP-requiring enzyme) did not show complete

Table 13. Effects of cofactors on synthesis of ALA by an extract of *E. coli* JL1268 processed through Sephadex G-25[a]

Assay mix	ALA synthesized nmol/reaction mix (% of control)
Complete	1.80 (45)
Complete + $tRNA^{glu}$ (0.5 U) (control)	4.00 (100)
Complete + $tRNA^{glu}$ (0.5 U):	
-glutamic acid	1.14 (29)
-ATP	0.40 (10)
-NADPH	1.12 (28)
-LA	1.56 (39)
-PLP	2.00 (50)

[a] The reaction mixture contained 250 µl of S (2) fraction (passed through Sephadex G-25 twice) (1.5 mg of protein), 50 µl of P fraction (0.8 mg of protein), 100 µl assay mix and buffer and additions to make 500 µl. The complete system contained 100 mM Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 1 mM NADPH, 3 mM glutamic acid, 5 mM levulinic acid and 20 µM pyridoxal phosphate. The mixture was incubated for 2 h. ALA was determined by the method of Weinstein & Beale (1985).

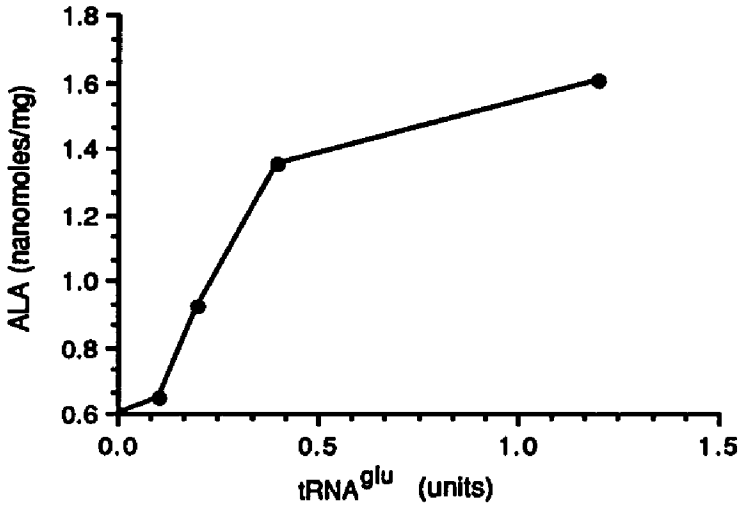


Figure 20. The effect of $tRNA^{glu}$ on ALA synthesis in JL1268. The reaction mixtures contained 250 µl of "S"(3) fraction (1.5 mg of protein) which had been run through Sephadex G-25 three times, 50 µl of P fraction (0.8 mg of protein), and 100 µl of assay mix, buffer and additions to make 500 µl, containing, for the complete system, 100 mM Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 1 mM NADPH, 3 mM glutamic acid, 5 mM levulinic acid and 20 µM pyridoxal phosphate. ALA content is expressed as nanomoles per milligram of total protein added.

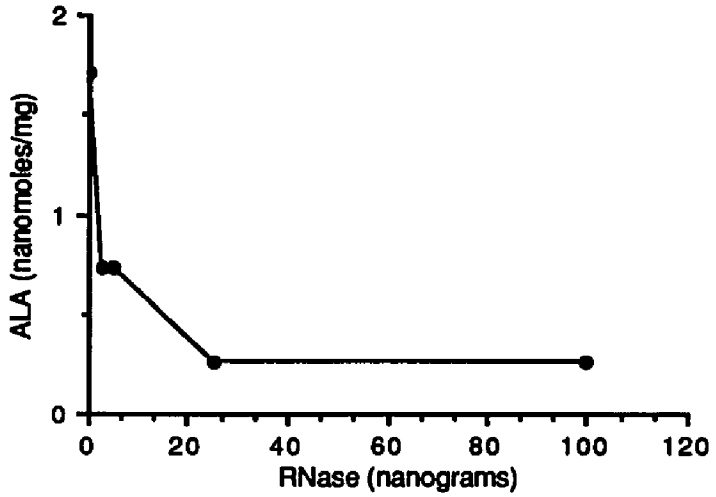


Figure 21. Effect of RNase treatment on the synthesis of ALA by JL1268. The reaction mixtures contained 250 µl of S (2) fraction (protein 1.5 mg of protein) run through Sephadex G-25 twice and 50 µl of P fraction (0.8 mg protein). $tRNA^{glu}$ (0.5 U) from *E. coli,* in buffer, was first mixed with RNase (Sigma, Type I-AS) and then incubated for 5 min at 37° C; cell extract was added, and the incubation was continued for 15 min more. Then the assay mix was added and incubation continued for 2 h. ALA is expressed as nanomoles per milligram of total protein added.

inhibition of ALA synthesis (Table 14). A similar result was seen for aminotransferase from greening barley leaves (Hoober et al., 1988). However, when the aminotransferase was first incubated with gabaculin in the presence of levulinic acid and then substrate added, the enzyme was an order of magnitude more sensitive to gabaculin. They found that sensitivity of aminotransferase to gabaculin was dependent upon the presence of keto compounds. They proposed that the coenzyme functions in the pyridoxamine form. The pyridoxal form, which is sensitive to gabaculin, is generated by incubation with a keto compound, in this case, GSA. The pyridoxamine is regenerated by removing the amino group from C-5 of modified GSA. These data may explain why the extract of JL1268 was not sensitive to gabaculin. Hemin had no effect on the synthesis of ALA by a crude extract of JL1268 (Table 15).

Table 14. Effect of gabaculin on the synthesis of ALA by an extract of *E. coli* JL1268[a]

Gabaculin (μM)	ALA (nmoles/mg) (%)
0	3.6 (100)
0.5	3.8 (105)
5	3.3 (92)
25	2.3 (61)
50	1.8 (50)
100	1.6 (44)

[a] The reaction mixture contained 250 μl of S (2) fraction (passed through Sephadex G-25 twice) (1.5 mg of protein), 50 μl of P fraction (0.8 mg of protein), 100 μl assay mix and buffer and additions to make 500 μl. The complete system contained 100 mM Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 1 mM NADPH, 3 mM glutamic acid, 5 mM levulinic acid, 20 μM pyridoxal phosphate, $tRNA^{glu}$ (0.5 U) and different amounts of gabaculin. The mixture was incubated for 2 h. ALA was determined by the method of Weinstein & Beale (1985).

Table 15. Effect of hemin on the synthesis of ALA by a crude extract of *E. coli* JL1268[a]

Hemin (μM)	ALA (nmoles/reaction) (%)
0	13.0 (100)
0.05	15.5 (119)
0.5	14.5 (111)
5.0	15.5 (119)
25.0	16.3 (125)

[a], The reaction mixture contained 0.1 M Tricine, pH 7.9, 25 mM $MgCl_2$, 5 mM DTT, 3 mM glutamic acid, 5 mM ATP, 1 mM NADPH, 20 μM pyridoxal phosphate, 5 mM levulinic acid, different amounts of hemin and 0.4 ml of cell extract. The blank was crude extract and buffer. ALA was determined with Burnham's method (1970).

CONCLUSIONS

1. Cloning and sequencing of the *hemB* gene

(1) The *hemB* gene has been cloned and sequenced. The cloned gene encodes a protein which has a subunit weight of 38 kDa as resolved on SDS-PAGE.

(2) The amino acid sequence of *E. coli* ALA D shows strong homology to ALA D for human, rat, and yeast. Therefore, ALA D is a highly conserved enzyme.

(3) Expression of the *hemB* gene is required for the appearance of PBG D activity although these two genes are far apart on the genome.

(4) Computer analysis of promoter strength and codon usage indicates that *hemB* gene is moderately-expressed.

2. Cloning and sequencing of the *hemA* gene

(1) An *E. coli hem* gene, which complements two independent *hemA* mutants of *E. coli* has been cloned. Both of the complemented strains overproduce or excrete ALA and fluoresce after twenty four hours.

(2) The cloned sequence encodes a 46 kDa protein, and expression of it by the maxicell procedure yields a 41 kDa protein as resolved on SDS PAGE.

(3) The amino acid sequence of this gene's product shows no significant homologies with any cloned ALA synthase genes nor with any protein in several databases searched. Computer analyses of the nucleotide sequence of the gene and its possible promoters indicate that it is probably expressed at a low level. In addition, the maxicell data demonstrate that it is probably a short-lived protein. This indicates that *hemA* may be a control step of the pathway.

(4) A second cloned gene fragment, which has its coding region 41 bp downstream of the coding region of the gene that complements *hemA*, has been identified as RFI. Thus the *hemA* allele of SASX41B has been established to map at min 26.7 and is very closely linked to RFI. These two genes may be controlled by the same promoter.

(5) Computer analysis of promoter strength and codon usage indicates that *hemA* gene is expressed at a low level.

3. ALA synthesis

(1) JL1268 overproduces ALA and excretes it into the medium.

(2) Glutamate, instead of succinyl CoA and glycine, is the precursor of ALA in *E. coli*. C-5 of ALA is derived from C-1 of glutamate.

(3) The synthesis of ALA is dependent on glutamate, ATP, NADPH, $tRNA^{glu}$, and pyridoxal phosphate.

(4) Two or more macromolecules may be required for the formation of ALA.

(5) It appears that bound PLP is the cofactor of aminotransferase because high concentration of gabaculin does not completely inhibit the synthesis of ALA.

(6) In contract to the case of ALA S in the C_4 pathway, hemin does not inhibit ALA synthesis in *E. coli*.

DISCUSSION

1. *HemB* GENE OF *E. COLI*

The enzyme The *hemB* gene of *E. coli* has been cloned by complementing a strain which harbors a mutant allele for *hemB*. Strains which carry the cloned plasmid produce 20 to 30 times more ALA D than wild type strains. The maxicell procedure demonstrates that the insert of pJL1002 encodes only one protein. This protein has a subunit size of 38 kDa, and has ALA D activity in at least two aggregated forms. This enzyme usually exists in octamer form. However, from the results of enzyme assay of nondenaturing gel slices, the smallest form among the three bands also had ALA D activity. It may be the monomer, instead of the octamer. Therefore the monomer may have enzyme activity, too.

The distance between hemB and lac The mutation mapped to the general region of where *hemB* is found on the *E. coli* map (Bachmann, 1983; Hadley et al., 1983; Powell et al., 1973). However, we could not confirm its precise location as tightly-linked to the *lac* locus. According to the published map, *lac* is located less than 0.25 min away from *hemB* (Bachmann, 1983). The original phasmid from which we subcloned the *hemB* gene contained a 17 Kb insert. Since one map minute contains 45 kb of DNA (Bachmann, 1980), it was possible that the phasmid would also carry the *lac* locus. This phasmid and a second one tested did not complement a *lacY* mutation. It is possible that the phasmids only carried *E. coli* genomic material downstream from *lac*, or the distance between *hemB* and *lac* is actually greater that than found previously.

In vivo control of PBG D by the hemB gene During the process of cloning of the *hemB* gene (which codes for ALA D, not for PBG D) in

Escherichia coli , a *hemB* mutant was isolated. This mutant has neither ALA D nor PBG D activity. When the mutant contains a single copy of the *hemB* gene on a plasmid it produces normal levels of ALA D and PBG D. When the mutant harbors a multi-copy plasmid containing the *hemB* gene it has 20-30 times more ALA D activity but a normal level of PBG D (Table 6) (Li et al., 1988a). The cloned plasmid which complements this strain overproduces ALA D (Figure 8), but only restores PBG D activity to normal levels. Since, as seen from the maxicell procedure, the plasmid produces only one protein, and this protein is ALA D, the gene cloned must be the structural gene for ALA D and not some regulatory locus. Therefore, it is obvious that the *hemB* mutant allele must be affecting the regulation of *hemC,* even through it maps a substantial distance away. This is the first *in vivo* demonstration that PBG D is controlled by *hemB* activity, or the product of ALA D, PBG. This contention is supported by other observations in this laboratory that a heme-permeable strain which is deficient in the *hemA* locus, when grown on heme, had no PBG D activity unless ALA was added to the growth medium (Umanoff, 1987). Presumably the ALA was converted to endogenous PBG. RP523 which has an absolute heme requirement for growth, and has a *hemB* mutation, cannot make PBG and therefore produces no PBG D activity in the presence of ALA. If PBG does, in fact, control the activity of PBG D, RP523 should synthesize PBG D when grown on medium supplemented with PBG. However, RP523 grew very poorly even in the presence of PBG at 10 μg/ml. The reason is that PBG is only slightly permeable to cells, as evidenced by the inability of *hemA* and *hemB* mutants to grow on medium supplemented with PBG at this concentration. When the concentration of PBG was increased to 40 μg/ml RP523 grew much better

and PBG D activity could be detected. When RP523 was mutagenized to be more permeable to PBG, normal PBG D activity was seen in the presence of 10 µg per ml of PBG. PBG did not affect the level of PBG D protein produced from *in vitro* transcription and translation of a plasmid harboring *hemC* (which encodes for PBG D) gene as an insert . Therefore, it seems that, in *E. coli*, PBG availability controls the activity of PBG D at some posttranscriptional level (Umanoff et al., 1988). Evidence supporting the idea includes: (1) Five different homogeneous forms of PBG D have been identified in human erythrocytes, the free monomeric enzyme, and four stable enzyme-substrate intermediates, which represent the mono-, di-, tri- and tetrapyrrole-enzyme complexes in the stepwise conversion of the monopyrrole to the linear tetrapyrrole. Using labeled PBG and electrophoretic resolution, it was shown that the binding between PBG and the enzyme is probably covalent (Anderson and Desnick, 1980). (2) PBG D is unstable in the absence of its substrate, PBG (Beaumont et al., 1986). (3) A di-PBG or dipyrrolmethane moiety, which does not normally become converted to product, but which acts as a coenzyme or anchoring group for the poly-PBG intermediates in the course of enzymatic catalysis, is irreversibly bound to PBG D (Hart et al., 1987; Jordan and Warren, 1987). Later studies showed that the dipyrromethane cofactor is bound to PBG D through the sulphur atom of cysteine-242 (Miller et al.,1988; Jordan et al., 1988; Scott et al., 1988). (4). Apo-porphobilinogen deaminase is synthesized in the *hemB* mutant, RP523, and incubation of purified apoenzyme with porphobilinogen results in reconstitution of the covalently attached dipyrromethane cofactor and a holoenzyme with about 50% of the expected catalytic activity, indicating no additional cofactor or enzymes are required for biosynthesis of holoenzyme (Scott et al., 1989).

Thus it appears that the tetrapyrrole or hydroxymethylbilane which is the precursor of uroporphyrinogen III, is assembled onto the dipyrrolemethane cofactor.

Conservation of the hemB gene Comparison of nt and amino acid sequence of *E. coli* ALA D with ALA D's from human liver, rat liver, and yeast shows that the gene for ALA D is indeed highly conserved both at DNA and protein levels (Li et al., 1989a), as is the gene for PBG D (Thomas and Jordan, 1986;), although these organisms are evolutionally far apart. The same conclusion was reached when Echelard et al. compared the deduced amino acid sequences of the *E. coli* and human genes (1988). This suggests that all ALA D's originate from the same early ancestral gene. Interestingly, it seems that the conservation of the heme biosynthetic pathway begins with the *hemB* gene, because this study shows that the *hemA* gene of *E. coli* has no homology to any known ALA S gene and that some organisms (plants, algae, strict anaerobic bacteria, and some oxygenic prokaryotes) utilize a different pathway to generate ALA.

Where transcription and translation start Analysis of the *hemB* gene sequence shows that there are two translation starting codons which lie in the same open reading frame. Each is preceded by a Shine-Dalgarno sequence CAGG (Figure 11). Also, two possible promoters are found. The question is: where do transcription and translation start. The results obtained in this study cannot answer this question. In a separate study by Echelard et al. (1988), two translation initiating codons were also found. Through mapping of the 5' end of *hemB* mRNA by primer extension it was found that transcription is initiated at nt 489 (nt 53 in their numbering system), which is immediately

downstream from the first translation starting site. Thus they claimed that translation must start at the second initiation codon (nt 516) instead of the first one (nt 483). This sounds very reasonable. However, it does not exclude the possibility that transcription and translation start at different upstream sites because (1) the fragment they sequenced is rather short and starts just 14 nt upstream from the -35 sequence, while the second promoter, which is closer to the translation start codon, obtained by the method of Mulligan and McClure (1986) was not even covered in their sequence. (2) The -35 sequence they obtained is atypical, although it might be the true -35 region. However, the method of Mulligan and McClure is assumed to be better than the one they used because this method incorporates more *E. coli* promoters and it also considers the flanking sequence of the consensus sequences at -10 and -35 regions which is known to affect transcription. (3) They argue that the fragment they sequenced has gene activity. However ALA D may still retain enzymatic activity even without the additional 11 N-terminal amino acids. This is supported by the observation that the N-terminal sequences of ALA Ds are not highly conserved (Figure 12, 13). The question can be resolved definitely by sequencing the N-terminal of purified ALA D. Mapping of 5' mRNA performed with a longer sequence at 5' end would give more reliable result.

2. *HemA* GENE OF *E. COLI*

Instability of the hemA gene product After cloning of the *hemA* gene, the maxicell procedure was used to identify the gene product. The plasmid pJL1269, which has a 2.1 kb insert, produced two more proteins in addition to the β-lactamase which was seen in the control. These two

proteins have subunit weights of 41 kDa and 23 kDa. The analysis of the sequence shows that the 23 kDa protein is a fusion protein of RFI and the vector. The 41 kDa protein must be encoded by ORFI and is the product of the *hemA* gene. Although the protein can be easily seen on the autoradiogram, Coomassie blue staining of the gel did not show overproduction of the protein. A cold methionine chase test after pulse-labeling of [^{35}S]methionine showed that the *hemA* gene product is not stable. The mechanism of the instability of *hemA* protein is unclear. A possible reason is that the *hemA* protein catalyzes a rate-limiting step of the heme biosynthetic pathway, and is stringently regulated. If the cell needs it, it will be synthesized. This idea is tentatively supported by qualitative evidence. Strain JL1268 overproduces the *hemA* gene product to a much lower level than the *hemB* gene product is produced in JL1002 (data not shown), but both accumulate porphyrins to the same extent. Thus, a small excess of the *hemA* gene product appears to result in the loss of regulation of the pathway. Unlike JL1268, HU2001, which contains a single copy *hemA* gene on the phasmid, behaves like wild type and does not overproduce ALA and porphyrins. The difference may be the result of gene dosage. However, it cannot exclude the possibility that the phasmid, which has 16.5 kb insert, may contain a regulatory gene in addition to the *hemA* gene. This gene may control the expression of the *hemA* gene. The removal of this control gene during the process of cloning of *hemA* gene may have resulted in overexpression of *hemA*.

Molecular size of the hemA protein The sequence analysis of the *hemA* gene gives a molecular weight of 46.3 kDa. The maxicell procedure with pJL69 shows a molecular weight of 41 kDa. Both methods did not give exactly the same weight. Such a difference in

molecular weights obtained by sequence analysis and other experiments are often seen. The *hemA* protein synthesized in an *in vitro* transcription-translation system with the large insert (pJL68) has a molecular weight of 46 kDa, which is consistent with the molecular weight predicted from the derived amino acid sequence (Umanoff, unpublished results). The *S. typhimurium hemA* protein shows a molecular weight of 44 and 45 kDa on SDS-PAGE gel, which is very close to 46 kDa predicted from its sequence (Elliot, 1989). Since the β-lactamase is an internal marker for both maxicell and *in vitro* transcription and translation results, an experimental error for the discrepancy may be excluded. Another possibility is that the two inserts give different molecular weight proteins. The large insert (in pJL68) has an additional 795 nt at the 5' end of the small insert (in pJL69). However, the smaller insert has the ability for heme prototrophy, and contains the whole coding region for the 46 kDa protein. The question is how two different size of proteins are generated from same gene. One possible reason is that the removal of the 795 nt at the left end results in different initiating sites of transcription and translation, so translation starts at an internal ATG , which is at nt 1447 and the resulting protein is 39 aa shorter and is close to 41 kDa. This appears unlikely because the promoter and first ATG codon are still in the shorter insert after removal of the 795 nt at 5' end. Another possibility is that the translation starts at the same place but is terminated early because of the absence of a regulatory effecter in the maxicell, which may be encoded by the *E. coli* chromosome or by a region of upstream from the *hemA*, that was removed by subcloning after Sal1 partial digestion.

Function of the hemA protein For many years ALA S activity could not be found in wild type *E. coli*. Even after cloning of the *hemA*

gene no ALA S activity is observed although the strain produces large amounts of ALA. Sequence analysis shows that the *hemA* protein has no apparent homology to any known ALA S, and also, that it is unlikely to be a regulatory gene. Recently, Hederstedt cloned the *hemA* gene of *B. subtilis* (personal communication, June,1989), Elliott cloned the *hemA* gene of *S. typhimurium* (1989), and Drolet cloned the *hemA* gene of *E. coli* (1989), but the function of the gene remained unknown until this work was done (Li et al., 1988b; Li et al., 1989b; Li et al., 1989c). This pathway for ALA synthesis requires three enzymatic steps. The synthesis of ALA is dependent on several cofactors, including $tRNA^{glu}$, ATP, NADPH and pyridoxal phosphate (Li et al., 1989b). The analysis of the secondary structure of the protein of the cloned gene reveals a nucleotide binding site, which has a βαβ motif. This structure is also seen in the *hemA* gene of *B. subtilis* (Hederstedt, personal communication, June of 1989) and *S. typhimurium* . This suggests that *hemA* gene encodes for a dehydrogenase which requires NADPH as a cofactor. Therefore it appears likely that this enzyme catalyzes the second reaction of the three step pathway (Figure 5), the conversion of glutamyl-tRNA to glutamic acid semialdehyde. Recent work by Avissar and Beale (1989) confirms that SASX41B, the original strain used for the cloning of the *hemA* gene in this study, is unable to carry out this reaction, but has enzymatic activities for the other two reactions (1989). These results strongly suggest that the *hemA* gene codes for the dehydrogenase. *S. typhimurium* and *B. subtilis* must share the same pathway for ALA synthesis because the three *hemA* genes show very strong homology; strikingly, *E. coli* and *S. typhimurium* show 94.5% homology of amino acid sequence.

HemA and RFI The DNA sequence analysis shows that the *hemA* gene is followed immediately by the RFI gene which codes for polypeptide release factor I with only a 41 nt space between the stop codon of *hemA* and the starting codon of RFI. No good promoter is found at the 5' end immediately upstream of RFI. Only one promoter, which is at the 5' end of the *hemA* gene, is seen for both *hemA* and RFI. However, in the maxicell procedure both *hemA* and RFI proteins were produced. Therefore, it is possible that both genes are under the control of one promoter. The same conclusion was reached by Elliot (1989) in the sequencing of the *S. typhimurium hemA* gene. Insertion of a transposon into the *hemA* gene is lethal to the cell, probably due a polar effect on transcription and translation of RF1, an essential protein. Therefore, these two genes are cotranscribed. However, since these two genes have different reading frames, the translation of RFI must be reinitiated. This brings up some interesting questions: what is the relationship of heme synthesis and protein synthesis, and what is the meaning of cotranscription of these two genes? It appears that they must both be constitutive because RF1 is an essential gene.

3. ALA SYNTHESIS

HemA does not encode ALA S For a long time it was believed that *E. coli* synthesizes ALA from succinyl CoA and glycine, catalyzed by ALA synthase. Efforts to detect ALA S were largely unsuccessful although some reports claimed ALA S activity in *E. coli* (Ishida and Hino, 1972; Schoenhaut and Curits, 1986; Tait et al., 1988). However, the ALA formed in these assays of ALA S was very low and was probably due to the stimulating effect of ATP existing in the assay as seen

in this study. Introduction of the ALA S gene from other organisms into *E. coli* mutants deficient in ALA synthesis complemented the mutations to heme prototrophy and caused an overproduction of ALA. However, this does not mean that *E. coli* has ALA S activity and uses succinyl CoA and glycine as precursors for ALA; it only demonstrated that, if given the proper enzyme, *E. coli* can convert succinyl CoA and glycine into ALA.

This report shows that *E. coli* does not use ALA S to synthesize ALA. The sequence of a gene required for ALA synthesis and its deduced amino acid sequence shows no homologies with any reported ALA S sequences (human , mouse , chicken liver, yeast , *Bradyrhizobium japonicum*), although all of these sequences share certain significant regions of homology. If *E. coli* uses succinyl CoA and glycine as precursors for ALA, [2-^{14}C] glycine should label ALA because the C-2 of glycine would become C-5 of ALA (Figure 5). But when JL1268 was grown with [2-^{14}C] glycine no ALA was labeled in spite of the large amount of ALA synthesized. Glycine was taken up very well by the strain (Table 10) so that radiolabeled glycine can get into the cells. It was concluded that ALA was not labeled because *E. coli* does not use the C_4 pathway.

Radiolabelling of ALA In contrast, when JL1268 was grown with [1-^{14}C] glutamic acid, ALA was labeled, and the C-1 of glutamic acid became C-5 of ALA. [1-^{14}C] glutamic acid was chosen to make sure that the labeling of ALA comes from glutamic acid, because C-1 of glutamic acid will is lost if it is converted to α-ketoglutarate and then to succinyl CoA in the citric acid cycle. If *E. coli* uses succinyl CoA as the precursor for ALA, ALA formed in this way will not be so labeled. The result of this experiment established that glutamic acid is the precursor of ALA.

The procedure used for the radiolabeling experiment has some advantages. It is necessary to purify the radiolabeled ALA obtained from cells in order to get a more reliable result. A few factors were considered in the design of the growth of JL1268: (1) To get enough ALA, a large volume of culture is necessary; (2) The complex mixture of metabolites in the cell might make isolation of pure labeled ALA difficult; (3) The incubation time after addition of radiolabeled precursor cannot be too long in order to avoid extensive metabolism. (4) The ALA accumulated in the cell and medium may inhibit further incorporation of radiolabeled precursor; (5) [1-^{14}C] glutamic acid is very expensive compared to other radiolabeled materials. JL1268 was chosen for the radiolabeling experiment for two reasons: (1) It has a multi-copy *hemA* gene, which can give large amounts of ALA; and (2) it excretes ALA into the medium, so ALA can be isolated from the growth medium. Minimal medium was chosen for its very simple composition so pure ALA could be readily obtained. Labeled precursor was added to strain JL1268 grown overnight, washed to remove any accumulated ALA in the medium, concentrated 5:1 in medium containing levulinic acid, the inhibitor of ALA D. Thus by using a high concentration of cells, high copy number of the gene, removal of existing ALA to prevent possible feedback inhibition on new synthesis and dilution of the radiolabel, and inhibition of ALA utilization, a large amount of newly synthesized, radiolabeled ALA was obtained.

