The Rhodobacter capsulatus hemA Region: Nucleotide Sequence and Functional Analysis.

John Joseph Eckert
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/5228

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
The *Rhodobacter capsulatus* hemA region: Nucleotide sequence and functional analysis

Eckert, John Joseph, Ph.D.

The Louisiana State University and Agricultural and Mechanical Col., 1991
THE Rhodobacter capsulatus hemA REGION:
NUCLEOTIDE SEQUENCE AND
FUNCTIONAL ANALYSIS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Microbiology

by

John Joseph Eckert
B.A., Skidmore College, 1985
December, 1991
Dedication

This work is gratefully dedicated to

Drs. Bob Mahoney and Bernie Possidente

for inspiring me to pursue my dream.
Acknowledgements

I find it hard to express my deep appreciation for all of the emotional support over the past three years provided by my greatest champion, my fiancé, Michelle. Your never-ending love and commitment guided me through some of the most difficult times of my life. Thank you.

Deepest heartfelt thanks to my parents for their love and support, both emotional and financial, without which I would not have made it through. Many thanks to my sister and brother-in-law for being there when I needed them.

To my dog, Cheryl, calm down! Your energy is infectious, and your presence cheered me up on the darkest of days.

Many thanks, also, to my Aunt Mary and Uncle Harold for providing me with a touch of home just down the street. I am only sorry that I didn’t get down to see you more often. My sincere gratitude to my cousin Mickey for dropping me off in Baton Rouge and leaving me; your assistance didn’t go unnoticed or unappreciated.

To my graduate advisor, Alan Biel, you have my deepest thanks and unending respect. Without you riding my butt, I would still be here in 1999! To Dr. Susan
Biel, thank you for all of your assistance in the lab when I first arrived. Your knowledge, candor and courage to venture out in the world of the self-employed have earned you my respect.

I would like to thank my committee members for putting up with me for so long: Drs. H.D. Braymer, J.M. Larkin, J.R. Battista, D.S. Shih and T.D. Dreesen; I may soon be out of your hair.

Several of my graduate student colleagues hold a special place in my heart: Maureen Wright, Matt Dobbins, Anne Johnson, their respective spouses, Greg, Danette and Greg (#2), and Jana Rajan, my fellow December graduate. It may sound a little "Bartles and Jaymesish," but thanks for your support!

To my tennis partner, Jerry Elliott, I owe a great deal of thanks. Over these past few years, we had some good talks, and tennis matches. I’ll miss both; but, wherever I end up, there will be lighted tennis courts, so you should come visit as often as you can.

I would like to express my deep appreciation to Laynette Spring for all of her assistance with DNA sequencing when I first started. I am glad that I did not have to learn alone! I would also like to thank Cindy Henk for help with slides and coffee!
My gratitude also goes out to "The Family" for making me feel like a "Bogaluser" from day one. I am indeed fortunate to be able to call Boonnark, Rubby, Mack and Frances "family". I would like to pay my respects to the late Mrs. Carnegie (Mamaw), you will never know how much it means to Michelle and me to have your blessing. We love and miss you very much!

I would like to take this opportunity to thank my roommate, Doug. For four years you have put up with a neurotic roommate; I don't know how you survived.

My deepest thanks to Dr. and Mrs. Dr. Soc for acting as my surrogate parents. Thank you for checking out Dr. Teague before he stitched my face! You are the greatest hosts I have ever met, I will truly miss you both.

JJE
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>Abstract</td>
<td>xii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>23</td>
</tr>
<tr>
<td>I  Bacterial Strains, Phage and Plasmids</td>
<td>23</td>
</tr>
<tr>
<td>II Media</td>
<td>23</td>
</tr>
<tr>
<td>III Plasmid and M13 Replicative Form Preparation</td>
<td>36</td>
</tr>
<tr>
<td>a. Minipreparation</td>
<td>36</td>
</tr>
<tr>
<td>b. Large Scale Preparation</td>
<td>38</td>
</tr>
<tr>
<td>IV Restriction Enzyme Digestion...</td>
<td>41</td>
</tr>
<tr>
<td>V  Agarose Gel Electrophoresis...</td>
<td>41</td>
</tr>
<tr>
<td>VI Ligation</td>
<td>42</td>
</tr>
</tbody>
</table>

vi
VII Transformation of Plasmid and M13 Replicative Form DNA........ 42

VIII Mating............................ 44

IX Sonication......................... 45

X Protein Determination.......... 45

XI Porphobilinogen Synthase Assay. 46

XII Porphobilinogen Deaminase Assay......................... 46

XIII DNA Sequencing..................... 47
   a. Template
      Preparation....... 48
   b. Primer Annealing.. 49
   c. Labeling Reaction. 49
   d. Termination
      Reaction......... 50
   e. Determination of
      Sequence Close to
      the Primer....... 50
   f. Determination of
      Sequence further
      from the Primer... 51
   g. Releaving G-C
      Compression....... 51
XIV Polyacrylamide Gel

Electrophoresis................. 52

   a. Preparation of
      the Plates............. 52

   b. Preparation of
      the Gel Mix.......... 53

   c. Pouring the
      Sequencing Gel..... 53

   d. Running the
      Sequencing Gel..... 54

   e. Fixing, Washing,
      Drying and Autoradiography of
      The Sequencing
      Gel................. 55

XV Computer Analysis of DNA

Sequence Data.................... 56

Results................................................. 58

I DNA Sequence of the R.
    capsulatus hemA Gene............ 58

II Confirmation that the 1,206
    Base Pair Open Reading Frame
    is hemA.................. 77

viii
List of Tables

1. Bacterial Strains used in this Study...... 24
2. Phage and Plasmids used in this Study..... 26
3. Primers used for Generation of *hemA*
   Sequence........................................ 65
4. Primers used for Generation of the Sequence
   of the 1.9 kb *EcoRI* Fragment............... 83
List of Figures

1. Tetrapyrrole Biosynthetic Pathway .......... 3

2. Partial Restriction Map of the AJB529 hemA Region, pCAP1 and pCAP17 ................. 59

3. Graphical Representation of the Strategy Employed to obtain DNA Sequence Data of the R. capsulatus hemA Region .............. 60

4. Subclones of the R. capsulatus hemA Region used for DNA Sequencing ..................... 62

5. Nucleotide Sequence of the R. capsulatus hemA Region .................................. 69

6. Codon Preference Plot of the Region Known to Contain the hemA Gene ...................... 78

7. Codon Preference Plot for the Region Encompassing orf-2 ................................ 84

8. Codon Preference Plot for the Entire hemA Region ........................................... 92

9. Alignment of δ-Aminolevulinate Synthases and Derived Consensus δ-Aminolevulinate Synthase Sequence ........................................ 101
Abstract

Most of the effort to elucidate the mechanism of oxygen-mediated regulation of tetrapyrrole biosynthesis in *Rhodobacter capsulatus* has focused on the first step in the biosynthetic pathway, the condensation of glycine with succinyl-CoA to form δ-aminolevulinic acid. This reaction, in *R. capsulatus*, is catalyzed by δ-aminolevulinate synthase, the product of the hemA gene. In order to facilitate further study of the hemA gene and its product, its nucleotide sequence was determined.

The open reading frame encoding δ-aminolevulinate synthase was determined to be 1,206 base pairs, encoding a 50,491 Da polypeptide of 401 amino acids. The DNA sequence was compared with other hemA genes and found to be between 60% and 70% identical. Likewise, the predicted amino acid sequence of this open reading frame was found to be 70% similar to other δ-aminolevulinate synthases. δ-Aminolevulinate synthase is a pyridoxal phosphate-requiring enzyme; thus, by comparing its amino acid sequence with that of other pyridoxal phosphate-requiring enzymes, a putative pyridoxal phosphate binding site has been identified.
When attempts were made to isolate a hemA strain of \textit{R. capsulatus} by Tn5 mutagenesis, no complete block in the hemA gene was obtained. Sequence data from the only aminolevulinate-requiring mutant revealed that the Tn5 inserted 20 bases into the coding region for \( \delta \)-aminolevulinate synthase. Transcription and translation initiating in Tn5, and continuing into the \textit{R. capsulatus} DNA would account for the low level \( \delta \)-aminolevulinate synthase activity observed in the mutant. This transcription could be necessary if hemA is part of an operon, and a complete block of transcription from the hemA promoter would have a polar effect on a downstream gene. Sequence data collected reveals that there are two open reading frames downstream of hemA. Both are transcribed in the same direction as hemA. The first begins 84 base pairs past the termination codon for hemA and extends for 1,239 base pairs. The product of this open reading frame has not been identified. Beginning 55 base pairs past the termination codon for orf-2 there is another unidentified open reading frame. The sequences of orf-2 and orf-3 were compared with all sequences in Genbank; no significant homology exists with any other gene.
Introduction

The purple, non-sulfur photosynthetic bacteria are among the most versatile of organisms, capable of sustaining growth under a wide variety of environmental conditions. *Rhodobacter capsulatus* will grow chemoheterotrophically in the presence of molecular oxygen and a small organic compound. Alternatively, in the absence of oxygen, but with a small organic compound supplied by the environment to serve as both carbon source and electron donor, *R. capsulatus* will grow photoorganotrophically. In the absence of both molecular oxygen and an organic carbon source, *R. capsulatus* will grow photoautotrophically using light as its energy source and carbon dioxide as its sole carbon source. Depending upon its mode of growth, *R. capsulatus* cells synthesize tremendously different quantities of tetrapyrroles.

Since the landmark studies by Cohen-Bazire et al. (21) which clearly demonstrated the dramatic effect of oxygen and light on formation of photosynthetic membranes and photopigments, much of the research into the mechanism of regulation of synthesis of the various tetrapyrroles has focused on the role that oxygen may
play. Working with *R. sphaeroides*, Jones (41) proposed that the synthesis of two tetrapyrroles, bacteriochlorophyll and heme, followed a common pathway until they branched after the production of protoporphyrin IX (Figure 1). If a magnesium atom is incorporated into the porphyrin ring, the ultimate product is bacteriochlorophyll; whereas, if an iron atom is inserted, the eventual product will be heme. Bacteriochlorophyll is the major light-harvesting photopigment and may accumulate up to 25 nanomoles per milligram dry weight under anaerobic conditions (50). Heme serves as the prosthetic group for several hemoproteins including cytochromes, catalase and peroxidase, and is required under all growth conditions.

Two other tetrapyrroles, siroheme and vitamin B-12, branch from the common portion of the tetrapyrrole biosynthetic pathway after the synthesis of uroporphyrinogen III. Siroheme, the prosthetic group of the enzyme sulfite reductase (24), is synthesized when uroporphyrinogen III is methylated and an iron atom is subsequently incorporated into the porphyrin ring. Vitamin B-12 has been shown to be required as a cofactor for methionine biosynthesis in *R. sphaeroides* (19). It is the ultimate product when uroporphyrinogen III is
Figure 1. Tetrapyrrole Biosynthetic Pathway.
methyalted and a cobalt atom is inserted into the porphyrin ring.

In contrast with bacteriochlorophyll which accumulates to 25 nanomoles per milligram dry weight when *R. capsulatus* is grown photosynthetically, heme accumulates to 0.3 nanomoles per milligram dry weight while 0.07 nanomoles of vitamin B-12 accumulate per milligram dry weight (50). When *R. capsulatus* is grown heterotrophically, bacteriochlorophyll synthesis is inhibited (21), while heme and vitamin B-12 are still synthesized. Oxygen, therefore, must play an important role in the regulation of tetrapyrrole biosynthesis.

Most of the research into this role for oxygen has focused on the first step in the common portion of the tetrapyrrole biosynthetic pathway, the condensation of glycine with succinyl CoA to form δ-aminolevulinic acid (45). In *R. capsulatus*, this reaction is catalyzed by the enzyme δ-aminolevulinate synthase which is encoded by the *hemA* gene. The regulation of *hemA* and δ-aminolevulinate synthase will be detailed later in this section.
The condensation of two molecules of δ-aminolevulinic acid with the release of two water molecules to form porphobilinogen is the next step in the pathway. This reaction is catalyzed by porphobilinogen synthase, the product of the hemB gene. In *R. capsulatus* and *R. sphaeroides* porphobilinogen synthase exists as a hexamer of about 240,000 daltons (62). The *R. sphaeroides* form of the enzyme requires the presence of metallic cations and thiols for full activity, and is 96% inhibited by 16 μM protoheme (42). The porphobilinogen synthase from *R. capsulatus* does not require metal cations or thiols for enzyme activity. In addition, 50 μM protoheme results in only 9% inhibition (42).

The gene encoding porphobilinogen synthase has been cloned from rat liver (12), human (99), yeast (61), *E. coli* (53) and recently from *R. sphaeroides* (22). Furthermore, Wetmur et al. (99) have sequenced the full length cDNA clone encoding the human δ-aminolevulinate dehydratase (porphobilinogen synthase) and found it to contain a 990 base pair open reading frame encoding a monomer of about 35,000 daltons. Echelard et al. (26) reported the sequence of the *E. coli* hemB gene and found it to be very similar to other hemB genes.
The next two steps in tetrapyrrole biosynthesis, the condensation and cyclization of four porphobilinogen molecules to yield the first tetrapyrrole, uroporphyrinogen III, are very closely linked. Porphobilinogen deaminase, encoded by the hemC gene, catalyzes the successive condensation of four porphobilinogen molecules to yield hydroxymethylbilane, a linear intermediate (33). Higuchi and Bogorad (34) have shown that porphobilinogen deaminase and uroporphyrinogen III cosynthase, the product of the hemD gene, are associated in a protein complex, and that the hydroxymethylbilane, upon synthesis, is passed directly to the active site of uroporphyrinogen III cosynthase. Uroporphyrinogen III cosynthase then catalyzes the cyclization of hydroxymethylbilane, releasing uroporphyrinogen III. In the absence of uroporphyrinogen III cosynthase, hydroxymethylbilane is non-enzymatically converted to uroporphyrinogen I, a non-physiological form (13).

Porphobilinogen deaminase has been purified 700-fold from R. sphaeroides and found to be a 36,000 dalton polypeptide (44). Jordan and Shemin proposed that porphobilinogen deaminase has a single catalytic site which is used four times in the synthesis of
hydroxymethylbilane (44). Uroporphyrinogen III cosynthase from *R. sphaeroides* has also been partially purified (13). This enzyme forms uroporphyrinogen III from hydroxymethylbilane, but, cannot use either porphobilinogen or uroporphyrinogen I. In *E. coli*, *hemC* and *hemD* form the Uro operon (77) located at 86 minutes on the *E. coli* chromosome (3). Both *hemC* (92) and *hemD* (77) from *E. coli* have been sequenced.

The conversion of uroporphyrinogen III to coproporphyrinogen III by the enzyme uroporphyrinogen III decarboxylase is the least studied step in the tetrapyrrole biosynthetic pathway. Uroporphyrinogen III decarboxylase is the product of the *hemE* gene, and catalyzes the successive removal of four carboxyl groups from side chain acetic acids of uroporphyrinogen III (42, 35). Hoare and Heath (35) partially purified the *R. sphaeroides* uroporphyrinogen III decarboxylase and provided evidence that the carboxyl groups are removed sequentially. Intermediates were isolated which possessed seven, six or five carboxylic acid groups. Romeo et al. (73) have reported the cloning and DNA sequence determination of the human *hemE* gene.
The oxidative decarboxylation of two carboxylic acid groups on coproporphyrinogen to vinyl side chains to yield protoporphyrinogen IX is catalyzed by coproporphyrinogen oxidase (69). This enzyme, encoded by the \textit{hemF} gene has been partially purified from \textit{R. sphaeroides} as a 44 kDa polypeptide (89). Sano and Granick demonstrated that the coproporphyrinogen oxidase from aerobically grown cultures required oxygen for activity (75). Moreover, their work revealed that only one side chain at a time is converted to a vinyl group. They isolated an intermediate which contained one vinyl side chain and three carboxylic acid side chains (75). Tait (88), working with \textit{R. sphaeroides}, and Ehteshamuddin (27), working with \textit{Pseudomonas}, showed that coproporphyrinogen oxidase activity could be obtained from extracts of aerobically and anaerobically grown cultures, with the latter requiring the addition of \textit{s}-adenosylmethionine to the reaction mix. The \textit{s}-adenosylmethionine does not act as a methylating agent, however (89). Moreover, upon subjecting the crude extracts from aerobically and anaerobically grown cultures to high speed centrifugation, the former required only the supernatant to express coproporphyrinogen oxidase activity, while the latter required both the supernatant and pellet (89). The gene
encoding coproporphyrinogen oxidase from yeast has been fully sequenced (107). The hemF strain used to isolate the wild type gene from E. coli has been lost, and there is question as to the integrity of the strain (B. Bachmann, personal communication).

