REGULATION OF HEMT EXPRESSION IN RHODOBACTER SPHAEROIDES WILD TYPE STRAIN 2.4.9

Natalie Coulianos

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2018

Committee:
Jill Zeilstra-Rylls, Advisor
I-Fen Lin
Graduate Faculty Representative
George Bullerjahn
Raymond Larsen
Paul Morris
Jill Zeilstra-Ryalls, Advisor

The metabolite 5-aminolevulinic acid (ALA) is essential to all cells, as it is the precursor to tetrapyrroles that include vitamin B_{12}, heme and bacteriochlorophyll. Among bacteria, formation of ALA via the C4 or Shemin pathway, in which succinyl-CoA is condensed with glycine in pyridoxal phosphate-dependent reaction catalyzed by ALA synthase, is limited to the \( \alpha \) class of proteobacteria, which include the facultative prototroph *Rhodobacter sphaeroides*. In fact, these bacteria have two ALA synthase isoenzymes, and relatives encode as many as four ALA synthases. Despite decades of studies, only recently has it come to light that the *R. sphaeroides* enzymes differ with respect to their sensitivity to feedback inhibition by heme (1). However, understanding the full significance of this finding requires knowledge as to the presence of the two enzymes in the cell. It is also necessary to explain how it is that the performance (growth) of wild type strains whose genomes only encode one enzyme appears to equal that of other wild type strains whose genomes encode both enzymes.

To learn more about the distinctive roles of the isoenzymes, the products of the *hemA* and *hemT* genes, *lacZ* transcription reporter plasmids were used to examine their expression in four wild type strains; 2.4.1, 2.4.9 and KD131, all of which encode both *hemA* and *hemT*, and 2.4.3 which has only *hemA*. It was found that, in all four strains, *hemA* is induced under anaerobic conditions, but that the induction levels differ. The *hemT* gene is transcriptionally silent in strain 2.4.1 under all growth conditions, including nitrosative stress, while it is actively transcribed in strains 2.4.9 and KD131, and strongly upregulated when cells are respiring on dimethyl sulfoxide.
(DMSO) compared to aerobic-dark and phototrophic conditions. The picture that emerged from these studies, together with the different susceptibilities of the enzymes to heme-mediated feedback inhibition, is that the bacteria employ different strategies to ensure that adequate amounts of ALA are available for synthesis of whatever kinds and levels of tetrapyrroles are needed according to growth conditions. In some strains hemA transcription is strongly upregulated in order to compensate for inhibition by heme; in other strains in which hemA transcription is weakly upregulated, the less sensitive HemT enzyme is present, when needed, to augment ALA synthase activity.

Further examination of hemT expression in strain 2.4.9 identified cis-acting sequences, as well as two extracytoplasmic function (ECF) sigma factors that are absent from strain 2.4.1, as being important for hemT transcription. Using electrophoretic mobility shift assays with purified sigma factor proteins it was determined that both ECF-type sigma factors directly transcribe hemT. EMSAs also confirmed that a second gene, which was suggested from transcriptomic data, is transcribed by these sigma factors. It encodes a periplasmic protein that binds C4-dicarboxylic acids, which are then transported into the cell by proteins whose genes are co-transcribed with the solute binding protein. For one of the ECF-type sigma factors evidence of the presence of a redox-active disulfide bond within the sigma factor itself was obtained, which explains, in part, upregulation of hemT transcription under reducing conditions. Since genes encoding a transporter of C4-dicarboxylic acids are transcribed by these sigma factors, the influence of one such compound (succinate) on the activities of the sigma factors was evaluated. The evidence suggests that succinate acts as an activating signal for the anti-sigma factor whose partner sigma factor contains the redox-active disulfide. While the signal for the anti-sigma
factor of the second sigma factor remains to be determined, it is clear that the transcriptional activity of that sigma factor is also responsive to changes in cellular redox.

**References**

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge my advisor Dr. Jill Zeilstra-Ryalls, who for the past ten years has been a dedicated and admirable mentor to me, without her guidance and continued support throughout my collegiate career I would not have achieved so much. I would like to acknowledge my committee members, Dr. Raymond Larsen, Dr. George Bullerjahn, and Dr. Paul Morris, for their constant encouragement throughout my time here at Bowling Green State University. I would like to also acknowledge my former and current lab members in the Zeilstra lab for their impeccable work which has furthered my own research efforts, for their willingness to always lend a helping hand, and who provided a relaxed, friendly working environment.

Finally, I would like to acknowledge my family and husband, without their encouragement and support from the time I chose my career path, I would not have made it to where I am today. They have always been there when I needed them the most, and for that I am eternally thankful.
# TABLE OF CONTENTS

## CHAPTER I: INTRODUCTION TO Rhodobacter sphaeroides

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purple bacteria</td>
<td>1</td>
</tr>
<tr>
<td><em>Rhodobacter sphaeroides</em></td>
<td>2</td>
</tr>
<tr>
<td>Tetrapyrrole biosynthesis</td>
<td>7</td>
</tr>
<tr>
<td>Project Overview</td>
<td>8</td>
</tr>
<tr>
<td>References</td>
<td>10</td>
</tr>
</tbody>
</table>

## CHAPTER II: ROLE AND REGULATION OF ALA SYNTHASE GENES *hemA* AND *hemT* IN WILD TYPE STRAINS OF *Rhodobacter sphaeroides*

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>15</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>22</td>
</tr>
<tr>
<td>Bacterial strains, plasmids and growth conditions</td>
<td>22</td>
</tr>
<tr>
<td>Measurements of culture densities</td>
<td>22</td>
</tr>
<tr>
<td>DNA treatments and manipulations</td>
<td>23</td>
</tr>
<tr>
<td>Transformations and conjugations</td>
<td>23</td>
</tr>
<tr>
<td>β-galactosidase and protein assays</td>
<td>23</td>
</tr>
<tr>
<td>Construction of JZ5018</td>
<td>23</td>
</tr>
<tr>
<td><em>hemT</em> response to sodium nitroprusside (SNP)</td>
<td>24</td>
</tr>
<tr>
<td>Bioinformatic analyses</td>
<td>24</td>
</tr>
<tr>
<td>Results</td>
<td>28</td>
</tr>
</tbody>
</table>

Survey of *hemA* and *hemT* transcription among wild type strains of

*R. sphaeroides* | 28 |
Analysis of the role of genetic background differences between wild type strains 2.4.1 and 2.4.9 with respect to *hemT* transcription.......................... 31
Transcriptional response of *hemT* to nitrosative stress.......................... 33
Investigation of media composition with respect to *hemT* transcription...... 37
Comparative growth of wild type and *hemT*-null mutants of *R. sphaeroides* 2.4.1 and 2.4.9..................................................................... 38
Discussion .......................................................................................... 43
References ......................................................................................... 47

CHAPTER III: ROLE OF SIGMA FACTORS RSPH17029_3536 AND RSPH17029_3603 IN TRANSCRIPTION OF THE *R. sphaeroides* *hemT* GENE......................................................... 52

Introduction .......................................................................................... 52
Transcription in Bacteria......................................................................... 52
Sigma factors involved in *hemT* gene expression.................................. 56
Materials and Methods........................................................................... 58
Bacterial strains, plasmids and growth conditions, measurements of culture densities, DNA treatments and manipulations, transformations and conjugations, β-galactosidase assays, and protein assays......................... 58
Construction of a *lacZ* transcription reporter plasmids with altered
*hemT* sequences .................................................................................. 58
Construction of mutant strains JZ5323, JZ5321..................................... 58
Construction of expression vectors, and purification of recombinant,
polyhistidine-tagged RSPH17029_3536 and RSPH17029_3603 proteins... 59
Electrophoretic mobility shift assays......................................................... 61
CHAPTER IV: INVESTIGATING THE ROLE OF ECF SIGMA FACTORS

RSPH17029_3536 AND RSPH17029_3603 ......................................................... 93

Introduction ........................................................................................................... 93

Materials and Methods .......................................................................................... 96

Bacterial strains, plasmids and growth conditions .................................................. 96

Measurements of culture densities, DNA treatments and manipulations, transformations and conjugations, β-galactosidase assays, and protein assays ................................................................. 96

Construction of plasmid pBBR::rsph17029_3536-C151S,C201S ......................... 96

Bioinformatic analyses ............................................................................................ 97

Results ..................................................................................................................... 101

Bioinformatic analyses of ECF sigma factor RSPH17029_3536 and
anti-ECF sigma factor RSPH17029_3537 ....................................................... 101
Bioinformatic analysis of ECF sigma factor RSPH17029_3603 and
anti-sigma factor RSPH17029_3602 .......................................................... 103
Bioinformatic analysis of the extended genomic "neighborhood" of
RSPH17029_3603 and 3536 ........................................................................ 107
Investigating the redox responsiveness of RSPH17029_3536 ................. 114
Investigating the role of C_4-dicarboxylic acid with respect to the
transcriptional activities of the ECF sigma factors ..................................... 116
Discussion .................................................................................................... 121
References ................................................................................................... 123
CHAPTER V: SUMMARY AND CONCLUSIONS ........................................... 127
References ................................................................................................... 133
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Common examples of tetrapyrrole species synthesized by <em>Rhodobacter sphaeroides</em></td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Simplified tetrapyrrole biosynthesis pathway in <em>R. sphaeroides</em></td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>Alignments of the upstream sequences of the <em>hemA</em> and <em>hemT</em> genes of the sequenced wild type strains of <em>R. sphaeroides</em></td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>Upstream sequence of <em>hemT</em> from <em>R. sphaeroides</em> wild type strain 2.4.1</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>Transcription of <em>hemA</em> and <em>hemT</em> in <em>R. sphaeroides</em> strains 2.4.1, 2.4.9, 2.4.3 and KD131, as reported by <em>lacZ</em> transcription fusion plasmids indicated</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>Analysis of <em>hemT</em> transcription</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>Responsiveness of <em>hemT</em> transcription, and photosynthesis spectral complexes to nitrosative stress</td>
<td>36</td>
</tr>
<tr>
<td>8</td>
<td>Effect of yeast extract and DMSO on <em>hemT</em> transcription</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>Growth profiles of <em>R. sphaeroides</em></td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>Pigmentation comparison of aerobic <em>R. sphaeroides</em> wild type strains 2.4.1 and 2.4.9</td>
<td>42</td>
</tr>
<tr>
<td>11</td>
<td>Schematic diagram of the canonical $\sigma^{70}$-RNA polymerase (RNAP) holoenzyme $(\alpha_2 \beta\beta'\omega\sigma)$ and its corresponding promoter structure</td>
<td>53</td>
</tr>
<tr>
<td>12</td>
<td>Extracytoplasmic function (ECF) $\sigma$-anti-$\sigma$ factor interaction and signaling</td>
<td>55</td>
</tr>
<tr>
<td>13</td>
<td>$\beta$-galactosidase activities of <em>hemT</em> in <em>R. sphaeroides</em> wild type 2.4.9 versus 2.4.9 sigma factor mutant strains</td>
<td>68</td>
</tr>
<tr>
<td>14</td>
<td>Growth profiles of <em>R. sphaeroides</em> wild type 2.4.9, the <em>hemT</em> null mutant strain</td>
<td></td>
</tr>
</tbody>
</table>
JZ5018, and the sigma factor mutants JZ5323 and JZ5321

15 ECF-like promoter sequences present within the upstream sequences of hemT

16 Investigation of an ECF-like promoter within the hemT upstream sequences

17 Investigation of ECF-like promoter sequences within the hemT upstream sequences

18 Upstream sequences of hemA containing the putative ECF-like promoter motif AAAAT-N_{17}-TTGT found upstream of hemT

19 Analysis of the role of the sigma factors in expression of 2.4.9 hemA

20 Detection of poly(his)-tagged fusion proteins RSPH17029_3536 and RSPH17029_3603

21 Electrophoretic mobility shift assays using purified RSPH17029_3536 and RSPH17029_3603 incubated with biotin-labeled double stranded hemT upstream sequences that include the ECF-like promoter sequence AAAAT-N_{19}-TTGT

22 Competition EMSA using RSPH17029_3536, labeled double stranded DNA corresponding to the hemT upstream sequences, and an unlabeled restriction fragment isolated from plasmid pHemT-UP(2.4.9) that contains these same sequences

23 Electrophoretic mobility shift assays using purified RSPH17029_3536 and RSPH17029_3603 incubated with biotin-labeled single stranded hemT upstream sequences that include the ECF-like promoter sequence AAAAT-N_{19}-TTGT

24 Alignment of upstream sequences from genes 2.4.9-hemT and rsph17029_0267 that were used in electrophoretic mobility shift assays

25 Electrophoretic mobility shift assays using purified RSPH17029_3536 and
RSPH17029_3603 incubated with biotin-labeled single stranded

rsph17029_0267 upstream sequences that include the ECF-like promoter

sequence AAAAT-N_{17}-TG ................................................................. 85

26 Amino acid sequence alignment of M. tuberculosis SigK and R. sphaeroides

2.4.9 ECF sigma factor RSPH17029_3536 .................................................. 102

27 Upstream sequences of rsph17029_3536. .................................................. 103

28 Upstream sequences of rsph17029_3603 and rsph17029_3602 ..................... 106

29 Genomic context of ECF sigma factor RSPH17029_3536 and

RSPH17029_3603 .................................................................................. 109

30 Upstream sequences of operons A – H (Figure 29) within the 200 kbp region

encompassing the ECF-sigma factors encoding products involved in the transport

of C4-dicarboxylic acids, amino acids, metals, chromate, or unknown substrates .... 111

31 β-galactosidase activities in extracts of R. sphaeroides strains with the

2.4.9-hemT::lacZ reporter plasmid pNC1 .................................................. 115

32 β-galactosidase activities of hemT in R. sphaeroides wild type 2.4.9 and

the 2.4.9 sigma factor mutant strains grown in pyruvate versus succinate media,

under (A) aerobic and (B) anaerobic-dark conditions with DMSO as the alternate

electron acceptor and supplemented with 0.1% peptone ................................ 118

33 β-galactosidase activities of hemT in response to C4-dicarboxylic acids with

respect to the redox signaling component of sigma factor RSPH17029_3536 ....... 120
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>98</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>131</td>
</tr>
</tbody>
</table>