E. coli uses the C_5 pathway for ALA synthesis When S was processed through Sephadex G-25 in order to remove small molecules and cofactors, the production of ALA became dependent on glutamic acid, ATP, NADPH, pyridoxal phosphate and $tRNA^{glu}$. Levulinate was also

required to prevent further utilization of ALA. $tRNA^{glu}$ stimulated ALA synthesis in a concentration-dependent manner. It is interesting to note that *E. coli* $tRNA^{glu}$ Type II acts as a specific inhibitor of the C_5 ALA-forming system in *Euglena*, even though the formation of glu-tRNA is normal or higher (Mayer et al., 1987). This suggests that the dehydrogenase is species specific for tRNA. In studies reported here, pre-treatment of $tRNA^{glu}$ with RNase lowered ALA production. There was no effect if RNase was added after tRNA had been allowed to interact with proteins in the extract. This suggests that when the tRNA is bound it is protected from RNase. Elimination of PLP from the assay mixture lowered ALA synthesis only partially. The crude extract and S (1 and 2) are not very sensitive to gabaculin, a competitor of PLP (Table 14). These results suggest that bound PLP rather than free PLP is acting as a cofactor. The above study demonstrate that *E. coli* utilizes the C_5 pathway for ALA synthesis. A later study by Avissar and Beale identified the enzymatic basis for ALA auxotrophy in SASX41B, the *hemA* mutant used for cloning in this study (1989) as glu-tRNA dehydrogenase, confirming these results.

Another indication of the operation of the C_5 pathway in *E. coli* is the following observation. Crude sonicates of JL1268 make appreciable quantities of ALA. The dialyzed P or gel-filtered S alone has no apparent ability to synthesis ALA. However, when both are combined the same amount of activity as crude extract was recovered. The requirement for both P and S means that at least two kinds of macromolecules, possibly two enzymes, are required in this pathway. But the possibility that the P may be an aggregated form of HemA cannot be excluded.

The patterns for uptake of labeled precursors, the requirement for at least two kinds of macromolecules and the requirement for cofactors

and $tRNA^{glu}$ provide compelling evidence for suggesting that ALA synthesis in *E. coli* follows the C_5 pathway by the series of steps shown in Figure 5.

Other hem mutants for ALA synthesis Since each step of the C_5 pathway is catalyzed by an enzyme as presently understood. More genetic loci responsible for these steps must exist. A number of ALA - requiring mutants of *E. coli* have been reported. Strain SASX41B, and SHSP19 represent different mutant alleles of the gene (Sasarman et al., 1968b). They map at 27 minutes on the genome, and both were complemented by pJL68. A different gene, *popC*, defined by an ALA auxotrophic mutation, was mapped at 4 minutes (Powell et al., 1973; Wulff, 1967). *popC* mutation may be in the gene for one of the other steps in ALA synthesis (glu-tRNA ligase or aminotransferase), but the strain carrying the mutation has been lost and cannot be studied. *E. coli* mutants deficient in other steps of ALA synthesis are being sought in our laboratory, but to date, only *hemA* mutants have been found. Mutations in the other two genes may be "very leaky" (Wulff, 1967), or if the gene is essential and the mutation is lethal, i e. glu-tRNA synthetase, and cannot be easily selected by ALA auxotrophy. In *S. typhimurium*, the counterpart of *popC* of *E. coli*, *hemL* is leaky.

Complex of glu-tRNA synthetase (glu-tRNA ligase) and hemA protein The first step of ALA synthesis is catalyzed by glu-tRNA synthetase. Glu-tRNA synthetase is also required for protein synthesis to incorporate glutamic acid into polypeptide. *E. coli* may have more than one glu-tRNA synthetase, one for protein synthesis, and one for heme synthesis. However, there is no biochemical evidence yet for more than glu-tRNA synthetase. In the purification of glu-tRNA synthetase from *E.*

coli and *B. subtilis*, a 46 kDa protein is often copurified in 1:1 ratio. The identity of this protein is unknown (Lapointe et al., 1985; Proulx and Lapointe, 1985). It may be a regulatory protein for glu-tRNA synthesis as suggested. However it may also be the 46 kDa *hemA* protein. If there is specific glu-tRNA synthetase for ALA synthesis the formation of a complex of the glu-tRNA synthetase and glutamyl-tRNA dehydrogenase (*hemA* protein) may accelerate the synthesis of ALA and provide efficient regulation of ALA and even heme synthesis.

Distribution of C_5 and C_4 pathways The distribution of the two pathways has evolutionary ramifications. The C_5 pathway is widely distributed in all of eubacterial subdivisions examined to date, and also appears in the archaebacterial species so far examined. The C_4 pathway appears to operate in eukaryotes except plants, the α subgroup of the purple bacteria, and a few other eubacterial genera. It is noteworthy that the subgroup of the purple bacteria has been proposed as the progenitor of mitochondria (Yang et al., 1985). The presence of the C_5 pathway in archaebacteria and in *Chloroflexus*, which appears to be the descendant of the earliest chlorophyll-containing organisms (Woese, 1987), indicates that the C_5 pathway is the older one and is present in organisms in which ALA S has not evolved or been acquired.

4. Regulation of heme biosynthetic pathway

ALA synthesis in livers of human and animals is a rate-limiting step (Kapas, 1983), and regulation is achieved by the inhibition by heme of transcription and translocation of ALA S into mitochondria. In developing erythroid cells, heme either increases or does not affect ALA S activity.

E. coli does not utilize the same pathway to generate ALA, and the regulation of ALA synthesis is undoubtedly different. There is no clear answer as to whether heme regulates the pathway in *E. coli*. In this study, heme had no obvious effect on the synthesis of ALA in crude extracts (Table 15). Heme also did not affect ALA D, PBG D and ferrochelatase activities in JL1002 when it was grown with hemin (Table 6). Moreover, hemin did not affect the synthesis of HemA, HemB and HemC proteins in an *in vitro* coupled transcription-translation system (Umanoff et al., 1988; Umanoff, unpublished data).

Javor found that various thiols (thioglycerol, dithiothreitol and cysteine) stimulated wild type *E. coli* to secrete porphyrins, riboflavin and other pigments (1985). How thiols are able to achieve this cellular response is not known. One explanation might be that excess thiols may tie up iron in the cell, thus resulting in the secretion of porphyrins. Javor and Kim (1989) isolated some *E. coli* mutants which do not respond to thiols. These genes do not seem to code for any enzyme of the heme biosynthetic pathway, since the aerobic growth rates of these mutants were comparable to that of their parent strains, but they may influence the rate of porphyrin production under favorable circumstances. Thus these genes may have some regulatory role on the heme pathway. The nature of these genes needs further investigation.

What is controlling ALA synthesis of *E. coli*? Introduction of the *hemA* multi-copy gene into HU227 and SHSP19 causes overproduction of ALA and porphyrins. The *hemA* gene appears to encode the rate-limiting step, at least, for ALA synthesis, and maybe for the entire heme pathway. Overproduction of ALA results in overproduction of porphyrins, which are detectable as pink fluorescence in UV light. The apparent

overproduction may be simply the consequence of loss of regulation of ALA synthesis in these strains because of the high dosage of the gene.

One question is what controls the activity of the HemA protein? the following observation may give a clue. When strain JL1268 was grown with very vigorous shaking in a large volume flask, cells grew very fast, but their extracts had low activity for ALA synthesis. Therefore, in order to obtain active extracts, cells were grown in ten 100 ml portions with slower shaking. Since vigorous shaking provides more oxygen a question to be answered is whether oxygen inhibits the synthesis or activity of the HemA protein. The procedure used here for detection of ALA synthesis by a crude extract is similar to the method of Oh-hama et al. (1988), which was used to study ALA synthesis in *Clostridium thermaceticum* (a strict anaerobe) and which was performed under anaerobic conditions. In yeast, oxygen and heme have been shown to be the control effecters of several respiratory genes which are either aerobically or anaerobically expressed. However, the effect of oxygen on ALA synthesis in *E. coli* is not clear and needs more investigation. Other possible factors which may control ALA synthesis are growth rate and accumulation of metabolites.

One documented mechanism of control occurs at the third step of heme synthesis. This is control of PBG D activity by the availability of PBG. However, excess PBG does not increase PBG D activity because JL1002 has only normal level of PBG D activity although it has much higher ALA D activity and thus overproduces PBG (Table 6). Therefore, this kind of regulation may not be the main regulating mechanism in *E. coli.* This is coincident with the fact that di-PBG is the cofactor of PBG D. When all PBG D molecules have been saturated more PBG will not increase its activity. PBG D may be the control enzyme in erythroid cells.

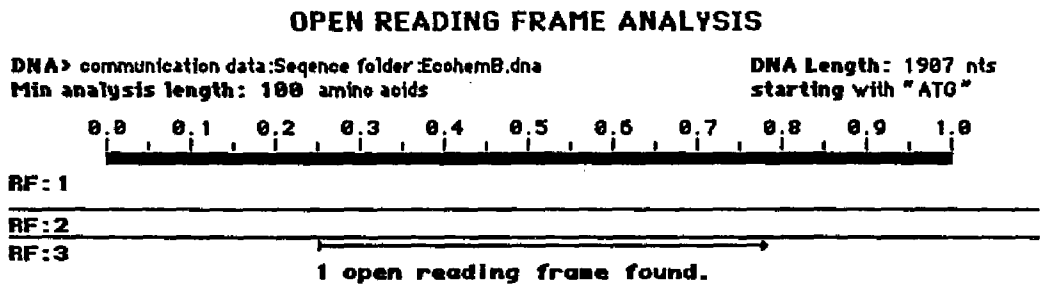
Alternative transcription and splicing of the PBG D gene has been shown for human erythroid cells. The same gene generates two isomers of PBG D. One is found in all cells acting as a housekeeping protein, the other is erythroid-specific. Other evidence also suggests that the synthesis of PBG D is a controlling step in erythroid tissue and that ALA synthetase is the controlling step in liver.

hemC, hemD, probably *hemG* and another gene form an operon (Alefounder et al., 1988). The genes are controlled under one promoter which is located at the 5' of *hemC.* This operon is responsible for the production of uroporphyrinogen III.

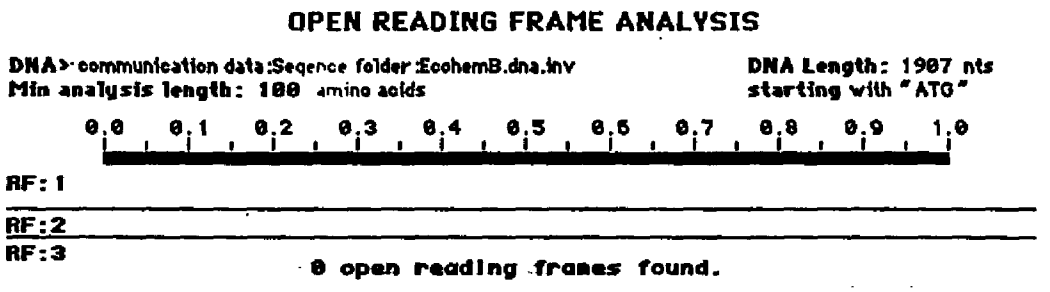
It appears that the regulation of the heme pathway in *E. coli* is not simple. All the *hem* genes may constitute a regulon, but no regulatory effecter has been found that can turn all the genes on or off. Regulation may occur at several different levels and steps.

Appendix A. Identification of *hemB* coding region by open reading frame search

a. Forward search:

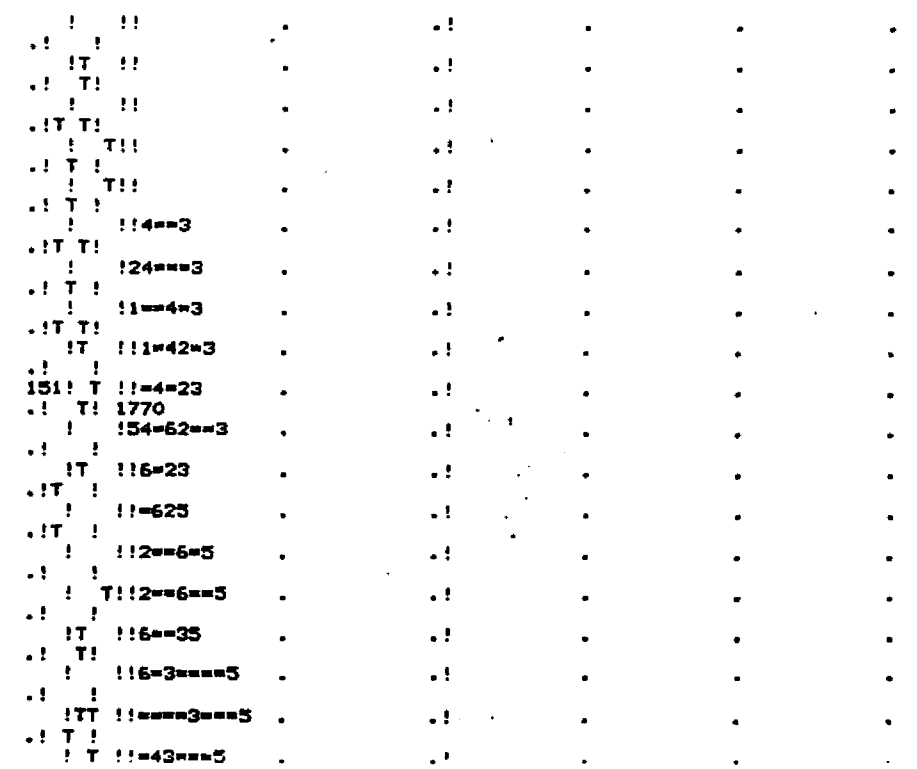


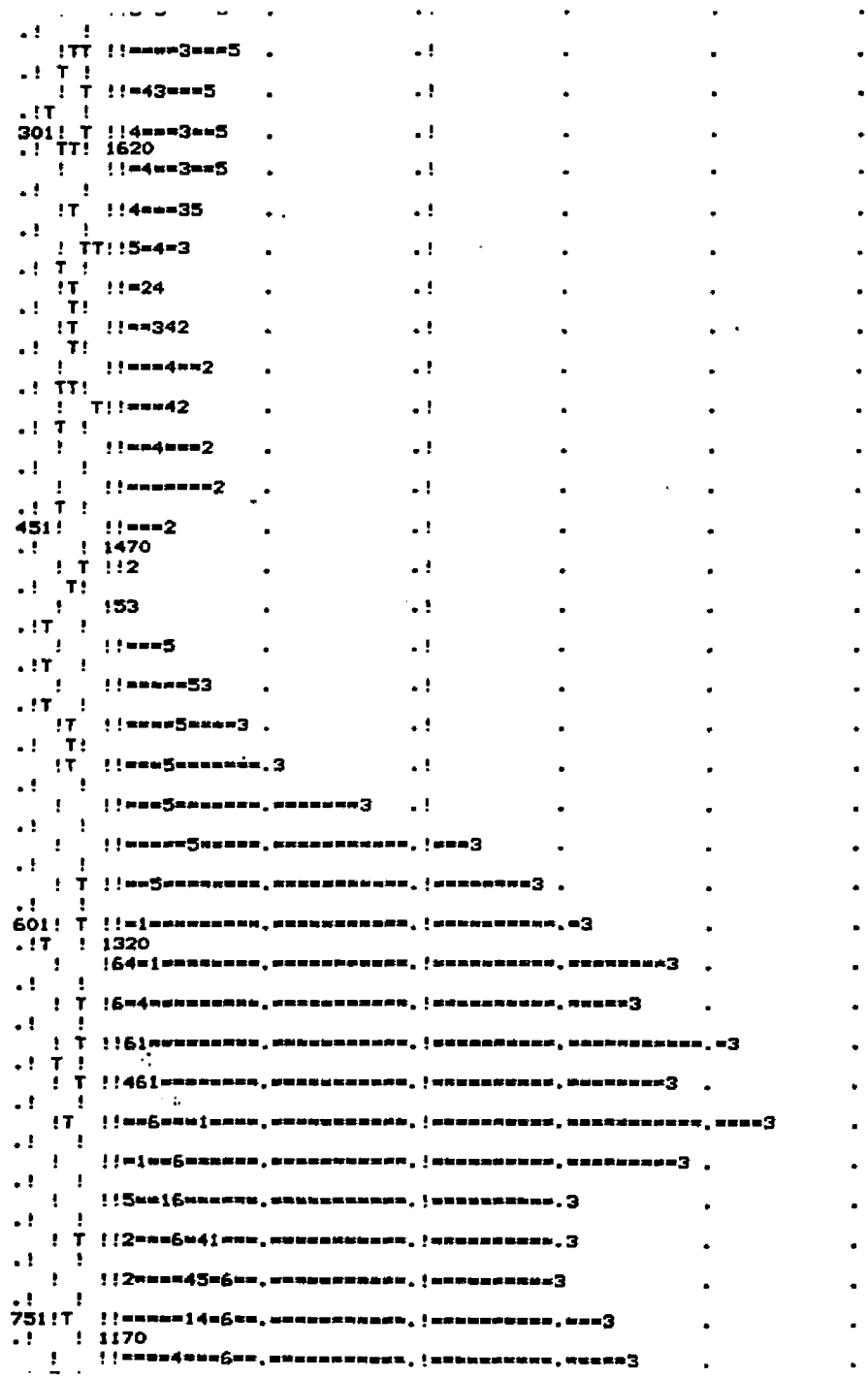
b. Inverse search:

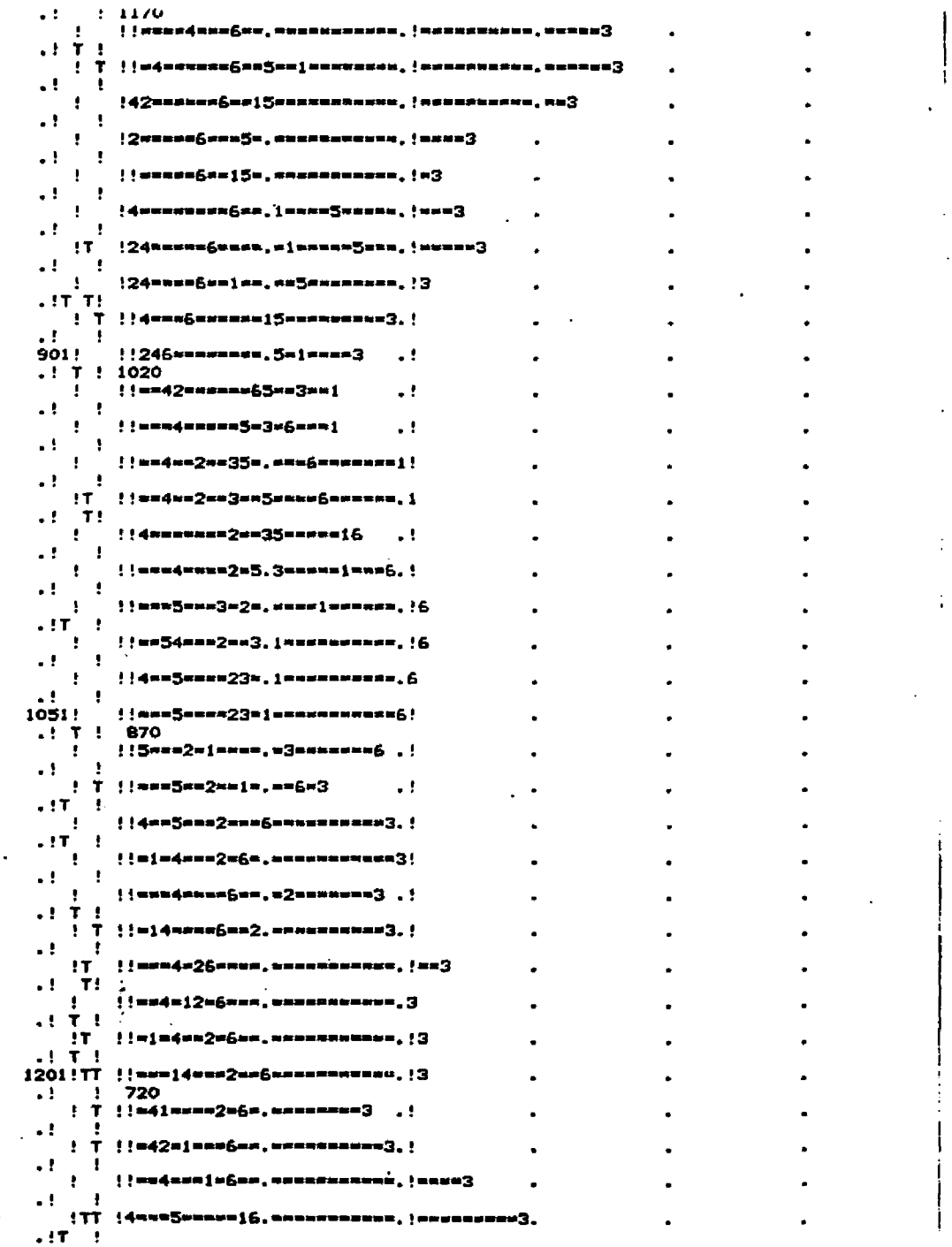


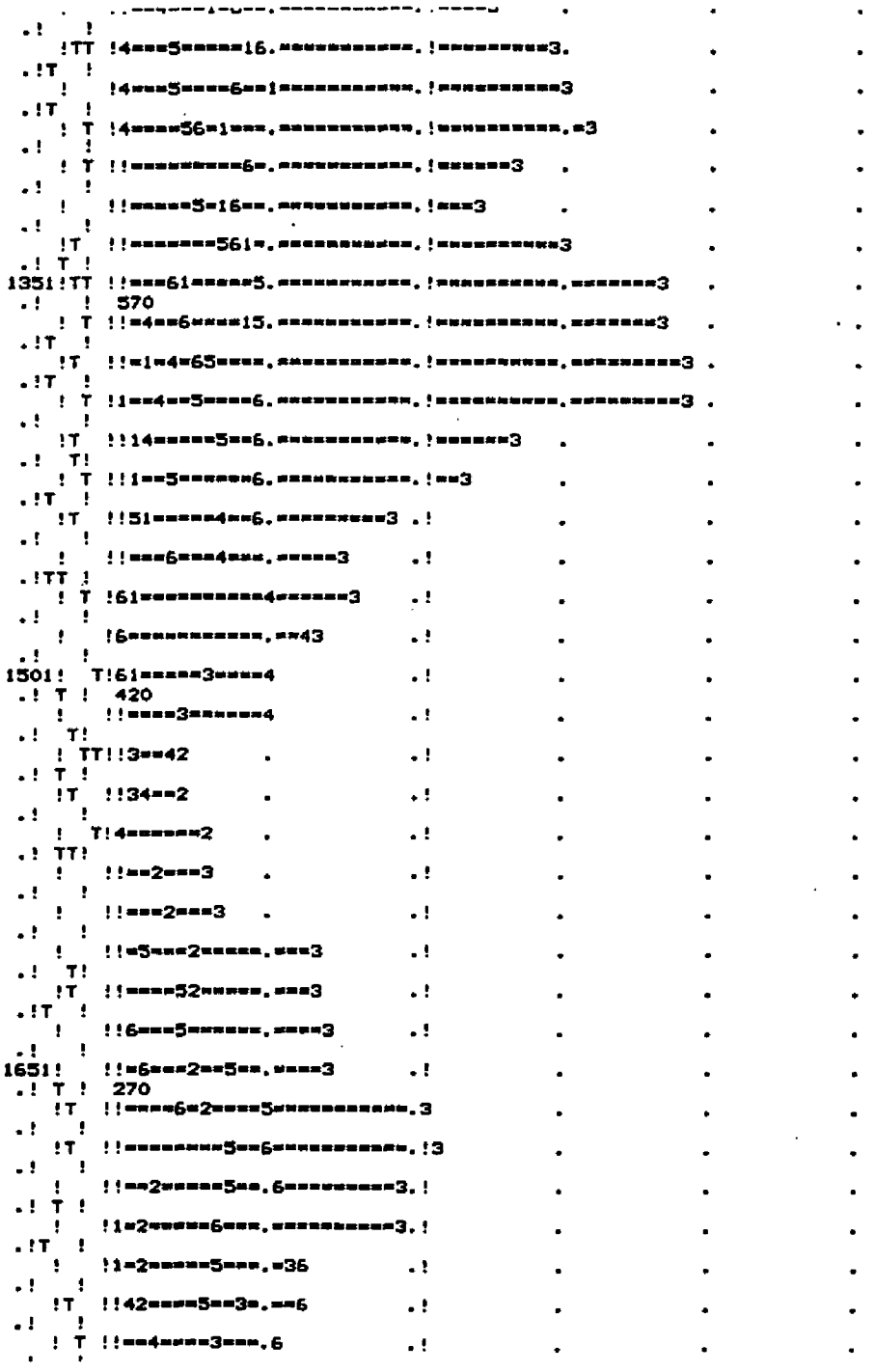
Appendix B. Protein coding region of the insert of pJL2 located by IBI protein coding region locator program

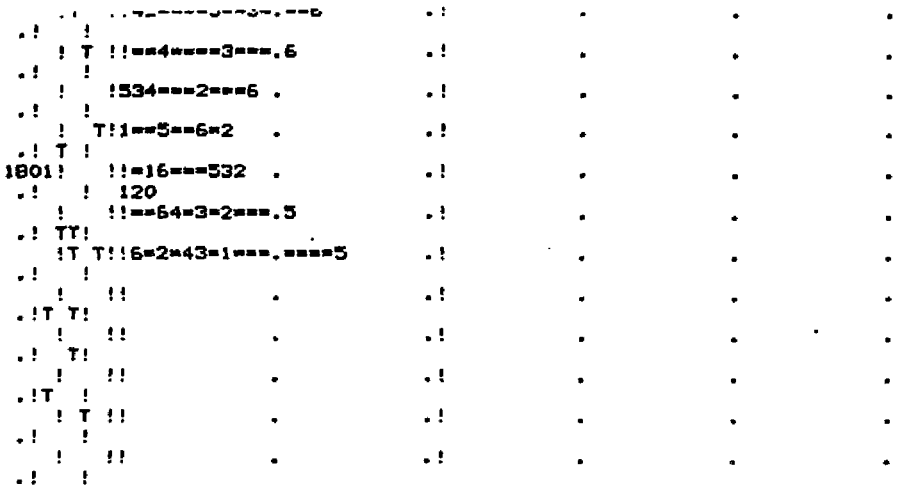
```
P U S T E L L  S E Q U E N C E  A N A L Y S I S  P R O G R A M S
International Biotechnologies, Inc.

PROTEIN CODING REGION LOCATOR  Version 2.0

Coding Prediction for BEFRAG7
From base      1 to base  1907
Compression=  5  Codon bias table LOWBIAS      used.
Values strand adjusted
Products over a range of 50 codons

    !123!R          1.5          2.M          2.5          3.0          3.5
  !456!
```











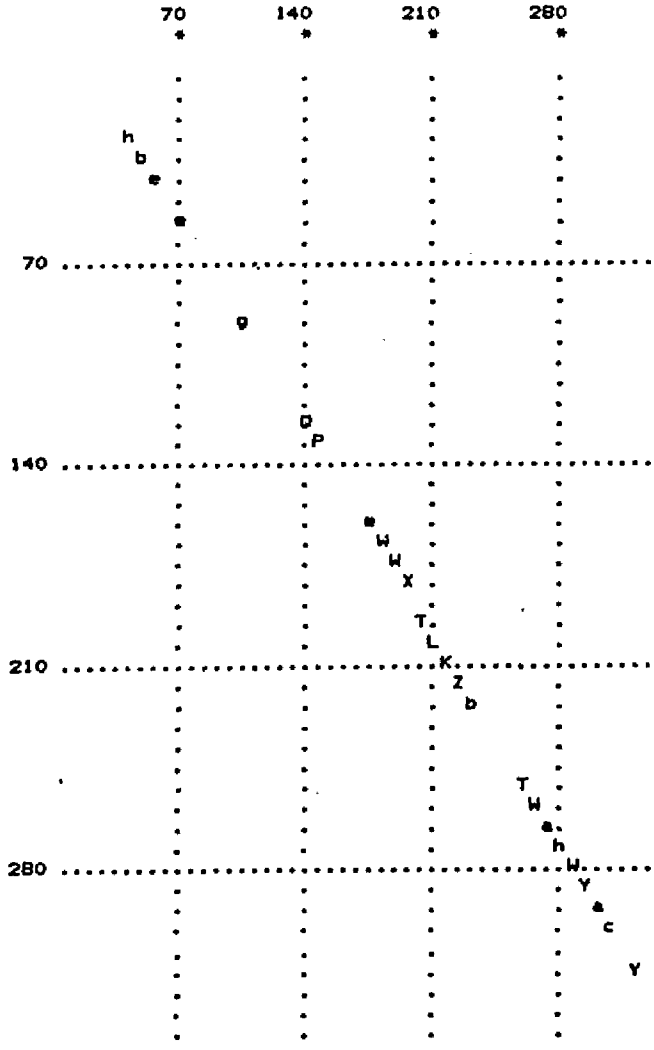
Appendix C. Homology matrix search of *E. coli* ALA D against human liver, rat liver and yeast ALA Ds

PUSTELL SEQUENCE ANALYSIS PROGRAMS
International Biotechnologies, Inc.