Protoporphyrinogen IX oxidase catalyzes the final step in the common portion of the tetrapyrrole biosynthetic pathway, the removal of six hydrogens from protoporphyrinogen IX to yield protoporphyrin IX (76). Protoporphyrinogen IX oxidase is a membrane bound enzyme encoded by the hemG gene. Under aerobic conditions, the enzyme has an absolute requirement for oxygen to serve as the hydrogen acceptor (68). Under anaerobic conditions in E. coli, however, Jacobs and Jacobs (38) demonstrated that fumarate, and to a lesser extent nitrate (39) will act as the final hydrogen acceptor. Since protoporphyrinogen IX oxidase is 90% inhibited by the quinone inhibitor 2-heptyl-4-hydroxyquinoline-N-oxide, Jacobs and Jacobs propose that it is coupled to electron transport (38). Extraction of quinones from R. sphaeroides has been shown to inhibit, by 80%, the activity of protoporphyrinogen IX oxidase; thus, linking the R. sphaeroides enzyme to electron transport as well (40). To date, no hemG sequences have been reported.
With the formation of protoporphyrin IX, the tetrapyrrole pathway branches off to either bacteriochlorophyll or heme synthesis. If the cell requires bacteriochlorophyll, magnesium chelatase will catalyze the insertion of a magnesium atom into the protoporphyrin IX ring, committing it to bacteriochlorophyll synthesis (31). Ferrochelatase will catalyze the incorporation of an iron atom if heme is required, likewise, committing it to heme synthesis (70).

The synthesis of tetrapyrroles from δ-aminolevulinic acid to protoporphyrin IX is identical in all organisms studied to date. The production of δ-aminolevulinic acid, however, can follow one of two pathways, the C-5 pathway (6), or the Shemin pathway (45). The C-5 pathway has been demonstrated in higher plants (6), cyanobacteria (64), and several bacterial species including Chlorobium vibrioforme (72), Bacillus subtilis (65, 67), Salmonella typhimurium (28, 29), and E. coli (52). The formation of δ-aminolevulinic acid via this pathway involves a three step process starting from glutamate (64). Glutamate is first activated by linking it with a tRNA^{Glu} by the activity of glutamyl-tRNA synthetase. Next, glutamyl-tRNA^{Glu} is reduced to glutamate-1-semialdehyde by glutamyl-tRNA reductase, the product of the hemA gene
(2). Finally, via a transamination reaction, δ-aminolevulinic acid is produced from glutamate-1-semialdehyde. This step is catalyzed by glutamate-1-semialdehyde aminotransferase, the product of the \textit{hemL} gene (29).

The \textit{hemA} gene from \textit{S. typhimurium} (28), \textit{B. subtilis} (67) and \textit{E. coli} (95) have been shown to be part of an operon. In each case, the \textit{hemA} gene was the first gene in the operon. The \textit{E. coli} and \textit{S. typhimurium} operons have, as the second gene in the operon, \textit{prfA} which encodes peptide chain release factor I (95, 28). The \textit{B. subtilis} operon has been shown to be more complex (67). Four open reading frames have been detected, the first being \textit{hemA}, followed by an unidentified ORF2, next is \textit{hemC} and finally, \textit{hemD}. Elliott (28) suggests that his inability to isolate a transposon insertion mutant of \textit{S. typhimurium} for the \textit{hemA} gene is due to a polar effect on an essential downstream gene. This phenomenon will be discussed shortly with respect to the isolation of a \textit{hemA} mutant strain of \textit{R. capsulatus}.

The alternative route for synthesis of δ-aminolevulinic acid, the C-4 or Shemin pathway, is known to function is humans (14) and other mammals (57), and
the purple non-sulfur photosynthetic bacteria (81). As mentioned earlier, this is a one step reaction characterized by the condensation of glycine with succinyl-CoA to form δ-aminolevulinic acid, and is catalyzed by δ-aminolevulinate synthase, the product of the hemA gene. There is confusion in the nomenclature regarding hemA genes. The basis for the confusion dates back to the isolation of an aminolevulinate-requiring strain of E. coli. At that time, E. coli was thought to use the Shemin pathway to synthesize aminolevulinate, thus the strain, based on its phenotype, was designated a hemA mutant. Subsequent work, however, has shown that E. coli uses the C-5 pathway to synthesize aminolevulinate (52).

Studies into the regulation of this initial step have been numerous. Some have suggested that oxygen may inhibit transcription of the hemA gene (63, 50). This theory has been dispelled by Wright et al. (102) who showed, with the use of a hemA-lacZ fusion, that transcription of the R. capsulatus hemA gene is only two-fold higher under low oxygen (3%) tension than under high oxygen (23%) tension. This small amount of regulation relative to the large increase in bacteriochlorophyll accumulation clearly demonstrates that repression of hemA
transcription is not the major site of regulation of the tetrapyrrole biosynthetic pathway.

The question then shifts; is δ-aminolevulinate synthase activity the site of oxygen-mediated regulation? Early work by Shemin showed that δ-aminolevulinate synthase from several sources is inhibited by hemin in vitro (81). Lascelles and Altschuler (49) demonstrated that δ-aminolevulinate synthase activity from R. sphaeroides increased four-fold upon a shift from high oxygen to low oxygen. Following a 1300-fold purification of δ-aminolevulinate synthase from R. sphaeroides, Warnick and Burnham found its activity was inhibited 50% by 4 mM protoheme (97). Further characterization of the purified enzyme estimated its molecular weight to be 57,000 daltons (97). Shortly thereafter, Yubisui and Yoneyama purified δ-aminolevulinate synthase from R. sphaeroides (106) and found that its activity was 50% inhibited by both 0.4 μM protoheme, and 2 μM mg-protoporphyrin.

Tuboi et al. (93), and later, Wider de Xifra et al. (100) found that, upon purification by DEAE-sephadex chromatography, R. sphaeroides δ-aminolevulinate synthase activity could be found in 2 fractions, designated
fraction I and fraction II. Fraction I consisted of both an active form and an inactive form of the enzyme. The inactive form could be rendered active by decreasing the oxygen tension and adding cystein trisulfide (92, 100). The inactive form from fraction II was made active by decreasing the oxygen tension and exposing it to bright light (93). The activities from both fractions were inhibited by high oxygen tension, and the molecular weight of the δ-aminolevulinate synthase from both fractions was estimated to be 100,000. An extension of this work by Fancia-Gaignier and Clement-Metral revealed that both fractions were 75% inhibited by 1 mM ATP (30).

None of these forms of regulation appears to be complete enough to account for the rapid decrease in bacteriochlorophyll synthesis upon shifting a culture from low oxygen tension to high oxygen tension. Lascelles proposed a model for oxygen regulation of tetrapyrrole synthesis (50) which hinges on competition for protoporphyrin IX between magnesium chelatase and ferrochelatase. In this model, magnesium chelatase has a higher affinity for protoporphyrin IX than does ferrochelatase. When the magnesium branch of the pathway, leading to the synthesis of bacteriochlorophyll synthesis, is inhibited, the protoporphyrin IX is then
used solely by ferrochelatase. This results in an increase in heme synthesis which feedback inhibits δ-aminolevulinate synthase activity (17). Recently, Bauer and Marrs (4) found that the *R. capsulatus* *puf* operon encodes a regulatory protein, designated PufQ, which appears to regulate synthesis of bacteriochlorophyll. In their model, Bauer and Marrs suggest that the actual substrate for magnesium chelatase is the complex between protoporphyrin IX and PufQ. Synthesis of PufQ is known to be highly regulated by oxygen (4), and more recently, Adams et al. (1) provided evidence that the mRNA for the *puf* operon is posttranscriptionally processed yielding discrete mRNAs that are degraded at different rates. The mRNA encoding PufQ was shown to be degraded fastest (1). The model, therefore, asserts that under low oxygen, PufQ is synthesized in high levels. The PufQ then complexes with protoporphyrin IX and serves as the substrate for magnesium chelatase, resulting in bacteriochlorophyll synthesis. Under high oxygen tension, however, PufQ synthesis is strongly repressed, and the mRNA is rapidly degraded; consequently, no complex is formed and protoporphyrin IX serves as the substrate for ferrochelatase. This results in an increase in heme synthesis which feedback inhibits δ-aminolevulinate synthase activity.
Experiments in which exogenous δ-aminolevulinic acid was added to cultures of an R. capsulatus strain unable to convert protoporphyrin IX to Mg-protoporphyrin monomethylester demonstrated that excess δ-aminolevulinic acid could not overcome oxygen-mediated regulation of protoporphyrin IX accumulation. This indicates that, in vivo, δ-aminolevulinic acid is not the main oxygen-regulated step in the common tetrapyrrole pathway (A. Biel, unpublished data).

The hemA gene encoding δ-aminolevulinate synthase has been cloned from human (5), chicken (54), mouse (78), yeast (94), R. sphaeroides (87) and R. capsulatus (10, 36, 102). Tai et al. (87) have reported the existence of two δ-aminolevulinate synthases from R. sphaeroides, the second, less active is encoded by the hemT gene. Wright et al. (102) used Tn5 mutagenesis to isolate a hemA mutant strain of R. capsulatus which lacked 95% of δ-aminolevulinate synthase activity. Of the 2,500 colonies which were scored for δ-aminolevulinic acid dependence, only one was found, and it was leaky. Recall that Elliott (28) reported the inability to isolate a hemA mutant in S. typhimurium using transposon mutagenesis due to its polar effect on an essential downstream gene. The inability to isolate a hemA mutant completely lacking δ-
aminolevulinate synthase activity from \textit{R. capsulatus} may also be due to a polar effect. A preliminary experiment to determine whether transcription exists downstream of the \textit{R. capsulatus} \textit{hemA} gene was performed (M. Wright, unpublished data). The chloramphenicol acetyl transferase (cat) gene from pCM7 (20) was inserted downstream of the \textit{hemA} gene, and cat activity was assayed as a measure of transcription of this region. The results obtained demonstrated that oxygen regulated transcription occurs 200 bases past the stop codon for \textit{hemA}. The aim of this research, therefore, was to determine the nucleotide sequence of the \textit{R. capsulatus} \textit{hemA} gene, identify the exact site of the Tn5 insertion in the \textit{hemA} mutant strain and explain the partial δ-aminolevulinate synthase activity observed in the mutant. To date, only 32 genes have been sequenced from \textit{R. capsulatus} (24). A partial explanation for this low number may be its ~70 mole \% G+C content (55). As will be discussed more fully soon, a high G+C content often results in DNA sequence artifacts, making the task of determining the sequence very arduous.

The method of DNA sequence determination used in this study was first described by Sanger et al. (74). It exploits the requirement of DNA polymerases for a 3'-
hydroxyl group on the ribose moiety of an extending chain in order to catalyze the formation of the next phosphodiester bond. The initial requirement for a 3'-hydroxyl group is met by hybridization of a synthetic primer to single stranded templated DNA upstream of the region to be sequenced. By adding 2',3'-dideoxynucleoside triphosphates to the reaction, along with the physiological 2'-deoxynucleoside triphosphates, upon incorporation of a dideoxynucleotide, extension of the newly synthesized chain will terminate (74). The DNA sequencing reactions are split into four tubes, each containing a specific 2',3'-dideoxynucleoside triphosphate. Each tube will contain fragments with the same 5' end; but, the 3' termini will be determined by the specific dideoxynucleoside triphosphate in that tube. The fragments are subjected to polyacrylamide gel electrophoresis in adjacent lanes to separate them on the basis of size. The DNA sequence is then determined by noting the relative position of bands on the resulting autoradiogram.

Until recently, most DNA sequencing reactions were performed using the large (Klenow) fragment of *E. coli* DNA polymerase I (74). Tabor and Richardson (84) described three major advantages of a chemically modified
version of phage T7 DNA polymerase over the Klenow fragment. First, T7 DNA polymerase is highly processive, able to catalyze the incorporation of thousands of nucleotides without dissociating, whereas the Klenow fragment will dissociate after an average of 10 nucleotides is incorporated. Second, the chemical modification of T7 DNA polymerase eliminated all exonuclease activity resulting in less variability in band intensity than seen with the *E. coli* DNA polymerase. Finally, *E. coli* DNA polymerase I large fragment discriminates 1000-fold against dideoxynucleoside triphosphates as well as other nucleotide analogs. The modified T7 enzyme was found to readily incorporate both dideoxynucleoside triphosphates and other base analogs. Subsequent *in vitro* mutagenesis work by Tabor and Richardson (85) resulted in a T7 DNA polymerase which was 9-fold more processive than the chemically modified enzyme. It, too, incorporates nucleotide analogs very efficiently.

A common problem with DNA sequencing from *R. capsulatus* is G-C compression. Formation of stable stem-loop structures during the synthesis steps of sequencing reactions often results in resolution artifacts when subjected to electrophoresis (USB, Cleveland, OH). A
region of inverted repeats containing a number of G-C base pairs is often not denatured under standard electrophoresis conditions. The result of such secondary structures is anomalous migration rates of DNA fragments through the acrylamide matrix. Fragments which encompass the entire stem-loop structure will migrate at the same rate as fragments which only contain one-half of the stem-loop sequence, thus masking the second half of the sequence (USB).

To eliminate G-C compression, a nucleotide analog of guanosine triphosphate is substituted in the sequencing reactions. The analog used is inosine triphosphate which has the same structure as guanosine triphosphate except it lacks the amino group bonded at the C-2 position. This amino group of guanine would form a hydrogen bond with the C-2 oxygen of cytosine. Since this amino group is missing on inosine, only two hydrogen bonds are able to form between cytosine and inosine. This makes disruption of secondary structure more favorable under standard electrophoresis conditions.

The single stranded template DNA is derived using the M13mp18/19 cloning vectors (59). M13 exists as a filamentous single stranded DNA phage which infects E.
coli cells via the F-pili. Once inside the host cell, the complementary strand of DNA is synthesized, and the DNA then exists in a double stranded, replicative form (RF). This replicative form of M13, when isolated, can be manipulated like a plasmid; thus, making cloning possible. The single stranded form can be isolated and serve as the template for DNA sequencing reactions.

Messing et al. (59) have described the construction of a series of M13 derivatives which combine multiple restriction enzyme recognition sites and an easy screening mechanism for recombinant phage. Their group cloned a fragment of the E. coli chromosome containing the lac regulatory region plus the coding region for the first 146 amino acids of β-galactosidase into M13 RF DNA. When the resulting phage infects a lac host which harbours a defective lacZ gene on its F-factor, the 146 amino acid peptide from the amino terminal end of β-galactosidase will complement the defective β-galactosidase. This is referred to as α-complementation. Upon infection of E. coli with these phage and incubation on plates containing 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and isopropyl-β-D-thiogalactopyranoside (IPTG), the resultant plaques will appear blue.
Cloned into the amino terminal portion of the gene for β-galactosidase on the M13 derivative is a segment of DNA containing nine unique restriction enzyme recognition sequences. The presence of this multicloning region within the coding region for β-galactosidase does not interfere with α-complementation. If a fragment of DNA is inserted in this multicloning region, however, α-complementation is disrupted and plaques from these recombinant phage will be white in the presence of X-gal and IPTG.

In addition to determining the sequence of the hemA gene, I wished to determine whether the difficulty encountered trying to isolate a completely disrupted hemA gene might be due to a polar effect on a downstream gene. This part of the study involved continued sequence determination downstream of hemA in an effort to uncover another open reading frame(s) which may be transcribed along with hemA from a common promotor.
Materials and Methods

Bacterial Strains, Phage and Plasmids

Tables 1 and 2 list the bacterial strains, phage and plasmids used in this study.

Media

*Escherichia coli* was routinely grown in LB medium (8) modified by omitting glucose and reducing the concentration of sodium chloride to 0.5%. *E. coli* strain NM522 was maintained in M9 minimal salts medium (32) supplemented with 10 μg ml⁻¹ thiamine. Medium for template preparation was 2X YT (0.5% bacto-yeast extract 0.8% bacto-tryptone; Difco Laboratories, Detroit, Mich.). *Rhodobacter capsulatus* was grown in either RCV minimal medium (98), or 0.3% peptone, 0.3% yeast extract (9).