1. Bacterial strains and plasmids used in this study
2. Bacterial strains and plasmids used in this study
3. Names and sequences of oligonucleotides, and their applications in this study
4. Bacterial strains and plasmids used in this study
5. Names and sequences of oligonucleotides, and their applications in this study
6. Summary of *hemA*, *hemT* and ECF sigma factor gene representation among wild type strains of *R. sphaeroides*
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>µg</td>
<td>Microgram</td>
</tr>
<tr>
<td>µl</td>
<td>Microliter</td>
</tr>
<tr>
<td>µM</td>
<td>Micromolar</td>
</tr>
<tr>
<td>ABC</td>
<td>ATP-binding cassette</td>
</tr>
<tr>
<td>ALA</td>
<td>5-aminolevulinic acid</td>
</tr>
<tr>
<td>Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Ampicillin resistant</td>
</tr>
<tr>
<td>bchl</td>
<td>Bacteriochlorophyll</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>Basepair</td>
</tr>
<tr>
<td>C</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>COG</td>
<td>Clusters of orthologous groups</td>
</tr>
<tr>
<td>Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Chloramphenicol resistant</td>
</tr>
<tr>
<td>CV</td>
<td>Column volume</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>ECF</td>
<td>Extracytoplasmic function</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>fmol</td>
<td>Femtomoles</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyronoside</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobasepair</td>
</tr>
<tr>
<td>Kᵢ</td>
<td>Inhibition constant</td>
</tr>
<tr>
<td>Kmᴿ</td>
<td>Kanamycin resistant</td>
</tr>
<tr>
<td>LH</td>
<td>Light harvesting complex</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>Mbp</td>
<td>Megabasepair</td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-Morpholino)propanesulfonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>ng</td>
<td>Nanogram</td>
</tr>
<tr>
<td>nM</td>
<td>Nanomolar</td>
</tr>
<tr>
<td>nm</td>
<td>Nanometer</td>
</tr>
<tr>
<td>nmol</td>
<td>Nanomole</td>
</tr>
<tr>
<td>NO</td>
<td>Nitric oxide</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PLP</td>
<td>Pyridoxal-5-phosphate</td>
</tr>
<tr>
<td>PYP</td>
<td>Photoactive yellow protein</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SA</td>
<td>Semi-aerobic</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SNP</td>
<td>Sodium nitroprusside</td>
</tr>
<tr>
<td>Sp^R</td>
<td>Spectinomycin resistant</td>
</tr>
<tr>
<td>St^R</td>
<td>Streptomycin resistant</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate EDTA</td>
</tr>
<tr>
<td>Te^R</td>
<td>Tetracycline resistant</td>
</tr>
<tr>
<td>Te^S</td>
<td>Tetracycline sensitive</td>
</tr>
<tr>
<td>Tp^R</td>
<td>Trimethoprim resistant</td>
</tr>
<tr>
<td>TRAP</td>
<td>Tripartite ATP-independent periplasmic transporter</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-violet</td>
</tr>
<tr>
<td>U</td>
<td>Units</td>
</tr>
<tr>
<td>V</td>
<td>Visible</td>
</tr>
<tr>
<td>W/m^2</td>
<td>Watts per meter squared</td>
</tr>
<tr>
<td>YE</td>
<td>Yeast extract</td>
</tr>
</tbody>
</table>
CHAPTER I: INTRODUCTION TO *Rhodobacter sphaeroides*

**Purple bacteria**

Purple bacteria are gram-negative prokaryotes capable of anoxygenic photosynthesis, the process of converting light energy into chemical energy via photophosphorylation without the generation of oxygen. To do so, they synthesize polypeptides, bacteriochlorophyll, and carotenoids that form so-called photosynthetic units that are housed within specialized intracytoplasmic membranes. Purple bacteria are important in the cycling of carbon in anoxic environments, both as primary producers (photoautotrophy), and as consumers of reduced organic compounds (photoheterotrophy) (16). When growing phototrophically, they can also fix atmospheric nitrogen, and so may require nothing beyond inorganic salts and light to proliferate. These bacteria are widely distributed in nature, and can be found in terrestrial, aquatic and wastewater environments. They are primarily found in freshwater ecosystems such as lakes, ponds, and estuaries where H$_2$S is present, and concentrations of oxygen are low with an abundance of light (16). Their competitive success in these environments is achieved by their ability to photosynthesize in the absence of oxygen, which comes about because expression of genes comprising components of the photosynthetic machinery is suppressed by molecular oxygen (16, 31). Purple bacteria have a major role as H$_2$S consumers in sulfidic environments, and this is ecologically probably more important than their role as primary producers, as they can successfully remediate H$_2$S, which is toxic to plants and animals, by oxidizing it to elemental S$^0$ and SO$_4^{2-}$ (16).

There are two major groups of purple bacteria, originally distinguished by their ability to tolerate and utilize H$_2$S, the purple sulfur bacteria (gamma-proteobacteria) and purple non-sulfur bacteria (alpha- or beta-proteobacteria). Purple sulfur bacteria were characterized by their ability
to tolerate millimolar concentrations of sulfide and store the oxidized sulfur globules internally, while purple non-sulfur bacteria were thought to do neither. However, subsequent studies revealed purple non-sulfur bacteria are capable of growth in the presence of up to 0.5 mM sulfide, and they also have the capacity to oxidize sulfide to $S^0$, $S_4O_6^{2-}$, and $SO_4^{2-}$. However, rather than storing these chemicals internally, the bacteria deposit them outside of the cell (16, 24). These two groups of purple bacteria are now differentiated on the basis of other aspects of their metabolic capabilities.

The environment in which one might find either purple sulfur or non-sulfur bacteria is intimately linked with their physiology. For example, for purple sulfur bacteria, light and sulfide concentrations are the main drivers for their persistence, while dark growth is limited to only a few species (16). Purple non-sulfur bacteria are relatively more metabolically versatile as they are able to grow photoautotrophically, photoheterotrophically, and also in the dark by aerobic or anaerobic respirations if an appropriate alternative electron acceptor is present. Some species have the additional options of using fermentation, chemolithotrophy, and/or denitrification for energy production (16). This breadth of metabolic versatility makes purple non-sulfur bacteria an attractive group to study in regards to not only photosynthesis, but also general molecular diversity.

*Rhodobacter sphaeroides*

*Rhodobacter sphaeroides* is a purple non-sulfur member of the α-3 subgroup of proteobacteria. It can obtain energy from aerobic and anaerobic respiration, fermentation, as well as from anoxygenic phototrophy. Under phototrophic conditions, it can also fix $CO_2$ and $N_2$. In addition it is capable of bioremediation of metal oxides and oxyanions (10, 18), biohydrogen production (2, 7, 8), production of bioplastics in the form of polyhydroxyalkonates
(8, 25), and it is currently being studied as potential therapeutic agents for the treatment of certain cancers by virtue of its ability to selectively target animal tumors. It has the distinct advantage over other bacteria being considered for such therapy, as they do not elicit an immune response. At the least, its autofluorescence properties make it useful for monitoring purposes (13).

*Rhodobacter sphaeroides* 2.4.1 was one of the original nine natural isolates of purple non-sulfur bacteria that were described by Cornelius van Niel in the 1940s (29). Some forty years later, Imhoff *et al.* systematically designated these bacteria as *R. sphaeroides* (9), and deposited them as such at the ATCC. Several of the van Niel strains have been studied throughout the world for decades due to their vast metabolic and genetic diversity. The first genome to be sequenced was that of wild type strain 2.4.1 (17), which confirmed earlier work showing that the bacterium has two chromosomes, an unprecedented discovery at the time (26, 27). The availability of its DNA sequence, as well as extensive transcriptomic data has made it so attractive for many investigations that it has become the workhorse throughout the world for such studies. They include studies of the evolution of photosynthesis, regulation of expression of photosynthesis genes (these are genes that encode structural polypeptides and enzymes that synthesize pigments of the light harvesting and reaction center complexes), the assembly of the photosynthetic machinery into intracytoplasmic membranes, electron transport, and regulated production of tetrapyrroles. The latter is now known to mirror the bacterium’s growth mode, providing a color-coded index of its physiological status.

The genome of *R. sphaeroides* 2.4.1 is highly complex. In addition to its two chromosomes, it includes five plasmids. At ~3.1 Mbp Chromosome I (CI) is larger than Chromosome II (CII), which is ~0.9 Mbp in size, and they encode approximately 3,106 and 874 open reading frames, respectively. Many essential and housekeeping genes are distributed
between CI and CII (1, 3, 4). There is also extensive duplication throughout the genome, which has probably facilitated the evolution of its wide range of metabolic capabilities in that duplications permit gene sequence divergence. A total of 2.7%, or approximately 111.7 kb, of the \textit{R. sphaeroides} 2.4.1 genome is duplicated (3). CI-CI and CII-CII sequence duplications encompass \~39 and \~18 kb, respectively, whereas CI-CII sequence duplications are \~55 kb. In all, a total of 234 genes are duplicated (1, 3); 83.8% of the duplications are chromosomal and 16.2% are localized on a chromosome and a plasmid, or on more than one plasmid (1). Analysis of the gene duplications showed 40.2% can be classified into generalized groups of Clusters of Orthologous Groups (COGs) related to metabolism, while the remaining 10.7%, 18.8% and 15.7% belong to COGs related to information processing, cellular processes, and poorly characterized functions, respectively (1). Of the duplicate protein pairs encoded by \textit{R. sphaeroides} 2.4.1, 77% (180 pairs) are more closely related to a homolog (ortholog) of a related species than to its paralog, with the majority of matching orthologs present within the \textit{\alpha}-proteobacteria group (1). Further analysis of the phylogenetic relationships of duplicate protein pairs to the nearest orthologs revealed that many of the gene duplications arose prior to speciation, indicating the origin of multiple chromosomes predates the origin of the species (1).

Before the sequencing of \textit{R. sphaeroides} strains other than 2.4.1 it was known that there was extensive chromosomal divergence between CI and CII among them, based on identifying specific gene loci using a combination of macro-restriction length polymorphisms and Southern hybridizations (21, 27). For example, the number of \textit{rrn} operons coding for ribosomal genes varies from two to five among the different strains examined (21). Optical mapping comparisons of strain 2.4.1, and two other van Niel isolates, 2.4.9 (ATCC 17029) and 2.4.3 (ATCC 17025), revealed the CIs are of similar size, while the CIIIs vary (34). This was confirmed by genome
sequencing (4). The relative evolution of CI and CII among these strains was also investigated by comparisons of sequence duplications and global sequence alignments. They indicate that the CIIs have evolved more rapidly than the CIs, and this rapid rate of evolution is due to higher numbers of genetic rearrangements and acquisition of new genetic elements via horizontal gene transfer (4). They also revealed that strain 2.4.1 is more closely related to strain 2.4.9 than to strain 2.4.3 (4). As a consequence of their genetic differences the metabolic diversity that exists among these three strains may play a role in their ability to exploit different niches. For example, *R. sphaeroides* strain 2.4.3 is capable of complete denitrification (12) but not anaerobic respiration using the alternative electron acceptor dimethyl sulfoxide (DMSO), while strains 2.4.1 and 2.4.9 are capable of anaerobic respiration using DMSO, but not of complete denitrification. Further, the interesting photoactive yellow protein (PYP) that is involved in regulating blue light-mediated phototaxis is present in wild type strain 2.4.9, but not in 2.4.1 or 2.4.3. It should be noted that the PYP was originally assigned to 2.4.1 (11); however, sequence analysis and other investigations (14, 21) have subsequently proved that the strain used in those studies is actually strain 2.4.9.

To support the broad metabolic capabilities of *R. sphaeroides*, the cells must be able to produce a variety of components necessary to carry out the various energy generating mechanisms. In order to maintain optimum growth the biosynthesis of these components must be regulated. For example, the production of components necessary for aerobic electron transport chains should take place when oxygen is available, while the molecular machinery for anoxygenic phototrophy should be present only under oxygen limiting conditions. It makes sense that this regulation extends to the biosynthesis of essential cofactors and other metabolites, including tetrapyrroles, which support those energy metabolisms. *R. sphaeroides* synthesizes
three major types of tetrapyrroles (Figure 1), heme, bacteriochlorophyll (bchl), and cobalamin, and their production is dependent upon various environmental stimuli, such as light intensity, oxygen availability, carbon source, etc. (31). Towards understanding how *R. sphaeroides* achieves the orderly production of these molecules, which models the metabolic changes taking place in the cell, a logical focus of investigation is the regulation of 5-aminolevulinic acid formation, the common precursor to all tetrapyrroles.

**Figure 1.** Common examples of tetrapyrrole species synthesized by *Rhodobacter sphaeroides*. Shown are (A) the basic structure of a tetrapyrrole ring which surrounds a central metal atom; (B) cobalamin, required for amino acid and bacteriochlorophyll synthesis, (C) heme type c, a necessary component of cytochromes and peroxidases; and (D) bacteriochlorophyll which is required for anoxygenic photosynthesis.
Tetrapyrrole biosynthesis

The biosynthetic pathway by which tetrapyrroles are formed in \textit{R. sphaeroides} is a branched pathway leading from the common precursor 5-aminolevulinic acid (ALA; Figure 2). The formation of ALA is via a pyridoxal phosphate-dependent condensation of glycine and succinyl-CoA. All animals have two nuclear-encoded ALA synthase isoenzymes that are both mitochondrially localized. The presence of duplicate ALA synthases is reasonably well explained by virtue of their distinctive roles. One gene is expressed in all somatic cells, but its presence is particularly notable in hepatocytes that are rich in mitochondria. Expression of the second gene is limited to erythrocytes where it is induced during erythropoiesis (reviewed in refs. 6 and 32). For many years, \textit{R. sphaeroides} was the only bacterium known to also have two ALA synthase isoenzymes, encoded by the \textit{hemA} and \textit{hemT} genes (28). Now, through DNA sequencing of other \(\alpha\)-proteobacteria, some species are known to have as many as four ALA synthase genes (31). This astounding number of isoenzymes has yet to be explained.
Figure 2. Simplified tetrapyrrole biosynthesis pathway in *R. sphaeroides*. Shown are the branch points leading from the precursor 5-aminolevulinic acid to the biologically relevant metal-bound end products cobalamin, heme, and bacteriochlorophyll. Enzymes participating in the synthesis of tetrapyrroles are shown in bold next to arrows linking the reactants to the products of the reaction they catalyze.

With knowledge of the distribution (21) and sequences of the ALA synthase genes for three sequenced van Niel isolates of *R. sphaeroides*, 2.4.1, 2.4.3, 2.4.9, and also an independent isolate KD131 (15), new experiments may be conducted to relate the presence and regulation of these two genes to the distinctive features of their individual physiologies, with an eye towards understanding why these bacteria have ALA synthase isoenzymes.

**Project overview**

Preliminary work profiling transcription of the two ALA synthase genes, *hemA* and *hemT*, in different wild type strains of *R. sphaeroides* established that some among them, such as wild type strain 2.4.1, rely exclusively on regulated expression of *hemA* to meet the variable needs for ALA required to support its diverse metabolisms, while other strains such as wild type
strain 2.4.9 require regulated expression of both hemA and hemT (5). The goals of this research are (i) to confirm and extend those findings, and (ii) to determine the factors required for the regulated transcription of the hemT gene in R. sphaeroides wild type strain 2.4.9. It was proposed that wild type 2.4.1 does not encode within its genome proteins necessary for hemT transcription, whereas they are encoded in the genome of wild type 2.4.9. Two unique extracytoplasmic function (ECF) sigma factor genes were identified in strain 2.4.9 by comparing its genome to that of strain 2.4.1. These sigma factors direct RNA polymerase to target genes in response to extracytoplasmic stimuli. Efforts will be directed toward determining their role in hemT expression. Efforts will also be directed toward determining the role of each sigma factor in strain 2.4.9, including identifying the signal prompting activation of each sigma factor.
References


CHAPTER II: ROLE AND REGULATION OF ALA SYNTHASE GENES \textit{hemA} AND \textit{hemT} IN WILD TYPE STRAINS OF \textit{Rhodobacter sphaeroides}

Introduction

The two ALA synthase genes of \textit{R. sphaeroides}, \textit{hemA} and \textit{hemT}, were first identified in wild type strain 2.4.1 by probing genomic DNA with partial sequences of the \textit{hemA} gene from \textit{Rhizobium meliloti} (37). The coding sequences of the \textit{hemA} and \textit{hemT} genes were found to be 65\% identical, and the deduced polypeptides are 53\% identical with a further 16\% similarity of the aligned amino acid sequences (25). Several advances have been made in defining the protein factors and mechanisms involved in regulation of \textit{hemA} transcription in strain 2.4.1. The gene is transcribed from two promoters (26), and \textit{hemA} transcription is positively regulated by two global regulatory proteins, PrrA and FnrL, in response to lowering oxygen tensions (12, 30, 31).