FORWARD HASH MATRIX Version 2.3

X AXIS=ALADECLI from A.A. no. 1 to A.A. no. 335
Y AXIS=ALADMAN from A.A. no. 1 to A.A. no. 331

Range=10 Scale= .90 Hash level= 2 Jump level= 1 Step= 1
Minimum value plotted= 35 Compressed 7 times

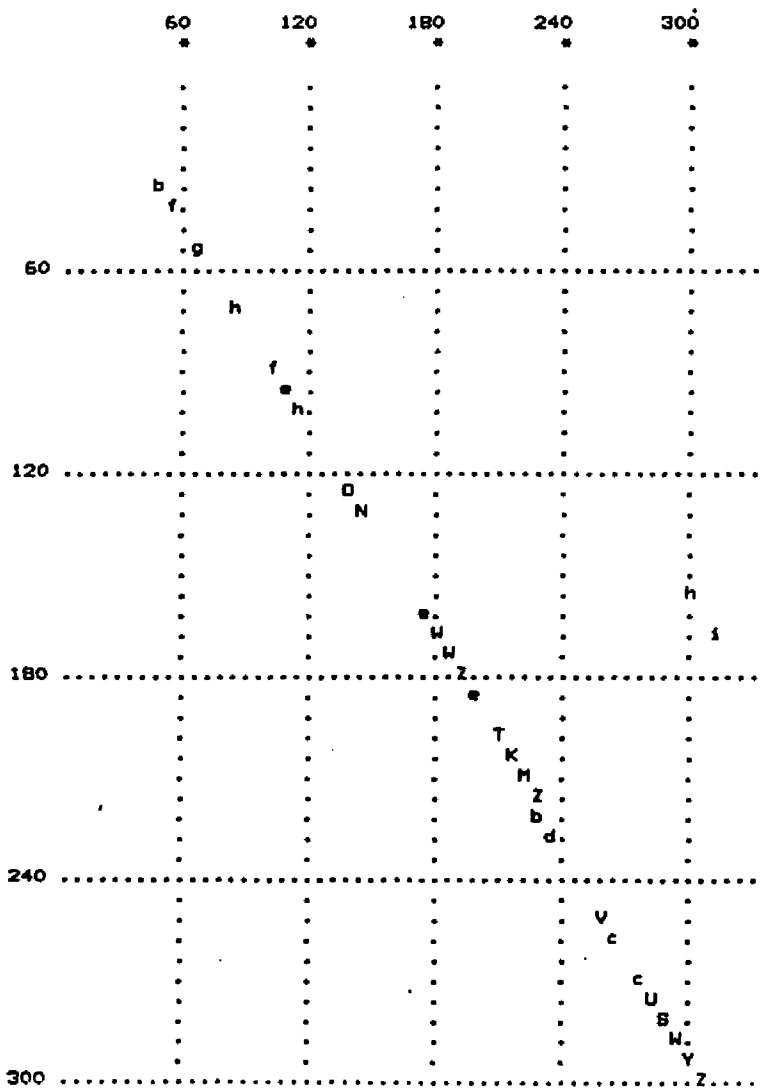


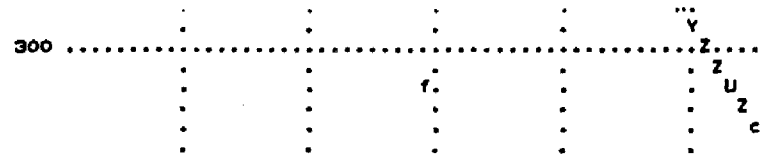
PUSTELL SEQUENCE ANALYSIS PROGRAMS
International Biotechnologies, Inc.

FORWARD HASH MATRIX Version 2.3

X AXIS=ALADECLI from A.A. no. 1 to A.A. no. 335
Y AXIS=RATALAD.PEP from A.A. no. 1 to A.A. no. 330

Range=10 Scale= .90 Hash level= 2 Jump level= 1 Step= 1
Minimum value plotted= 33 Compressed 6 times



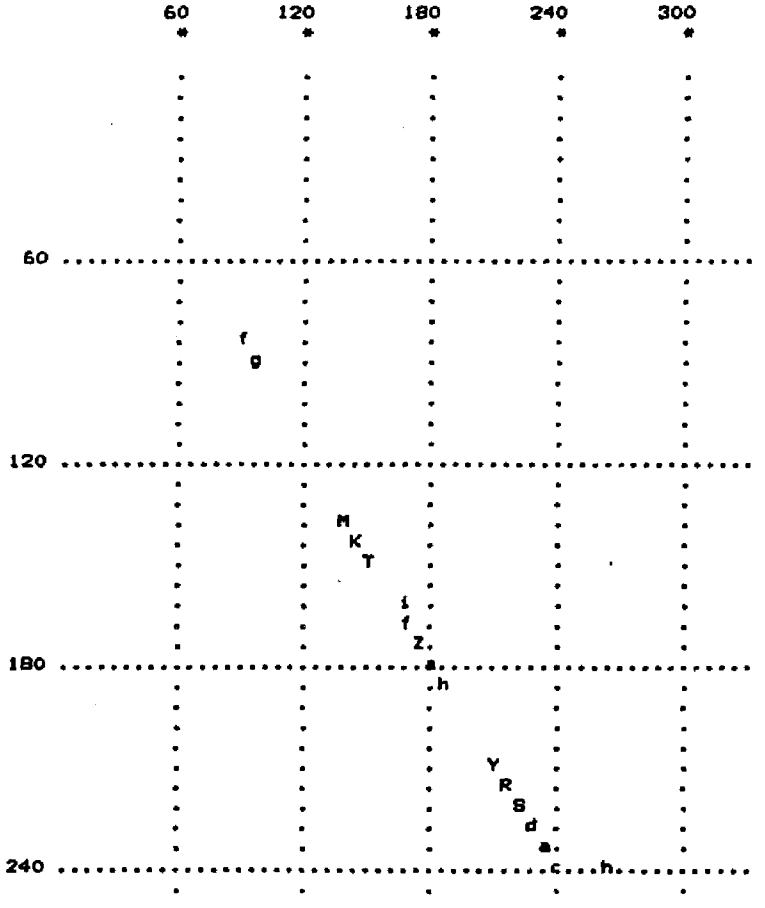


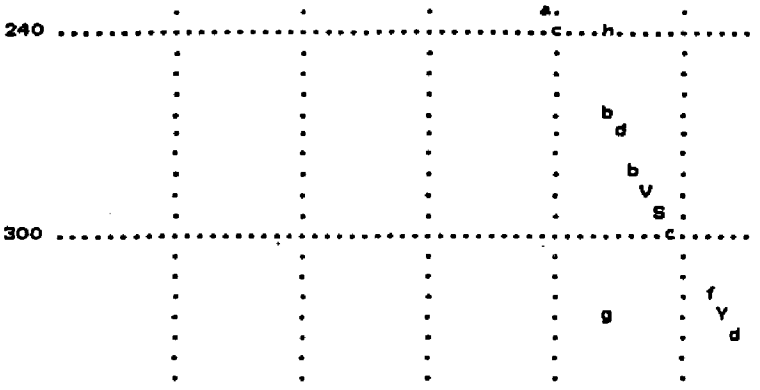
P U S T E L L S E Q U E N C E A N A L Y S I S P R O G R A M S
International Biotechnologies, Inc.

FORWARD HASH MATRIX Version 2.3

X AXIS=ALADECLI from A.A. no. 1 to A.A. no. 335
Y AXIS=ALADYEST from A.A. no. 1 to A.A. no. 342

Range=10 Scale= .90 Hash level= 2 Jump level= 1 Step= 1
Minimum value plotted= 33 Compressed 6 times




240
300
a
c
h
b
d
b
V
s
c
f
Y
d
g

Appendix D. Secondary structures of ALA Ds of human, rat and *E. coli*

```
Print detailed analysis data in the output file? (<CR>=NO) NO
Do you want verbose commentary in the output file? (<CR>=NO) NO
^LPEP - IntelliGenetics              19-May-88 10:31 pm

<Px> = average propensity for x = a(lpha), b(eta), or t(urn).

("*" in the <Pa> column means that <Pa> was adjusted according to
Conditions II.A.5 in (Chou and Fasman 1978)).

Key:    AAAAAA  Alpha helix
        BBBBB   Beta sheet
        TTTT    Turn
```

Human ALAD

Secondary Structures predicted by Chou and Fasman algorithm:

Begin	End	Type	<Pa>	<Pb>	<Pt>
4	8	beta	1.030	1.144	
8	11	turn	.758	.960	1.270
11	18	beta	1.026	1.079	
12	20	alpha	1.079*	1.070	
13	16	turn	.998	1.005	.913
19	28	beta	1.023	1.108	
27	30	turn	.883	.840	1.303
28	33	beta	.973	1.140	
30	40	beta	.945	1.265	
40	43	turn	.913	.832	1.235
44	48	beta	.934	1.208	
50	53	turn	.853	1.075	1.043
54	57	turn	.915	.995	1.078
57	66	alpha	1.164	.944	
64	69	beta	1.080	1.138	
68	76	alpha	1.090	1.068	
72	79	beta	.993	1.269	
81	84	turn	.845	.983	1.100
84	90	alpha	1.039*	.823	
86	89	turn	1.063	.550	1.183
88	91	turn	1.017	.648	1.178
92	99	alpha	1.148	.649	
95	98	turn	1.200	.508	1.093
97	112	alpha	1.149	.936	
105	109	beta	1.096	1.200	
111	118	beta	1.013	1.143	
112	115	turn	.895	1.030	1.068
116	124	beta	1.008	1.216	
124	127	turn	.698	1.100	1.203
127	130	turn	.793	.890	1.225
129	132	turn	.818	.920	1.163
131	135	beta	.938	1.082	
136	139	turn	.880	.690	1.323
139	160	alpha	1.196	.918	
153	157	beta	1.160	1.226	
159	162	turn	.963	.878	1.105
160	166	beta	1.049	1.157	
163	172	alpha	1.091*	.981	
166	169	turn	.943	.668	1.268
170	185	alpha	1.231	.926	
170	173	turn	1.120	.848	1.055
189	194	beta	.998	1.170	
194	197	turn	.920	1.005	1.150
196	202	alpha	1.051	.964	
200	205	beta	.973	1.167	
203	206	turn	.773	1.198	1.123

206	209	turn	.813	.903	1.158
207	214	alpha	1.058*	.819	
220	223	turn	.918	.898	1.138
221	227	beta	.891	1.067	
227	230	turn	.783	.670	1.315
229	232	turn	.885	.815	1.183
230	244	alpha	1.121	.961	
239	242	turn	1.015	.928	1.093
242	252	alpha	1.171	.996	
252	255	turn	.938	.773	1.173
255	258	turn	.980	1.093	.963
257	262	beta	1.005	1.257	
258	263	alpha	1.142	1.073	
264	267	turn	1.098	.930	.995
266	269	turn	.935	.675	1.235
268	271	turn	.948	.815	1.130
270	275	alpha	1.080	1.037	
271	278	beta	1.028	1.215	
277	287	alpha	1.109	1.022	
277	280	turn	.850	1.017	1.110
284	288	beta	1.062	1.068	
289	297	alpha	1.161	.922	
293	310	alpha	1.232	.988	
312	327	beta	1.013	1.219	
322	331	alpha	1.211	.998	

```
        10         20         30         40         50         60         70
MQPQSVLHSGYFHPLLRAWQTATTTLNASNLIYPIFVTDVPDDIQPITSLPGVARYGVKRLEEMLRPLVE
           AAAAAAAAA                                          AAAAAAAAAA AAA
   BBBBB  BBBBBBBBBBBBBBBBBBBBBBBBBBBBB    BBBBB                    BBBBBB
       TTTT TTTT           TTTT          TTTT       TTTTTTTT
```

```
        80         90        100        110        120        130        140
EGLRCVLIFGVPSRVPKDERGSAADSEESPAIEAIHLLRKTFPNLLVACDVCLCPYTSHGHCGLLSENGA
AAAAAA        AAAAAAA AAAAAAAAAAAAAAAAAAAAA                              AA
 BBBBBBBB                          BBBBB BBBBBBBBBBBBBB       BBBBB
           TTTT TTTTTT   TTTT               TTTT         TTTTTTTTT   TTTT
```

```
       150        160        170        180        190        200        210
FRAEESRQRLAEVALAYAKAGCQVVAPSDMMDGRVEAIKEALMAHGLGNRVSVMSYSAKFASCFYGPFRD
AAAAAAAAAAAAAAAAAAAA  AAAAAAAAAAAAAAAAAAAAAAA          AAAAAAA    AAAA
              BBBBB  BBBBBBB                          BBBBBB     BBBBBB
                     TTTT   TTTTTTTT                       TTTT     TTTTTTT
```

```
       220        230        240        250        260        270        280
AAKSSPAFGDRRCYQLPPGARGLALRAVDRDVREGADMLMVKPGMPYLDIVREVKDKHPDLPLAVYHVSG
AAAA              AAAAAAAAAAAAAAAAAAAAAAA     AAAAAA      AAAAAA AAAA
          BBBBBBB                                BBBBBB       BBBBBBBB
         TTTT    TTTTTT       TTTT          TTTTTTT     TTTTTTTT     TTTT
```

```
       290        300        310        320        330
EFAMLWHGAQAGAFDLKAAVLEAMTAFRRAGADIIITYYTPQLLQWLKEE.
AAAAAAA AAAAAAAAAAAAAAAAAAAAA           AAAAAAAAAA
   BBBBB                          BBBBBBBBBBBBBBBBB
```

Rat ALA D

Secondary Structures predicted by Chou and Fasman algorithm:

Begin	End	Type	<Pa>	<Pb>	<Pt>
1	6	alpha	1.065	1.057	
1	7	beta	1.086	1.091	
8	11	turn	.758	.960	1.270
11	18	beta	1.026	1.079	
12	20	alpha	1.079*	1.070	
13	16	turn	.998	1.005	.913
18	23	beta	.973	1.038	
22	25	turn	.750	.920	1.218
25	32	beta	.984	1.181	
28	40	beta	.972	1.226	
40	43	turn	.913	.832	1.235
44	48	beta	1.048	1.156	
45	50	alpha	1.099*	1.022	
50	53	turn	.853	1.075	1.043
53	61	beta	.974	1.186	
54	57	turn	.915	.995	1.078
58	66	alpha	1.190	1.001	
64	69	beta	1.080	1.138	
68	75	alpha	1.083	1.046	
71	79	beta	1.040	1.220	
81	84	turn	.845	.983	1.100
84	90	alpha	1.057*	.847	
86	89	turn	1.063	.550	1.183
89	92	turn	.990	.743	1.178
92	99	alpha	1.085	.670	
96	99	turn	1.015	.603	1.265
97	111	alpha	1.075	.953	
99	102	turn	.813	1.023	1.095
104	109	beta	1.143	1.165	
111	119	beta	.996	1.181	
117	124	beta	.983	1.205	
124	127	turn	.698	1.100	1.203
127	130	turn	.793	.890	1.225
129	132	turn	.818	.920	1.163
131	135	beta	.938	1.082	
136	139	turn	.880	.690	1.323
139	160	alpha	1.207	.935	
153	157	beta	1.160	1.226	
159	162	turn	.963	.878	1.105
160	166	beta	1.049	1.157	
163	172	alpha	1.091*	.981	
166	169	turn	.943	.668	1.268
170	185	alpha	1.194	.964	
170	173	turn	1.120	.848	1.055
189	194	beta	.998	1.170	
194	197	turn	.920	1.005	1.150
196	202	alpha	1.051	.964	
200	205	beta	.973	1.167	
203	206	turn	.773	1.198	1.123
206	209	turn	.813	.903	1.158
207	214	alpha	1.051*	.864	
220	223	turn	.918	.898	1.138
221	227	beta	.891	1.067	
227	230	turn	.783	.670	1.315
229	232	turn	.885	.815	1.183
230	244	alpha	1.159	.985	
242	252	alpha	1.151	1.053	
247	251	beta	1.162	1.238	
252	255	turn	.878	.835	1.170
255	258	turn	.920	1.155	.960
257	262	beta	1.088	1.193	

258	268	alpha	1.158	.968	
264	267	turn	1.098	.930	.995
266	269	turn	.935	.675	1.235
268	271	turn	1.073	.773	.950
270	277	alpha	1.098	1.078	
271	278	beta	1.041	1.244	
277	287	alpha	1.119	1.043	
278	281	turn	.978	.893	1.058
284	288	beta	1.062	1.068	
289	297	alpha	1.147	.903	
293	310	alpha	1.153	1.014	
294	298	beta	1.032	1.068	
312	320	beta	1.082	1.227	
319	331	alpha	1.175	.952	
322	327	beta	1.163	1.185	

```
        10         20         30         40         50         60         70
MHHQSVLHSGYFHPLLRAWQTTPSTVSATNLIYPIFVTDVPDDVQPIASLPGVARYGVNQLEEMLRPLVE
AAAAAA      AAAAAAAAA                              AAAAAA        AAAAAAAAA AAA
BBBBBBB    BBBBBBBBBBBBB BBBBBBBBBBBBBBBB    BBBBB     BBBBBBBBB  BBBBBB
       TTTT TTTT      TTTT                TTTT       TTTTTTTT

        80         90        100        110        120        130        140
AGLRCVLIFGVPSRVPKDEQGSAADSEDSPTIEAVRLLRKTFPTLLVACDVCLCPYTSHGHCGLLSENGA
AAAAA         AAAAAAA AAAAAAAAAAAAAAAAAAAA                                AA
BBBBBBBBB                        BBBBBB BBBBBBBBBBBBBB       BBBBB
          TTTT TTTTTTT   TTTTTTT                          TTTTTTTTT   TTTT

       150        160        170        180        190        200        210
FLAEESRQRLAEVALAYAKAGCQVVAPSDMMDGRVEAIKAALLKHGLGNRVSVMSYSAKFASCFYGPFRD
AAAAAAAAAAAAAAAAAAA  AAAAAAAAAAAAAAAAAAAAAA          AAAAAAA    AAAA
            BBBBB  BBBBBBB                          BBBBBB     BBBBBB
                   TTTT   TTTTTTTT                     TTTT     TTTTTTT

       220        230        240        250        260        270        280
AAQSSPAFGDRRCYQLPPGARGLALRAVARDIQEGADILMVKPGLPYLDMVQEVKDKHPELPLAVYQVSG
AAAA                AAAAAAAAAAAAAAAAAAAAAA     AAAAAAAAAAA AAAAAAAAAAA
          BBBBBBB                       BBBBB     BBBBBB         BBBBBBBB
         TTTT   TTTTTT                      TTTTTTT     TTTTTTTT      TTT

       290        300        310        320        330
EFAMLWHGAKAGAFDLRTAVLESMTAFRRAGADIIITYFAPQLLKWLKEE.
AAAAAAA AAAAAAAAAAAAAAAAAAAAA         AAAAAAAAAAAAA
   BBBBB      BBBBB              BBBBBBBBB BBBBBB
T
```

E.coli ALA D

Secondary Structures predicted by Chou and Fasman algorithm:

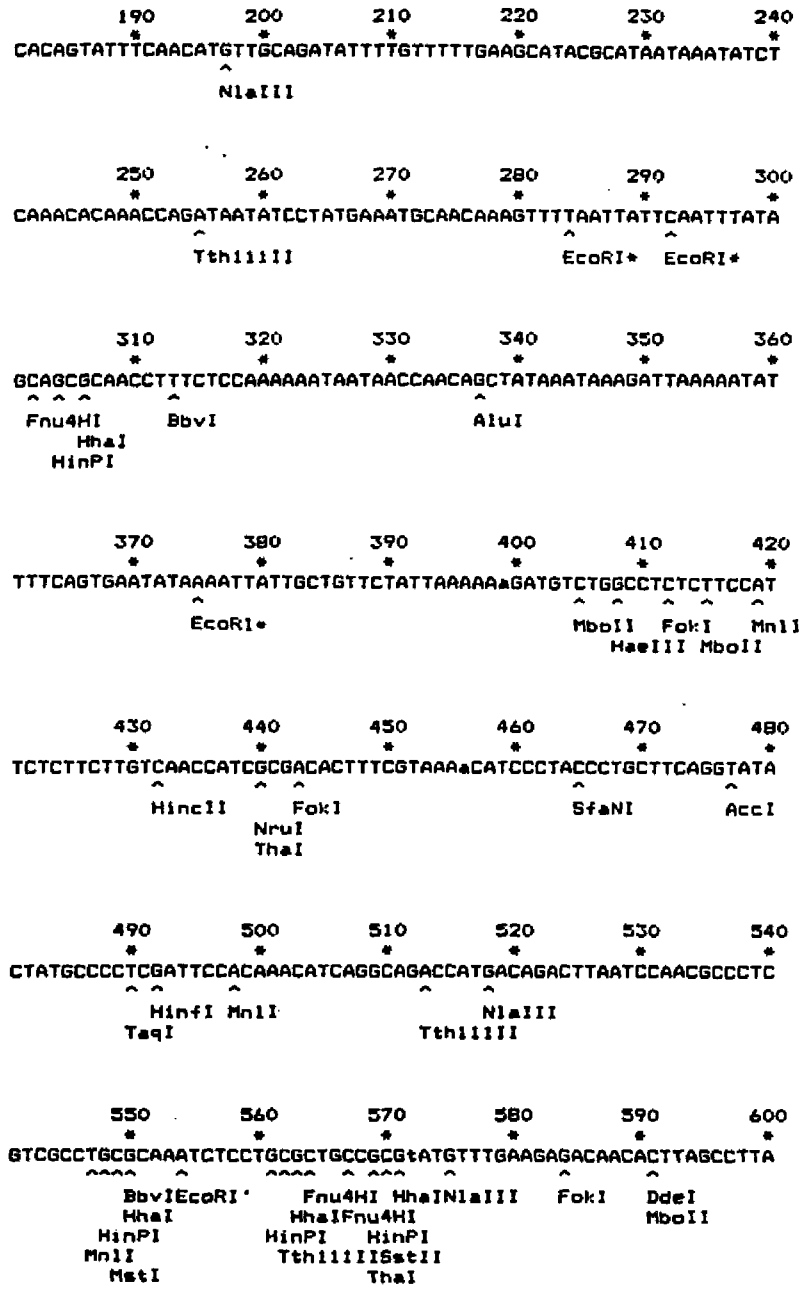
Begin	End	Type	<Pa>	<Pb>	<Pt>
5	8	turn	.838	1.108	1.105
6	13	beta	.973	1.143	
10	18	beta	1.068	1.111	
18	21	turn	.878	.835	1.093
27	39	alpha	1.125	.962	
35	40	beta	.920	1.103	
41	54	alpha	1.140	.990	
42	49	beta	1.104	1.238	
52	61	alpha	1.181	.967	
53	56	turn	.968	.823	1.268
61	64	turn	.913	1.013	1.045
63	67	beta	1.028	1.206	
64	80	alpha	1.176*	.975	
81	84	turn	.935	1.017	1.063
84	90	beta	1.043	1.229	
95	98	turn	1.088	.743	1.028
98	101	turn	.920	.765	1.173
100	103	turn	.943	.718	1.278
101	107	alpha	1.111*	.761	
105	114	alpha	1.096	.915	
105	108	turn	1.018	.648	1.178
116	122	beta	.930	1.153	
118	125	alpha	1.096	1.038	
124	128	beta	1.162	1.230	
127	130	turn	1.015	.883	1.113
129	133	beta	.874	1.098	
129	132	turn	.918	1.075	1.052
136	139	turn	.793	.890	1.225
138	141	turn	.818	.920	1.163
140	145	beta	.873	1.167	
145	148	turn	.945	.795	1.110
150	153	turn	1.028	.700	1.285
152	158	alpha	1.123	.917	
160	168	alpha	1.277	1.043	
160	164	beta	1.162	1.214	
166	174	alpha	1.210	.936	
172	180	alpha	1.141*	.929	
174	177	turn	1.045	.740	1.068
179	182	turn	1.035	.860	1.150
181	186	beta	1.058	1.180	
182	193	alpha	1.196	1.058	
195	202	alpha	1.106	1.010	
198	203	beta	1.040	1.148	
201	204	turn	.920	1.005	1.150
205	211	alpha	1.030	1.003	
209	212	turn	.840	1.088	1.150
213	216	turn	.813	.903	1.158
214	219	alpha	1.172	.815	
222	229	alpha	1.035*	.823	
223	226	turn	.988	.832	1.155
227	230	turn	.900	.973	1.133
229	233	beta	.938	1.052	
234	237	turn	.918	.855	1.158
237	243	alpha	1.213	.811	
241	251	alpha	1.158	.951	
251	254	turn	.713	.758	1.433
253	258	alpha	1.098	1.087	
253	257	beta	1.086	1.156	
262	268	beta	1.016	1.246	
266	276	alpha	1.169	.999	