Antibiotics and other supplements were added at the following final concentrations (μg ml⁻¹): ampicillin, 50; chloramphenicol, 25; kanamycin, 10; spectinomycin, 25 for *E. coli* and 5 for *R. capsulatus*; tetracycline, 10 for
### Table 1. Bacterial Strains used in this Study

<table>
<thead>
<tr>
<th>Designation</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM522</td>
<td>supE thi Δ(lac-proAB) hsd5</td>
<td>Gough and Murray (32)</td>
</tr>
<tr>
<td></td>
<td>F'[proAB lacI lacZΔM15]</td>
<td></td>
</tr>
<tr>
<td>MC1061</td>
<td>hsdR2 araD139 Δ(araABC-leu)7679 Cassadaban</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Δ(lac)X74 galU galK rpsL thi</td>
<td>and Cohen (18)</td>
</tr>
<tr>
<td>GM2163</td>
<td>F' ara-14 leuB6 thi-1</td>
<td>Woodcock et al.</td>
</tr>
<tr>
<td></td>
<td>fhuA31 lacY1 tsx-78 galK2</td>
<td>(101).</td>
</tr>
<tr>
<td></td>
<td>galT22 supE44 hisG4 rpsL136</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Str') xyl-5 mtl-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dam13::Tn9 (Cam') dcm-6 mcrB1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hsdR2 (rK-mK+) mcrA</td>
<td></td>
</tr>
</tbody>
</table>
## Table 1. Continued.

<table>
<thead>
<tr>
<th>Designation</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HB101</strong></td>
<td>$hsdS20\ (r_B^b\ m_B^b)\ recA13$</td>
<td>Boyer and Roulland-</td>
</tr>
<tr>
<td></td>
<td>$leu-6\ thi-1\ ara-14\ proA2$</td>
<td>Dussiox (15).</td>
</tr>
<tr>
<td></td>
<td>$lacY1\ galK2\ rpsL20\ (Sm')$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$xyl-5\ mtl-1\ supE44\ \lambda$</td>
<td></td>
</tr>
<tr>
<td><strong>C600</strong></td>
<td>$supE44\ hsdR\ thy-1\ thr-1$</td>
<td>Taylor and Cohen (90).</td>
</tr>
<tr>
<td></td>
<td>$leuB6\ lacY1\ tonA21$</td>
<td></td>
</tr>
</tbody>
</table>

**R. capsulatus**

<table>
<thead>
<tr>
<th>Designation</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PAS100</strong></td>
<td>$hsd-1\ str-2$</td>
<td>Taylor et al. (91).</td>
</tr>
<tr>
<td><strong>AJB529</strong></td>
<td>$hemA1::Tn5\ hsd-1\ str-2$</td>
<td>Wright et al. (102).</td>
</tr>
<tr>
<td>Phage or Plasmid</td>
<td>Relevant characteristics*</td>
<td>Reference or Description</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>pNM481 and pNM482</td>
<td>Ap'</td>
<td>Minton (60)</td>
</tr>
<tr>
<td>pDPT51</td>
<td>Ap', mob', tra'</td>
<td>Taylor et al. (91).</td>
</tr>
<tr>
<td>pBR322Ω</td>
<td>Ap', Tc', Sp'</td>
<td>Omega cartridge cloned</td>
</tr>
<tr>
<td></td>
<td></td>
<td>from pH45Ω (71)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>into the EcoRI site</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of pBR322. Donahue.</td>
</tr>
<tr>
<td>pCAP1</td>
<td>Ap', Km', Tc'</td>
<td>S. Biel et al. (10)</td>
</tr>
<tr>
<td>pCAP17</td>
<td>Tc', mob', tra'</td>
<td>S. Biel et al. (10)</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCAP22</td>
<td>Ap', Tc', Cm'</td>
<td>0.6 kb EcoRI fragment from pCAP17 cloned into pSUP202. S. Biel.</td>
</tr>
<tr>
<td>pCAP40</td>
<td>Tc', mob', tra'</td>
<td>pCAP17 with a deletion of the 1.0 kb PstI fragment. S. Biel.</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCAP41</td>
<td>Tc', mob', tra'</td>
<td>Deletion of the 2.3 kb EcoRI fragment from pCAP17. S. Biel.</td>
</tr>
<tr>
<td>pCAP47</td>
<td>Tc', mob', tra'</td>
<td>pCAP17 with the 1.9 kb EcoRI fragment deleted. S. Biel.</td>
</tr>
<tr>
<td>Phage or Plasmid</td>
<td>Relevant Characteristics</td>
<td>Reference or Description</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>pCAP77</td>
<td>Ap', hemA*</td>
<td>Translational fusion between orf-2 and lacZ of pNM481 at the HindIII site. (This study).</td>
</tr>
<tr>
<td>pCAP80</td>
<td>Ap', hemA*</td>
<td>4.0 Kb BglII fragment containing hemA and orf-2 cloned from into BamHI digested pUC8 (96) (This study).</td>
</tr>
<tr>
<td>pCAP81</td>
<td>Ap', Tc', Km'</td>
<td>1.5 kb Kan cartridge cut from pUC-4k with HincII, ligated with Smal digested pCAP27 (This study).</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCAP83</td>
<td>Ap'</td>
<td>1.9 kb EcoRI fragment from pCAP27 cloned into pKK223-3 (16). (This study).</td>
</tr>
<tr>
<td>pCAP84</td>
<td>Ap', <em>hemA</em></td>
<td>2.0 kb SmaI-PstI fragment from pCAP41 cloned into SmaI and PstI digested pNM482 forming a translational fusion between orf-2 and <em>lacZ</em> (This study).</td>
</tr>
</tbody>
</table>

Phage

- M13mp18 and *lacZ*'
- M13mp19

Messing (59)
<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13mp18.1</td>
<td>Km(^{r})</td>
<td>3.7 kb BamHI-EcoRI fragment from pCAP1 cloned into M13mp18. (This study).</td>
</tr>
<tr>
<td>M13mp19.1</td>
<td>Km(^{r})</td>
<td>M13mp18.1 with insert in opposite orientation. (This study).</td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics*</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13mpl8.44+</td>
<td>1.4 kb EcoRI fragment from pCAP44 cloned into M13mpl8 such that sequence generated is in the same direction as transcription of hemA. (This study).</td>
<td></td>
</tr>
<tr>
<td>M13mpl8.44-</td>
<td>Same as M13mpl8.44+, but insert in opposite orientation. (This study).</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics*</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13mpl8.416+</td>
<td>416 base pair NsiI-EcoRI fragment from pCAP44 cloned into PstI and EcoRI digested M13mpl8. (This study).</td>
<td></td>
</tr>
<tr>
<td>M13mpl9.X600</td>
<td>1.0 Kb BclI-HindIII fragment from pCAP17 cloned which spans the 0.6 kb EcoRI fragment cloned into BamHI-HindIII digested M13mpl9. (This study).</td>
<td></td>
</tr>
<tr>
<td>Phage or Plasmid</td>
<td>Relevant Characteristics*</td>
<td>Reference or Description</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>M13mp18.44a</td>
<td>M13mp18.44 with the 416 base pair NsiI-EcoRI fragment deleted. (This study).</td>
<td></td>
</tr>
<tr>
<td>M13mp19.44a</td>
<td>Same as M13mp18.44a, but insert in opposite orientation. (This study).</td>
<td></td>
</tr>
<tr>
<td>M13mp18.22-1</td>
<td>0.6 kb EcoRI fragment from pCAP22 cloned in M13mp18. (This study).</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Continued.

<table>
<thead>
<tr>
<th>Phage or Plasmid</th>
<th>Relevant Characteristics*</th>
<th>Reference or Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13mp18.22-2</td>
<td>Same as M13mp18.22-1 insert in opposite orientation. (This study).</td>
<td></td>
</tr>
<tr>
<td>M13mp18-1.9+ orf-2'</td>
<td>1.9 kb EcoRI fragment from pCAP27 cloned into M13mp18. (This study).</td>
<td></td>
</tr>
<tr>
<td>M13mp18-1.9- orf-2'</td>
<td>Same as M13mp18-1.9+ insert in opposite orientation. (This study).</td>
<td></td>
</tr>
</tbody>
</table>

*Antibiotics are abbreviated as follows: Ap=ampicillin, Cm=chloramphenicol, Km=kanamycin, Tc=tetracycline.
E. coli and 1 for R. capsulatus; 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), 40; Isopropyl-β-D-thiogalactopyranoside, 5. δ-Aminolevulinic acid was added at 10 μM (102).

Bacterial strains were stored at -70°C in 10% glycerol. Bacteriophage M13 stocks were stored at 4°C as lysates in LB medium.

**Plasmid and M13 Replicative Form Preparation**

**Minipreparation**

Minipreparations of plasmid DNA were performed using the procedure described by Birnboim and Doly (11). Colonies containing the plasmid of interest were picked from an agar plate using a tooth pick and transferred to 2.0 ml of broth medium with the appropriate antibiotic(s). The cultures were incubated overnight with shaking at 37°C. One and one-half milliliters of the culture was spun in a microcentrifuge for 1 minute to pellet the cells. The supernatant was removed and discarded, and the pellet was resuspended in 100 μl of 25 mM Tris·HCl (pH 8.0), 50 mM glucose, 10 mM disodium ethylenediaminetetraacetic acid (EDTA) and 5 mg lysozyme
per ml. The resuspended cells were incubated for 5 minutes at room temperature. Two hundred microliters of 0.2 N NaOH plus 1% sodium dodecyl sulfate was added and the tube was shaken gently to thoroughly mix the contents. Following incubation at 4°C for 5 minutes, 150 μl of 5 M potassium acetate (pH 4.6) was added and the tube was shaken gently and incubated for an additional 10 minutes at 4°C.

Following this incubation, the tube was spun in a microcentrifuge for 5 minutes to pellet the cell debris. The supernatant was transferred to a new tube which was then filled with cold 95% ethanol and placed at -70°C. The DNA was pelleted by centrifugation for 5 minutes. The supernatant was aspirated and the DNA pellet was resuspended in 100 μl of 0.1 M sodium acetate, 50 mM Tris•HCl (pH 8.0) and 300 μl of cold ethanol. The tube was incubated at -70°C for 5 minutes, then centrifuged for 5 minutes. This wash step was repeated once, then the DNA pellet was dried in a vacuum desiccator for 15 minutes, or in the Speed-Vac (Savant) for 5 minutes.
The procedure for minipreparation of M13 replicative form (RF) DNA was the same as that for plasmid DNA, with the following exceptions. A mixture of NM522 and phage from a lysate was layered atop an LB agar plate and incubated overnight at 37°C. The following day, a plaque was picked with a pasteur pipet and transferred to 2 ml of LB broth which contained 200 μl of early log phase NM522. This culture was incubated for 6 hours at 37°C with vigorous shaking. M13 RF DNA was isolated from the infected cells as described above.

**Large Scale Preparation**

Large scale cesium chloride plasmid preparations were performed as described by Sidikaro and Nomura (9). One liter of LB broth containing the appropriate antibiotic(s) was inoculated with 10 ml of a 3 X 10⁸ cells ml⁻¹ starter culture and incubated overnight at 37°C, or for 6 hours in the case of M13. The cells were harvested by centrifugation in a Sorvall RC-5B (E.I. DuPont, Wilmington, DE) for 10 minutes at 12,000 x g. The cell paste was spread on parafilm, wrapped in aluminum foil and placed at -70°C for at least 30 minutes. The frozen cells were then thawed on ice in 15
ml of 50 mM Tris·HCl (pH 8.0), 25% sucrose per liter of original culture. Powdered lysozyme, 30 mg per liter of culture, was added to the thawed cell suspension and this was incubated for 10 minutes at 4°C. Following this incubation, 5 ml of 0.25 M EDTA in 50 mM Tris·HCl (pH 8.0) was added and incubation was continued for another 10 minutes at 4°C. Seven milliliters of 5 M sodium chloride and 3 ml of 10% SDS was added, the beaker was swirled gently to mix the contents, and incubation was continued for 20 minutes at 4°C.

Following this incubation, the mixture was transferred to a centrifuge tube and centrifuged for 1 hour at 20,000 x g in a Sorvall GSA rotor. The supernatant was transferred to a teflon centrifuge tube and an equal volume of cold 20% polyethylene glycol 8000, 6% sodium chloride was added, and the tube was incubated for at least 20 minutes at 4°C. The tube was then centrifuged at 8,000 x g for 5 minutes. The supernatant was completely decanted and the pellet was resuspended in 3.5 ml of 10 mM Tris·HCl (pH 8.0), 1 mM EDTA (TE). An equal volume of chloroform was added, the contents were gently mixed by inversion, and the tube was centrifuged
at 8,000 x g for 10 minutes. The aqueous layer was removed to a 25 ml beaker and the volume was adjusted to 4 ml. Four grams of cesium chloride was dissolved and 300 µl of 10 mg ml⁻¹ ethidium bromide was added. This mixture was then transferred to an ultracentrifuge tube which was then heat sealed and centrifuged at 50,000 rpm overnight in a Beckman VTi65 rotor (Beckman Instruments, Inc. San Ramon, CA).

The plasmid band was collected and transferred to a Corex tube and the ethidium bromide was removed from the DNA by extracting 3 to 5 times with an equal volume of secondary butanol. The DNA was then precipitated by adding 3 volumes of cold ethanol and incubating the mixture at -70°C for 20 minutes, followed by centrifugation at 10,000 x g for 15 minutes. The resulting pellet was resuspended in 1 ml of deionized water to which 100 µl of 5 M potassium acetate (pH 4.6) and 3.3 ml of cold ethanol was added. The tube was incubated at -70°C for 20 minutes and then centrifuged at 10,000 x g for 15 minutes. The DNA pellet was dried under vacuum and resuspended in 250 µl of sterile deionized water. The concentration of the DNA solution
was determined using a Perkin-Elmer Lambda 3B spectrophotometer. A 50 µg ml⁻¹ solution of DNA will yield an absorbance of 1 O.D. unit at 260 nm.

**Restriction Enzyme Digestion**

Typically, 1 µg of DNA was digested in a 10 µl volume which contained 1 µl of the appropriate 10X reaction buffer (Bethesda Research Laboratories, Gaithersburg, MD.) plus 0.5-1.0 µl of the restriction enzyme. Digestions were carried out for 1-4 hours at 37°C in a dry block incubator. If the DNA was to be subjected to electrophoresis, 3.0 µl of loading buffer (0.42% bromphenol blue, 40% sucrose in water) was added to the reaction mix just prior to electrophoresis.

**Agarose Gel Electrophoresis**

Agarose gel electrophoresis was carried out in a Hoefer Submarine Gel Unit, Model HE 99, or a Hoefer Minnie Gel Unit, Model HE 33 powered by a Hoefer PS 500x DC Power Supply. A 0.7% agarose gel was prepared in 1 X Tris-acetate (0.04 M Tris-acetate, 1 mM EDTA) to which 0.5 µg ethidium bromide ml⁻¹ was added. Electrophoresis was conducted for 1 hour at 90 volts, or several hours at
60 volts. The DNA bands were then visualized using a UV transilluminator and DNA fragment sizes were estimated by comparison with molecular weight markers obtained by digesting lambda DNA with HindIII.

Ligation

Following restriction enzyme digestion, DNA to be ligated was precipitated by adding 1.0 μl of 5 M potassium acetate and 33 μl of cold ethanol. The tube was placed in a -70°C freezer for 5 minutes, then centrifuged for 5 minutes. The supernatant was aspirated off and the pellet was dried in a vacuum desiccator for 15 minutes or in a Savant Speed Vac for 5 minutes. The dried pellet was then resuspended in 7.0 μl of sterile deionized water and 2.0 μl of 5X T4 DNA ligase buffer and 1.0 μl of T4 ligase (Bethesda Research Laboratories, Gaithersburg, MD.) were added. The reaction was incubated overnight in a 12.5°C waterbath for cohesive end ligations, or at 25°C for blunt end ligations.

Transformation of Plasmid and M13 Replicative Form DNA

Transformations were performed using the procedure of Kushner (48). A 10 ml culture of E. coli was
incubated with vigorous shaking at 37°C until a cell density of 5 \times 10^7 \text{ cells ml}^{-1} was achieved. A microfuge tube was then filled with 1.5 ml of this culture and spun for 2 minutes to pellet the cells. The cell pellet was resuspended in 1.0 ml of 10 mM morpholinopropane sulfonic acid (MOPS) (pH 7.0), plus 10 mM rubidium chloride to wash the cells. This suspension was immediately centrifuged for 2 minutes. The resulting pellet was resuspended in 1.0 ml of 100 mM MOPS (pH 6.5), 10 mM rubidium chloride and 50 mM calcium chloride. The cell suspension was incubated at 4°C for 15 minutes, then spun for 2 minutes. The pellet was resuspended in 200 µl 100 mM MOPS (pH 6.5), 10 mM rubidium chloride and 50 mM calcium chloride. Three microliters of dimethylsulfoxide and 1-200 ng of DNA was added to the suspension and incubation was continued for 30 minutes at 4°C. Following the 30 minute incubation, the suspension was heat shocked at 43-44°C for 1 minute.