PrrA is the partner regulator protein of the PrrBA two component system that senses and, via a phospho-relay process, responds to changes in external conditions that affect the intracellular redox status, such as lowering oxygen tensions (40). Transcriptomic analyses indicate that in \textit{R. sphaeroides} 2.4.1 the expression of approximately 1/4 of all genes are regulated by PrrA (11). They include photosynthesis genes, and in the absence of PrrA, the bacteria are unable to grow phototrophically (9). There are two PrrA binding sites within the upstream sequences of \textit{hemA} (30). Site I is preferentially bound by PrrA in its unphosphorylated state, and this is important for basal transcription of \textit{hemA} from the so-called P1 promoter. PrrA binding site II has greater affinity for the phosphorylated form of PrrA, and binding of PrrA\textsuperscript{P} is thought to be responsible for anaerobic induction of transcription from the P2 promoter (31).
The role of R. sphaeroides FnrL is regarded as analogous to that of its Escherichia coli homolog Fnr. Based on sequence similarities to E. coli Fnr it is thought that, by means of an oxygen-labile 4Fe-4S center within its effector domain, the protein senses changes in oxygen concentrations and is only competent to bind DNA in its absence (18, 41). In both E. coli and R. sphaeroides, this regulator alters transcription of genes that enables the cell to switch between aerobic and anaerobic energy metabolisms. The global role of FnrL in R. sphaeroides is demonstrated by the fact that, when the fnrL gene is disabled, the cells are incapable of either anaerobic respiration in the dark or (anoxygenic) phototrophic growth in the light (12, 41). With respect to hemA, within the upstream region of the gene there exists an FNR consensus-like sequence that overlaps the +1 site of the P1 promoter (25). FnrL binding activates transcription from the P2 promoter, but with respect to the P1 promoter, FnrL stimulation of transcription is indirect, possibly working through PrrA (31).

Evidence has been available for some time indicating that ALA synthase gene representation is variable among wild type strains of R. sphaeroides (27). This variability was confirmed when genomes of several strains became available (6). It is now known that, as is true of R. sphaeroides 2.4.1, wild type strain 2.4.9 also encodes within its genome both hemA and hemT, while wild type strain 2.4.3 has only hemA. Recently the genome of a fourth wild type strain, KD131, has been sequenced and its genome also encodes both the hemA and hemT genes (21). This information made it possible to consider the significance and consequences of differences in ALA synthase gene representation among strains of R. sphaeroides.

Inspection of an alignment of the upstream sequences of hemA and hemT from 2.4.1, 2.4.3, 2.4.9 and KD131 (Figure 3) revealed that there are a number of sequence differences, indicating that transcription of these genes may not be the same in all strains. For example,
within the *hemA* upstream sequences there are numerous alterations in the binding sites for PrrA. Most sequence differences are within the PrrA binding site I, and several are within the 9-base identical repeated sequence that is present in both binding sites (31; Figure 3). However, the FNR consensus-like sequence is invariant among all the strains. With knowledge of ALA synthase gene distribution, and the availability of their sequences, it became possible to examine transcription of *hemA* and *hemT* in strains other than strain 2.4.1 toward understanding the significance of having one or two ALA synthases.
Figure 3. Alignments of the upstream sequences of the *hemA* and *hemT* genes of the sequenced wild type strains of *R. sphaeroides*. (A) The *hemA* upstream sequences of wild type strains 2.4.1, 2.4.3, 2.4.9 and KD131 are aligned; differences from the 2.4.1 sequences are shaded grey.
Features of interest are regions protected by PrrA from DNaseI cleavage (31; yellow; identical 9-base sequence shaded teal), the FNR consensus-like sequence (magenta), the positions of the two +1 start sites of transcription (25), and the translation initiation site. (B) The hemT upstream sequences of wild type strains 2.4.1, 2.4.9 and KD131 are aligned; strain 2.4.3 lacks the hemT gene. Differences from the 2.4.1 sequences are shaded in grey. The translation initiation sites of the divergently arranged rdxA and hemT genes are shown in bold and underlined.

As to the expression or regulation of hemA in strains of R. sphaeroides other than 2.4.1, preliminary efforts revealed that hemA is induced in response to lowering oxygen tensions, but to varying degrees. In strain 2.4.3, a 27-fold induction of hemA under anaerobic-dark conditions was reported, in strain 2.4.9 hemA induction was 2.7-fold and in strain 2.4.1 anaerobic induction was 7.4-fold (7). More recently, it was determined that the strain of 2.4.3 in that study was a mixed population of strains 2.4.1 and 2.4.3. Strain 2.4.3 is, in fact, incapable of anaerobic-dark respiration as it lacks a DMSO reductase gene. Therefore, the regulated transcription of hemA in 2.4.3 was not resolved in that study. Also not known was how and why such a large difference in hemA anaerobic induction was observed in strain 2.4.1 compared to strain 2.4.9. Since both strains 2.4.1 and 2.4.9 have a full complement of genes encoding the PrrBA and FnrL regulatory systems, other factors may be responsible.

While studies focused on regulated hemA transcription in strain 2.4.1 have provided insights into its role in R. sphaeroides, little progress has been made in understanding the role of hemT as it is transcriptionally silent in that strain under all growth conditions investigated (aerobic, anaerobic-dark with DMSO as the electron acceptor, photosynthetic, diazotrophic, and autotrophic) (25, 43). However, in a mutant strain in which the hemA gene had been inactivated
by insertion of a transcription-translation termination (Ω) cassette (29), hemT transcripts were
detected in both aerobic and phototrophic cells, and HemT was able to fulfill the necessary ALA
requirements for growth under both conditions (26). Therefore, it was known that, when
expressed, the hemT gene codes for a fully functional enzyme.

Although the 5' ends of hemT transcripts were identified in this same mutant strain (25)
the inability to detect them in the wild type parent strain (25) suggested that, in addition to the
engineered disruption of the hemA gene, a second mutational event may have occurred.
Confirmation of this possibility came from a subsequent study in which hemA was carefully
disabled in a manner that revealed that a loss of a functional hemA gene confers ALA auxotrophy,
but prototrophs were easily obtained (40). This has been interpreted to mean that the ALA
prototrophy of the original hemA-null mutant strain, HemA1, was due to either the presence of a
trans-acting mutation that resulted in increased hemT transcription, or cis-acting mutations
within the hemT upstream sequence that created a fortuitous promoter functional under the
growth conditions used. That trans-acting mutations were possible was demonstrated by the fact
that ALA prototrophic mutants of hemA-null mutants can be isolated; that cis-acting mutations
were possible was demonstrated by means of a selection process involving a plasmid having a
transcription fusion between the upstream sequences of hemT and the coding sequences of the
Km resistance gene of Tn903 (40). Exconjugants of the plasmid were incapable of growth on
kanamycin at concentrations above 5 µg/ml. However, spontaneous mutants capable of growth
on 25 µg/ml kanamycin and above were successfully isolated and shown to have mutations
within the hemT upstream sequences that led to increased transcription of hemT in wild type
2.4.1 cells (Figure 4; 40).
Figure 4. Upstream sequence of hemT from R. sphaeroides wild type strain 2.4.1. Shown is the entire nucleotide sequence between the divergently arranged rdxA and hemT genes. Features of interest include the +1 site of transcription determined in a 2.4.1-derived mutant strain expressing hemT (25), the locations and sequences of two spontaneous "up" mutants (40), and in bold underneath the E. coli σ^{70} promoter consensus sequence (13).

To extend and clarify the previous preliminary findings with respect to ALA synthase gene transcription among wild type strains of R. sphaeroides (7), lacZ transcription reporter plasmids were used. β-galactosidase activities in extracts of cells with the reporter plasmids that were grown under different conditions were determined. Any transcriptional differences observed were then further characterized based on differences in the respective upstream sequence or the genetic backgrounds.
Materials and Methods

Bacterial strains, plasmids and growth conditions. *R. sphaeroides* was cultured in Sistrom’s succinate minimal medium A (35) at 30°C, and *Escherichia coli* was grown in Luria-Bertani medium (32) at 37°C. The following are the final concentrations of antibiotics used for strain selection and plasmid maintenance: for *R. sphaeroides*, 50 μg/ml each of Spectinomycin (Sp) and Streptomycin (St), 0.8 μg/mL Tetracycline (Tc), and 50 μg/mL Kanamycin (Km); for *E. coli*, the same concentrations of Sp, St, and Km were used, 15 μg/ml of Tc, and 100 μg/ml of Ampicillin (Ap) were also used. *R. sphaeroides* cultures of 100 ml in milk dilution bottles were continuously sparged with gas mixtures in the ratios 30% O₂:68% N₂:2% CO₂ for aerobic conditions, and 2% O₂:97% N₂:1% CO₂ for micro-aerobic conditions. Semi-aerobic conditions were established by inoculating *R. sphaeroides* into 100 ml of medium in 250 mL Erlenmeyer flasks that were incubated at 30°C in a New Brunswick Gyratory Shaker (Model G76) at 90 rpm. For cultures grown anaerobically in the dark, screw-capped tubes were completely filled with Sistrom’s succinate minimal A medium supplemented with dimethyl-sulfoxide (DMSO; final concentration was 0.06 M) as alternate electron acceptor and 0.1% yeast extract. For growth under photosynthetic conditions, screw-capped tubes were completely filled with Sistrom’s succinate minimal A medium and incubated under medium-high light intensity (approximately 10 W/m²). Antibiotics and other fine chemicals were purchased from Sigma-Aldrich (St. Louis, MO).

Measurements of culture densities. Optical densities of cultures were determined by measuring at 660 nm using a U-2010 UV/V spectrophotometer (Hitachi High Technologies America, Inc., Schaumburg, IL), a Spectronic 20 Genesys spectrophotometer, or by using a Klett
colorimeter fitted with a 66-red filter. As necessary, samples of the cultures were diluted to
densities that were within the linear range of the instruments.

**DNA treatments and manipulations.** Plasmid DNA isolations were performed using
GenElute Plasmid Miniprep kit (Sigma-Aldrich, St. Louis, MO). Total genomic DNA was
isolated using the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich). Restriction
endonuclease treatments, and other enzymatic treatments of DNA fragments and plasmids were
performed according to manufacturer’s instructions or according to standard laboratory
procedures (32). Enzymes were purchased from New England BioLabs, Inc. (Beverly, MA),
GIBCO-BRL/Life Technologies, Inc. (Gaithersburg, MD), and Promega (Madison, WI). DNA
was analyzed by standard electrophoretic techniques (Mol. Cloning, Cold Spring Harbor), and
isolation of DNA fragments was performed using a Zymoclean purification kit (Zymo Research
Co., Orange, CA).

**Transformations and conjugations.** *E. coli* cells were prepared for transformations by
CaCl₂ treatment (32). Mobilizations of plasmids into *R. sphaeroides* were performed by
diparental matings using strain S17-1 (Table 1), or triparental matings using the helper strain
HB101(pRK2013), as described by Davis et al. (8).

**β-galactosidase and protein assays.** Crude lysates were prepared and assayed for β-
galactosidase activities following routine laboratory procedures (37). Protein assays of crude
lysates were performed using Pierce BCA® reagents (Thermo Scientific, Rockfield, IL) or using
the Bradford protein assay (Bio-Rad, Hercules, CA) according to manufacturer’s instructions
with bovine serum albumin as a standard.

**Construction of JZ5018.** The 2.4.9-*hemT* null mutant strain JZ5018 was constructed
using the suicide vector plasmid pUI1009 (26; Table 1). The plasmid was mobilized into *R.*
sphaeroides 2.4.9, and double crossover candidates were identified by scoring exconjugants that were Sp/St resistant for Tc sensitivity. The integrity of one such candidate, JZ5018, was verified by sequencing the 2.6 kb PCR product generated by using primers that flank the hemT-deletion::ΩSp^R/St^R insertion site. To confirm that selection for hemT null mutants did not result in mutations in the hemA gene, the complete hemA sequence in strain JZ5018 was also amplified and sequenced.

**hemT response to sodium nitroprusside (SNP).** *R. sphaeroides* strains 2.4.1 and 2.4.9 carrying pUI1098 and pNC1 transcription reporter plasmids, respectively (Table 1), were cultured in triplicate under micro-aerobic conditions, which were established by sparging liquid cultures with a mixture of 97% nitrogen, 2% oxygen, and 1% carbon dioxide to reproduce the conditions reported by Arai, *et al.* (2). At 0.2 OD660, either 1 ml of 100 mM of SNP prepared in 50 mM MOPS buffer, pH 7.0, or 50 mM MOPS equivalent volume, was added to 100 ml of liquid culture for a final concentration of 1 mM SNP. After 15 minutes, 1 hour, and 4 hours, 25 ml samples were removed. Crude lysates were prepared and assayed for β-galactosidase activity. Spectra were recorded by scanning crude lysates between wavelengths of 350-900 nm using a Hitachi U-2010 UV/Vis Spectrophotometer (Hitachi High Technologies America, Inc., Schaumburg, Illinois). Bacteriochlorophyll a-protein complex concentrations were calculated from the spectral data using the method of Meinhardt *et al.* (23). Pigments were extracted from equivalent numbers of whole cells using 7:2 acetone:methanol. Bacteriochlorophyll a and carotenoid concentrations were determined as previously described (23).

**Bioinformatic analyses.** Software used included the BLAST server at NCBI (www.ncbi.nlm.nih.gov; 1, 13), ClustalO (http://www.ebi.ac.uk/Tools/msa/clustalo; 14, 33), ECF
finder (http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder; 36) and GENEDOC (http://www.nrbsc.org/gfx/gedoc; 28).
### Table 1. Bacterial strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strain/Plasmid</th>
<th>Relevant Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rhodobacter sphaeroides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4.1</td>
<td>wild type</td>
<td>35</td>
</tr>
<tr>
<td>2.4.3 (ATCC17025)</td>
<td>wild type</td>
<td>ATCC</td>
</tr>
<tr>
<td>2.4.9 (ATCC17029)</td>
<td>wild type</td>
<td>ATCC</td>
</tr>
<tr>
<td>KD131</td>
<td>wild type</td>
<td>21</td>
</tr>
<tr>
<td>T1</td>
<td>2.4.1 derivative, *hemT::*ΩSp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;</td>
<td>26</td>
</tr>
<tr>
<td>JZ5018</td>
<td>2.4.9-*hemT::*ΩSp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;; Tc&lt;sup&gt;S&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Escherichia coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>F- (φ80lacZΔM15) recA1 endA1 hsdR17 supE44 thi-1 gyrA96 relA1 deoR Δ(lacZYA-argF)U169</td>
<td>17</td>
</tr>
<tr>
<td>DH5aphe</td>
<td>DH5α with *phe::*Tn10d; Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>9</td>
</tr>
<tr>
<td>HB101</td>
<td>F- Δ(gpt-proA)62 leuB6 supE44 ara-14 galK2 lacY1</td>
<td>3</td>
</tr>
<tr>
<td>S17-1</td>
<td>Conjugal donor; C600::RP4 2-(Tc::Mu) (Km::Tn7) pro res&lt;sup&gt;+&lt;/sup&gt; mod&lt;sup&gt;+&lt;/sup&gt; (Tp&lt;sup&gt;R&lt;/sup&gt; Sp&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>34</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBBR1-MCS2</td>
<td>Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>19</td>
</tr>
<tr>
<td>pUI1009</td>
<td>pSUP202-derived *hemT::*ΩSp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;</td>
<td>26</td>
</tr>
<tr>
<td>pUI1098</td>
<td>2.4.1-<em>hemT::lacZ</em> in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>40</td>
</tr>
<tr>
<td>Vector</td>
<td>Description</td>
<td>Resistance</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>pNC1</td>
<td>2.4.9-hemT::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>7</td>
</tr>
<tr>
<td>pNC3</td>
<td>2.4.9-hemA::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>7</td>
</tr>
<tr>
<td>pNC5</td>
<td>2.4.3-hemA::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>7</td>
</tr>
<tr>
<td>pSB4</td>
<td>2.4.1-hemA::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>30</td>
</tr>
</tbody>
</table>
Results


To evaluate hemA and hemT transcription lacZ transcription reporter plasmids were used (Table 1); pUI1098 is the 2.4.1-hemT::lacZ reporter plasmid (40) and pNC1 is the analogous 2.4.9-hemT::lacZ reporter plasmid (7); pSB4 is the 2.4.1-hemA::lacZ reporter plasmid (30) and pNC3 and pNC5 are the analogous 2.4.9 and 2.4.3-hemA::lacZ plasmids, respectively (7). R. sphaeroides wild type strains with their “own” reporter plasmids were grown aerobically, anaerobically-dark with dimethyl sulfoxide (DMSO) as an alternate electron accepter and anaerobically in the light (phototrophically), except for KD131 which was grown with 2.4.9-hemA reporter plasmid since there are no sequence differences. All cultures were then assayed for β-galactosidase activity.