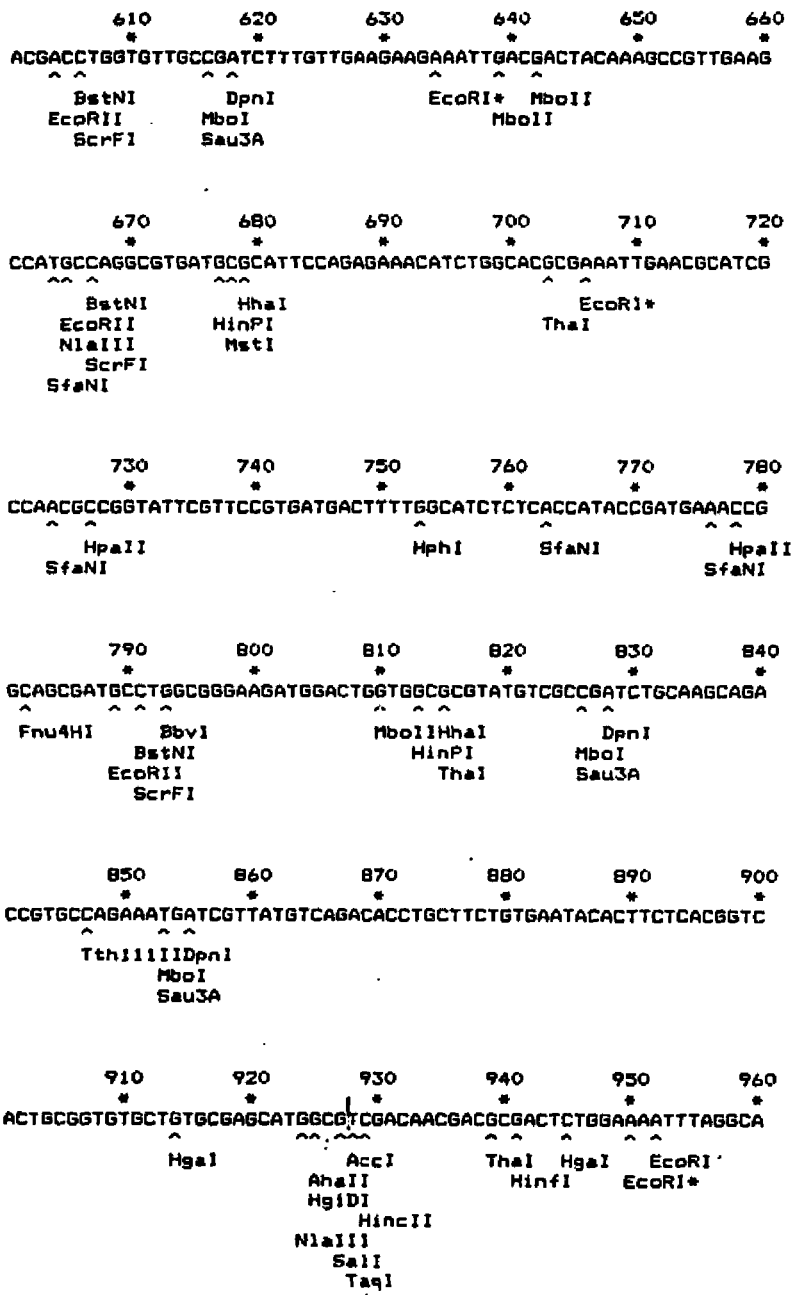
278	283	beta	.988	1.242	
283	286	turn	.978	.893	1.058
286	296	alpha	1.265	1.021	
288	309	alpha	1.218	.987	
311	320	alpha	1.070	.987	
319	326	beta	1.080	1.251	
324	335	alpha	1.208	.988	

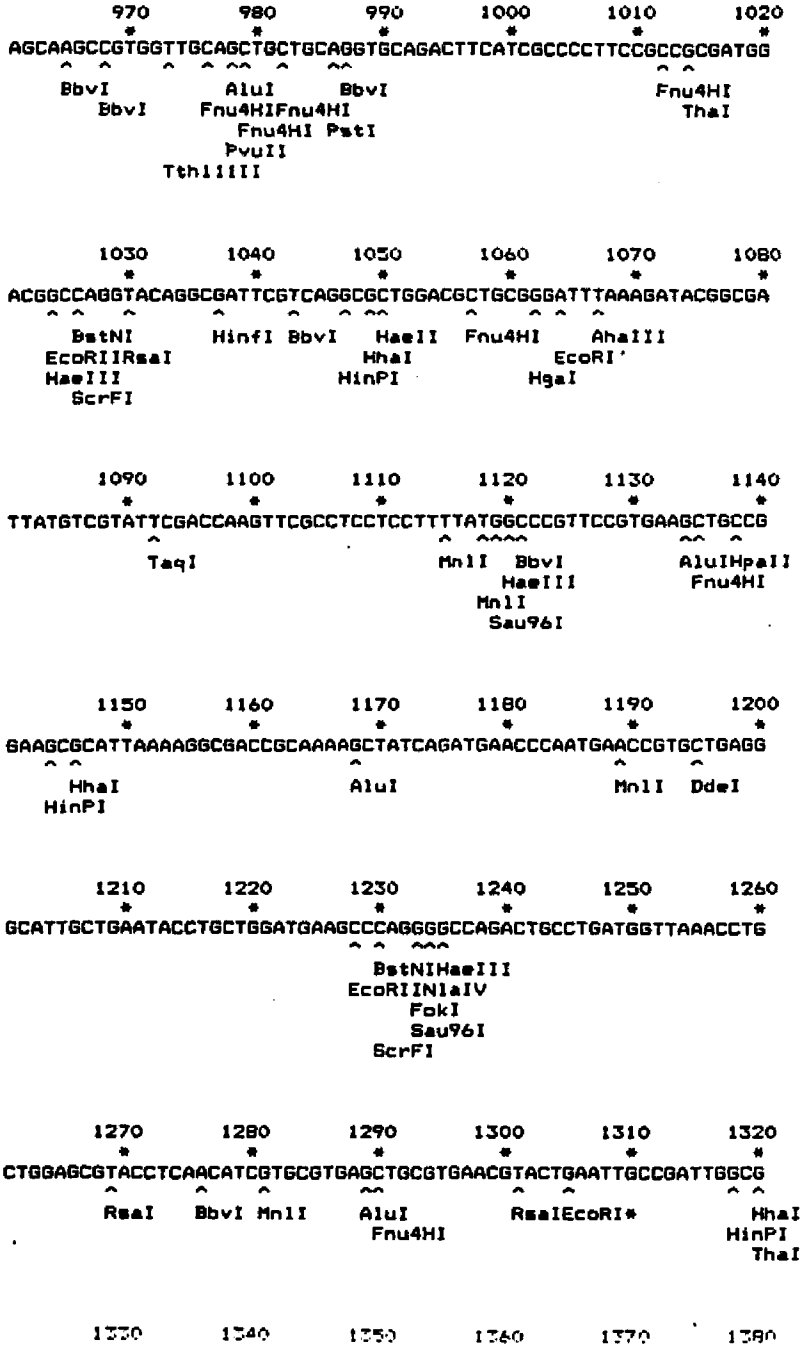
```
        10        20        30        40        50        60        70
MPLDSTNIRQTMTDLIQRPRRLRKSPALPRMFEETTLSLNDLVLPIFVEEEIDDYKAVEAMPGVMRIPEK
                              AAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAA  AAAAAAA
     BBBBBBBBBBBBB                  BBBBBB BBBBBBBB              BBBBB
    TTTT         TTTT                                  TTTT    TTTT

        80        90       100       110       120       130       140
HLAREIERIANAGIRSVMTFGISHHTDETGSDAWREDGLVARMSPICKQTVPEMIVMSDTCFCEYTSHGH
AAAAAAAAAA                    AAAAAAAAAAAAAA   AAAAAAAA
             BBBBBBB                          BBBBBBB BBBBBBBBBB      B
          TTTT           TTTTTTTTT TTTT                  TTTTTT   TTTTT

       150       160       170       180       190       200       210
CGVLCEHGVDNDATLENLGKQAVVAAAAGADFIAPSAAMDGQVQAIRQALDAAGFKDTAIMSYSTKFASS
           AAAAAAA AAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAA AAAAAAAA  AAAAAA
BBBBB              BBBBB                   BBBBBB            BBBBBB
T   TTTT TTTT                       TTTT TTTT                    TTTT    TT

       220       230       240       250       260       270       280
FYGPFREAAGSALKGDRKSYQMNPMNRAEGIAEYLLDEAQGPDCLMVKPAGAYLNIVRELRERTELPIGA
A  AAAAAA  AAAAAAAA       AAAAAAAAAAAAAAAA AAAAAA       AAAAAAAAAAA
                  BBBBB                      BBBBB    BBBBBBB          BBB
TTTTTT       TTTTTTTT   TTTT              TTTT

       290       300       310       320       330
YQVSGEYAMIKFAALAGAIDEEKVVLESLGSIKRAGADLIFSYFALDLAEKKILR
    AAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAA   AAAAAAAAAAAA
BBB                                  BBBBBBBB
  TTTT
```

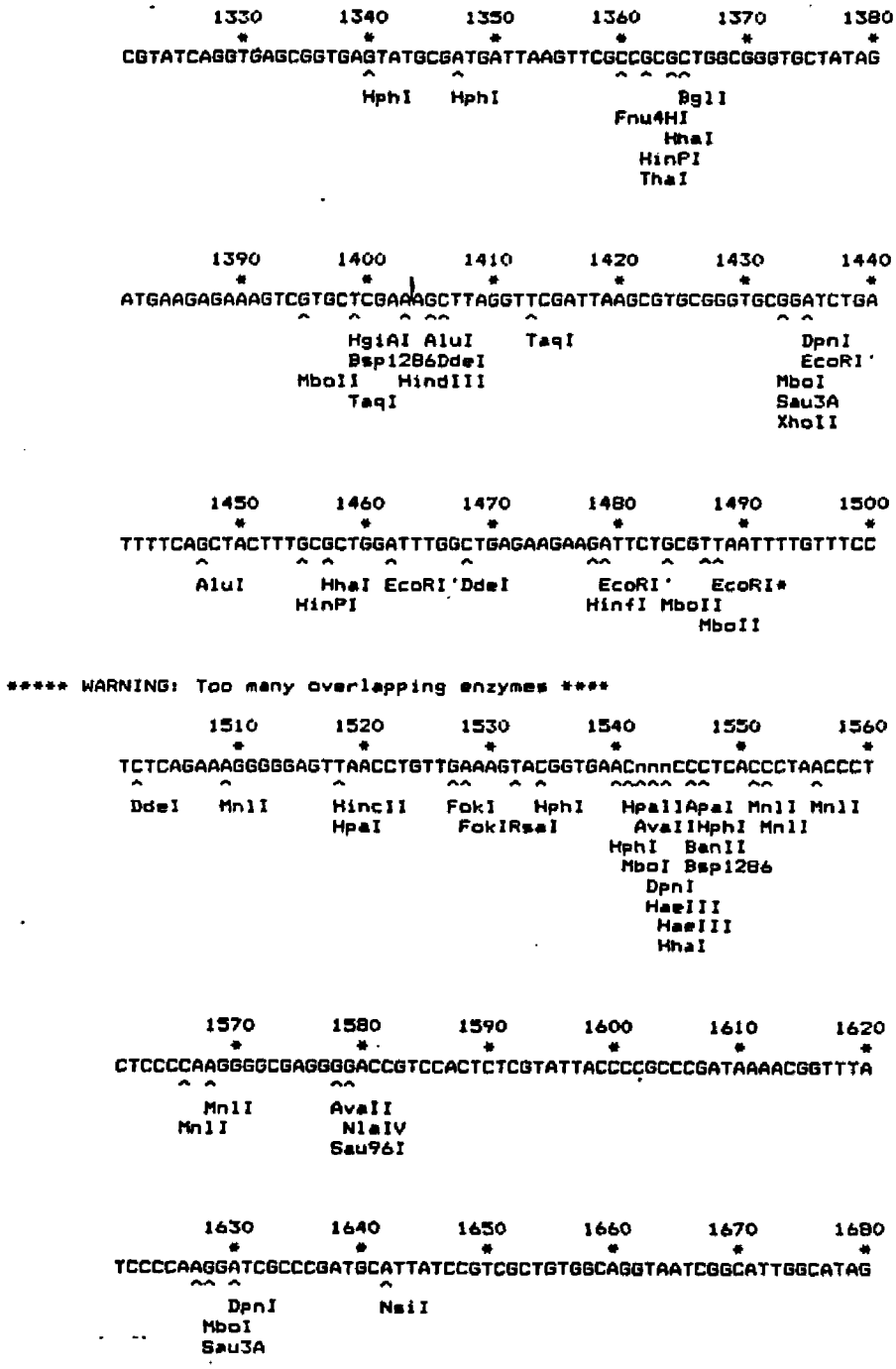
Appendix E. Restriction map of the insert of pJL2

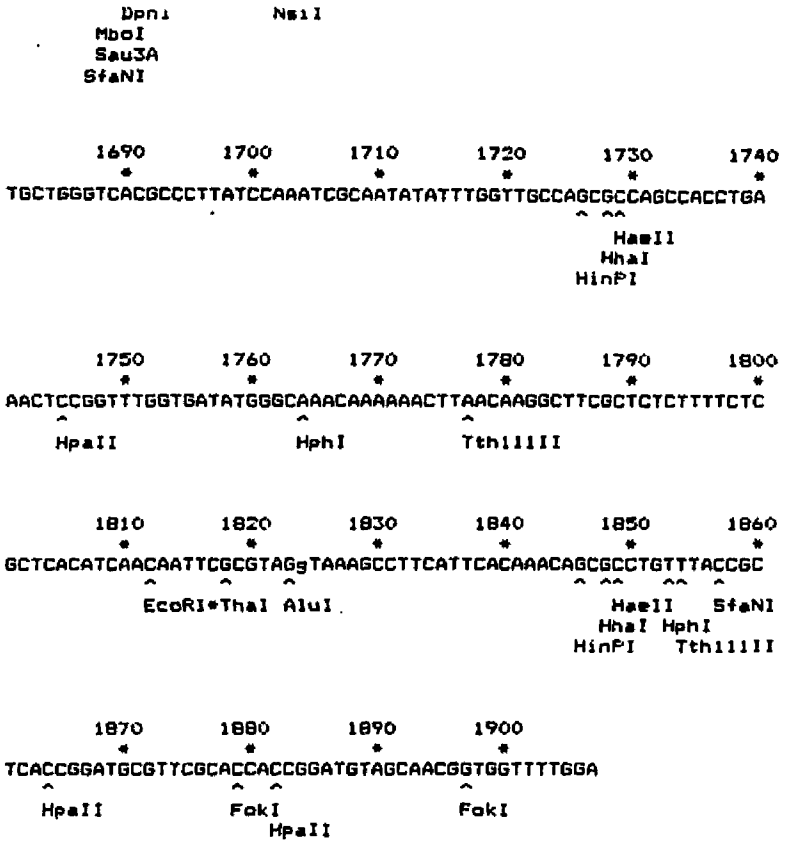
```
P U S T E L L  S E Q U E N C E  A N A L Y S I S  P R O G R A M S
International Biotechnologies, Inc.
-------------------------------------------------------------
RESTRICTION SITE SEARCH PROGRAM          Version 5.1

RESTRICTION SITES OF BEFRAG7

from base no.     1 to base no. 1907
Positions numbered from base no.     1
All enzymes listed are commercially available
WARNING: There are ambiguous bases in the following positions:
      127,  399,  456,  572, 1542, 1543, 1544, 1824,
    Restriction sites spanning these positions should be viewed as tenative


^ appears below base just preceeding restriction cut
If cut site unknown, mark is placed in center of site
First letter of enzyme name is below ^
- Note that the cut for many enzymes with asymmetric
recognition sequences will be distant from that sequence

          10         20         30         40         50         60
           *          *          *          *          *          *
TGCATGCCTGCAGGTCGACTCTAGAGGATCCATCAATTATCAATGCATAGAAAACATATT
    ^     ^ ^^^^  ^     ^ ^     ^          ^
   NlaIII   AccI XbaI  BamHI   EcoRI*      NsiI
   SphI  PstIHincII      DpnI
            HinfI        EcoRI'
          MnlI         MboI
        SalI             NlaIV
         TaqI          Sau3A
                       XhoII

          70         80         90        100        110        120
           *          *          *          *          *          *
ATGTTTCGCGCTAACGATTATGCTGCGTTACAGACGAATGGTTTCAAATTATTCTTTCCC
   ^^^           ^                    ^    ^          ^
    BbvI         Fnu4HI               XmnI EcoRI*     DdeI
     HhaI
   HinPI
   ThaI

          130        140        150        160        170        180
           *          *          *          *          *          *
TAAGTCcGAATGTGGACATATATAATTTATGGAGAAAAACACGCATCAAATTCATAAAAG
                     ^                              ^ ^ ^
                     EcoRI*                          EcoRI'
                                                  EcoRI*
                                                     SfaNI

          190        200        210        220        230        240
```


190 200 210 220 230 240
CACAGTATTTCAACATGTTGCAGATATTTTGTTTTTGAAGCATACGCATAATAAATATCT
NlaIII
250 260 270 280 290 300
CAAACACAAACCAGATAATATCCTATGAAATGCAACAAAGTTTTAATTATTCAATTTATA
TthIIIII
EcoRI* EcoRI*
310 320 330 340 350 360
GCAGCGCAACCTTTCTCCAAAAAATAATAACCAACAGCTATAAATAAAGATTAAAAATAT
Fnu4HI BbvI AluI
HhaI
HinPI
370 380 390 400 410 420
TTTCAGTGAATATAAAATTATTGCTGTTCTATTAAAAAaGATGTCTGGCCTCTCTTCCAT
EcoR1*
MboII FokI MnlI
HaeIII MboII
430 440 450 460 470 480
TCTCTTCTTGTCAACCATCGCGACACTTTCGTAAAaCATCCCTACCCTGCTTCAGGTATA
HincII FokI
NruI
ThaI
SfaNI AccI
490 500 510 520 530 540
CTATGCCCCTCGATTCCACAAACATCAGGCAGACCATGACAGACTTAATCCAACGCCCTC
HinfI MnlI
TaqI
NlaIII
TthIIIII
550 560 570 580 590 600
GTCGCCTGCGCAAATCTCCTGCGCTGCCGCGtATGTTTGAAGAGACAACACTTAGCCTTA
BbvI EcoRI' Fnu4HI HhaI NlaIII FokI DdeI
HhaI HhaI Fnu4HI MboII
HinPI HinPI HinPI
MnlI TthIIIII SstII
MstI ThaI

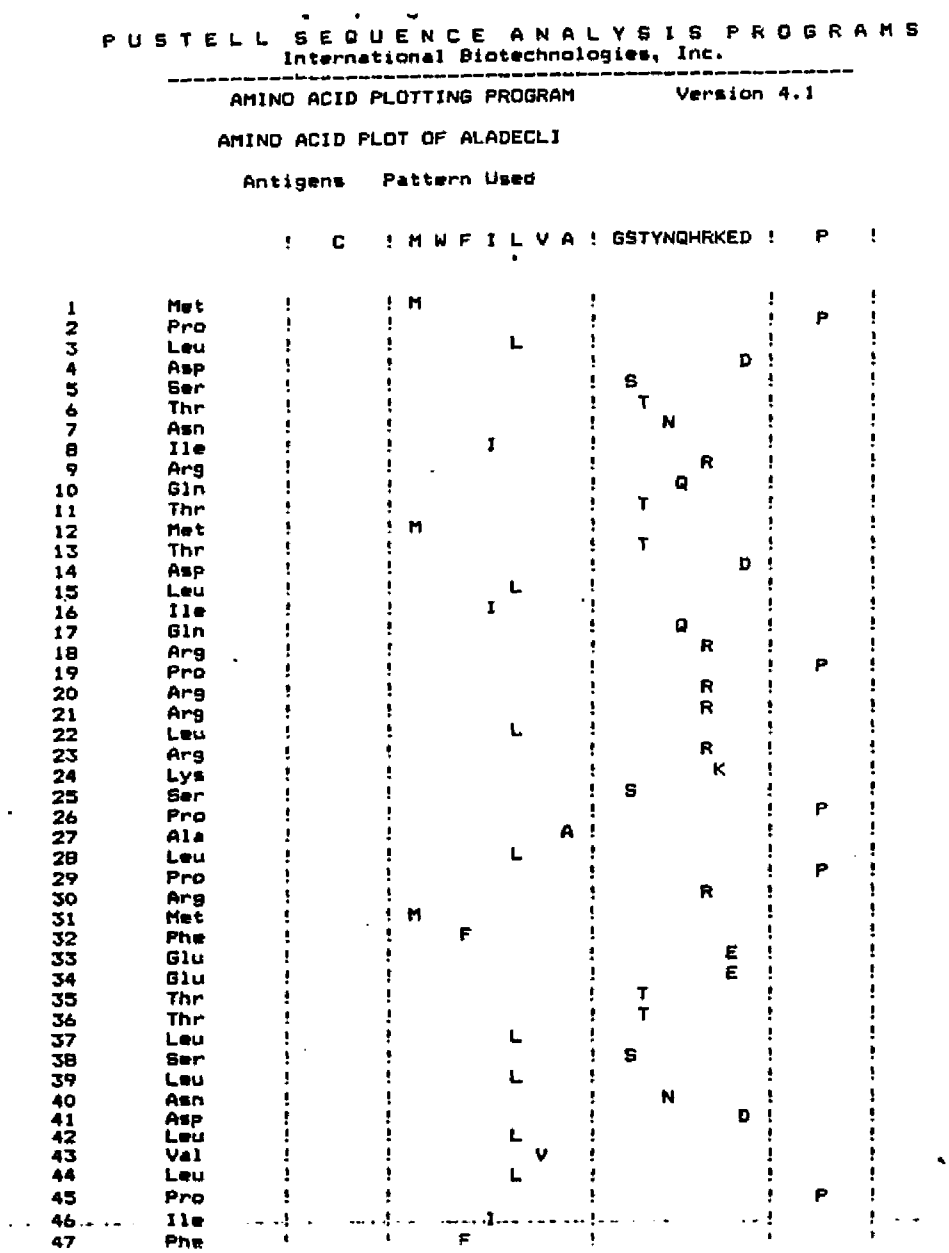

610 620 630 640 650 660
ACGACCTGGTGTTGCCGATCTTTGTTGAAGAAGAAATTGACGACTACAAAGCCGTTGAAG
BstNI DpnI EcoRI* MboII
EcoRII MboI MboII
ScrFI Sau3A
670 680 690 700 710 720
CCATGCCAGGCGTGATGCGCATTCCAGAGAAACATCTGGCACGCGAAATTGAACGCATCG
BstNI HhaI EcoRI*
EcoRII HinPI ThaI
NlaIII MstI
ScrFI
SfaNI
730 740 750 760 770 780
CCAACGCCGGTATTCGTTCCGTGATGACTTTTGGCATCTCTCACCATACCGATGAAACCG
HpaII HphI SfaNI HpaII
SfaNI SfaNI
790 800 810 820 830 840
GCAGCGATGCCTGGCGGGAAGATGGACTGGTGGCGCGTATGTCGCCGATCTGCAAGCAGA
Fnu4HI BbvI MboIIHhaI DpnI
BstNI HinPI MboI
EcoRII ThaI Sau3A
ScrFI
850 860 870 880 890 900
CCGTGCCAGAAATGATCGTTATGTCAGACACCTGCTTCTGTGAATACACTTCTCACGGTC
TthIIIIIDpnI
MboI
Sau3A
910 920 930 940 950 960
ACTGCGGTGTGCTGTGCGAGCATGGCGTCGACAACGACGCGACTCTGGAAAATTTAGGCA
HgaI AccI ThaI HgaI EcoRI
AhaII HinfI EcoRI*
HgiDI
HincII
NlaIII
SalI
TaqI

