Riboswitching of Upstream UTR of Heme Biosynthetic Genes and Cloning of *hemA* Gene of *Escherichia coli*

A

M. Sc Thesis

Submitted to

Central Department of Biotechnology

Tribhuvan University

Damodar Parajuli

Roll No. : BT 005/066

T.U. Regd. No. : 5-2-37-537-2005

Supervisor

Dr. Rajani Malla

Head of Department

Central Department of Biotechnology

Tribhuvan University

Kirtipur, Kathmandu, Nepal

Feb. 05, 2015

RECOMMENDATION

This is to certify that Mr. Damodar Parajuli has successfully completed the dissertation work entitled **"Riboswitching of Upstream UTR of Heme Biosynthetic Genes and Cloning of *hemA* Gene of *Escherichia coli*"** under my supervision.

This thesis work was performed for the partial fulfillment for award of Master of Science in Biotechnology Under the course title BT621. The results presented here is his own findings. I hereby, recommend this thesis for final evaluation.

…….

Dr Rajani Malla

(Supervisor)

Head of Department

Central Department of Biotechnoligy

Tribhuvan University

Kirtipur Kathmandu, Nepal

CERTIFICATE OF EVALUATION

This is to certify that thesis entitled **"Riboswitching of Upstream UTR of Heme Biosynthetic Genes and Cloning of *hemA* Gene of *Escherichia coli*"** presented to evaluation committee by **Mr. Damodar Parajuli** is found satisfactory for the partial fulfillment of Master of Science in Biotechnology.

……

Dr. Rajani Malla
(Supervisor)
Head of Department
Central Department of
Biotechnology
Tribhuvan University
Kirtipur, Kathmandu, Nepal

……

Dr. Megha Raj Banjara
(External Examiner)
Central Department of
Microbiology
Tribhuwan University
Kirtipur, Kathmandu, Nepal

……

Mr. Bal Hari Poudel
(Internal Examiner)
Central Department of
Biotechnology
Tribhuvan University
Kirtipur, Kathmandu, Nepal

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere and earnest compliments to my respected teacher, supervisor and head of the department **Dr. Rajani Malla** for her scholarly guidance, indispensable suggestion, inspiration, and worthy complements. I am also equally indebted to my respected supervisor Dr. Pramod Aryal, for his expert guidance, suggestion and instructions. I also highly appreciate him for kindly providing the place and resources to conduct the study.

I am highly obliged to all my respected teaches at the Central Department of Biotechnology for their valuable suggestions, encouragement and support.

My whole hearted thanks goes to all the team of supra inventica where all of the research has been conducted, with teamwork, guidance, encouragement, motivation and kind co-operation this work would have been incomplete.

I am really grateful to all the staffs at the Central Department of Biotechnology for their support and help.

I would like to express many thanks to my all friends of Central Department of Biotechnology for their valuable, suggestions, co-operation and help.

I owe special thanks to my family members and relatives for their precise help and support. I would like to additionally thank Keshav Raj Budha, Sushil Khanal, Sujina Mali, Rumika Maharjan, Suman Lama, Samidha Shah and Ayesha Dhungel for being part in Laboratory.

Finally I would like to take this opportunity to express my gratitude and respect to my family members for their deep love, inspiration, support and encouragement in each and every step of my life.

Date: Feb. 05, 2015

Damodar Parajuli

ABSTRACT

5-Aminolevulinic acid (ALA) is the main precursor for tetrapyrrole containing biological molecules that include chlorophyll, cytochrome C for respiration, cytochrome P450 for drug metabolism, seroheme for cysteine biosynthesis, corrin ring of vitamin B12 molecule, having implication in diverse biological system from eubacteria to plants and animals including mammals. One of the limitations in the pharmaceutical or chemical industries is to study the drug metabolism profiles of new pharmaceutical entities or new chemical compound that requires either rat liver microsome, S9 mix of phenobarbital (carcinogen) induced liver enzyme system, and eventually human liver microsomes. This system has been studied for the metabolism of carcinogens and their genotoxicity in bacterial system. However, in these bacterial expression systems the expression amount is compromised as it requires expensive ALA as the precursor. However due to tight regulation by different mechanisms with or without involving enzymes, these critically important molecules are very rarely synthesized by Microorganisms. Many works have been done to overexpress gene biosynthesizing ALA for expression of cytochrome P450 isoforms. There might also be RNA involved regulation in ALA biosynthetic genes and subsequent stages of heme biosynthesis by Riboswitch like mechanism. Here if new regulatory mechanism is stipulated the protein and chromosomal engineering will be designed for sequence playing role in Riboswitch like mechanism to mimic this effect. However, very less of the works are involved that has looked in to the untranslated RNA region and role of their folding structure that would reduce amount of ALA and final product. Here mfold structure is developed for 5'UTR of heme biosynthetic genes and their thermal stability and GC content is analysed. We selected upstream UTR of 10 heme biosynthetic genes as mentioned in ecocys. Out of 10 genes transcript detail of only five genes were available and for rest of five genes upstream 100 nucleotide from translation initiation site were selected for structural study. In heme biosynthetic pathway *hemN* and *hemH* showed folding structure that could regulate heme biosynthesis by Riboswitch like mechanism. Similarly UTR of *hemB* gene also showed more stable structure than that of other with length of nucleotide and GC percentage comparable. There was no sequence similarity of UTR of all heme Biosynthetic genes in *Escherichia coli* when studied through BLAST (alignment tool). We designed PCR

amplification primers of hemA gene of *E coli*. After PCR amplification heme gene is also been successfully cloned in pET15b expression vector and pUC19.

Key words: heme, hemA, UTR, riboswitch

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Glossary Acronyms

µ	micro
ALA	5-Aminolevulinic acid
ATP	Adenosine Tri Phosphate
BLAST	Basic Local Alignment Search Tool
bp	base pairs
cat	chloramphenicol acetyltransferase gene
CYP	Cytochrome P450
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
dsDNA	double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylene Diamine Tetra Acetic Acid
GLB	Gel Loading Buffer
GluTR	Glutamyl tRNA Reductase
hNPR	human NADPH-P450 Reductase
IDT	Integrated DNA Technologies
IPTG	Isopropyl Thiogalactopyranoside
kb	kilo base pairs
L	Liter
LB	Luria Bertani
m	milli

M	Molar
min	minute
NADPH	Nicotinamide Adenosine Diphosphate (H)
NCBI	National Center for Biotechnology Information
NEB	New England's Biolabs
OD_{600}	Optical Density at 600 nm
orf	open reading frame
ori	origin of replication
PALP	Pyridoxal Phosphate
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
rbs	ribosome binding site
RNase	Ribonuclease
rpm	revolutions per minute
RT	Room Temperature
SOC	Super Optimal Broth With Catabolite
TB	Terrific Broth medium
T_m	Melting Temperature
Tris	Tris-(hydroxymethyl)-aminmethane
U	Units
UV	Ultraviolet
v/v	volume per volume
w/v	weight per volume

CHAPTER ONE

INTRODUCTION

1.1 Background

Bioinformatics is an emerging interdisciplinary area of Science & Technology encompassing a systematic development and application of IT solutions to handle biological information by addressing biological data collection and warehousing, data mining, database searches, analyses and interpretation, modeling and product design. Being an interface between modern biology and informatics it involves discovery, development and implementation of computational algorithms and software tools that facilitate an understanding of the biological processes with the goal to serve primarily agriculture and healthcare sectors with several spinoffs.

In a developing country like Nepal, bioinformatics has a key role to play in areas like agriculture where it can be applied for increasing the nutritional content, increasing the volume of the agricultural produce and implanting disease resistance, pathway modeling, disease identification etc. In the pharmaceutical sector (van der *et al*., 1999), it can be used to reduce the time and cost involved in drug discovery process particularly for third world diseases, to custom design drugs and to develop personalized medicine (van der *et al*., 1999).

Further, bioinformatics can be used to formulate experimental strategy and could be useful for the prediction of possible outcome of formulated strategy. Information derived from certain group of organism about the regulatory mechanism of cellular process can be implied for similar group of organism and related phenomenon can be stipulated. Bioinformatics includes major areas such as Genome Analysis, Protein folding, Drug design, Metablomics, Pathway modulation etc which has enabled researchers for rational designing of experiments maintaining the cellular homeostasis.

These days new term metabolic engineering has developed as a multidisciplinary field which interlinks between the principles of chemical engineering, biochemistry, molecular and cell biology, computational sciences and mathematical modeling for targeted and

purposeful alteration of metabolic pathways found in an organism (Nudler *et al*., 2002 and 2004). This allows better understanding of cellular pathways and technology and utilize this information for modification to develop desired trait through recombinant DNA technology (Lessard 1996, Blazek and Alper 2010, Tang and Zhao 2009). In recent years tremendous progress towards the understanding of microbial metabolism on higher level of the entire functional system have been feasible (Kohlstedt et al 2010) with the advancement of X-ome technologies (transcriptome, metablome, proteome, fluxome) and advanced mathematical modeling tools such as genome scale metabolic engineering (Otero and Nielsen 2010) and bioinformatics is the key tool for this advancement.

Heme is the prosthetic group of hemoglobin, myoglobin, & cytochromes. Because of the methylation side chains in position within the tetrapyrrole ring structures, heme can be defined as an asymmetric molecule (Caughey *et al*., 1975). The tetrapyrrole ring system of heme could be synthesized from two pathaways, one depends on the precursor molecule such as glycine and succinyl-CoA known as C4 pathway and another utilizes precursor molecule glutamyl tRNA known as C5 pathway (Petricek *et al*., 2006). The tetrapyrrole biosynthetic pathway, which starts with glutamate is found in plants, many bacteria (including Escherichia coli), and the archaea. A second tetrapyrrole biosynthetic pathway, C4 pathway, is found in animals, fungi, certain protozoans and members of the α-proteobacteria, starts with glycine and succinyl-coA (Fig: 2.1). Regardless whether the starting point is glycine or glutamate, both pathways converge at the intermediate 5-amino-levulinate and proceed through the important intermediate uroporphyrinogen-III, which is a major branch point that leads to biosynthesis of different tetrapyrrole compounds, such as the corrinoid cobalamine (vitamin B12), the methanogenic coenzyme F430 , siroheme, heme D (Frankenberg and Lagarias 2003).

In C4 pathway, it was found that N and C atoms of heme are derived from glycine and acetate through experiment using isotopic tracers (Samuel *et a*l., 1974). In subsequent works conducted by using radiolabeled carbon by Paul *et al*. (1975), it was found that the labeled acetate enters Krebs Cycle as acetyl-CoA ultimately leading to biosynthesis of succinyl-COA, immediate precursor of heme. It has been now established that in C4 mechanism, which involves 4 carbon atoms, heme synthesis begins with condensation of glycine and succinyl-CoA, further by decarboxylation reaction results synthesis of δ -

aminolevulinic acid (ALA) (Moser *et al*., 1999) which is the main precursor for tetrapyrrole ring structure which found in many biologically important molecule such as sirroheme, chlorophyl, cobalamin ring of vitamin B12 etc (Paul *et al*., 1975).

ALA synthesis is the committed step of the heme synthesis pathway, and is usually rate-limiting for the overall pathway. Regulation of this step occurs through at transcriptional level in which heme is found to be controlling transcription by feedback inhibition mechanism and heme has ultimately repressing the transcription of the ALA synthase gene in most cells (Kannangara *et al*., 1988).

Most of bacteria like *Escherichia coli* synthesize ALA through C5 pathway from Glutamate. The synthesis of ALA from glutamate is a more complex process, requiring three separate enzymes (Kannangara *et al*., 1988). The first step in C-5 pathway is the charging of a glutamate-accepting tRNA (tRNAGlu) with glutamate, a step identical to that observed in protein synthesis. The reaction is performed by glutamyl-tRNA synthetase (EC 6.1.1.17) and requires ATP (Rauxa et al., 1880-1883), glutamate and tRNAGlu. There is report where over expression of glutamyl-tRNA synthetase could express cytochrome P450 isoforms without exogenous supplementation of ALA indicating that endogenous ALA synthesis can supplement for heme synthesis (Cynthia *et al*., 2007). The next step is a unique reaction seen nowhere else in any other metabolism, which is the reduction of the aminoacylated tRNAGlu to glutamate-1-semialdehyde (GSA). The enzyme that catalizes this reaction is a glutamate-tRNA dehydrogenase and requires NADPH as a coenzyme (Moser *et a*l., 1999). The final step in the synthesis of ALA from glutamate is a transamination reaction catalyzed by the enzyme GSA aminotransferase (EC 5.4.3.8)(Rauxa *et al*., 1880-1883).

In *Salmonella typhimurium* and *Escherichia coli*, heme plays a role both in respiration and in defense against the toxic oxygen metabolite H_2O_2. Heme b (Fe protoporphyrin IX) and various modified hemes are cofactors for a number of cytochromes as well as two catalases (Anraku *et al.,* 1987, Chiu *et al* 1989, Mogi et al 1994). The heme biosynthetic pathway also branches to produce two other tetrapyrroles: siroheme, the cofactor for sulphite and nitrite reductases (Kredich *et al* 1987, Warren *et al*., 1994), and cobalamin (vitamin B12). *S. typhimurium* synthesizes cobalamin de novo, but only during growth under anaerobic or low-oxygen conditions (Ailion *et al*., 1993, Jeter *et a.,l* 1987). Thus, the products of this

branched pathway have a variety of functions related to oxygen, respiration, and electron transfer.

1.2 Current Studies

The biochemistry of heme synthesis is well established in many organism, and, with the exception of the initial reactions leading to 5-aminolevulinic acid (ALA), the pathway is conserved among all organisms that make heme (reviewed by Beale 1966). In mammals, the mitochondrial enzyme d-aminolevulinate synthase (ALAS) (E.C. 2.3.1.37) provides the rate limiting step in the heme biosynthetic pathway. ALAS exists in two forms, an erythroid specific isozyme (eALAS), and a nonerythroid isozyme common to the liver and other tissues (ALAS1) (Marver 1966, Nakakuk 1980, Whiting *et al*., 1972 and Warnick *et al*., 1971). Natural regulation of the (hepatic) heme synthesis pathway occurs by end-product inhibition of the synthesis and mitochondrial translocation of ALAS (Ades *et al*., 1987, and Hayashi *et al*., 1969). The mitochondria exports d-aminolevulinic acid (ALA) to the cytoplasm for the subsequent steps in heme synthesis: tetrapyrrole assembly and side chain decarboxylation (Schobert *et al*., 2002). The final stages in heme synthesis: side chain oxidation and dehydrogenation, return to the mitochondria. Iron enters the heme cycle when mitochondrial ferrochelatase (E.C. 4.99.1.1) incorporates ferrous iron, Fe(II), into protoporphyrin IX (PPIX), generating protoheme (ferroprotoporphyrin IX, PPIX-Fe(II). The heme iron reversibly oxidizes to the ferric form, ferriprotoporphyrin IX, PPIX-Fe(III), which forms two complexes: hematin, PPIX-Fe(III)-OH^-, and hemin, PPIX-Fe(III)-Cl^- (Warnick *et al*., 1971).

But in case of many bacteria like *Escherichia coli*, the committed precursor in the heme pathway, δ-aminolevulinic acid (ALA) is formed in three steps from the five-carbon skeleton of glutamate (C5 pathway). The key C5 enzyme, glutamyl-tRNA reductase converts charged glutamyl-tRNAGlu to glutamate-1-semialdehyde (GSA) or its cyclic form. GSA is then converted to ALA by the hemL-encoded enzyme, glutamate-1-semialdehyde aminotransferase (reviewed by Pane et al 2002). Only a small fraction of the charged tRNAGlu of the cell is used to make heme, the reductase reaction is considered to be the first committed step in heme and tetrapyrrole biosynthesis.

Mechanism of riboswitch is also well established these days and till date more than 15 types of riboswitch has been identified as a new mechanism of gene regulation. (Nahvi *et al*., 2002, Henkin *et al*., 2011). In riboswitch mechanism there is no role of protein. Although there are different tools for identifying already existing Riboswitch (Ribosw, riboswitch finder), all tools are based on sequence of already discovered riboswitches. In this mechanism mostly 5'UTR region of gene is involved to control regulation by their secondary and tertiary structure in association with some other ligands. Majority of ligands are found to be the product of that gene.

In this study we cloned *hemA* gene of *E coli* into pUC19 and pET15b. pET15b is an expression vector on which we cloned *hemA* gene. Based on bioinformatics we examined UTR of all heme biosynthetic genes. *hemA* is considered as one of the most important gene in all tetrapyrole biosynthesis including heme. In laboratory basis we designed primers for hemA gene of E. coli and cloned in expression vector. We studied 5'UTR of all heme biosynthetic genes of same organism by employing different bioinformatics tools such as mfold, BLAST, STRING, ecocyc etc.

Generally this study analyzes *hemA* gene of *Escherichia coli*. *Escherichia coli* is most used organism in laboratory research and in biotechnology purposes. Riboswitch is a newly discovered regulatory mechanism and has posited widened paradigm for unraveling the mechanism of regulation of different genes. We had tried to compare structure of some known riboswitch with the folding pattern and structure of folding of 5' UTR of *hemA* gene. We primarily used mfold (default) to determine the folding structure of UTR and we compared the folded structure if the structure resembles with any of the previously discovered riboswitch structure. Current study mainly focused on determining the folding pattern of 5' untranslated region of ten genes involved in heme biosynthesis pathway and all of the structure have been analyzed for their stability on the basis of free energy value of secondary structure. Furthermore, GC percentage has also been analyzed to determine the level of stability in folding form. In addition, free availability of ribosome binding site (rbs) for ribosome in folded form has also been studied. On the basis of stability of secondary structure of UTR and free availability of rbs site, possible regulation of studied genes by riboswitch mechanism has been stipulated.

On the basis of this study we tried to generalize if riboswitch like regulation mechanism is involved in heme biosynthesis and B12 biosynthesis as both of the metabolites share same pathway up to uroprophyrinogen III.

1.3 Objectives

1.3.1 General Objective

This study mainly aims to study the 5' UTR of different genes involved in heme biosynthesis pathway so that a novel regulatory mechanism of heme biosynthesis could be stipulated, particulary focusing on riboswitch regulation mechanism. Further, this study also aims to employ different molecular biology tools for expression analysis *hemA* gene.

1.3.2 Specific Objectives

- Mechanism of Riboswtch is understood with some relationship with heme biosynthetic genes.
- Basic techniques of Primer designing, PCR programming and technique of cloning of gene into vector is carried out
- Stability of all structure UTR of heme biosynthetic gene is compared on the basis of GC content and Gibbs free energy value.
- These studies explore the possibility of comparing nucleotide sequences of organism which do not directly translated to protein but transcribed to RNA.

1.4 Rationale and Scope and limitation

hemA gene mediates the formation of 5 ALA in *Escherichia coli* by C5 pathway. Synthesis of ALA is highly regulated and it is considered to be a rate limiting stage in biosynthesis of B12, heme, chlorophyll as well as expression of cytochrome P450. Thus regulation of ALA biosynthesis at *hemA* gene level is of major interest for higher production of heme for cytochrome P450 expression and vitamin B12 biosynthesis. Cytochrome p450 plays important role in plant secondary metabolite synthesis as many terpenoid biosynthesis pathways are catalyzed by CYP. Furthermore, vitamin B12 is the most valuable commercial commodity. Therefore the proper understanding of regulation mechanism for heme biosynthesis could be helpful to posit a rational prototype for production of mentioned compound in desired quantity.

The 5’ UTR region of ALA synthetic gene (*hemA*) of *Escerichia coli*, upon structuring in MFold, exhibited stoichemistry that could possibly prevent access of 3’ end of mRNA of transcribed 5’-UTR to DNA template region for RNA polymerase which could possibly cause premature abortion of mRNA elongation resulting in truncated mRNA hence preventing translation of the protein. Comparision of 5’UTR of all genes responsible for heme biosynthesis provides clear role of the structure of UTR and regulation of reaction catalyzed by respected gene.

Currently many researches are focused on nucleotide sequences which are translated into protein. 5’ UTR of transcript that has critical role in regulation of expression of same gene without involvement of protein is because of its RNA structure, which can interact with some ligands or structure itself in presence of certain concentration of ligands, becomes more stabilized and prevents binding of ribosome in RBS region. Mechanism of riboswitch is because of folding of RNA but there are very few researches in UTR region of heme Biosynthetic genes. Being heme is critically important metabolite this research aids to explore the galaxy of benefits in riboswitch in drug designing that targets these genes.

CHAPTER TWO

LITERATURE REVIEW

2.1 Tetrapyroles

Tetrapyrrole class compounds are characterized by their four five-membered pyrrole rings, most often linked together via single atom bridges. The four rings of the macrocycle are labelled clockwise A–D starting with the first of the three symmetric rings with regard to the ring substituents. Two classes of cyclic tetrapyrroles are found in nature. The porphyrins, including heme, bacteriochlorophyll and chlorophyll are characterized by their completely saturated ring system. The second class of more reduced cyclic tetrapyrroles, the porphinoids, includes vitamin B12 (corrinoids), siroheme, coenzyme F430 and heme d1. In cyclic tetrapyrroles, the nitrogen atoms of the four pyrrole rings are used to chelate a variety of divalent cations. Tetrapyrroles are very distinct in color and amongst the most abundant molecules on earth. This is most visible in the green color of plants, which is due to their high chlorophyll content. In photosynthetic bacteria and plants, Mg-chelating chlorophylls and bacteriochlorophylls are the key pigments in trapping sunlight for photosynthesis (Beale 1999; Vavilin *et al*., 2002).

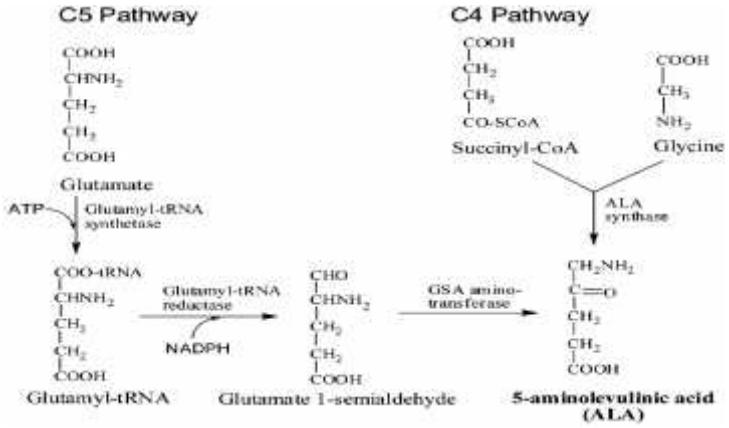


Fig 2.1(a) Synthesis of 5 ALA from two pathway (Moser *et al*., 1999)

The nickel-containing yellow coenzyme F430 is the prosthetic group of methyl-coenzyme M reductase that catalyzes the final step of methane formation in methanogenenic archaea (Thauer and Bonacker 1994). The pink cobalt-containing vitamin B12 derivatives are the

most complex known tetrapyrroles (Martens *et al*. 2002). They are involved in multiple enzymatic reactions, e.g., radical dependent nucleotide reduction and methyl transfer. The iron-chelating yellow-greenish siroheme is required for the six electron transfer reactions during assimilatory nitrite or sulfite reduction (Raux *et a*l. 2003). The green pigment heme d1, is part of the dissimilatory nitrite reductase in *Pseudomonas* and is a typical porphinoid that differs significantly in structure and color from the other porphyrin-based hemes (Chang 1994).

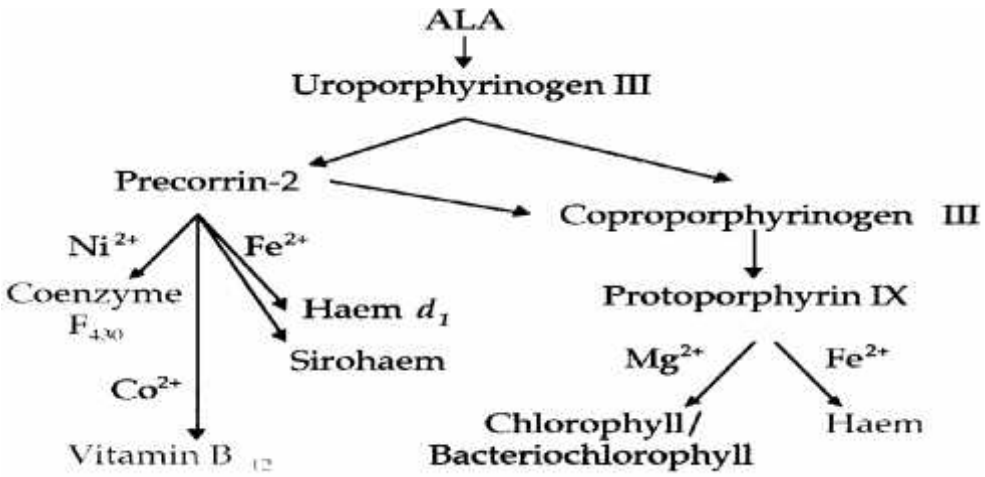


Fig 2.1(b) Fate of ALA to different products

The other commonly found tetrapyrrole structures are the open-chain molecules that are all derived from cleaved macrocycles. These can be divided into bile pigments, chlorophyll degradation products and phycobilins. The latter function as the direct chromophore precursor of the photoreceptor phytochrome and the cyanobacterial lightharvesting phycobiliproteins (for reviews see Hortensteiner 1999; Frankenberg and Lagarias 2003). Although most linear tetrapyrroles are derived from heme by oxidative cleavage, their nomenclature is not consistent with that of the porphyrin system. The various rings of linear tetrapyrroles are also labelled A–D, but starting from the cleavage position of the heme macrocycle.

In eukaryotes, the synthesis of tetrapyrroles is restricted to heme, siroheme, chlorophyll and bilins. Prokaryotes additionally form most complicated tetrapyrroles, such as corrinoids, heme d1 and coenzyme F430 (Vavilin and Vermaas 2002).

Among all tetrapyrole vitamin B12 has unique features in many aspects. Exceptional in comparison of vitamin B12 to other vitamins and coenzymes is as of its complexity, which is also reflected in its biosynthetic requirements such that somewhere around thirty genes are necessary for its complete de novo synthesis. Secondly, B12 is unique amongst the vitamins in that its synthesis is restricted to certain microorganisms. There is no genetic evidence that any eukaryote is able to make cobalamin. In contrast, in the prokaryotic world genome sequencing studies have revealed that the archae and certain eubacteria possess the genetic software that encodes the cobalamin biosynthetic enzymes.

2.2 ALA biosynthetic gene

Avissar and Beale (1989) first identified enzymatic basis for ALA auxotrophy in *hemA* mutant *E. coli*. This mutant strain was able to convert charged glutamyl tRNA to glutamate-1-semialdehyde and subsequently formed ALA (Figure 2.2) only when glutamyl tRNA reductase was exogenously supplemented. This confirmed the important role of *hemA* gene product in ALA biosynthesis. The involvement of glutamyl $tRNA^{Glu}$ as precursor molecule for biosynthesis of tetrapyrrole containing molecules also signified potential role of charged tRNA in metabolite biosynthetic pathway apart from universal dogma of protein biosynthesis in inserting respective amino acids in elongating peptide chain.

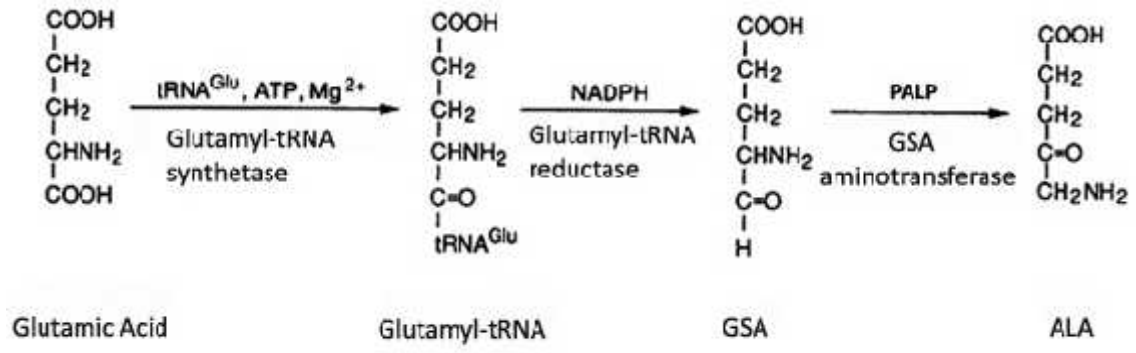


Figure 2.2: C5 pathway of ALA biosynthesis in *E. coli*. Chemical structure of intermediates, cofactors and enzymes involved in the reactions are illustrated (Adopted from Avissar and Beale, 1989).

Verkamp and Chelm (1989) reported that the open reading frame of E. coli glutamyl tRNA reductase (GluTR or HemA protein) encoded by hemA gene consists of 418 amino acids with a calculated molecular mass of 46 kDa (Figure 2.3). However, two proteins with distinct glutamyl tRNA reductase activities having molecular weights of 85 kDa (GluTR 85) and 45 kDa (GluTR 45), respectively have been reported in *E. coli* (Jahn et al., 1991). Later, it was elucidated that Glu-tRNA reductase can exist in multiple multimeric forms, but dimeric structure (GluTR 85) represents functionally active enzyme (Moser et al., 2001).

SN	**Title**	**Information**
1	Gene symbol	*hem*A
2	Gene description	glutamyl tRNA reductase (GTR reductase, GluTR, HemA)
3	Organism	*Escherichia coli* str. K-12 substr. MG 1655
4	Lineage	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Escherichia*
5	Also known as	ECK1198; gtrA; JW1201; b1210
6	Annotation	NC_000913.2 (1262937…1264193)
7	Length	1257 bp
8	Gene ID	945777
9	GI	16129173
10	ECOCYC	EG10427
11	EcoGene	EG10427
12	Protein	NP_415728.1
13	UniProtKB	P0A6X1, http://www.uniprot.org/uniprot/P0A6X1
14	Amino acid	418
15	Molecular weight	46.3kDa (calculated based on amino acid sequence in uniprot P0A6X1)

Table 2.1: *E. coli hemA* gene information derived from NCBI

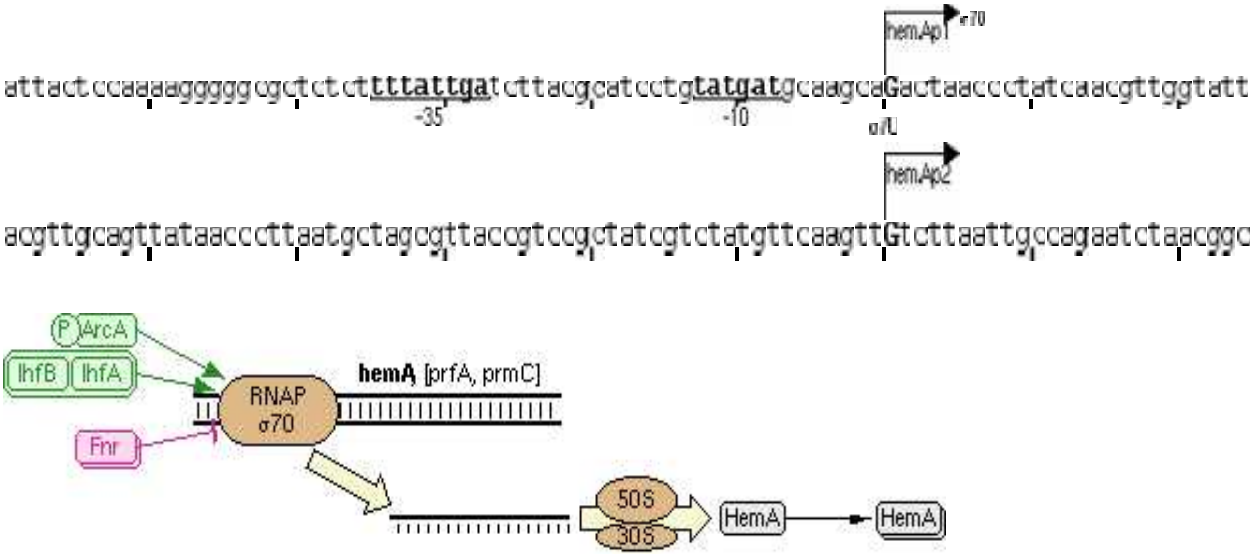


Figure 2.3: *hemA*, Sequence showing two promoters and gene map *E. coli hemA g*ene information

Some applications of ALA can be listed in the following table

Field	Application	References
Agriculture	Biodegradable herbicide	Rebeiz *et al*., 1984
	Biodegradable insecticide	Rebeiz *et al*., 1988
	Herbicide accelerator	Sasaki *et al*., 1998
	Growth promoting factor and yield enhancement	Tanaka *et al*., 1992; Hotta and Watanabe, 1999
	Salt tolerance	Kuramochi *et al*., 1997
	Cold temperature tolerance	Hotta and Watanabe, 1999
	Quality improvement of vegetables	Hotta and Watanabe, 1999
	Color intensifying effects	Hotta and Watanabe, 1999
	Growth and green-color maintenance of grass	Hotta and Watanabe, 1999
Medicine	Diagnosis of heavy-metal poisoning	Takeya *et al*., 1997
	Diagnosis of porphyria	Takeya *et al*., 1997

	Cancer treatment	Sasaki *et al*., 1998
	Diagnosis of brain tumor	Kaneko *et al*., 1998
	Treatment of rheumatoid arthritis	Trauner and Hassen, 1994
	Restore hair growth and prevent hair loss	Ishino *et al*., 1999
	Treatment of mycosis	Peter *et al*., 1994
	Cosmetic and dermatological applications	Uhlmann *et al*., 1996
Biotechnology	Porphyrin production	Miyachi *et al*., 1998
	Vitamin B12 production	Vladimir *et al*., 1997
	Plant cell culture	Masuda *et al*., 1994
	Animal cell culture	Fujita *et al*., 1997
	Genetically treated bacteria culture	Nakayashiki and Inokuchi, 1996
	Heme-containing enzyme production	Imai *et al*., 1993

Table 2.2: Potential applications of ALA in agriculture, medicine and biotechnology. (Sasaki *et al*., 2002)

2.3 Overview of heme biosynthetic genes and enzymes

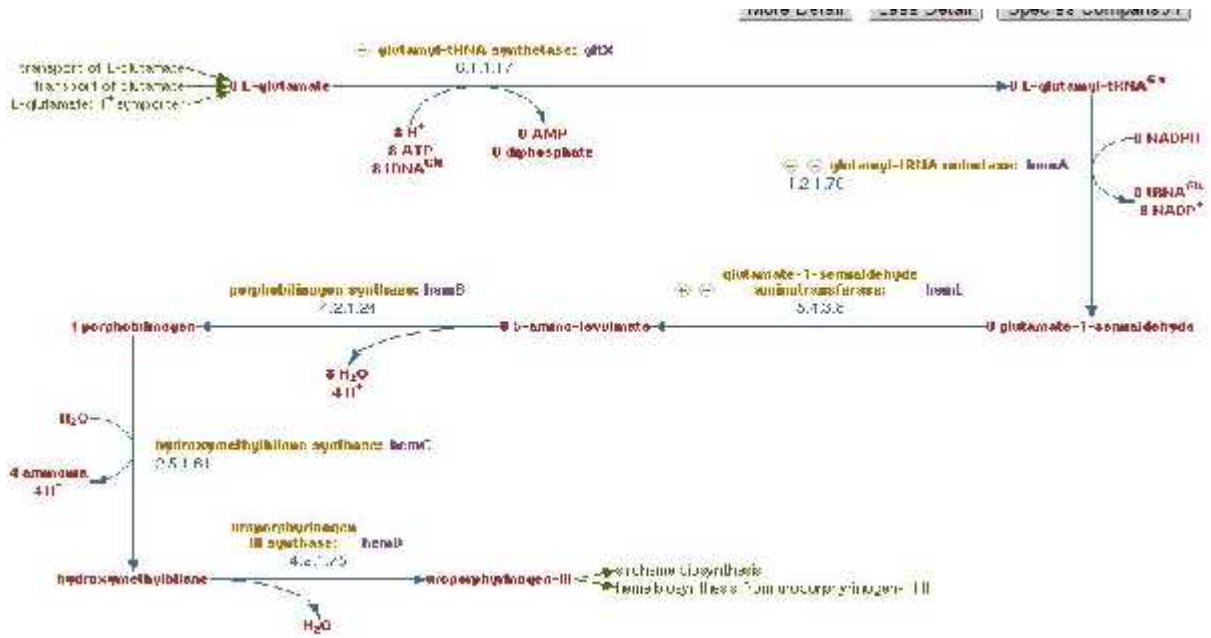


Fig 2.4 Biosynthesis of Uroporphyrinogen III from L Glutamate in *E coli*

Heme biosynthesis can be divided into two parts, first from glutamate to uroporphyrinogen III involving 6 genes and second from uroporphyrinogen III to protoheme IX. In Escherichia coli second part occurs via two different pathways differing in coproporphyrinogen III to protoprophyrin IX, where difference occurs only at two genes and enzymes only. In first pathway hemF coding coproporphyrinogen III oxidase converts coproporphyrinogen III to protoporphyrinogen IX with the removal of two carbondioxide and water each but in second pathway same conversion occurs via hemN coding dehydrogenase enzyme and uses two molecules of SAM. Next different mechanism is subsequent reaction protoporphyrinogen IX to protoporphyrin IX where in first pathway involves protoporphyrinogen oxidase coded by hemG gene and in second pathway with same gene and protein but with different mechanism with dehydrogenation. In heme biosynthesis from uroporphyrinogen-III this step occurs with fumarate as the final electron acceptor, the reaction is dependent on menaquinone as the electron carrier and is independent of cytochromes (Jacobs *et al* 1978).