According to the assays, the pattern of hemA transcription was different in all wild type strains examined (Figure 5A). In strains 2.4.1 and 2.4.9, as reported previously (7, 26, 30), hemA was upregulated in response to lowering oxygen tensions. This is presumed to be necessary in order to meet the increased requirement for ALA required for synthesis of the large amounts of bacteriochlorophyll a needed for anoxygenic photosynthesis. For the same reason, it was hypothesized that hemA would also be upregulated in 2.4.3 in response to the lowering of oxygen tensions. While it is true that, in all strains, hemA was indeed induced as oxygen tensions decrease, the extent of induction differed among them. In strain 2.4.1 hemA transcription was induced 7.4-fold under anaerobic-dark conditions and 13.3-fold under phototrophic conditions. However, in strains 2.4.3, 2.4.9, and KD131, the magnitude of the induction was less than that in strain 2.4.1. In strain 2.4.3, hemA transcription was induced only 2.0-fold, raising question as to how this strain meets the increased ALA demand under phototrophic conditions with such
relatively limited induction of its only ALA synthase gene. In strain 2.4.9 hemA transcription was induced 3.0-fold and 1.8-fold under anaerobic-dark and phototrophic conditions, respectively. In strain KD131, the level of induction was similar to that in strain 2.4.9, with increases of 1.8-fold under anaerobic-dark conditions, and 2.4-fold under phototrophic conditions, relative to levels under aerobic conditions. Since both 2.4.9 and KD131 have a hemT gene, examining its expression is necessary in order to formulate hypotheses as to how these bacteria surmount an increased need for ALA.

The analysis of hemT transcription (Figure 5B) indicates that, unlike in wild type strain 2.4.1, hemT is transcribed under all conditions, and this transcription is regulated in wild type strain 2.4.9; it is maximal under anaerobic-dark with DMSO conditions. This strain difference with respect to hemT expression has been confirmed by immunoblot analysis in which HemT protein was detected in samples of R. sphaeroides 2.4.9 but not in samples of strain 2.4.1 (7).

To determine whether or not regulated hemT transcription is unique to strain 2.4.9, β-galactosidase assays were also performed in KD131(pNC1) extracts. (The upstream sequences of the 2.4.9- and KD131-hemT genes differ by only nine nucleotides.) Results from these assays showed that hemT is actively transcribed in this strain as well (Figure 5B), with the highest level of transcription taking place in cells cultured under anaerobic-dark conditions, as is true of strain 2.4.9.

On the basis of these results it is hypothesized that there exists a transcription factor or factors that is (are) present in wild type strains 2.4.9 and KD131, but absent in 2.4.1, which is (are) required for hemT expression, and that the need for hemT expression is most acute under anaerobic-dark conditions.
Figure 5. Transcription of hemA and hemT in R. sphaeroides strains 2.4.1, 2.4.9, 2.4.3 and KD131, as reported by lacZ transcription fusion plasmids indicated (see Table 1 for plasmid descriptions). Panel A: β-galactosidase activity of hemA was assayed in strains 2.4.1, 2.4.9, 2.4.3 and KD131 using extracts of cultures grown under 30% oxygen (black bars), anaerobic-dark with DMSO as an alternate electron accepter (dark grey bars; except strain 2.4.3) and anoxygenic photosynthetic (light grey bars) conditions with reporter plasmids as indicated. Aerobic values for hemA expression in strain 2.4.1 are from (31). Panel B: β-galactosidase activity of hemT was
assayed in wild type strains 2.4.1, 2.4.9, and KD131 using extracts of cultures grown under 30% oxygen (black bars), anaerobic-dark with DMSO as an alternate electron acceptor (dark grey bars) and anoxygenic photosynthetic (light grey bars) conditions with reporter plasmids as indicated. Values are the average of at least three replicates, with ranges indicated in parentheses.

**Analysis of the role of genetic background differences between wild type strains 2.4.1 and 2.4.9 with respect to hemT transcription.** Preliminary data indicated that there are differences in strains 2.4.1 and 2.4.9 that are responsible for the differences in hemT transcription in the two strains. To confirm those results, additional β-galactosidase activity assays of strains 2.4.9 with the 2.4.1-hemT::lacZ reporter plasmid pUI1098 (Table 1) and strain 2.4.1 with the 2.4.9-hemT::lacZ reporter plasmid pNC1 (Table 1) were performed using extracts from cultures grown under anaerobic-dark conditions with DMSO as the alternate electron acceptor.
Figure 6. Analysis of *hemT* transcription. Assays of β-galactosidase activities were performed using extracts of wild type strains 2.4.1 and 2.4.9 with reporter plasmids indicated (details are provided in Table 1), that were grown under anaerobic-dark conditions with dimethyl sulfoxide as alternate electron acceptor. For comparative purposes, values for wild type 2.4.1 (pUI1098) and wild type 2.4.9 (pNC1) presented in Figure 1 are reproduced here. Values are the average of at least three replicates, with ranges indicated.
These results for the “cross-conjugated” strains confirm that *hemT* transcription differences in strain 2.4.1 and 2.4.9 are not due to differences in the upstream sequences of the 2.4.1 and 2.4.9 *hemT* genes; rather they are due to differences in the strains themselves (Figure 6). In strain 2.4.1, regardless of which plasmid was used, *hemT* was reported to be transcriptionally silent as the level of β-galactosidase activity is below that which is reliably detectable (β-galactosidase activity in the presence of the empty vector was 2 U). By contrast, in strain 2.4.9 both plasmids report that *hemT* is transcribed and to approximately the same extent. On the basis of these results, it is hypothesized that strain 2.4.9 encodes within its genome a transcription factor or factors that are required for transcription of *hemT* that is not present within the 2.4.1 genome.

**Transcriptional response of *hemT* to nitrosative stress.** In 2013, Arai et al. (2) reported that *hemT* is induced in response to nitrosative stress, specifically nitric oxide (NO), under micro-aerobic conditions in strain 2.4.1. This would explain why *hemT* transcripts were not detected under routine laboratory growth conditions (25). In that study, *hemT* transcript levels in wild type strain 2.4.1 were measured using Affymetrix genechips and RNA that was isolated from cultures which were exposed to NO for 15 min due to addition of 1 mM sodium nitroprusside (SNP) to the media (2). A 465-fold increase in *hemT* mRNA levels over levels present in cells not exposed to NO was reported (2); it should be noted that the absence of detectable *hemT* transcripts in cells grown in the absence of SNP makes the true magnitude of this induction unknowable.

In light of the differences in *hemT* transcription in strains 2.4.1 versus 2.4.9 observed in previous transcription reporter assays involving β-galactosidase (7), it was important to investigate this NO response in both strains, and to also investigate the dynamics of this
response. To do so, cultures of wild type 2.4.1 with the 2.4.1-\textit{hemT::lacZ} reporter plasmid pUI1098 and 2.4.9 with the 2.4.9-\textit{hemT::lacZ} reporter plasmid pNC1 (Table 1) were grown under those same micro-aerobic conditions, and exposed to SNP in the same manner as described by Arai \textit{et al.} (2). Samples were taken 15 minutes, 1 hour, and 4 hours after exposure to 1 mM SNP and assayed for β-galactosidase activity. By way of comparing the overall level of "stress" experienced by the two strains the amounts of photosynthetic pigment-protein complexes were also determined since NO can generate reactive nitrogen species that are capable of inhibiting electron transport (38), disrupting iron-sulfur clusters by binding to metals, and nitroslyating thiol groups (2).

Based on the β-galactosidase activity assays \textit{hemT} transcription is not induced in strain 2.4.1 (Figure 7A). Furthermore, for strain 2.4.1, SNP does not affect the formation and amounts of photosynthetic Light Harvesting complexes I and II (LHI and LHII) present in the cells (Figure 7B), which are responsible for the capturing of light energy at 875 nm, and at 800 and 850 nm, respectively. Results of this assay are in agreement with previous work by Wu \textit{et al.} (39) that demonstrated by disk diffusion susceptibility assays strain 2.4.1 is not sensitive to NO generated by SNP. This may be due to the ability of strain 2.4.1 to detoxify exogenous NO, as it encodes within its genome the \textit{nor} operon, which codes for proteins involved in nitric oxide reduction (20). Therefore, although this strain is not capable of complete denitrification, the \textit{nor} operon could serve to detoxify NO, which is important in natural environs where the bacteria could co-exist complete denitrifiers (20).

In strain 2.4.9 there is at the most a 2-fold increase in \textit{hemT} transcription after 4 hours of exposure to 1 mM SNP, which is far less than the nearly 11-fold difference in expression in cells grown aerobically \textit{versus} anaerobically in the dark with DMSO. Exposure to 1 mM SNP also
resulted in an approximately 2-fold decrease in LHI and LHII in strain 2.4.9. It should be noted that this represents a substantial change in the levels of these complexes. While strain 2.4.9 also encodes the nor operon, its expression in response to exogenous NO has yet to be investigated. It is possible that expression of the nor operon may be different in 2.4.9 such that the ability of this strain to cope with exogenous NO is less than that of 2.4.1; however, this remains to be determined. Collectively, these findings argue that, while hemT transcriptional activity is unaffected (Figure 7B), other genes that apparently include photosynthesis genes are affected by SNP in strain 2.4.9.
Figure 7. Responsiveness of hemT transcription, and photosynthesis spectral complexes to nitrosative stress. Panel A: β-galactosidase activities in *R. sphaeroides* strains indicated after exposure to 1 mM of SNP for 15 minutes, 1 hour and 4 hours. For comparative purposes
cultures were treated with 50 mM MOPS buffer equivalent volumes as a control. Values are the average of at least three replicates with ranges indicated. Panel B: Concentrations of light harvesting complexes LHII and LHI (nmol bacteriochlorophyll a/mg protein) present after 15 minutes of exposure to SNP.

Investigation of media composition with respect to hemT transcription. According to the analysis of hemT transcription in strain 2.4.9, the gene in maximally transcribed under anaerobic-dark conditions when the cells are obtaining energy using a respiratory chain that terminates with the reduction of DMSO. The bacteria are routinely cultured under those conditions in the presence of yeast extract (YE) to generally reduce the metabolic burden on the cell, but in the case of strain 2.4.9, it is also required because it is conditionally auxotrophic for certain amino acids (4). To determine whether DMSO and/or the components of YE affect hemT transcription, cultures of strain 2.4.9 bearing plasmid pNC1 (Table1) were grown under semi-aerobic conditions, which are not inducing for hemT transcription, but do support growth without amino acid supplementation, in the presence or absence of DMSO and YE, singly or in combination, and extracts of cultures samples were assayed for β-galactosidase activity. Based on the results (Figure 8), under these conditions, hemT transcription is not influenced by the presence of YE or DMSO alone or in combination, as β-galactosidase activity varied at most 2-fold. This variance is far below the approximately 11-fold difference in hemT transcription reported in aerobically versus anaerobically-dark with DMSO cells (7). The conclusion based on these results, and also based on the absence of induction under phototrophic conditions (Figure 5), is that hemT transcription is not responsive to changes in oxygen concentrations, nor to the presence of DMSO and/or YE per se. Rather, the inducing signal for hemT transcription seems
to arise from the specific metabolic state of the cells that prevails when they are obtaining energy by means of an anaerobic electron chain ending with the reduction of DMSO.

Figure 8. Effect of yeast extract and DMSO on hemT transcription. Cultures of 2.4.9 with the 2.4.9-hemT::lacZ reporter plasmid pNC1 (Table 1) were grown under semi-aerobic (SA) conditions in the absence or presence of yeast extract (YE) and dimethyl sulfoxide (DMSO), or in combination as indicated. Values are the average of at least three replicates, with ranges indicated.

Comparative growth of wild type and hemT-null mutants of R. sphaeroides 2.4.1 and 2.4.9. Toward determining the role of hemT in strain 2.4.9, growth of the hemT null mutant strain JZ5018 (Table 1) was compared to that of the wild type parent strain 2.4.9. Growth of wild type strain 2.4.1 and its corresponding hemT null mutant T1 (26) were similarly profiled.
Because \textit{hemT} is not represented in the 2.4.3 genome, it was of interest to profile the growth of this strain as well. In all cases, aerobically grown precultures were used as inoculums for cultures that were either incubated anaerobically in the dark with DMSO or phototrophically. Thus, this was, in fact, an analysis of the ability of these strains to transition from relying upon aerobic energy metabolism to relying upon anaerobic energy metabolism using an alternate electron acceptor, or to relying upon phototrophic energy metabolism.

While strains 2.4.1 and T1 are able to rapidly transition to anaerobic-dark growth, wild type strain 2.4.9 took approximately 24 hours longer to achieve exponential growth and this lag-period was prolonged an additional 72 hours for mutant strain JZ5018 (Figure 9A and B). However, once the \textit{hemT} mutant bacteria have successfully transitioned, their growth rate is comparable to that of the parent wild type strain 2.4.9; both having a doubling time of approximately 3 hours, which is that same as that of 2.4.1. That the observed growth of 2.4.9 and JZ5018 is not a consequence of selection for mutants was determined by isolating single colonies from the cultures, re-culturing them aerobically, and then examining growth of anaerobic-dark inoculums; the same profile was observed (results not shown).
Figure 9. Growth profiles of *R. sphaeroides*. (A) Growth profiles of wild type strain 2.4.1 and the 2.4.1 *hemT*-null mutant T1 (24) when transitioning from aerobic to anaerobic-dark (with DMSO) conditions. (B) Growth profiles of wild type strain 2.4.9 and the 2.4.9 *hemT*-null mutant strain JZ5018 when transitioning from aerobic to anaerobic-dark conditions. (C) Growth profiles of *R. sphaeroides* strains 2.4.9, JZ5018, 2.4.1, and 2.4.3 transitioning from aerobic to anoxygenic phototrophic conditions. (D) Comparison of lag-periods of *R. sphaeroides* strains 2.4.9, JZ5018, 2.4.1, and 2.4.3 transitioning from aerobic to anoxygenic phototrophic conditions. All values represent the average of triplicate growth experiments with ranges indicated.
Unlike the slow transition to anaerobic-dark growth, strains 2.4.9 and JZ5018 readily transition from aerobic to phototrophic growth with a brief lag of 4 hours or less. However, cultures of strains 2.4.1 and mutant strain T1 undergo a 12-hour lag period before cells enter exponential growth when undergoing this same transition (Figure 9C); strain 2.4.3 had a lag phase that was 9 hours longer than that of strain 2.4.1 (Figure 9C and D).

Based on the growth profiles, it is hypothesized that the role of hemT is to provide additional ALA necessary for the de novo synthesis of specialized cytochromes that are components of the anaerobic-dark respiratory chain ending in DMSO reductase (24), a requirement that apparently the hemA gene alone cannot fulfill in this strain. This requirement is unique to anaerobic-dark with DMSO growth, because no such delay was observed as these bacteria transitioned to phototrophic growth. That this lag is not observed for strain 2.4.1 in which hemT is transcriptionally silent, nor for the hemT null mutant strain T1, indicates that this strain has evolved a different strategy for meeting their ALA requirements, and it is one that does not rely upon hemT. This strategy could work through hemA, which is more strongly upregulated in wild type strain 2.4.1 (more than 7-fold) than in strain 2.4.9 (approximately 3-fold).