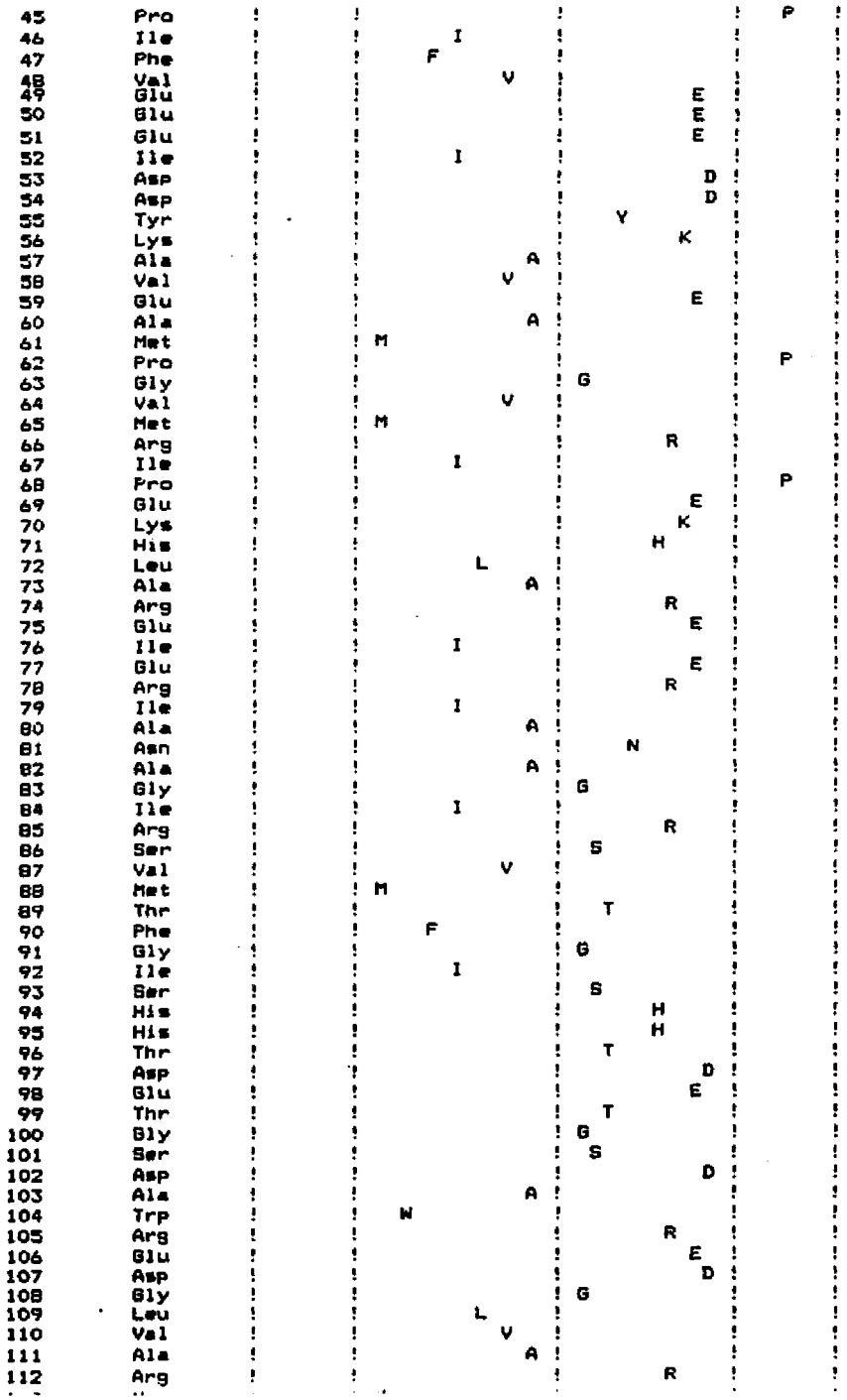

970 980 990 1000 1010 1020
AGCAAGCCGTGGTTGCAGCTGCTGCAGGTGCAGACTTCATCGCCCCTTCCGCCGCGATGG
BbvI AluI BbvI Fnu4HI
BbvI Fnu4HIFnu4HI ThaI
Fnu4HI PstI
PvuII
TthIIIII
1030 1040 1050 1060 1070 1080
ACGGCCAGGTACAGGCGATTCGTCAGGCGCTGGACGCTGCGGGATTTAAAGATACGGCGA
BstNI HinfI BbvI HaeII Fnu4HI AhaIII
EcoRIIRsaI HhaI EcoRI
HaeIII HinPI HgaI
ScrFI
1090 1100 1110 1120 1130 1140
TTATGTCGTATTCGACCAAGTTCGCCTCCTCCTTTTATGGCCCGTTCCGTGAAGCTGCCG
TaqI MnlI BbvI AluIHpaII
HaeIII Fnu4HI
MnlI
Sau96I
1150 1160 1170 1180 1190 1200
GAAGCGCATTAAAAGGCGACCGCAAAAGCTATCAGATGAACCCAATGAACCGTGCTGAGG
HhaI AluI MnlI DdeI
HinPI
1210 1220 1230 1240 1250 1260
GCATTGCTGAATACCTGCTGGATGAAGCCCAGGGGCCAGACTGCCTGATGGTTAAACCTG
BstNIHaeIII
EcoRIINlaIV
FokI
Sau96I
ScrFI
1270 1280 1290 1300 1310 1320
CTGGAGCGTACCTCAACATCGTGCGTGAGCTGCGTGAACGTACTGAATTGCCGATTGGCG
RsaI BbvI MnlI AluI RsaIEcoRI* HhaI
Fnu4HI HinPI
ThaI
1330 1340 1350 1360 1370 1380

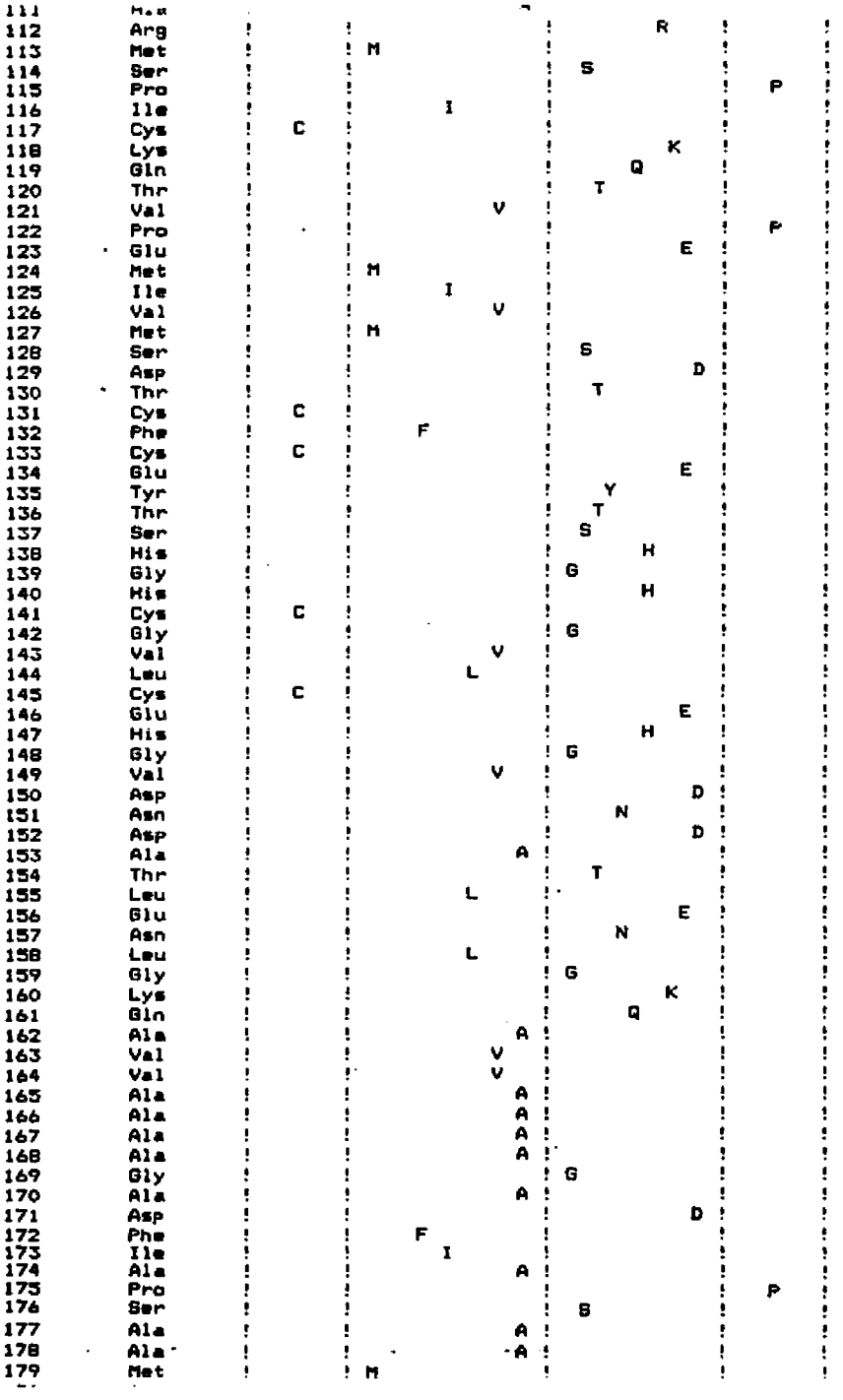

1330 1340 1350 1360 1370 1380
CGTATCAGGTGAGCGGTGAGTATGCGATGATTAAGTTCGCCGCGCTGGCGGGTGCTATAG
HphI HphI BglI Fnu4HI HhaI HinPI ThaI
1390 1400 1410 1420 1430 1440
ATGAAGAGAAAGTCGTGCTCGAAAGCTTAGGTTCGATTAAGCGTGCGGGTGCGGATCTGA
HgiAI AluI Bsp1286DdeI HindIII TaqI MboII TaqI TaqI DpnI EcoRI MboI Sau3A XhoII
1450 1460 1470 1480 1490 1500
TTTTCAGCTACTTTGCGCTGGATTTGGCTGAGAAGAAGATTCTGCGTTAATTTTGTTTCC
AluI HhaI EcoRI DdeI HinPI EcoRI HinfI MboII EcoRI MboII
***** WARNING: Too many overlapping enzymes ****
1510 1520 1530 1540 1550 1560
TCTCAGAAAGGGGGAGTTAACCTGTTGAAAGTACGGTGAACnnnCCCTCACCCTAACCCT
DdeI MnlI HincII HpaI FokI FokI HphI RsaI HpaII ApaI MnlI MnlI AvaII HphI MnlI HphI BanII MboI Bsp1286 DpnI HaeIII HaeIII HhaI
1570 1580 1590 1600 1610 1620
CTCCCCAAGGGGCGAGGGGACCGTCCACTCTCGTATTACCCCGCCCGATAAAACGGTTTA
MnlI MnlI AvaII NlaIV Sau96I
1630 1640 1650 1660 1670 1680
TCCCCAAGGATCGCCCGATGCATTATCCGTCGCTGTGGCAGGTAATCGGCATTGGCATAG
DpnI MboI Sau3A NsiI

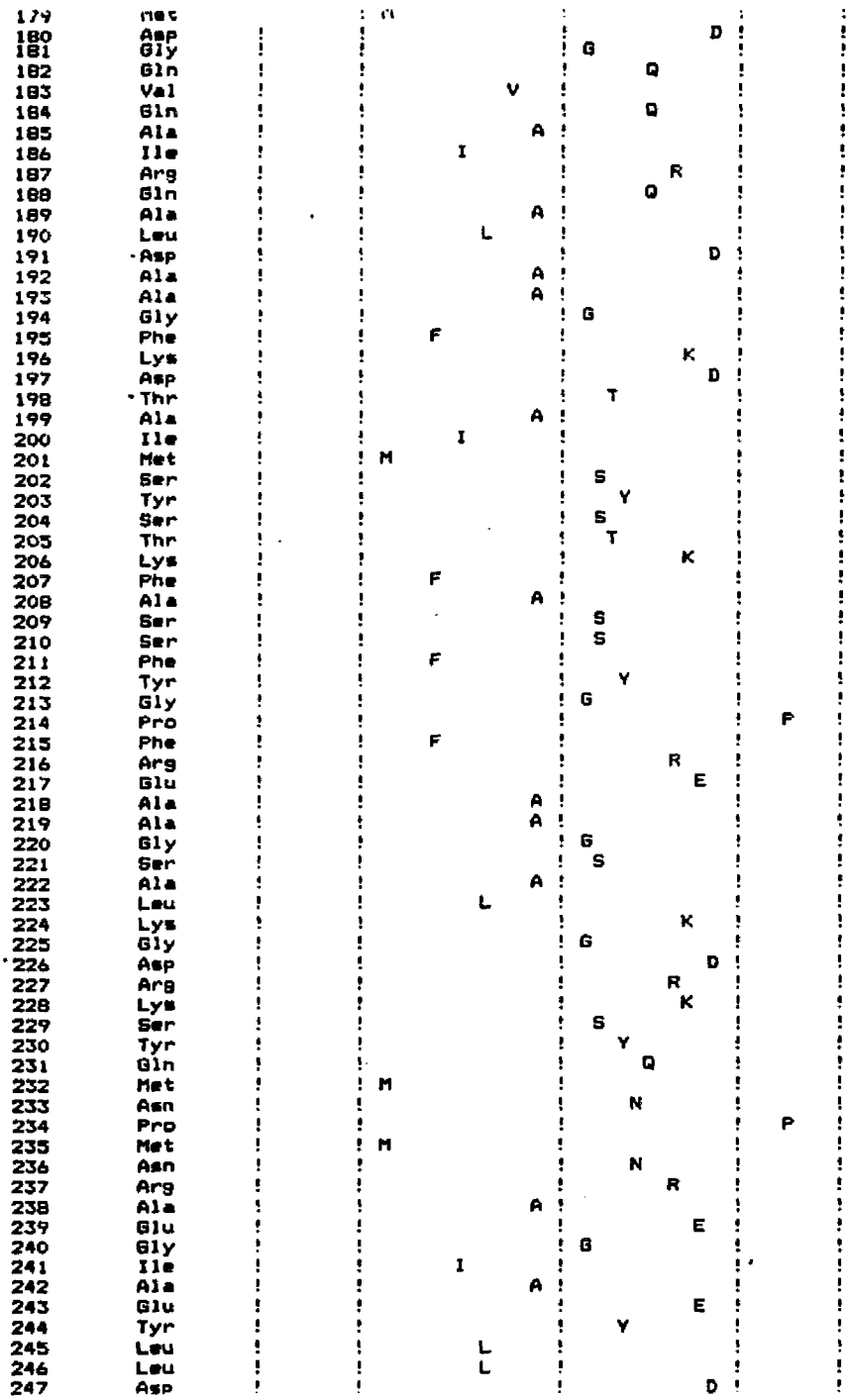

DpnI
NsiI
MboI
Sau3A
SfaNI
1690
1700
1710
1720
1730
1740
TGCTGGGTCACGCCCTTATCCAAATCGCAATATATTTGGTTGCCAGCGCCAGCCACCTGA
HaeII
HhaI
HinPI
1750
1760
1770
1780
1790
1800
AACTCCGGTTTGGTGATATGGGCAAACAAAAAACTTAACAAGGCTTCGCTCTCTTTTCTC
HpaII
HphI
TthIIIII
1810
1820
1830
1840
1850
1860
GCTCACATCAACAATTCGCGTAGgTAAAGCCTTCATTCACAAACAGCGCCTGTTTACCGC
EcoRI*ThaI
AluI
HaeII
SfaNI
HhaI
HphI
HinPI
TthIIIII
1870
1880
1890
1900
TCACCGGATGCGTTCGCACCACCGGATGTAGCAACGGTGGTTTTGGA
HpaII
FokI
HpaII
FokI

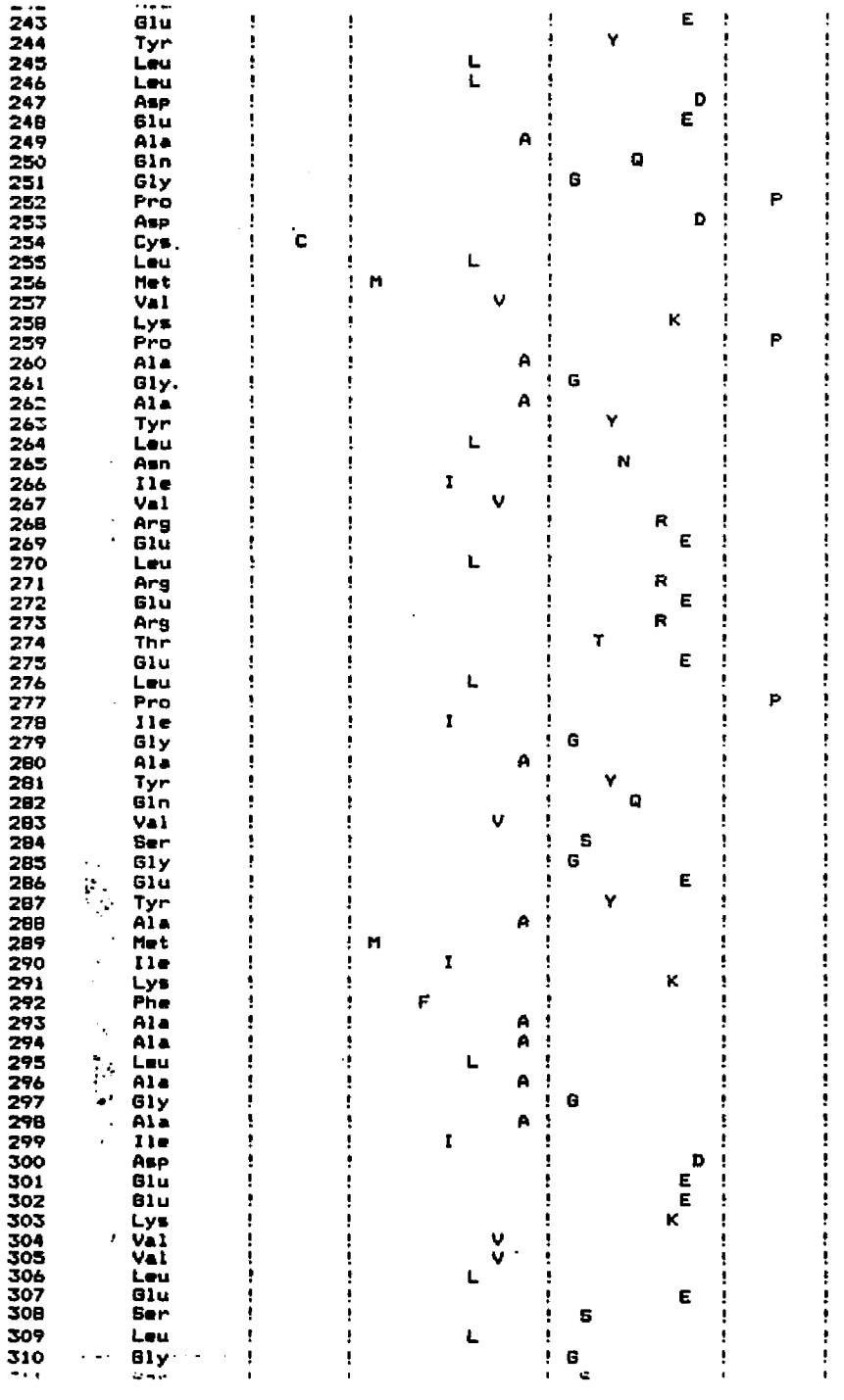
Appendix F. Antigen plot of *E. coli* ALA D

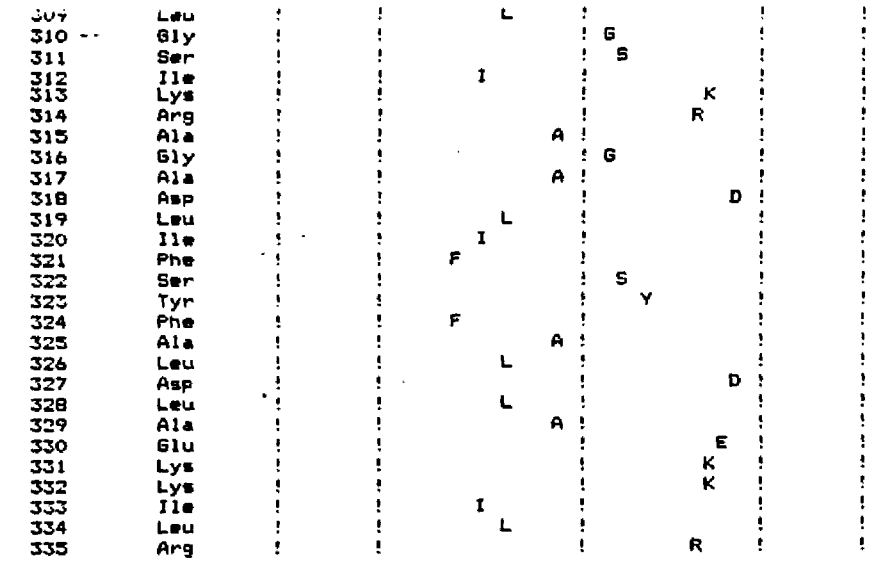



45 Pro
46 Ile
47 Phe
48 Val
49 Glu
50 Glu
51 Glu
52 Ile
53 Asp
54 Asp
55 Tyr
56 Lys
57 Ala
58 Val
59 Glu
60 Ala
61 Met
62 Pro
63 Gly
64 Val
65 Met
66 Arg
67 Ile
68 Pro
69 Glu
70 Lys
71 His
72 Leu
73 Ala
74 Arg
75 Glu
76 Ile
77 Glu
78 Arg
79 Ile
80 Ala
81 Asn
82 Ala
83 Gly
84 Ile
85 Arg
86 Ser
87 Val
88 Met
89 Thr
90 Phe
91 Gly
92 Ile
93 Ser
94 His
95 His
96 Thr
97 Asp
98 Glu
99 Thr
100 Gly
101 Ser
102 Asp
103 Ala
104 Trp
105 Arg
106 Glu
107 Asp
108 Gly
109 Leu
110 Val
111 Ala
112 Arg


111 Met
112 Arg
113 Met
114 Ser
115 Pro
116 Ile
117 Cys
118 Lys
119 Gln
120 Thr
121 Val
122 Pro
123 Glu
124 Met
125 Ile
126 Val
127 Met
128 Ser
129 Asp
130 Thr
131 Cys
132 Phe
133 Cys
134 Glu
135 Tyr
136 Thr
137 Ser
138 His
139 Gly
140 His
141 Cys
142 Gly
143 Val
144 Leu
145 Cys
146 Glu
147 His
148 Gly
149 Val
150 Asp
151 Asn
152 Asp
153 Ala
154 Thr
155 Leu
156 Glu
157 Asn
158 Leu
159 Gly
160 Lys
161 Gln
162 Ala
163 Val
164 Val
165 Ala
166 Ala
167 Ala
168 Ala
169 Gly
170 Ala
171 Asp
172 Phe
173 Ile
174 Ala
175 Pro
176 Ser
177 Ala
178 Ala
179 Met


179 met
180 Asp
181 Gly
182 Gln
183 Val
184 Gln
185 Ala
186 Ile
187 Arg
188 Gln
189 Ala
190 Leu
191 Asp
192 Ala
193 Ala
194 Gly
195 Phe
196 Lys
197 Asp
198 Thr
199 Ala
200 Ile
201 Met
202 Ser
203 Tyr
204 Ser
205 Thr
206 Lys
207 Phe
208 Ala
209 Ser
210 Ser
211 Phe
212 Tyr
213 Gly
214 Pro
215 Phe
216 Arg
217 Glu
218 Ala
219 Ala
220 Gly
221 Ser
222 Ala
223 Leu
224 Lys
225 Gly
226 Asp
227 Arg
228 Lys
229 Ser
230 Tyr
231 Gln
232 Met
233 Asn
234 Pro
235 Met
236 Asn
237 Arg
238 Ala
239 Glu
240 Gly
241 Ile
242 Ala
243 Glu
244 Tyr
245 Leu
246 Leu
247 Asp


243 Glu
244 Tyr
245 Leu
246 Leu
247 Asp
248 Glu
249 Ala
250 Gln
251 Gly
252 Pro
253 Asp
254 Cys.
255 Leu
256 Met
257 Val
258 Lys
259 Pro
260 Ala
261 Gly.
262 Ala
263 Tyr
264 Leu
265 Asn
266 Ile
267 Val
268 Arg
269 Glu
270 Leu
271 Arg
272 Glu
273 Arg
274 Thr
275 Glu
276 Leu
277 Pro
278 Ile
279 Gly
280 Ala
281 Tyr
282 Gln
283 Val
284 Ser
285 Gly
286 Glu
287 Tyr
288 Ala
289 Met
290 Ile
291 Lys
292 Phe
293 Ala
294 Ala
295 Leu
296 Ala
297 Gly
298 Ala
299 Ile
300 Asp
301 Glu
302 Glu
303 Lys
304 Val
305 Val
306 Leu
307 Glu
308 Ser
309 Leu
310 Gly



Plotted from AA 1 to AA 335
Numbered from AA 1

Appendix G. Amino acid composition, molecular weight, pI, and cleavage map with CNBr and trypsin of *E. coli* ALA D

PUSTELL SEQUENCE ANALYSIS PROGRAMS
International Biotechnologies, Inc.