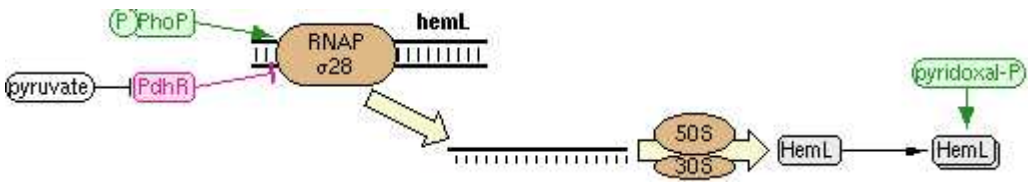
2.3.1 *gltx*(Glutamyl tRNA Reductase)

gltX is first gene involved in heme biosynthetic pathway codes for Glutamyl-tRNA synthetase (GluRS). GluRS is a member of the family of aminoacyl-tRNA synthetases, which interpret the genetic code by covalently linking amino acids to their specific tRNA molecules. The reaction is driven by ATP hydrolysis. GluRS belongs to the Class IB aminoacyl-tRNA synthetases (Eriani *et al*., 1990, Landes *et a*l., 1995). GluRS charges tRNAGlu for both protein and δ-aminolevulinic acid (ALA) synthesis. ALA is the first committed tetrapyrrole precursor (Neidhardt *et al.,* 1996). According to different research temperature-sensitive alleles of gltX have been isolated (Russell *et al*., 1971, Kaplan *et al.,* 1973, Masuda *et al* 1993). Further it has also been reported that expression of this gene may have regulated post transcriptionally (Brun *et al.,* 1990).

2.3.2 *hemL*

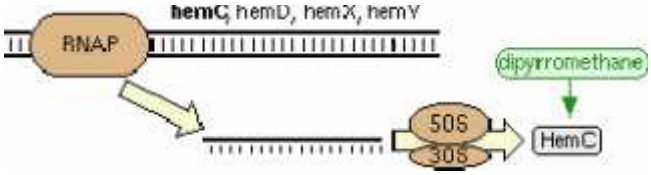
hemL in *E. coli* codes for Glutamate-1-semialdehyde aminotransferase which catalyzes the pyridoxal 5'-phosphate-dependent reaction and converts GSA to delta-aminolevulinate (ALA). ALA is the first committed precursor of porphyrin biosynthesis (Ilag *et al* 1991). The enzyme is homodimeric (Ilag *et al.,* 1991).

HemL forms a tight complex with glytamyl-tRNA reductase, the preceding enzyme in the pathway, suggesting metabolic channeling of the highly reactive intermediate glutamate-1-semialdehyde (Luer *et al*., 2005). According to (Minagawa *et al.,* 2003) Transcription of *hemL* is regulated by Mg2+ and PhoP .



2.3.3 *hemB*, *hemC* and *hemD*

Less research has been conducted on *hemB* which codes for porphobilinogen synthase and is active at octameric form. Porphobilinogen synthase (PBGS) catalyzes the synthesis of porphobilinogen from two molecules of δ-aminolevilinate via a Schiff-base intermediate (Spencer *et al*., 1993). PBGS is a metalloenzyme that requires Zn^{++} for activity. It is stimulated by Mg^{++} which is distinct from the zinc ions (Spencer *et al.,* 1993, Mitchell *et al.,* 1993, Jaffe *et al.,* 1995). Subsequent reaction is catalyzed by hemC and *hemD* to convert 4 molecules of PBG to hydroxymethylbilane and to uroporphyrinogenIII respectively. Both *hemC* and *hemD* are located on same operon in *E. coli* MG 1655. Both the *hemC* and *hemD* are overlapped with 3 nucleotides.



2.3.4 *hemE*

Second part of heme biosynthetic pathway starts with action of uooporphyrinogen III decarboxylase, a product of *hemE*. Uroporphyrinogen decarboxylase(UPD) catalyzes the decarboxylation of all four acetate residues of uroporphyrinogen III to generate coproporphyrinogen III. This is the first committed step after the branch point between heme and siroheme biosynthesis (Sasarman *et al*., 1975, Nishimura *et al.,* 1993, Ineichen *et al.,* 1993).

2.3.5 *hemF* and *hemN*

After action of hemE two enzymes coded by different genes are involved in conversion of coproporphyrinogen III to prorotoporphyrinogen IX, *hemF* and *hemN*. Former one is strictly aerobic and later one is activated on anaerobic condition. Coproporphyrinogen III oxidase (*HemF*) catalyzes the decarboxylation of coproporphyrinogen III to protoporphyrinogen IX. The enzyme is strictly aerobic, requiring molecular oxygen as the electron acceptor and

producing hydrogen peroxide (Troup *et al*., 1995, Breckau *et al*., 2003). His and Trp residues essential for catalytic activity were identified by kinetic analysis of mutant enzymes (Breckau *et al*., 2003).

E. coli contains two coproporphyrinogen III oxidases, one that is active under aerobic conditions, the hemF gene product, and one that is active under anaerobic conditions, the hemN gene product (Troup *et al*., 1995).

Over expression of *hemF* can suppress the growth defect of a *hemG* mutant; thus, *HemF* was hypothesized to be able to oxidize protoporphyrinogen IX as well (Narita *et al*., 1999). However, no such activity was found in vitro (Breckau *et al*., 2003).

hemN is monomeric when overproduced and purified under anaerobic conditions and might be membrane-associated in vivo. The enzyme contains an oxygen-sensitive [4Fe-4S] iron-sulfur cluster; mutations in specific residues important for iron-sulfur cluster coordination and catalysis have been analyzed and a catalytic mechanism has been proposed (Layer *et al*., 2002). HemN has been classified as a "Radical SAM enzyme" (Sofia *et al*., 2001). S-adenosylmethionine (SAM) is consumed during catalysis, and the role of a second SAM binding site in HemN has been investigated by site-directed mutagenesis (Layer *et al*., 2005). The catalytic mechanism has been studied in detail (Layer *et al*., 2006).

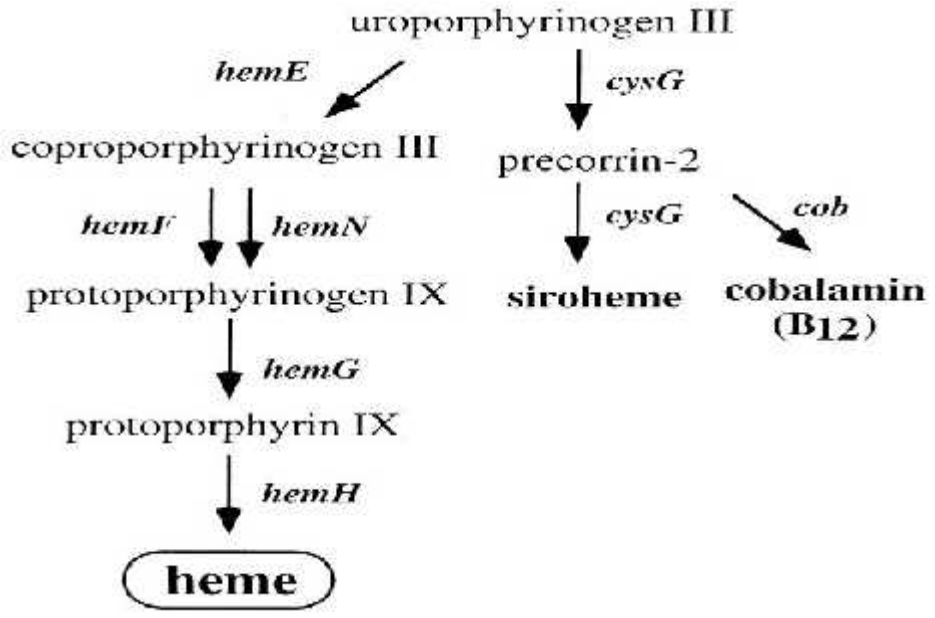
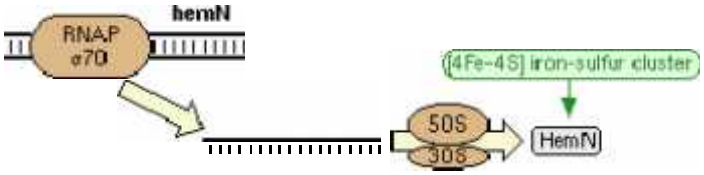
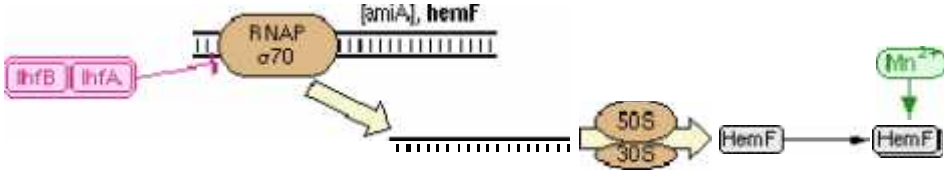


Fig 2.5: Diversion of Uroporphyrinogen III to different tetrapyroles

Expression of *hemN* is increased under anaerobic growth conditions and decreased under conditions of limited iron availability (Troup *et al*., 1995). hemN can complement a *Salmonella typhyimurium* hemF hemN double mutant under both aerobic and anaerobic growth conditions (Troup *et al*., 1995). A hemN mutant is defective for cytochrome b and cytochrome c synthesis only under anaerobic growth conditions (Tyson *et al*., 1997a).





2.3.6 *hemG*

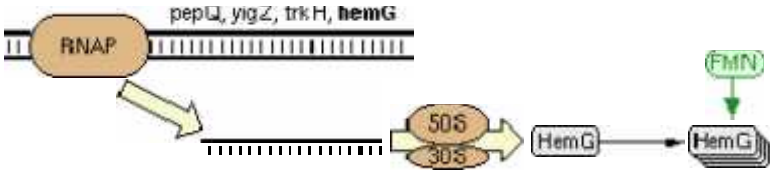
Protoporphyrinogen oxidase (PPO) catalyzes the six-electron oxidation of protoporphyrinogen IX to protoporphyrin IX. The enzyme belongs to the flavodoxin family of proteins, and the long chain insert loop which distinguishes it from other flavodoxins may be responsible for PPO activity (Boynton *et al.,* 2009).

Until recently (Boynton *et al.,* 2009), the E. coli enzyme had only been assayed in crude extracts, and the activity was associated with the membrane fraction (Jacobs *et al.,* 1978).

Under aerobic conditions, the electron acceptor for the protoporphyrinogen oxidase reaction was thought to be molecular oxygen (Jacobs *et al.,* 1976). Under anaerobic conditions, the reaction can be coupled to nitrate or fumarate reduction within the anaerobic electron transport chain (Jacobs *et al*., 1976, Jacobs *et al*., 1977). Protoporphyrinogen oxidase is inactivated by treatment with detergent, suggesting that the

enzyme requires an intact electron transport system for activity (Jacobs *et al* 1984). In vitro, the purified enzyme can use menadione as the electron acceptor (Boynton *et al*., 2009).

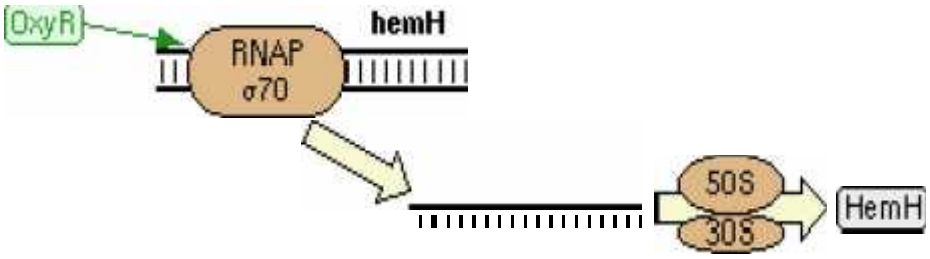
The presence of a flavodoxin motif in the sequence indicates that the enzyme may utilize FMN as a cofactor (Nishimura *et al*., 1995). A probable FMN cofactor was detected by MALDI-TOF, measurement of the UV-visible spectrum, and the EPR signal during redox titrations (Boynton *et al*., 2009).



2.3.7 *hemH*

Ferrochelatase is the terminal enzyme in the heme biosynthesis pathway and catalyzes the insertion of Fe^{2+} into protoporphyrin IX. Both eukaryotic and some microbial ferrochelatases have been shown to contain a [2Fe-2S] cluster (Dailey *et al*., 2002); however, no studies have been performed with the *E. coli* enzyme.

A *hemH* mutant is sensitive to visible light due to the accumulation of protoporphyrin IX. This is similar to the defect observed in human protoporphyria (Nakahigash *et al*., 1991, Miyamoto *et al*., 1992). The defect in iron incorporation into protoporphyrin IX in a *hemH* mutant results in a defect in assembly of membrane-associated succinate-ubiquinone reductase (Nihei *et al.,* 2001) Expression of *hemH* is mildly regulated in response to heme availability (Mc Nicholas et al 1997a) and is activated by OxyR (Zheng *et al*., 2001).



2.4 Function of Heme

Hemes are involved in many facets of oxidative metabolism, including O_2 transport, electron-transport dependent oxidative phosphorylation, oxidative stress responses, O_2-sensing, oxygenation reactions and detoxification. They are also involved in the synthesis or sensing of other diatomic gases such as carbon monoxide and nitric oxide (Rodgers 1999). As a regulatory molecule, heme mediates gene expression at the level of transcription (Schmitt 1999, Hach *et al*., 1999, Ogawa *et al*., 2001, Reddy *et al.,* 1996), translation (Chen & London, 1995), protein targeting (Lathrop & Timko, 1993), protein stability (Qi et al1999, Wang *et a*l., 1999) and differentiation (Nakajima *et al*., 1999).

Cytochromes are quantitatively the predominant heme proteins in most prokaryotes, and are essential for respiration. Prokaryotes can carry out oxidative respiration in the absence of oxygen using nitrate, sulfate or other oxidants. Cytochromes mediate electron transfer and the final reduction of the terminal electron acceptor as occurs with O_2 based respiration. Also, some prokaryotes respond to hypoxia by expressing cytochrome oxidases with high affinity for oxygen, and therefore oxygen limitation actually increases the demand for heme synthesis. Thus the need for heme is not restricted to aerobic metabolism at ambient O_2 levels as it is in eukaryotes; rather, heme contributes to metabolic adaptation. Indeed, hemes are found throughout the Bacteria and Archaea, and in most cases the demand is met by biosynthesis rather than by acquisition from the environment (Jacobs *et al*., 1996).

2.5 Untranslated Region

This term is commonly linked to mRNA. Typically this is the part of mRNA which is not translated to protein. The recent analysis of the human genome (Venter *et al*., 2001, Lander *et al*., 2001) and the data available about other higher eukaryotic genomes have revealed that only a small fraction of the genetic material which is about 1.5 % codes for protein. Indeed, most genomic DNA is involved in the regulation of gene expression, which can be exerted at either the transcriptional level, controlling whether a gene is transcribed or not and to what extent, or the post-transcriptional level, controlling the fate of the transcribed RNA molecules, including their stability, the efficiency of their translation and their sub

cellular localization. Matured mRNA is formed and regulated from several regulated factors like RNA polymerase and a series of cis-acting elements located in the DNA, such as promoters, enhancers, silencers and locus-control elements, organized in a modular structure and regulates the production of pre-mRNA molecules, which undergo several steps of processing before they become functional mRNAs. Introns are removed, a 7-methyl-guanylate (m7G) cap structure is added at the 5’ end of the first exon, and a stretch of 100-250 adenine residues (the poly(A) tail) is added at the 3’ end of the last exon, which is itself generated by endonucleolytic cleavage of the primary transcript. Sometimes the sequence of the mRNA is also altered in a process called mRNA editing, and the resulting coding sequence of the mature RNA differs from the corresponding sequence in the genome.

The resultant mature mRNA, in eukaryotes, has a tripartite structure consisting of a 5’ untranslated region (5’UTR), a coding region made up of triplet codons that each encode an amino acid and a 3’ untranslated region (3’UTR). Thus 5' UTR is the portion of a matured mRNA from the 5' end to the position of the first codon used in translation. The 3' UTR is the portion of an mRNA from the 3' end of the mRNA to the position of the last codon used in translation (Van der *et al*., 1999).

2.5.1 Role of UTR

UTRs are known to play crucial roles in the post-transcriptional regulation of gene expression, including modulation of the transport of mRNAs out of the nucleus and of translation efficiency (van der Velden *et al*., 1999), subcellular localization (Jansen 2001) and stability(Bashirullah *et al*., 2001), Regulation by ligand binding (Mironov *et al*., 2002). We are focusing mainly on ligand binding and stability functions, but UTRs may also play other roles, such as the specific incorporation of the modified amino acid selenocysteine at UGA codons of mRNAs encoding selenoproteins in a process mediated by a conserved stem-loop structure in the 3’ UTR (Walczak *et al*., 1996). The importance of UTRs in regulating gene expression is underlined by the finding that mutations that alter the UTR can lead to serious pathology (Conne *et al*., 2000).

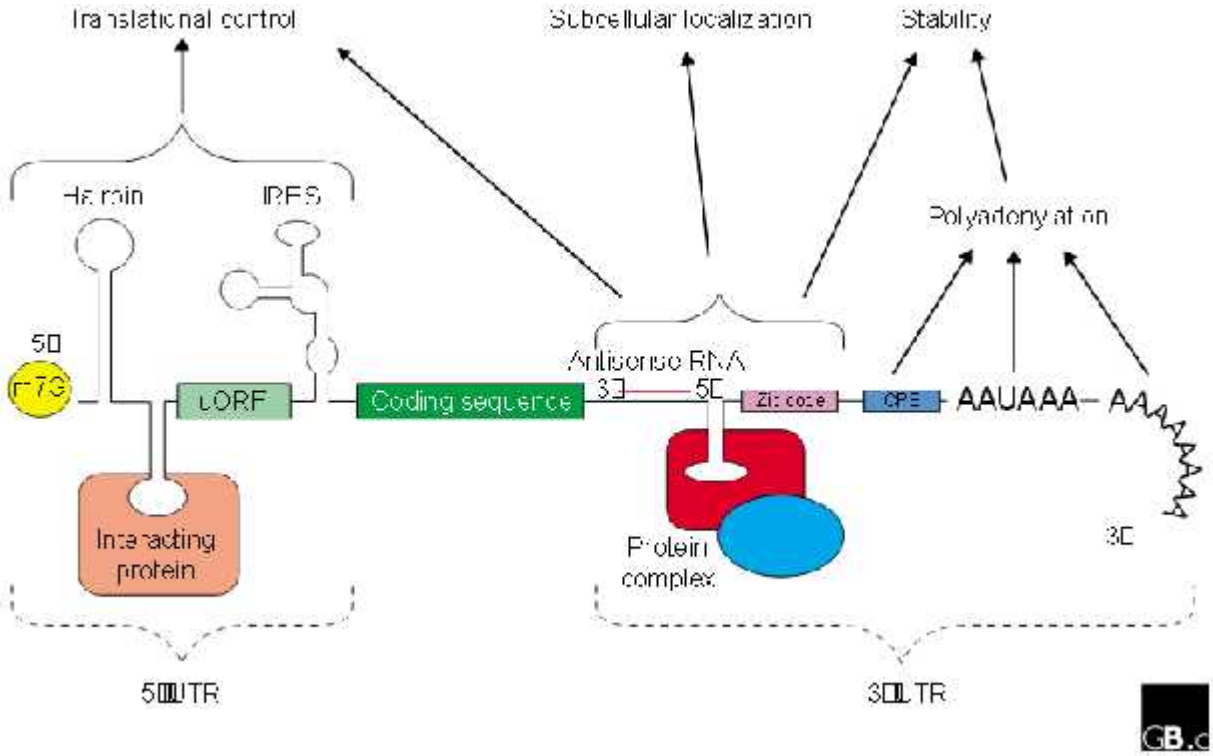


Fig 2.6 The generic structure of a eukaryotic mRNA, illustrating some post-transcriptional regulatory elements that affect gene expression.

2.6 Riboswitches

In early decade of 21st centuary first comprehensive proofs of multiple classes of riboswitches were published, including protein-free binding assays, and metabolite-binding riboswitches were established as a new mechanism of gene regulation (Mironov *et al*., 2002, Winkler *et al*., 2002a and 2002b, Nahvi *et al*., 2002). Subsequent studies on the expression of genes for the biosynthesis of nucleotides (purines), and vitamins (thiamin, riboflavin, cobalamin) in Bacillus subtilis and E. coli have given insights into a riboswitch mechanism of gene regulation in these bacteria (Vitreschak *et al*., 2004). Riboswitches are made up of the three-dimensional structure of RNA, in which RNA can undergo two mutually exclusive conformations in response to an environmental signal in the form of a metabolite.

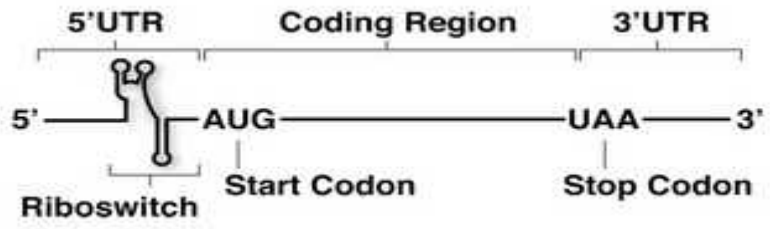


Fig 2.7: Tripartite Structure of gene showing location of UTR

Genetic regulation by RNA is widespread in bacteria. One common form of riboregulation in bacteria is the use of ribonucleic acid sequences encoded within mRNA that directly affects the expression of genes encoded in the full transcript (called cis-acting elements because they act on the same molecule they're coded in). These regulatory elements are known as riboswitches and are defined as mRNA elements that bind metabolites or metal ions as ligands and regulate mRNA expression by forming alternative structures in response to this ligand binding (Nudler & Mironov 2004; Tucker & Breaker 2005; Winkler 2005). To date, scientists have discovered that a variety of ligands are sensed by riboswitches; the group includes magnesium ions, nucleic acid precursors, enzyme cofactors, and amino acid residues (Westhof 2007).

More than 2% of the genes in some species are regulated by riboswitches. Riboswitches have a remarkable affinity for their cognate ligands and can discriminate against even closely related analogues, as shown by biochemical experiments (Kaempfer 2003). Riboswitches typically contain disordered regions in their conserved aptamer cores that become structured upon metabolite binding. These changes may trigger rearrangements in additional expression platform structures located outside of the aptamer, such that two alternative conformations with mutually exclusive base-paired architectures exist for the entire riboswitch. Some riboswitches operate at thermodynamic equilibrium (Rieder *et al*., 2007). They are able to interconvert between these ligand-bound and ligand-free structures in the context of the full-length RNA. Regulation by other riboswitches is kinetically controlled (Wickiser *et al*., 2005, Gilbert *et al*., 2006, Wickiser *et al*., 2005a, Lemay *et al*., 2006). The relative speeds of transcription and co-transcriptional ligand binding dominate a one-time decision as to which folding pathway to follow. The active and inactive conformations of these riboswitches are trapped in the final RNA molecule and do not

readily interconvert on a time scale that is relevant to the gene control system (Mandal *et al*., 2004).

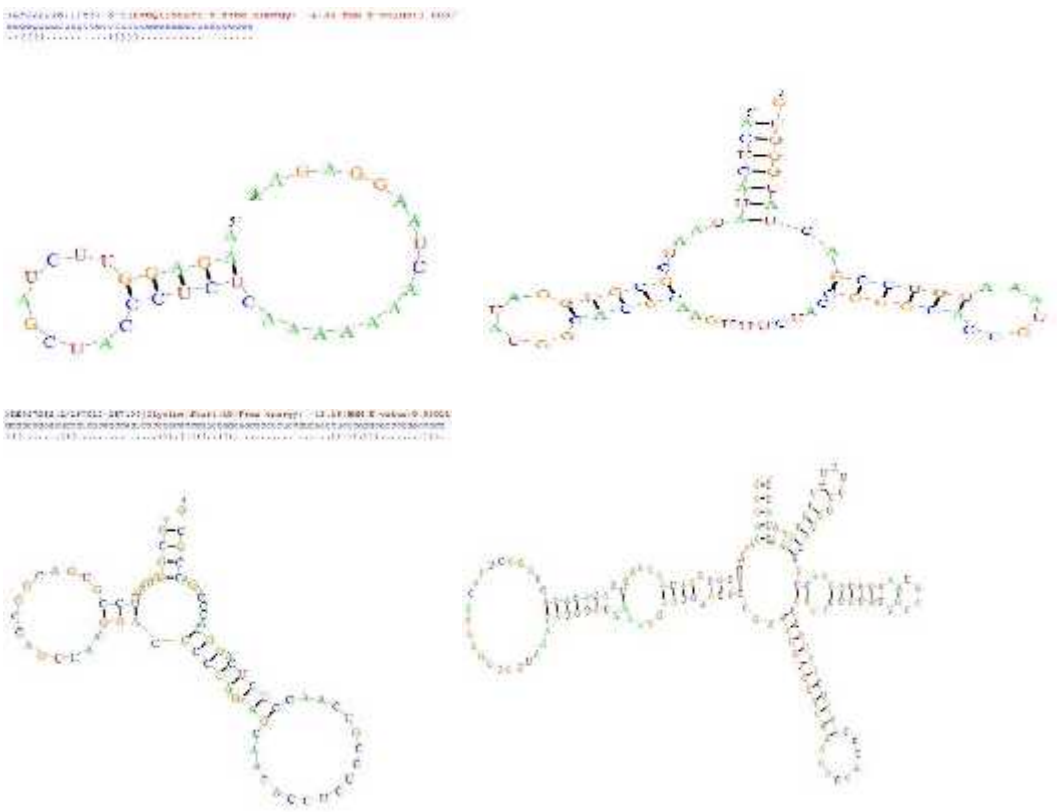


Fig: 2.8 Structure of PreQ, Purine1, Glyciene and lysiene Riboswitches

2.6.1 Location of Riboswitches

Riboswitches are most often located in the 5' untranslated region (5' UTR; a stretch of RNA that precedes the translation start site) of bacterial mRNA. There they regulate the occlusion of signals for transcription attenuation or translation initiation. However, not all riboswitches are at the 5' UTR; scientists have discovered that, in some eukaryotic mRNA, the thiamine pyrophosphate (TPP) riboswitch regulates splicing at the 3' end (Wachter *et al*., 2007).

2.6.2 Structure of Riboswitches

Riboswitches are composed of two domains: the aptamer domain and the expression platform (Figure 1; Tucker & Breaker 2005). The aptamer domain acts as a receptor that

specifically binds a ligand (Figure below). The expression platform acts directly on gene expression through its ability to toggle between two different secondary structures in response to ligand binding. Common to both domains is something called the switching sequence, and its placement in the aptamer domain or the expression platform ultimately dictates the expression outcome of the mRNA (Figure 2.9). Specifically, if metabolite binding to the riboswitch stabilizes incorporation of the switching sequence into the aptamer domain, the expression platform must fold into a specific structure. Riboswitches that control transcriptional repression have a switching sequence that directs formation of a Rho-independent transcriptional terminator, a short stem-loop structure (followed by six or more uridine residues) that signals RNA polymerase to abort transcription (Hammann & Westhof 2007). Other riboswitches that regulate translational initiation utilize a switching sequence that can expose or occlude a ribosome-binding site (called the Shine-Dalgarno sequence; Hammann & Westhof 2007).

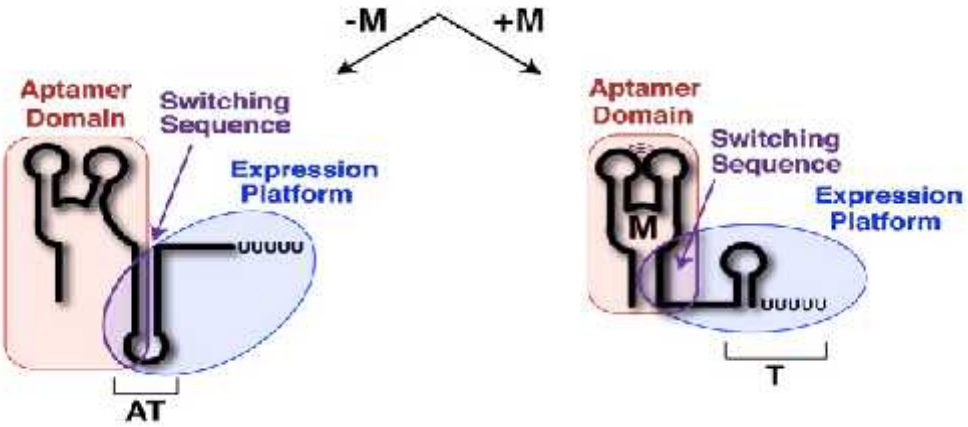


Figure 2.9: Typical structure of Riboswitch with presence and absence of ligand

2.6.3 Occurance of Riboswitch System

Riboswitches are autonomous noncoding RNA elements that monitor the cellular environment and control gene expression. More than a dozen classes of riboswitches that respond to changes in the concentrations of specific small molecule ligands ranging from amino acids to coenzymes are currently known. These metabolite-binding riboswitches are classified according to the architectures of their conserved aptamer domains, which fold

into complex three-dimensional structures to serve as precise receptors for their target molecules. Riboswitches have been identified in the genomes of archaea, fungi, and plants; but most examples have been found in bacteria.

2.6.4 Mechanism of Riboswitchs

Regulation by riboswitches does not require any macromolecular factors other than an organism's basal gene expression machinery. Metabolite binding to riboswitch aptamers typically causes an allosteric rearrangement in nearby mRNA structure that result in a gene control response. For example, bacterial riboswitches located in the 5' untranslated regions (UTRs) of messenger RNAs can influence the formation of an intrinsic terminator hairpin that prematurely ends transcription or the formation of an RNA structure that blocks ribosome binding. Most riboswitches inhibit the production of unnecessary biosynthetic enzymes or transporters when a compound is already present at sufficient levels. However, some riboswitches activate the expression of salvage or degradation pathways when their target molecules are present in excess. Certain riboswitches also employ more sophisticated mechanisms involving self-cleavage (Winkler *et al*., 2004), cooperative ligand binding (Mandal *et al*., 2004), or tandem aptamer arrangements (Welz *et al*., 2007).

In case of B. subtilis, the rib, thi, and xpt–pbuX operons (responsible for the synthesis of riboflavin, thiamin and guanine respectively) have been shown to be regulated by riboswitches. In case of the rib-operon, the 5’ UTR (untranslated region; ~300 nts) of the nascent transcript can fold into two mutually exclusive conformations in relation to its binding to FMN (flavin mononucleotide). In the low concentration/absence of FMN, the nascent transcript comes out of the RNA polymerase and folds into a confirmation that allows the transcription read through the proximal intrinsic termination site. As a result, the full transcript is formed and the genes in the operon are expressed. However, in the presence of FMN (1 mM), it’s binding to the nascent transcript folds the RNA in a way that results into transcription termination at the proximal intrinsic termination site (Ray 2004). This premature transcription termination results into no expression of genes in the operon. This is a new mechanism of attenuation mediated by riboswitch. Similar mechanism is used to control the gene expression in the thi-operon and xpt–pbuX operon in response to thiamin pyrophosphate (TPP) and guanine, respectively (Wickiser *et al*., 2005).

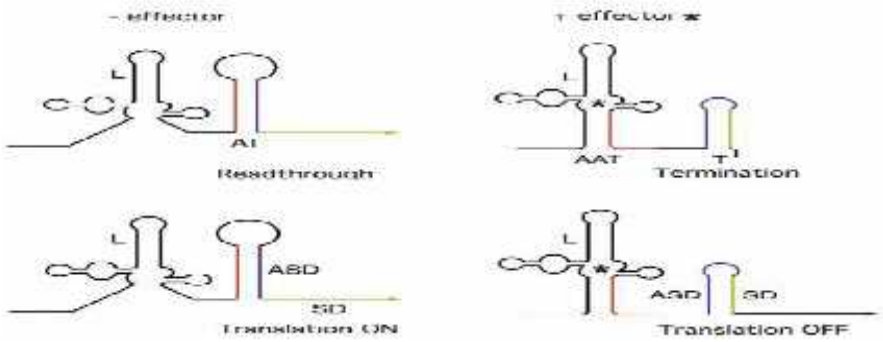


Fig 2.10: Mechanism of riboswitch action

2.6.5 Classification of Riboswitches

Riboswitches are organized into families and classes according to two features: the type of ligand they bind, and their secondary structure (the arrangement of Watson-Crick paired helices; Hammann & Westhof 2007; Montange & Batey 2008). A family of riboswitches is typically a group of RNAs related by the ligands they recognize. For example, the SAM riboswitch family recognizes the compound S-adenosylmethionine (SAM). Within a family, there may be distinct classes of riboswitches, each class distinguished by a common sequence pattern that usually defines the ligand-binding pocket, as well as features required for folding the RNA into a three-dimensional shape. The SAM riboswitch family contains at least five known classes (Wang & Breaker 2008). These classes are distinguished from one another by their architectural features. For example, the SAM-I class forms a four-way helical junction, SAM-II forms a classic (H-type) pseudoknot, and SAM-III is defined by a three-way junction (Montange & Batey 2006; Gilbert *et al*., 2008, Lu *et al*., 2008).