Evidently, hemT is not required for strain 2.4.9 to transition to phototrophic growth since both wild type parent and its hemT null mutant derivative JZ5018 are equally facile at undergoing the transition. However, it is clear that strains 2.4.1, 2.4.3, and 2.4.9 differ in their abilities to undergo this transition. This could be explained by considering the differences in pigmentation of aerobic cultures of wild type strains (Figure 10A); cultures of wild type strains 2.4.1 (and 2.4.3; not shown) are unpigmented while cultures of 2.4.9 are pigmented. It suggests that photosynthesis pigments, and possibly the photosystem itself, are present even in the
presence of oxygen in strain 2.4.9. Quantitation of bacteriochlorophyll \( a \) and carotenoid levels present in aerobically grown cells confirmed this, as wild type 2.4.9 had concentrations of pigments that are 3 times higher than those present in strain 2.4.1 (Figure 10B). Therefore, the differences in gene expression between strains 2.4.1, 2.4.3, and 2.4.9 extend beyond \( \text{hemT} \) alone.

The overall picture that emerges from these growth studies is that wild type strain 2.4.1 has evolved to respond rapidly to changes in oxygen availability while wild type strain 2.4.9 has evolved to respond rapidly to changes in light availability. With respect to strain 2.4.3, one could argue that its transitioning capabilities are best suited to an organism capable of complete denitrification, but this remains to be determined.

Figure 10. Pigmentation comparison of aerobic \( R. \text{sphaeroides} \) wild type strains 2.4.1 and 2.4.9. (A) Image of aerobic cultures of wild type strains 2.4.1 and 2.4.9. (B) Total bacteriochlorophyll \( a \) and carotenoid levels in equivalent numbers of wild type strains 2.4.1 and 2.4.9 cells.
Discussion

The experiments presented here were designed to investigate the differences in regulated transcription of the two ALA synthase genes, \textit{hemA} and \textit{hemT}, among wild type strains of \textit{R. sphaeroides}. They revealed that transcription \textit{hemA} and \textit{hemT} is different amongst the four sequenced strains of \textit{R. sphaeroides} studied here. While transcription of the \textit{hemA} gene in strains 2.4.9 and KD131 increases when oxygen becomes unavailable, the increase is far less than that observed in strain 2.4.1; this limited oxygen responsiveness of \textit{hemA} transcription is likewise true of strain 2.4.3.

Toward understanding the \textit{hemA} transcription differences between these four wild type strains, a bioinformatic analysis of proteins known to regulate \textit{hemA} expression, FnrL and PrrA, was performed. This was extended to include PrrB and C, as they are necessary components of the signal-transduction system to which PrrA belongs (7, 10). Amino acid sequence comparisons of proteins FnrL, PrrA, B, and C from wild type strain 2.4.9 and KD131 were identical to those of strain 2.4.1, while those of 2.4.3 differ substantially (7). However, it is unlikely the diminished induction of \textit{hemA} in 2.4.3 relative to that which occurs in strain 2.4.1 is due to any sequence differences among FnrL, PrrB, or PrrA as important residues involved in DNA binding of FnrL and PrrA are conserved. Likewise, the transmembrane sensor domain responsible for receiving the redox signal of PrrB is conserved (7). In the case of PrrC, which is thought to relay to PrrBA changes in cellular redox, there exist a total of 31 amino acid residue differences (7). The sheer number of amino acid sequence differences within the protein suggests that 2.4.3-PrrC may differ in its activity relative to the 2.4.1 or 2.4.9 proteins. Thus, if PrrC does transmit the inducing signal to PrrB, diminished activity of PrrC would effectively lead to minimal induction of \textit{hemA} transcription due to low levels of phosphorylated PrrA under
phototrophic conditions. In combination with the hemA sequence differences, this could explain the pattern of transcription suggested by the β-galactosidase assay results for 2.4.3, and also the prolonged lag-period of strain 2.4.3 relative to strain 2.4.1 (Figure 9D) as cells transition from aerobic to phototrophic (anaerobic) conditions.

While the amino acid sequences of the known regulators of hemA transcription are identical between strains 2.4.1, 2.4.9, and KD131, and the FNR consensus-like sequence within the hemA upstream sequences is completely conserved, the PrrA target sequences are not conserved. Thus, it may be that the differences in the patterns of hemA transcription between those strains are due to differences in PrrA binding.

Until now, little progress had been made in understanding the role of hemT in the cell due to the fact that hemT is transcriptionally silent under all growth conditions in strain 2.4.1. While an investigation of nitrosative stress had suggested that hemT transcription is responsive to this stressor (2), those findings could not be confirmed here. That both cis- and trans-acting mutations are known to increase hemT transcription (40) might explain the inconsistency between that report and the present findings. The need for a hemT gene in order to mount a response to nitrosative stress is also inconsistent with the profile of hemT transcription in wild type strain 2.4.9, in which the major difference was the strong upregulation of hemT transcription when grown under anaerobic-dark conditions with DMSO.

In a study that paralleled this investigation, the kinetic properties of ALA synthase enzymes were determined. The specific activity of 2.4.3-HemA is nearly 2-fold greater than that of 2.4.1- or 2.4.9-HemA (22). Since hemT seems to be specifically required for anaerobic-dark growth with DMSO, a growth modality that is not available to this strain, apparently mild
anaerobic induction of the \textit{hemA} gene encoding a more active enzyme is sufficient to meet increasing ALA requirements for phototrophic growth.

This study also revealed that all HemA enzymes are sensitive to feedback inhibition by heme while HemT enzymes are far less inhibited; for 2.4.9-HemA, the apparent $K_i$ for heme is 13-fold lower than that of 2.4.9-HemT (22). Therefore, the presence of HemT would clearly mitigate a reduction in HemA activity in cells having inhibitory levels of heme. Thus, it would seem that the role of HemT is to augment ALA production in the presence of heme concentrations that are inhibitory for HemA activity. In order to understand why the need for HemT is apparently limited to anaerobic-dark growth conditions, it is necessary to consider the types of heme that are present in aerobic \textit{versus} anaerobic cells. The A form of heme is essential for aerobic respiration, as the electron transport chain terminates in an \textit{aa}_3\text{-type cytochrome }c\text{ oxidase}. Respiration with DMSO relies upon an electron transport chain having only B-type heme, which predominates in the anaerobic cell because heme A is only synthesized in the presence of oxygen (16). If feedback regulation for HemA is exclusive to the B form of heme, then the anaerobic cytoplasm would be far more inhibitory for HemA enzyme activity than the aerobic cytoplasm, which then explains the need for HemT under anaerobic conditions.

However, this scenario of varying ratios of heme A:heme B dictating the requirement for different kinds and amounts of ALA synthases remains to be tested since heme A is not commercially available. It is noteworthy that an early study of ALA synthase activity purified from \textit{R. sphaeroides} demonstrated that, unlike heme B, bacteriochlorophyll \textit{a} had no effect on enzyme activity (5). This suggests that the presence of an isoprene tail precludes inhibition of the enzyme by the tetrapyrrole ring. If true, then heme A might likewise be unable to bind and inhibit HemA, since it is also a tetrapyrrole modified by an isoprene tail. Such a finding would
be important not only in terms of understanding *R. sphaeroides* physiology, but also for other organisms that rely upon ALA synthases for heme synthesis, which extends to humans (16).
References


12. **Fales, L., L. Nogaj, and J. Zeilstra-Ryalls.** 2002. Analysis of the upstream sequences of the *Rhodobacter sphaeroides* 2.4.1 *hemA* gene: *in vivo* evidence for the presence of two promoters that are both regulated by fnrL. Photosynthesis Research **74:**143-151.


CHAPTER III: ROLE OF SIGMA FACTORS RSPH17029_3536 AND RSPH17029_3603 IN TRANSCRIPTION OF THE *R. sphaeroides* hem*T* GENE

**Introduction**

**Transcription in Bacteria.** Transcription is a three-stage process consisting of initiation, elongation, and termination. For initiation to occur the core RNA polymerase (RNAP), consisting of subunits ββ’α2ω, must interact with a sigma factor protein. Described here are events involving the essential housekeeping σ70 sigma factor, which transcribes genes necessary for vegetative growth.

During the initiation stage, sigma factors, including σ70, recruit RNAP from a limited pool to their target genes through DNA sequence recognitions of the promoter, which is located upstream of the transcriptional +1 start site (Figure 10). In this way, the sigma factor correctly orients the RNAP on the DNA. Once RNAP holoenzyme (ββ’α2ωσ) has bound to the DNA template, the sigma factor causes key conformational changes to occur, including the melting of the duplexed DNA at the -10 promoter region and insertion of the template strand into the RNAP active site, before elongation can begin (17). The elongation stage begins with the formation of the first phosphodiester bond and promoter escape, after which the sigma factor dissociates from the complex and diffuses away to recruit a second RNAP to initiate another round of transcription (17).
Figure 11. Schematic diagram of the canonical $\sigma^{70}$-RNA polymerase (RNAP) holoenzyme ($\alpha_2 \beta \beta' \omega \sigma$) and its corresponding promoter structure (adapted from 7). The canonical $\sigma^{70}$ consists of four conserved regions $\sigma_1$-$\sigma_4$ (colored). When free within the cell region $\sigma_1$ (blue) is positioned such that DNA binding is occluded. Region $\sigma_2$ (green) is required for recognition of the promoter -10 element (green), while $\sigma_4$ (red) recognizes the -35 element (red). Region $\sigma_3$ (orange) is important in stabilizing the RNAP holoenzyme-DNA binding complex on DNA sequences comprising promoters having a -35 element that varies considerably from the sequence recognized by $\sigma^{70}$. In such cases, region $\sigma_3$ interacts with a dinucleotide sequence (TG) directly upstream of -10, which is often referred to as the extended -10 element (orange). Thus, the sigma factor protein correctly positions and orients the catalytic core of RNAP so that it overlaps with the transcriptional start site (+1) (17).
In addition to $\sigma^{70}$, Bacterial genomes encode a variable number of other sigma factor proteins. These so-called alternate sigma factors belong to one of two families, the $\sigma^{70}$ family or the $\sigma^{54}$ family. In the model organism *Escherichia coli* there a total of seven sigma factor subunits, six belong to the $\sigma^{70}$ family and one belongs to the $\sigma^{54}$ family (21). Each of these recognize different promoter sequences; however, all $\sigma^{70}$ family members recognize -35 and -10 elements, while $\sigma^{54}$ proteins recognize consensus sequences positioned -24 and -12 basepairs upstream of the transcription start site.

Proteins within the group I $\sigma^{70}$ family are found in all Bacteria. Members can be divided into subfamilies (or groups) I-IV based on the presence or absence of four conserved domains, $\sigma_1$ to $\sigma_4$ (Figure 11). Domains $\sigma_2$ and $\sigma_4$ are necessary for the recognition of the -10 and -35 promoter elements respectively, while domain $\sigma_3$ is thought to participate in binding to the core RNAP, as well as aid in recognition of promoters with an “extended -10” element (T GnTATAAT; 4, 17;). Domain $\sigma_1$ is present only in group I sigma factors, and it acts to occlude region $\sigma_4$, effectively inhibiting free $\sigma^{70}$ from binding to DNA (17, 35, 38, 43).

Members of group II are closely related to $\sigma^{70}$ in structure, but are dispensable for growth. The group III $\sigma^{70}$ factors are very distantly related to $\sigma^{70}$, and they lack domain $\sigma_1$ and as well as a non-conserved segment which links $\sigma_1$ to domain $\sigma_{2-4}$. Members of the group III subfamily transcribe flagellar, sporulation, and heat shock genes (28, 34, 41). Group IV $\sigma^{70}$ proteins have only the two domains necessary for promoter recognition, $\sigma_2$ and $\sigma_4$ (35). Members of group IV, also called extracytoplasmic function $\sigma$-factors (ECF), transcribe genes in response to extracytoplasmic environmental signals, such as misfolded proteins in the periplasmic space (6, 35). Typically, the activity of an ECF sigma factor is regulated by its cognate membrane-localized anti-sigma factor through formation of an $\sigma$-anti-$\sigma$ complex, which
effectively inactivates the ECF sigma factor by binding to the DNA-binding domains $\sigma_2$ and $\sigma_4$ (Figure 12). Release of the ECF sigma factor from the complex occurs when an appropriate signal is received by the anti-sigma protein. Typically causing a conformational change by which the sigma factor protein is “untethered” and free to interact with RNAP core.

Figure 12. Extracytoplasmic function (ECF) $\sigma$-anti-$\sigma$ factor interaction and signaling. ECF sigma factors ($\sigma^{ECF}$) are held inactive through protein-protein interactions with its cognate membrane-localized anti-sigma factor (Anti- $\sigma^{ECF}$). The anti-sigma factor, upon receiving an appropriate external signal, undergoes conformational changes that result in the release of the ECF sigma from the protein-protein complex. The sigma factor is then free to interact and recruit RNA polymerase (RNAP) to its target genes (16).
Members of the $\sigma^{54}$ family are found in a wide range of bacteria. In some there is more than one $\sigma^{54}$ gene, while in others there are none (24). The $\sigma^{54}$ protein, also called $\sigma^N$ because it was first identified through analysis of its role in expression of glutamine synthetase and regulation of nitrogen metabolism in enteric bacteria (18, 20). It has since been shown to be involved in pathways seemingly unrelated to one another, such as bacterial virulence, flagellar synthesis, fermentation, and carbon-utilization (37). $\sigma^{54}$ proteins are distinctive from the $\sigma^{70}$ family in that $\sigma^{54}$ sigma factors absolutely require an activator protein with ATPase activity in order to initiate transcription. Collectively, these sigma factors direct RNAP to transcribe a wide range of genes for which no underlying functional theme has yet been identified (3).

The final stage of transcription is termination and relies upon properties of the nascent RNA, or interaction of the RNA with accessory proteins. The former involves sequences within the RNA that fold into a stable (GC-rich) hairpin loop, which is followed by a stretch of uracil-rich RNA. These sequences comprise the so-called intrinsic terminator. An example of the latter is Rho protein-dependent termination in *E. coli*, in which the RNA:DNA hybrid is destabilized by Rho (17). Both kinds of termination result in the dissociation of core RNAP from its template.

The ability of any gene to be transcribed depends upon availability of the sigma factor recognizing the promoter. Furthermore, transcription levels are affected by the relative affinity of a sigma factor for the limited amount of RNA polymerase core enzyme in the cell.

**Sigma factors involved in hemT gene expression.** While the *hemT* gene of *R. sphaeroides* wild type strain 2.4.1 codes for a fully functional ALA synthase enzyme (31), it is not expressed under any of the standard growth conditions used for culturing this organism, which include aerobic and anaerobic conditions in the dark, as well as phototrophic conditions
The hypothesis previously put forth to explain the presence of the hemT gene in that strain was that it fulfills a specialized role in the cell, and so is only transcribed under conditions that were not known (31, 32, 45). Recently it was reported that hemT transcription was responsive to nitrosative stress based on transcriptomic data; however, this could not be confirmed using a hemT::lacZ reporter plasmid (Chapter II, Figure 7). Therefore, any special conditions that may be required for the expression of hemT in strain 2.4.1 are unknown. The findings presented in the previous chapter using hemT::lacZ transcription reporter assays in wild-type strain 2.4.1 versus wild type 2.4.9 now provide an alternate explanation for the absence of hemT transcription in strain 2.4.1 as is it now known that there are genetic differences between the two strains which are responsible for the differences in hemT expression (Chapter II, Figure 6).