At this point, if transforming M13 RF DNA, 10 µl and 100 µl aliquots of the suspension were added to 3.0 ml of 0.7% water top agar containing 200 µl of NM522, 40 µg
ml\textsuperscript{1} X-gal and 5 \(\mu\text{g}\) ml\textsuperscript{1} IPTG. The tube was vortexed and its contents were layered atop a dry LB agar plate. If plasmid DNA was being used for transformation, the heat shocked mixture was diluted to 5.0 ml with LB broth and incubated standing for 1 hour at 37°C, then 10 \(\mu\text{l}\) and 100 \(\mu\text{l}\) aliquots were spread over the surface of a selective agar medium.

**Mating**

Transfer of DNA into *R. capsulatus* strains was accomplished using the plate mating method described by Marrs (56). For triparental matings, 1 ml of the *R. capsulatus* recipient strain, grown, standing overnight at 37°C, was mixed with 200 \(\mu\text{l}\) of HB101 (15) containing the mobilization helper plasmid pRK2013, and 200 \(\mu\text{l}\) of the *E. coli* donor strain. This mixture was then centrifuged to pellet the cells. The supernatant was decanted and the cells were resuspended in the remaining liquid. The resulting suspension was spotted on a sterile nitrocellulose filter disk which had previously been placed on the surface of an RCV agar plate. Conjugation was allowed to continue for 3 hours to overnight. The mating mixture was then spread across the plate with 100
μl of RCV broth or appropriate antibiotic and the plate was incubated at 37°C.

**Sonication**

Cultures of *E. coli* or *R. capsulatus* were incubated in the appropriate medium until the culture achieved early log phase. The cells were then harvested by centrifugation at 12,000 x g for 10 minutes. The cell pellet was resuspended in 1.0 ml of 0.1 M Tris·HCl (pH 8.2) per gram wet weight of cells and the resulting suspension was subjected to a 20 second burst from a Heat Systems Sonicator, Model W-220 (Ultra Sonics, Inc.) followed by a 30 second cooling period. This sonication was repeated 3 to 4 times until the suspension cleared. The cell debris was removed by centrifugation at 12,000 x g for 10 minutes and the supernatant was removed to a new tube.

**Protein Determination**

The protein concentration of cell extracts to be used for enzyme assays was determined using the Bio-Rad Protein Assay Reagent using bovine serum albumin as a standard.
Porphobilinogen Synthase Assay

Porphobilinogen synthase activity was measured using the method described by Shemin (81). Cell extract containing 1 mg of protein was added to 150 μl of 1 M Tris-HCl (pH 8.5), 75 μl of 1 M KCl, 50 μl of 0.1 M aminolevulinic acid (pH 7.0), 0.6 μl of β-mercaptoethanol and the volume was adjusted to 1.5 ml. The reaction mixture was incubated for 30 minutes at 37°C. The reaction was terminated by adding 0.5 ml of 20% trichloroacetic acid, 0.1 M HgCl₂. The reaction tube was spun for 2 minutes in a microcentrifuge and the supernatant was transferred to a tube which contained 2 ml of modified Ehrlich's reagent (0.5 g p-dimethylamino-benzaldehyde, 17 ml glacial acetic acid and 8 ml of 70% perchloric acid) and the reaction was incubated for 5 minutes at 25°C. The absorbance of the reaction mixture was measured at 556 nm. The absorbance at 556 nm of a 1 molar solution of porphobilinogen is 6.2 X 10⁴ (79).

Porphobilinogen Deaminase Assay

Porphobilinogen deaminase activity was assayed using the method of Jordan (43). Crude extract containing 0.5 mg of protein was incubated with 0.1 M Tris-HCl (pH 8.2)
in a total volume of 900 µl, at 37°C for 2 minutes to prewarm the reaction mixture. To the reaction tube, 100 µl of 0.5 mg ml⁻¹ porphobilinogen was added to start the reaction. One hundred microliters of the reaction mixture was immediately removed to a tube containing 100 µl of 10% trichloroacetic acid (TCA) which was subsequently placed on ice. The reaction tube was placed in a 37°C dry block incubator, and 100 µl aliquots were removed to 10% TCA at 10 minute intervals over a total of 40 minutes and stored on ice. The tubes containing the terminated reactions were centrifuged for 10 minutes to remove proteins and the supernatants were transferred to 300 µl of sterile deionized water. An equal volume of modified Ehrlich's reagent was added to the mixture, and the reactions were incubated for 10 minutes at 25°C. The absorbance for each time interval was measured at 556 nm to determine the concentration of porphobilinogen.

**DNA Sequencing**

DNA sequence data was obtained using the method described by Sanger et al. (74). The Sequenase Version 2.0 kit (United States Biochemical Corp, Cleveland, OH)
was found to provide the most reliable sequence and was used in this work.

**Template Preparation**

Single stranded template DNA for sequencing was prepared by the method of Kristensen et al. (47) and purified using the Sephaglas M13 Miniprep Kit (Pharmacia, Piscataway, N.J.). Approximately 100 plaque forming units from a lysate of recombinant M13 were mixed with 200 μl of fresh overnight NM522 and 3.0 ml of 0.7% water agar. The mixture was vortexed and layered atop an LB agar plate which was incubated at 37°C overnight.

A single plaque was picked from the plate with a pasteur pipet and added to 2.0 ml of 2X YT broth plus 200 μl of early log phase NM522 in a 50 ml Erlenmeyer flask. This mixture was shaken vigorously at 37°C for 5 to 8 hours. One and one-half milliliters of this culture was transferred to a microfuge tube and spun for 2 minutes to pellet the cells. M13 phage from 1.0 ml of the supernatant was purified according to directions supplied with the Sephaglas M13 Miniprep Kit.
**Primer Annealing**

Four microliters of template DNA was mixed with 1.0 μl (1.0 pmol) of primer DNA (Operon Technologies, Austin, TX) 2.0 μl of reaction buffer (200 mM Tris·HCl pH 7.5, 100 mM MgCl\(_2\), and 250 mM NaCl) and 3.0 μl of deionized water. The mixture was placed in a 72°C water bath or dry block, held at 72°C for 2 minutes then allowed to cool for 30 minutes to below 35°C.

**Labeling Reaction**

One microliter of 100 mM dithiothreitol (DTT), 2.0 μl of diluted labeling mix (1:5 dilution of 7.5 μM dGTP, dCTP and dTTP), 0.5 μl dATPα-thio ester (\(^{35}\)S, New England Nuclear,) and 2.0 μl of Sequenase Version 2.0 (85) (diluted 1:8 in 10 mM Tris·HCl ph 7.5, 5 mM DTT, 0.5 mg/ml BSA) was spotted on the side of the tube containing the annealed template and primer reaction tube. The reaction tube was spun to mix the reagents together at the bottom of the tube. These reactions were incubated at room temperature for 5 minutes.
Termination Reaction

While the annealing step was in progress, four tubes for each reaction were labeled T, C, G and A respectively. To each tube, 2.5 μl of the appropriate termination mix (80 μM of all four dNTP's, 8.0 μM of the specific ddNTP and 50 mM NaCl) was added. These tubes were stored in a 37°C water bath until needed. When the labeling reaction was complete, 3.5 μl of the labeling reaction mix was spotted on the side of each of the termination reaction tubes and spun to mix the reagents at the bottom of the tube. These reactions were incubated at 37°C or 42°C for 5 minutes. At the completion of the termination step, 4.0 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol FF) was added to the reactions which were then stored in a frost free freezer until they were subjected to electrophoresis.

Determination of Sequence Close to the Primer

Under the standard reaction conditions, sequence within 30 nucleotides of the primer is rarely obtained (86). Tabor and Richardson have shown that in the presence of manganese, T7 DNA polymerase incorporates
dideoxynucleoside triphosphates 4-fold more efficiently than when incubated in the presence of magnesium (86). One microliter of manganese buffer (150 mM sodium isocitrate and 100 mM MnCl₂) was added to the reaction tube in the labeling reaction and the subsequent reactions were incubated as previously described.

**Determination of Sequence Further from the Primer**

Under standard reaction conditions, labeling of fragments fades out about 650 nucleotides from the primer (85). By decreasing the dideoxynucleoside triphosphate:deoxynucleoside triphosphate ratio in the termination step it is possible to increase the labeling of fragments further from the primer. Typically 1.5 μl of sequence extending mix (180 mM each dTTP, dCTP, dGTP, dATP and 50 mM NaCl) plus 1.0 μl of the standard termination mix were mixed in the appropriate termination reaction tube. Reaction conditions were the same as previously described.

**Relieving G-C Compression**

Deoxyinosine triphosphate and dideoxyinosine triphosphate are substituted for dGTP and ddGTP in the
labeling and termination reactions, respectively. Instead of adding 2.0 \mu l of GTP labeling mix, 2.0 \mu l of the ITP labeling mix (15 \mu M dITP, 7.5 \mu M dCTP, 7.5 \mu M dTTP) was added. In the termination reaction, ddI termination mixes were substituted for ddG termination mixes. Reaction times were reduced to 4 and 3 minutes for the labeling and termination reactions, respectively.

Polyacrylamide Gel Electrophoresis

Preparation of the Plates

The 30 cm wide x 35 cm long front plate which was bevelled at the top was coated with Sigmacote (Sigma Chemical Company, St. Louis, MO) 15 minutes before the plates were taped together. After the excess Sigmacote was wiped from the top plate, it was laid atop the 30 cm wide x 40 cm long bottom plate, with the side containing the sigmacote down. Between the two plates at the side edges were wedge spacers with the 0.2 mm end at the top and the 0.6 mm end at the bottom. The plates were taped together at the sides first and then at the bottom to prevent the acrylamide from leaking out from between them while the gel was being cast. Additionally, 4 to 6
bulldog clamps were placed along the edges on both sides to prevent leakage.

**Preparation of the Gel Mix**

A six percent gel was used for polyacrylamide gel electrophoresis, and prepared by mixing 5.7 g of acrylamide (AMRESCO, Solon, OH), 0.3 g of N,N'-methylenebisacrylamide (bis) (Bethesda Research Laboratories, Gaithersburg, MD), 48 g of urea (AMRESCO, Solon, OH), 10 ml of 10X TBE (890 mM Tris-borate, 20 mM disodium EDTA) and 40 ml of deionized water on a magnetic stirrer until the acrylamide and urea went into solution. The volume was adjusted to 100 ml. The acrylamide solution was transferred to a vacuum flask and the solution was degassed.

**Pouring the Sequencing Gel**

Just prior to pouring, 1.0 ml of freshly prepared 10% ammonium persulfate (Bio-Rad, Richmond, CA) and 45 μl of N,N,N',N'-tetramethylethylenediamine (TEMED) was added to the acrylamide mix. The mix was poured between the plates from a plastic bottle with a spout allowing control of the rate at which acrylamide was introduced.
between the plates. When the gel was fully cast, the gel
sandwich was laid down with the long plate on the bottom
and the flat edge of a shark tooth comb was inserted
between the plates to a depth of 1 cm. The gel sandwich
was wrapped with plastic wrap until used, to prevent
desiccation.

Running the Sequencing Gel

After complete polymerization had occurred, the comb
was removed and the well that was formed was cleaned with
deonized water to remove loose acrylamide and urea, and
the comb was reinserted teeth-down. The tape was removed
from the bottom of the gel sandwich and the bulldog
clamps were removed from the sides. The gel sandwich was
placed in the Poker Face model SE1500 Sequencer (Hoefer
Scientific Instruments, San Fransisco, CA) with the small
top plate toward the inside of the instrument, and the
clamps were tightened. Two liters of 1X TBE was poured
between the small plate and the back wall of the
sequencer to form the upper buffer chamber, and 2.0
liters of 1X TBE was poured into the lower buffer
chamber.
Current was set at 60 milliamps using the Hoeffer PS2500 DC Power Supply and the gel was run for approximately 30 minutes to allow the gel to achieve a 45°C temperature before samples were loaded. Prior to loading samples, the wells were cleaned of urea by forcing buffer from a pasteur pipet across them. Samples were boiled for 2 minutes, then loaded in the order T, C, G, A with a blank well between adjacent reaction sets. The power supply was set to constant power at 50 watts and the samples were subjected to electrophoresis for 3.5 hours for a "short" run, 6 hours for a "regular" run and 9 hours for a "long" run.

When electrophoresis was complete, the gel sandwich was removed from the sequencer, dismantaled and the gel remained on one of the plates.

Fixing, Washing, Drying and Autoradiography of the Sequencing Gel

Any unused sections of the gel were trimmed away and the remaining gel, still on one plate was submerged in 2.0 liters of a fixing solution of 5% acetic acid and 5% methyl alcohol in deionized water, for 30 minutes. After
30 minutes, the fixing solution was drained off and replaced with 2.0 liters of deionized water. The gel was then washed in the water for 15 minutes. The water was drained off and the gel was transferred from the plate to a piece of Whatman 3MM paper, covered with plastic wrap and dried under vacuum at 80°C for 45 minutes using a Bio-Rad gel dryer. When dry, the plastic wrap was removed and the gel was placed in a light tight cassette and x-ray film (Kodak XAR) was applied. After exposure, the film was developed according to the manufacturer's instructions.

**Computer Analysis of DNA Sequence Data**

The sequence derived from each gel was entered on the computer (DEC420, Digital Equipment Company) using the University of Wisconsin Genetics Computer Group (UWGCG) program Seqed (23). Comparison of different gel readings covering the same region were made using the program Bestfit. Overlapping regions were consolidated into a consensus sequence using the program Assemble. Identification of open reading frames from sequence data were made using the program Map. The program Codonpreference was used to examine the codon usage in
all possible reading frames and compare those with a
codon frequency table for *Rhodobacter capsulatus* (YDL-1).
The output file from Codonpreference was plotted on a
Hewlett-Packard 7475-A Plotter. This plot would indicate
the likelihood that a particular open reading frame would
be expressed in *R. capsulatus*. 
Results

DNA Sequence of the *R. capsulatus* hemA Gene

The initial goal of this project was to determine the sequence of the hemA gene from the hemA⁺ *Rhodobacter capsulatus* strain SB1003 (56), and to determine the Tn5 insertion point in the hemA mutant, strain AJB529. Previous work indicated that the hemA gene had one endpoint in the 1.4 kb EcoRI fragment and the other endpoint in the 0.6 kb EcoRI fragment (10) (Figure 2, pCAP17); therefore, the sequencing could be limited to these two fragments. It was also known that the Tn5 insertion did not completely eliminate δ-aminolevulinate synthase activity, suggesting that the insertion was either in the promoter or in the carboxyl terminus of the gene (102). Figure 3 depicts the strategy used to determine the DNA sequence of the hemA region. The first primer used hybridized 80 bases from the 3' end of Tn5 (7) and allowed determination of the adjacent *R. capsulatus* DNA sequence. The template for these sequencing experiments was derived from pCAP1 (Figure 2). The plasmid was digested with BamHI and EcoRI, mixed with similarly digested M13mp18 replicative form (RF) DNA and ligated. In M13mp18 the BamHI recognition site is located 5' to the EcoRI site relative to the
Figure 2: Partial Restriction Map of the AJB529 hemA Region, pCAP1 and pCAP17. The 0.4 and 2.3 kb EcoRI fragments in pCAP17 represent unlinked R. capsulatus DNA. B=BamHI, E=EcoRI and all measurements are in kilobases.
Figure 3. Graphical Representation of the Strategy Employed to Obtain DNA Sequence Data of the *R. capsulatus* *hemA* Region.

B=BamHI, C=BclI, E=EcoRI, H=HindIII, N=NsII
universal primer annealing site. This situation forced the orientation of the insert such that the Tn5 sequence was upstream of the *R. capsulatus* sequence.

*Escherichia coli* strain NM522 was transformed with approximately 200 ng of DNA from the ligation mix using the method described by Kushner (48). Recombinant phage yield white plaques in the presence of X-gal and IPTG while wild-type phage plaques are blue due to α-complementation as described in the Introduction. Several white plaques were picked and purified as described previously. RF DNA was isolated by the alkaline lysis method of Birnboim and Doly (11), and scored for the presence of a 3.3 kb *BamHI*-EcoRI insert. One recombinant phage which possessed such an insert was designated M13mpl8.1 (Figure 4).

Template DNA from M13mpl8.1 was isolated as described by Kristensen (47) and sequencing reactions were conducted using the Tn5 primer. The data generated using this primer provided the sequence of bases from 163 through 411 (Figure 5) (base numbers are relative to the 5' end of the sequenced region), identified an open reading frame, transcribed in the same direction as sequence data was
Figure 4. Subclones of the *R. capsulatus* hemA Region used for DNA Sequencing.