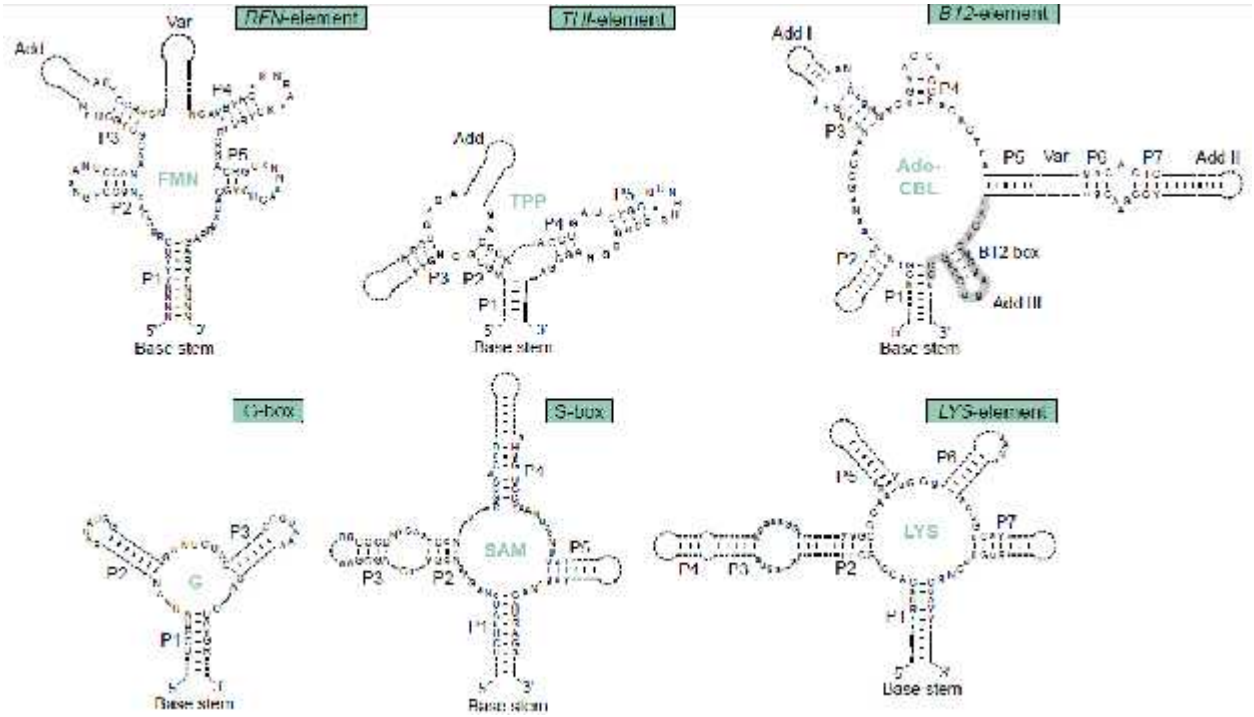


Fig 2.11: Different type of Riboswitches

Many aspects of riboswitch regulation have not yet been critically and quantitatively surveyed. The expanded riboswitch sequence alignments resulting from previous searches include newly identified variants that provide valuable information about their conserved aptamer structures (Ray 2004). The updated structures reveal that certain riboswitch aptamers utilize previously unrecognized examples of common RNA structure motifs as components of their conserved architectures. They also highlight new base-base interactions predicted with a procedure that estimates the statistical significance of mutual information scores between alignment columns (Winkler *et al*., 2004, Welz *et al*., 2007).

The function of riboswitches is tied to the ability of RNA to form a diversity of structures. The most basic of these is the double-stranded helix, similar to that found in DNA. However, since most RNAs, unlike DNA, do not need to maintain perfect Watson-Crick base pairing, they can form other types of structures. For example, a single strand of RNA can fold back on itself to form a hairpin, which is composed of a helix capped by a loop. These elements of structure are called secondary structure. In larger RNAs, the helices and hairpins pack together into a specific pattern, referred to as the tertiary structure.

Riboswitch	Ligand	Description	Structure
FMN	Flavin mononucleotide (FMN)	Mironov et al., 2002; Winkler et al., 2002a	
THI box	Thiamine pyrophosphate (TPP)	Mironov et al., 2002; Winkler et al., 2002b	Edwards and Ferre-D'Amare 2006; Serganov et al., 2006; Thore et al., 2006
B12	Adenosylcobalamin	Nahvi *et al*., 2002	
S box (SAM-I)	S-adenosylmethionine (SAM)	Epshtein *et al*., 2003; McDaniel *et al*., 2003; Winkler *et al*., 2003	Montange and Batey 2006
SAM-II	S-adenosylmethionine (SAM)	Corbino *et al*., 2005	Gilbert *et al*., 2008
SMK box (SAM-III)	S-adenosylmethionine (SAM)	Fuchs *et al*., 2006	Lu *et al*., 2008
SAH	S-adenosylhomocysteine (SAH)	Wang *et al*., 2008	
L box	Lysine	Grundy *et al*., 2003; Sudarsan *et al*., 2003b	Garst *et al*., 2008; Serganov *et al*.,2008
Glycine	Glycine	Mandal *et al*., 2004	
Purine	Guanine/adenine	Mandal *et al*., 2003; Mandal and Breaker 2004	Batey *et al*., 2004; Serganov *et al*., 2004
dG	Deoxyguanosine	Kim *et al*., 2007	
cyclic di-GMP	Cyclic di-GMP	Sudarsan *et al*., 2008	
glmS	Glucosamine-6-phosphate	Winkler *et al*., 2004	Klein and Ferre-D'Amare 2006; Cochrane *et al*., 2007
preQ1	7-Aminoethyl 7-deazaguanine	Roth *et al*., 2007	

Mg	Magnesium	Cromie *et al*., 2006; Dann *et al*., 2007	Dann *et al*., 2007

Source: Henkin TM, 2011

Table 2. 3: Classification of Riboswitch

2.7 Identification of riboswitch systems

Most riboswitch RNAs were initially recognized based on the identification of conserved RNA sequence and structural features in the 5' region of sets of bacterial genes known (or predicted) to respond in concert to a specific physiological signal. Identification of these patterns was initially done manually (e.g., Grundy and Henkin 1993, 1998) and gradually employed bioinformatics tools of increasing complexity (e.g., Gelfand *et al.,* 1999; Miranda-Rios *et al.,* 2001; Vitreschak *et al.,* 2002, 2003). Global searches for new RNA patterns that could potentially act as riboswitches (e.g., Bengert and Dandekar 2004; Abreu-Goodger and Merino 2005; Weinberg *et al.,* 2007) have expanded the list of riboswitch candidates and revealed variations on the pattern for known riboswitch RNAs. Most searches have focused on the 5'- untranslated region (UTR) in bacterial genomes; the identification of Thi-box elements in the 3_-UTR of plant genes (Sudarsan *et al.,* 2003a; Cheah *et al.,* 2007) illustrates the need for expanded searches, especially in eukaryotic genomes. The physical linkage of riboswitches to their target coding sequences greatly facilitates their characterization; for trans-acting regulatory RNAs, identification of the target gene(s) can present a major challenge (Vogel and Wagner 2007).

Heme and cobalamin share some common pathways and also have corrin ring in their structure. The ligand associated with corrin ring of cobalamin and heme are cobalt (Z=27) and iron (Z=26), respectively, Both are located on VIIB group of periodic table and on same period. Their paulings electronegativity value is also similar. During their synthesis after divergence from uroprophyrinogen III, cobalamin synthesis is regulated by riboswitch mechanism. In addition, the corrin ring needs to be methylated during ring formation by S-adenosylmethionine (SAM), and heme is also known to have methylation in its ring structure. SAM biosynthesis is regulated by riboswitch and carries either S-box

corresponding to SAM and B12 box corresponding to adenosylcobalamin. As heme also has to be methylated in its ring formation, and also carries tetrapyrrole ring, it can be assumed that the steps leading to heme formation from precorrin 2 or coprophyrorinogen could have similar mechanism riboswitch as that of B12 box which could be possibly termed as H box. Similarly, there might be some effect of heme concentration in heme biosynthesis as concentration cobalamin blocks its own synthesis by binding at 5' UTR of cobalamin synthesizing mRNA.

In accord with the fact that most riboswitches were identified initially by investigation of individual sets of genes or pathways, initial studies focused on analysis of the conditions to which each set of genes responds and the regulatory response (Grundy and Henkin 1993, Nou and Kadner 2000, Miranda-Rios *et al*., 2001). The recent identification of subclasses of elements that are present only in small groups of organisms (e.g., SMK box and SAM-II) (Corbino *et al*., 2005, Fuchs *et al*., 2006) or that exhibit extensive structural diversity and low conservation (e.g SMK box) (Fuchs *et al*., 2006) suggest that additional riboswitch elements remain to be uncovered. The rapidly growing data set of genomic sequences and improved bioinformatics techniques provide the necessary tools for their identification.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study site

Lab works were performed on Supra Inventica Laboratory Bansbari and Central Department of Biotechnology T.U. Kirtipur. Dry Laboratory work was performed using bioinformatics tools as mentioned on site for respective methods.

3.2 Ethical Consideration

Study was performed with the permission from Central Department of Biotechnology.

3.3 Chemicals and Restriction Enzymes

The sources of chemicals, restriction enzymes, DNA polymerase and other DNA modifying enzymes employed in this study were from following sources: 5-Aminolevulinic acid (ALA), Ribonuclease A and IPTG were purchased from Sigma-Aldrich, USA. Restriction enzymes (EcoRI-HFTM, NdeI, DpnI and HindIII-HFTM), Phusion High-Fidelity (HF) DNA polymerase and T4 DNA ligase were from New England Biolabs Inc. USA. DNA amplifying primers were purchased from Integrated DNA technologies (IDT), USA. Other chemicals and reagents were obtained from commercial suppliers, Himedia and Merck.

Escherichia coli HME5 tna<>cat (Yu et al., 2000 was generous gift from Dr. F. P. Guengerich, (Director, Center in Molecular Toxicology and Interim Chair, Department of Biochemistry, Vanderbilt University School of Medicine, TN, USA) was used to isolate chromosomal DNA for hemA gene amplification. *E. coli* K12 ER2420/pACYC184 (Chang et al., 1978) and pUC19 plasmid was obtained from New England Biolab Inc. USA.

3.4 Primer designing

The hemA gene amplification primers (kindly provided by Sujina Mali) were designed manually using different web based tools discussed below and using the nucleotide sequence deposited in the GenBank for *E. coli* K-12 MG1655 strain (NC_000913.2). Reverse primer was designed downstream of stop codon of hemA gene that would amplify the gene few bases downstream at 3'-end of the coding sequences to meet the general criteria's of

primer designing regarding length, melting temperature (Tm), GC content, secondary structure, self and hetero dimer, etc. BLAST- search (non-redundant database in NCBI) was performed for each primer to ensure that the designed primers would not prime to non-specific regions. EcoRI and XbaI were the sites inserted in FP and RP respectively. The basic information of designed primers is presented in Appendix.

During primer designing following criteria were taken into consideration

- Specificity
- Length
- Melting temperature
- Hairpin formation
- GC content
- Self dimer and hetero dimer formation
- GC clamp
- Self complementary

To check specificity of primer NCBI BLAST tool was used and primer with 100% specific only for targeted sequence was chosen. Length of primer was generally taken between 18- 25 bases for priming site but it was variable for the primer with restriction site or for mutation primers. Important consideration was primer melting temperature, the temperature at which half of the primer melts to form single strand DNA. T_m was taken from 55 to 65 degree Celsius. There was not any restriction for GC content the matter is only that it is directly linked to primer melting temperatures. Generally GC content is good for the range of 40 to 60 %. But primers set with melting temperature difference more than 3 degree Celsius was not taken. To calculate Tm,

Oligo calc: oligonucleotide properties calculator:

(http://www.basic.northwestern.edu/biotools/oligoc alc.) and

Biomath: (http://www.promega.com/techserv/tools/biomath/calc11.html) were used.

Similarly secondary structures of each primer were taken into consideration. Integrated DNA technology and mfold/DNA folding form was used as tool to study

secondary structures. Self dimer and hetero dimer was analysed using the tools available in Integrated DNA technology. Self complementary can be checked easily in oligo calc.

Bioinformatics Tools used for Primer Designing

a. NCBI for whole genome and genome map of *E. coli*
   http://www.ncbi.nlm.nih.gov/genome
   http://www.ncbi.nlm.nih.gov/guide/genomes-maps/
b. Integrated DNA technologies; Oligo analyzer 3.1 for Tm, hetero/homo dimer
   http://eu.idtdna.com/analyzer/applications/oligoanalyzer/default.aspx
c. The mfold Web server- DNA folding form for folding structure of primer DNA
   http://mfold.rna.albany.edu/?q=mfold/dna-folding-form
d. NCBI- Standard Nucleotide BLAST

   http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Nucleotides

e. New Englands Biolabs webtool: NEB cutter V2.0 for restriction map analysis of amplicon
   http://tools.neb.com/NEBcutter2/cutshow.php?name=35a4abff-

***hem*A gene amplification primers**

SSMPOA6X1F01: CTAC**GAATTC**TCCAAA **CATatg** acc ***AAA AAA*** ct c tta gca ctc gg
*Eco*RI *Nde*I

SSMPOA6X1R01: AA ACC ACT **GAATTC** CTG ATG **TCTAga** tgt act gct act cc
*Eco*RI *Xba*I

3.5 Primer Dilution

Primers were purchased from Integrated DNA technologies (IDT) in lyophilized form. The primers were provided in standard desalting purification process and used without any further purification. The freezer and working stock solutions of primers were prepared as described below.

First, the lyophilized primer containing vials were briefly centrifuged for few minutes before opening the vial for first time to avoid loss of the DNA pellet. Then 1 mM stock solution was prepared by adding a volume of TE buffer (pH 8.0) equal to the number of nano moles of DNA present in the tube. For example, 77.4 nMoles of oligo was dissolved in 77.4 µl of TE buffer (pH 8.0) to make a stock 1 mM solution. Similarly, 50 µM stock solutions were prepared by adding 5 µl from 1 mM stock solution in 95 µl of TE buffer (pH 8.0). Finally 10 µM working stock solution was prepared by adding 10 µl of 50 µM stock solution in 40 µl of MilliQ water. For long term storage, 1 mM and 50 µM stock solutions were stored at -40°C and working solution of 10 µM was stored at -20°C.

3.6 Genomic DNA Isolation

3.6.1 Preparation of cell

E. coli HME5 *tna<>cat* strain was revived and streaked on LB/chloramphenicol (10 µg/ml) plate and incubated overnight at 37°C. Then, a single isolated colony was inoculated in 2 ml of LB medium and incubated at 37°C for 18 hours at 200 rpm. From this overnight culture 1.5 ml of the culture was transferred to sterilize eppendorf tube and centrifuged at 5,000 rpm for 5 minutes at 4°C. Immediately, supernatant was discarded by aspiration and remaining overnight culture (0.5 ml) was also added to same tube containing the cell pellet and centrifuged again at 5,000 rpm for 5 minutes at 4°C. Then, supernatant was removed as much as possible without disturbing the cell pellet.

3.6.2 Cell Lysis

The cell pellet was re-suspended in 450 µl of TE1 buffer by gentle pipetting. The solution was split into two fresh sterilized eppendorf tubes by transferring 225 µl of above suspension to each tube. To each tube 180 µl of lysozyme (1 mg/ml) was added. Both tubes were incubated at 37°C for 30 minutes gently mixing the solution by inverting the tube every 5 minutes for proper cell lysis. Then 45 µl of STEP solution was added in both tubes and incubated in ice for 45 minutes or until the solution became clear due to cell lysis, with gentle inversion in between the ice incubation period. Equal volume of chilled phenol (450 µl) was added and mixed by vortexing. The mixture was centrifuged at 13,000 rpm for 10 minutes. The upper aqueous layer containing DNA was transferred to fresh sterilized eppendorf tube without carryover of lower organic phase. Again, equal volume of chilled

phenol: chloroform: isoamyl alcohol (25:24:1) was added to above aqueous solution, mixed by vortexing and centrifuged at 13,000 rpm for 10 minutes at 4°C. After collection of aqueous layer in a fresh tube, equal volume of chloroform was added and vortexed. The mixture was then centrifuged at 13,000 rpm for 2 minutes and the aqueous phase was collected in a fresh eppendorf tube.

3.6.3 Genomic DNA recovery

To the aqueous solution (450 µl) containing genomic DNA, 100 µl of 3M chilled sodium acetate (pH 5.2) and double volume of 95% ethanol (1,100 µl) was added and incubated at -20°C for 30 minutes. Then the mixture was centrifuged at 13,000 rpm for 20 minutes at 4°C. The supernatant was poured off and pellet was washed with 250 µl of 70% ethanol without disturbing the pellet. Then the solution was centrifuged at 13,000 rpm for 10 minutes at 4°C. After draining the supernatant, remaining ethanol was removed by keeping the tube open in room temperature for 5- 10 minutes. Care was taken not to over dry the DNA pellet. The genomic DNA was re-suspended in 100 µl of autoclaved MilliQ water or TE buffer (pH 8.0) and stored at -20°C until use.

3.6.4 Genomic DNA Quantification

The genomic DNA quantification and purity assessment was done by UV-visible Spectrophotometer. The absorbance of the sample was measured at 260 nm and 280 nm wavelengths. The concentration of genomic DNA was calculated using the following formula

DNA concentration (µg/ml) = A_{260} X 50 X Dilution factor

A_{260} of 1 corresponds to approximately 50 µg/ml of DNA .

DNA purity = A_{260} / A_{280}

Where, A_{260} and A_{280} are Absorbance of DNA or RNA at wavelength 260 nm and protein at wavelength 280 nm, respectively.

Pure DNA sample has A_{260} / A_{280} ratio equals to 1.8, the ratio above 1.8 indicates RNA contamination whereas the ratio below 1.8 suggests protein contamination in the DNA sample.

3.7 PCR amplification of hemA gene

The chromosomal DNA sample isolated from *E. coli* HME5 *tna<>cat* strain was used as the template DNA for PCR amplification of *hem*A gene using specifically designed forward (SSMPOA6X1F01) and reverse (SSMPOA6X1R01) primers designed by Sujina Mali. The DNA sample was quantified spectrophotometrically and diluted in MilliQ water to prepare template DNA with concentration of 10-20 ng/µl. PCR amplification was performed with Phusion High-Fidelity (HF) DNA polymerase, NEB. The general protocol provided by NEB technical reference was followed to prepare the reaction mixtures and to determine PCR conditions and are presented in Appendix.

The two step reaction conditions was used for *hem*A gene amplification. For the first 5 PCR cycles, the annealing temperature targeted to only priming region (3'-end nucleotide complimentary to template DNA) was used whereas for rest 25 PCR cycles, annealing temperature (T_m– 5°C) was determined regarding the whole primer sequence. This strategy increases overall PCR efficiency and yield of PCR product. PCR was performed using Thermal cycler: Bioer-XP technology, China. Then 5 µl of PCR product was subjected to 1% agarose gel electrophoresis along with 1 kb NEB DNA ladder in a separate well as described previously.

3.8 PCR product purification

After confirmation of correct amplicon size (1.3 kb DNA fragment) from gel electrophoresis, PCR product was purified by chilled phenol-chloroform method to remove Phusion DNA polymerase and buffers that might interfere with subsequent restriction digestion and ligation steps.

Briefly, 155 µl of MilliQ water was added to the remaining PCR product (45µl) to increase the final volume to 200 µl. Then equal volume of chilled phenol: chloroform (1:1 v/v) solution (200 µl) was added, briefly vortexed and centrifuged at 13,000 rpm for 2 minutes at 4°C. The aqueous layer was carefully collected in a fresh eppendorf tube avoiding carry off of lower organic phase. Equal volume of chloroform (200 µl) was added to this aqueous solution and centrifuged at 13,000 rpm for 2 minutes at 4°C. Again the aqueous layer containing DNA was collected in a fresh tube. To this solution, 25 µl of 3M Chilled sodium acetate (pH 5.2) ($1/10^{th}$ volume of aqueous solution) and equal volume of isopropanol (250

µl) was added and incubated at -20°C for 30 minutes. Then the DNA was recovered by centrifuging the mixture at 13,000 rpm for 15 minutes at 4°C and the pellet was washed with 1 ml of 70% ethanol without disturbing the pellet. The mixture was then centrifuged at 13,000 rpm for 2 minutes at 4^0C and supernatant was removed and allowed to air dry for 5-10 minutes until alcohol was evaporated. Then the precipitated DNA was re-suspended in 20 µl of MilliQ water and used for restriction digestion.

3.9 Isolation of pUC19 plasmid

Plasmid pUC19 was recovered from E. col*i* K12 (Provided by New England Biolab) strain by Alkaline lysis method and purified using HiPurA Silica kit (HiMedia) following manufacturer's protocol.

3.9.1 Bacterial culture and cell harvest

Briefly, *E. coli* K12 strain was revived by streaking on Luria Bertani (LB)/Ampicilline (5 µg/ml) agar plate and incubated overnight at 37°C. A single isolated colony was inoculated in 3 ml of LB/Ampicillin (10 µg/ml) medium and incubated at 37°C for 18 hours at 200 rpm. Then, 1.5 ml of the overnight culture was transferred to sterilized eppendorf tube and centrifuged at 5,000 rpm for 5 minutes at 4°C. The supernatant was poured off and remaining culture (1.5 ml) was added to same tube with cell pellet and again centrifuged at 5,000 rpm for 5 minutes at 4°C. Then supernatant was removed as much as possible without disturbing the cell pellet and the cells were re-suspended in 500 µl of ice cold STE solution and again centrifuged at 5,000 rpm for 5 minutes at 4°C.

3.9.2 Cell lysis: Alkaline lysis method

The cell pellet after washing with STE solution was re-suspended in 200 µl of alkaline lysis solution I and mixed well by pipetting. To this solution, 200 µl of freshly prepared alkaline lysis solution II was added and mixed by gentle inversion of tube for three times. The bacterial solution becomes clear as the cell lysis occurs. To this mixture 200 µl of ice cold alkaline lysis solution III was added and mixed by inverting the tube gently for several times. During this step, a cloudy precipitate becomes visible due to precipitation of proteins. The mixture was then centrifuged at 13,000 rpm for 10 minutes at 4°C. Then supernatant solution consisting plasmid DNA was transferred to a fresh eppendorf tube.

3.9.3 Plasmid DNA purification by HiPurA Silica kit

To the supernatant containing plasmid DNA, 100 µl of chaotropic salt solution (provided by manufacturer) and 5 µl of glass powder suspension (GPS) was added and left at room temperature for 10 minutes with intermittent mixing. The solution was centrifuged at 10,000 rpm for 1 minute at 4°C. Then supernatant was removed without disturbing the pellet. The pellet was re-suspended by pipetting after addition of 500 µl of diluted wash solution (provided by manufacturer). Then it was centrifuged at 10,000 rpm for 1 minute at 4°C. Then the supernatant was again discarded and the pellet was washed two more times with 100 µl of diluted wash solution and centrifuged as above. Then supernatant was removed and the pellet was allowed to air dry for 5- 10 minutes at room temperature. The pellet was dissolved in 20 µl of MilliQ water by pipetting and incubated at 55^0C in a water bath for 10 minutes. The solution was mixed in every 2-3 minutes to properly extract DNA from silica gel. Then it was centrifuged at 10,000 rpm for 1 minute at room temperature and the supernatant solution containing plasmid DNA was transferred to a fresh eppendorf tube without touching the pellet and stored at -20°C until use.

3.10.1 Restriction digestion of hemA gene

*hem*A gene was digested with *EcoRI* and *Xba*I to clone into pUC19 vector for further analysis which has been digested and further EcoNI digestion. When *hem*A gene was analysed by NEB Cutter tool (Appendix 3), *Eco*NI site was found to be present in internal coding sequence at 3 places(196, 355 and 1102(CCTNN↓NNNAGG) creating 5' overhang from *hem*A start codon). Thus, *hem*A gene was completely digested with *Eco*RI and XbaI.

Complete digestion with *Eco*RI and XbaI

DNA solution	20 µl
*Eco*RI-HF™ (5 U/ul)	2 µl (4U)
XbaI	2 µl(4U)
NEB buffer 4 (10X)	3 µl
BSA(10X)	3 µl
Total	30 µl

The reaction mixture was incubated at 37°C for 4 hours and proceeded for gel run for DNA extraction.

3.10.2 1% Low melting Agarose Gel electrophoresis

For gel electrophoresis and to recover proper size digested DNA containing *hem*A gene, 1% low melting agarose gel was prepared by boiling 0.2 gm of low melting agarose (HiMedia) dissolved in 20 ml 0.5X TAE buffer. Once the gel was cooled to about 50^0C- 60^0C, 1 µl of EtBr solution (5 mg/ml) was added and gel was casted. Then it was placed in gel tank containing 0.5X TAE buffer. The restriction digested DNA solution pooled to 15 µl that consisted insert DNA sample was mixed with 10X gel loading buffer and was loaded in a well. In another well 6 µl of 1X NEB 1kb DNA ladder was loaded and run for nearly 2 hours at 50V. The gel was visualized under Wise DOC and correct size insert (1.3 kb) DNA band was cut with clean knife leaving other fragments in the gel for gel extraction.

3.10.3 DNA extraction from low-melting gel

DNA recovery efficiency from low melting gel is greater than from high melting gel. Thus, multiple DNA bands of partially digested insert DNA sample was separated using low melting agarose gel as above and the DNA extraction protocols employed were as described below.

Briefly, the weight of gel was measured and approximately 1.5 volume of TE buffer (pH 8.0) was added (For example, for 100 mg of agarose gel, 150 µl of TE buffer, pH 8.0 was added). The gel was melted by boiling the gel at 65^0C for 5-10 minutes. Then, equal volume of phenol (250 µl) was added to the solution, briefly vortexed and centrifuged at 13,000 rpm for 10 minutes at 4°C. The aqueous layer was carefully collected in fresh eppendorf tube avoiding carry off of lower organic phase and DNA was recovered by alcohol precipitation. The solution was mixed with 25 µl of 3M chilled sodium acetate (pH 5.2), 1 µl of glycogen (2 mg/ml) and 280 µl isopropanol solution. The mixture was incubated at room temperature for few minutes and centrifuged at 13,000 rpm for 15 minutes at 4°C. The pellet was washed with 1 ml of chilled 70% ethanol and again centrifuged at 13,000 rpm for 2 minutes at 4^0C. Ethanol was evaporated by allowing the tube to air dry for 5-10 minutes. Then recovered DNA was re-suspended in 10 µl of MilliQ water and directly used for ligation reaction.

3.11 Preparation of *E. coli* DH5α competent cells

E. coli DH5α competent cells were prepared by Rubidium chloride (RuCl) method. First, the bacterial strain was revived by streaking on LB agar plate and incubated overnight at 37°C. After incubation, a single isolated colony was inoculated in 2 ml of LB medium and incubated at 37^0C for 18 hours at 200 rpm. Then 20 µl of overnight culture was 100 times diluted in 2 ml of SOC medium and incubated at 37°C with shaking at 200 rpm for an hour or until OD_{600} reaches 0.4-0.6. The bacterial cells were pelleted by centrifuging at 10,000 rpm for 5 minutes at 4°C. Then the cells were re-suspended into 800 µl of ice cold RF1 solution (Note: add 0.4 volume of RF1 to original volume of culture) and incubated in ice bath for 15 minutes. Again the cells were collected by centrifuging at 10,000 rpm for 5 minutes at 4°C. The supernatant was discarded and cell pellet was re-suspended in 80 µl of ice cold RF2 solution (Note: add 0.04 volume of RF2 to original volume of culture) and incubated in ice bath for 15 minutes. Thus prepared *E. coli* DH5α competent cells were ready for transformation or stored at -40°C until use.

3.12 Cloning of *hem*A in plasmid pUC19

The *hemA* gene recovered after complete digestion with *Eco*RI and partial digestion with *Nde*I was sub-cloned in plasmid pUC19, previously digested with same restriction enzymes, employing T4 DNA ligase (NEB). For the ligation reaction, vector: insert molar ratio was kept within range of 1:3 to 1:10 according to known quantity of vector and insert DNA. The concentration of vector and insert DNA was determined by gel quantification against the known amount of standard DNA ladder. The amount of 4 kb vector DNA required to be added to 25 ng of 1.3 kb insert (*hemA*) DNA with molar ratio of 1:5 was calculated using following formula.

$$\text{ng of vector to be added} = \frac{\text{ng of insert X kb size of vector X Molar ratio of vector: insert}}{\text{kb size of insert}}$$

Ligation reaction mixture

*hem*A (insert)	8 µl	(25 ng)
pUC19 (vector)	6 µl	(15 ng)
T4 DNA ligase (1U/µl)	4 µl	
10X T4 DNA ligase buffer	2 µl	

Total	20 µl

The ligation mixture was incubated at 16°C for overnight and the whole ligation mixture was used for transformation of *E. coli* DH5α competent cells by heat shock method as described.

3.13 Transformation

Purified *Dpn*I digested and ethanol precipitated PCR amplified plasmid DNA sample was used for transformation of *E. coli* DH5α competent cells by heat shock method. Briefly, 10 µl of plasmid DNA was added to 80 µl of *E. coli* DH5α competent cells, mixed well and placed in ice bath for 30 minutes. Then it was subjected to heat shock by keeping in water bath at 42°C for 2 minutes and immediately transferred to ice bath and kept for 10 minutes. Then 1 ml of SOC medium was added to the transformed competent cells and incubated at 37°C for 90 minutes keeping the tube perpendicular and slightly shaking (100 rpm) to stabilize the plasmids in transformed cells. After incubation, 100 µl of transformed cells were directly plated into LB/ampicilline (5 µg/ml) agar plates whereas remaining cells were centrifuged at 10,000 rpm for 10 minutes at 4°C and cells were re-suspended in 100 µl of sterilized ice cold phosphate buffer (pH 7.0) and spread in LB/ ampicilline (5 µg/ml) agar plates.

Both plates were incubated at 37^0C for overnight. Then few of the isolated tetracycline resistant transformed colonies were selected to inoculate 3 ml of LB/ ampicilline (10 µg/ml) medium and incubated as described previously

3.14 Confirmation of Transformants

3.14.1 Plasmid Extraction

The newly transformed *E. coli* DH5α cells with ligation mixture was plated in LB/Ampicilline (5 µg/ml) agar plate. Then, the Ampicilline resistant colonies were sub-cultured in 2 ml of LB/Ap (5 µg/ml) medium and incubated at 37°C with shaking at 200 rpm. The plasmid was isolated as described previously by Alkaline lysis method, purified by chilled phenol-chloroform mixture and recovered after ethanol precipitation. Isolated Plasmid was re confirmed with EcoRI digestion.

3.15 Restriction digestion of pUC –hemA with EcoNI

Puc19 vector cloned with hemA were digested with EcoNI . hemA gene contains 3 EcoNI digestion sites. After digestion vector was followed by ligation generating hemA knockout pUC19.

Complete digestion with EcoNI

DNA solution	1.5 µl(150 ng)
*Eco*NI (5 U/ul)	1.5 µl(stock)
NEB buffer 4 (10X)	1.5 µl
MilliQ water	10.5 µl
Total	15 µl

The restriction digestion mixtures were incubated at 37°C for 2 hours and electrophoresed in 1% agarose gel as described previously

3.16 Cloning of HemA into pET15B vector

With same procedure hemA was cloned in expression vector pET15b with restriction digestion NdeI and EcoRI followed by ligation. Reaction mixture for restriction digestion and ligation is as follows. For digestion of Vector

Template	20 µl
Buffer	4 µl
EcoRI	5 µl
NdeI	5 µl
MilliQ waterH2O	6 µl
Total	40 µl

HemA digestion with EcoRI and NdeI for cloning into pET15b vector

DNA	18 µl (250 ng)
*Nde*I (1U/µl)	4.5 µl
*Eco*RI-HF^{TM} (1U/µl)	4.5 µl
NEB buffer 4 (10X)	4 µl
Total	40 µl

Both bands were electrophorised and subjected into ligation after gel purification as same method described above (pUC19).

Ligation mixture was as follows

Insert:	(20 ng whole)
Vector:	2ul 9ng/ µl)
Ligase buffer:	1.5 µl
Enzyme:	3 µl
DD water:	8.5 µl
Total:	15 µl

3.17.1 Gene identification for heme biosynthesis

Gene of heme biosynthesis are analyzed by grouping in two groups. Six genes responsible for uroporphyrinogen III synthesis from L-glutamate are grouped in one group and protoheme biosynthesis from uroprophyrinigen III are grouped in another group. Most of gene information taken using reference source encyclopedia of Escherichia coli (http://ecocyc.org/ECOLI/).

3.17.2 Upstream analysis of all genes of heme biosynthesis

Transcription initiation sites of five genes are mentioned clearly in (http://ecocyc.org/ECOLI/) and transcription initiation sites of remaining genes are not available till this date. In same database translation initiation is available and we choose 100 upstream nucleotides from translation initiation site. Thus structures are analyzed from 100 nucleotides upstream from first gene. For gene which are in operon with other gene and are not first gene, upstream region from first amino acid are considered.

3.17.3 Blast search for nucleotide similarity

NCBI blast (special BLAST) align sequence nucleotide BLAST(http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&SHOW_DEFAULTS=on&BLAST_SPEC=blast2seq&LINK_LOC=align2seq) was done for upstream 80 nucleotide from first gene of all operons of heme

biosynthesis where query sequence was taken for upstream 80 nt of hemA and subject sequence were taken from all other remaining gene with same length and same position as of hemA. This was done for searching if similarity is available for the 5' UTR of heme biosynthetic genes or not.

3.17.4 Structure identification

RNA structure of upstream UTR of all transcripts of all genes were identified using mFOLD RNA biology(http://mfold.rna.albany.edu/?q=mfold). Pdf version of structures were extracted and stability were analysed by comparing gibbs free energy value. More negative value of Gibbs free energy(-ΔG), more is stability of the structure. Stability of transcripts of hemA gene will be then compared with other structure which are involved in heme biosynthesis.

3.17.4 Stability analysis

For stability we mostly considered negative gibbs free energy change (-dG) value of the structure as obtained from the mfold web server(http://mfold.rna.albany.edu/?q=mfold). For gene whose detail of transcript is not available on ecocys(http://ecocyc.org/) upstream 100 nucleotide from first gene (-100) of the operon has been selected for analysis. For the transcripts whose detail information like transcription initiation site, TATA, GC box, number of transcript is available all upstream sequence from first gene along with upstream 100 nucleotides from first gene were analyzed.

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Primer designing for *hem*A gene

Using different free access web based tool *hemA* gene amplification primers were designed to amplify genomic *hemA* gene. *E. coli* HME5 tnaA<>cat strain is more efficient than other *E. coli* strains in expression of cytochrome P450 (Wu *et al*., 2005), and is also efficient in homologous recombination and is *E. coli*K12 derivative (Yu *et al*., 2000) thus the primer was designed to amplify hemA gene for cloning in pUC19 vector.