Among the possible sequence differences between strains 2.4.1 and 2.4.9 that could account for the differences in hemT transcription are those encoding any sigma factor that is present in 2.4.9 (also known as ATCC17029) but absent from 2.4.1. A survey of the genomic sequences of wild type strain 2.4.1 and 2.4.9 with respect to their sigma factor genes revealed that wild type strain 2.4.9 has two such genes, rsph17029_3536 and rsph17029_3603, that are not present in wild type strain 2.4.1. Recently, a fourth wild type strain of R. sphaeroides, KD131, has been sequenced (26). As is true of both 2.4.1 and 2.4.9, it also has a hemA and a hemT gene. One of the two sigma factor genes present in 2.4.9 and absent in 2.4.1, the product of rsph17029_3603, is also represented in KD131 and is 99% identical at the amino acid sequence level.
The present study is directed toward determining whether one or the other, or both, of the 2.4.9 sigma factors contribute to hemT expression. It will also investigate whether their role in hemT transcription, if any, is direct or indirect.

Materials and Methods

Bacterial strains, plasmids and growth conditions, measurements of culture densities, DNA treatments and manipulations, transformations and conjugations, β-galactosidase assays, and protein assays. Some of the strains and plasmids used in this study have been described previously (Table 1, Chapter II), and others are listed in Table 2. Media, culturing methods, and measurements of cell growth were as previously described (Chapter II), as were protocols used for DNA treatments and manipulations, bacterial genetics, and biochemical assays.

Construction of lacZ transcription reporter plasmids with altered hemT sequences. Plasmid pLD1 (Table 2) was constructed by isolating an EagI-XbaI fragment having 2.4.9-hemT upstream sequences that had been altered using oligonucleotide directed mutagenesis (p2.4.9-hemT*NarI; 8) and inserting them into NotI- and XbaI-restricted pCF1010 (27). The sequence changes are shown in Figure 15. Plasmid pLD2 (Table 2) was constructed by an internal deletion such that the hemT*NarI upstream sequences present in pLD1 beginning 105 bp upstream of the translation initiation codon was positioned in from of lacZ in pCF1010 (8).

Construction of mutant strains JZ5323 and JZ5321. The gene sequences encompassing 316 bp of the rsph17029_3536 (NCBI Gene nomenclature) open reading frame were amplified by PCR using primers Sigma3536-UP and Sigma3536-DOWN (Table 3). The resulting PCR product was ligated in the EcoRV site of pUI1087. Following sequence
confirmation of this recombinant plasmid, pNC31 (Table 2), a KpnI-PstI fragment was isolated and inserted into KpnI- and PstI-treated pKNOCK-Km, creating plasmid pNC33 (Table 2). An analogous approach was used for construction of pNC32 (Table 2), by first inserting a 345 bp region of the rsph17029_3603 (NCBI Gene nomenclature) open reading frame amplified from genomic DNA using primers Sigma3603-UP and Sigma3603-DOWN (Table 3) into the EcoRV site of pUI1087. The resulting plasmid, pNC30 (Table 2) was then digested with PstI and KpnI to release the fragment, which was ligated with PstI- and KpnI- restricted pKNOCK-Km.

The suicide vectors pNC33 and pNC32 were mobilized individually into strain 2.4.9. Candidate exconjugants in which the entire plasmids were integrated into the chromosomal copy of rsph17029_3536 or rsph17029_3603 via recombination between the homologous sequences located on the plasmid, were selected for on plates containing kanamycin. The correct mutants were confirmed by amplification of the region, followed by sequencing of the PCR product.

All PCR reactions were performed using MyCycler thermocycler (Bio-Rad, Hercules, CA). Oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA).

Construction of expression vectors, and purification of recombinant polyhistidine-tagged RSPH17029_3536 and RSPH17029_3603 proteins. Artificial genes encoding C-terminus polyhistidine-tagged versions of RSPH17029_3536 and RSPH17029_3603 proteins were purchased from IDT and inserted into the pET3a plasmid vector, in which expression of the sigma factor genes relies upon transcription from a T7 RNA polymerase-dependent promoter that itself is expressed from a lac promoter with a lacI operator sequence (42). The plasmids were then transformed into E. coli strain BL21(λDE3)(pLysS). In this strain, the T7 polymerase gene is carried on the integrated lambda phage DE3. The pLysS plasmid encodes a T7 lysozyme
that inhibits T7 RNA polymerase activity in the cell, keeping T7-dependent transcription maximally suppress until such time as it is induced by the addition of isopropyl β-D-1-thiogalactoside (IPTG) (30, 42).

The plasmid-bearing *E. coli* strains were inoculated from fresh overnight cultures and grown at 37°C until an OD600 of 0.6 was achieved. Expression was induced by the addition of 0.5 mM IPTG, and the cultures were subsequently incubated for 2 more hours at 37°C. The cells were pelleted and then re-suspended in lysis buffer (20 mM tricine, 150 mM NaCl, 20 mM imidazole, 10% glycerol; pH 7.4) and lysed by passage through a French Pressure Cell (Thermo-Scientific, Inc.) at 700 lbs/in². Insoluble debris was removed by centrifugation. To confirm the presence of RSPH17029_3536 and RSPH17029_3603 proteins, 15 µl of cell free lysate was subjected to SDS-PAGE using an 8-16% Tris-Glycine protein gel. The gel was subsequently stained with Coomassie Blue, or proteins were transferred from the gel to BioTrace NT, pure nitrocellulose blotting membrane (Pall Corp.) for detection of the polyhistidine-tag using the SuperSignal West HisProbe Kit (Pierce Biotechnology) following manufacturer's instructions.

The remainder of the cell-free lysate was incubated with 1 ml of nickel-affinity resin (Clontech) in a sealed disposable column for 1 hour at 4°C on an orbital shaker. The column was then clamped vertically, and the following purification scheme was applied: Unbound protein was removed by washing the column with 10 column volumes (CV) of wash buffer (20 mM tricine, 150 mM NaCl, 40 mM imidazole, 10% glycerol; pH 7.4). The bound protein was eluted using elution buffer (20 mM tricine, 150 mM NaCl, 350 mM imidazole, 10% glycerol; pH 7.4) and collecting 1 ml fractions. Peak protein fractions were concentrated as needed using Amicon Ultra Centrifugal 10K MW Cutoff filters at 4°C.
Electrophoretic mobility shift assays. EMSAs involving double-stranded DNA were performed using annealed, complementary end-labeled oligonucleotides purchased from IDT. The 60-mer oligonucleotides corresponding to hemT upstream sequences were HemT-UP and its complement HemT-DOWN (Table 3). Annealing was performed as stated previously (43) using a thermocycler. Annealing of unlabeled competitor hemT DNA having the same sequence as the biotin-labeled 60-mer corresponding to the hemT upstream sequences were generated as described previously (43). For EMSAs involving single-stranded DNA, HemT-UP-Biotin and the 51-mer oligonucleotide corresponding to the rsph17029_0267 upstream sequence was UP-0267-Biotin (Table 3) were used; competitor DNA was an identical unlabeled hemT oligonucleotide.

Mobility shift assays were carried out according to manufacturer’s instruction using the LightShift™ Chemiluminescent EMSA Kit (ThermoFisher-Scientific, Waltham, MA) with 50 ng/µl poly(dA·dT) (Sigma-Aldrich Inc.) as recommended for high G+C genomes to reduce nonspecific binding. The binding reactions mixtures contained 10 fmols of biotin-labeled double stranded-DNA, or 20 fmols of single stranded-DNA oligonucleotides, 1-2µg of total protein for RSPH17029_3536 and RSPH17029_3603 and 50 ng/µl poly(dA·dT) in a total volume of 20 µl. A 22-fold, 100-fold and 800-fold molar concentration of unlabeled competitor DNA was used in competition assays when stated. All reactions were incubated for 20 minutes at room temperature followed by addition of 5 µl loading buffer (LightShift™ Chemiluminescent EMSA kit). Electrophoresis and transfer to nylon membrane were carried out at room temperature following manufacturer’s instructions in TBE buffer (ThermoFisher-Scientific, Waltham, MA). Fluorescence detection was by exposing the processed membranes to Kodak BioMax XAR film (Sigma-Aldrich Inc.).
The unlabeled hemT competitor DNA used in some EMSAs was generated by restriction with EagI and XbaI of plasmid p2.4.9-hemT-UP, which is pUI1087 with 404 bp of hemT sequences upstream of the translation start site (Table 2). A total of 2 pmol of the 410 bp EagI-XbaI product was recovered using Zymoclean DNA Gel Recovery Kit (Zymo Research, Irvine, CA) based upon the Kodak 1D Image Analysis Software and using a total of 1 µg of NEB 2.0-log marker as a standard.

**Bioinformatic analyses.** Software used included the BLAST server at NCBI ([www.ncbi.nlm.nih.gov; 1, 13](http://www.ncbi.nlm.nih.gov)), and ECFfinder ([http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder; 41](http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder)).
**Table 2.** Bacterial strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rhodobacter sphaeroides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4.9 (ATCC17029)</td>
<td>wild type</td>
<td>ATCC</td>
</tr>
<tr>
<td>JZ5018</td>
<td>hemT::ΩSp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;; Tc&lt;sup&gt;S&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>JZ5323</td>
<td>2.4.9::(pKNOCK-Km::rsph17029_3536)</td>
<td>This study</td>
</tr>
<tr>
<td>JZ5321</td>
<td>2.4.9::(pKNOCK-Km:: rsph17029_3603)</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Escherichia coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>F- (ϕ80dlacZΔM15) recA1 endA1 hsdR17 supE44 thi-1 gyrA96 relA1 deoR Δ(lacZYA-argF)U169</td>
<td>22</td>
</tr>
<tr>
<td>DH5aphe</td>
<td>DH5α with phe::Tn10d; Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>9</td>
</tr>
<tr>
<td>HB101</td>
<td>F- Δ(gpt-proA)62 leuB6 supE44 ara-14 galK2 lacY1</td>
<td>2</td>
</tr>
<tr>
<td>S17-1</td>
<td>Conjugal donor; C600::RP4 2-</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>(Tc::Mu)(Km::Tn7) pro res&lt;sup&gt;−&lt;/sup&gt; mod&lt;sup&gt;+&lt;/sup&gt; (Tp&lt;sup&gt;R&lt;/sup&gt; Sp&lt;sup&gt;R&lt;/sup&gt;)</td>
<td></td>
</tr>
<tr>
<td>BL21(λDE3)</td>
<td>pLysS, Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>30, 42</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBSIISK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>ColE1; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Agilent</td>
</tr>
<tr>
<td>pUII1087</td>
<td>BSIISK&lt;sup&gt;+&lt;/sup&gt; with modified polylinker; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>45</td>
</tr>
<tr>
<td>pCF1010</td>
<td>RSF1010 derivative; used for creating lacZ transcriptional fusions; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>26</td>
</tr>
<tr>
<td>Vector/Expression Vector</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>pUI1009</td>
<td>pSUP202-derived <em>hemT</em>::ΩSp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pSUP202</td>
<td>pBR325 derivative, Mob&lt;sup&gt;+&lt;/sup&gt;, Ap&lt;sup&gt;R&lt;/sup&gt; Cm&lt;sup&gt;R&lt;/sup&gt; Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pET-3a</td>
<td>N-terminal T7-Tag expression vector, Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pKNOCK-Km</td>
<td>Mobilizable vector for insertional mutagenesis; Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pNC1</td>
<td>2.4.9-<em>hemT</em>::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pNC3</td>
<td>2.4.9-<em>hemA</em>::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>phemT-UP(2.4.9)</td>
<td>PCR-amplified <em>hemT</em>-UP sequences from 2.4.9 inserted into the EcoRV site of pUI1087; Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pLD1</td>
<td>p2.4.9-<em>hemT</em>&lt;sup&gt;+&lt;/sup&gt;Narl&lt;sup&gt;+&lt;/sup&gt;::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pLD2</td>
<td>p2.4.9-<em>ΔhemT</em>&lt;sup&gt;+&lt;/sup&gt;::lacZ in pCF1010; Sp&lt;sup&gt;R&lt;/sup&gt;/St&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pET3a::*rsph17029_3536</td>
<td>Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pET3a::*rsph17029_3603</td>
<td>Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pNC30</td>
<td>322 bp of <em>rsph17029_3603</em> inserted into the EcoRV site of pUI1087</td>
<td></td>
</tr>
<tr>
<td>pNC31</td>
<td>316 bp of <em>rsph17029_3536</em> inserted into the EcoRV site of pUI1087</td>
<td></td>
</tr>
<tr>
<td>pNC32</td>
<td>PstI-KpnI fragment from pNC30 inserted into pKNOCK-Km</td>
<td></td>
</tr>
<tr>
<td>pNC33</td>
<td>PstI-KpnI fragment from pNC31 inserted into pKNOCK-Km</td>
<td></td>
</tr>
</tbody>
</table>

This study
Table 3. Names and sequences of oligonucleotides, and their applications in this study.

<table>
<thead>
<tr>
<th>Oligonucleotide name/number</th>
<th>Sequence</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigma3536-UP/357</td>
<td>GCCGGTCGGTATCGGGTGACC</td>
<td>With #444, amplification of <em>rshp17029</em>_3536 sequences for construction of pNC30</td>
</tr>
<tr>
<td>Sigma3536-DOWN/444</td>
<td>GCCATGTGCAGCAGTGTTACAGC</td>
<td>With #357, amplification of <em>rshp17029</em>_353 sequences for construction of pNC30</td>
</tr>
<tr>
<td>Sigma3603-UP/362</td>
<td>GGTCTGGACTGTCATTGGCGC</td>
<td>With #359, amplification of <em>rshp17029</em>_3603 sequences for construction of pNC31</td>
</tr>
<tr>
<td>Sigma3603-DOWN/359</td>
<td>CGTCGCCGAATTGCGCCGCCC</td>
<td>With #362, amplification of <em>rshp17029</em>_3603 sequences for construction of pNC31</td>
</tr>
<tr>
<td>HemT-UP (with and without 5' biotinyl)</td>
<td>CGTCAGGCGGGCGAAAAATGTCCGT</td>
<td>Biotin-labeled: EMSA binding reactions; unlabeled: EMSA competitor reactions</td>
</tr>
<tr>
<td>HemT-DOWN (with and without 5' biotinyl)</td>
<td>TTTGTATGGCCTACAATCCGGATAAAC</td>
<td>Biotin-labeled: EMSA binding reactions; unlabeled: EMSA competitor reactions</td>
</tr>
<tr>
<td>UP-0267 (5') biotinylated</td>
<td>CCGTCTGCCGAAAAAGGGAAAAATGCCCACT</td>
<td>EMSA binding reactions</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td>TAGGCACTCATGCGCAGAAT</td>
<td></td>
</tr>
</tbody>
</table>
Results

The role of RSPH17029_3536 and RSPH17029_3603 in hemT transcription. To determine whether the sigma factors RSPH17029_3536 and RSPH17029_3603 play a role in hemT transcription, mutant strains JZ5321 and 5323 in which each sigma factor were singly disrupted were constructed (Table 2). The 2.4.9-hemT::lacZ reporter plasmid was then mobilized into the mutant bacteria.