B=BamHI, C=BclI, E=EcoRI, H=HindIII, N=NsiI
generated, and located an NsiI restriction enzyme recognition site.

When DNA is digested with NsiI, the resulting cohesive ends are identical to those left when DNA is digested with PstI. Plasmid pCAP1 DNA was digested with both NsiI and EcoRI and ligated with PstI and EcoRI digested M13mp18 RF DNA. NM522 was transformed as before and phage RF DNA from white plaques was scored for the presence of a 416 bp EcoRI-NsiI insert. One phage which possessed the desired insert was designated M13mp18.416+. The universal primer #1212, (USB, Cleveland, OH), which hybridizes with M13 DNA upstream of the multicloning region, was used for the subsequent sequencing reactions. The sequence of nucleotides 337 through 590 was determined. A 74 base overlap with Tn5 primer generated sequence was detected, and the open reading frame was extended 179 bases. Contained within the sequence derived using the M13mp18.416+ template was the sequence 'TGATCA'. This is the recognition sequence of BclI which catalyzes DNA cleavage between the T and G residues and leaves the same cohesive ends as BamHI. The sequence 'GATC' in the middle of the BclI recognition sequence also serves as a site for the E. coli adenine methylation system. The
activity of BclII is severely inhibited by methylated adenine; therefore, in order to use BclII for digestion, it was necessary to transform pCAP17 (Figure 2) into E. coli strain GM2163, which contains a Tn9 insert in the dam gene. Plasmid DNA isolated from this strain lacked methylated adenine and could be digested with BclII. Previous restriction enzyme mapping analysis of pCAP17 revealed the presence of a HindIII recognition site in the 1.9 kb EcoRI fragment (S. Biel, unpublished data). The 1.1 kb BclII-HindIII fragment was excised from pCAP17 and ligated into BamHI and HindIII digested M13mp19 to yield M13mp19.X600. With this template and using the universal primer, the sequence of nucleotides 529 through 768 was determined. The common EcoRI recognition site between the 1.4 kb and 0.6 kb EcoRI fragments was found at 735.

Plasmid pCAP22, which is pSUP202 plus the 0.6 kb EcoRI fragment, was digested with EcoRI and the 0.6 kb fragment was inserted into EcoRI digested M13mp18 RF DNA. Two recombinant phage were purified, each containing the 0.6 kb EcoRI insert; but, in opposite orientations to yield M13mp18.22-1 and M13mp18.22-2. Sequence data from these templates derived by using the universal primer demonstrated that the insert in M13mp18.22-1 was oriented
Table 3. Primers used for Generation of hemA Sequence

<table>
<thead>
<tr>
<th>Name</th>
<th>5' end*</th>
<th>Orientationb</th>
<th>Sequence</th>
<th>Productivityc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal</td>
<td>NA</td>
<td>+/-</td>
<td>GTTTTCCCAGTCACGAC</td>
<td>NA</td>
</tr>
<tr>
<td>TnS within TnS</td>
<td>+</td>
<td>TGTTAGGAGGTCACATG</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>1.4-1+</td>
<td>10</td>
<td>+</td>
<td>ATTAGAACTCCCATT</td>
<td>211</td>
</tr>
<tr>
<td>1.4-1-</td>
<td>275</td>
<td>-</td>
<td>GCACCAGACGGGTGAT</td>
<td>253</td>
</tr>
<tr>
<td>1.4-2-</td>
<td>452</td>
<td>-</td>
<td>GACAAGCGCCGTTCTTG</td>
<td>201</td>
</tr>
<tr>
<td>0.6-1+</td>
<td>830</td>
<td>+</td>
<td>CCGCATCGACATCTT</td>
<td>235</td>
</tr>
<tr>
<td>0.6-2+</td>
<td>982</td>
<td>+</td>
<td>CGCAGGCCTCGATCG</td>
<td>188</td>
</tr>
<tr>
<td>0.6-3+</td>
<td>1264</td>
<td>+</td>
<td>ACCTGAAACAGATCG</td>
<td>118</td>
</tr>
<tr>
<td>0.6-1-</td>
<td>1106</td>
<td>-</td>
<td>GCCATGGTCGATGAT</td>
<td>180</td>
</tr>
<tr>
<td>0.6-2-</td>
<td>949</td>
<td>-</td>
<td>GAGAAGATGAAGCCC</td>
<td>181</td>
</tr>
</tbody>
</table>

* Location of the 5' end of the primer relative to the sequence in Figure 5.

b Orientation of the sequence generated by the specific primer relative to the direction of transcription of the hemA gene. "+" indicates that sequence is generated in the same direction as transcription, while "-" primers generate sequence data from the complementary strand.

c Number of nucleotides sequenced with the specific primer.
properly to continue extension of the putative hemA open reading frame. This combination of primer/template provided the sequence of bases from 735 through 952. A primer, 0.6-1+ (Table 3) was ordered whose 5' end hybridized with M13mp18.22-1 at 830. From this primer, sequence data from 911 through 1,145 was collected, and the putative hemA open reading frame was still open.

Primer 0.6-2+ was ordered to extend the sequence further. As with the previous primers used to generate sequence data from this template, only limited resolution was possible. From the 0.6-2+ primer, the sequence of bases 1,118 through 1,305 was determined. No stop codon was detected for the putative hemA open reading frame; consequently, primer 0.6-3+ was ordered which provided the sequence of bases from 1,286 to the EcoRI cloning site at 1,403. Contained within this newly compiled sequence was the nonsense codon, "TGA", at 1,326 which closed the putative hemA open reading frame.

Although the sequence of a 1,100 base open reading frame in the region of DNA known to contain the hemA gene from R. capsulatus had been amassed, no initiation codon
had been found. To locate the initiation codon, it was necessary to insert the 1.4 kb *EcoRI* fragment from pCAP44 into M13mp18 yielding M13mp18.44-. A primer, 1.4→1- was ordered which hybridized at 275 and would provide sequence data from the complementary strand. From this primer the sequence of bases 252 through 1 was determined. An initiation codon was found for the putative *hemA* open reading frame at base 121. Initiation from this "ATG" results in a 1,206 base pair open reading frame which codes for a protein of approximately 51 kDa. Data obtained using the 1.4→1- primer in concert with the M13mp19.1 template demonstrated the exact location of the Tn5 insertion in the *hemA* mutant AJB529 to be 20 bp into the coding region.

Since the sequence of the putative *hemA* gene had been determined from one strand of DNA, it needed to be confirmed by carrying out experiments employing primer/template combinations which would generate data from the complementary strand. Primer 1.4→1+ which hybridized at base 10 and extended sequence data through the initiation codon was ordered. Data obtained from this primer, annealed to the M13mp18.44+ template confirmed the sequence of bases from 89 through 299. The 1.0 kb *NsiI*-
EcoRI fragment from pCAP44 was subcloned into M13mp18 and the resulting template, M13mp18.44a, was used in conjunction with the universal primer to confirm sequence data from 309 through 200.

To confirm the sequence generated when M13mp18.416+ was used as the template, primer 1.4→2− was ordered and used in conjunction with the M13mp18.44− template in the subsequent sequencing experiment. This primer/template pair verified the nucleotide sequence from 434 through 234. When the universal primer was employed along with M13mp18.44−, the nucleotide sequence data from the EcoRI recognition site at 740 through 418 was corroborated.

M13mp18.22−2 was the only template used to confirm the sequence of the 0.6 kb EcoRI fragment. It is the same insert as that in M13mp18.22−1, but in the opposite orientation. Sequence data obtained using the universal primer confirmed the sequence from 1,403 through 1,039. Primer 0.6→1− was ordered and used to further extend the compiled confirmatory sequence from 1,081 through 902. Finally, Primer 0.6→2− validated the sequence from 915 through 773.
Figure 5. Nucleotide Sequence of the *R. capsulatus* hemA Region.

The coding region for δ-aminolevulinate synthase is from nucleotide 121 through nucleotide 1,326. The coding region of orf-2 is from nucleotide 1,466 through 2,682. The predicted amino acid sequences is printed below the nucleotide sequence. The DNA sequence data presented in lower case is that which has been determined by other members of the laboratory, and has not been confirmed by deriving the sequence from the complementary strand. The amino acid sequence encoded by orf-3 has not been included since there are two obvious frame shift errors in the data. The site of the Tn5 insertion in strain AJB529 is noted by the arrow above the sequence.
GGATGCGACAATTTAGAACTCCCATTTTTTTGGCGATGCGCG
TAAACCTGTGCGCCGACAGGGCCGAAGAGCCGACCCGGC
AAACCAAAACCAGCAGCAAGCCAGGGGTGACGCACATGGAC
YNLALDKAIQKLD
TACAATCTGCGGCTCGACAAAGCGATCCAGAAACTCCACGAC
YNLALDKAIQKLD
GAGGGACGTTACCAGCACGTTACATCGACATCGAACGCGAGAG
EGRYRTFIDIEREK
GGCCGCTTCCCCAAAGGCGCGGATGGAACCAGCCCGATGGGC
GAFPKAQWNRPDDG
GAGGACCATCACCAGTCGTTGCGCCGAAACGACTATCTGGGC
KQDITVWCGNBDYL
ATGGGCGCAGCCAGGCCTGCTGGTCGCGGCGATGCATGAGGCG
MGQHPVVLAAMHEA
CTGGAGAGCCGCGGTGTTGCCGCGGCACCACCAGACATCG
LEAVGAGSGGTRNI
TCGGGCACCCAGCGCTATCAACCAGCCTGATGGAAGCCGAGAATC
SGTATYHRREAEI
CCGATCTGCAAGGAAAGCGCGCTTGTCTTTCTCTCTCG
ADLHGKEAALVFS
Figure 5, continued