Using Oligo analyzer 3.1, the self complimentary and other parameters were determined. Then the DNA folding structure was checked using mfold Web server- DNA folding form (figure 4.1) with Tm of the secondary structures (30^0C for forward, and 39°C for reverse) which is lower than Tm (63.6°C for forward and 63.2°C for reverse primer) of the whole primer used during PCR. 3' region of primer did not show any secondary structure formation at default condition of mfold tool. Thus 3'-end was completely free of folding structure, which was assumed to facilitate proper annealing and will permit access to polymerase for polymerization. Similarly, heterodimer structure was also analyzed and found to have structure having ΔG = -8.51 kcal/mole which is very low energy to make the primer primer homodimer and heterodimer formation. Further, primer was analyzed by using BLAST to determine any non specific binding, however no such non specific binding was found. They were given name as SSMPOA6X1F01 for forward primer in which at the 5'-end had restriction sites created for recognition by restriction enzymes *EcoRI* followed by *NdeI*. Last three base pair of NdeI which is ATG was overlapped with the ATG of gene itself while designing the primer. Similarly, name of reverse primer was given as SSMPOA6X1R01 in which *EcoRI* and *XbaI* restriction sites were inserted at 5' end of the primer.

4.2 PCR amplification of hemA gene of E. Coli

Genomic DNA was extracted from *E. coli* HME5 *tna<>cat* strain and its concentration and purity were determined spectrophotometrically. The concentration of genomic DNA was

calculated to be 720µg/ml with purity of 1.67(A_{260}/A_{280}). Strain containing mutant *hem*A gene, insensitive to heme mediated regulation, developed by Jones and Elliott (2010) was provided to us and hemA gene was amplified by using primers (SSMPOA6X1F01 and SSMPOA6X1R01) form the genomic DNA of the same mutant strain provided. High Fidelity Phusion DNA polymerase was used for higher accuracy. The PCR product gave a clear band corresponding to around 1.3 kb DNA fragment on comparing with 1 kb NEB DNA ladder (Figure 4.1). This DNA band was similar to nucleotide size of *hem*A gene as calculated from genome sequence of *E.* coli MG1655 strain (Appendix). Taking the size of the amplified DNA and the specificity of both primers as confirmed from NCBI BLAST results, it was assumed that *hem*A gene has been successfully amplified.

4.3 Restriction digestion of *hem*A gene

The PCR product that gave 1.3 kb DNA fragment size and presumed to be *hem*A gene was purified by phenol:chloroform method. After purification, the amplified DNA was completely digested with *EcoRI* and *XbaI* subsequent cloning in pUC19 digested with *EcoRI* and *XbaI*. In order to purify DNA digested with restriction enzymes, the restriction digested PCR sample was run in 1% low melting agarose gel (Figure 4.3 lane 2) to separate DNA. After gel run corresponding DNA band (~1.3 Kb) was cut with clean knife and preceded for gel purification.

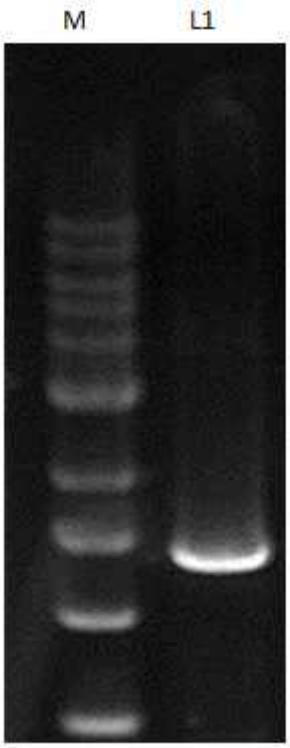


Figure 4.3: Restriction digested *hem*A gene in 1% Low melting agarose gel, Lane M: 1 KB DNA ladder and lane L1: *hemA* band after restriction digestion with *EcoRI* and *XbaI*

4.4 Restriction digestion of pUC19 with *EcoRI* and *XbaI*

The cloning vector pUC19 was prepared for sub-cloning of *hemA* gene by double digestion with *EcoRI* and *XbaI* and whole reaction mixture was subjected to gel electrophoresis in 1% agarose gel. The linearized vector corresponding to 2.6 kb DNA fragment (Figure 4.4) was subjected to gel purification.

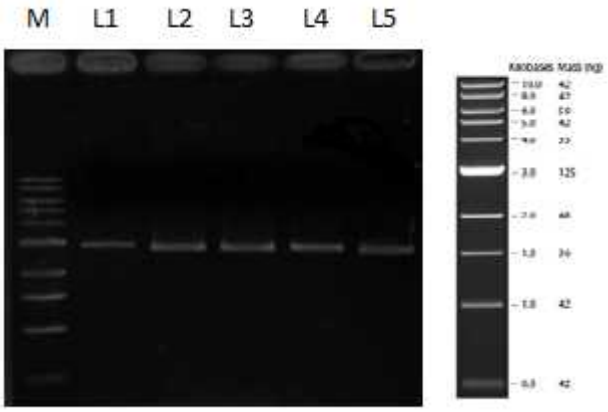


Figure 4.4: Restriction digested vector of pUC19 with *EcoRI* and *XbaI* (Lane L1-L5), 1 KB DNA ladder (right)

4.5 Cloning of hemA into pUC19 vector

Both PCR amplified *hem*A gene fragment and pUC19 digested with *EcoR*I and *XbaI*, were electrophoresed and their respective quantities were empirically calculated by visualization in 1% agarose gel comparing with known quantity of 1 kb NEB DNA marker. With the formula for subcloning as described in materials and methods, respective DNA was subjected to ligation. After ligation, the reaction mixture was used to transform *E. coli* DH5α competent cells by heat shock method as described. Sub-cloning of *hemA* gene (1.3 kb) in pUC19 vector (2.6 kb) was confirmed by restriction map analysis of isolated plasmids from transformants.

Plasmid DNA was isolated from positive clones surviving in ampicilline plate and was confirmed by restriction digestion for insertion of *hemA* gene in pUC19. The linearized plasmid showed 3.9 kb DNA band with *EcoRI* digestion corresponding to cumulative base pairs of pUC19 (2.6 kb) and PCR amplified *hem*A gene (1.3 kb). Furthermore when cloned vector is allowed to *EcoRI* and *XbaI* double digestion band corresponding to pUC19 (2.6KB) and corresponding to *hemA* (1.3 KB) (Figure 4.5) was obtained. This indicated that *hemA* gene was successfully cloned in plasmid. The newly developed plasmid was given name as pPAhemA.

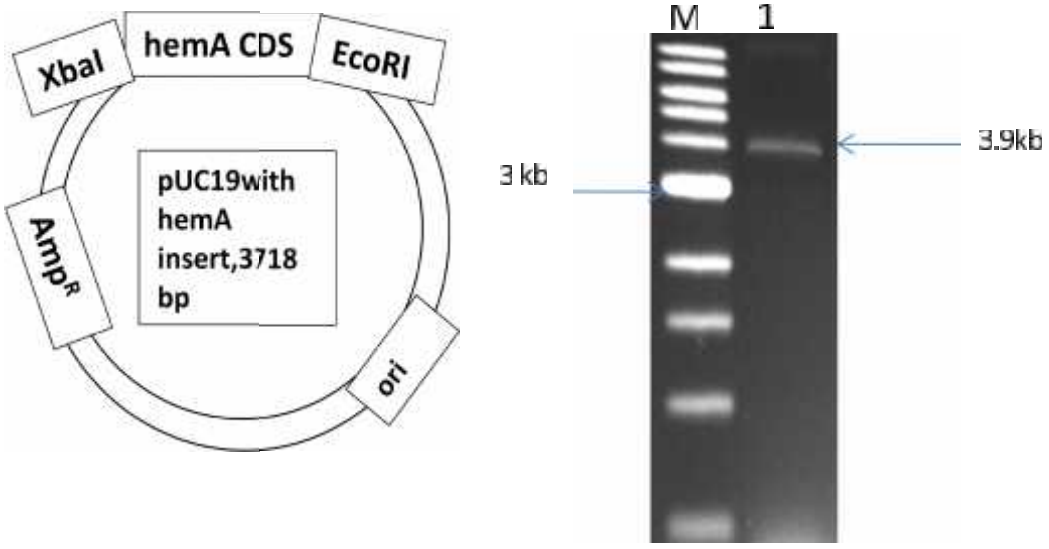


Fig 4.5: pUC19 cloned with *hemA* lane1 (Above) and Map of pUC19 after cloning of hemA (below)

4.6.1 Restriction digestion of *hemA* and pET15b by *NdeI* and *EcoRI*

The PCR product that gave 1.3 kb DNA fragment size and presumed to be *hemA* gene was purified by phenol:chloroform method. After purification, the amplified DNA was completely digested with *EcoRI* and *NdeI* subsequent cloning in pET15b digested with *NdeI* and *EcoRI*. In order to purify DNA digested with restriction enzymes, the restriction digested PCR sample was run in 1% low melting agarose gel (Figure 4.3 lane 2) to separate DNA. After gel run corresponding DNA band of hemA (~1.3 Kb) anf pET15b was cut with clean knife and proceeded for gel extraction.

4.6.2 Cloning of *hemA* into pET15b Vector

hemA gene has also been cloned into pET15b vector vector kindly provided by Dr L Sreerama under the T7 promoter. This cloning was performed with digestion of NdeI and EcoRI followed by ligation. The size of the pET15b vector is 5.7 KB and after cloning size about 7KB was obtained as shown in figure.

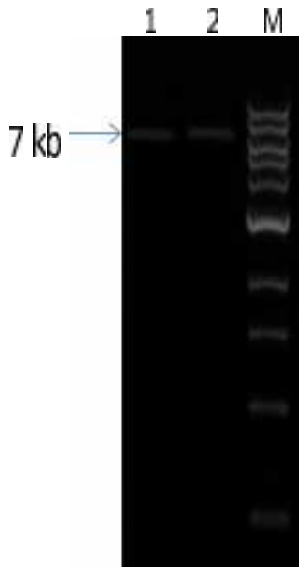


Fig 4.5: Band of pET15b cloned with hemA with NdeI and EcoRI. Lane 1 XbaI digested and 2 NdeI digested

4.7 Digestion of pUC19hemA with EcoNI

hemA gene has 3 restriction sites for *EcoNI* within its gene. We have carried out experiment on digestion of EcoNI with pUC19 vector cloned with hemA. As pUC19 vector does not contain EcoNI restriction site and only it is with hemA gene , it would be possible to make hemA knockout vector present with initial region and terminal region of hemA gene. After digestion of hemA cloned vector and followed by ligation of vector it would be possible to make hemA knockout vector. But result was unexpected although it was performed 3 times.

After EcoNI digestion pUC19hemA vector must lose 906 nucleotides as shown above. When we obtain gel electrophoresis after EcoNI digestion only band size equivalent to 750 BP was obtained.

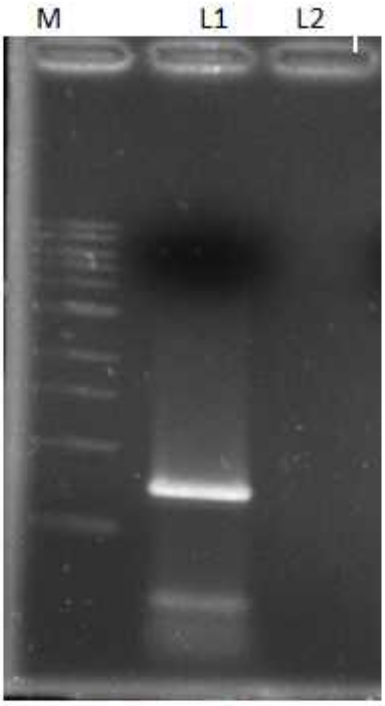


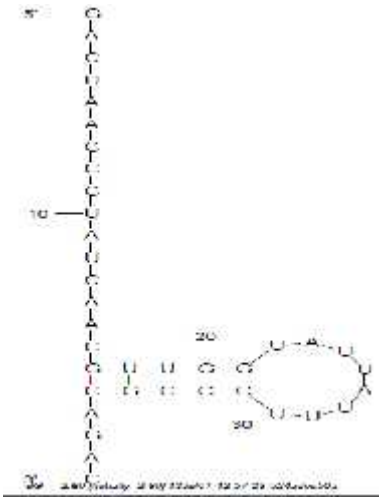
Fig. 4.6 Band of HemA after digestion with EcoNI (Lane L1) and 1KB DNA ladder (M).

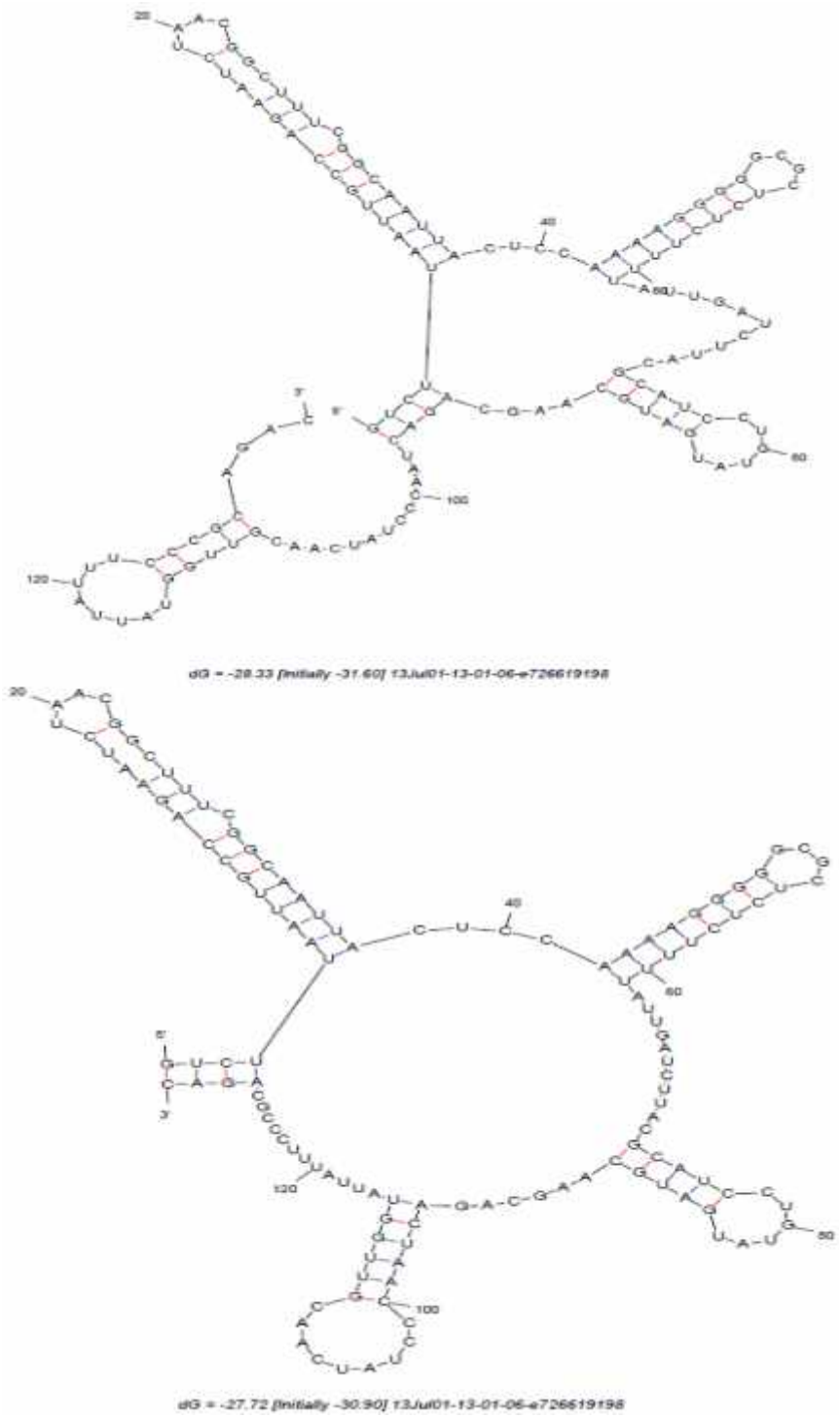
4.8 Structure identification and Stability analysis

4.8.1 hemA gene

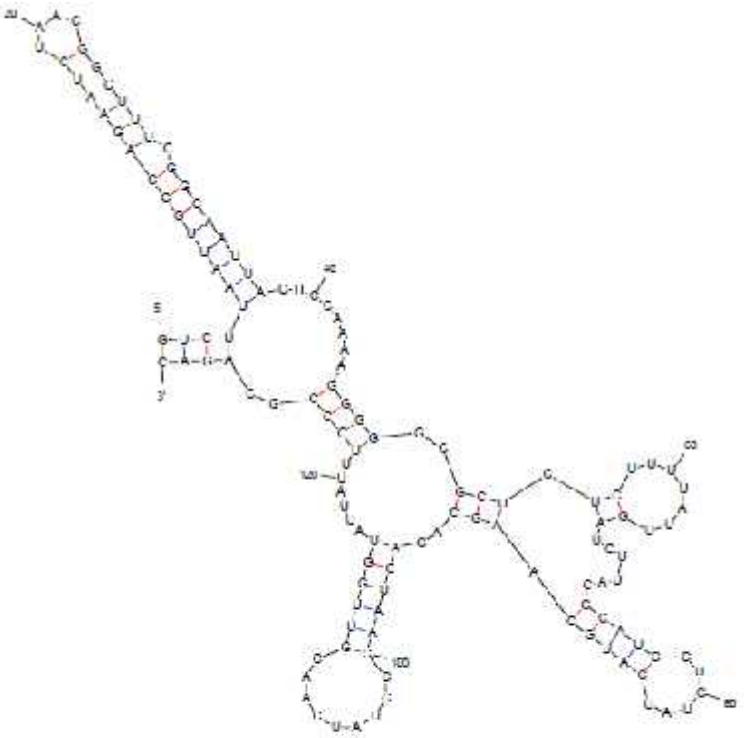
Two transcripts with promoter hemAp1 and hemAp2 has been reported in ecocys first with upstream UTR length of 38, initiating from map position 1262899 to 1262936 and next with upstream UTR length 131bp initiating from map position 1262805 in Escherichia coli genome. Mfold RNA showed two structures with first transcript and four structures with second transcripts and it was four for upstream 100 from first gene hemA.

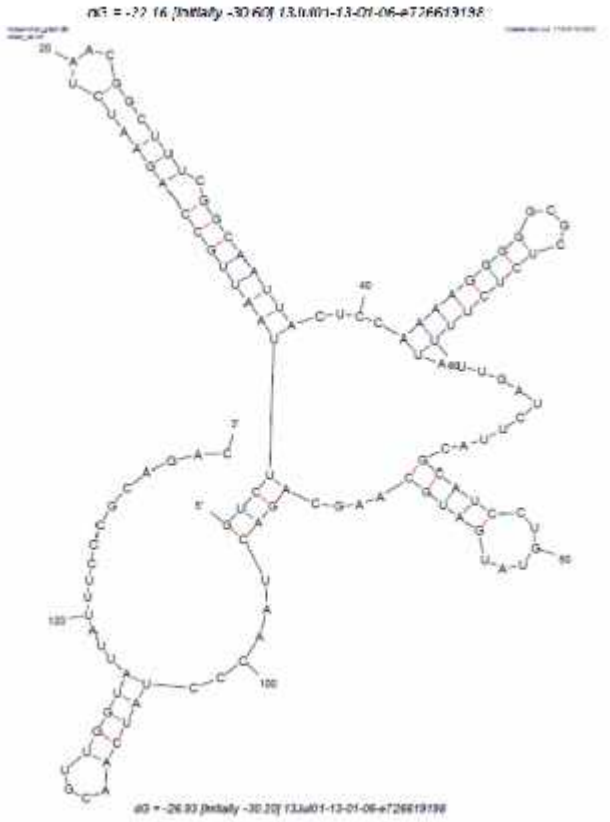
HemA					
SN	Promoter	Structure	dG(-ve)	Free nucleotide at 3'end	GC content(%)
1	hemAP1(-38)	hemA38.17	3.5	16	45
2		hemA38.18	2.6	4	
3	hemAP2(-131))	hemA131.19	28.33	4	44
4		hemA131.20	27.72	0	
5		hemA131.21	22.16	0	
6		hemA131.22	26.93	15	
7	upstream 100	hemA100.23	14.32	1	44
8		hemA100.24	13.98	2	
9		hemA100.25	17.3	16	
10		hemA100.26	17.1	4	

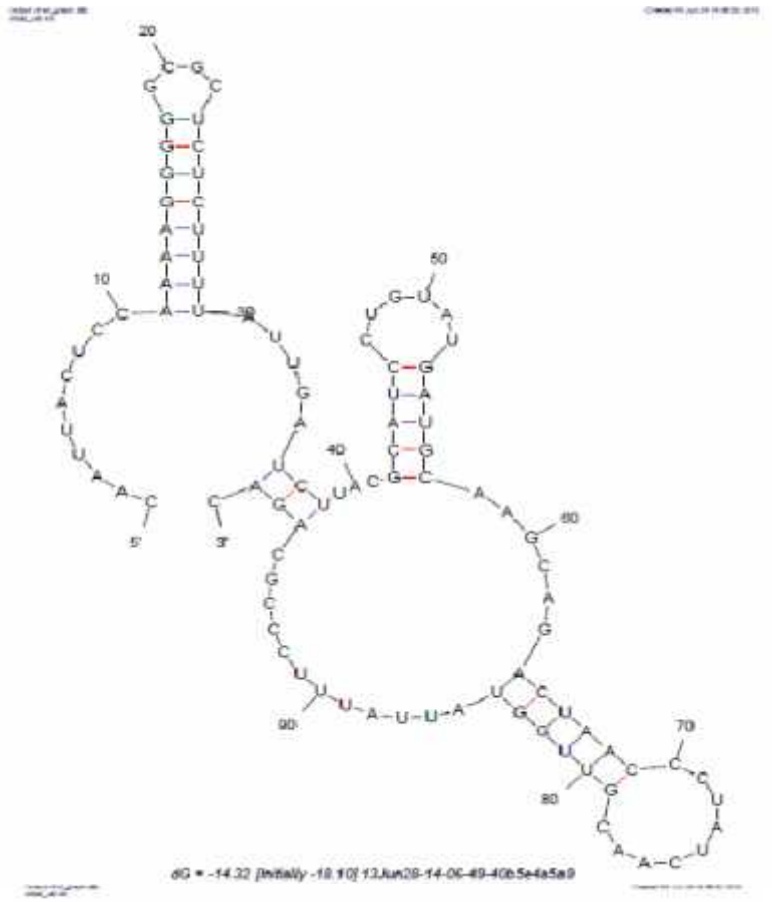


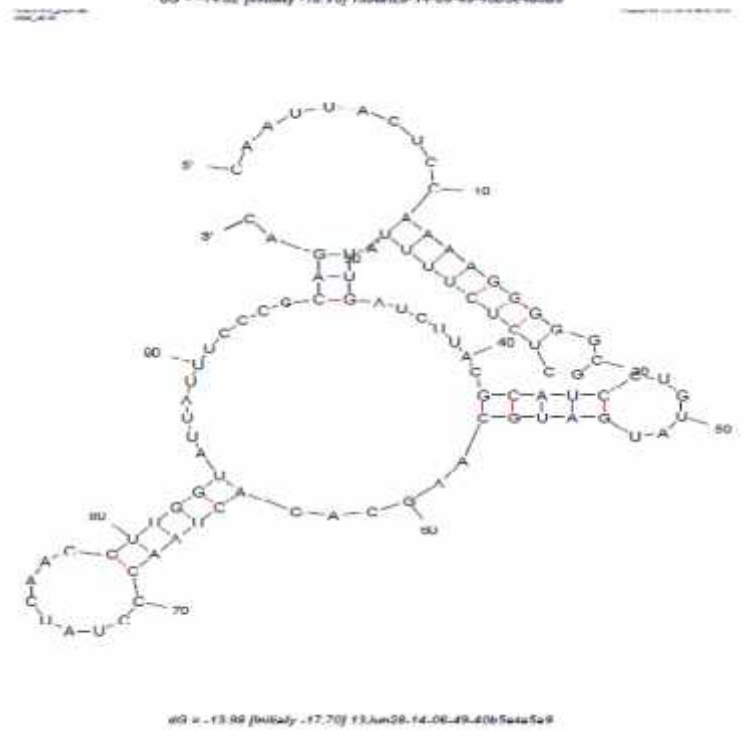


Structure of upstream UTR hemA 38.17, hemA38.18, hemA131.19 hemA131.20

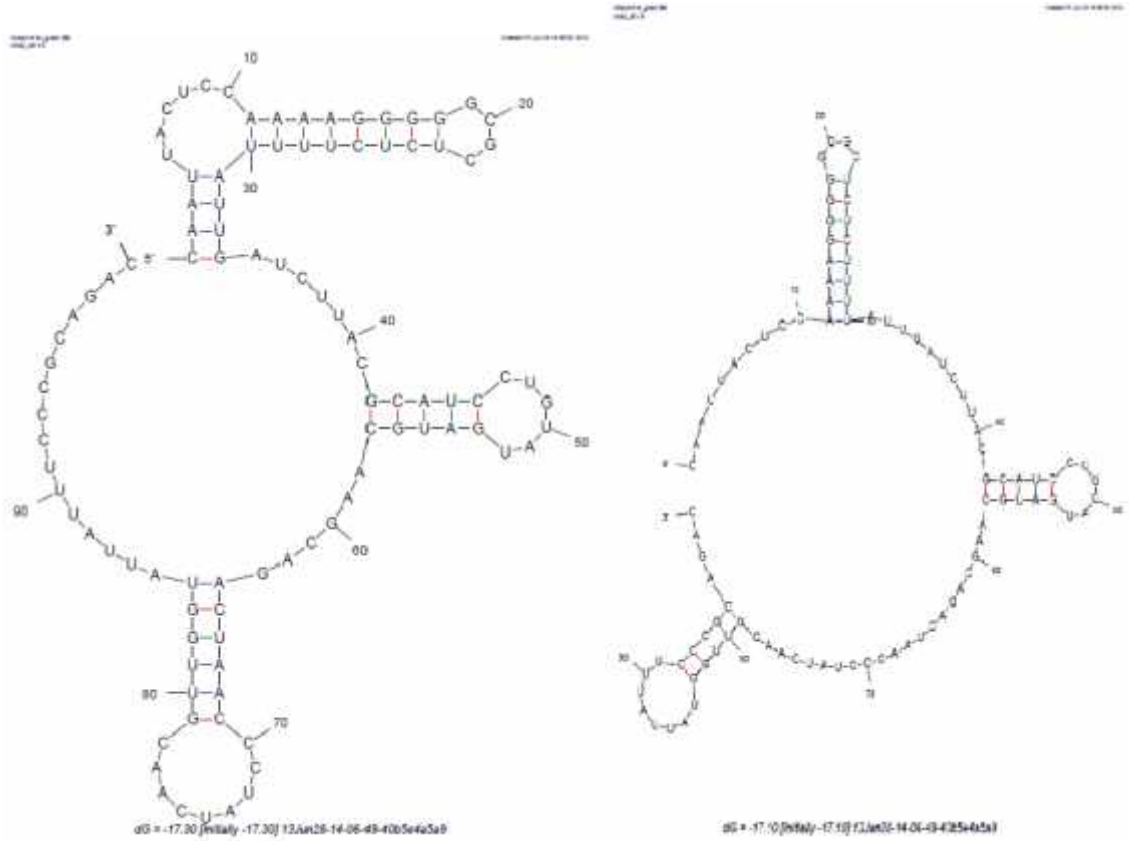








Structure of upstream UTR hemA131.21 hemA131.22, hemA100.23, hemA100.24

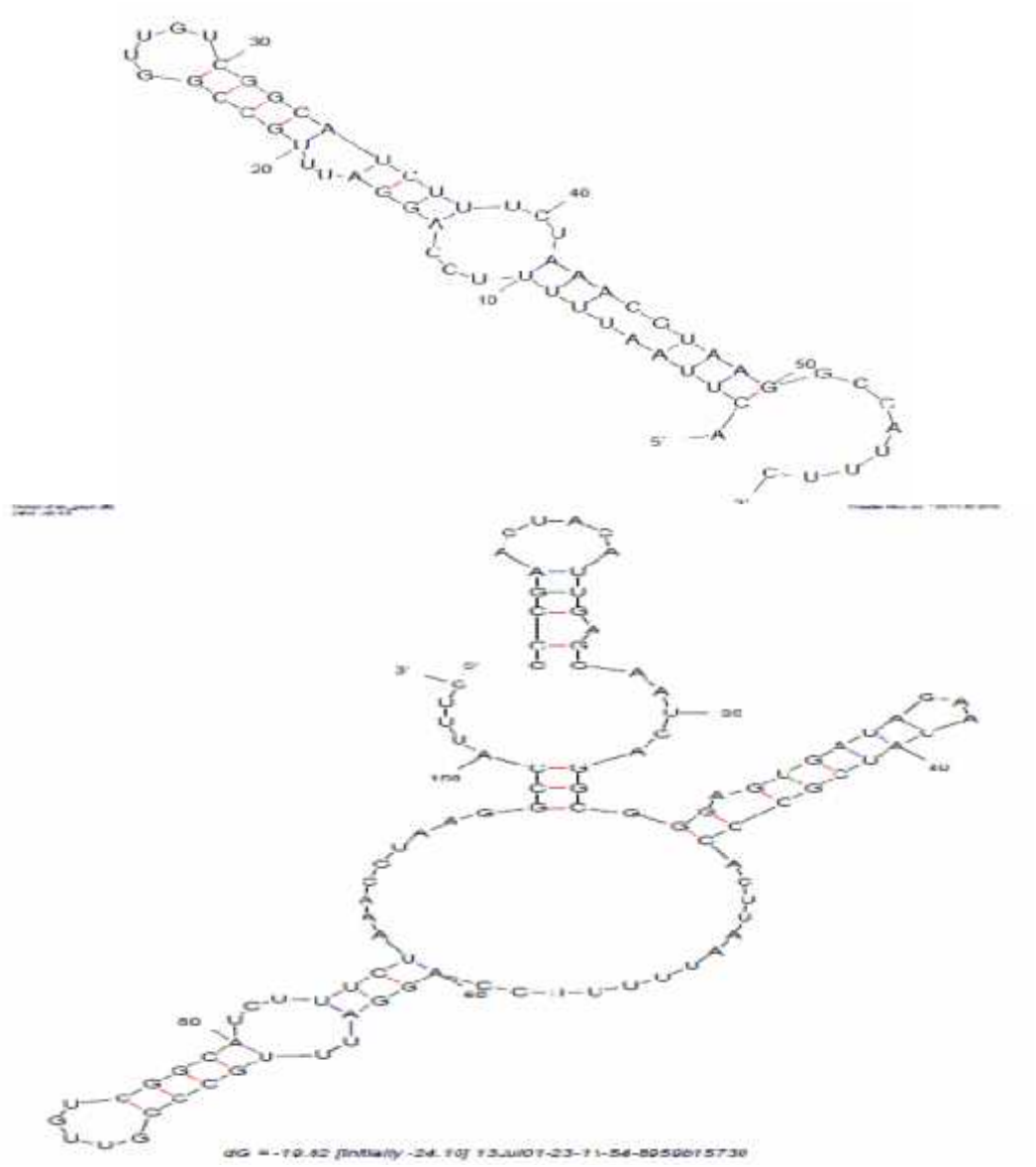


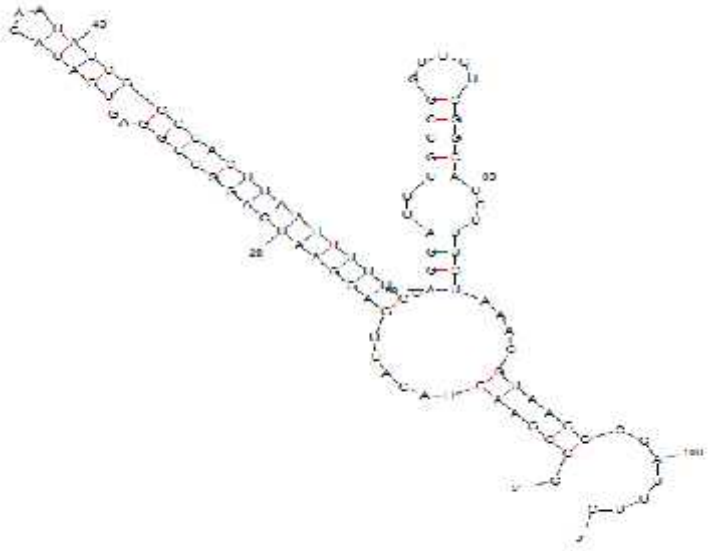
Structure of upstream UTR hemA 100.25, hemA100.26