PEPTIDE ANALYSIS OF ALADECLI
from AA 1 to AA 335

Amino Acid Composition **********

Non-polar:		Number	Percent
Ala	A	42	12.537
Val	V	18	5.373
Leu	L	28	8.358
Ile	I	22	6.567
Pro	P	15	4.478
Met	M	15	4.478
Phe	F	12	3.582
Trp	W	1	.299
Polar:		**Number**	**Percent**
Gly	G	23	6.866
Ser	S	20	5.970
Thr	T	15	4.478
Cys	C	6	1.791
Tyr	Y	10	2.985
Asn	N	8	2.388
Gln	Q	10	2.985
Acidic:		**Number**	**Percent**
Asp	D	21	6.269
Glu	E	27	8.060

Glu	E	27	8.060

Basic:		Number	Percent
Lys	K	15	4.478
Arg	R	21	6.269
His	H	6	1.791

Total AAs = 335
Calculated Molecular Weight = 36745.400
Estimated pI = 5.79

Cleavage Map **********
C = CNBr T = Trypsin

```
          10        20        30        40        50        60
          *         *         *         *         *         *
MPLDSTNIRQTMTDLIQRPRRLRKSPALPRMFEETTLSLNDLVLPIFVEEEIDDYKAVEA
^       ^  ^     ^ ^^ ^^      ^^                         ^
C       T  C     T TT TT      TC                         T

          70        80        90       100       110       120
          *         *         *         *         *         *
MPGVMRIPEKHLAREIERIANAGIRSVMTFGISHHTDETGSDAWREDGLVARMSPICKQT
^   ^^   ^   ^   ^      ^  ^                ^       ^^    ^
C   CT   T   T   T      T  C                T       TC    T

         130       140       150       160       170       180
          *         *         *         *         *         *
VPEMIVMSDTCFCEYTSHGHCGVLCEHGVDNDATLENLGKQAVVAAAAGADFIAPSAAMD
  ^  ^                                  ^                   ^
  C  C                                  T                   C

         190       200       210       220       230       240
          *         *         *         *         *         *
GQVQAIRQALDAAGFKDTAIMSYSTKFASSFYGPFREAAGSALKGDRKSYQMNPMNRAEG
       ^         ^    ^    ^         ^        ^  ^^   ^  ^ ^
       T         T    C    T         T        T  TT   C  C T

         250       260       270       280       290       300
          *         *         *         *         *         *
IAEYLLDEAQGPDCLMVKPAGAYLNIVRELRERTELPIGAYQVSGEYAMIKFAALAGAID
                ^ ^           ^  ^ ^               ^ ^
                C T           T  T T               C T

         310       320       330
          *         *         *
EEKVVLESLGSIKRAGADLIFSYFALDLAEKKILR
  ^         ^^                ^^  ^
  T         TT                TT  T
```

Cleavage Fragments: **********

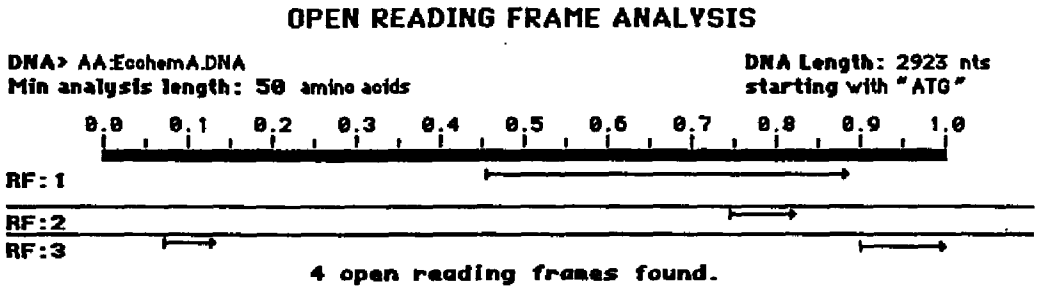
CNBr:	From	To	Mol. Wt.	Residues	pI
	128	179	5269.459	52	5.13
	290	335	4994.369	46	7.00
	257	289	3695.020	33	7.00
	32	61	3473.660	30	4.17
	202	232	3438.510	31	11.10
	89	113	2788.860	25	5.32
	66	88	2659.950	23	11.20
	180	201	2319.450	22	5.32
	236	256	2308.430	21	4.72

180	201	2319.450	22	5.32
236	256	2308.450	21	4.72
13	31	2304.620	19	12.57
1	12	1406.570	12	7.00
114	124	1232.380	11	7.00
62	65	402.500	4	7.00
125	127	361.490	3	7.00
233	235	360.420	3	7.00

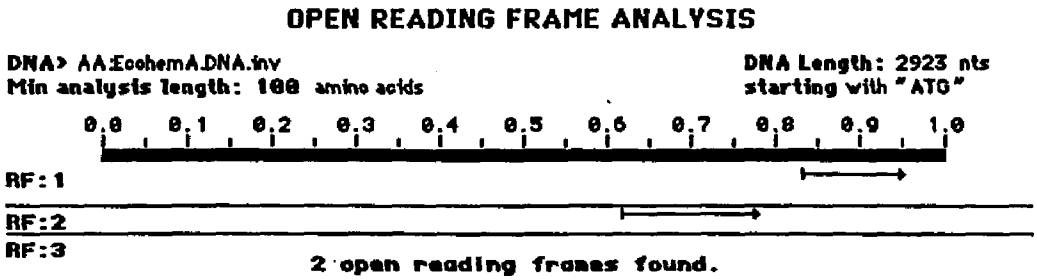
Trypsin:	From	To	Mol. Wt.	Residues	pI
	119	160	4584.859	42	5.13
	31	56	3103.280	26	4.17
	161	187	2599.780	27	5.32
	238	258	2265.410	21	4.72
	86	105	2234.290	20	5.71
	274	291	1970.130	18	5.32
	315	331	1843.930	17	4.72
	292	303	1234.250	12	4.72
	207	216	1178.250	10	11.20
	229	237	1140.260	9	11.20
	197	206	1116.150	10	7.00
	10	18	1105.230	9	7.00
	259	268	1073.200	10	11.20
	57	66	1060.250	10	7.00
	1	9	1046.150	9	7.00
	304	313	1044.160	10	7.00
	188	196	919.940	9	7.00
	106	112	758.790	7	5.32
	217	224	745.740	8	7.00
	79	85	713.790	7	11.20
	113	118	677.800	6	10.19
	25	30	639.710	6	11.20
	75	78	545.570	4	5.32
	71	74	495.560	4	11.20
	67	70	485.520	4	7.00
	269	271	416.460	3	7.00
	333	335	400.500	3	11.20
	225	227	346.330	3	7.00
	272	273	303.310	2	7.00
	22	23	287.350	2	11.20
	19	20	271.310	2	11.20
	21	21	174.200	1	11.20
	314	314	174.200	1	11.20
	24	24	146.150	1	10.19
	228	228	146.150	1	10.19
	332	332	146.150	1	10.19

Appendix H. Identification of *hemA* coding region by open reading frame search

a. Forward search:

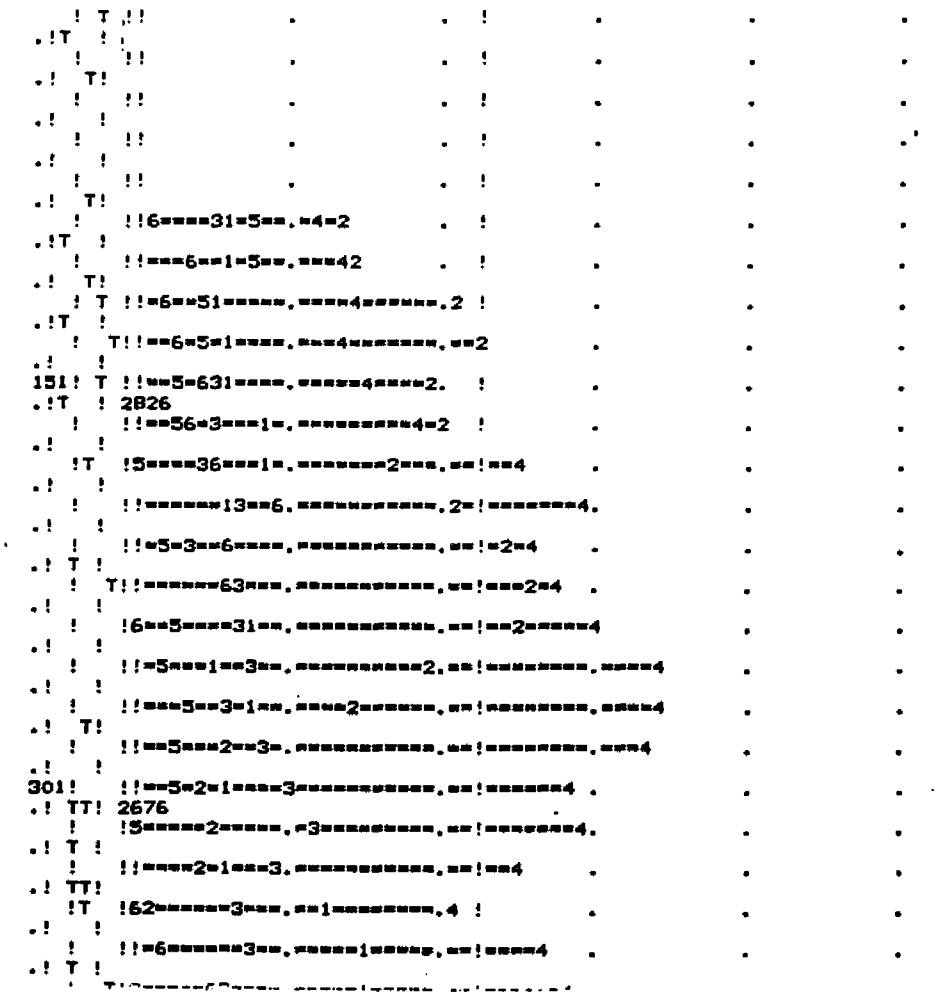


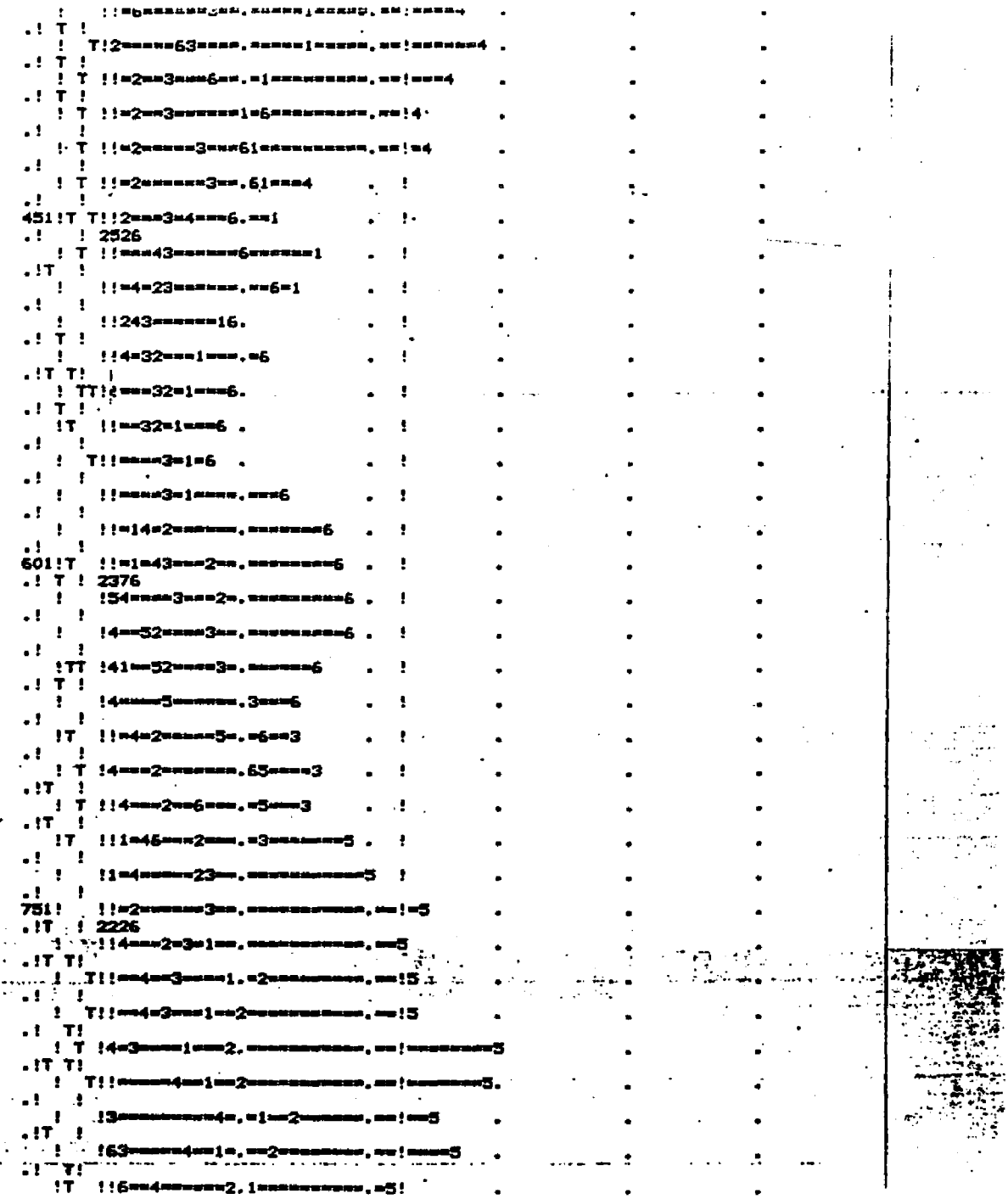
b. Inverse search:

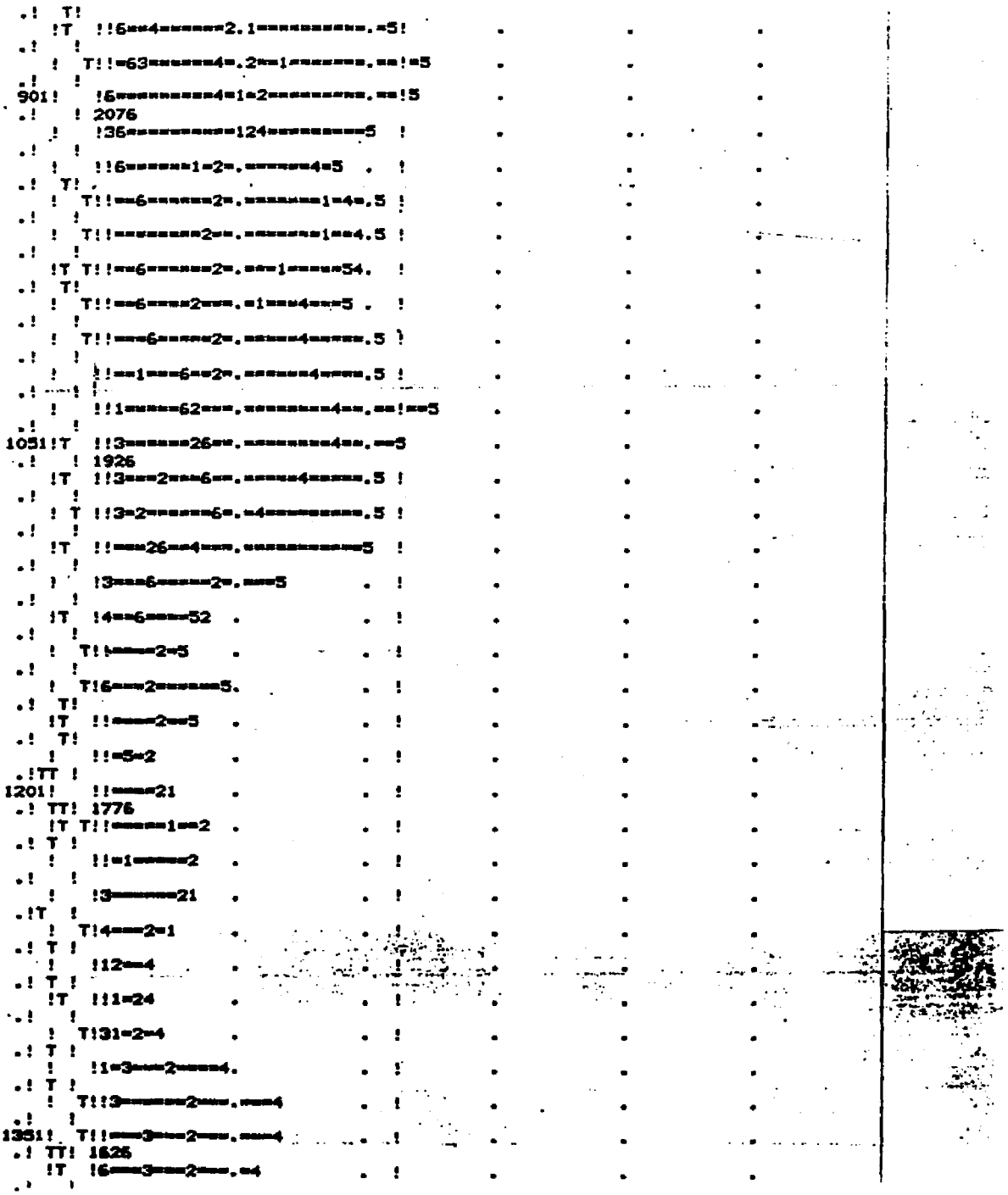


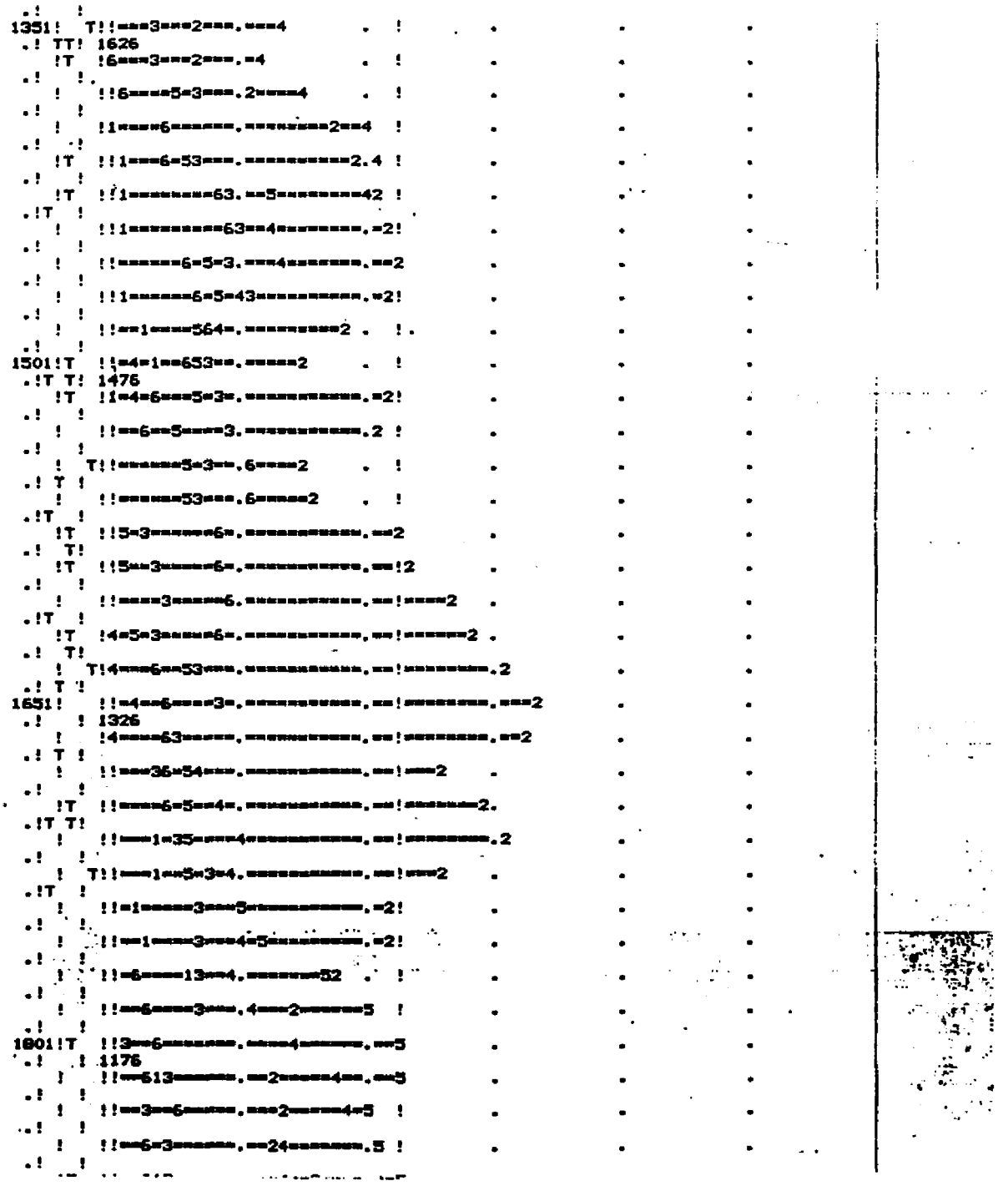
Appendix I. Protein coding regions of the insert of pJL68 located by IBI protein coding region locator program

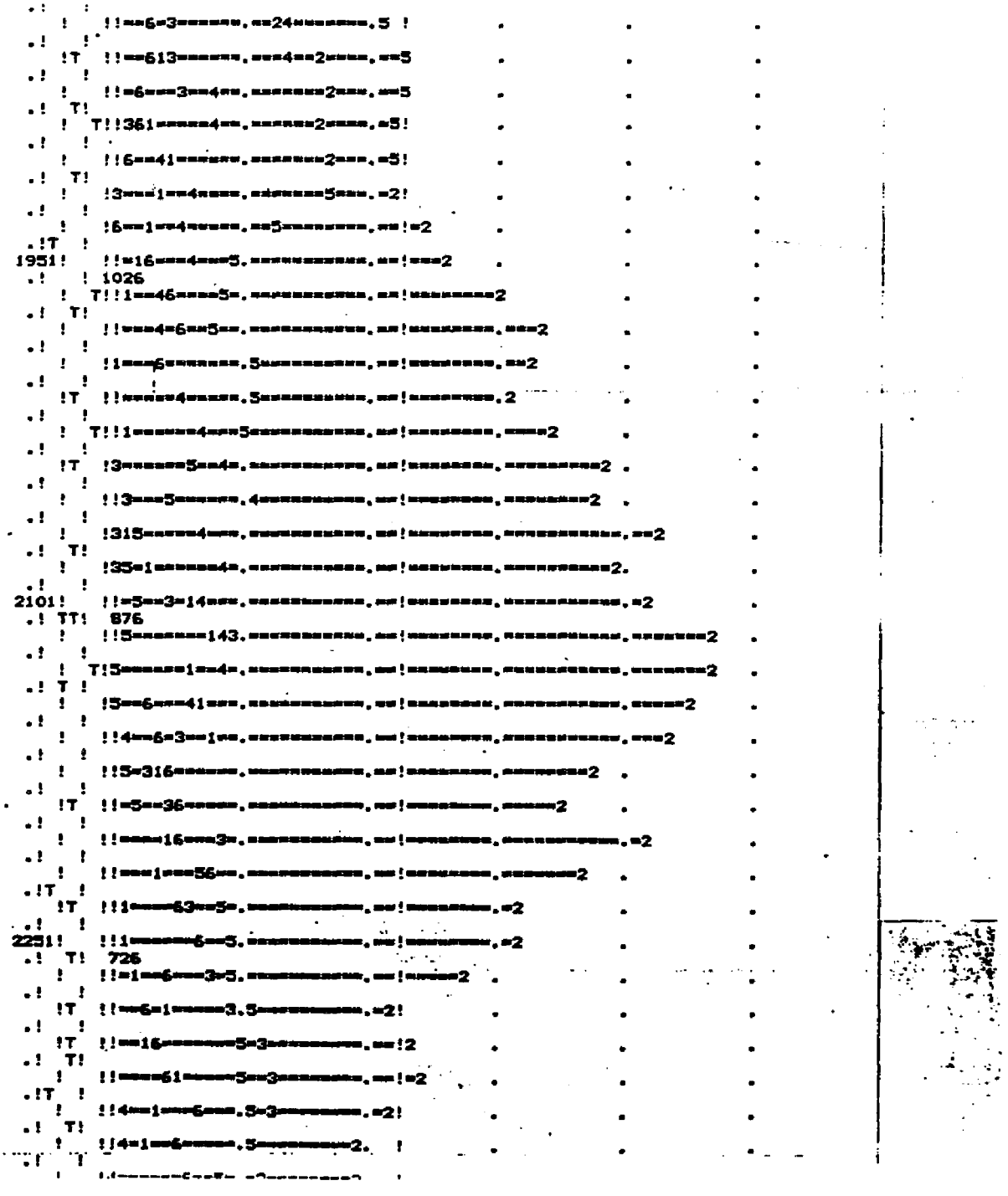
```
P U S T E L L  S E Q U E N C E  A N A L Y S I S  P R O G R A M S
                International Biotechnologies, Inc.
-------------------------------------------------------------
PROTEIN CODING REGION LOCATOR  Version 2.0

Coding Prediction for BSFRAG4
From base      1 to base  2963
Compression=  5  Codon bias table RADMBIAS     used.
Values strand adjusted
Products over a range of 50 codons

    !123!R          1.5          2.0 M        2.5          3.0          3.5
   !456!
```