463  GCCTATATGCACAAATGACGCAGCTCTCGACGCTGCGGCTG  504  A Y I A N D A T L S T L R L

505  CTTTTCCGCCGCTTGCATCATCTATCTTCCGACAGCTGAACCAC  546  L F P G L I I Y S D S L N H

547  GCCTCGATGATGAGGGGATCAAGCGCAATGCCGGGCCGAAG  588  A S M I E G I K R N A G P K

589  CGGATCTTCCGTCACAATGACGTCGCCCATCTGCGCGAGCTG  630  R I F R H N D V A H L R E L

631  ATCGCCGCTGATGATCGCCGCTGCCAGCTGATCGCCCTTC  672  I A A D D P A A P K L I A F

673  GAATCGGTCTATTCGATGGATGGCGACTTCGGCCCGATCAAG  714  E S V Y S M D G D F G P I K

715  GAAATCTGCGACATCGCCGATGAATTCGGCGCGCTGACCTAT  756  E I C D I A D E F G A L T Y

757  ATCGACGAAATCTGCGACATCGCCGATGAATTCGGCGCGCTGACCTAT  798  I D E V H A V G M Y G P R G

799  GCGGGCGTGGCCGAGCGTGACGGTCTGATGCACCGCATCGAC  840  A G V A E R D G L M H R I D

841  ATCTTCAACGGCAACGCTGGCGGAAGCGCTATGCGTCTTCCGCC  882  I F N G T L A K A Y G V F G

883  GGCTACATCGCCGCTTCCGCGGAAGATGCTGATGCTCGCGCGC  924  G Y I A A S A K M V D A V R
Figure 5, continued

925  TCCATATGGCCGCGGCTTCATTTCTGACCTCGCTGCCGCCG
     SYAPGFIFSSTSLPP

967  GCGATCGCCGCTGGCCAGCAGGCCTCGATCGCGTTTTGAAA
     AIAAGAQAASIAFLK

1008 GCGATCGCCGCTGGCGCGCAGGCCTCGATCGCGTTTTGAAA
     AIAAGAQAASIAFLK

1050 ACCGCCGAAGGGCAAGCTGCGCCGACGCGCAACAGATGCACT
     TAEGQKLRDAQQMH

1092 GCGAAGGTGCTGAAATGCACGGCTCAAGGCCTGGGATGCGCG
     AKVLKMRLKALGMP

1134 ATCATCGACCATTGCCACGCAATCGTCCGGTGATCGCTCGT
     IIDHGSHTPVVIG

1176 GACCCCGTGCACACCAAGGCCTGGATGCTTCTCGTCT
     DPVHTKAHSVDMLLS

1218 GATTACGGCGTTTACGTCAGCCGACGATCAAATCCTCCCGACG
     DYGVYYVQPINFTPV

1260 CCGCGCGGCCACCGAAGGGCTGCCCTTCACCCCCACCGCGGGTG
     PRGTERLRFTPSPV

1302 CATGACCTGAAACAGATCGACCGGCTGGTGTCATGCCCATGGA
     HDLKQIDGLVHAMD

1344 CTGCTCTGGGCGGCGTCTGGGATACTCGCGCCGAGGCTCTG
     LLWARCA*

1386 CCTGAGCCATTCTGCGGATGCACCGCCCTCGGGATTATGCTA

1428 GGGATTTCGTAAGAAATTCGGGACCAGACAGTTTCGAAATCGGAC
**Figure 5, continued**

1429  AAAACCACACCGAGGGCCGGACAGGCAGATGATGAGTTGAAAG  1470  M I R W K

1471  GGCAGAACCTCTGACGAGGGCGGGACAGGCAGATGATGAGTTT  1512  G K P S T Q V E A A Q K G F

1513  GACGACTTCGACATGAAAGCTTTGCCGACATGCTTCGCGGAA  1554  D D F D M K L G M L R G E

1555  CGTGCGGACCCTGTCAATACTGCTTCTTATGTGCAAGCGCGAG  1596  R A T L S K S L D V Q R E

1597  CTGCGCATCAAGGCGCTCTATATATTGCCGCTACGAAATTTCG  1638  L R I K A V Y I A A I E N C

1639  GACCCCCAGCCTTTCGAAAGCCCTCCTTCTCCCGCGGAGTTTC  1680  D P S A F E S P S F I A G F

1681  GTGCGGTCTTATGCGCGCCTGTACGCCGGCCTGCGGACTGGG  1722  V R S Y A R Y L G L D P D W

1723  GCCCTTTGAACGCTTCTGCAAGGAAATCGGGGCTTTACCCCAACG  1764  A F D R F C K E S G F T P T

1765  CATGGCCTGGCCGCTGCCGCTGGCCGCTGGCCGCAAAGCAACCGCGG  1806  H G L A A A A S G P K A P R

1807  CGCACCCGCGATCTGGGCAATCGCTGCTGGCCAATCCAATGCC  1848  R T A D L A S S L A N P N A

1849  TCCTTCATCCCGAAGACAGGCGCCTGCTGGACCCAAATCGAC  1890  S F I P K Q A P I W T K I D
Figure 5, continued

1891 CCGCGCGCGGGTGGGTTCGCTGGCGCGCGTCGGTGGTATGGTGTCG 1932
P R A V G S L A V V V M V A

1933 GGCAGCATCGGCTATGGCGCCCTGGTGCTGGTGGTGCAGGAAGTG 1974
G G I G Y G A W S V L Q E V

1975 CAGCCGCCGCTGACGCTGACCGCCTGGTATGGTGTCG 2016
Q R V Q L T P V E Q A P G V

2017 GTGGCCCGAGATTGACCGCCTGCAGCCCGTCGCCCCCCGGTGAGG 2058
V A E I D P L Q P V A P V E

2059 ATGCCCTCGGCGCCCGATGTCCTGCCCCCGCCGACGTTGCTGCC 2100
M A S A A D V L P A D V L P

2101 GATGCCGCCACCGGAGAGGCTTTGAGGCCCCTGAAAACTGCCGCC 2142
D A A T E E A F D R L Y R P

2143 GCCGCGCTGGAAAGCCCGGCTGATCGGCAGCGTCGCTGGCGAG 2184
A A L E T P V L I A R D G P

2185 ATGCGGCCGATCGACCGCAGCGCGCCGATGGTTCTGGGAGG 2226
I A A I D P R E A G V L P G

2227 ACGGCAGCGCGATGACCAGCCCGCGAGGCTGGTCTGGCGCACGG 2268
T A Q A M T S A P M V L A E

2269 GCCGCGCGCGACGGCGGCACCTGGGACGGTCTGCTGGCGACGGTCTGGC 2310
A G P T P A E P G Q L P V V

2311 CATGCCTGGGCGCAGCGGAGGTCGAGGTGCTGCCGACGGTCTGGCG 2352
H A L G A E P A E V E V L A
Figure 5, continued

2353  GTGCGGCCGTCCAGTGGGTGCAAATCGCTCCGACCGGTGTCG  
V R P S W V R I T A S D G S  2394

2395  GTGATCTTCGAAAAGGTGATGAGTGCAGCGAGGCGGAGTCTGCCC  
V I F E K V M D A G E R F A  2436

2437  GTGCGGAAGCTGGAGAAGCGCCCGGAAGCTGCGCACCAGGGAA  
V P K L E E P P K L R T G E  2478

2479  TCGGGCGCAGATCTCTTTGCGGTGAACGGCGTCGCGCATGGG  
S G A I Y F A V N G V A H G  2520

2521  CCGGTGGGCTCAGAGTGGCGGCTATGCCGTGCGCATG  
P V G S K G A V T K N V E L  2562

2563  TCGGCCGCTGGGGTTGGCCAGGCGCTATGCCGTGCGCATG  
S A Q A L G T A Y A V A D M  2604

2605  GCCGCGGATGCGGAGCTTGGCCAGATGGGCGCGGTGGCCGAT  
A A D G D L A K M V A V A D  2646

2647  GCCCTCTCGGCTGTGCCGGGAAGCCCGAGGCACTAAACAGCTC  
A S S V V P G T P E Q *  2688

2689  TTTGAAAGCCAGCAGCAGGCGGCGGTGGCCCTTTGGCGCGGCTG  
2730

2731  ACCTATTTGTGGGCGAGTGCAACGAGGCCTTCCCATTGTGCG  
2772

2773  TCAATCCCCATCGGCTGGGAGGACGACACAGCCGCGATG  
2814

2815  cgcggcaggatctgggctggcagattgctggcgggcgccagt  2856
Figure 5, continued

2857  cgccgatctcggtgcagacgatgaccaacaccatcaccCTCGG 2898
2899  ATGTCAACCGCAGCGCTGGGGCAGGTGCTGCGGGCGGCCGATG 2940
2941  TGGGCGCCGACATCGGCCGTTCTCGGTCCCGGATGGAGGATG 2982
2983  CGACCCGGCCACTGAAGGAAATCTGCCGCCAAAGCCCCGGTGC 3024
3025  CGATCGTGCCGCACATCCATTTCCACTACAAGCCCGGCGATCG 3066
3067  AGGCCGAGGGGCCGCCCGCTGCTGCGCATCAACCCGGGCAA 3108
3109  TATCGGCGATGCGGCCCAGCGTGCGTGAGGTGATCAAGGCCG 3150
3151  CCAAGGATCAGGCCGCTGCGATGCGGGGTCAATGCCG 3192
3193  GGTCGCTGGAAAAGCACCTTCTGGACAAATACGGCGAGCCCT 3234
3235  GCCCGGAGGCGATGCTCGAAAGCGGTCTCGATCATATCAAGA 3276
3277  TCCTGCAGGACAACGATTTCCACGAATTC 3304
Confirmation That the 1,206 Base Pair Open Reading Frame is hemA

The DNA sequence data was examined using the computer program CODONPREFERENCE (23) (Figure 6) which analyzes codon usage for all reading frames, and compares it to a codon frequency table for Rhodobacter capsulatus derived from the data provided by Youvan and Marrs (105). The only reading frame which was open for an extended length and bore a favorable codon preference profile was the 1,206 base pair open reading frame. This alone was suggestive evidence that this open reading frame would be translated in R. capsulatus.

To study transcriptional regulation of the R. capsulatus hemA gene, Wright et al. (102) inserted the 1.4 kb EcoRI fragment from pCAP44 into pNM481. If the coding region for δ-aminolevulinate synthase corresponded to the 1,206 base pair open reading frame, a fusion protein of 142,000 should be formed between the N-terminal portion of δ-aminolevulinate synthase and β-galactosidase. A fusion protein of 147,000 was observed, which is in very close agreement with the
Figure 6. Codon Preference Plot of the Region Known to Contain the \textit{hemA} Gene.

The codon preference profile for \textit{hemA} is reflected in panel 1.

Open reading frames are shown as an open box beneath plot in the panel. Rare codons are shown as vertical hatch marks in the box below each panel. The numbering system along the top and bottom of the figure are relative to the DNA sequence presented in Figure 5. The dashed line in each panel represents the rare codon threshold, a plot above which signifies a region of DNA in that reading frame with few rare codons.
estimated molecular weight, and B-galactosidase activity was detected in the *R. capsulatus* strain harboring this plasmid. The fact that a fusion protein of the expected size was observed proves translation of the 1,206 base pair open reading frame.

The sequence of the *hemA* gene from *R. capsulatus* was independently derived by Hornberger et al. (36). The sequence data that they reported is identical to that presented here. Finally, as detailed in the Discussion section, the predicted amino acid sequence of the 1,206 base pair open reading frame was found to be between 69% and 80% similar to that of δ-aminolevulinate synthases from other organisms.

**DNA Sequence Analysis of the 1.9 kb EcoRI Fragment**

Analysis of the sequence data downstream of the *hemA* gene revealed the existence of another open reading frame beginning at 1,362. This is 36 bases downstream of the stop codon for *hemA*. This second open reading frame remained open to the EcoRI recognition site at 1,403. The chloramphenicol acetyl transferase (cat)
cartridge from pCM7 (20) was inserted into the HindIII site located within the 1.9 kb EcoRI fragment (M. Wright, unpublished data). Assay of cat activity upon introduction of this plasmid into R. capsulatus demonstrated that oxygen regulated transcription does occur downstream of hemA.

The 1.9 kb EcoRI fragment from pCAP27 was inserted into M13mp18 RF DNA which had been digested with EcoRI. Two plaques, upon purification, yielded RF DNA bearing the 1.9 kb EcoRI insert; one in each orientation. These recombinant phage were designated M13mp18-1.9+ and M13mp18-1.9-, where the '+' indicated the clone from which DNA sequence data would extend the open reading frame, and the '-' template would generate data from the complementary strand. Sequence data obtained using the universal primer in conjunction with the 1.9+ template extended the open reading frame 335 bases from 1,397 through 1,737. The HindIII recognition sequence used for cloning the cat cartridge was found to be at 1,533.

Two primers, complementary in sequence, were ordered (Table 4). Primer 1.9→1+ extended the sequence
data from 1,638 through 1,977 while primer 1.9→1- confirmed bases 1,588 through 1,397. Again, two primers were ordered, one to extend the sequence data, the other to confirm the data previously obtained. Primer 1.9→2+ extended the compiled sequence of the open reading frame from 1,946 through 2,275. With the 1.9→3- primer, the sequence from 1,882 through 1,585 was verified.

Primers 1.9→3+ and 1.9→4+ failed to generate sequence data. It is assumed that strong secondary structure in the region which these primers should hybridize either prevented hybridization, or inhibited the polymerase activity. Primer 1.9→5+, on the other hand, allowed determination of the base sequence from 2,108 through 2,484, extending the open reading frame 209 bases. From the 1.9- template, using the 1.9→4- primer, the sequence of bases from 2,163 through 1,866 was confirmed.

Finally, with primer 1.9→6+ the sequence was extended from 2,466 through 2,789, and the sequence 'TAA' was found at 2,682 which closed the open reading frame. Primer 1.9→6- was used to confirm the sequence
<table>
<thead>
<tr>
<th>Name</th>
<th>5' end</th>
<th>Orientation</th>
<th>Sequence</th>
<th>Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universal</td>
<td>NA</td>
<td>+/-</td>
<td>GTTTTCCCAGTCACGAC</td>
<td>NA</td>
</tr>
<tr>
<td>1.9-1+</td>
<td>1615</td>
<td>+</td>
<td>ATATTGCGGCCATCG</td>
<td>340</td>
</tr>
<tr>
<td>1.9-2+</td>
<td>1937</td>
<td>+</td>
<td>GCATCGGCTATGGCG</td>
<td>330</td>
</tr>
<tr>
<td>1.9-3+</td>
<td>2206</td>
<td>+</td>
<td>GAGGCAGGTGCTGCTG</td>
<td>None</td>
</tr>
<tr>
<td>1.9-4+</td>
<td>2225</td>
<td>+</td>
<td>GCACGTCGAGCGCAG</td>
<td>None</td>
</tr>
<tr>
<td>1.9-5+</td>
<td>2080</td>
<td>+</td>
<td>CTGCCCGCCGACGTG</td>
<td>377</td>
</tr>
<tr>
<td>1.9-6+</td>
<td>2443</td>
<td>+</td>
<td>AAGCTGGAAGAGCAGC</td>
<td>324</td>
</tr>
<tr>
<td>1.9-1-</td>
<td>1630</td>
<td>-</td>
<td>CGATGGCGGCAATAT</td>
<td>192</td>
</tr>
<tr>
<td>1.9-2-</td>
<td>3143</td>
<td>-</td>
<td>GATCACCTCACGCAC</td>
<td>213</td>
</tr>
<tr>
<td>1.9-3-</td>
<td>2951</td>
<td>-</td>
<td>CGCCATAGCCGATGC</td>
<td>298</td>
</tr>
<tr>
<td>1.9-4-</td>
<td>3220</td>
<td>-</td>
<td>CAGCACACCTGCCTC</td>
<td>298</td>
</tr>
<tr>
<td>1.9-5-</td>
<td>3239</td>
<td>-</td>
<td>TCGCCTGCGACGTGC</td>
<td>None</td>
</tr>
<tr>
<td>1.9-6-</td>
<td>3457</td>
<td>-</td>
<td>CGGCTCTTCCAGCTT</td>
<td>239</td>
</tr>
<tr>
<td>1.9-7-</td>
<td>3744</td>
<td>-</td>
<td>CATTCAGCTGCCCA</td>
<td>386</td>
</tr>
</tbody>
</table>

* Location of the 5' end of the primer relative to the sequence in Figure 5.  
* Orientation of the sequence generated by the specific primer relative to the direction of transcription.  
* Number of nucleotides sequenced with the specific primer.
Figure 7. Codon Preference Plot for the Region Encompassing orf-2.
The codon preference for orf-2 is shown in panel 1.

Open reading frames are shown as an open box beneath plot in the panel. Rare codons are shown as vertical hatch marks in the box below each panel. The numbering system along the top and bottom of the figure are relative to the DNA sequence presented in Figure 5. The dashed line in each panel represents the rare codon threshold, a plot above which signifies a region of DNA in that reading frame with few rare codons. The small vertical hatch marks below the figure point out, from left to right, the position of the putative start codon for orf-2 and the HindIII recognition site, respectively.
from 2,365 through 2,127, and primer 1.9-7 verified the base sequence from 2,669 through 2,284.

Translation of orf-2

Sequence data from the region downstream of hemA was examined using the program CODONPREFERENCE (Figure 7). The plot suggests that orf-2 is translated, however, translation probably does not initiate at the first 'ATG' after hemA. Instead, translation appears to initiate within the 1.9 kb EcoRI fragment. There is an 'ATG' codon at position 1,466, within the open reading frame, which appears to be the site of translation initiation. This will result in an open reading frame of 1,239 base pairs coding for a 412 amino acid product. The molecular weight of the polypeptide encoded by orf-2, based on the predicted amino acid sequence is 50,671.

Attempts to construct a translational fusion between orf-2 and lacZ (pCAP77 and pCAP84) have not been successful. It appears that AJB529 will not function as a recipient strain in this mating system.

Attempts to Identify orf-2

Other members of the lab had previously attempted to determine if additional hem genes resided on pCAP17 by introducing the plasmid into R. capsulatus strain PAS100 and looking for a gene dosage affect by assaying
the tetrapyrrole biosynthetic enzymes (R. Cardin and A. Biel, unpublished data). When extracts from PAS100 (pCAP17) were incubated with δ-aminolevulinic acid, the rate of protoporphyrin synthesis was twice that of the parental strain, indicating that there may be a second \textit{hem} gene on pCAP17. When the same extracts were used to assay for δ-aminolevulinate synthase, the specific activity for PAS100 was found to be 21.0 nmoles of δ-aminolevulinic acid formed/hr/mg protein, whereas that for PAS100 (pCAP17) was 42.2. This was not unexpected since pCAP17 was known to possess the \textit{hemA} gene.

Porphobilinogen synthase activity was measured for each of the extracts and found to be 223 nmoles of porphobilinogen formed/hr/mg protein for PAS100, and 287 for PAS100 (pCAP17). This is not a significant difference, indicating that the second \textit{hem} gene may not be \textit{hemB}. Similarly, the activity of the product of the \textit{hemC} gene, porphobilinogen deaminase, was measured and found to be the same for both PAS100 and PAS100 (pCAP17): 1,983 and 1,781 nmoles of uroporphyrin formed/hr/mg protein, respectively. Protoporphyrinogen oxidase activity, encoded by the \textit{hemG} gene, was found to be the same in both cases as well: 54.6 and 58.7 nmoles
of protoporphyrin formed/hr/mg protein for PAS100 and
PAS100 (pCAP17), respectively.

I repeated the assays for porphobilinogen synthase
and porphobilinogen deaminase; however, instead of
PAS100 as the control, PAS100 (pCAP42) was used.
Plasmid pCAP42 is the same as pCAP17, but, the 1.9 kb
EcoRI fragment has been deleted (S. Biel, unpublished
data). Again, no significant differences were observed
when the activities of the products of hemB and hemC
were compared between the extracts from cells possessing
pCAP17 and the control. Attempts were also made to
measure the activity of coproporphyrinogen oxidase, the
product of the hemF gene, but, for reasons which will be
addressed in the Discussion, no meaningful results were
obtained.

In an effort to increase the synthesis of the
product of orf-2, the 4.0 kb BglII fragment from pCAP41,
containing all of hemA and orf-2, was cloned into BamHI
digested pUC8, downstream of the lac promoter. An
extract of E. coli cells harboring this plasmid, pCAP80,
was used to assay the product of the hemF gene. Again,
the coproporphyrinogen oxidase assay yielded unclear results. One more clone was constructed in an effort to boost synthesis of the product of orf-2. The 1.9 kb EcoRI fragment from pCAP27 was cloned downstream of the tac promoter on the expression vector pKK223-3 (16), yielding pCAP83 (A. Biel, unpublished data). The activities of porphobilinogen synthase and porphobilinogen deaminase were determined to be nearly the same between MC1061 (pKK223-3) and MC1061 (pCAP83). Had the product of orf-2 been either of these two enzymes, one would expect a marked increase in the activity of one of them.

These results strongly suggest that orf-2 is not hemB, hemC, or hemG. However, the possibility still exists that orf-2 may be hemD, hemE, or hemF. Further work will be necessary to determine whether orf-2 is in fact one of these genes.

Since enzyme assays did not yield the identity of orf-2, attempts were made to identify this open reading frame by the phenotype of a mutant, deficient for the product of orf-2. The kanamycin resistance cartridge
from pUC-4K was cloned into the Smal site within orf-2 in pCAP27. The resulting plasmid construct, pCAP81, will not be maintained in *R. capsulatus*, thus, in order for an exconjugant to exhibit kanamycin resistance, there must be a recombination of the kan cartridge into the *R. capsulatus* chromosome. Several attempts at deriving a mutant strain of *R. capsulatus* using pCAP81 did not yield any kanamycin resistant exconjugants.

**Analysis of the DNA Sequence Data Downstream from orf-2**

The DNA sequence of the entire 1.9 kb EcoRI fragment has been determined (Figure 5). Careful examination of this data revealed the existence of a third open reading frame beginning 84 bases downstream of the termination codon for orf-2. This third open reading frame, designated orf-3, extends 539 bases to the EcoRI restriction site used for cloning. Undoubtedly, orf-3 extends further into the adjacent 1.7 kb EcoRI fragment. Figure 8 presents a plot of the output of the program CODONPREFERENCE used to analyze the DNA sequence of the *hemA* region. There are obviously two errors in the sequence data obtained for the orf-3 region. This is apparent from the dramatic
drop in the codon preference profile in panel 3, and the sharp rise in panel 2 at the same point in the sequence. It is clear that orf-3 is likely to be translated in R. capsulatus based on the results of this analysis.
Figure 8. Codon Preference Plot for the Entire \textit{hemA}
Region.

The codon preference profiles of \textit{hemA} and orf-2 are reflected in panel 1. That for orf-3 starts in panel 3.

Open reading frames are shown as an open box beneath the plot in the panel. Rare codons are shown as vertical hatch marks in the box below each panel. The numbering system along the top and bottom of the figure are relative to the DNA sequence presented in Figure 5. The dashed line in each panel represents the rare codon threshold, a plot above which signifies a region of DNA in that reading frame with few rare codons. The small vertical hatch marks below the figure point out, from left to right, the position of the putative start codon for orf-2 and the \textit{HindIII} recognition site, respectively.
Discussion

Due to their central role in energy production, tetrapyrroles rate as one of the most important classes of molecules on earth. Consequently, a great deal of research has been conducted on the subject of tetrapyrrole biosynthesis. The purple, non-sulfur photosynthetic bacteria have proven to be a model system for this endeavor since they produce the four main tetrapyrroles: heme, bacteriochlorophyll, siroheme and vitamin B-12. The major study by Cohen-Bazire et al. (21) in 1957 is viewed as the landmark work on the regulation of tetrapyrrole biosynthesis. Their research clearly demonstrated a link between environmental factors such as oxygen tension and light intensity, and the suppression of photosynthetic membrane and photopigment formation.