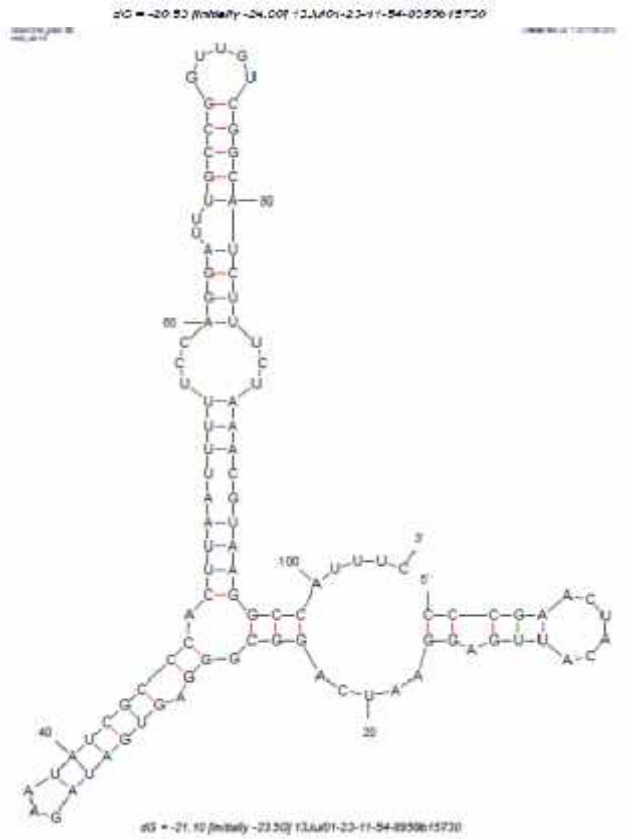
4.8.2 gltX gene

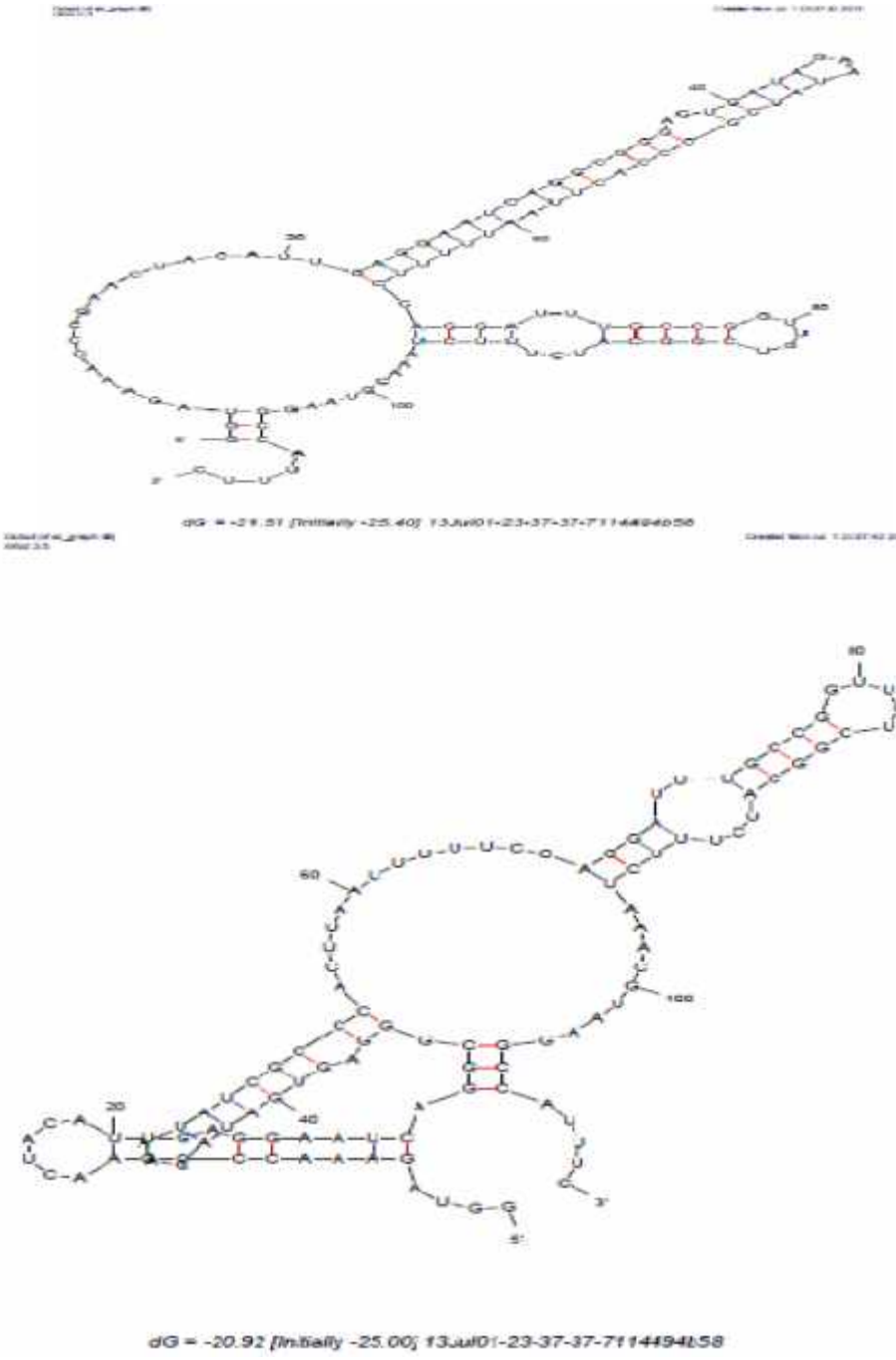
For this gene 3 initiation sites are reported initiating from map position 2518806(reverse complement) for gltXp1, 2518798(reverse complement) for gltXp2 and 2518752(reverse complement) for gltXp3 to 2518695. They showed 8, 3, and 1 structures with mfold and upstream 100 has 4 structures which was as compared to gltXp1 on the basis of GC content and free 3' nucleotides.

gltX					
SN	Promoter	Structure	dG value(-ve)	Free nucleotide at 3'end	GC content (%)
1	gltXP1	gltX112.4	21.51	5	46
2		gltX112.5	20.92	5	
3		gltX112.6	20.32	5	
4		gltX112.7	21.08	10	
5		gltX112.8	22.2	5	
6		gltX112.9	20.12	0	
7		gltX112.10	20.83	0	
8		gltX112.11	19.22	5	
9	gltXP2	gltX104.1	19.82	5	46
10		gltX104.2	20.53	7	
11		gltx104.3	21.1	5	
12	gltXP3	gltX58.16	11.1	8	41
13	upstream 100	gltX100.12	19.88	10	44
14		gltX100.13	18.32	5	
15		gltX100.14	18.12	0	
16		gltX100.15	18.02	5	

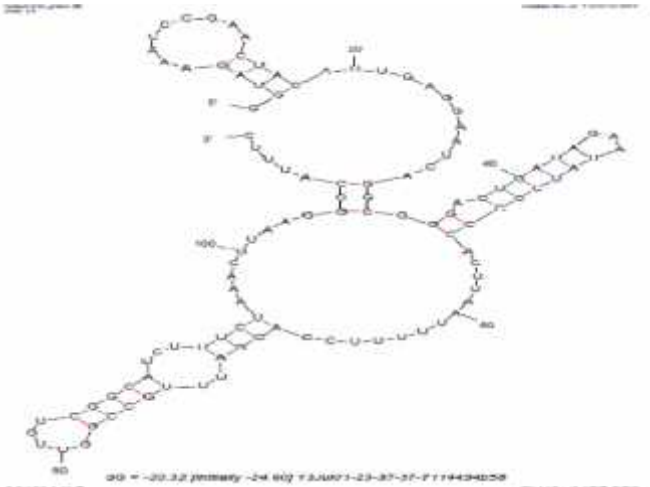

dG = -19.82 [initially -24.10] 13Jul01-23-11-54-895901573

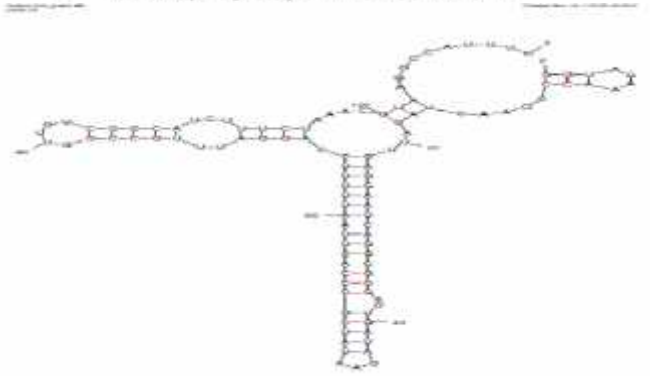


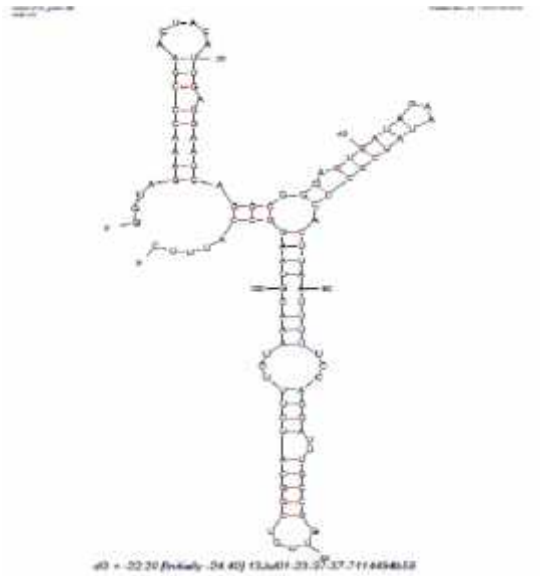


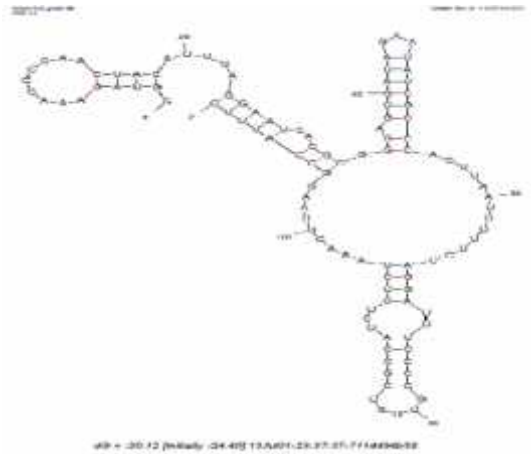


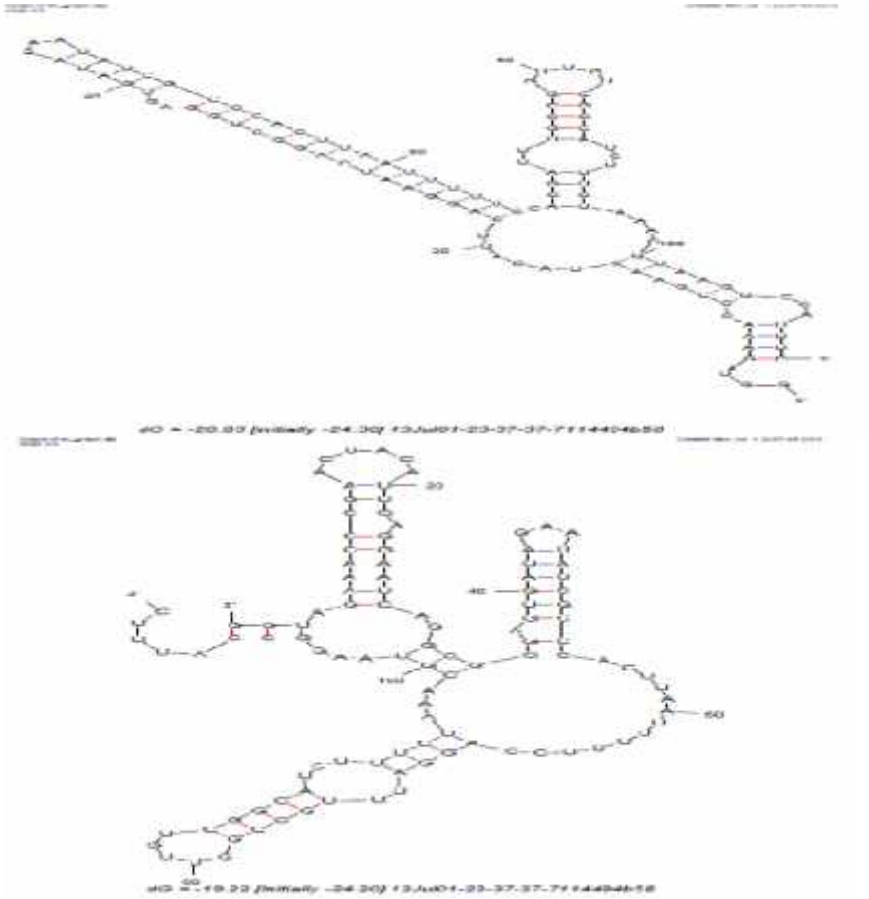
Mfold structure of 5’UTR of gltXp3(gltX58.16 first), gltXp2 (gltX104.1 second,104.2 third, and 104.3, fourth), gltXp1 (gltX112.4-fifth and gltX112.5 sixth.



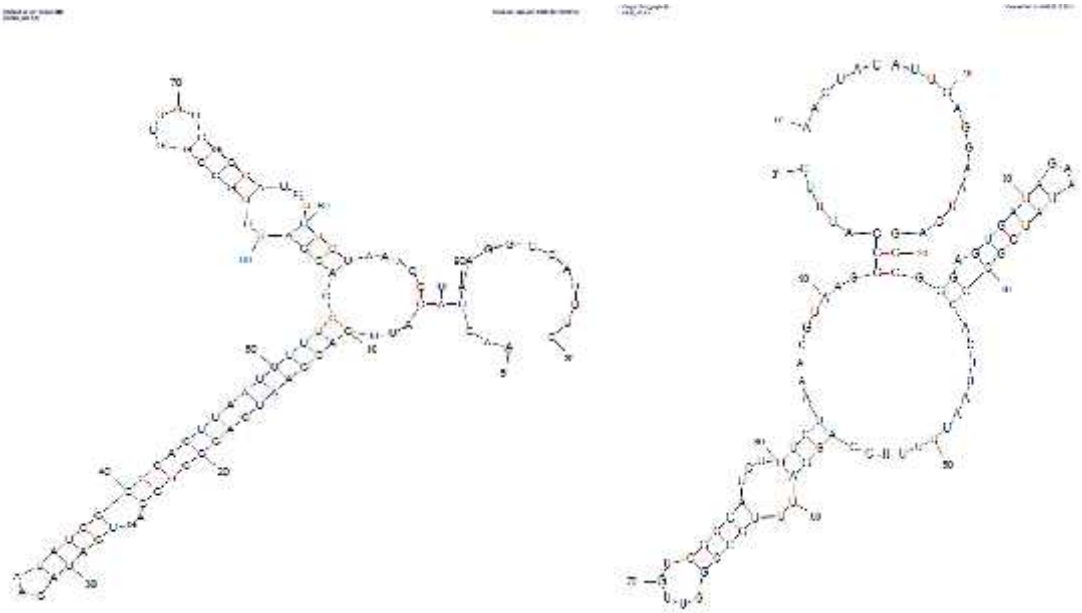


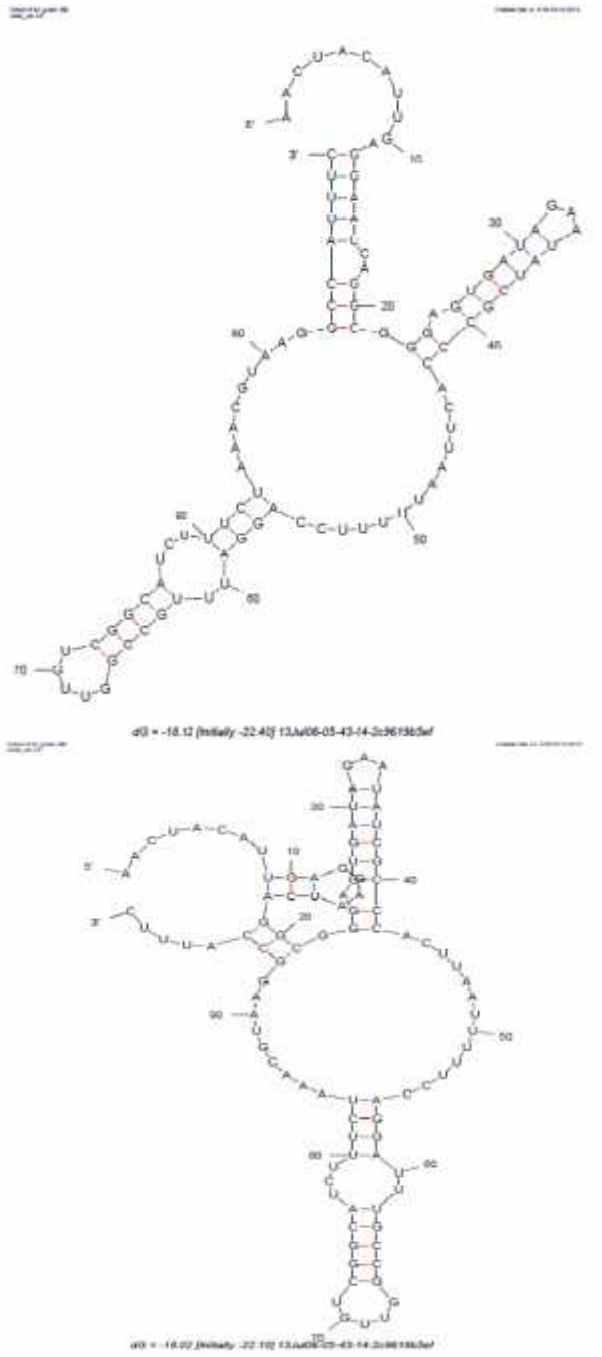






Mfolf structure of 5’ UTR of gltXP1(gltx112.6 to gltX112.11)





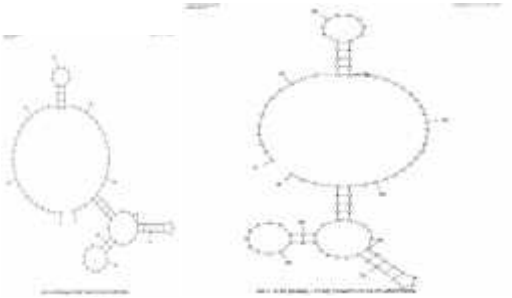
Mfold structures Upstream 100 nucleotides of gltxgene (gltx100.12, gltX100.13, gltX100.14, gltX100.15)

4.8.3 hemB gene(porphobilinogen synthase)

No details of hemB was available on ecocys and we selected 100 nucleotides from hemB gene in exact map position from 389051 to 388950 and 2 structures were obtained from mfold with comparable stability and only 3 nucleotide free at 3' region. As per database available for complete genome of Escherichia coli hemB is located as reverse complement.

GC content of upstream region was somewhat higher as compared to hemA and gltX. GC content was further concentrated near translation initiation site which was 52 % for upstream 50 nucleotide, 55% for upstream 29 nt 58% for upstream 31 nucleotide.

HemB					
SN	Promoter	Structure	dG(-ve)	Free nucleotide at 3'end	GC content (%)
1	Upstream 100	hemB100.27	9.08	3	47
2		hemB100.28	8.08	3	

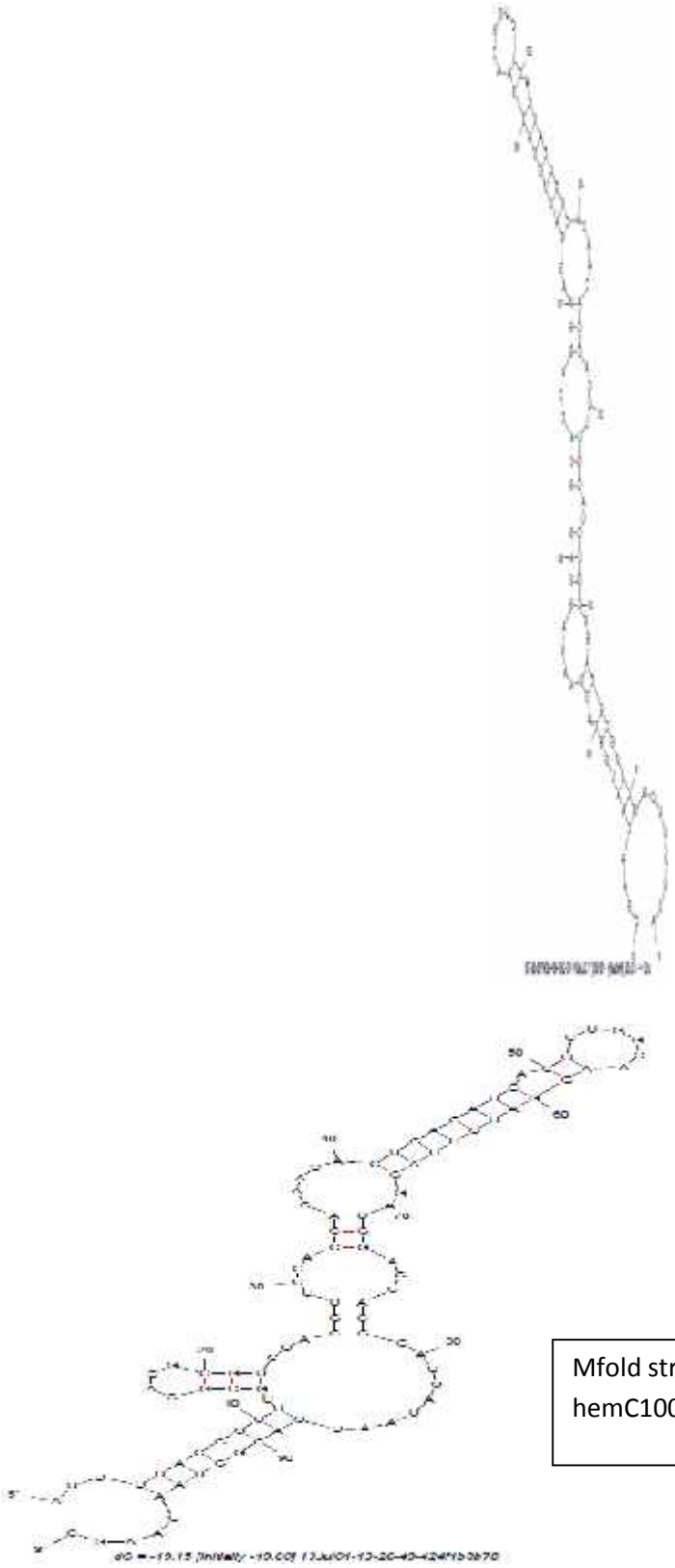


Mfold structure of UTR of hemB(hemB100.27,hemB100.28)

4.8.4 hemC(hydroxymethylbilane synthase)

hemC gene is in operon with hemChemDhemXhemY and is first gene in operon. There is one transcript reported but detail was not available. 100 nucleotide region between 3988989 to 3988890 (reverse complement) from database were selected for analysis. Although GC content of 5' UTR is same as of hemB but GC content was not concentrated on near 50 nucleotide from translation initiation site. GC content of upstream 50 nucleotides was same that is 48% and it was 50 % for upstream 30 nucleotide.

HemC					
SN	Promoter	Structure	dG(-ve)	Free nucleotide at 3'end	GC content(%)
1	Upstream 100	hemC100.29	20.6	5	48
2		hemC100.30	19.8	4	
3		hemC100.31	19.15	5	

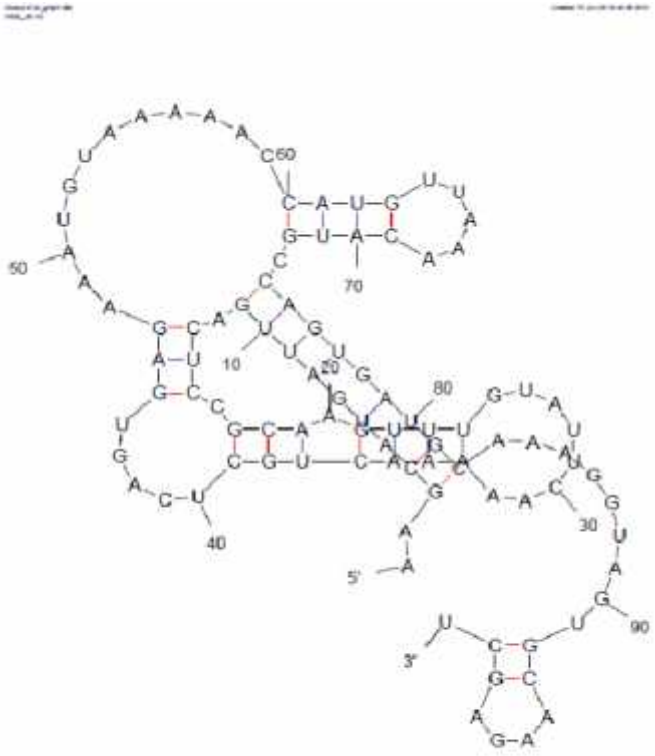


Mfold structure of upstream UTR of hemC 100.29, hemC100.30, and hemC100.31

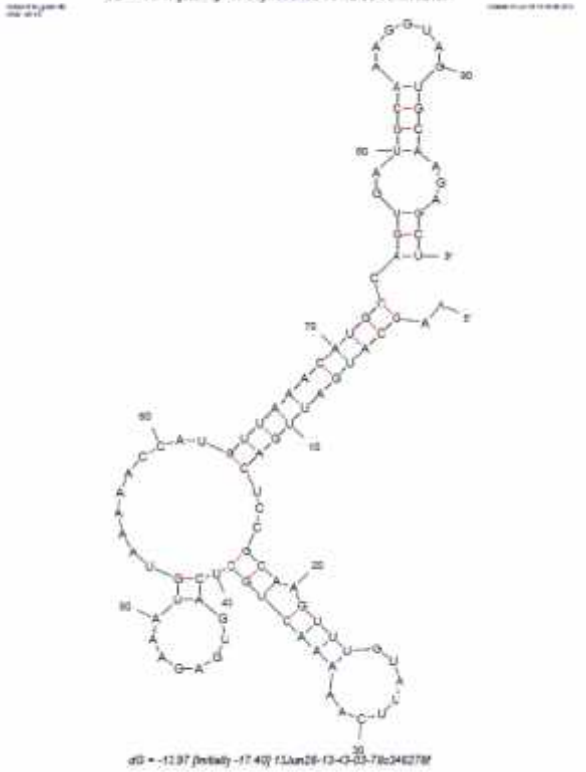
4.8.5 hemE gene(uroporphyrinogen decarboxylase)

We found unique result for *hemE*. Nearly no free 3' nucleotide for all 6 structures obtained from mfold and least GC content but higher stability as shown by Gibbs free energy. No details available for this transcript in ecocys and region from 4194826 to 4194795 of *E. coli* database. In *Escherichia coli* hemE gene found on operon with nudChemEnfi, where hemE in the middle. Sequence for mfold structure was taken from upstream of nudC gene, the first gene of nudChemEnfi operon.

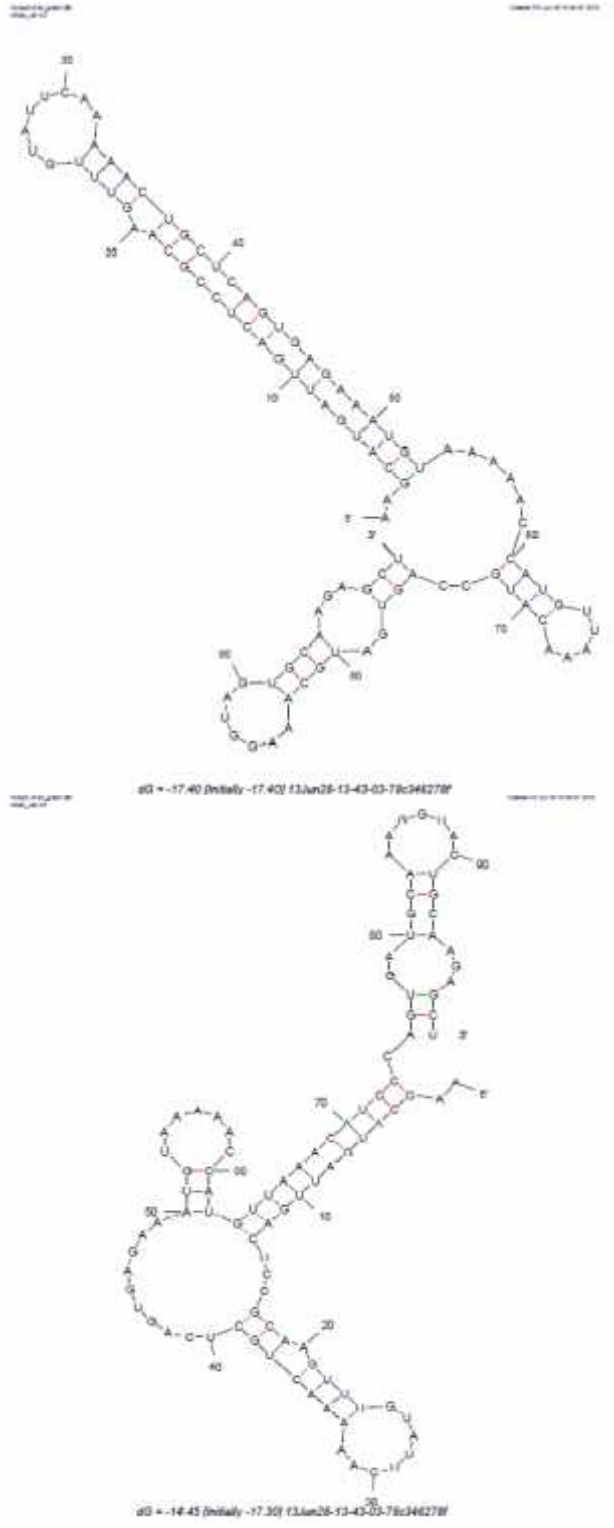
HemE					
SN	Promoter	Structure	dG(-)	Free nucleotide at 3'end	GC content(%)
1	Upstream 100	hemE100.32	14.17	1	39
2		hemE100.33	13.97	0	
3		hemE100.34	17.4	0	
4		hemE100.35	14.45	0	
5		hemE100.36	13.04	0	
6		hemE100.37	14.5	0	



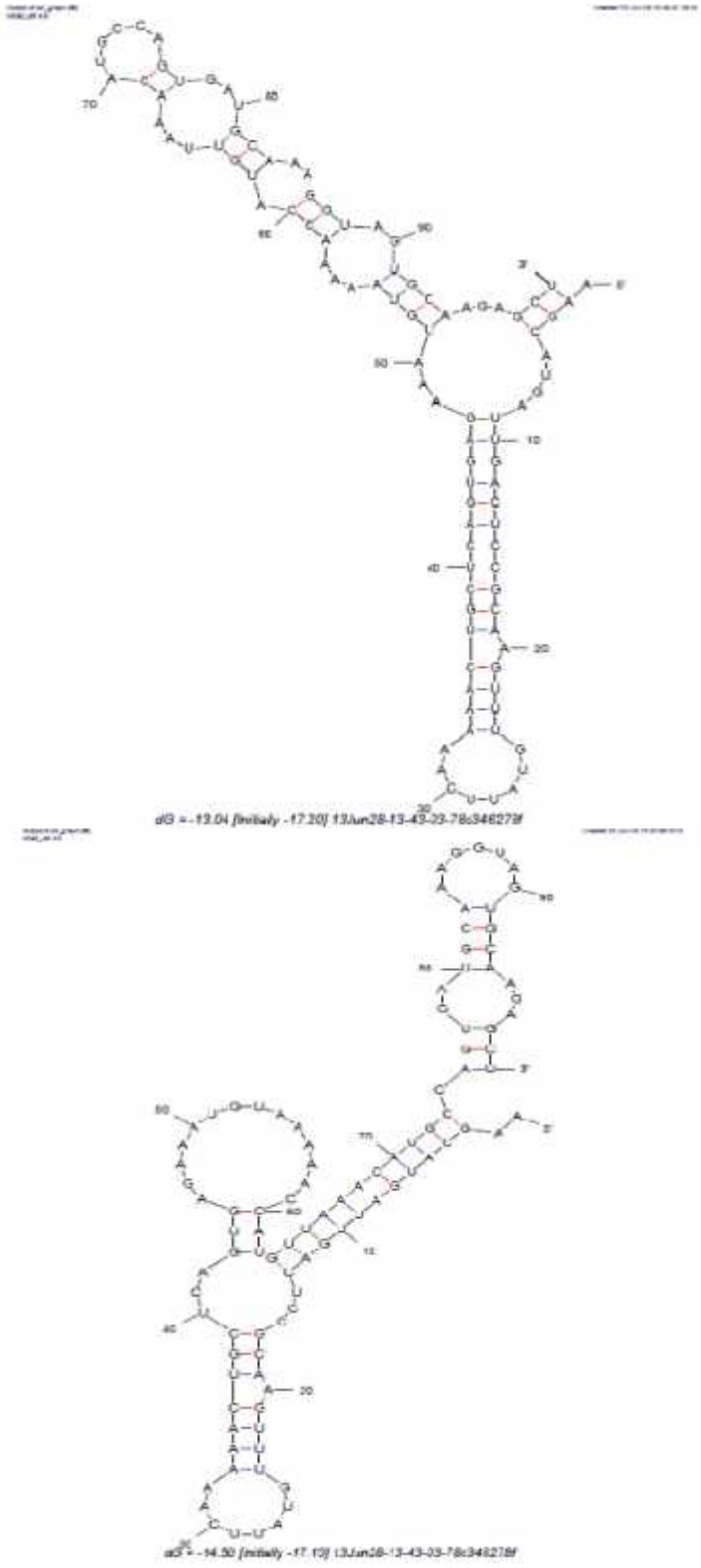
dG = -14.17 [initially -17.50] 13Jun28-13-43-03-78c34627ff



dG = -12.97 [initially -17.40] 13Jun28-13-43-03-78c34627ff



Mfold structure of hemE100.32- hemE100.35



Mfold structure of hemE100.36 and hemE100.37

4.8.6 hemG gene(protoporphyrinogen oxidase)

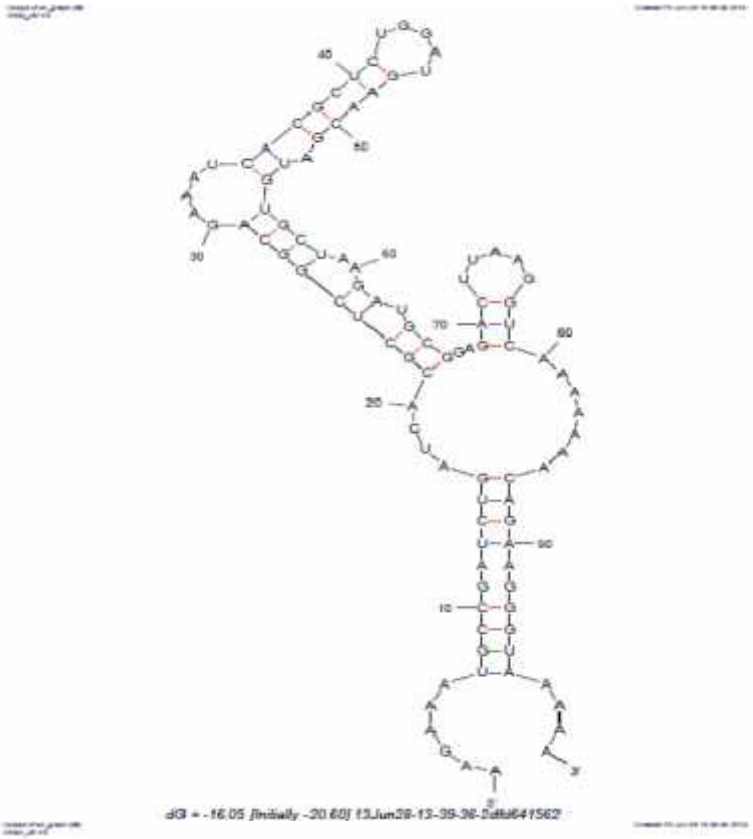
hemG is last gene of pepQyigZtrkHhemG operon, no details of this transcript was available in ecocys till this date. 5' UTR of 100 nucleotide showed 2 structures with 44% GC content and more GC content on -51 to -100 from translation initiation site(50%) and only 38% for location -1 to -50 from translation initiation site.