β-galactosidase assays of the exconjugants (Figure 13) supported the premise that both sigma factors contribute to hemT expression, since enzyme activities were 7-fold lower in extracts of the individual sigma mutant strains compared to activities in wild type extracts. These assays were performed using samples of cells cultured under anaerobic-dark conditions, which are optimal for hemT expression (Chapter II).
Figure 13. β-galactosidase activities of \textit{hemT} in \textit{R. sphaeroides} wild type 2.4.9 versus 2.4.9 sigma factor mutant strains. In all cases, the reporter plasmid was 2.4.9-\textit{hemT}:\textit{lacZ} and the strains are as indicated. All cultures were grown under anaerobic-dark conditions supplemented with yeast extract and dimethyl sulfoxide as alternate electron acceptor. Wild type 2.4.9 \textit{hemT}:\textit{lacZ} values are reproduced here for clarity. Values are the average of at least three replicates and ranges are as indicated.
Growth of the sigma factor mutants as they transition from aerobic to anaerobic-dark with DMSO conditions was profiled. Consistent with the transcription evidence of a role of both sigma factors in hemT expression, and the need for hemT during the transition (Figure 13), the profiles of the sigma factor mutants were comparable to the profile of the hemT null mutant strain JZ5018 (Figure 14). All of the mutant strains have a lag phase that is prolonged by several days relative to the wild type (Figure 14).

![Growth profiles of R. sphaeroides wild type 2.4.9, the hemT null mutant strain JZ5018, and the sigma factor mutants JZ5323 and JZ5321. Depicted are representative profiles of cultures inoculated with aerobically grown cells that were then incubated under anaerobic-dark conditions.](image)
Collectively, these results argue that both sigma factors are involved in \textit{hemT} transcription. Despite numerous attempts constructing a double mutant, in which both sigma factor genes are disabled, was unsuccessful. Although it remains to be proven, this negative outcome argues that the absence of both sigma factors is incompatible with viability. If true, it would suggest that the sigma factors are required for expression of essential genes, in addition to \textit{hemT}, which is not essential.

**Identification of upstream sequences important for \textit{hemT} transcription.** Alignments of -35 elements by several ECF sigma factors, some of which have known functions, revealed that many have conserved AA dinucleotides at the -33 and -32 positions (25). Unlike the interactions between other $\sigma^{70}$ proteins and their -35 elements, the ECF sigma factors do not appear to make direct contact with these dinucleotides; rather, it is thought that the AA sequence induces a specific DNA geometry that is required for ECF sigma factor binding (25, 33), which explains the conservation in DNA sequence among ECF promoters, despite the lack of amino acid sequence similarity of the $\sigma_4$ domains of these proteins.

Within the upstream sequences of \textit{hemT} there are multiple AA dinucleotides, but the most striking is the AAAAT that is also appropriately positioned upstream of a sequence corresponding to a -10-like element for many ECF sigma factors, a dinucleotide motif of TG or GT (12, 15, 29; Figure 15). To test the possibility that these sequences comprise a promoter recognized by one or the other of the 2.4.9-specific sigma factors, the 2.4.9-\textit{hemT}:\textit{lacZ} reporter plasmid pLD1 was constructed in which the AAAAT sequence was mutated to GCCAT (Table 2).
Figure 15. ECF-like promoter sequences present within the upstream sequences of hemT.

Shown is an alignment of *R. sphaeroides* 2.4.1 and 2.4.9 hemT upstream sequences with the putative ECF-like promoter highlighted in yellow. The altered sequences introduced by oligonucleotide-directed mutagenesis are shown under the alignment, with the changes indicated in blue.

β-galactosidase assays of wild type 2.4.9 and sigma factor mutant exconjugants of plasmid pLD1 were performed. The sequence alterations resulted in a 6-fold decrease in transcription in wild type strain 2.4.9 (Figure 16), suggesting the AAAAT-N₁₉-TTGT motif is a possible promoter for one or the other, or both ECF sigma factors. A further 4-fold decrease in transcription was observed in each sigma factor mutant strain relative to the wild type (Figure 16), indicating that the sequences are important for transcription involving both of the sigma factors, a finding that predicts both RSPH17029_3536 and 3603 directly transcribe hemT.
Figure 16. Investigation of an ECF-like promoter within the hemT upstream sequences. Assays of β-galactosidase activities were performed using extracts of wild type strain 2.4.9, and sigma factor mutant strains JZ5323 and JZ5321 (Table 2) with the lacZ transcription reporter plasmid pLD1 (Table 2) carrying mutant sequences in the putative -35 promoter region (depicted in Figure 15). All cultures were incubated under anaerobic-dark conditions with dimethyl sulfoxide as alternate electron acceptor. Wild type 2.4.9 hemT::lacZ values are reproduced here for clarity. Values are the averages of at least three replicates with standard deviations indicated.
To eliminate any potential promoter within the sequences upstream of the NarI site in pLD1, the NarI site was used to truncate the hemT sequences, creating pLD2 (Table 2) in which the SpR/StR transcription-translation termination Ω cassette is positioned immediately adjacent to the remaining hemT sequences. Plasmid pLD2 was then also mobilized into wild type strain 2.4.9 and the sigma factor mutant strains JZ5323 and JZ5321 (Table 2), and β-galactosidase assays were performed in extracts of exconjugants grown under anaerobic-dark conditions with DMSO. Results in the wild type strain with pLD2 revealed a further 3-fold decrease in hemT expression relative to 2.4.9 with pLD1; no further decrease in activity was observed when pLD2 was introduced into the sigma factor mutant backgrounds (Figure 17). These results reinforce the idea that the A-rich sequence upstream of hemT is important for transcription by both sigma factors, and suggests that sequences upstream of the NarI site are not involved.
Figure 17. Investigation of ECF-like promoter sequences within the hemT upstream sequences. Assays of β-galactosidase activities were performed using extracts of wild type strain 2.4.9, and sigma factor mutant strains JZ5323 and JZ5321 (Table 2) with the lacZ transcription reporter plasmid pLD2 (Table 2). All cultures were grown under anaerobic-dark conditions with dimethyl sulfoxide as alternate electron acceptor. Wild type 2.4.9 hemT::lacZ values are reproduced here for clarity. Values are the averages of at least three replicates with standard deviations indicated.
The upstream sequences of *hemA* also contain an ECF-like promoter sequence (Figure 18). To determine whether the sigma factors RSPH17029_3536 and RSPH17029_3603 play a role in 2.4.9-*hemA* transcription, β-galactosidase assays were performed in extracts of the wild type and each of the sigma factor mutants bearing the 2.4.9-*hemA::lacZ* reporter plasmid pNC3 (Table 2) that were cultured anaerobically in the dark with DMSO.

![Figure 18. Upstream sequences of hemA containing the putative ECF-like promoter motif AAAAT-N17-TTGT found upstream of hemT (yellow).](image)

β-galactosidase activities in extracts of wild type *versus* sigma factor mutant bacteria with plasmid pNC3 did not significantly differ from each other (Figure 18), based on a T-test comparing the wild type as a control to each of the individual mutant strains; the *p*-value for JZ5323(pNC3) versus 2.4.9(pNC3) was 0.67, and the *p*-value for or JZ5321(pNC3) *versus* 2.4.9(pNC3) was 0.87. Therefore, *hemA* transcription is apparently independent of these sigma factors.
Figure 19. Analysis of the role of the sigma factors in expression of 2.4.9 hemA. Assay of β-galactosidase activity were performed with extracts of wild type 2.4.9, sigma factor mutant strains JZ5323 and JZ5321 (Table 2) with reporter plasmid pNC3 carrying 2.4.9-hemA upstream sequences positioned in front of a promoterless lacZ gene (Table 2). All bacteria were cultured under anaerobic-dark conditions with DMSO as an alternate electron acceptor. Values are the averages of at least three replicates with standard deviations indicated.

Direct binding of RSPH17029_3536 and RSPH17029_3603 to hemT upstream sequences. The data presented above are compatible with the possibility that the ECF sigma factors of R. sphaeroides 2.4.9, RSPH17029_3536 and RSPH17029_3603, are binding to overlapping promoters, as (1) the absence of either sigma factor results in a reduction of hemT transcription, and (2) mutations targeted to the AAAAT region of the hemT upstream sequences
further reduce \textit{hemT} transcription in each sigma factor mutant. To examine this possibility, \textit{in vitro} biochemical experiments using electrophoretic mobility shift assays (EMSAs) were performed.

Synthetic sigma factor genes, modified by the addition of DNA sequences coding for C-terminal poly-histidine tags were inserted into the pET3a expression plasmid, in which the genes are transcribed from an IPTG-inducible T7 polymerase-dependent promoter. The sigma factor genes were expressed in \textit{E. coli}, and the recombinant proteins were purified using a nickel-affinity gravity column. Confirmation that intact, polyhistidine-tagged sigma factor protein was being expressed was obtained by probing membranes prepared from cell free lysates with a probe that is nickel (Ni$^{2+}$)-activated and conjugated to a derivative of horseradish peroxidase (HRP) enabling direct, IMAC-based detection of His-tagged proteins (Figure 20).

For the gel shift assays, each purified sigma factor was incubated with (biotin-labeled) annealed complementary oligonucleotides (Table 3) that were 60 base pairs in length and correspond to \textit{hemT} upstream sequences that encompass the AAAAT-N$_{19}$-TTGT region (Figure 24).
Figure 20. Detection of poly(his)-tagged fusion proteins RSPH17029_3536 and RSPH17029_3603. (A) Coomassie Blue stained SDS-PAG (8-16% tris-glycine gradient gel) of cell free lysates of *E. coli* strain BL21(λDE3)(pLysS) with plasmids pET3a-rsp17029_3536-6XHis and pET3a-rsp17029_3536-6XHis (Table 2). The polyhistidine-tagged RSPH17029_3536 protein has a calculated molecular mass of approximately 20 kDa in size and the calculated molecular mass of polyhistidine-tagged RSPH17029_3603 is approximately 21 kDa. Included as a control is a sample of the cell free lysate of *E. coli* strain BL21λDE3(pLysS) with the pET3a vector. The lane labeled "X" denotes a sample lane that is not relevant; M: BioRad Kaleidoscope Precision Plus Marker with sizes indicated. (B) Detection of polyhistidine-tagged protein present the same cell free lysates as in (A) by immunoblot using a nickel activated His-probe conjugated to horseradish peroxidase; M: Benchmark polyhistidine-tagged protein standard (Invitrogen, Thermo-Fisher Sci.) with sizes indicated.
When either sigma factor was incubated with labeled double-stranded DNA corresponding to the hemT upstream sequences, a shift in DNA mobility through the gel was observed, indicating that both sigma factors are directly binding to hemT upstream sequences (Figure 21, Lanes 2 and 5). To confirm sequence specificity, each sigma factor was also incubated with a 800-fold molar excess of unlabeled DNA corresponding to the hemT sequences (Table 3), as well as the hemT biotin-labeled DNA.

Figure 21. Electrophoretic mobility shift assays using purified RSPH17029_3536 and RSPH17029_3603 incubated with biotin-labeled double stranded hemT upstream sequences that include the ECF-like promoter sequence AAAAT-N19-TTGT. (A) EMSA using purified RSPH17029_3536 protein. Lane 1: 0.8 nM labeled hemT only, lane 2: RSPH17029_3536 purified protein and 0.8 nM labeled hemT DNA, lane 3: RSPH17029_3536 purified protein and 0.8 nM labeled hemT DNA with 320 nM of unlabeled hemT DNA (800-fold molar excess). (B) EMSA using purified RSPH17029_3536 protein. Lane 4: 0.8 nM labeled hemT only, lane 5: RSPH17029_3603 purified protein and 0.8 nM labeled hemT DNA, lane 6: RSPH17029_3603 purified protein and 0.8 nM labeled hemT DNA with 320 nM of unlabeled hemT DNA (800-fold molar excess).
The signal generated by RSPH17029_3536 bound to the labeled DNA in the absence of competitor (Figure 21 lane 2) was low. However, sequence-specific binding by this sigma factor was confirmed by performing an EMSA in which the competitor DNA used was a restriction fragment of *phemT*-UP(2.4.9) (Table 2) that contains the same upstream sequences of *hemT* (Figure 22) as the annealed oligonucleotides.

![Figure 22. Competition EMSA using RSPH17029_3536, labeled double stranded DNA corresponding to the *hemT* upstream sequences, and an unlabeled restriction fragment isolated from plasmid *phemT*-UP(2.4.9) that contains these same sequences. Lane 1: RSPH17029_3536 purified protein and 10 nM labeled *hemT* DNA, lane 2: RSPH17029_3536 purified protein and 10 nM labeled *hemT* DNA with 220 nM of unlabeled *hemT* DNA generated by restriction digest of *phemT*-UP(2.4.9) (22-fold molar excess), lane 3: 10 nM labeled *hemT* only.](image)
Assays using double stranded labeled and unlabeled hemT oligonucleotides proved difficult to reliably reproduce. In transcription initiation, sigma factors contribute to DNA melting (7, 19). Studies interested in promoter DNA recognition by sigma factors have revealed that, while free sigma factors are unable to initiate transcription without RNAP, they can bind to DNA in its absence (10, 11). Research by Huang, et al. concluded that in Bacillus subtilis when RNAP is complexed with σ^A it preferentially bound single stranded DNA over double stranded DNA, and that the -10 element of the promoter region was key to this interaction (19). Therefore, it is possible that the purified proteins can bind double-stranded DNA species, and also single stranded DNA that might be present in the labeled and/or unlabeled DNA in the reaction mixture. Toward improving the ability to reliably reproduce the results, the EMSA assays were repeated using single-stranded DNA.

Although there were multiple shifted species observed for both RSPH17029_3536 and RSPH17029_3603 (Figure 23), only one of them could be competed by unlabeled DNA. This species corresponds to the sigma factor that is sequence-specifically bound to hemT labeled DNA (Figure 23, arrow).
Figure 23. Electrophoretic mobility shift assays using purified RSPH17029_3536 and RSPH17029_3603 incubated with biotin-labeled single stranded hemT upstream sequences that include the ECF-like promoter sequence AAAAT-N$_{19}$-TTGT. In all reactions the 10 nM of labeled oligonucleotide HemT-UP-Biotin (Table 3) was used. For reactions with competitor DNA 5 mM of the unlabeled single stranded hemT oligonucleotide having the same DNA sequence as HemT-UP-Biotin at 500-fold molar excess was used. The arrow identifies the shift due to sequence specific-binding that is absent when the assay is performed in the presence of unlabeled single stranded hemT sequences.
Survey of transcriptomic data for wild type 2.4.9 versus sigma factor mutants

**JZ5321 and JZ5323.** An RNA-seq analysis of wild type 2.4.9 and mutant strains JZ5321 and JZ5323 has been performed (Table 2; 46). As is true of hemT, which is transcribed by ECF sigma factors RSPH17029_3635 and 3603, the hemT transcript levels are low in both JZ5321 and JZ5323. The transcript levels of other genes that are similarly transcribed in *R. sphaeroides* 2.4.9 should also be low, or absent from both the sigma factor mutant strains.

The set of genes which are statistically significantly down-regulated in the absence of either ECF sigma factor included hemT as expected, and also rsp17029_0267-0269 that code for polypeptide components of a C₄-dicarboxylic acid TRAP transporter. The product of the first gene in the operon, rsph17029_0267, codes for an extracytoplasmic solute binding protein that is thought to deliver its substrate to its partner tripartite ATP-independent periplasmic (TRAP) transporter proteins. The C₄-dicarboxylates include succinate, fumarate and malate (24). As yet, no other metabolic theme that might include hemT has been discerned among the other genes in the set.

An alignment of the upstream sequences of hemT with the upstream sequences of rsph17029_0267 revealed that the AAAAT sequence at the -35, and the TG sequence at the -10 of the hemT putative promoter region is also present within the rsph17029_0267 upstream sequences (Figure 20). However, it should be noted that, despite the presence of a similar sequence within the hemA upstream sequences, there was no difference in hemA mRNA levels according to the RNA-seq data; this is consistent with assays using the hemA::lacZ transcription reporter plasmid pNC3 (Figure 19).
Figure 24. Alignment of upstream sequences from genes 2.4.9-hemT and rsph17029_0267 that were used in electrophoretic mobility shift assays. The DNA corresponding to the hemT upstream sequences is 60 nucleotides long; that corresponding to the rsph17029_0267 upstream sequences is 51 nucleotides long. Highlighted regions are the sigma factor putative -35 and -10 promoter binding sites.