Since that study, considerable effort has gone toward identifying the major point of oxygen regulation of tetrapyrrole biosynthesis. Most of that effort has concentrated on the first step in the tetrapyrrole biosynthetic pathway, the condensation of glycine with succinyl-CoA to form δ-aminolevulinic acid. It was widely
believed that transcription of the hemA gene, encoding δ-aminolevulinate synthase, was the primary target of the regulation.

As a first step in elucidating what, if any, effect oxygen may have on expression of the hemA gene in Rhodobacter capsulatus, Wright et al. (102) isolated a δ-aminolevulanic acid dependent mutant strain designated AJB529. This mutant was isolated via Tn5 mutagenesis as described by Simon et al. (82). AJB529 was found to possess ~10% of the normal δ-aminolevulinate synthase activity. However, no mutant was obtained which completely lacked activity. The question of why this was the case will be addressed in detail shortly; but, two possible explanations were put forth. First, as in R. sphaeroides, there may be a second gene encoding δ-aminolevulinate synthase. Second, the Tn5 may have inserted in either the beginning of the gene, perhaps in the promoter region, or in the distal portion of the gene resulting in a truncated, partially active form of the enzyme. A cosmid containing an insert which complemented this mutation was isolated and designated pCAP17 (10). Through deletion analysis using pCAP17, the hemA gene was found to lie completely on two EcoRI fragments, one of 1.4
kb, and the other 0.6 kb. From southern hybridization analysis, the Tn5 was shown to have inserted into the 1.4 kb EcoRI fragment (10).

In order to facilitate further study of regulation of the cloned hemA gene, it was necessary that its DNA sequence be determined. The dideoxynucleotide chain termination method of Sanger et al. (74) was chosen, and template DNA was prepared from recombinant M13 phage. The amassed DNA sequence data from this research is presented in Figure 5. A 1,206 base pair open reading frame was found which spanned the EcoRI site in common between the 1.4 kb EcoRI fragment and the 0.6 kb EcoRI fragment. The 'ATG' start codon was found on the 1.4 kb EcoRI fragment at position 121 (Figure 5). The coding region for δ-aminolevulinate synthase extends into the 0.6 kb EcoRI fragment, and terminates at position 1,326 with the nonsense codon 'TGA'.

The sequence data was used to construct a hemA-lacZ translational fusion at the central EcoRI site at position 740 (103). If this open reading frame does correspond to the hemA gene, then by cloning the 1.4 kb EcoRI fragment into EcoRI digested pNM481, an in frame fusion between
hemA and lacZ would be made. This construct was made, and upon its introduction into *R. capsulatus* strain PAS100 (Lac), β-galactosidase activity was conferred on the strain. This proved that the 1,206 base pair open reading frame is translated in *R. capsulatus*. Moreover, when the molecular weight of the δ-aminolevulinate synthase-β-galactosidase fusion protein was determined by subjecting cell extracts to polyacrylamide gel electrophoresis, it was in very close agreement with the estimated molecular weight based on the sequence data (103). This suggests that translation does initiate with the 'ATG' at position 121.

When *R. capsulatus* strain PAS100 carrying the fusion plasmid was incubated under varying oxygen tension, and β-galactosidase activity was measured as an indicator of transcription, β-galactosidase activity was found to be only 2-fold higher under low (3%) oxygen tension than when incubated at high (23%) oxygen tension (103). This indicates that, while there is some small effect on transcription exerted by oxygen, it is not at the magnitude necessary to account for the dramatic inhibition of bacteriochlorophyll synthesis under high oxygen. These
results were subsequently corroborated by Hornberger et al. (37).

A codon frequency table for *R. capsulatus* was constructed from data presented by Youvan and Marrs (105) and used in conjunction with the computer program CODONPREFERENCE (23) to examine the sequence of the 1,206 base pair open reading frame. The resulting codon preference plot is shown in Figure 6. It is clear that the 1,206 base pair open reading frame is the only open reading frame in this region that is likely to be translated in *R. capsulatus*. In fact, as indicated by the black hatch marks below the plot, there are only 6 rare codons used in the entire 1,206 base pair open reading frame as compared with the other two possible reading frames. Note, also, the number of rare codons in the same panel in the regions bordering the 1,206 base pair open reading frame. This is convincing evidence that the coding region for δ-aminolevulinate synthase does correspond to the 1,206 base pair open reading frame. In addition, the DNA sequence data was used to select regions for PCR primers to clone *hemA* into the expression vector pKK223-2 (16). To the 5'-end of the upstream primer, a unique *BamHI* recognition sequence was incorporated. A
unique HindIII recognition sequence was added on to the 5'-end of the downstream primer. When the clone was transformed into *E. coli* strain S905 (hemA41), δ-aminolevulinic acid independence was conferred to the strain. This confirmed that the 1,206 open reading frame does encode δ-aminolevulinate synthase.

This open reading frame codes for a 401 amino acid polypeptide with a molecular weight of 50,491. This is in good agreement with the estimated molecular weight reported by Warnick and Burnham of 57,000 for δ-aminolevulinate synthase from *R. sphaeroides* (97). Hornberger et al. (36) have also reported the DNA sequence of the *R. capsulatus* hemA gene, and it is identical to that which I had obtained.

Figure 9 shows the alignment of the *R. capsulatus* δ-aminolevulinate synthase with the δ-aminolevulinate synthases from *Bradyrhizobium japonicum* (58), *Saccharomyces cerevisiae* (94), rat (104) and human (5). In Genbank, the product of the *Bacillus subtilis* hemA gene is identified as a δ-aminolevulinate synthase, therefore, it was included in the alignment program. Obviously, this sequence is not homologous to the δ-aminolevulinate
synthases. Petricek et al. (67) indicates that this hemA gene in fact encodes glutamyl-tRNA reductase. When the B. subtilis glutamyl-tRNA reductase sequence is removed, 5 of the 8 gaps are no longer necessary to achieve optimal alignment between the various δ-aminolevulinate synthases. When compared with the other hemA genes, the R. capsulatus sequence is between 60% and 72% identical. The predicted amino acid sequence of the R. capsulatus δ-aminolevulinate synthase is from 69% to 73% similar, and 49-55% identical.

δ-Aminolevulinate synthase requires the cofactor pyridoxal phosphate for its catalytic activity. The function of the pyridoxal phosphate is to form a Schiff base intermediate with the amino group of glycine in order to position the reactant properly in the active site of the enzyme, as well as act as an electron sink during the subsequent reaction (83). Upon formation of the Schiff base intermediate, the glycine is decarboxylated, followed by condensation with succinyl-CoA which is also present in the active site (83). The resulting δ-aminolevulinic acid-pyridoximine phosphate is hydrolyzed to yield free δ-aminolevulinic acid and pyridoxal phosphate.
Figure 9. Alignment of δ-Aminolevulinate Synthases and Derived Consensus δ-Aminolevulinate Synthase Sequence.

Rcap= *Rhodobacter capsulatus*, Brady= *Bradyrhizobium japonicum* (accession # M16751), Yeast= *Saccharomyces cerevisiae* (accession # M26329), Bacil= *Bacillus subtilis* (accession # M32130), Cons= Consensus δ-aminolevulinate synthase sequence. The numbering is in reference to the *R. capsulatus* δ-aminolevulinate synthase sequence. The boxed sequence in the conserved sequence is the proposed pyridoxal phosphate binding site consensus sequence. The boxed lysine is the lysine which will form a Schiff base with the pyridoxal phosphate in the absence of glycine.
1  
Rcap  mDYnlaLDKA IqklhdEgRY ......RtFi dIEReKgaFP  
Brady  mDYsqFFnSA ldRlhtErrY ......RVFa DleRmAGrFP  
Yeast  FDYEgLdIsE qkKrDkSY ......RyFn nINRlAKeFP  
Human  FqYDRFFDDE IDEKKnDhtY ......RVFK tVNRRaQlFP  
Rat  FqYDhFFEKK IDEKKnDhtY ......RVFK tVNRRaQlFP  
Bacil  .tCNrTeiyA VvDqlhtgRY yikkfladWf qlsKeelspf  
Cons  FDYDRFF-KA IDEK-ND-RY ------RVFK --NR-AGlFP  

34  
Rcap  KA.gWnrpdg gKqdtItVWcG NDYLGmGqHP v...VlaAmH  
Brady  hA.iWhspk. gKrdVvIWcS NDylAsKHP k...VvGAmv  
Yeast  LA.hRqr... eadKvTvWCS NDylAskHP evld...AmH  
Human  MADDYsdsLI tKKqVsVWcS NDYlgMsrHP r...VCGAmv  
Rat  MADDYtdsLI tKKqVsVWcS NDYlgMsrHP r...VCGAmv  
Bacil  L.t.Fyesdaa vehifrVaCg lDsMbiGerq ilgQVrdsfk  
Cons  -ADDY---LI -KK-V-VWcS NDYLGM-HP ----VCGAm-  

70  
Rcap  EaleavGAGs GGTRNISGtt aYHrREAEi A DLHGEKAAAL  
Brady  ETatrvGtGA GGTRNiAGt hplVqLEAEi A DLHGEKAAAL  
Yeast  kTidkyGcGA GGTRNiAGhN ipl1nLEAEi A tlhlKKeGAl  
Human  DTlKqHgAGA GGTRNISGTS KFHvDErEL A DLHgKDAAL  
Rat  ETVqKqHAGA GGTRNISGTS KFHvLEQEL A DLHgKDAAL  
Bacil  taqkteitiGt ifnelfkgqav tvgrkhAE. tDigsnavsv  
Cons  ET--Q-GAGA GGTRNISGTS KFHv-LEAEi A DLHgKDAAL  

110  
Rcap  vFSSaYIAND ..aTLStLrl LFPGliIYSD SlnHAsMIEG  
Brady  LFtSgYVsNq ...tgiaTiAK LiPnC1lISD eLNHnSmIEG  
Yeast  vFSSCVyVAND avL...SLlgq KMkd1vIFSD eLNHAsMIVG  
Human  LFSSCFVAND ..SLTFLAK MMPGCEIYSD SgNHAsMIQG  
Rat  LFSSCFVAND ..STLFtLAK MMPGCEIYSD SgNHAsMIQG  
Bacil  sYaavelAkK ifgnLSskhi Li1Gagkmge 1aenlhgQg  
Cons  LFSSCVYVAND --STLFtLAK MMPGC-EIYSD SlnHAsMIQG  

148  
Rcap  IKrnrqPKrI FRHNDVhHLR ELiaadDPaa PKLIAFEsvY  
Brady  IRqSgcqerqV FRHNDladLe aLLKaagAr PKLIAcEsvY  
Yeast  IKhAnVKhI FkHNDInELe qLLQsypkSV PKLIAFEsvY  
Human  IRNRSVPKkyI FRHNDvShLR ELLQrDSLPS PK1vAFetVh  
Rat  IRNRSVPKkyI FRHNDvHlr ELLQrDSLPS1 PK1vAFetVh  
Bacil  IgkvtVinrY ......lkaK ELAdRfsgae rsLrnqLESaL  
Cons  IRNRSVPK--I FRHNDV-HLR ELLQR-DPS- PKLIAFEsvY
### Figure 9, continued

<table>
<thead>
<tr>
<th></th>
<th>SMDGdfgPik</th>
<th>EICDlADEFG</th>
<th>ALTYIDEVH.</th>
<th>......AVGMY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rcap</td>
<td>SMDGdVaPLa</td>
<td>kICDlAEkYn</td>
<td>AMTYVDEVH.</td>
<td>......AVGMY</td>
</tr>
<tr>
<td>Brady</td>
<td>SMDGdVaPle</td>
<td>E1CDvAhEFG</td>
<td>A1TFVDEVH.</td>
<td>......AVGLY</td>
</tr>
<tr>
<td>Yeast</td>
<td>SMDGAVcPLE</td>
<td>E1CdfAhEFG</td>
<td>A1TFVDEVH.</td>
<td>......AVGLY</td>
</tr>
<tr>
<td>Human</td>
<td>SMDGAVcPLE</td>
<td>E1CdfAhEFG</td>
<td>A1TFVDEVH.</td>
<td>......AVGLY</td>
</tr>
<tr>
<td>Rat</td>
<td>SMDGAVcPLE</td>
<td>E1CdfAhEFG</td>
<td>A1TFVDEVH.</td>
<td>......AVGLY</td>
</tr>
<tr>
<td>bacil</td>
<td>aeadilisst</td>
<td>gasEfYvskm</td>
<td>mMenanklrk</td>
<td>grplfMvdia</td>
</tr>
<tr>
<td>Cons</td>
<td>SMDG-V-PLE</td>
<td>EICD-ADEFG</td>
<td>A-TFVDEVH-</td>
<td>-----AVGLY</td>
</tr>
</tbody>
</table>

|    | GPRGaGVAER | D. ............. | ............ | GlMhR.I |
| Brady | GPRGGIAER | D. ............. | ............ | GVMhR.I |
| Yeast | GPRGaGVAER | cDEFshrasg iatptndkg gaktVMdR.V |
| Human | GPRGGGIAER | .D. ............. | ............ | GVMpk.m |
| Rat | GPRGGGIAER | .D. ............. | ............ | GVMpk.m |
| Bacil | vPRdldpLan | .DLE. ............ | ............ | GVflydI |
| Cons | GPRGGGIAER | -D-E-------- | *********** | ---GVM-R-I |

|    | DIFnGTLafffo | DIleGTLaK* |
| Brady | DlsGTLCKk | DmitGTLC K 5 |
| Yeast | DlsGTLCKk | DlisGTLGK | fc |
| Human | DlsGTLCKk | DlisGTLCKk |
| Rat | DlsGTLCKk | DdleGivean |
| Bacil | DdleGivean | mkerretaeK vellIEetiv efkqwmnTlg |
| Cons | D--GTLCKA | FGCVGGYIAA | S-LID-VRS | YAPGFIFTTS |

|    | LPPalAAGAQ | ASIafLKTAE | GqkLRdaq.Q | mhaKvLkmlL |
| Brady | LPPalsaAt | AalKhLKTss | ..weReR.HQ | draavkaiL |
| Yeast | LPPsVMAGAt | AalRyqrchi | dlrtsqKht | myVKkafheL |
| Human | LPPmlLAGAl | esVRILKsAE | GRvLRrQ.HQ | rNVK1MrQML |
| Rat | LPPmlLAGAl | esVRILKsnE | GRaLRrQ.HQ | rNBK1MrQML |
| Bacil | vvPvIsAlre | kalaIqsetm | dsieRkIhP1 | strekkkl..L |
| Cons | LPP-ILAGA- | A-IRILK- | GR-LR-Q-HQ | -NVK-M-QML |
Figure 9, continued

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>319</td>
<td>358</td>
</tr>
<tr>
<td>Rcap</td>
<td>kalGMPIdh gSHIVPVvIg DpvhtkaVSD mLLSdygVVY</td>
</tr>
<tr>
<td>Brady</td>
<td>naAGLPVmss dtHIVP1f1g DAeKckqaSD lLLeHgIYI</td>
</tr>
<tr>
<td>Yeast</td>
<td>gipviPnpsh ivpVlignad lAKqasDiIi nkhqiYvqaI</td>
</tr>
<tr>
<td>Human</td>
<td>mdAGLPVvHc PSHIIPvVa DAAKnTEICD ILMtrHnIYV</td>
</tr>
<tr>
<td>Rat</td>
<td>mdAGLPVIhc PSHIIPVrVa DAAKnTEICD ELMtrHnIYV</td>
</tr>
<tr>
<td>Bacil</td>
<td>nkhktsIInq mlr.dPIlkv kelaadadSE EKL...alFm</td>
</tr>
<tr>
<td>Cons</td>
<td>--AGLPVI-- PHSI-PV--- DAAK--E-SD ELLS-H-IYV</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>359</td>
<td>396</td>
</tr>
<tr>
<td>Rcap</td>
<td>QpINFPTVPR GtErLRftPS P..vHdlkqI dgLvHAmDLl</td>
</tr>
<tr>
<td>Brady</td>
<td>QpINYPTVak GsErLRITPS P..yHddgLI dqLaEALLqV</td>
</tr>
<tr>
<td>Yeast</td>
<td>nfptvargte rltiptygth t.ndlsdILl navdDvFn.</td>
</tr>
<tr>
<td>Human</td>
<td>QaINYPRVPR GEELLRIaPt P..hHtoqMm nYFlEnLLvt</td>
</tr>
<tr>
<td>Rat</td>
<td>QaINYPTVPR GEELLRIaPt P..hHtpqMm nFFlEkLLLt</td>
</tr>
<tr>
<td>Bacil</td>
<td>Qidfiieaa. GrqMMktvSe sqkvHsfkkka eskaqfslv</td>
</tr>
<tr>
<td>Cons</td>
<td>Q-INYPTVPR GEELLRI-PS P---H----I ----EALLI-</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>397</td>
<td>401</td>
</tr>
<tr>
<td>Rcap</td>
<td>WaRcA.....</td>
</tr>
<tr>
<td>Brady</td>
<td>WdRLGlplkq</td>
</tr>
<tr>
<td>Yeast</td>
<td>..........</td>
</tr>
<tr>
<td>Human</td>
<td>WkqVG.....</td>
</tr>
<tr>
<td>Rat</td>
<td>WkRVG.....</td>
</tr>
<tr>
<td>Bacil</td>
<td>se........</td>
</tr>
<tr>
<td>Cons</td>
<td>W-RVG-----</td>
</tr>
</tbody>
</table>
Pyridoxal phosphate is maintained in the active site by several non-covalent interactions (46) including hydrophobic interactions and hydrogen bonding. One amino acid must be conserved in all pyridoxal phosphate enzymes; that is lysine (66). When the pyridoxal phosphate is not forming a Schiff base with glycine, it must form one with the ε-amino group of a specific lysine to stabilize the enzyme-pyridoxal phosphate complex (46). Otsuka et al. (66) have proposed that the sequence: P(D/E)(I/L)L(X), G(A/G)(L/F) represents a consensus sequence for a pyridoxal phosphate binding site based on amino acid homology between several pyridoxal phosphate enzymes. Upon comparing the consensus δ-aminolevulinate synthase amino acid sequence with the sequence of the pyridoxal phosphate binding site, one region was found to bear high homology to the consensus sequence. This region contains a fully conserved lysine residue (boxed in on Figure 9) at position 248. Examination of the amino acids surrounding the conserved lysine reveals a high degree of identity with the consensus pyridoxal phosphate binding sequence.