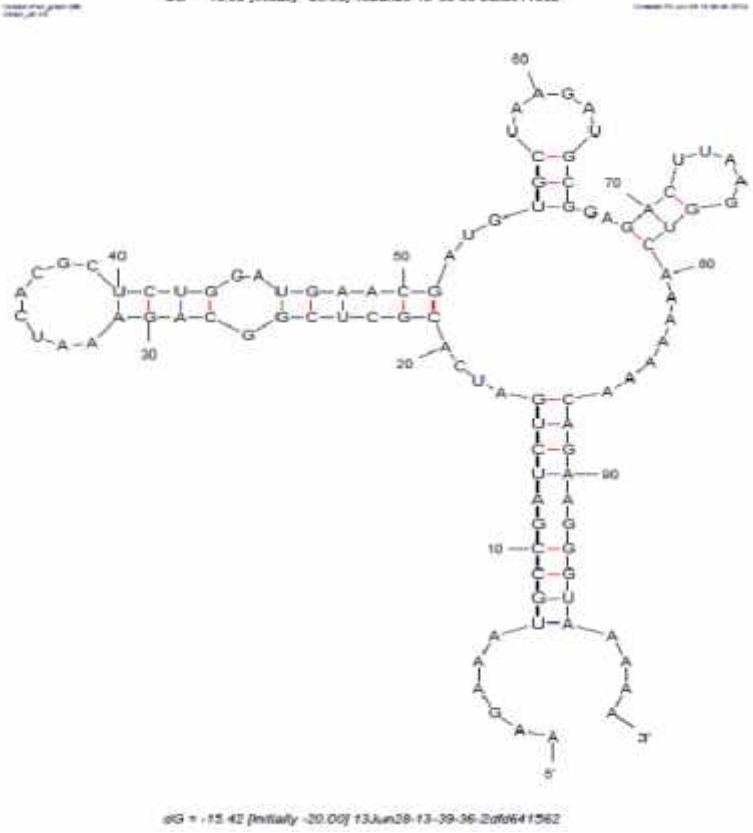
HemG					
SN	Promoter	Structure	dG(-ve)	Free nucleotide at 3'end	GC content(%)
1	Upstream 100	pepQ100.50	16.05	4	44
2		pepQ100.51	15.42	4	

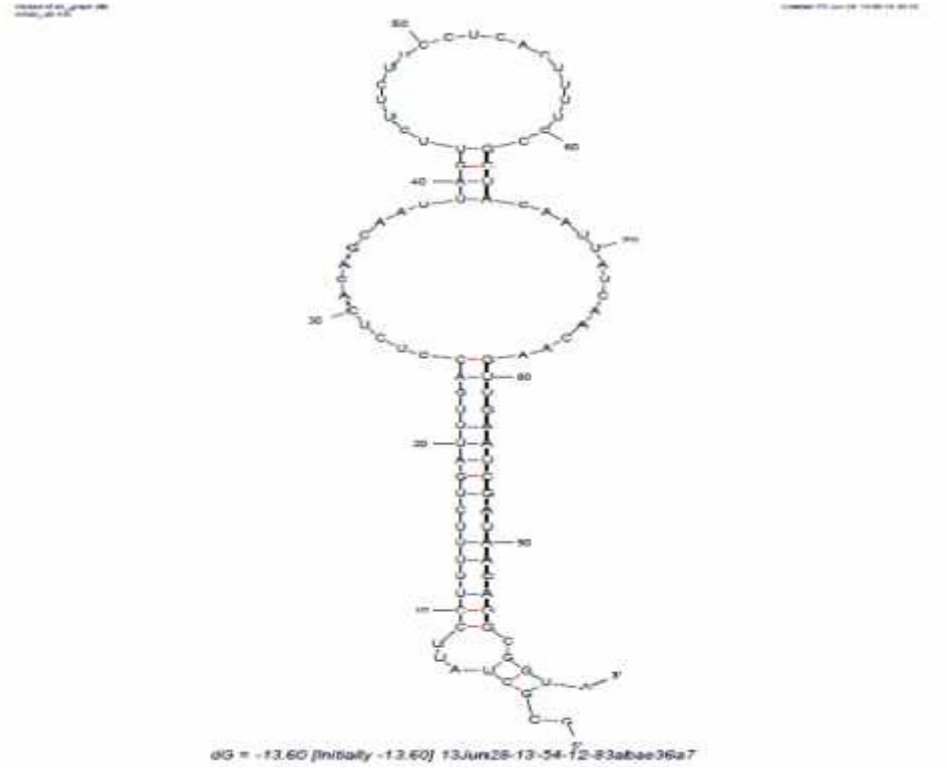
4.8.7 *hemH* gene(ferrochelatase)

This gene codes for the enzyme responsible for protoheme synthesis critically important gene for heme biosynthesis. Only 1 structure is possible from upstream 100 nucleotides from transcription initiation site of *hemH* and with only one nucleotide free for obtained structure. GC content was equally distributed in both half out of 100 upstream nucleotides which was equal to 40%.

hemH					
SN	Promoter	Structure	dG(-ve)	Free nucleotide at 3'end	GC content(%)
1	Upstream 100	hemH100.52	13.6	1	40


dG = -16.05 [Initially -20.60] 13Jun28-13-39-36-2dfd641562


dG = -15.42 [Initially -20.00] 13Jun28-13-39-36-2dfd641562

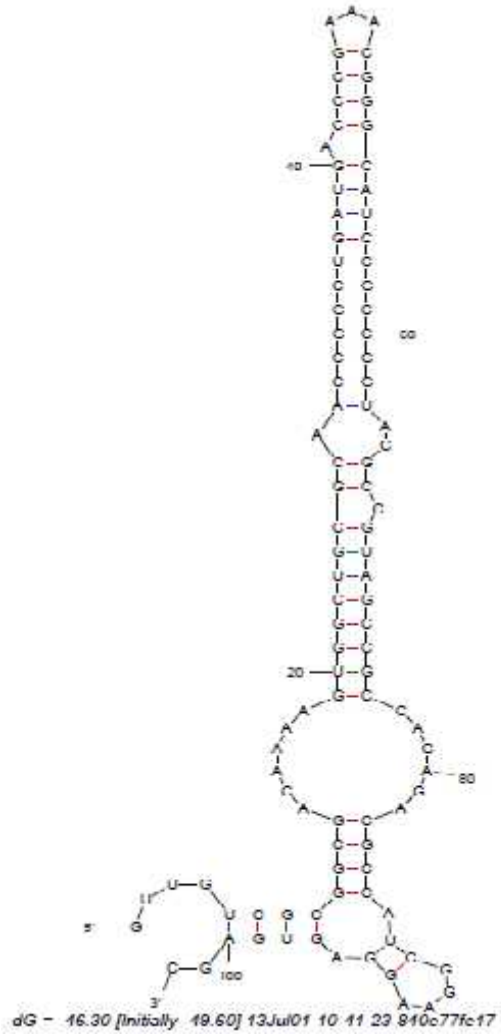


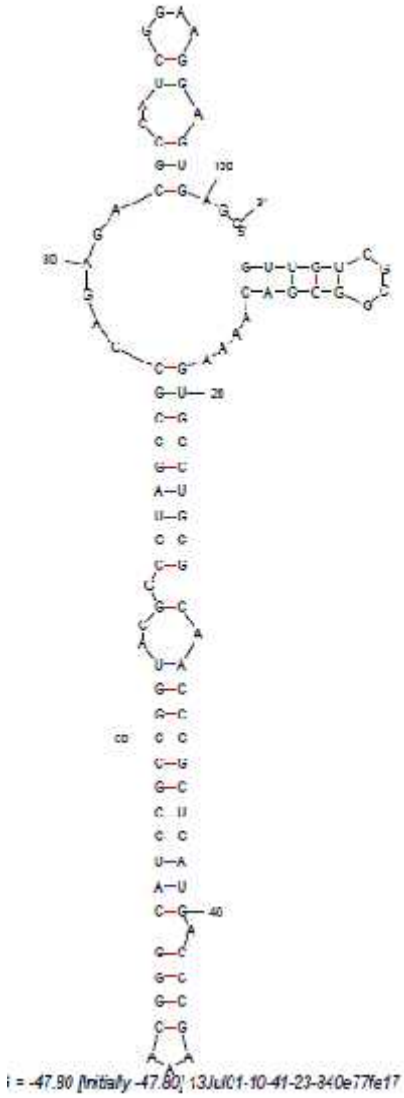
Mfold structure of pepQ100.50, pepQ100.51and hemH100.52

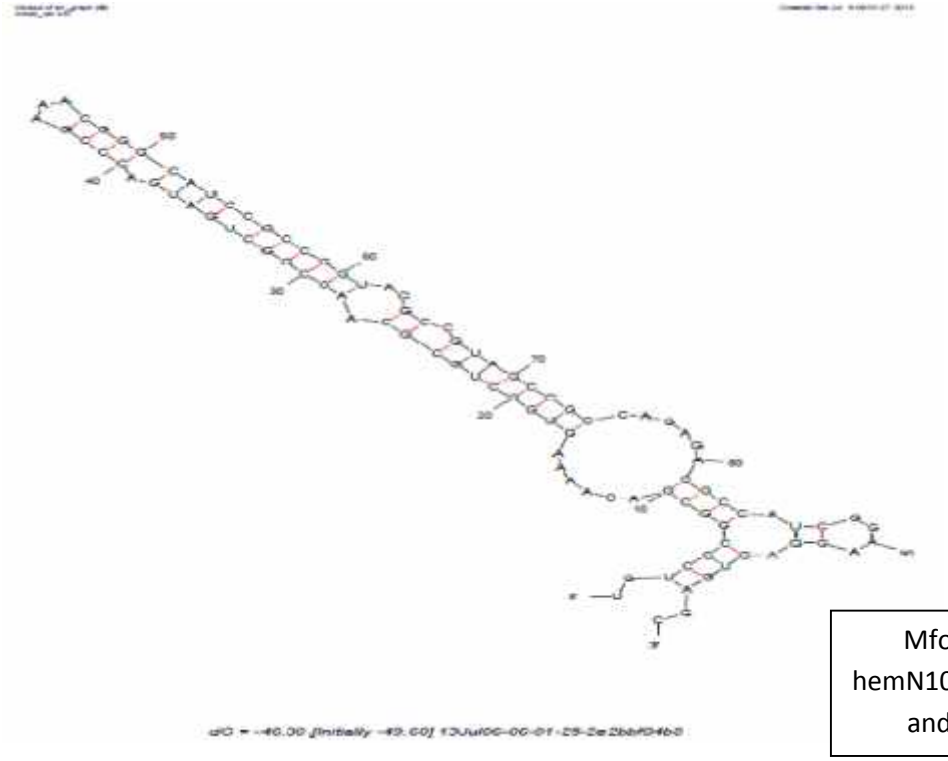
4.8.8 *hemN* gene (coproporphyrinogen III dehydrogenase)

Exceptional gene of heme biosynthesis genes having exceptionally high content of GC content(66% for all transcript 102 nt and 100 nt)and most stable structure as confirmed by gibbs free energy value. Upstream UTR for hemNP is 102 in length and initiating from 4049966.to 4050067 Free nucleotide at 3' end of mfold structure were also less i.e. 2 and 3 for two structures. GC content for nearest 50 upstream nucleotide from translation initiation site was 68%.

HemN					
SN	Promoter	Structure	dG(-)	Free nucleotide at 3'end	GC content(%)
1	hemNP	hemN102.65	46.3	2	66
2		hemN102.66	47.8	3	
3	upsream 100	hemN100.67	46.3	2	66







Mfold structure of hemN102.65, hemN102.66 and hemN111.67

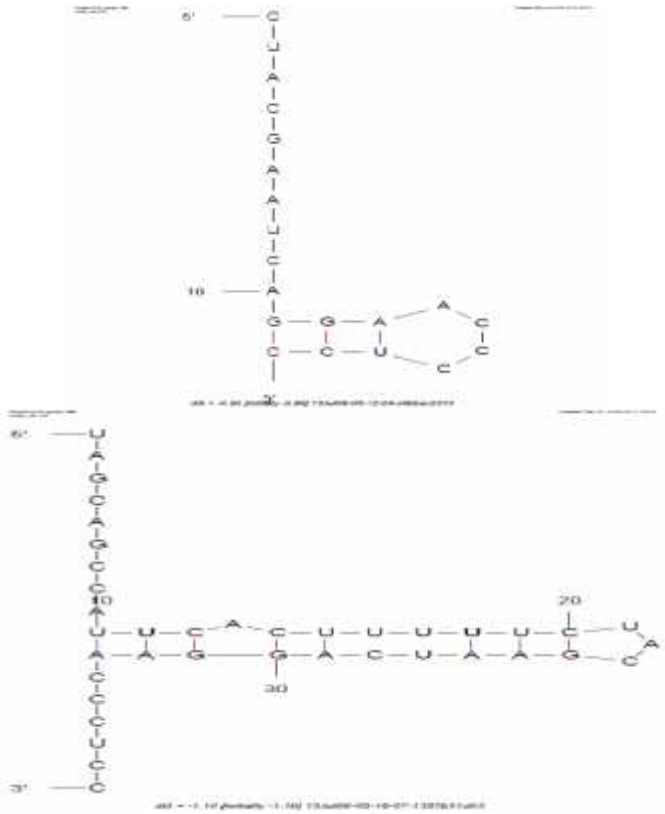
4.8.9 hemL gene(glutamate-1-semialdehyde aminotransferase)

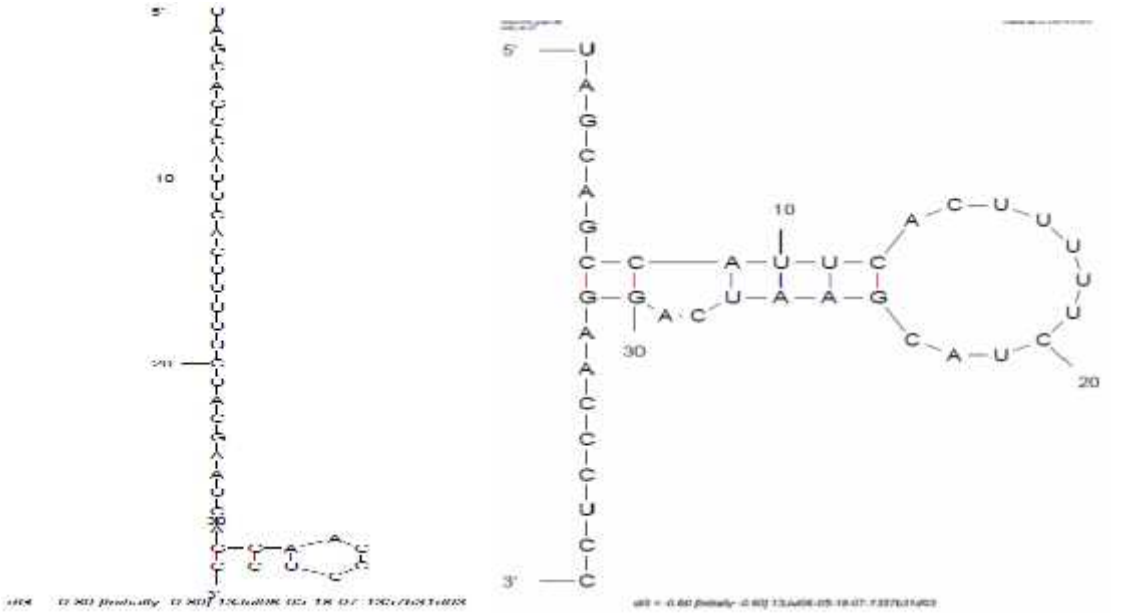
hemL gene has 3 known transcripts and it catalyse the reaction to convert GSA to delta amino levulinic acid, next reaction catalysed by hemA so called most regulated reaction in heme biosynthesis. hemLp1, hemLp2, hemLp3 transcripts has 1, 4 and 5 mfold structures respectively. GC content wes drastically higher for first transcript for hemLp1 (upstream 20 nucleotides from translation initiation site).

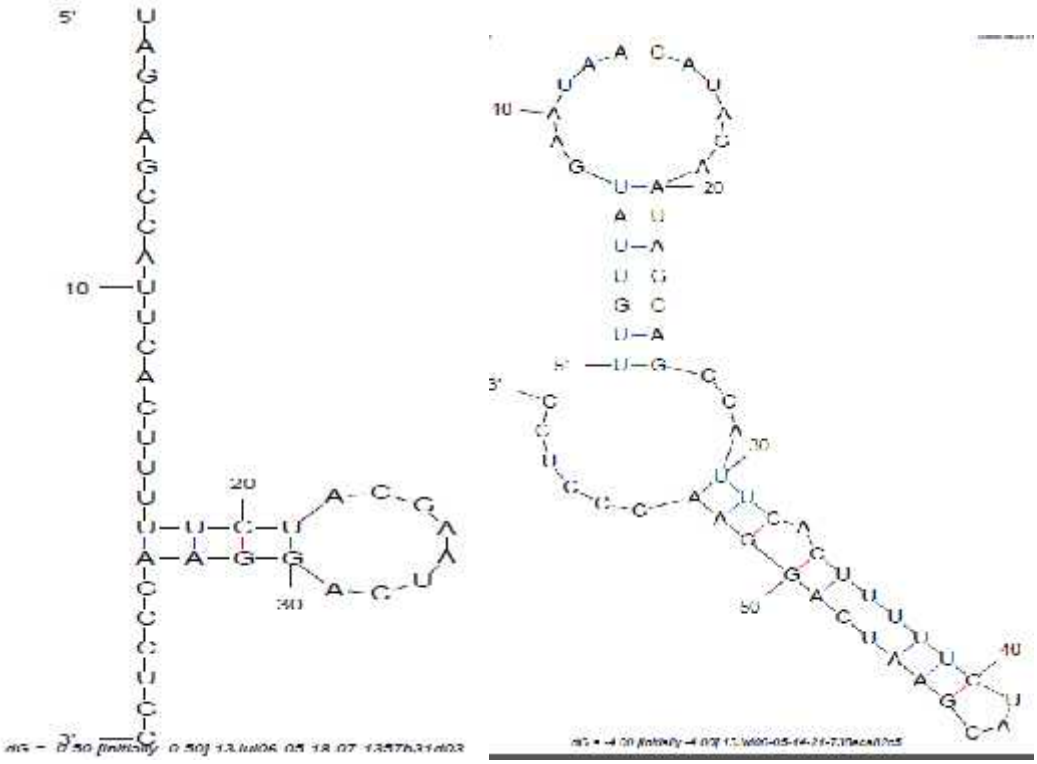
4.8.10 hemF gene(coproporphyrinogen III oxidase)

hemF catalyses the reaction catalysed by *hemN* responsible for formation for protoporphrinogen IX. Two known transcripts are known for this gene one with promotor hemFp, 20 nucleotide upstream from hemF and next with promoter amiP, 111 nucleotide upstream from *amiA* gene of amiAhemF operon.

HemL					
SN	Promoter	Structure	dG(-ve)	Free nucleotide at 3'end	GC content (%)
1	hemLp1(-20)	hemL20.53	0.8	0	55
2	hemLp2(-39)	hemL39.54	1.1	6	46
3		hemL39.55	0.8	0	
4		hemL39.56	0.6	8	
5		hemL39.57	0.5	6	
6	hemlp3(-59)	hemL59.58	4	6	37
7		hemL59.59	3.7	0	
8		hemL59.60	3.4	6	
9		hemL59.61	4.7	8	
10		hemL59.62	3	0	
11	Upstream 100	hemL100.63	14.79	8	40
12		hemL100.64	13.85	8	

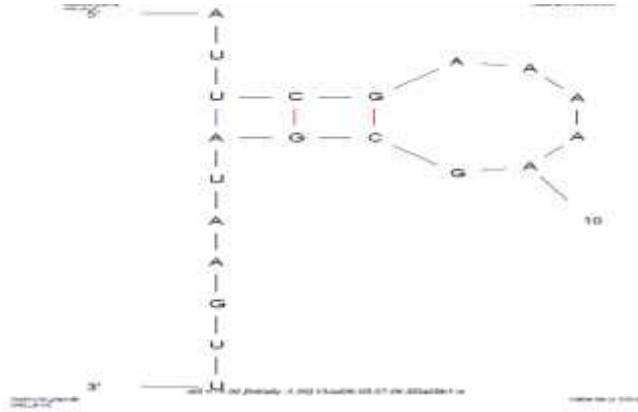


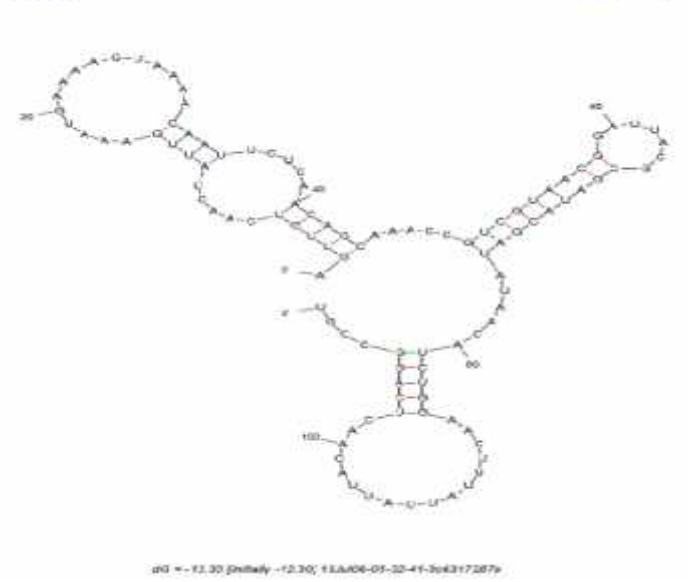


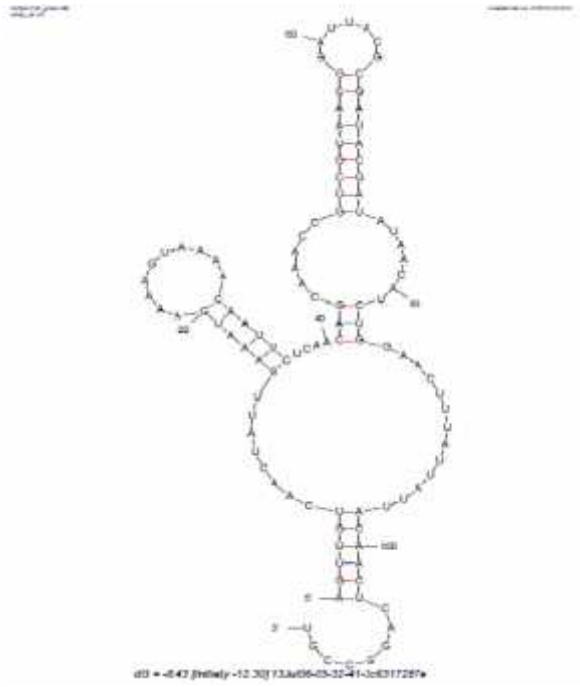


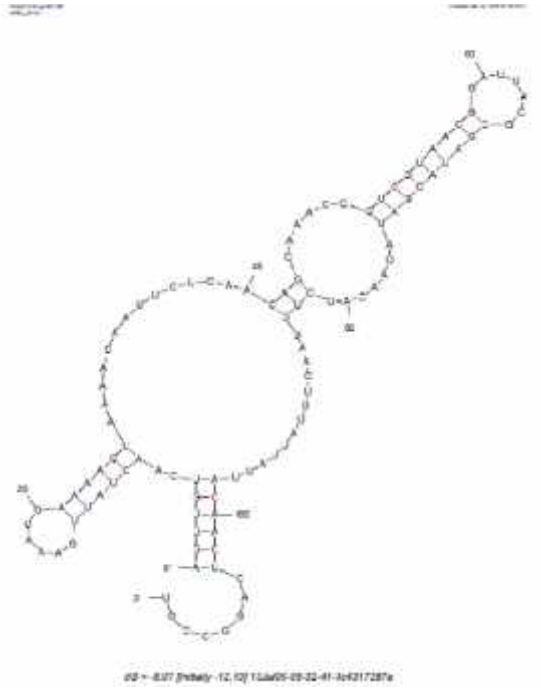
Mfold structure of hemL20.53, hemL39.54 to hemL39.57 and hemF59.58

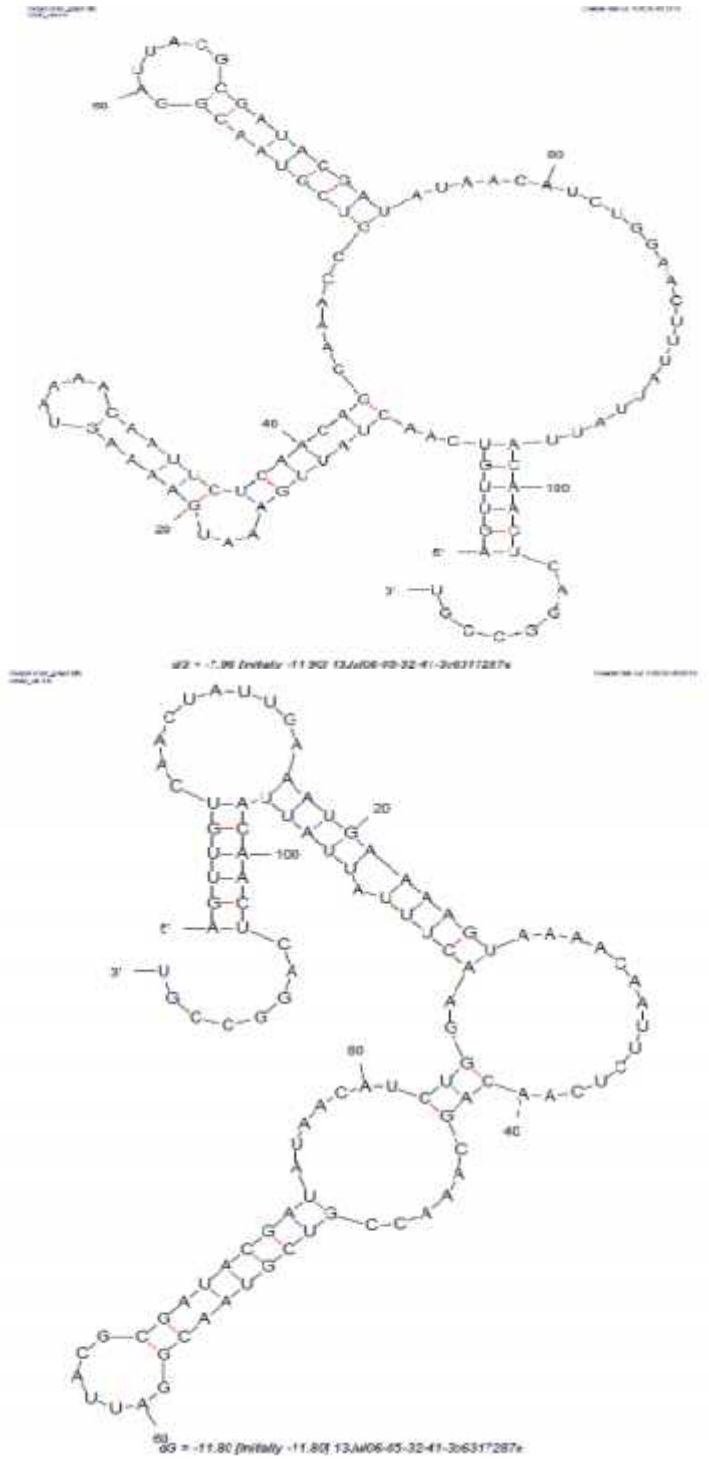
HemF					
SN	Promoter	Structure	dG(-)	Free nuvleotide at 3'end	GC content(%)
1	hemFp(-20)	hemF20.38	1	6	30
2	amiP(-111)	amiA111.39	12.3	4	36
3		amiA111.40	8.43	8	
4		amiA111.41	8.07	8	
5		amiA111.42	7.96	8	
6		amiA111.43	11.8	8	
7		amiA111.44	11.7	3	
8	upstream 100	hemF100.45	10.8	10	38
9		hemF100.46	5.93	10	
10		hemF100.47	10	6	
11		hemF100.48	9.8	5	
12		hemF100.49	5.63	10	



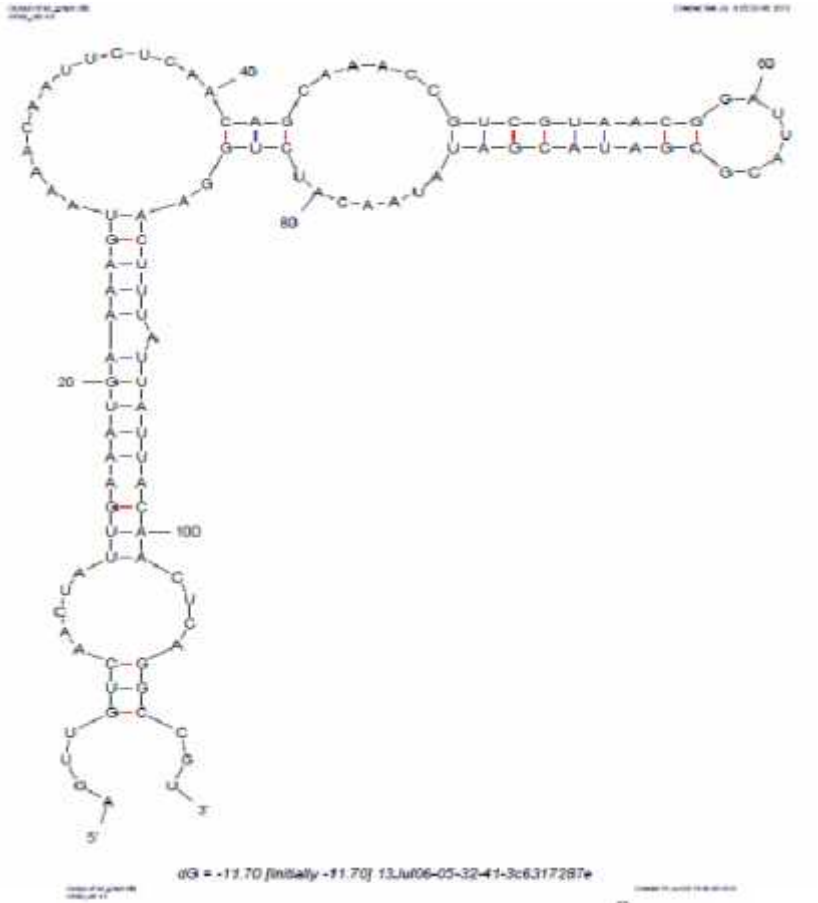




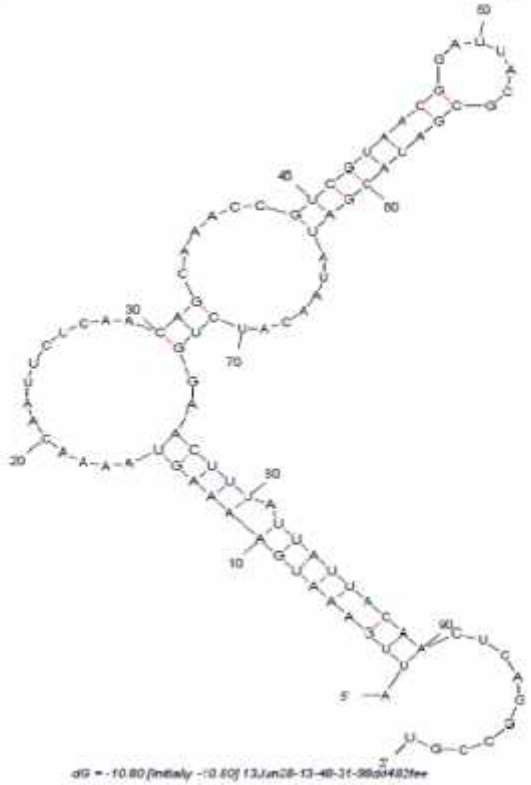




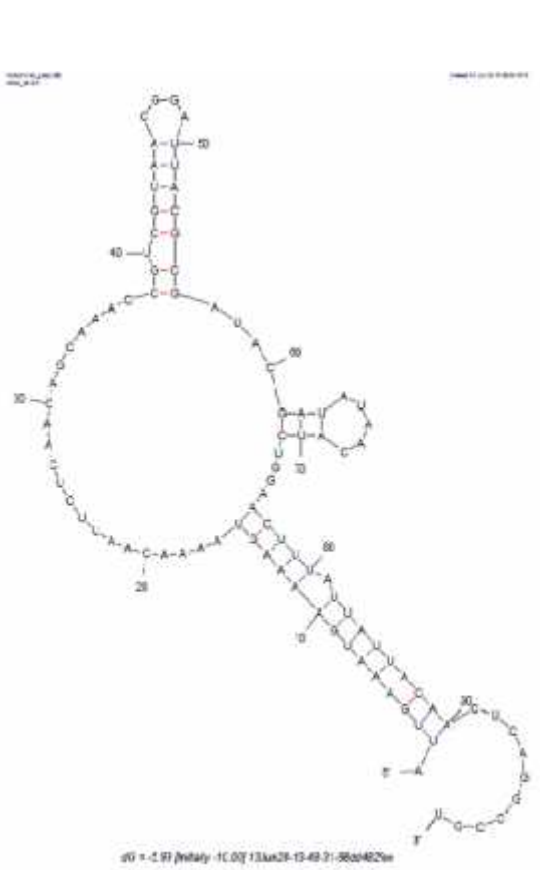
Mfold structure of hemF20.38, ami111.39 to ami111.43



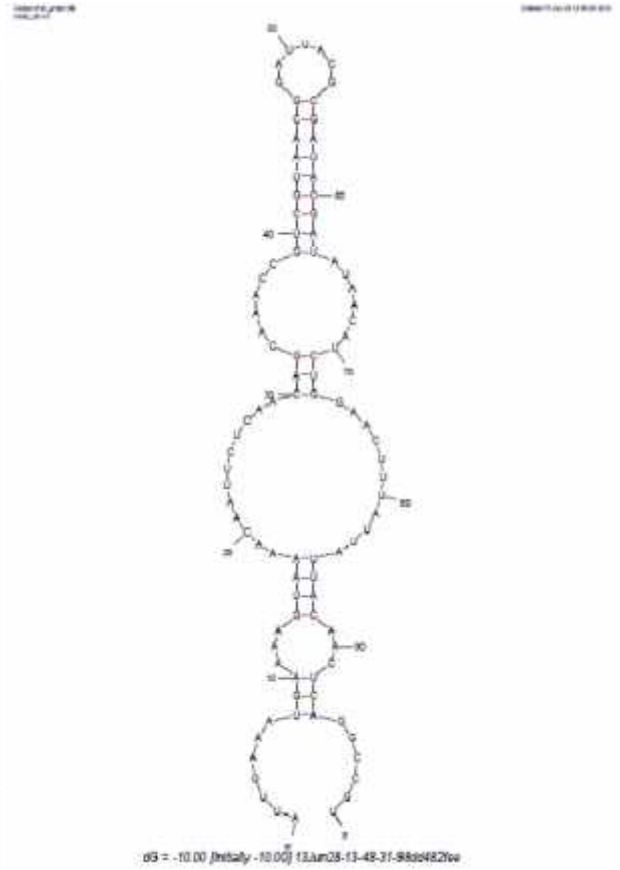
dG = -11.70 [initially -11.70] 13Jul06-05-32-41-3c6317287e