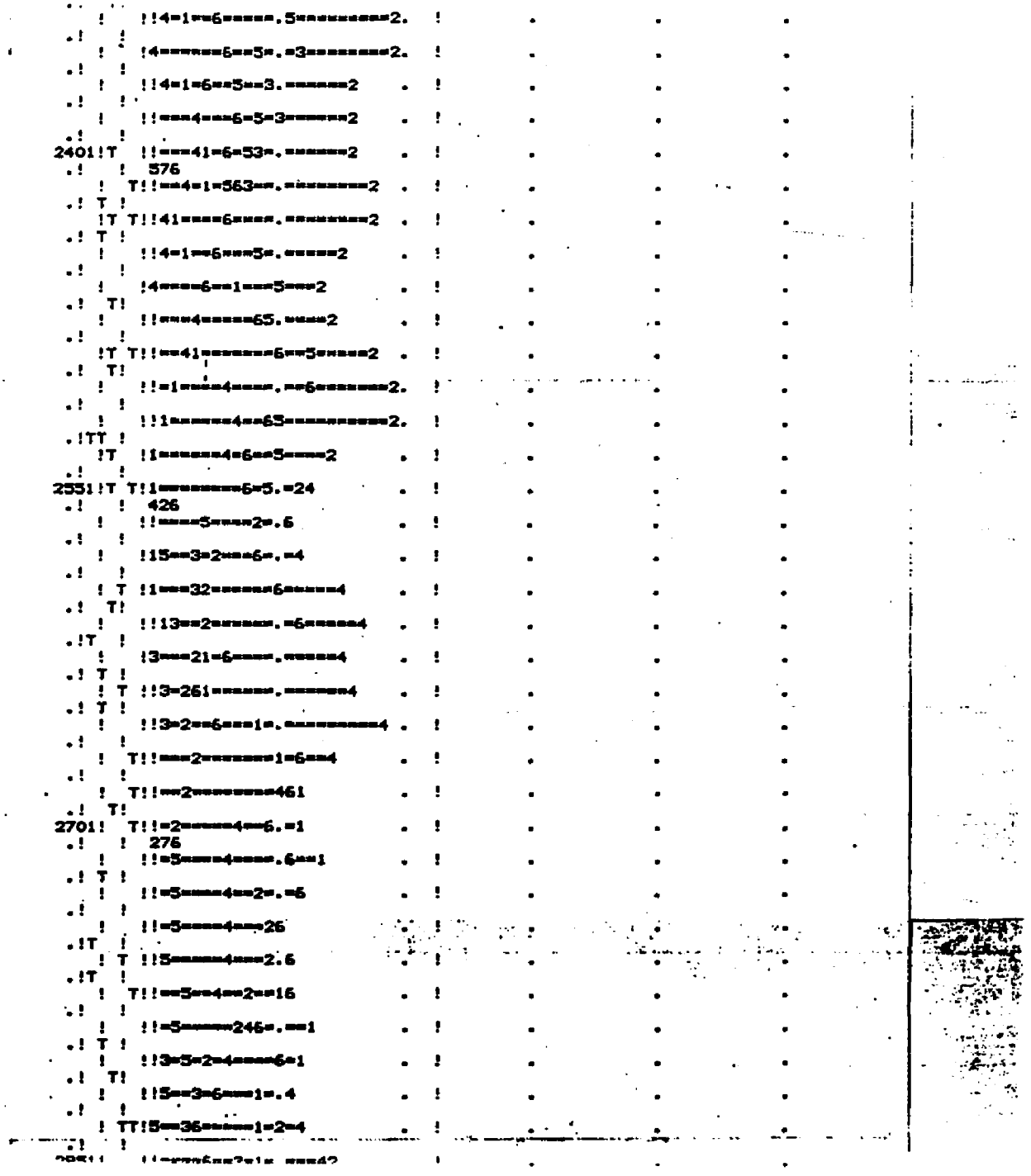


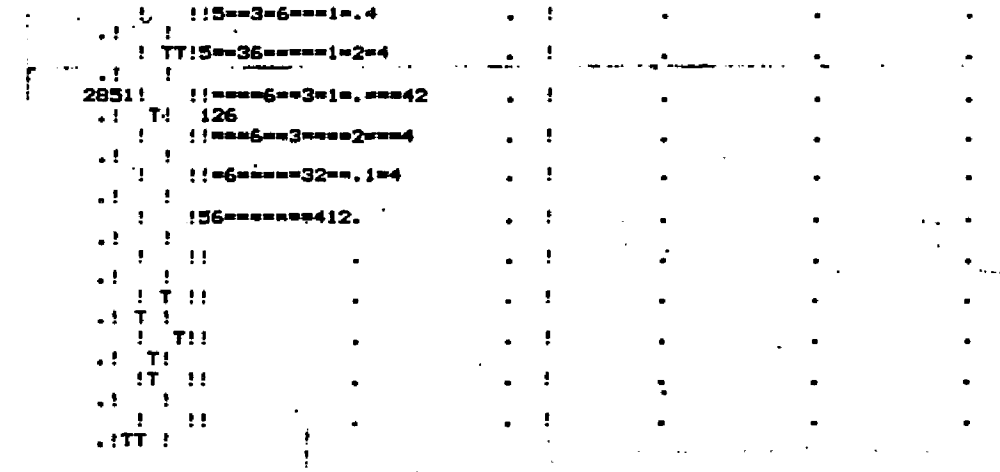




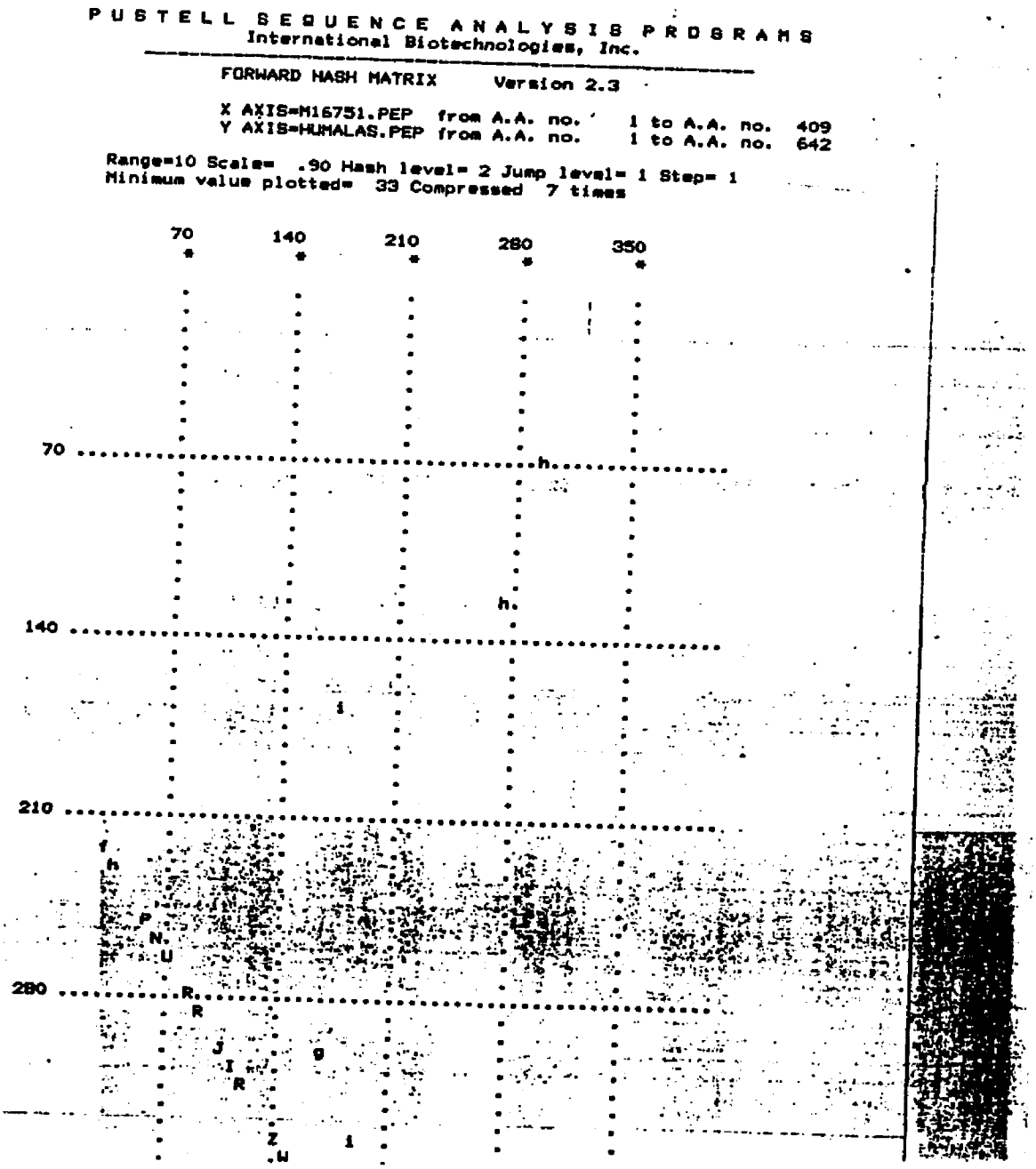


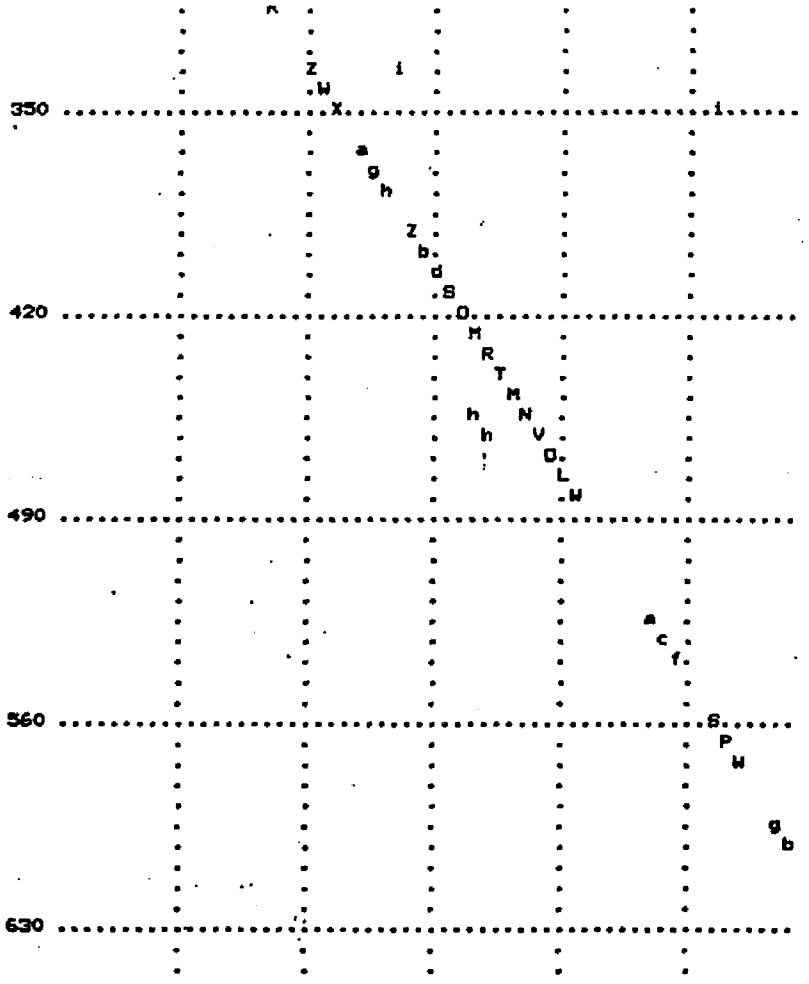





!!5==3=6===1=.4
!TT!5==36=====1=2=4
2851!!!====6==3=1=.===42
T! 126
!!===6==3====2===4
!!=6======32==.1=4
!56=======412.

Appendix J. Comparison of the human and *B. japonicum* ALA synthases with homology matrix




350
420
490
560
630

Appendix K . Top 50 scores obtained by a Genbank datadase search with IFIND protocol on Bionet

RANK	SEQUENCE	LENGTH	SCORE	SD*
1	ECORF1X	L=1440	539	886.33
2	M15423	L=5614	68	85.23
3	NPANPVP	L=1143	24	10.39
4	ECOASPAW	L=2901	21	5.29
5	NPAPH	L=525	20	3.59
6	ECOAROL	L=810	20	3.59
7	M11436	L=1034	20	3.59
8	M10240	L=1406	20	3.59
9	RATPKCI	L=3113	20	3.59
10	M16167	L=3316	20	3.59
11	M13465	L=3725	20	3.59
12	YSCADE3	L=4883	20	3.59
13	ECOLAC	L=7477	20	3.59
14	HUMHP1G1	L=643	19	1.89
15	TRN1721RE	L=761	19	1.89
16	RATOSP	L=1457	19	1.89
17	SPICPTGI	L=1956	19	1.89
18	HS4EDLI	L=2640	19	1.89
19	YSDSTA1	L=2753	19	1.89
20	M17424	L=2889	19	1.89
21	DROTNP	L=2907	19	1.89
22	HS4MP	L=3407	19	1.89
23	ECOSDHACD	L=3614	19	1.89
24	ECOGYRBF	L=3789	19	1.89
25	DRODDCG	L=4835	19	1.89
26	ECOKDPABC	L=4933	19	1.89
27	HS1ATIF1	L=7241	19	1.89
28	ECOGLTA	L=1-10000	19	1.89
29	HUMHPARS1	L=1-10000	19	1.89
30	CHKNCAM01	L=266	18	.19
31	CELMSPPS	L=430	18	.19
32	MUSGKAL	L=500	18	.19
33	PWMBREPS	L=569	18	.19
34	MUSKALLR3	L=847	18	.19
35	MUSNGFG	L=849	18	.19
36	MUSEGFBPB	L=869	18	.19
37	M17419	L=999	18	.19
38	BNOPIL2S	L=1033	18	.19
39	M13500	L=1210	18	.19
40	ECOPFKBK	L=1249	18	.19
41	NEUMTCO3G	L=1254	18	.19
42	PHOCHL	L=1431	18	.19
43	ASNMTTGRN	L=1550	18	.19
44	YSCLTE1	L=1700	18	.19
45	YSCHKA	L=1716	18	.19
46	MUSIGCF	L=1836	18	.19
47	RABRFVHOM	L=1902	18	.19
48	M11639	L=2014	18	.19
49	M11638	L=2015	18	.19
50	M11641	L=2015	18	.19

*, SD: STANDARD-DEVIATION-FROM-MEAN.

Appendix L. Secondary structures of *E. coli hemA* protein

Print detailed analysis data in the output file? (<CR>=No)
Do you want verbose commentary in the output file? (<CR>=No)
HEMA.PEP

Secondary Structures predicted by Chou and Fasman algorithm:

Begin	End	Type	<Pa>	<Pb>	<Pt>
1	7	alpha	1.129	1.103	
2	6	beta	1.176	1.184	
11	19	alpha	1.057*	.929	
23	28	alpha	1.048*	.877	
24	27	turn	.878	.645	1.355
26	38	alpha	1.143	.936	
38	42	beta	1.060	1.100	
41	44	turn	.828	1.075	1.150
45	52	beta	.910	1.206	
48	51	turn	.743	1.005	1.285
53	63	alpha	1.129	.951	
53	57	beta	1.090	1.126	
61	75	alpha	1.121	.982	
62	65	turn	1.000	.958	1.148
68	74	beta	1.097	1.217	
74	77	turn	.850	1.017	1.185
79	88	alpha	1.124	.849	
87	92	beta	.977	1.143	
90	97	alpha	1.045	.980	
91	94	turn	.948	.850	1.238
93	96	turn	1.028	.700	1.285
97	101	beta	1.098	1.134	
98	105	alpha	1.083	1.023	
104	107	turn	.993	.908	1.060
107	114	alpha	1.069	1.001	
116	124	alpha	1.098	1.096	
116	130	alpha	1.099	1.013	
116	121	beta	1.023	1.258	
128	131	turn	.903	.835	1.245
132	158	alpha	1.128	.966	
151	156	beta	1.032	1.052	
160	169	alpha	1.189	1.043	
160	181	alpha	1.119	1.051	
163	168	beta	1.143	1.198	
170	177	beta	1.058	1.190	
180	188	beta	1.027	1.348	
190	212	alpha	1.166	1.025	
193	197	beta	1.138	1.232	
206	215	beta	1.084	1.163	
214	226	alpha	1.162	.971	
224	252	alpha	1.108	.957	
230	234	beta	1.108	1.236	
243	247	beta	1.134	1.234	
258	261	turn	.938	.823	1.183
259	270	alpha	1.104	.918	
269	272	turn	.935	.963	1.110
270	277	beta	1.033	1.103	
272	278	alpha	1.089	1.077	
278	286	alpha	1.078*	.973	
281	284	turn	.905	.930	1.108
288	293	alpha	1.155	.948	
292	298	beta	1.044	1.156	
301	306	alpha	1.032	.972	
303	307	beta	1.050	1.270	
306	347	alpha	1.167*	.977	
323	327	beta	1.100	1.284	

344	348	beta	1.018	1.112	
348	351	turn	.888	1.063	1.125
349	369	alpha	1.209	.916	
368	371	turn	.940	.743	1.210
371	385	alpha	1.139	1.005	
382	391	beta	1.064	1.102	
384	387	turn	.923	1.078	1.015
391	394	turn	.995	.828	1.038
393	402	alpha	1.102	.931	
400	403	turn	.995	.763	1.158
402	405	turn	.815	.680	1.510
406	414	alpha	1.047	.957	
407	412	beta	1.022	1.158	
411	414	turn	.993	.880	1.108
413	416	turn	.890	.835	1.260

<Px> = average propensity for x = a(lpha), b(eta), or t(urn).
("*" in the <Pa> column means that <Pa> was adjusted
according to Conditions II.A.5 in (Chou and Fasman 1978)).

Secondary structure of *hemA* protein predicted by Chou and Fasman algorithm:

Key: AAAAAA Alpha helix
BBBBB Beta sheet
TTTT Turn

```
        10        20        30        40        50        60        70
MTLLALGINHKTAPVSLRERVSFSPDKLDQALDSLLAQPMVQGGVVLSTCNRTELYLSVEEQDNLQEALI
AAAAAAA   AAAAAAAAA   AAAAAAAAAAAAAAAA              AAAAAAAAAAAAAAAAAA
 BBBBB                              BBBBB  BBBBBBBBBBBBB           BBB
                          TTTT             TTTT   TTTT          TTTT

        80        90       100       110       120       130       140
RWLCDYHNLNEEDLRKSLYWHQDNDAVSHLMRVASGLDSLVLGEPQILGQVKKAFADSQKGHMKASELER
AAAAA   AAAAAAAAAA AAAAAAAAAAAAAAAA AAAAAAAA AAAAAAAAAAAAAAA AAAAAAAAA
BBBB             BBBBBB    BBBBB               BBBBBB
   TTTT             TTTTTT       TTTT                    TTTT

       150       160       170       180       190       200       210
MFQKSFSVAKRVRTETDIGASAVSVAFAACTLARQIFESLSTVTVLLVGAGETIELVARHLREHKVQKMI
AAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA        AAAAAAAAAAAAAAAAAAAA
          BBBBBB       BBBBBB BBBBBBBB  BBBBBBBBB    BBBBB         BBBBB


       220       230       240       250       260       270       280
IANRTRERAQILADEVGAEVIALSDIDERLREADIIISSTASPLPIIGKGMVERALKSRRNQPMLLVDIA
AA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA      AAAAAAAAAAA AAAAAAAAA
BBBBB               BBBBB        BBBBB                    BBBBBBBB
                                                TTTT       TTTT

       290       300       310       320       330       340       350
VPRDVEPEVGKLANAYLYSVDDLQSIISHNLAQRKAAAVEAETIVAQETSEFMAWLRAQSASETIREYRS
AAAAAA AAAAAA       AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA
          BBBBBBB    BBBBB                BBBBB                  BBBBB
TTTT                                                              TTT

       360       370       380       390       400       410
QAEQVRDELTAKALAALEQGGDAQAIMQDLAWKLTNRLIHAPTKSLQQAARDGDNERLNILRDSLGLE
AAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAA       AAAAAAAAA   AAAAAAAAA
                                  BBBBBBBBBB              BBBBBB
T                 TTTT            TTTT   TTTT     TTTTTT     TTTTTT
```

BIBLIOGRAPHY

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

Anderson, P. M., and Desnick, R. J. (1979) Purification and properties of Δ-aminolevulinate dehydrase from human erythrocytes. J. Biol. Chem. **254**:6924-6930

Anderson, P., and Desnick, R. (1980) Purification and properties of uroporphyrinogen I synthase from human erythrocytes: identification of stable enzyme-substrate intermediate. J. Biol. Chem. **255**:1993-1999

Aota. S.-I., Gojobori, T., Ishibashi, F., Maruyama, T. and Ikemura, T. (1988) Codon usage tabulated from the GenBank genetics sequence data. Nucl. Acids Res. **16**:r315-r402.

Avissar, Y. (1980) Biosynthesis of 5-aminolevulinate from glutamate in *Anabaena variabilis* . Biochem. Biophys. Acta **613**: 220-228

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

Bachman, B. J. and Low, K. B. (1980) Linkage map of *Escherichia coli* K–12. Edition 6. Microbiol. Rev. 44:1–56

Bachmann, B. J. (1983) Linkage map of *Escherichia coli* K-12, edition 7. Microbiological Reviews June 1983, p. 180-230

Barnard, G. F., Itoh, R., Hohbergr, L. H., and Shemin, D. (1977) Mechanism of porphobilinogen synthase. Possible role of essential thiol groups. J. Biol. Chem. **252**:8965-8974

Barnes, W. M. (1987) Sequencing DNA with dideoxyribonucleotides chain terminators: Hints and strategies for big projects. Methods in Enzymology **152**:538-556

Battersby, Fookes, C., Matcham, C., and McDonald, E. (1980) Biosynthesis of pigments of life: formation of macrocycle. Nature (London) **285**:17-21

Battersby, A., Fookes, G., Mc Donald, E., and Meegan, M. (1978) Biosynthesis of type-III porphyrins: proof of intact enzymic conversion of the head-to-tail bilane into uro'gen-III by intramolecular rearrangement. J. C. S. Chem. Commun. **5**:185-186

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

Beale, S. I. (1970) The biosynthesis of δ-aminolevulinic acid in *Chlorella*. Plant Physiol. **45**:504-506

Beale, S. I. (1971) Studies on the biosynthesis of and metabolism of δ-aminolevulinic acid in *Chlorella*. Plant Physiol. **48**:316-319

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

Beale, S. I., Gough, S. P., and Granick, S. (1975) Biosynthesis of δ-aminolevulinic acid from the intact carbon skeleton of glutamic acid in greening barley. Proc. Nat. Acad. Sci. U. S. A. **73**:2719-2723

Beale, S. I. (1976) The biosynthesis of δ-aminolevulinic acid in plants. Phil. Trans. R. Soc. Lond. B. **273**: 99-108.

Beaumont, C., Grandchamp, B., Bogard, M., Verneuil, H. de, and Nordmann, Y. (1986) Porphobilinogen deaminase is unstable in the absence of its substrate. Biochim. Biophys. Acta **882**:384-399

Beljanski, M., and Beljanski, M. (1957) Formation of respiratory enzymes in a streptmycin-resistant *Escherichia coli* mutant without respiratory activity. Ann Inst. Pasteur **92**:396-412

Bishop, T. R., Cohen, P. J., Boyer, S. H., Noyes, A. N., and Frelin, L. P. (1986a) Isolation of a rat liver Δ-aminolevulinate dehydrase (ALAD) cDNA clone: evidence for unequal ALAD gene dosage among inbred mouse strains. Proc. Natl. Acad. Sci. U. S. A. **83**:5568-5572

Bishop, T. R., Frelin, L. P., and Boyer, S. H. (1986b) Nucleotide sequence of rat liver δ-aminolevulinic acid dehydratase cDNA. Nucl. Acids Res. **14**:10115

Bishop, T. R., Hodes, Z. I., Frelin, L. P., and Boyer, S. H. (1989) Cloning and sequence of mouse erythroid δ-aminolevulinate dehydratase cDNA. Nucl. Acids Res. **17**:1775

Borthwick, I. A., Srivastava, G., Hobbs, A. A., Pirola, B. A., Brooker, J. D., May, B. K., and Elliott, W. H. (1984) Molecular cloning of hepatic 5-aminolevulinate synthase. Eur. J. Biochem. **144**:95-99

Borthwick, I. A., Srivastava, G., Day, A. R., Pirola, B. A., Snowell, M. A., May, B. K., and Elliott, W. H. (1985) Complete nucleotide sequence of hepatic 5-aminolevulinate synthase precursor. Eur. J. Biochem. **1150**:481-484

Bottomly, S. S., and Smithee, G. A. (1968) Characterization and measurement of Δ-aminolevulinate synthetase in bone marrow cell mitochondria. Biochim. biophys. Acta **159**:27-37

Brennan, M. J. W., and Cantrill, R. C. (1979) δ–Aminolevulinic acid is a potent agonist for GABA autoreceptors. Nature **280**:514-515

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

Burnham, B. F. (1970) δ-Aminolevulinic acid synthase. Methods Enzymol. **17**:195-204

Cauthen, S. E., Foster, M. A., and Woods, D. D. (1966) Methionine synthesis by extracts of *Salmonella typhimurium*. Biochem. J **98**:630-635

Chang, G. W., and Chang, J. T. (1975) Evidence for the B_{12}-dependent enzyme ethanolamine deaminase in *Salmonella* . Nature **254**:150-151

Chartrand, P., Tardif, D., and Sasarmann, A. (1979) Uroporphyrin- and coproporphyrin I-accumulating mutant of *Escherichia coli* K12. J. Gen. Microbiol. **110**:61-66

Cheh, A. and Neilands, J. B. (1966) Zinc, an essential metal ion for beef liver Δ-aminolevulinate dehydratase. Biochem. Biophys. Commun. **55**:1060-1063

Chou,P. Y., and Fasman, G.D. (1974) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochem. **13**:212

Chretien, S., Dubart, A., Beaupain, D., Raich, N., Grandchamp, B., Rosa, J., Goossens, M., and Romeo, P.-H. (1988) Alternative transcription and splicing of the human porphobilinogen deaminase gene result either in tissue-specific or housekeeping expression. Proc. Natl. Acad. Sci. U. S. A. **85**:6-10

Coleman, D. L. (1966) Purification and properties of Δ-aminolevulinic acid dehydrase from tissues of two strains of mice. J. Biol. Chem. **241**:5511-5517

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

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

Dale, R. M. K., McClure, B. A., and Houchins, J. P. (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18s rDNA. Plasmid **13**:31-40

Davis, R. W., Botstein, D., and Roth, J. R. (1980) Advanced Bacterial Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. pp. 140–141

de Bruijn, F. J., and Ausubel, F. M. (1981) The cloning and transposon mutagenesis of the *glnA* region of *Klebsiella pneumoniae*: identification of *glnR*, a gene involved in the regulation of the *nif* and *hut* operons. Mol. Gen. Genet. **18**:3289–297

Doyle, D., and Schimke, R. T. (1969) The genetic and developmental regulation of hepatic Δ-aminolevulinic acid dehydrase in mice. J. Biol. Chem. **244**:5449-5459

Drolet, M., Peloquin, L., Echelard, Y., Cousineau, L., and Sasarman, A. (1989) Isolation and nucleotide sequence of the *hemA* gene of *Escherichia coli.* Mol. Gen. Genet. **216**:347-352

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

Elberg, H., Mohr, J., and Nielsen, L. S. (1983) δ-Aminolevulinate dehydrase:synteny with ABO-AKI-ORM (and assignment to chromosome 9). Clin. Genet. **23**:150-154

Elledge, S. J., and Walker, G. C. (1985) Phasmid vectors for identification of genes by complementation of *Escherichia coli* mutants. J. Bacteriol. **162**:777-783

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

Epstein, O., Lahva, M., Schoenfeld, N., Nemesh, L., Shakil, M., and Atsmon, A. (1983) Erythrocyte uroporphyrinogen synthase

activity as a possible diagnostic aid in the diagnosis of lymphoproliferative disease. Cancer **52**:828-832

Fagard, R., and London, I. M. (1981) Relationship between phosphorylation and activity of heme-regulated eukaryotic initiation factor 2α–kinase. Pro. Natl Acad. Sci. U. S. A. **78**:866-870

Ford, S. H., and Friedmann, H. C. (1979) Formation of δ-aminolevulinic acid from glutamic acid by a partially-purified enzyme system from wheat leaves. Biochim. Biophys. Acta. **569**: 153-159

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

Fischer, H., and Orth, H. (1937) *Die Chemie des Pyrrols,* Vol. II.1 Akademische Verlagsgesellschaft, Leipzig

Fischer, H., and Stern, A. (1940) *Die Chemie des Pyrrols,* Vol. II.1 Akademische Verlagsgesellschaft, Leipzig

Gellerfors, P. L., Saltzgaber-Muller, J., and Douglas, M. G. (1986) Selection by genetic complementation and characterization of the gene coding for the yeast porphobilinogen deaminase. Biochem. J. **240**:673-677

Gibbs, P. N. B., and Jordan, P. M. (1986) Identification of lysine at the active site of human 5-aminolevulinate dehydratase. Biochem. J. 236:447-451

Gibbs, P. N. B., and Jordan, P. M. (1981) 5-Aminolevulinate dehydratase: zinc-65 binding and exchange with the enzyme from human erythrocytes. Biochem. Soc. Trans. **9**:232-233

Gibson, K. D., Neuberger, A., and Scott, J. J. (1955) The purification and properties of Δ-aminolevulinic acid dehydrase. Biochem. J. **61**:618-629

Gibson, K. D., Laver, W. D., and Neuberger, A. (1958) Initial stages in the biosynthesis of porphyrin. II. Formation of δ-aminolevulinic acid from glycine and succinyl CoA by particles from chicken erythrocytes. Biochem. J. **61**:618-620

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

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

Grandchamp, B., De Verneuil, H., Beaumont, C., Chretien, S., Walter, O., Nordmann, Y. (1987) Tissue-specific expression of porphobilinogen deaminase. Eur. J. Biochem. **162**:105-110

Granick, S. (1954) Enzymatic conversion of Δ-aminolevulinic acid to porphobilinogen. Science **120**:1105-1106

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier,R. (1981) Codon catalogue usage is a genome strategy modulated for gene expressivity. Nucl. Acid Res. **9**:r43-r73

Gicquel-Sanzey, B., and Cossart, P. (1982) Homologies between different prokaryotic DNA-binding regulatory proteins and between their sites of action. EMBO J. **1**:591-595

Grosjean, H., and Fiers, W. (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199-209.