As previously mentioned, AJB529 still possesses ~10% of the normal δ-aminolevulinate synthase activity. Two possible explanations were proposed for this observation;
either there is a second gene encoding δ-aminolevulinate synthase, or the mutation is leaky. Since AJB529 is unable to sustain growth without exogenous δ-aminolevulinic acid provided, it is not likely that there is a second δ-aminolevulinate synthase in R. capsulatus (102). Additionally, when R. capsulatus chromosomal DNA is subjected to southern hybridization with either the hemT gene from R. sphaeroides, or the hemA from Rhizobium meliloti (51), the banding patterns were identical, thus proving the existence of only one gene (36).

An explanation for the 10% δ-aminolevulinate synthase activity was uncovered when the DNA sequence data revealed the exact location of the Tn5 insertion (Figure 5, 103). The Tn5 inserted 20 bases into the coding region for δ-aminolevulinate synthase. Tn5 has an outward facing promoter (7). Although translation could not initiate at the normal 'ATG', there is a 'TTG' codon in Tn5 from which translation could initiate (83). The resultant δ-aminolevulinate synthase would be a hybrid with the first six N-terminal amino acids encoded by Tn5 sequence, and the remainder of the protein encoded by hemA. Efficiency of translation initiation from a 'TTG' codon is only 5% of that using a traditional 'ATG' codon (83). This would
explain the low level of activity while transcription could be maintained, but, at a reduced level.

This basal level of δ-aminolevulinate synthase activity coupled with the fact that no mutant which completely lacked δ-aminolevulinate synthase activity could be isolated, suggests that there may be a polar effect on an essential gene downstream of hemA. Analysis of the DNA sequence data downstream of hemA revealed an 'ATG' codon at position 1,381 (Figure 5), 55 base pairs past the stop codon for hemA. This open reading frame extended through the EcoRI cloning site at position 1,403. Upon insertion of the chloramphenicol acetyl transferase (cat) cartridge from pCM7 (20) into the HindIII site at 1,527, transcription was assessed by measuring cat activity. Two-fold higher cat activity was observed when cells were incubated with low oxygen tension than those incubated at high oxygen tension (M. Wright, unpublished data). These results prove that oxygen regulated transcription, identical in scale to that for hemA, exists downstream of the hemA gene.

Hornberger et al (37) have cloned the 1.58 kb Smal-HindIII fragment containing all of hemA and the downstream
region up to the HindIII recognition site into a promoter search vector. This placed the HindIII site just upstream of a promoterless lacZ gene. β-Galactosidase activity was observed when the plasmid was transferred into R. capsulatus, though no quantitation of β-galactosidase activity from this plasmid construct was reported. Activity was also observed when the 0.78 kb SmaI-EcoRI fragment containing the hemA promoter and one-half of the structural gene was cloned in the promoter search vector. They report that β-galactosidase activity was two-to-three-fold higher upon shifting cells from high oxygen to low oxygen conditions. This is in agreement with the data obtained by Wright et al. (103). A 0.4 kb fragment, one end of which was generated by exonuclease III digestion of most of the hemA gene, the other end corresponded to the HindIII site on the 1.9 kb EcoRI fragment, was cloned into the promoter search vector. They report that β-galactosidase activity was observed, and that it was unaffected by oxygen tension. This result suggests that there is a promoter between hemA and the HindIII site which is not regulated by oxygen.

The latter finding is in direct conflict with the results obtained when the cat cartridge was inserted in
the HindIII site, as discussed earlier. Recall, in that experiment, a two-fold regulation of transcription by oxygen was observed at the HindIII site. This discrepancy may be due to the differences between the two test plasmids. When the cat cartridge was inserted in the HindIII site, the hemA gene and its promoter were still present upstream. In the Hornberger experiment, the hemA promoter had been deleted. There may be a weak promoter between hemA and the HindIII recognition site, transcription from which is virtually undetectable in the presence of the hemA promoter. The hemA promoter, presumably, is much stronger, resulting in transcription extending from it through the HindIII site. The fact that no quantitative data regarding β-galactosidase activity under high and low oxygen is presented by Hornberger for the plasmid that contained the 1.58 kb insert, leaves this question unresolved.

To derive the sequence of the open reading frame downstream of hemA, the 1.9 kb EcoRI fragment (Figure 2) was cloned into the EcoRI recognition site of M13mp18. The resultant recombinant phage were used to provide template DNA for the subsequent sequencing experiments. The open reading frame was found to extend 1,302 base
pairs, from 1,380 to 2,681. As is clear from the codon preference plots (Figures 7 and 8), translation is not likely to initiate at the first ATG codon past hemA.

There are three ATG codons shortly past the first one in the same reading frame. The first of the three, located at 1,466 is 140 base pairs beyond the stop codon for hemA. The second initiation codon is located at 1,525, immediately prior to the HindIII recognition sequence, and the third is at 1,540. Codon preference predicts that initiation of translation of orf-2 probably occurs at the first of the three ATG codons (Figure 7). Initiation at this point results in an open reading frame of 1,239 base pairs, encoding a product of 412 amino acids. The product of orf-2 has an estimated molecular weight of 50,671. Hornberger et al. (37) also suggests that translation of orf-2 initiates with the ATG at 1,466.

Their claim that translation initiates with this start codon is based on the synthesis of a 12 kDa polypeptide in an in vitro transcription-translation system. When they cloned the 2.5 kb BglII-HindIII fragment containing all of hemA, orf-2 up to the HindIII
site and 1,120 bases of DNA upstream of hemA into a pGEM plasmid, two major products were obtained. The first was a 44 kDa polypeptide that they claim is δ-aminolevulinate synthase, the second is the 12 kDa polypeptide. They propose that the 12 kDa polypeptide is a fusion peptide consisting of the first 26 amino acids encoded by orf-2, and 79 amino acids encoded by the vector.

To test this hypothesis, they deleted 947 base pairs of DNA from upstream of hemA, and inserted this Smal-HindIII fragment into the vector, but in the opposite orientation from the previous insert. When this plasmid was used as the template in the in vitro system, only the 44 kDa polypeptide was visualized. Hornberger claims that a 5 kDa polypeptide is synthesized, corresponding to Orf-2, that cannot be detected with a 15% polyacrylamide gel. They never report having used a higher percentage polyacrylamide gel to attempt to visualize the smaller product. Moreover, they never report if they cloned this Smal-HindIII fragment in the opposite orientation to determine if the 12 kDa polypeptide is formed. This leaves open the possibility that the small product is encoded in the 947 base pair region which had been deleted. From codon preference and the data presented by
Hornberger et al (37), however, it does appear that translation of orf-2 initiates with the 'ATG' at 1,466; although, definitive proof is still lacking.

Preliminary data obtained when cell extracts from PAS100 (pCAP17) were incubated with δ-aminolevulinic acid demonstrated that the rate of protoporphyrin synthesis was higher than that for PAS100 without the plasmid (R. Cardin, unpublished data). This led to the question of whether orf-2 may be another hem gene. Assays for the various enzymes involved in tetrapyrrole biosynthesis did not yield convincing evidence that this is the case. However, some of these assays are plagued by difficulties which make the interpretation of the results somewhat unreliable. Most notably, the coproporphyrinogen oxidase assay which relies on the ability to distinguish protoporphyrin from coproporphyrin following the reaction (80). The separation techniques described in the protocols do not remove all of the coproporphyrin from the assay mixture. Coproporphyrin has an absorption maximum at 401 nm, while that for protoporphyrin is at 408 nm (89). Theoretically, even in a mixture of the two, one should be able to quantitate each, since their absorption maxima are quite distinct. In practice, however, when
coproporphyrin is present in solution with protoporphyrin, only one peak is observed, and its center is at 401 nm. This makes detection of protoporphyrin impossible, thus, any results obtained from such an assay are questionable. The assay for uroporphyrinogen III cosynthase requires the purification of the enzyme from cell extracts.

When the DNA sequence of orf-2 was compared with all other hem genes in Genbank (23), no homology to any of them was detected. Moreover, when the DNA sequence of this open reading frame was compared with all DNA sequences in the Genbank data base (23), no significant homology was observed. When the predicted amino acid sequence encoded by orf-2 was used to search the data base of protein sequences, SwissProt (23), likewise, no homology was detected. In fact, the highest degree of homology was with collagen proteins from various sources. This similarity was limited to only a short stretch of the protein, and only bore between 27 and 35% identity.

In an effort to identify the gene corresponding to orf-2 by the phenotype of a mutant for that gene, a kanamycin resistance cartridge from pUC-4K was cloned into the SmaI recognition site within the gene at position
2,007. The resulting plasmid, pCAP81, will not be maintained in *R. capsulatus*, thus, the only kanamycin resistant exconjugants will result from a recombinational event between the plasmid and chromosome. Several attempts were made to mate this plasmid into *R. capsulatus* strain PAS100 on rich medium containing hemin, methionine and cysteine which should supplement *hem* mutants, but, no kanamycin resistant exconjugants were obtained. We are currently unable to explain the inability to get recombination of the kan cartridge into the PAS100 cromosome. It is possible that the resulting mutation is lethal, or that the medium used will not support growth of the mutant strain.

From the sequence data obtained downstream of orf-2, a third open reading frame was found. This open reading frame initiates 84 base pairs past the termination codon for orf-2 and is transcribed in the same direction. This reading frame is open from 2,766 through the EcoRI recognition site used for cloning at 3,304. Figure 8, the codon preference plot, shows that this open reading frame is likely to be translated in *R. capsulatus*. Note, also, from the codon preference plot, two mistakes have been entered in the compiled sequence. One of the mistakes
involves the insertion of one base which is not truly present, the other is the omission of one base, both probably obscured by G-C compression. These mistakes will need to be resolved by deriving the DNA sequence in this region from the complementary strand.

When the sequence of this open reading frame was compared to all published sequences in Genbank (23), no significant homology was detected, just as with orf-2. The predicted amino acid sequence of the product of orf-3 is not included in Figure 5 because of the obvious mistakes which exist in the DNA sequence data.

In summary, therefore, the DNA sequence of the hemA gene from R. capsulatus was determined. The open reading frame which encodes δ-aminolevulinate synthase was found to be 1,206 base pairs, and would code for a product of 50,491. This open reading frame was proven to be translated in R. capsulatus by Wright et al. (103), and the polymerase chain reaction (PCR) was used to amplify the open reading frame. The product of the PCR reaction was cloned into an expression vector and transformed into E. coli strain S905 (hemA41). Once the plasmid construct was introduced into S905, the strain was no longer
dependent on exogenous δ-aminolevulinic acid for growth (A. Biel, unpublished data). This unequivocally proves that the 1,206 base pair open reading frame encodes δ-aminolevulinate synthase. A pyridoxal phosphate binding site was proposed based on sequence similarity between δ-aminolevulinate synthases and other pyridoxal phosphate enzymes. Site directed mutagenesis is needed to determine whether the proposed site is necessary for pyridoxal phosphate binding.

Enzyme assays have shown that the rate of protoporphyrin synthesis increases when pCAP17 is present in R. capsulatus, indicating that there may be a second hem gene on pCAP17. Immediately downstream of the hemA gene, a second open reading frame was found. This open reading frame, designated orf-2, is 1,239 base pairs long, and would encode a product of 412 amino acids with a deduced molecular weight of 50,671. The codon preference profile for orf-2 strongly suggests that this open reading frame is likely to be translated in R. capsulatus. Enzyme assays of strains which contain this region on expression vectors have shown that the product of this open reading frame is neither porphobilinogen synthase, porphobilinogen deaminase nor protoporphyrinogen oxidase; the products of
the \textit{hemB}, \textit{hemC} and \textit{hemG} genes, respectively. However, it is not possible to say whether orf-2 may correspond to the \textit{hemD}, \textit{hemE} or \textit{hemF}. Attempts to construct an insertion mutant for orf-2 were not successful. Additional work will be necessary in this area to determine what orf-2 encodes.

Finally, 55 base pairs downstream of orf-2, a third open reading frame was uncovered. Like orf-2, it shows no significant homology to any of the sequences currently present in the Genbank database. This open reading frame is likely to be translated in \textit{R. capsulatus} as is evident from its codon preference profile (Figure 8). In this case also, more work is required to identify the product of this open reading frame. In addition to identifying the products of orf-2 and orf-3, more work is needed to determine whether an operon exists between these three genes. The fact that three genes, which are transcribed in the same direction, are clustered so close on the chromosome, and that no complete block of the \textit{hemA} gene was obtained suggests that \textit{hemA} is part of an operon. Translational fusion constructs (pCAP77, pCAP84) have been made between orf-2 and \textit{lacZ}, however, I have been unable to mate them into \textit{R. capsulatus} strain AJB529. It now
appears that use of the pNM481/pDPT51 mating system is not possible when the recipient strain is AJB529. Once this obstacle is overcome, it will be possible to directly test whether hemA is part of an operon by blocking transcription from hemA and looking at what effect the block has on β-galactosidase activity. If β-galactosidase activity is also disrupted by blocking transcription from the hemA promoter, this would be convincing evidence that hemA and orf-2 are transcriptionally coupled.


35. Hoare, D.S., and H. Heath. 1959. The biosynthesis of porphyrins from porphobilinogen by Rhodopseudomonas sphaeroides. 2. The partial purification and some properties of


70. Porra, R.J., O. Klein, and P.E. Wright. 1983. The proof by $^{13}$C-NMR spectroscopy of the predominance of the C$_3$ pathway over the Shemin pathway in chlorophyll biosynthesis in higher plants and of the formation of the methyl ester groups of chlorophyll from glycine. Eur. J. Biochem. 130:509.


Vita

John Joseph Eckert was born on October 6, 1963 in Saratoga Springs, New York. He graduated from Saratoga Springs Senior High School in June, 1981. In May, 1985 John was awarded the degree of Bachelor of Arts in Biology and Chemistry from Skidmore College in Saratoga Springs, New York. In August of 1986, John entered the doctoral program in the Department of Microbiology at the Louisiana State University, Baton Rouge under the direction of Alan Jay Biel as his graduate advisor. John is currently a Ph.D. candidate in the Department of Microbiology at L.S.U.
Candidate: John Joseph Eckert

Major Field: Microbiology

Title of Dissertation: THE Rhodobacter capsulatus hemA REGION: NUCLEOTIDE SEQUENCE AND FUNCTIONAL ANALYSIS

Approved:

[Signatures]

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination: Thursday, November 7, 1991