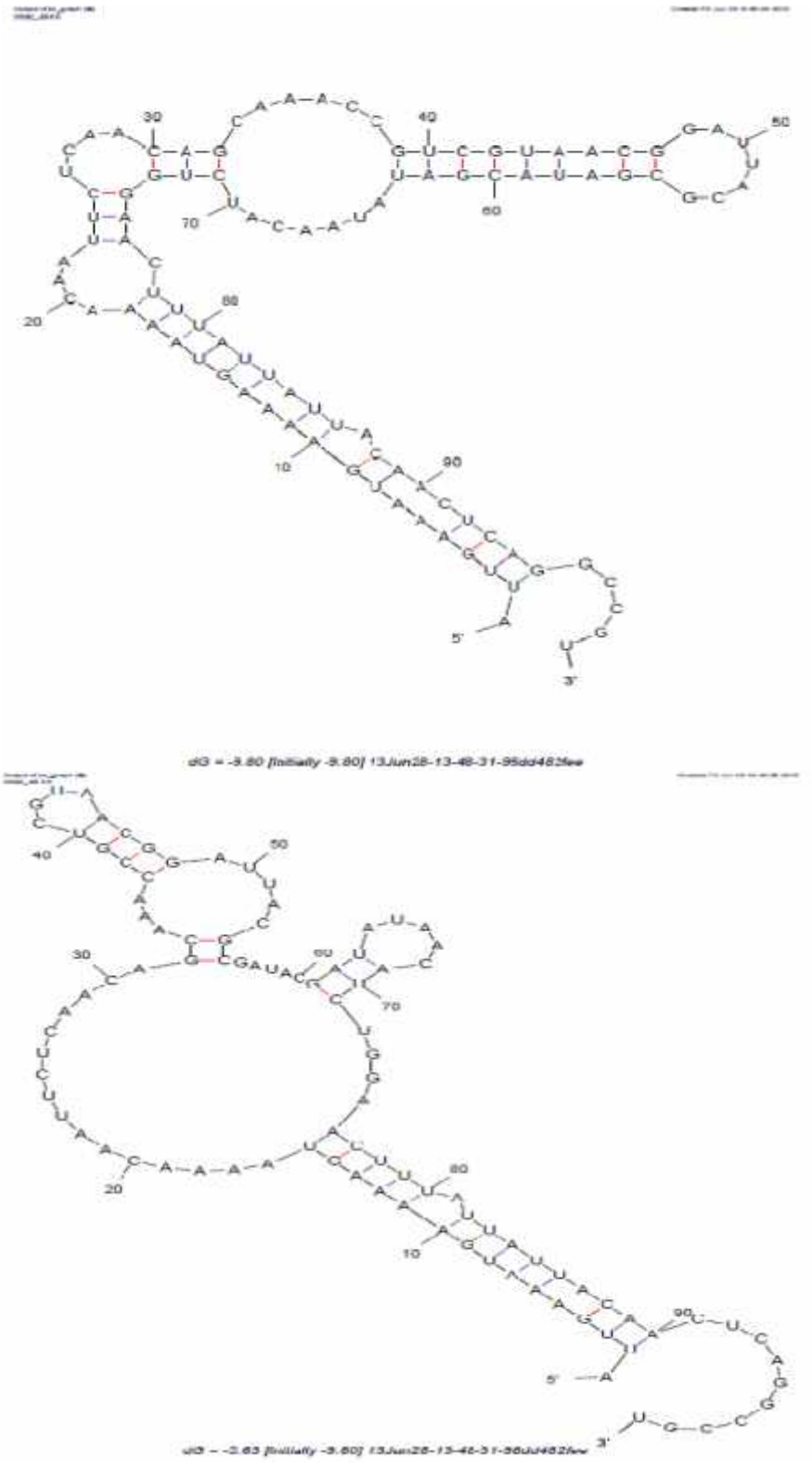
dG = -10.80 [initially -10.80] 13Jun28-13-48-31-98d0482fee



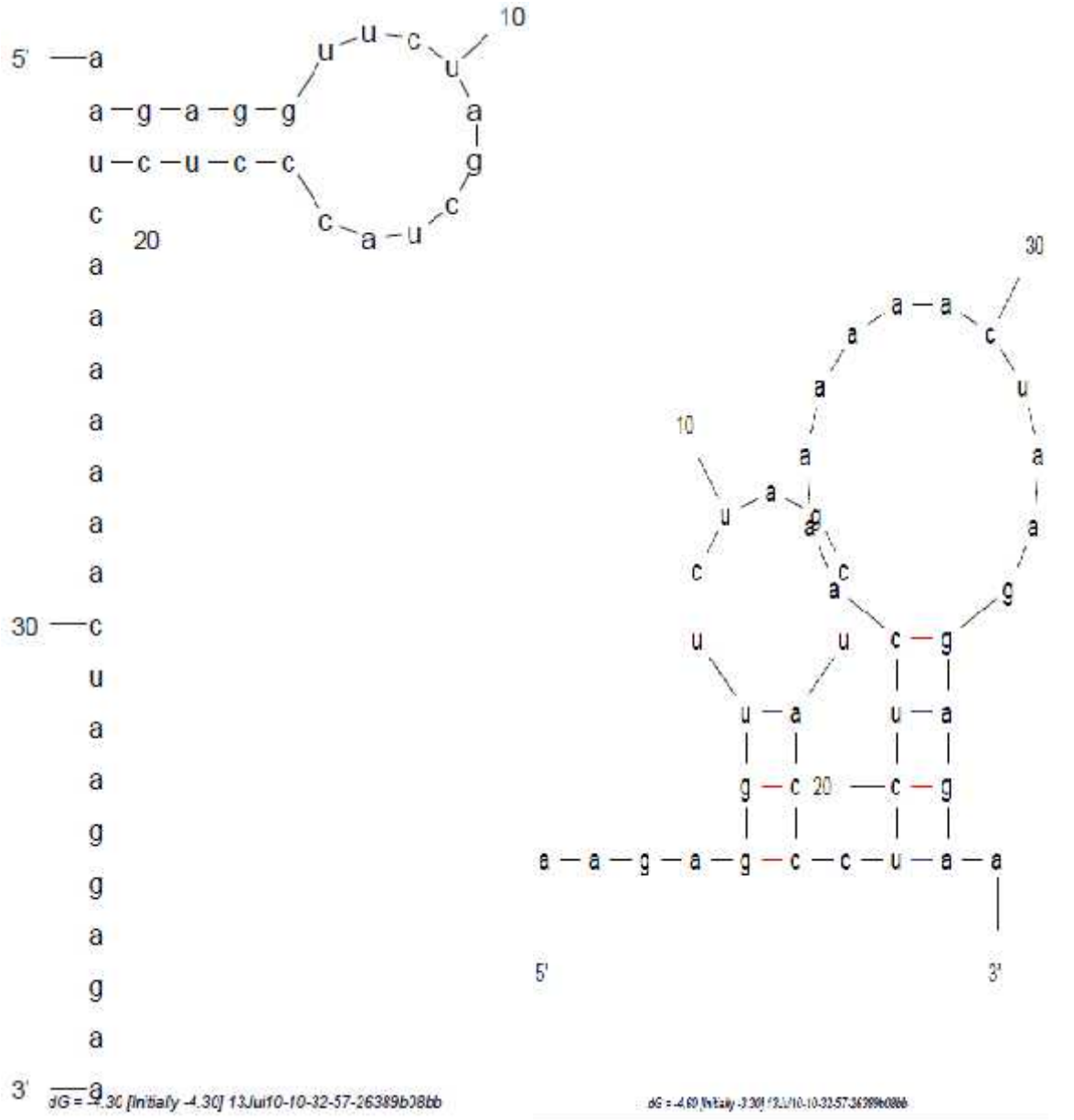
dG = -5.90 [Initially -10.00] 13Jun28-13-48-31-98dd482be



dG = -10.00 [Initially -10.00] 13Jun28-13-48-31-98dd482be



Mfold structure of ami111.44 and hemF100.45 to hemF100.49



Mfold structure of known riboswitch PreQ

4.9 Sequence similarity study of upstream UTR of heme biosynthetic genes

We compared all UTR of heme biosynthetic genes of *Escherichia coli* from NCBI BLAST/sequence alignment. Input sequence was taken for UTR of hemA which was compared with other genes subsequently. Length of 100, 80 and 50 were taken for hemA and same length was taken for other genes as hemA was taken.

During our analysis no similarity was obtained for any genes with hemA when all parameter was taken highly similar, more similar, and somewhat similar from BLAST page. In each cases no similarity were found.

4.10 Discussion of Structure analysis

In this study, 5' untranslated region (UTR) of all of the ten genes involved in heme biosynthesis has been analyzed using mfold. We have focused on the secondary structure stability on the basis of thermodynamic properties GC percentage in UTR region and free availability of ribosome binding site during formation of secondary structure. Recently a novel regulation mechanism for regulation of gene has been proposed, which explains the presence of riboswitch at 5' region leader sequence of the gene. Riboswitch regulates the gene expression at transcription and translation level. Riboswitch is basically a folded secondary structure which contains aptamer binding domain and expression platform. Aptamer binding domain is in the form of stem loop structure and is the site for binding of ligand. Ligand, in most of the case is the final metabolite of particular metabolic pathway or the intermediate metabolite. Once the ligand is bound to aptamer region the secondary structure is even more stabilized, thus repressing the gene expression. Riboswitch regulation is mostly negative and results either truncated mRNA or it prevents translation by preventing access of ribosome to ribosome binding site. This study has tried to stipulate if such riboswitch like mechanism is involved in heme biosynthesis pathway. Different secondary structure of 5' region leader sequence including almost 100 bp have been developed and stability of those secondary structures has been analyzed on the basis of different parameters such as free energy value, overall GC percentage of leader sequence and secondary folding at ribosome binding site to prevent access to ribosome.

While performing the folding of almost 100 bp UTR of *hemB* by web based tool mfold, we found two possible secondary structures. In both of the structures all of the base pairs at extreme 3' end are complementary to form a stem loop structure and there are only three nucleotides free at 3' region of the leader sequence. Thus this suggests that ribosome binding site, which very near to the initiation codon is hidden in secondary structure and suggests that ribosome is hindered for getting access to the ribosome binding site. Thus we can stipulate that this gene might be under regulation at translational level. However,

stability of the secondary structure is less as the free energy of structure is comparatively high (Gibbs free energy -9). In addition 5' UTR of *hemB* contains quite higher GC percent which is 47% which may depict that secondary structure could be quite stable. However, the sequence is not in the pattern of inverted repeat sequence, which therefore is making secondary structure quite unstable despite having higher GC percentage.

Similarly transcription initiation site of *hemC* is not characterized yet. Therefore, we performed the folding of UTR taking almost 100 base pair region upstream to start codon. Folding resulted three possible structures and all three structures seems to be highly stable while taking consideration of free energy value. These structures have very low free energy value which has been found to be around -20. Furthermore, possible ribosome binding site GACGGUAA bound to form stem loop structure in all three structures. Similarly, GC percent is 48 which is quite same as *hemB*. However inverted repeated pattern arrangement of sequence is making secondary structure more stable then the secondary structure of UTR of *hemB.* Thus form all these fact we can assumed the presence of riboswitch like regulation mechanism in *hemC,* more confidently than in case of *hemB*.

Regulation of *hemA* gene has been extensively studied and regulation is found to be dependent on feedback inhibition (Elliot *et al*., 1996; Beck *et al*., 2005; Luo *et al*., 2012). HemA has two transcripts hemP1 and hemP2. Both of the transcripts were taken for developing the secondary structure of UTR region. In addition secondary structures of 100 bp UTR region were also developed. Total of 10 structures were obtained and free energy value of secondary structure developed from the UTR of hemP2 was found to be very low in comparison to all other structures. Thus hemp2 seems to be more regulated via secondary structure mechanism. As hemp1 is abundantly found in cellular system in aerobic condition (Olson *et a 2008l*), this further substantiates that hemp2 is more tightly regulated and regulation mechanism might be similar to that of riboswitch.

From all of the structure developed for different ten gene involved in heme biosynthesis pathway, we found that structure of UTR region of two genes *hemN* and *hemH* are very much similar to the previously developed riboswitch structures. In case of *hemN* total of three structures were obtained. Free energy value is exceptionally low in comparison to all other structures and GC percent is exceptionally high. In addition, 3' region where ribosome

binding sequence is located is not free. All these fact depicts the very high chance of regulation of this gene by secondary loop structure which might be riboswitch and we also assumed more strongly that this could be a possible regulatory region for *hemN* gene and rate limiting step in heme biosynthesis pathway. However further detail study and experimental validation is needed.

In Escherichia coli hemN calalyzes the formation of protoporphyrinogen IX from coproporphyrinogen III. protoporphyrinogen IX can also be synthesized from precorrin-2, a precursor to vitamin B12, Siroheme andcoenzyme F430. Thus after deviation from UroporphyrinogenIII to B12 and heme pathway there could be bypass pathway to heme synthesis from precorrin-2 to coproporphynogen III. From this intermediate there is hemN to convert to protoporphyrinogen IX and being exceptional feature in hemN nucleotide there is critical regulation for heme synthesis. In this step there is critical regulation beside hemA. This regulation is possibly to control of coproporphynogen III synthesis from precorrin 2. In bypass pathway there could be feedback regulation by coproporphynogen III and to limit heme synthesis.

Similarly in case of *hemH* only one secondary structure was found which form a perfect stem loop structure. Thermodynamic stability on the basis of free energy seems to be quite more stable and GC percent is 40% and nucleotide at ribosome binding site region is completely bounded. Though the free energy of secondary structure of UTR of *hemH* is quite higher than the secondary structure of UTR of *hemN*, the stem loop structure is very much similar to some of the other riboswitch like SAM (Montange et al 2006).

From all these structural analysis we have assumed that the regulation of two gene hemN and hemH could be via riboswitch. Although we have employed dry lab methodology for analysis of structure, it gives quite confident assumption for presence of riboswitch in regulation of *hemN* and *hemH* gene. However wet lab research much be conducted for further validation of assumption of posited by present study.

CHAPTER FIVE

CONCLUSION

Gene control by Riboswitchs mechanisms are new mechanism and till now observed mostly in microorganism. It is regulated by metabolite that directly binds to mRNA and changes the confirmation of mRNA and alters gene regulation either by terminating transcription and translation. As riboswitch mediated regulation of gene is dependent on binding of ligand to the aptamer region of riboswitch, artificial ligand could be developed which would mimic the cellular metabolite ultimately repressing the gene. This could be a new paradigm for developing new efficient drugs. As riboswitch mechanism is based mostly on secondary folding structure at 5' UTR, we initially studied the secondary structure of the 5' UTR of almost ten genes involved in heme biosynthesis pathway. We selected 5' UTR region of all biosynthetic genes in this study. On the analysis of structures of all UTR we found exceptional structure of hemN with very high GC content (66%) and highest stability of structure that have new role in gene regulation and regulation might be similar to the riboswitch mechanism. Least GC content (by %) is with hemF and hemL(second promoter). HemLp1 gene UTR has compareteveley less negative value of change in Gibbs free energy although having high GC content. Therefore from this study we have stipulated that *hemN* and *hemH* gene have been regulated by riboswitch mechanism.

Furthermore, we designed primers manually without using any web based tools for PCR amplification of *hemA*. We amplified *hemA* gene from Escherichia coli genome and successfully cloned it in to pUC19 plasmid with digestion of *EcoRI* and *XbaI*. The same gene was cloned into pET15b which is a expression vector plasmid with digestion of *EcoRI* and *NdeI*.

Beside Laboratory techniques like PCR, restriction digestion, cloning, transformation, gel purification we mostly focused on untranslated region of heme biosynthetic genes considering that there is significant role of UTR region in gene regulation. Three dimensional structure of any macromolecule depend on two dimensional structures. Thus 2D structure has also very important role in stability, prediction of gene regulation, intermolecular interaction etc.

References

Abreu-Goodger C and Merino E (2005) RibEX: A Web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res*. **33**: W690–W692. doi: 10.1093/nar/gki445.

Ades IZ, Stevens TM, Drew PD (1987) Biogenesis of embryonic chick liver d-aminolevulinate synthase: regulation of the level of mRNA by hemin. *Arch. Biochem. Biophys.* **253**:297–304.

Ailion M, Bobik TA, Roth JR (1993) Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of Salmonella typhimurium. *J. Bacteriol*. **175**:7200–7208.

Anraku Y, Gennis RB (1987) The aerobic respiratory chain of Escherichia coli. *Trends Biochem. Sci***. 12**:262–266.

Bashirullah A, Cooperstock RL, Lipshitz HD (2001) Spatial and temporal control of RNA stability. *Proc Natl Acad Sci USA* ,**98**:7025-7028.

Beck CF, Vasileuskaya Z, Oster U (2005) Mg-Protoporphyrin IX and Heme Control *HEMA*, the Gene Encoding the First Specific Step of Tetrapyrrole Biosynthesis, in *Chlamydomonas reinhardtii.* doi: 10.1128/EC.4.10.1620-1628.2005

Bengert, P and Dandekar T (2004). Riboswitch finder—A tool for identification of riboswitch RNAs. *Nucleic Acids Res*. 32: W154–W159. doi: 10.1093/nar/gkh352.

Boynton TO, Daugherty LE, Dailey TA, Dailey HA (2009) Identification of Escherichia coli HemG as a novel, menadione-dependent flavodoxin with protoporphyrinogen oxidase activity. *Biochemistry* **48**(29):6705-11. PMID: 19583219

Breckau D, Mahlitz E, Sauerwald A, Layer G, Jahn D (2003). Oxygen-dependent coproporphyrinogen III oxidase (HemF) from Escherichia coli is stimulated by manganese. *J Biol Chem* **278**(47);46625-31

Brun YV, Sanfacon H, Breton R, Lapointe J (1990) Closely spaced and divergent promoters for an aminoacyl-tRNA synthetase gene and a tRNA operon in Escherichia coli. Transcriptional and post-transcriptional regulation of gltX, valU and alaW. *J Mol Biol* **214**(4);845-64. PMID: 2201777

Caughey WS, Smythe GA, O'Keeffe DH, Maskasky JE, Smith MI (1975) Heme A of Cytochrome c Oxidase Structure And Properties: Comparisons With Hemes B, C, And S And Derivatives. *J. Biol. Chem.* **250** (19): 7602–7622. PMID 170266

Cheah MT, Wachter A, Sudarsan N, and Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**: 497–500.

Chen JJ,& London IM (1995) Regulation of protein synthesis by heme-regulated eIF-2 alpha kinase. *Trends Biochem Sci* **20**,105-108.

Chiu JT, Loewen PC, Switala J, Gennis RB, Timkovich R (1989) Proposed structure for the prosthetic group of the catalase HPII from Escherichia coli. *J. Am Chem. Soc*. **111**:7046–7050.

Conne B, Stutz A, Vassalli JD (2000) The 3' untranslated region of messenger RNA: a molecular 'hotspot' for pathology? *Nat Med* **6**: 637-641.

Corbino KA, Barrick JE, Lim J, Welz R, Tucker BJ, Puskarz I, Mandal M, Rudnick ND, and Breaker RR (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in _-proteobacteria. *Genome Biol*. **6**: R70. doi: 10.1186/gb-2005-6-8-r70.

Corbino, K A, Barrick, J E, Lim, J, et al (2005) Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biology*, 6:R70

Cynthia L et al (2007) Heme Concentration Dependence and Metalloporphyrin Inhibition of the System I and II Cytochrome c Assembly Pathways, *Journal Of Bacteriology*, p 455–463.

Dailey TA, Dailey HA (2002) Identification of [2Fe-2S] clusters in microbial ferrochelatases. *J Bacteriol* 184(9);2460-4. PMID: 11948160.

E Rauxa et al (2000) Biosynthesis of cobalamin (vitamin B12): a bacterial conundrum. CMLS, *Cell. Mol. Life Sci*. **57**: 1880–1893.

Edwards AL & Batey RT (2010) Riboswitches: A Common RNA Regulatory Element. *Nature Education* **3**(9):9.

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**(6289);203-6. PMID: 2203971.

Fuchs RT, Grundy FJ, and Henkin TM (2006) The SMK box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol*. **13**: 226–233.

Gelfand MS, Mironov AA, Jomantas J, Kozlov YI, and Perumov DA (1999) A conserved RNA structure element involved in the regulation of riboflavin biosynthesis genes. *Trends Genet*. **15**: 439–442.

Gilbert SD, Stoddard CD, Wise SJ, Batey RT (2006) Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol*, **359**:754-768.

Grundy FJ and Henkin TM (1993). tRNA as a positive regulator of transcripton antitermination in B. subtilis. *Cell* **74**: 475–482.

Grundy FJ., Winkler WC, and Henkin TM (2002) tRNAmediated transcription antitermination in vitro: Codon–anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci*. **99**: 11121–11126.

Grundy FJ and Henkin TM (1993) tRNA as a positive regulator of transcripton ntitermination in B. subtilis. *Cell* **74**:475–482.

Grundy FJ, and Henkin TM (1998). The S box regulon: A new global transcripton termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. Mol. Microbiol. 30: 737–749

Hach A, Hon T, Zhang L (1999) A new class of repression modules is critical for heme regulation of the yeast transcriptional activator Hap1. *Mol Cell Biol* **19**, 4324-4333.

Hammann, C. & Westhof (2007) E. Searching genomes for ribozymes and riboswitches. *Genome Biology* **8,** 210

Hayashi N, Yoda B, Kikuchi G (1969) Mechanism of allylisopropylacetamide-induced increase of d-aminolevulinic acid synthase in the liver mitochondria IV. Accumulation of the enzyme in the soluble fraction of the rat liver. *Arch. Biochem. Biophys*. **131**:83–91.

Henkin TM (2011) Riboswitch RNAs: using RNA to sense cellular metabolism, *Genes Dev*. **22**: 3383-3390

Ilag LL, Jahn D, Eggertsson G, Soll D (1991) The Escherichia coli hemL gene encodes glutamate 1-semialdehyde aminotransferase. *J Bacteriol* ; **173**(11);3408-13

Ineichen G, Biel AJ (1993) Location of the hemE gene on the physical map of Escherichia coli. *J Bacteriol* **175**(23); 7749-50.

Jacobs JM, Jacobs NJ (1984) Protoporphyrinogen oxidation, an enzymatic step in heme and chlorophyll synthesis: partial characterization of the reaction in plant organelles and comparison with mammalian and bacterial systems. *Arch Biochem Biophys* **229**(1);312-9.

Jacobs NJ, Jacobs JM (1976) Nitrate, fumarate, and oxygen as electron acceptors for a late step in microbial heme synthesis. *Biochim Biophys Acta* **449**(1);1-9. PMID: 788792

Jacobs NJ, Jacobs JM (1977) Evidence for involvement of the electron transport system at a late step of anaerobic microbial heme synthesis. *Biochim Biophys Acta* **459**(1);141-4. PMID: 318855.

Jacobs NJ, Jacobs JM (1978) Quinones as hydrogen carriers for a late step in anaerobic heme biosynthesis in Escherichia coli. *Biochim Biophys Acta* **544**(3):540-6.

Jaffe EK, Ali S, Mitchell LW, Taylor KM, Volin M, Markham GD (1995) Characterization of the role of the stimulatory magnesium of Escherichia coli porphobilinogen synthase. *Biochemistry* **34**(1); 244-51.

Jansen RP (2001) mRNA localization: message on the move. *Nat Rev Mol Cell Biol* **2**:247-256.

Jeter RJ. Escalante-Semerena C, Roof D, Olivera B and Roth J (1987) Synthesis and use of vitamin B12, p. 551–556. In Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, and Umbarger HE (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology. *American Society for Microbiology*, Washington, D.C.

Kaempfer, R (2003) RNA sensors: novel regulators of gene expression. EMBO reports **4**, 1043–1047.

Kannangara CG, Gough SP, Bruyant P, Hoober JK, Kahn A and Wettstein D von (1988) tRNA(Glu) as a cofactor in delta-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis. *Biochem. Sci*. **13**: 139–143.

Kaplan S, Atherly AG, Barrett A (1973) Synthesis of stable RNA in stringent Escherichia coli cells in the absence of charged transfer RNA. *Proc Natl Acad Sci* U S A **70**(3);689-92. PMID: 4577134.

Kredich NM (1987) Biosynthesis of cysteine, p. 419–428. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), Escherichia coli and Salmonella typhimurium: cellular and molecular biology. American Society for Microbiology, Washington, D.C.

Lander ES, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. *Nature*.**409**:860-921.

Landes C, Perona JJ, Brunie S, Rould MA, Zelwer C, Steitz TA, Risler JL (1995) A structure-based multiple sequence alignment of all class I aminoacyl-tRNA synthetases. Biochimie **77**(3);194-203. PMID: 7647112

Lathrop JT, Timko MP (1993) Regulation by heme of mitochondrial protein transport through a conserved amino acid motif. Science 259, 522525.

Layer G, Grage K, Teschner T, Schunemann V, Breckau D, Masoumi A, Jahn M, Heathcote P, Trautwein AX, Jahn D (2005) "Radical S-adenosylmethionine enzyme coproporphyrinogen III oxidase HemN: functional features of the [4Fe-4S] cluster and the two bound S-adenosyl-L-methionines." *J Biol Chem* **280**(32);29038-46

Layer G, Pierik AJ, Trost M, Rigby SE, Leech HK, Grage K, Breckau D, Astner I, Jansch L, Heathcote P, Warren MJ, Heinz DW, Jahn D (2006) The substrate radical of Escherichia coli oxygen-independent coproporphyrinogen III oxidase HemN. *J Biol Chem* **281**(23);15727-34

Layer G, Verfurth K, Mahlitz E, Jahn D (2002). Oxygen-independent coproporphyrinogen-III oxidase HemN from Escherichia coli. J Biol Chem 277(37);34136-42

Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA (2006) Folding of the adenine riboswitch. *Chem Biol*, **13**:857-868.

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

Luer C, Schauer S, Mobius K, Schulze J, Schubert WD, Heinz DW, Jahn D, Moser J (2005) Complex formation between glutamyl-tRNA reductase and glutamate-1-semialdehyde 2,1-aminomutase in Escherichia coli during the initial reactions of porphyrin biosynthesis. *J Biol Chem* **280**(19); 18568-72.

Luo T, Fan T, Liu Y, Rothbart M, Yu J, Zhou S, Grimm B, Luo M (2012) Thioredoxin Redox Regulates ATPase Activity of Magnesium Chelatase CHLI Subunit and Modulates Redox-Mediated Signaling in Tetrapyrrole Biosynthesis and Homeostasis of Reactive Oxygen Species in Pea Plants. Plant Physiol. **159**(1): 118–130

Maines MD(1992) Heme oxygenase: clinical applications and functions. Boca Raton: CRC Press.

Mandal M, Breaker RR (2004) Gene regulation by riboswitches. Nat Rev Mol Cell Biol : **5**:451-463.

Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**:275-279.

Marver HS, Tschudy DP, Perlroth MG, Collins A. d-aminolevulinic acid synthetase (1966) I. Studies in liver homogenates. *J. Biol. Chem*. **241**:2803–2809.

Masuda Y, Tsuchimoto S, Nishimura A, Ohtsubo E (1993) Isolation of temperature-sensitive aminoacyl-tRNA synthetase mutants from an Escherichia coli strain harboring the pemK plasmid. *Mol Gen Genet* **238**(1-2);169-76. PMID: 8479423

Minagawa S, Ogasawara H, Kato A, Yamamoto K, Eguchi Y, Oshima T, Mori H, Ishihama A, Utsumi R (2003) Identification and molecular characterization of the Mg^{2+} stimulon of Escherichia coli. *J Bacteriol* **185**(13);3696-702

Miranda-Rios J Navarro M, and Soberon M (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl. Acad. Sci.* **98**: 9736–9741.

Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111** (5): 747–56. PMID 12464185.

Mitchell LW, Jaffe EK (1993) Porphobilinogen synthase from Escherichia coli is a Zn(II) metalloenzyme stimulated by Mg(II). *Arch Biochem Biophys* **300**(1) 169-77

Miyamoto K, Nishimura K, Masuda T, Tsuji H, Inokuchi H (1992) Accumulation of protoporphyrin IX in light-sensitive mutants of Escherichia coli. FE*BS Let*t **310**(3);246-8. PMID: 1397280

Mogi T, Saiki K, Anraku Y (1994) Biosynthesis and functional role of haem O and haem A. *Mol. Microbiol*. **14**:391–398.

Montange RK, Batey RT (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element Nature. **441** (7097): 1172 -1175.

Moser J, Lorenz S, Hubschwerlen C, Rompf A and Jahn D (1999). Methanopyrus kandleri glutamyl-tRNA reductase. *J. Biol. Chem.* **274**: 30679–30685.

Nahvi, A, Sudarsan N, Evert MS, Zou X, Brown KL, and Breaker RR (2002) Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**: 1043–1049.

Nakahigashi K, Nishimura K, Miyamoto K, Inokuchi H (1991) Photosensitivity of a protoporphyrin-accumulating, light-sensitive mutant (visA) of Escherichia coli K-12. Proc Natl Acad Sci U S A **88**(23);10520-4. PMID: 1835790

Nakajima O, Takahashi S, Harigae H, Furuyama K, Hayashi N, Sassa S, Yamamoto M (1999) Heme defciency in erythroid lineage causes differentiation arrest and cytoplasmic iron overload. *EMBO J* **18**, 6282-6289.

Nakakuki M, Yamauchi K, Hayashi N, Kikuchi G (1980) Purification and some properties of delta-aminolevulinate synthase from the rat liver cytosol fraction and immunochemical identity of the cytosolic enzyme and the mitochondrial enzyme. *J. Biol. Chem*. **255**:1738–1745.

Narita S, Taketani S, Inokuchi H (1999) Oxidation of protoporphyrinogen IX in Escherichia coli is mediated by the aerobic coproporphyrinogen oxidase. *Mol Gen Gene*t **261**(6);1012-20.

Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low Jr KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (1996) Escherichia coli and Salmonella, Cellular and Molecular Biology, Second Edition." *American Society for Microbiology*, Washington, D.C.

Nihei C, Nakayashiki T, Nakamura K, Inokuchi H, Gennis RB, Kojima S, Kita K (2001) Abortive assembly of succinate-ubiquinone reductase (complex II) in a ferrochelatase-deficient mutant of Escherichia coli. *Mol Genet Genomics* **265**(3);394-404. PMID: 11405622.

Nishimura K, Nakayashiki T, Inokuchi H (1993) Cloning and sequencing of the hemE gene encoding uroporphyrinogen III decarboxylase (UPD) from Escherichia coli K-12. *Gene* **133**(1);109-13

Nishimura K, Nakayashiki T, Inokuchi H (1995). Cloning and identification of the hemG gene encoding protoporphyrinogen oxidase (PPO) of Escherichia coli K-12(1995) *DNA Res*;**2**(1);1-8. PMID: 7788523.

Nou X and Kadner RJ (2000) Adenosylcobalamin inhibits ribosome binding to btuB RNA. *Proc. Natl. Acad. Sci*. **97:** 7190–7195.

Nudler, E & Gottesman, ME (2002)Transcription termination and anti-termination in E. coli. *Genes to Cells* **7**, 755–768

Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. Trends in Biochemical Sciences **29**, 11–17 (2004)

Ogawa K, Sun J, Taketani S, 8 other authors (2001) Hememediates derepression of Maf recognition element through direct binding to transcription repressor Bach1. *EMBO J* **20**, 2835-2843.

Olson BR, Zeilstra-Ryalls JH (2008) Regulation of the *Rhodobacter sphaeroides* 2.4.1 *hemA* Gene by PrrA and FnrL. *J Bacteriol.* **190**(20): 6769–6778

Paul A. Castelfranco, Owen TG Jones (1975) Protoheme Turnover and Chlorophyll Synthesis in Greening Barley Tissue .*Plant Physiol*. **55**, 485-490.

Petricek M, Petrickova K, Havlicek L, and Felsberg J (2006) Occurrence of Two 5-Aminolevulinate Biosynthetic Pathways in Streptomyces nodosus subsp. asukaensis Is Linked with the Production of Asukamycin, *J Bacteriol.* **188**(14): 5113–5123.

Qi Z, Hamza I, O'Brian MR (1999) Heme is an effector molecule for iron-dependent degradation of the bacterial iron response regulator (Irr) protein. *Proc Natl Acad Sci SA* **96**, 13056-13061.

Ray SK (2004) Riboswitch: A new mechanism of gene regulation in bacteria. *Current Science,* **87**, No. 9.

Reddy SV, Alcantara G, Roodman D, Boldt DH (1996) Inhibition of tartrate-resistant acid phosphatase gene expression by hemin and protoporphyrin IX. Identifcation of a heminresponsive inhibitor of transcription. *Blood* **88**, 2288-2297.

Rieder R, Lang K, Graber D, Micura R (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem* **8**:896-902.

Rodgers KR (1999) Heme-based sensors in biological systems. *Curr Opin Chem Biol* **3**, 158-167.

Russell RR, Pittard AJ (1971) Mutants of Escherichia coli unable to make protein at 42 degree C. *J Bacteriol* **108**(2);790-8. PMID: 4942764

Samuel I, Beale A Castelfranco, Paul A Castelfranco (1974) The Biosynthesis of a-Aminolevulinic Acid in Higher Plants (Formation Of 14c-6-Aminolevulinic Acid From Labeled Precursors In Greening Plant Tissues') *Plant Physiol*. **53**, 297-303.

Sasaki K, Watanabe M and Tanaka T (2002) Biosynthesis, biotechnological production and applications of 5-aminolevulinic acid. *Appl Microbiol Biotechnol*. **58**(1): 23-29

Sasarman A, Chartrand P, Proschek R, Desrochers M, Tardif D, Lapointe C (1975) Uroporphyrin-accumulating mutant of Escherichia coli K-12. J Bacteriol **124**(3);1205-12

Schmitt MP (1999) Identifcation of a two-component signal transduction system from Corynebacterium diphtheriae that activates gene expression in response to the presence of heme and hemoglobin. *J Bacteriol* **181**, 5330-5340.

Schobert M Jahn D (2002) Regulation of Heme Biosynthesis in Non-Phototrophic Bacteria, *J. Mol. Microbiol. Biotechnol.* **4**(3): 287–294.

Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* **29**(5);1097-106. PMID: 11222759

Spencer P, Jordan PM (1993) Purification and characterization of 5-aminolaevulinic acid dehydratase from Escherichia coli and a study of the reactive thiols at the metal-binding domain. *Biochem J* **290** (Pt 1);279-87.

Sudarsan, N Barrick JE, and Breaker RR (2003a) Metabolitebinding RNA domains are present in the genes of eukaryotes. RNA 9: 644–647.

Troup B, Hungerer C, Jahn D (1995) Cloning and characterization of the Escherichia coli hemN gene encoding the oxygen-independent coproporphyrinogen III oxidase. *J Bacteriol* **177**(11);3326-31

Tucker BJ Breaker RR (2005) Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology* **15**, 342–348

Tyson K, Metheringham R, Griffiths L, Cole J (1997a) Characterisation of Escherichia coli K-12 mutants defective in formate-dependent nitrite reduction: essential roles for hemN and the menFDBCE operon. *Arch Microbiol* **168**(5);403-11. PMID: 9325429

Van der Velden AW, Thomas AA (1999), The role of the 5' untranslated region of an mRNA in translation regulation during development. Int J Biochem Cell Biol,31: 87-106.

Venter JC, Adams MD, Myers EW and 274 more author (2001) The sequence of the human genome. *Science*, **291**:1304-1351.

Vitreschak AG, Rodionov DA, Mironov AA, and Gelfand MS (2003) Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA* **9**: 1084–1097.

Vitreschak AG, Rodionov DA, Mironov AA, and Gelfand MS(2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res*. **30**: 3141–3151.

Vitreschak, AG, Rodionov, DA, Mironov, AA and Gelfand, MS (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet*; **20**, 44–50.

Vogel J and Wagner EG (2007) Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol*. **10**:262–270.

Wachter, A. et al (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cel*l **19**, 3437–3450

Walczak R, Westhof E, Carbon P, Krol A (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs.*RNA*, **2**:367-379.

Wang L, Elliott M, Elliott T (1999) Conditional stability of the HemA protein (glutamyl-tRNA reductase) regulates heme biosynthesis in Salmonella typhimurium. *J Bacteriol* **181**, 1211-1219.

Warnick GR, Burnham B (1971) Regulation of porphyrin biosynthesis: purification and characterization of d-aminolevulinic acid synthase. *J. Biol. Chem*. **246**:6880–6885.

Warren, MJ, Bolt EL, Roessner CA, Scott AI, Spencer JB and Woodcock SC (1994) Gene dissection demonstrates that the Escherichia coli cysG gene encodes a multifunctional protein. *Biochem. J.* **302**:837–844.

Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, et al. (2007) Identification of 22 candidate structured RNAs in bacteria using the Cmfinder comparative genomics pipeline. *Nucleic Acids Res*. **35:** 4809–4819.

Welz R, Breaker RR (2007) Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. *RNA*, **13**:573-582.