Guest, J. R. (1969) Biochemical and genetic studies with nitrate reductase C-gene mutants of *Escherichia coli* . Mol. Gen. Genet. **105**:285-297

Gurba, P. E., Sennett, R. E., and Kobes, R. (1972) Studies on the mechanism of action of Δ-aminolevulinate dehydratase from bovine and rat liver. Arch. Biochem. Biophys. **150**:130-136

Hadley, R. G., Hu, M., Simmons, M., Yun, K., and Deonier, R. C. (1983) A partial restriction snap of the *proA–purE* region of the *E. coli* K–12 chromosome. Gene **22**:281-287.

Harel, E. and Neeman, E. (1983) Alternative routes for the synthesis of 5-aminolevulinic avid in maize leaves. Plant Physiol. **72**:1062-1067

Hart, G. J., Miller, A. D., Leeper, F. G., and Battersby, A. R. (1987) Biosynthesis of the natural porphyrins: proof that hydroxymethylbilane synthase (porphobilinogen deaminase) uses a novel binding group in its catalytic action. J. Chem. Soc. Chem. Commun. **109**:1762-1765

Hoober, J. K., Kahn, A., Ash, D. E., Gough, S., and Kannangara, C. G. (1988) Biosynthesis of Δ-aminolevulinate in greening barley leaves. IX. structure of the substrate, mode of gabaculin

inhibition, and the catalytic mechanism of glutamate 1-semialdehyde aminotransferase. Carlberg Res. Commun. **53**:11-25

Huang, D.-D., and Wang, W.-Y. (l986) Chlorophyll biosynthesis in *Chlamydomonas* starts with the formation of glutamyl-tRNA. J. Biol.Chem. **261**: 13451-13455

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

Javor, G. T. (1985) Thiol-stimulated secretion of riboflavin and porphyrins by *Escherichia coli.* FEMS Microbiology Letters **27**:243-245

Javor, G. T., and Kim, H. (1989) A simple method for screening porphyrin secretion by colonies of *Escherichia coli* . FEMS Microbiology Letters **56**:195-198

Jeter, R., Escalante-Semerena, J. C., Roof, D., Olivera, B., and Roth, J. (1987) Synthesis and use of vitamin B_{12} in *Salmonella* and *E. coli.* In: Ingraham, J. L., Low, K. B., Neidhardt, F. C., Magasanik, B., Schaechter, M., and Umbarger, H. E. (eds) *Escherichia coli* and *Salmonella typhimurium.* Cellular and molecular biology, vol 1. American Society for Microbiology. Washington,DC, pp 551-556

Jordan, P., and Berry, A. (1980) Preuroporphyrinogen, a universal intermediate in the biosynthesis of uroporphyrinogen III. FEBS Lett. **112**:86-88

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

Jordan, P. M., and Warren, M. J. (1987) Evidence for a dipyrromethane cofactor at the catalytic site of *E. coli* porphobilinogen deaminase. FEBS Lett. **225**:87-92

Jordan, P.M., Mgbeje, B.I.A., Thomas, S.D., and Alwan, A.F. (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., Warren, M. J., Williams, H. J., Stolowich, N. J., Roessner, C. A., Grant, S. K., Scott, A. I. (1988) Identification of a cysteine residue as the binding site for the dipyrromethane cofactor at the active site of *Escherichia coli* porphobilinogen deaminase. FEBS Lett. **235**:189-193

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

Kannangara, C. G., and Gough, S. P. (1977) Synthesis of Δ-aminolevulinic acid and chlorophyll by isolated chloroplasts. Carlsberg Res. Commun. **42**:441-457

Kannangara, C. G., and Gough, S. P. (1978) Biosynthesis of Δ-aminolevulinate in greening barley leaves: glutamate 1-semialdehyde transferase. Carlsberg Res. Commun. **43**:185-194

Kannangara, C. G., Gough, S. P., Bruyant, P., Hoober, J. K., Kahn, A., and Wettstein, D. (1988) $tRNA^{glu}$ as a cofactor in δ-aminolevulinate biosynthesis. Trends in Biochem Sci. **13**:139-143

Kappas, A., Sassa, S., and Anderson, K. E. (1983) The porphyrias. In Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. and Brown, M. S. (Eds) The metabolic basis of inherited diseases, 5th ed. McGraw-Hill, New York, pp.1301-1384

Kaslow, D. C. (1986) A rapid biochemical method for purifying lambda DNA from phage lysates. Nucl. Acids Res. **14**:6767

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

Kikuchi, G., Kumar, A., Talmage, P., and Shemin, D. (1958) The enzymatic synthesis of δ-aminolevulinic acid. J. Biol. Chem. **233**: 1214-1219.

Klein, O. and Senger, H. (1978a) Biosynthetic pathways to δ-aminolevulinic acid induced by blue light in the pigment mutant C-2A' of *Scenedesmus obliquus.* Photochem. Photobiol. **27**:203-208

Klein, O., and Senger, H. (1978) Two biosynthetic pathways to δ-aminolevulinic acid in a pigment mutant of the green alga, *Scenedesmus obliquus.* Plant Physiol. **62**:10-13

Kredich, N. M. (1983) Regulation of cysteine biosynthesis in *Escherichia coli* and *Salmonella typhimurium.* In: Herrmann, K. M., Somerville, R. L. (eds) Amino acids: biosynthesis and genetic regulation. Addison Wesley, Reading, MA, pp115-132

Kushner, S. R., Nagaishi, H., Templin, A., and Clark, A. J. (1971) Genetic recombination in *Escherichia coli* : the role of exonuclease I. Proc. Natl. Acad. Sci. U. S. A. **68**:824-827

Kuster, W. (1912) Beitrage zur kenntnis des bilirubins and hamins. Hoppe-Setler's Z. Physiol. Chem. **82**:463-483

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**:680-685

Lapointe, J., Levasseur, S., and Kern, D. (1985) Glutamyl-tRNA synthetase fron *Escherichia coli.* Methods in Enzymology **113**:42-49

Lascelles, J. (1964) Tetrapyrrole Biosynthesis and Regulation. W. A. Benjamin Inc., New York

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

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

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

Li, J.-M., Cosloy, S., and Russell, C. (1988b) 5-aminolevulinic acid synthesis in *Escherichia coli.* J. Cell Biol. 107:617a

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

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

Li, J.-M., Russell, C. S., and Cosloy, S.D. (1989c) Cloning and structure of *hemA* gene of *Escherichia coli* K-12. Gene. **82**:209-217

Liedgens, W., Lutz, C., and Schneider, H. A. W. (1983) Molecular properties of 5-aminolevulinic acid dehydratase from *Spinacia oleraiea.* Eur. J. Biochem. **135**:75-79

Lowry, C. V., and Lieber, R. H. (1986) Negative regulation of the *Saccharomyces cerevisiae ANB1* gene by heme, as mediated by the *ROX* 1 gene product. Mol. Cellular Biol. **6**:4145-4148

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

Maralihalli, G. B., Rao, S. R., and Bhagwat, A. S. (1985) Histidine residues at the active site of maize Δ-aminolevulinic acid dehydratase. Phytochemistry **24**:2533-2536

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

Mau, Y.-H., Wang, W.-Y., Tamura, R. N., and Chang, T.-E. (1987) Identification of an intermediate of δ-aminolevulinate biosynthesis in *Chlamydomonas* by high-performance liquid chromatography. Arch. Biochem. Biophys. **255**:75-79

Mauzerall, D., and Granick, S. (1956) The occurence and determination of δ-aminolevulinic acid and porphobilinogen in urine. J. Biol. Chem. **219**:435-446

Mayer, S. M., Beale, S. I., and Weinstein, J. D. (1987) Enzymatic conversion of glutamate to δ-aminolevulinic acid in soluble extracts of *Euglena gracilis.* J. Biol. Chem. **159**:12541-12549

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

McConville, M., and Charles, H. P. (1979a) Mutants of *Escherichia coli* K-12 accumulating porphobilinogen: a new locus, *hemC.* J. Gen. Microbiol. **111**:193-200

McConville, M., and Charles, H. P. (1979b) Isolation of haemin-requiring mutants of *Escherichia coli* K-12 . J. Gen. Microbiol. **113**:155-164

Meyer, V., Strand, L., Doss, M., Rees, A., and Marver, H. (1972) Intermittent acute porphyria: demonstration of a genetic defect in porphobilinogen metabolism. N. Engl. J. Med. **286**:1277-1282

Mignotte, V., Wall, L., deBoer, E., Grosveld, F., and Romeo, P.-H. (1989) Two tissue -specific factors bind the erythroid promotor of the human porphobilinogen deaminase gene. Nucl. Acids Res. **17**:37-54

Miller, A. D., Hart, G. J., Packman, L. C., and Battersby, A. R. (1988) Evidance that the pyrromethane cofactor of hydroxymehtylbilane synthase (porphobilinogen deaminase) is bound to the protein through the sulphur atom of cysteine-242. Biochem. J. **254**:915-918

Moss, G. P. (1988) IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), Nomenclature of tetrapyrroles. Recommendations 1986. Eur. J. Biochem. **178**:277-328

Mulligan, M. E., and McClure, W. R. (1986) Analysis of the occurence of promoter-sites in DNA. Nucl. Acids Res. **14**:109-126.

Myers, A., Crivellone, M. D., Koerner, T. J., and Tzagoloff, A. (1987) Characterization of the yeast *HEM2* gene and transcriptional regulation of *COX5* and *COR1* by heme. J. Biol. Chem. **262**:16822-16829

Nandi, D. L., and Waygood, E. R. (1967) Biosynthesis of porphyrin in wheat leaves. II Δ-aminolevulinate hydro-lyase. Can. J. Biochem. **45**:327-336

Nandi, D. L., and Shemin, D. (1968) Δ-Aminolevulinic acid dehydratase of *Rhodopseudomonas spheroides.* J. Biol. Chem. **243**:1231-1235

Nandi, D. L., and Shemin, D. (1973) Δ-Aminolevulinic acid dehydratase of *Rhodopseudomonas capsulata.* Arch. Biochem. Biophys. 158:305-311

Oh-hama, T., Seto,H. and Miyachi, S. (1986) ^{13}C-NMR evidence for bacteriochlorophyll *c* formation in the C_5 pathway in green sulfur bacterium, *Prosthecochloris.* Eur. J. Biochem. **159**: 189-194

Oh-hama, T., Seto, H. and Miyachi, S. (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. , Stolowich, N. J., and Scott, A. I. (1988) 5-aminolevulinic acid formation from glutamate via the C_5 pathway in *Clostridium thermoaceticum.* FEBS Lett. **228**:89-93

O'neill, G. P., Peterson, D. M., Schon, A., Chen, M.-V., and Soll, D. (1988) Formation of the chlorophyll precursor δ-aminolevulinic acid in *Cyanobacteria* requires aminoacylation of $tRNA^{glu}$ species. J. Bacteriol. **170**:3810-3816

Pabo, C. O., and Sauer, R. T. (1984) Protein-DNA recognition. Ann. Rev. Biochem. **53**:293-321

Pasanen, A. V. O., Vuopio, P., Borgatrom, G. H., and Tenhunen, R. (1981) Biosynthesis in refractory sideroblastic anemia associated with the preleukemic syndrome. Haematol. **27**:35-44

Perbal, B. A. (1984) Practical Guide to Molecular Cloning, John Wiley & Sons, New York, N.Y. pp 529–532

Peterson, D., Schon, A., and Soll, D. (1988) The nucleotide sequences of barley cytoplasmic transfer RNAs and structural features essential for formation of δ-aminolevulinic acid. Plant Mol. Biol. **11**:293-299

Porra, R. J. and Jones, O. T. G. (1963) Studies on ferrochelatase I. Assay and properties of ferrochelatase from a pig–liver mitochondrial extract. Biochem. J. **87**:181–185

Porra, R. J., Irving, E. A., and Tennick, A. M. (1972). The detection of δ-aminolaevulic acid synthetase in anaerobically-grown *Torulopsis utilis*. Arch. Biochem. Biophys. **149**: 563-565.

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

Proulx, M., and Lapointe, J. (1985) Purification of glutamyl-tRNA synthetase from *Bacillus subtilis*. Methods in Enzymology **113**:50-54

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

Raich, N., Romeo, P. H., Dubart, A., Beaupain, D., Cohen-Solal, M., and Goossens, M. (1986) Molecuar cloning and complete primary sequence of human erythrocyte porphobilinogen deaminase. Nucl. Acids Res. **14**:5955-5968

Rebeiz, C. A., Montazer-Zouhoor, A., Hopen, H. J., and Wu, S. M. (1984) Photodynamic herbicides.1. Concept and phenomenology. Enzyme Micro. Technol. **6**:390-401

Rebeiz, C. A. (1987) Ever green: Nature knows more than one way to paint her favorite color. The Sciences, Sept/Oct, 40-45

Rhee, H.-I., Murata, K., and Kimura, A. (1987) Formation of the herbicide, δ-aminolevulinate, from L-alanine and 4,5-dioxovalerate by *Pseudomonas riboflavina*. Agric. Biol. Chem. **51**:1701-1702

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

Rosenberg, M., and Court, D. (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. **13**:319-353

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

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

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

Sasaki, K., Ikeda, S., Nishizawa, Y., Hayashi, M. (1987) Production of 5-aminolevulinic acid by photosynthetic bacteria. J. Ferment. Technol. **65**:511-515

Sasarman, A., and Horodniceau, T. (1963) Utilisation de la neomycin pour la selection des variantes naines de *Salmonella.* Arch. Roum. Pathol. Exp. Microbiol. **22**:1101–1110

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

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

Sasarman, A., Surietta, M., and Horodniceanu, T. (1968b) Locus determing the synthesis of δ-aminolevulinic acid in *Escherichia coli* K-12. J. Bacteriol. **96**:1882-1884

Sasarman, A., Sanderson, K. E., Surdeanu, M., and Sonea, S. (1970) Hemin–deficient mutants of *Salmonella typhimurium.* J. Bacteriol. **102**:531–536.

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

Sasarman, A., Chartrand, P., Lavole, M., Tardif, D.,Proschek, R., and Lapointe, C. (1979) Mapping of a new *hem* gene in *Escherichia coli* K12. J. Gen. Microbiol. **113**:297-303

Sasarman, A., Nepveu, A., Echelard, Y., Dymetryszym, J., Drolet, M. and Goyer, G. (1987) Molecular cloning and sequencing of the *hemD* gene of *Escherichia coli* K-12 and preliminary data on the *uro* operon. J. Bacteriol. **169**:4257-4262

Sassa, S. (1982) Δ–Aminolevulinic acid dehydratase assay. Enzyme **28**:133–145.

Sassa, S., and Kappas, A. (1981) Genetic, metabolic, and biochemical aspects of the porphyrias. in *Advances in Human Genetics* (Harris, H., and Hirschhorn, K., eds) pp.121-231, Plenum Press, New York

Schleif, R. C., and Wensink, P. C. (1981) Practical Methods in Molecular Biology, Springer Verlag, New York pp.196–197

Schmid, R., and Shmin, D. (1955) The enzymatic formation of porphobilinogen from Δ-aminolevulinic acid and its conversion to protoporphyrin. J. Am. Chem. Soc. **77**:506-507

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

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

Schoenfeld, N., Mamet, R., Leibovici, L., Epstein, O., Teitz, Y., and Atsmon, A. (1988) Growth rate determines activity of porphobilinogen deaminase both in nonmalignant and malignant cell lines. Biochem. Med. and Metabolic Biol. **40**:213-217

Scott, A. I., Roessner, C. A., Stolowich, N. J., Karuso, P., Williams, H. J., Grant, S. K., Gonzalez, M. D., and Hoshino, T. (1988) Site-directed mutagenesis and high-resolution NMR spectroscopy of the active site of porphobilinogen deaminase. Biochem. **27**:7984-7990

Scott, A. I., Clemens, K. R., Stolowich, N. J., Santander, P. J., Gonzalez, M. D., and Roessner, C. A. (1989) Reconstitution of apo-porphobilinogen deaminase: structural changes induced by cofactor binding. FEBS Lett. **242**:319-324

Sharp, P.M., and Li, W-H. (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucl. Acid Res. **14**: 7737-7749.

Shemin, D., and Russell, C. S. (1953) δ-Aminolevulinic acid, its role in the biosynthesis of porphyrins and purines. J. Am. Chem. Soc. **75**:4873-4874

Shemin, D., Abramsky, T., and Russell, C. S. (1954) The synthesis of protoporphyrin from δ-aminolevulinic acid in a cell-free extract. J. Am. Chem. Soc. **76**:1204-1205

Shemin, D., Russell, C. S., and Abramsky, T. (1955) The succinate-glycine cycle. I. The mechanism of pyrrole synthesis. J. Biol. Chem. **215**:613-626

Shetty, A. S., and Miller, G. W. (1969) Δ-Aminolevulinate dehydratase from ox liver and tobacco leaves. Biochim. Biophys. Acta **185**:458-460

Shibata, H., and Ochiai, H. (1976) Purification and properties of Δ-aminolevulinate dehydratase from radish cotyledons. Plant Cell Physiol. **18**:421-429

Siegel, L. M., Murphy, M. J., and Kamin, H. (1973) Reduced nicotinamide adenine dinucleotide phosphate-sulfite reductase of enterobacteria. I. The *Escherichia coli* hemoflavoprotein: molecular parameters and prosthetic groups. J. Biol. Chem. **248**:251-264

Stella, A. M., and Batlle, A. M. del C. (1978) Porphyrin biosynthesis. Immobilized enzymes and ligands. VIII. Studies on the purification of Δ-aminolevulinate dehydrase from *Euglena gracilis.* Plant Sci. Lett. **11**:87-92

Strand, L., Felsher, B., Redeker, A., and Marver, H. (1970) Heme biosynthesis in intermittent acute porphyria: decreased hepatic conversion of porphobilinogen to porphyrins and increased delta aminolevulinic acid synthetase activity. Proc. Natl. Acad. Sci. U. S. A. **67**:1315-1320

Stubnicer, A.-C., Picat, C., and Grandchamp, B. (1988) Rat porphobilinogen deaminase cDNA: nucleotide sequence of the erythropoietic form. Nucl. Acids. Res. **16**:3102

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

Talt, G. H. (1972) 5-Aminolevulinate synthetase of *Micrococcus denitrificans.* Biochem. J. **128**:32P.

Tait, G. H. (1973) Aminolevulinate synthetase of *Micrococcus denitrificans*. Biochem. J. **131**: 389-403

Taylor, W. (1987) Protein structure prediction. In Nucleic Acid and Protein Sequence Analysis. pp285-322 IRL Press, Oxford

Thomas, S. D., and Jordan, P. M. (1986) Nucleotide sequence of the *hemC* locus envoding porphobilinogen deaminase of *Escherichia coli* K12. Nucl. Acids. Res. **14**:6215-6226

Troxler, R. F., and Offner, G. D. (1979) δ-Aminolevulinic acid synthesis in a *Cyanidium caldarium* mutant unable to make chlorophyll and phycobiliproteins. Arch. Biochem. Biophys. **195**:53-65

Tsukamoto, I., Yoshinaga, T., and Sano, S. (1979) The role of zinc with special reference to the essential thiol groups in Δ-aminolevulinic acid dehydratase of bovine liver. Biochim. Biophys. Acta **570**:167-168

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

Urban-Grimal, D., Volland, C., Garnier, T., Dehoux, P., and Labbe-Bois, R. (1986) The nucleotide sequence of the *HEM1* gene and evidence for a precursor form of the mitochondrial 5-aminolevulinate synthase in *Saccharomyces cerevisiae.* Eur. J. Biochem. **156**:511-519

Volland, C., and Urban-Grimal, D. (1988) The presequence of yeast 5-aminolevulinate synthase is not required for targeting to mitochondria. J. Biol. Chem. **263**:8294-8299

Wang, A.-L., Arredondo, F. X., Giampietro, P. F., Smith, M Anderson, W. F., and Desnick, R. J. (1981) Regional gene assignment of human porphobilinogen deaminase and esterase A4 to chromosome 11q23->11qter. Proc. Natl. Acad. Sci. U. S. A. **78**:5734-5738

Wang, A.-L., Astrin, K. H., Anderson, W. F., and Desnick, R. J. (1985) δ-aminolevulinate dehydratase induced expression and regional assignment of the human gene to chromosome 9q13->qter. Hum. Genet. **70**:6-10

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

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

Warnick, G. R. and Burnham, B. F. (1971). Regulation of porphyrin biosynthesis. Purification and characterization of δ-aminolevulinic acid synthase. J. Biol. Chem. **246**:6880-6885.

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

Weinstein, J. D., and Beale, S. I. (1985a) RNA is required for enzymatic conversion of glutamate to δ-aminolevulinate by extracts of *Chlorella vulgarie.* Arch. Biochem. and Biophys. **239**:87-93

Weinstein, J. D., and Beale, S. I. (l985b) Enzymatic conversion of glutamate to δ-aminolevulinate in soluble extracts of the unicellular green alga, *Chlorella vulgaris.* Arch. Biochem. Biophys. **237**:454-464

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

Weissberg, J. B., and Voytek, P. E. (1974) Liver and red cell porphobilinogen synthase in the adult and fetal guinea pig. Biochim. Biophys. Acta **364**:304319

Wilson, E. L., Burger, P. E., and Dowdle, E. B. (1972) Beef liver Δ-aminolevulinic acid dehydrase. Eur. J. Biochem. **29**:563-571

Wetmur, J. G., Bishop, D. F., Ostasiewicz, L., and Desnick, R. J. (1986a) Molecular cloning of a cDNA for human Δ-aminolevulinic acid dehydratase. Gene **43**:123-130

Wetmur, J. G., Bishop, D. F., Cantelmo, C., and Desnick, R. G. (1986b) Human Δ-aminolevulinic acid dehydratase: nucleotide sequence of a full-length cDNA clone. Proc. Natl. Acad. Sci. U. S. A. **83**:7703-7707

Wierenga, R. K., and Hol, W. G. J. (1983) Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant. Nature **302**:842-844

Woese, C. R. (1987) Bacterial evolution. Microbiol. Rev. **51**:221-271

Wright, M. S., Cardin, R. D. and Biel, A. (1987) Isolation and characterization of an aminolevulinate-requiring *Rhodobacter capsulatus* mutant. J. Bacteriol. **169**: 961-966.

Wu, W. H., Shemin, D., Richards, K. E., and Williams, R. C. (1974) The quaternary structure of Δ-aminolevulinic acid dehydrase from bovine liver. Proc. Natl. Acad. Sci. U. S. A. **71**:1767-1770

Wulff, D. L. (1967) δ-Aminolevulinic acid-requiring mutant from *Escherichia coli* . J. Bacteriol. **93**:1473-1474

Yamasaki, H., and Moriyama, T. (1971) Δ-Aminolevulinic acid dehydratase of *Mycobacterium phlei.* Biochim. Biophys. Acta **227**:698-705

Yamauchi, K., Hayashi, N., and Kikuchi, G. (1980) Cell-free synthesis of rat liver δ-aminolevulinate synthase and possible occurence of processing of the enzyme protein in the course of its translocation from the cytosol into the mitochondrial matrix. FEBS Lett. **115**:15-18

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G. J., Woese, C. R. (1985) Mitochondrial origins. Proc. Natl. Acad. Sci. U. S. A. **82**:4443-4447

Yanisch-Perron, C., Viera, J., and Messing, J. (1983) Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene. **33**:103-119.

Zagorec, M., and Labbe-Bios, R. (1986) Negative control of yeast coproporphyrinogen oxidase synthesis by heme and oxygen. J. Biol. Chem. **261**:2506-2509