Whiting MJ, Elliott WH (1972) Purification and properties of solubilized mitochondrial d-aminolevulinic acid synthetase and comparison with the cytosol enzyme, *J. Biol. Chem.* **247**:6818–6826.

Wickiser JK, Cheah MT, Breaker RR, Crothers DM (2005a) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, **44**:13404-13414.

Wickiser JK, Winkler WC, Breaker RR, Crothers DM(2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell*, **18**:49-60.

Winkler WC, Cohen-Chalamis S, and Breaker R (2002a) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci*. **99**: 15908–15913.

Winkler WC, Nahvi A, and Breaker RR (2002b) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952–956.

Winkler W, Nahvi A, Breaker RR (2002a) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419** (6910): 952–956. PMID 12410317.

Winkler WC, (2005) Metabolic monitoring by bacterial mRNAs. Arch Microbiol, 183:151-159.

Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol*, **59**:487-517

Winkler WC, Cohen-Chalamish S, Breaker RR (2002b) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci USA* **99** (25): 15908–13. PMID 12456892.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR(2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**:281-286.

XU K Delling J, Elliott T (1992) The Genes Required for Heme Synthesis in Salmonella typhimurium Include Those Encoding Alternative Functions for Aerobic and Anaerobic Coproporphyrinogen Oxidation. *Journal of Bacteriology* **174**(12) 3953-3963

Zheng M, Wang X, Templeton LJ, Smulski DR, LaRossa RA, Storz G (2001) DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. *J Bacteriol* **183**(15);4562-70. PMID: 11443091

Annexure

(A) Deduced amino acids sequence of Glutamyl-tRNA reductase

```
        10         20         30         40         50         60
MTLLALGINH KTAPVSLRER VSFSPDKLDQ ALDSLLAQPM VQGGVVLSTC NRTELYLSVE
        70         80         90        100        110        120
EQDNLQEALI RWLCDYHNLN EEDLRKSLYW HQDNDAVSHL MRVASGLDSL VLGEPQILGQ
       130        140        150        160        170        180
VKKAFADSQK GHMKASELER MFQKSFSVAK RVRTETDIGA SAVSVAFAAC TLARQIFESL
       190        200        210        220        230        240
STVTVLLVGA GETIELVARH LREHKVQKMI IANRTRERAQ ILADEVGAEV IALSDIDERL
       250        260        270        280        290        300
READIIISST ASPLPIIGKG MVERALKSRR NQPMLLVDIA VPRDVEPEVG KLANAYLYSV
       310        320        330        340        350        360
DDLQSIISHN LAQRKAAAVE AETIVAQETS EFMAWLRAQS ASETIREYRS QAEQVRDELT
       370        380        390        400        410
AKALAALEQG GDAQAIMQDL AWKLTNRLIH APTKSLQQAA RDGDNERLNI LRDSLGLE
```

(B) Sequence to be analysed

To obtaina analyse structure following sequence was taken as input sequence in mfold web server.

gltX UTR

GG TAGAAACCCG AACTACATTG AGGAATCAGG CGGGAGTGAT AGAATATCGC CCACTTAATT TTTCCAGGAT TTGCCGGTTG TCGGCATCTT TCTAAACGTA AGGCCATTTC

hwmA UTR

G TCTTAATTGC CAGAATCTAA CGGCTTTCGG CAATTACTCC AAAAGGGGGC GCTCTCTTTT ATTGATCTTA CGCATCCTGT ATGATGCAAG CAGACTAACC CTATCAACGT TGGTATTATT TCCCGCAGAC

hemB UTR

TCCATTCTCT TCTTGTCAAC ATCGCGACAA CTTTCGTAAA ACATCCCTAC CCTGCTTCAG
GTATACTATG CCCCTCGATT CCACAAACAT CAGGCAGACC

hemE UTR(nudC)

AAGCATGATT GACTCCGCAA GTTTGTATTC AAAAACTGCT CAGTGAGAAA TGTAAAAACC
ATGTTAAACA TGCCAGTGAT GCAAAGGTAG TGCAAGAGCT

hemF

T TAGGCACGGC GGCGTTTCGT CAGAAAATCG CCACAGCGAT TGCTGAAGGC GTGATCAGTT
ATTTCCACTG GTTCGACAAC CAGAAAGCAC ATTCGAAAAA GCGATAAGTT

hemG

AAGAAATGCC GATCTGATCA CGCTCGGCAG AAATCACGCT CTGGATGAAC GATGTGCTAA
GATGCGGAGA CTTAAGGTCA AAAAAACAGA AGGGTAAAAA

hemH

GCGCTATTCC TTTTTCTGAT TTGACCTCTC ACAGCAATTA GTTCTTCTTC CTCACTTTTC
CGCTACAATT ATCAACAAGT TGAATCGATA AGAGGCGGTA

hemL

TGCACCAGTA CAAGCAGCCT GATGTTTGAC GAGTATTTAA CTTGTTATGA ATAACATAGA
ATAGCAGCCA TTCACTTTTT CTACGAATCA GGAACCCTCC

hemN

GT TGTCGCGGCG ACAAAAGTGG CTGCGCAACC GGCTGATGAC CCGAAACGGG CATCCGCCGG
TACGCCGTAG CCGCCAGAGA CGCCATCGGA AGGAGTGAGC

Appendix 1.1 : Primers used for *hem*A gene amplification

S.No	Specifications	Sequence	
1	Forward primer**(SSMPOA6X1F01)**	CTAC**GAATTC**TCCAAA **CATatg** acc ***AAA AAA*** ct c tta gca ctc gg	
	Reverse primer **(SSMPOA6X1R01)**	AA ACC ACT **GAATTC** CTG ATG **TCTAga** tgt act gct act cc	
		Forward Primer	**Reverse Primer**

2	T_m	Basic (whole/anealing)	63.6^0 C/56.3^0 C	63.2^0 C/45.2^0 C
3	G +C content(whole/anealing)		40%/55%	42.5%/50 %
4	No. of base pairs(whole/anealing)		45/20	40/16
5	Secondary structure		-	-
	T_m		19.4^oC	28.8^0 C
	dG(Kcal/mol)		0.35	-0.26
6	Self dimer: with restriction site		dG= = -8.51Kcal/ mol	dG = -8.51 Kcal/ mol
	without restriction site		dG =	dG =
7	Hetero dimer		dG = = -8.51, <9.00 Kcal/ mol (in negative figure)	
8	Approximate amplicon size		1315 bp	

Appendix 1.2 PCR reaction mixture: *hem*A gene

Components	Quantity	Final concentration
Template: Genomic DNA of *E. coli* HME 5	1.5 µl	25 ng
2mM dNTPs each	3.2 µl	250 µM dNTPs of each
10µM Forward primer (SSMPOA6X1F01)	2.5 µl	1 µM
10µM Reverse primer (SSMPOA6X1R01)	2.5 µl	1 µM
5 X Reaction Buffer	5 µl	1X
0.5U/ µl DNA polymerase	1 µl	0.5U
Nuclease free water	Up to 25 µl	

Appendix 1.3 PCR conditions: *hem*A gene amplification

Step	Cycle	Temp	Time
Initial Denaturation	1 Cycle	95^0C	5 Min
Denaturation	5 Cycles	95^0C	30 sec
Annealing		52^0C	30 sec
Extension (1 min/kb)		72^0C	1 min 38 sec
Denaturation	25 cycles	95^0C	30 sec
Annealing		59^0C	30 sec
Extension		72^0C	1 min 38 sec
Final Extension	1 Cycle	72^0C	5 Min

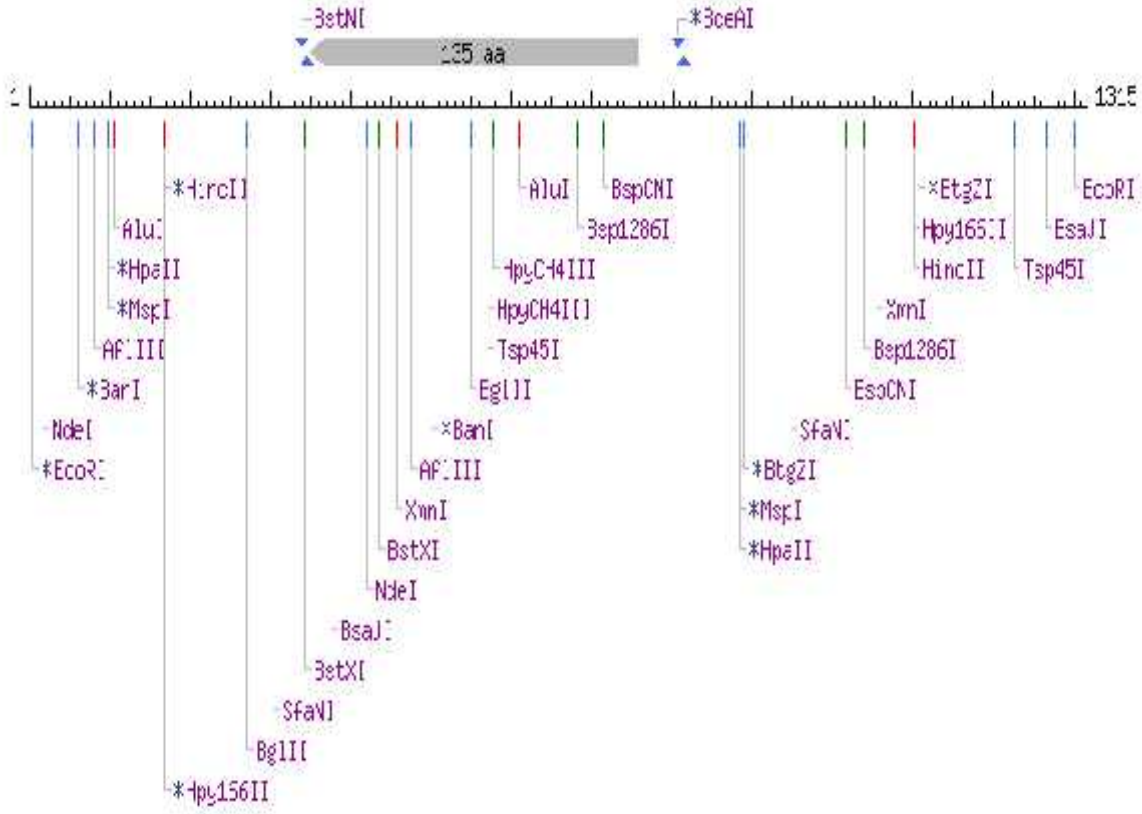
Hold	1 Cycle	4^0C	∞

Appendix 2: 20 Amino acids, their single-letter data-base codes, and their corresponding DNA codons

Amino Acid	SLC	DNA codons
Isoleucine	I	ATT, ATC, ATA
Leucine	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	V	GTT, GTC, GTA, GTG
Phenylalanine	F	TTT, TTC
Methionine	M	ATG
Cysteine	C	TGT, TGC
Alanine	A	GCT, GCC, GCA, GCG
Glycine	G	GGT, GGC, GGA, GGG
Proline	P	CCT, CCC, CCA, CCG
Threonine	T	ACT, ACC, ACA, ACG
Serine	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Y	TAT, TAC
Tryptophan	W	TGG
Glutamine	Q	CAA, CAG
Asparagine	N	AAT, AAC
Histidine	H	CAT, CAC
Glutamic acid	E	GAA, GAG
Aspartic acid	D	GAT, GAC
Lysine	K	AAA, AAG
Arginine	R	CGT, CGC, CGA, CGG, AGA, AGG
Stop codons	Stop	TAA, TAG, TGA

APPENDIX 3: Analysis of Restriction site present in *hem*A gene sequence by using NEBcutter V2.0

(http://tools.neb.com/NEBcutter2/cutshow.php?name=35a4abff-)



Appendix 4: Preparation of Stock solutions, Buffers and Reagents commonly used in Molecular Biology

EDTA (0.5 M, pH 8.0)

18.61 g of disodium EDTA.$2H_20$ was added to 80 ml of MilliQ water, stirred vigorously on magnetic stirrer and pH was adjusted to 8.0 with NaOH. The final volume of the solution was made upto 100 ml and sterilized by autoclaving and stored at 4^0C.

Tris-Cl (1 M, pH 7.5 and 8.0)

Tris-Cl buffer was prepared by adding 12.11 g of Tris base in 80 ml of MilliQ water and pH was adjusted to 7.5 or 8.0 by adding concentrated HCl. Then final volume was maintained to 100 ml. The solution was sterilized by autoclaving and stored at 4^0C.

Tris EDTA (TE) buffers

pH 7.5

10 mM Tris-Cl (pH 7.5)

1 mM EDTA (pH 8.0)

pH 8.0

10 mM Tris-Cl (pH 8.0)

1 mM EDTA (pH 8.0)

Potassium Phosphate buffer (0.1 M, pH 7.0)

1 M K_2HPO_4	61.5 ml
1 M KH_2PO_4	38.5 ml
MilliQ water	upto 1000 ml

Preparation of solutions for chromosomal DNA isolation

TE1 solution

50 mM Tris-Cl (pH 7.5)

50 mM EDTA (pH 8.0)

STEP solution

50 mM Tris-Cl (pH 7.5)

0.2 mM EDTA (pH 8.0)

0.5% SDS

1 mg/ml Proteinase K

Lysozyme (1 mg/ml) (Freshly prepared)

1 mg of lysozyme dissolved in 10mM Tris-Cl (pH 8.0) with final volume of 1 ml. This solution was prepared immediately before use.

Proteinase K (20 mg/ml)

The lyophilized proteinase powder dissolved at a concentration of 20 mg/ml in sterilized 50 mM Tris (pH 8.0), 1.5 mM calcium acetate. The solution was stored at -20^{0}C.

Sodium Acetate (3 M, pH 5.2)

40.83 g of sodium acetate.3$H_2$0 was dissolved in 80 ml of MilliQ water and pH was adjusted to 5.2 with glacial acetic acid and final volume was made upto 100 ml. The solution was sterilized by autoclaving and stored at 4^{0}C.

Preparation of solutions for Plasmid DNA extraction by Alkanline-lysis method

Alkaline Lysis Solution I

50 mM glucose

25 mM Tris-Cl (pH 8.0)

10 mM EDTA (pH 8.0)

The solution was sterilized by autoclaving and stored at 4^0C.

Alkaline Lysis Solution II (Freshly prepared)

0.2 N NaOH

1% (w/v) SDS

Alkaline Lysis Solution III

5 M Potassium acetate	60.0 ml
Glacial acetic acid	11.5 ml
H20	28.5 ml

STE

10 mM Tris-Cl (pH 8.0)

0.1 M NaCl

1 mM EDTA (pH 8.0)

Preparation of solutions for Transformation

RF1

30 mM KOAc (Potassium acetate)

100 mM RbCl

10 mM $CaCl_2$

50 mM $MnCl_2$

15% Glycerol

pH of the solution was adjusted to 5.8 with acetic acid and filter sterilized (0.22 µm) and stored in dark bottle at 4^0C.

RF2

10 mM MOPS

75 mM $CaCl_2$

10 mM RbCl

15% Glycerol

pH of the solution was adjusted to 6.5 with KOH and and filter sterilized (0.22 µm) and stored in dark bottle at 4^0C.

Preparation of solutions for Agarose gel electrophoresis

10 X TAE Buffer

Tris base	4.84 g
Glacial acetic acid	1.142 ml
0.5 M EDTA (pH 8.0)	2 ml
MilliQ water	upto 100 ml

The working solution (0.5 X) of TAE buffer was prepared by diluting 17.5 ml of 10 X TAE stock solutions with MilliQ water to make the final volume upto 350 ml.

10X Gel loading buffer

Bromophenol blue	10.5 mg
0.5 M EDTA (pH 8.0)	2 ml
98% Glycerol	2.5 ml
MilliQ water	upto 5 ml

Ethidium Bromide (10 mg/ml)

100 mg of ethidium bromide was weighed and dissolved in 10 ml of MilliQ water. The solution was protected from sun light by wrapping with aluminum foil and stored at room temperature.

Preparation of Media and Additives

IPTG (1 M)

238 mg of IPTG powder was dissolved in 1 ml of MilliQ water, filter sterilized (0.22 µm) and stored at -20^{0}C.

5-Aminolevulinic acid (1 M)

167.59 mg of 5- Aminolevulinic acid was dissolved in 1 ml of MilliQ water, filter sterilized (0.22 µm) and stored at -20^{0}C.

Indole (5 M)

585.75 mg of indole was dissolved in 1 ml of DMSO and stored at 4^{0}C.

Thiamine (1 M)

337.27 mg of thiamine was dissolved in 1 ml of MilliQ water, filter sterilized (0.22 µm) and stored at 4^{0}C.

Trace elements

$FeCl_3$. $6H_20$	27.0 g
$ZnCl_2$. $4H_20$	2.0 g

$CoCl_2$. 6 H_20	2.0 g
Na_2MoO_4.2 H_20	2.0 g
$CaCl_2$. $2H_20$	1.0 g
$CuCl_2$	1.0 g
H_3B0_3	0.5 g
30% HCl	100 ml

The mixture was dissolved in MilliQ water and final volume was adjusted to 1000 ml.

Appendix 5: PCR amplified *hem*A gene sequence (GenBank: NC_000913.2)

```
0001 ctac***gaattc*** tccaaa***cata tg***accaaaaa actcttagca ctcggtatca accataaaac
0061 ggcacctgta tcgctgcgag aacgtgtatc gttttcgccg gataagctcg atcaggcgct
0121 ttgacagcct gcttgcgcag ccgatggtgc agggcggcgt ggtgctgtcg acgtgcaacc
0181 gcacggaact ttatcttagc gttgaagagc aggacaacct gcaagaggcg ttaatccgct
0241 ggctttgcga ttatcacaat cttaatgaag aagatctgcg taaaagcctc tactggcatc
0301 aggataacga cgcggttagc catttaatgc gtgttgccag cggcctggat tcactggttc
0361 tgggggagcc gcagatcctc ggtcaggtta aaaaagcgtt tgccgattcg caaaaaggtc
0421 atatgaaggc cagcgaactg gaacgcatgt tccagaaatc tttctctgtc gcgaaacgcg
0481 ttcgcactga aacagatatc ggtgccagcg ctgtgtctgt cgcttttgcg gcttgtacgc
0541 tggcgcggca gatctttgaa tcgctctcta cggtcacagt gttgctggta ggcgcgggcg
0601 aaactatcga gctggtggcg cgtcatctgc gcgaacacaa agtacagaag atgattatcg
0661 ccaaccgcac tcgcgaacgt gcccaaattc tggcagatga agtcggcgcg gaagtgattg
0721 ccctgagtga tatcgacgaa cgtctgcgcg aagccgatat catcatcagt tccaccgcca
0781 gcccgttacc gattatcggg aaaggcatgg tggagcgcgc attaaaaagc cgtcgcaacc
0841 aaccaatgct gttggtggat attgccgttc cgcgcgatgt tgagccggaa gttggcaaac
```

0901 tggcgaatgc ttatctttat agcgttgatg atctgcaaag catcatttcg cacaacctgg

0961 cgcagcgtaa agccgcagcg gttgaggcgg aaactattgt cgctcaggaa accagcgaat

1021 ttatggcgtg gctgcgagca caaagcgcca gcgaaaccat tcgcgagtat cgcagccagg

1081 cagagcaagt tcgcgatgag ttaaccgcca aagcgttagc ggcccttgag cagggcggcg

1141 acgcgcaagc cattatgcag gatctggcat ggaaactgac taaccgcttg atccatgcgc

1201 caacgaaatc acttcaacag gccgcccgtg acggggataa cgaacgcctg atattctgcg

1261 cgacagcctc gggctggagt agcagtaca***t ctaga***catca g***gaattc***agt ggttt

Appendix 6: pUC19 plasmid information from NEB

pUC19

GenBank Accession #: L09137

Feature	Coordinates	Source
lacZα	469–146	–
origin	1452–867	pMB1 (mutant)
bla (Ap^R)	2486–1626	*Tn3*

ori – origin of replication
Ap – ampicillin

There are no restriction sites for the following enzymes: AarI(x), AflII, AflIII, AgeI, AlwI, ApaI, AscI, AsiSI, AvrII, BaeI, BbsI, BbvCI, BclI, BglII, BlpI, BmgBI, BmtI, BpuY08, BsaAI, BsaBI, BsaHI, BsgI, BsiWI, BsmFI, BsmI, BspDI, BspEI, BsrGI, BssHII, BstBI, BstEII, BstXI, BtrZ17I, Bsu36I, StgI, StgZI, ClaI, CspCI, DraIII, EagI, EcoNI, EcoRV, FseI, FspAI(x), HpaI, I-CeuI, I-SceI, MfeI, MluI, MscI, NaeI, NcoI, NgoMIV, NheI, NotI, NruI, NsiI, PI-PspI, PI-SceI, PacI, PaeR7I, PflFI, PflMI, PmeI, PmlI, PpuMI, PshAI, PsiI, PspGMI, PspXI, RsrII, SacII, SanDI(x), SexAI, SfiI, SgrAI, SnaBI, SpeI, SrfI(x), StuI, StyI, SwaI, TliI, Tth111I, XcmI, XhoI

(x) = enzyme not available from NEB

pUC19 is a small, high-copy number *E. coli* plasmid cloning vector containing portions of pBR322 and M13mp19 (1). It contains the pMB1 origin of replication from pBR322, but it lacks the *rop* gene and carries a point mutation in the RNAII transcript (G 2975 in pBR322 to A 1308 in pUC19, 2). These changes together result in a temperature-dependent copy number of about 75 per cell at 37°C and >200 per cell at 42°C (2, 3). The multiple cloning site (MCS) is in frame with the *lacZα* gene, allowing screening for insertions using α-complementation.

pUC18 is identical to pUC19 except that the MCS region (nt 397–454) is inverted.

pNEB193 is also identical to pUC19 except for the addition of several restriction endonuclease sites to the MCS. Its total length is 2713 bp.

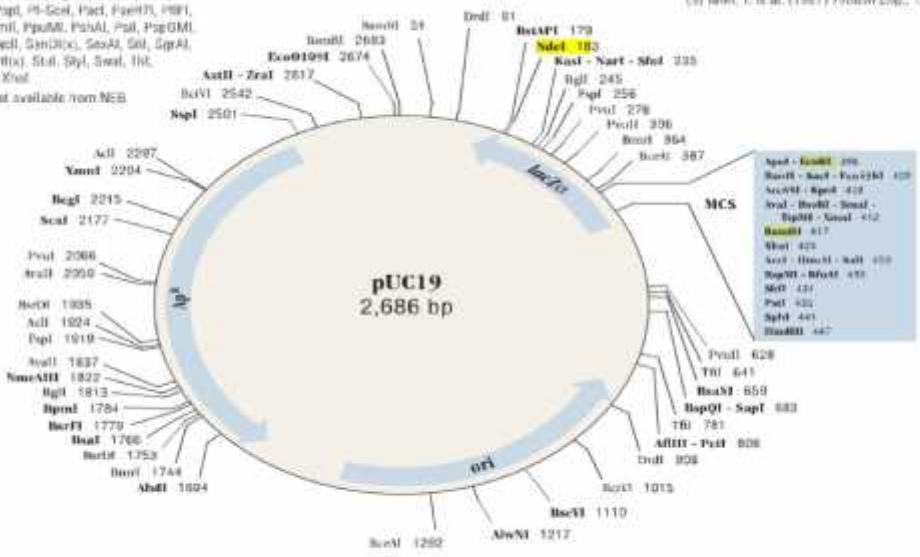
Enzymes with unique restriction sites are shown in **bold** type, and enzymes with two restriction sites are shown in regular type. Location of sites of all NEB restriction enzymes can be found on the NEB web site (choose Technical Reference > DNA Sequences and Maps). Restriction site coordinates refer to the position of the 5′-most base on the top strand in each recognition sequence.

Open reading frame (ORF) coordinates are in the form "translational start – translational stop"; numbers refer to positions on the top (clockwise) strand, regardless of the direction of transcription and include the start and stop codons.

Origin of replication coordinates include the region from the –35 promoter sequence of the RNAII transcript to the RNA/DNA switch point. *bla* (Ap^R) gene coordinates include the signal sequence.

References

(1) Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, 33, 103–119.
(2) Lin-Chao, S., Chen, W.-T. and Wong, T.-T. (1992) *Mol. Microbiol.*, 6, 3385–3393.
(3) Miki, T. et al. (1987) *Protein Eng.*, 1, 327–332.



pUC19 MCS

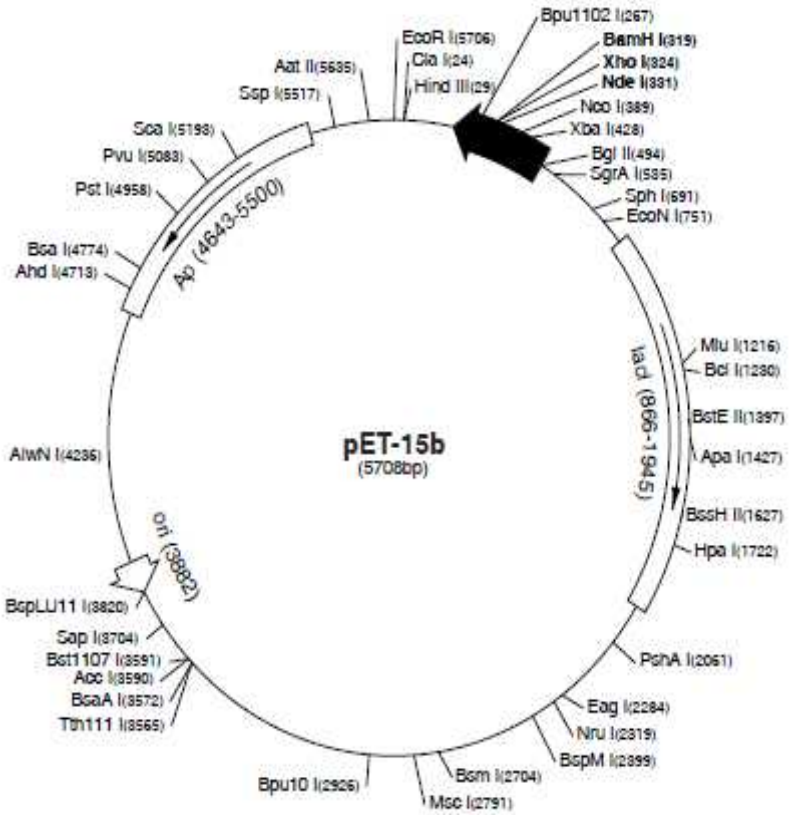
```
        SacI      SmaI      XbaI      SbfI
   EcoRI     KpnI      BamHI     SalI  PstI  SphI  HindIII
agtgAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGcgtaatcatggtcat
      400       410       420       430       440       450       460
...S  N  S  S  P  V  R  P  D  E  L  T  S  R  C  A  H  L  S  P  T  I  M  T  M
                                                                    lacZα translational start
```

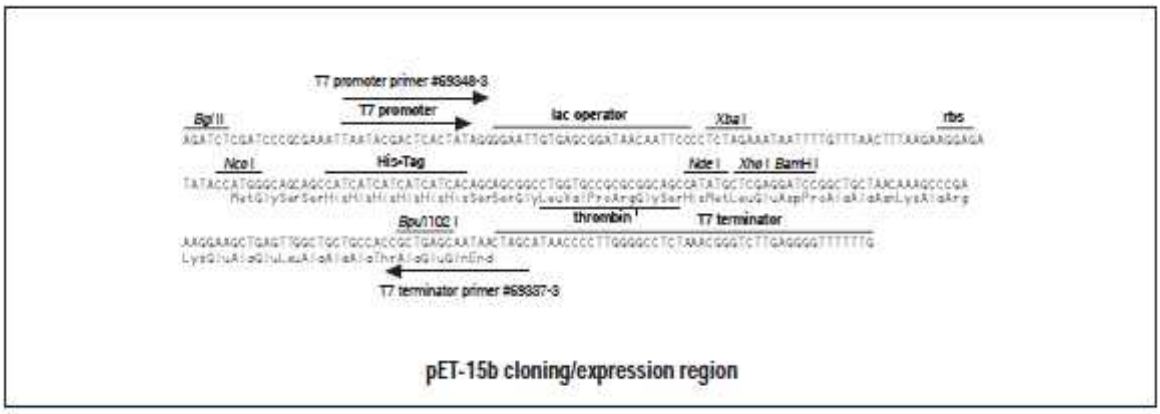
pNEB193 MCS

```
        SacI      SmaI   BssHII      PacI                    SbfI
   EcoRI     KpnI      AscI    BamHI        XbaI  SalI   PmeI    PstI  SphI  HindIII
agtgAATTCGAGCTCGGTACCCGGGGCGCGCCGGATCCTTAATTAAGTCTAGAGTCGACTGTTTAAACCTGCAGGCATGCAAGCTTGGcgtaatcatggtcat
      400       410       420       430       440       450       460       470       480       490
...S  N  S  S  P  V  R  P  R  A  P  D  K  I  L  D  L  T  S  D  K  F  R  C  A  H  L  S  P  T  I  M  T  M
                                                                                    lacZα translational start
```

Appendix 7 : pET15B Map Detail

pET-15b sequence landmarks	
T7 promoter	453-469
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500





pET-15b Restriction Sites

TB045 12/98

Enzyme	# Sites	Locations				
AatII	1	5635				
Acc I	1	3590				
AccIII	1	983	1711	2042	3329	3470
		3772	5012			
Aci I	91					
AflII	2	1216	3620			
AlwI	24					
AlwN I	16					
AlwNI	8	716	1200	2523	2814	3688
		4136	5299	5384		
ApaLI	4	1196	3634	4134	5380	
ApaOI	1	4236				
ApaI	1	1427				
ApaBI	2	800	2397			
ApoI	2	1491	5706			
AscI	2	324	2770			
AseI	9	1758	2144	2232	2481	2784
		2826	3106	4651	5013	
BamH I	1	313				
BanI	13					
BanII	3	600	614	1427		
BbsI	5	1362	1701	2015	2936	5681
BbvI	31					
BccI	16					
BceAI	7	208	2030	2200	3911	4209
		4450	5318			
BcefI	5	735	1076	1703	2512	4322
BcgI	8	1508	1542	2042	2076	3357
3437	5218	5252				
BciI	1	1230				
BfaI	6	257	429	2834	4315	4568
		4903				
BglI	3	2288	2514	4833		
BglII	1	494				
BmgI	1	1405				
BpmI	6	1054	1543	2171	2731	3347
		4782				
Bpu10I	1	2026				
Bpu1102I	1	267				
BsaI	1	4774				
BsaAI	1	3572				
BsaBI	3	483	499	3017		
BsaHI	8	539	560	674	1133	1856
		2557	5250	5632		
BsaJI	11					
BsaWI	7	189	1535	2038	3009	4026
		4173	5004			
BsaXI	1	1875				
BsbI	2	3536	5256			
BscCI	13					
BsgI	3	1067	1767	2980		
BsiI	3	3993	5371	5684		
BsiEI	6	2007	2261	2736	4160	5083
		5232				
BsiI	22					
BsmI	1	2704				
BsmAI	7	913	1318	1444	1831	3461
		4774	5550			
BsmBI	2	1831	3461			
BsmFI	4	677	2218	2443	3091	
BsoFI	57					
Bsp24I	12					
Bsp1286I	11					
BspEI	2	189	3009			
BspGI	3	2404	2461	3346		
BspLU11I	1	3820				
BspMI	1	2388				
BsrI	25					
BsrBI	3	449	3753	5554		
BsrDI	4	1263	1629	4774	4948	
BstFI	8	160	526	535	902	2714
		2774	2828	4793		

Enzyme	# Sites	Locations				
BstHI	1	1627				
BstYI	1	3591				
BstEII	1	1397				
BstXI	3	1018	1147	1270		
BstYI	11					
CacSI	41					
CjeI	25					
CjePI	26					
ClaI	1	24				
CviJI	97					
CviRI	26					
DdeI	11					
DpnI	28					
DraI	3	4579	4668	5290		
DrdI	2	3513	3928			
DrdII	1	839				
DsaI	3	389	653	2702		
EaeI	6	524	656	1890	2284	2789
		5101				
EagI	1	2284				
EarI/11051	1	4713				
EarI	3	834	3704	5508		
EciI	5	983	2740	3694	4040	4868
Eco47II	3	621	2122	3014		
Eco57I	2	4368	5380			
EcoNI	1	751				
Eco0109I	5	240	646	2764	2826	5689
EcoRI	1	5706				
EcoRII	11					
EcoRV	2	197	1666			
FauI	18					
FokI	14					
FspI	3	2703	2801	4695		
GdiII	5	524	656	1890	2284	5101
HaeI	8	944	2265	2337	2394	2791
		3835	3946	4256		
HaeII	13					
HaeIII	28					
HgaI	15					
HgiEII	2	874	4406			
HhaI	45					
HinfI	5	16	1115	2486	4712	4786
HincII	2	1722	5254			
HindIII	1	29				
HinfI	14					
HpaI	1	1722				
HphI	17					
MaeI	12					
MaeII	18					
MboII	15					
MluI	1	1216				
MnlI	2	4035	4219			
MnlI	34					
MscI	1	2701				
MseI	24					
MslI	10	1268	1556	1586	2376	2807
		3002	3393	4965	5124	5483
MspI	35					
MspA1I	11					
MwoI	44					
NaeI	5	539	562	674	1856	2551
NciI	14					
NcoI	1	389				
NdeI	1	331				
NgoAIV	4	526	2714	2274	2628	
NlaIII	31					
NlaIV	29					
NruI	1	2219				
NspI	4	691	3165	3457	3824	
PflT1108I	2	2103	4731			
PflMI	3	788	2666	2715		

Enzyme	# Sites	Locations				
PleI	7	477	765	852	1648	3714
		4199	4752			
PshAI	1	2067				
PspGI	2	2704	2826			
PspT406I	5	878	2246	3145	4939	5312
PstI	1	4868				
PvuI	1	5093				
PvuII	3	1916	1989	3411		
RcaI	4	674	4640	5548	5653	
RsaI	4	165	1363	3626	5193	
SapI	1	3704				
Sau96I	21					
Sau3AI	29					
ScaI	1	5193				
ScrFI	25					
SfaNI	24					
SfcI	5	138	462	4088	4276	4954
SgrAI	1	535				
SghI	1	691				
SgsI	1	5517				
SlyI	3	244	389	2714		
TaqI	13					
TaqII	8	1124	1342	3015	3122	5067
SZ46	5390	5416				
TfiI	7	1886	2197	2357	2649	2830
		3374	3755			
ThaI	41					
TsaI	31					
Tsp45I	9	124	1387	2225	2492	3259
		3472	3567	4669	5180	
Tsp509I	16					
Tth111I	1	3565				
Tth111II	7	1055	1746	3281	4410	4417
		4449	5705			
UbaU	26					
VspI	4	477	1901	1960	4885	
XbaI	1	426				
XcmI	3	1077	1588	1626		
XhoI	1	324				
XmnI	2	3378	5312			

Enzymes that do not cut pET-15b

AflII	AgeI	AscI	AvrII	BalI
BsaWI	BstZI	Bsu36I	DraIII	FseI
KpnI	MluNI	NheI	NotI	NsiI
NspV	PacI	PmeI	PmlI	RsrII
SacI	SacII	SalI	SbfI	SexAI
SfiI	SgfI	SmaI	SnaBI	SpeI
SrfI	Sse8387I	SstI	SunI	SwaI