EMSAs were performed to determine whether or not the sigma factors RSPH17029_3536 and RSPH17029_3603 act directly in transcription of rsph17029_0267. Using single-stranded biotin-labeled rsph17029_0267 a shifted complex was observed for both sigma factors (Figure 25), and each of these complexes were competed by unlabeled single-stranded DNA corresponding to the hemT upstream sequence (Figure 25).
Figure 25. Electrophoretic mobility shift assays using purified RSPH17029_3536 and RSPH17029_3603 incubated with biotin-labeled single stranded *rsph17029_0267* upstream sequences that include the ECF-like promoter sequence AAAAT-N₁₇-TG. In all reactions the 10 nM of labeled oligonucleotide UP-0267-Biotin was used. For reactions with competitor DNA 5 mM of the unlabeled single stranded *hemT* oligonucleotide for 500-fold molar excess was used when indicated. The arrow identifies the shift due to sequence specific-binding that is absent when the assay is performed in the presence of unlabeled single stranded DNA corresponding to the *rsph17029_0267* sequences.

Since the sigma factors were not purified to homogeneity (Figure 20), any of the additional shifted bands could be due to interactions between proteins other than the sigma factors and the DNA. Possible contaminating proteins that might be present include *E. coli* sigma factors, as well as single-stranded binding proteins, all of which are similar in molecular
weights (37) and would give rise to similarly shifted complexes with approximately the same 
migrations. As is apparently true of the \textit{R. sphaeroides} sigma factors, these protein contaminants 
would likely be capable of binding to single-stranded DNA, but this binding would be sequence 
non-specific. Therefore, only the complexes that are competed away by the presence of non-
labeled DNA are the relevant \textit{R. sphaeroides} sigma factor-DNA complexes.

\textbf{Discussion}

This study was prompted by the discovery that the \textit{hemT} gene is actively transcribed in \textit{R. sphaeroides} strain 2.4.9, but not in strain 2.4.1 (presented in Chapter II). The possibility that two 
sigma factors that are present in strain 2.4.9 but absent from strain 2.4.1 might be responsible for 
this difference in \textit{hemT} expression was investigated.

The level of \textit{hemT} transcription measured in mutant strains disabled in one or the other of 
the sigma factor genes was significantly reduced relative to that in wild type strain 2.4.9 
indicating both sigma factors are important. That the same alteration to the \textit{hemT} upstream 
sequences negatively impacts transcription in either sigma factor mutant suggests that both 
proteins are directly involved in \textit{hemT} transcription, and that those sequences encompass the 
\textit{hemT} promoter recognized by the sigma factor proteins.

As would be expected if \textit{hemT} transcription requires these sigma factors, the sigma factor 
mutants have a prolonged lag phase in transitioning from aerobic to anaerobic-dark respiration 
on DMSO. This is the same pattern of growth as was observed for the \textit{hemT} null mutant strain 
(Chapter 2). Thus, consistent with the transcription measurements, the sigma factors are 
implicated in expression of \textit{hemT}. 
The inability to construct a sigma factor double mutant in strain 2.4.9 indicates that the presence of at least one of the two sigma factors is required for survival. Since hemT is not essential, it suggests additional genes are transcribed by one or the other, or both of these sigma factors that are indispensable for growth. The presence of such additional genes was supported by RNA-seq data, in that there are several genes whose RNA levels are significantly lower in the sigma factor mutants relative to levels in wild type cells. One of those is rsph17029_0267, which is annotated as a C4-dicarboxylate periplasmic binding protein. Whether this gene, or other genes whose normal transcription requires these sigma factors based on RNA-seq data, is essential remains to be determined.

The EMSAs fully support the in vivo evidence that both sigma factors directly transcribe the hemT gene, and also rsph17029_0267. Sequence similarities within the upstream sequences of these genes suggest that the sigma factors bind to an AA-rich -35 region. This is as expected, since this sequence confers a specific conformation to the DNA that is associated with ECF sigma factor binding (25). The identification of two such sequences will be useful in bioinformatic approaches, combined with the RNA-seq analysis, toward determining the entire complement of genes having promoters transcribed by these sigma factors. Presumably, by knowing all such genes, unifying themes as to the function of their products can guide the identification of the specific signal to which the sigma factors respond, thereby more fully defining their roles in R. sphaeroides.
References


CHAPTER IV: INVESTIGATING THE ROLE OF ECF SIGMA FACTORS

RSPH17029_3536 AND RSPH17029_3603

Introduction

The members of the Group IV subfamily of sigma 70 proteins were named the extracytoplasmic function (ECF) sigma factors by Lonetto et al. in 1994 (18). This group is distinguished from other groups of sigma 70 proteins by virtue of having only the two DNA-binding domains among the four protein domains that are present in the archetypal housekeeping sigma 70 protein. The ECF sigma factor group is the largest and most diverse group of sigma factors, consisting of 43 distinct sub-groups, with many still unclassified as to function using Hidden Markov Model-based analysis (8, 27). ECF sigma factors are involved in transcribing sets of genes that make it possible for the cell to effectively and appropriately respond to environmental cues generated outside of the cell or within the periplasm. These cues can range from changes in nutrient availability, such as the presence of useful forms of iron (E. coli FecI; 5, 19) to "life-threatening" stressors such as oxidative stress (R. sphaeroides σE; 3) and cell envelope stressors like heat shock (E. coli σE; 20, 24) and alkaline shock (B. subtilis σW; 28).

One way in which the cues are communicated to the ECF sigma factors, and by which their activity is regulated, is via membrane-localized anti-sigma factor proteins. By binding to the DNA binding domains of their cognate sigma factor, they tether the sigma factors to the membrane, making them unavailable to interact with core RNA polymerase. In this way, transcription of genes having promoter sequences recognized by the ECF sigma factors is prevented. The sigma factor becomes available when a specific signal is received by the anti-sigma factor, which results in the release of the sigma factor. This release can involve induced
conformational changes in the anti-sigma protein (21), proteolysis of the anti-sigma factor (21), or so-called "partner switching" in which another protein that is a molecular mimic of the sigma factor becomes available to compete for binding to the anti-sigma factor protein (6); the latter mechanism has only recently been identified and is known to be present in α-proteobacteria (6). The genes encoding anti-sigma factor proteins are typically located directly downstream of, and are often co-transcribed with their partner ECF sigma factor, which is useful in studies to understand the role of the sigma factors.

For many ECF sigma factors, an additional layer of control of their activities is achieved at the level of transcription regulation. For example, σE of E. coli (24) and σX and σM of B. subtilis (4), auto-regulate their own transcription via positive feedback loops. By transcribing their own gene, and thus their cognate anti-sigma factor as well, the cell is able to amplify the response to the signal which they are receiving, as well as ensure increased expression of the anti-sigma factor protein for rapid inactivation of the sigma factor once the stress has been relieved (12). Transcription of ECF sigma factor genes can also be regulated by other DNA binding proteins. For example, the E. coli FecI sigma factor gene is negatively regulated by Fur which, when loaded with Fe^{3+}, binds to the promoter region upstream of the FecI sigma factor structural gene and so represses transcription (5). Thus, in addition to responding to extracytoplasmic changes that control ECF sigma factor availability and activity via anti-sigma factor signaling events, the level of transcription by ECF sigma factors can also be responsive to cytoplasmic changes that alter transcription of the ECF sigma factor genes themselves.

The number of ECF sigma factors an organism may encode can vary widely, ranging from zero, such as in obligate intracellular bacteria including Chlamydia (27, 29), to hundreds, such as in the planctomycete Gemmata obscuriglobus (13). In organisms that encode multiple
ECF sigma factors the question of specificity presents itself; i.e. is it possible that, under the right circumstances, these sigma factors can transcribe promoters other than their primary targets? This question has been investigated by several groups. The available evidence suggests that, for sigma factors belonging to different subgroups, as defined by Staron et al., such "crosstalk" between ECF sigma factors and promoters is apparently low (8). However, instances of crosstalk are known. For example, in B. subtilis in response to cell envelope stress, ECF sigma factors from different sub-groups all transcribe a subset of core genes, with additional genes that are uniquely transcribed by each individual ECF sigma factor in response to a specific type of cell envelope stress, such as alkaline or oxidative stress (12). The overlapping of genes transcribed by multiple ECF sigma factors establishes a hierarchy of transcription responsiveness, in which genes whose products are effective in coping with multiple stressors are transcribed, while others are transcribed by individual ECF sigma factors in response to specific stressor types.

The goal of this work was to learn more about the roles of two ECF sigma factors that both directly transcribe the hemT gene of R. sphaeroides strain 2.4.9. To do so, efforts were directed toward understanding how the activities of these sigma factors are controlled. First, in order to gather information that might provide insights about these proteins and their putative roles, bioinformatic analyses were performed based on what is known in other organisms about these sigma factors and their cognate anti-sigma factors, as well as analyses of the genomic context of the sigma factors. Findings stemming from these analyses were used to formulate hypotheses that were then experimentally tested. Second, the transcriptional responsiveness of these sigma factors to C4-dicarboxylic acids was considered since, in addition to hemT, they both transcribe genes involved in transport of these compounds (Chapter III).
Materials and Methods

**Bacterial strains, plasmids and growth conditions.** Strains and plasmids used in this study are listed in Table 4. *E. coli* were grown as previously described (Chapter II). *R. sphaeroides* cultures were grown under 30% oxygen conditions as previously described (Chapter II) in Sistrom’s minimal A medium with either 34 mM succinate or with 34 mM pyruvate. For anaerobic-dark growth screw-capped tubes were completely filled with Sistrom’s succinate minimal A medium with either 34 mM succinate or with 34 mM pyruvate, and with dimethyl-sulfoxide (DMSO; final concentration was 0.06 M) as alternate electron acceptor. Where indicated, the anaerobic-dark medium was supplemented with either 0.1% yeast extract or 0.1% Bacto-Peptone (Becton, Dickinson & Co., Sparks, MD).

**Measurements of culture densities, DNA treatments and manipulations, transformations and conjugations, β-galactosidase assays, and protein assays.** Culture densities were determined as previously described (Chapter II). All DNA treatments and manipulations, transformations and conjugations, and biochemical analyses were also performed as described previously (Chapter II).

**Construction of plasmid pBBR::rsph17029_3536-C151S,C201S.** A synthetic gene fragment clone was purchased from Integrated DNA technologies in which the *rsph17029_3536* sequences encode serines rather than cysteines at positions 151 and 201. In order to facilitate the identification of recombinants during cloning, an ΩSp/St cassette (23) was first positioned downstream of the *rsph17029_3536* sequences in this plasmid. Then a ZraI-EcoRV fragment encompassing the *rsph17029_3536* coding sequences and the Sp/St resistance cassette was ligated into ZraI and EcoRV-treated pBBR::*rsph17029_3536* plasmid, which removes that portion of the wild type sequences that includes the two cysteine residues and replaces them with
the sequences coding for a mutant protein with serines at positions 151 and 201. The Sp/St resistance cassette was subsequently removed by treatment with HindIII followed by circularization of the vector fragment.

**Bioinformatic analyses.** Software used included the BLAST server at NCBI (www.ncbi.nlm.nih.gov; 2, 9), ClustalW (16), ECFfinder (http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder/; 27), and tools available at jgi.doe.gov.
Table 4. Bacterial strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strains and plasmids</th>
<th>Relevant characteristics</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rhodobacter sphaeroides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4.9 (ATCC17029)</td>
<td>Wild type</td>
<td>ATCC</td>
</tr>
<tr>
<td>JZ5323</td>
<td>2.4.9::(pKNOCK-Km::rsph17029_3536)</td>
<td>This study</td>
</tr>
<tr>
<td>JZ5321</td>
<td>2.4.9::(pKNOCK-Km:: rsph17029_3603)</td>
<td>This study</td>
</tr>
<tr>
<td>NC360</td>
<td>2.4.9 with plasmids pNC1 and pBBR-MCS2</td>
<td>This study</td>
</tr>
<tr>
<td>NC392</td>
<td>2.4.9 with plasmids pNC1 and pBBR::rsph17029_3536-C151S,C201S</td>
<td>This study</td>
</tr>
<tr>
<td>NC397</td>
<td>2.4.9 with plasmids pNC1 and pBBR::rsph17029_3536)</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Escherichia coli</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>F- (φ80dlacZΔM15) recA1 endA1 hsdR17 supE44 thi-1 gyrA96 relA1 deoR Δ(lacZYA-argF)U169</td>
<td>25</td>
</tr>
<tr>
<td>DH5αphe</td>
<td>DH5α with phe::Tn10d; CmR</td>
<td></td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCF1010</td>
<td>RSF1010 derivative; used for creating lacZ transcriptional fusions; SpR/StR, TcR</td>
<td>16</td>
</tr>
<tr>
<td>pNC1</td>
<td>2.4.9-hemT::lacZ in pCF1010; SpR/StR, TcR</td>
<td>29</td>
</tr>
<tr>
<td>pBBR1-MCS2</td>
<td>KmR</td>
<td>13</td>
</tr>
<tr>
<td>pBBR::rsph17029_3536</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td><em>rsph17029_3536</em> sequences, from 335 bp upstream of the translation start site to 44 bp downstream of the nonsense codon, amplified from 2.4.9 using primers #360 and 361, ligated into Ecl136II (pBBR-MCS2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pBBR::rsph17029_3536</th>
<th>This study</th>
</tr>
</thead>
<tbody>
<tr>
<td>-C151S,C201S</td>
<td>This study</td>
</tr>
<tr>
<td><em>rsph17029_3536</em> sequences encoding cysteines at positions 151 and 201 are replaced by sequences encoding serines; Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>
**Table 5.** Names and sequences of oligonucleotides, and their applications used in this study.

<table>
<thead>
<tr>
<th>Oligonucleotide name/number</th>
<th>Sequence</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigma3536-UP2/360</td>
<td>CATAGCCCAGCAGCTGCCCCTGGT</td>
<td>With #361, amplification of <em>rsph17029_3536</em> sequences for construction of pBBR::<em>rsph17029_3536</em></td>
</tr>
<tr>
<td>Sigma3536-DOWN2/361</td>
<td>CGCGAGCATGTCATCGGTTGTCG</td>
<td>With #360, amplification of <em>rsph17029_3536</em> sequences for construction of pBBR::<em>rsph17029_3536</em></td>
</tr>
</tbody>
</table>
Results

Bioinformatic analyses of ECF sigma factor RSPH17029_3536 and anti-ECF sigma factor RSPH17029_3537. BLAST searches for ECF sigma factor RSPH17029_3536 relatives in other organisms revealed that its distribution is limited to the α-Proteobacteria group, and no information for any relative as to function is available. However, BLAST searches of the anti-sigma factor RSPH17029_3537 identified it as member of the anti-sigma factor protein RskA superfamily. In Mycobacterium tuberculosis and M. bovis, RskA is the anti-sigma factor of ECF sigma factor SigK, which transcribes genes involved in maintaining redox-homeostasis (25).

While the actual signal to which the anti-sigma factor responds remains unknown, Shukla, et al. found that SigK itself is redox-responsive, and this stems from two cysteines that form a disulfide bond in the σ₄K domain of the protein under oxidizing conditions (25). This disulfide bridge keeps SigK anchored to the membrane by stabilizing the SigK-RskA interactions, thus inhibiting the activity of the sigma factor. Upon reduction of the disulfide bond the SigK-RskA complex is destabilized, thereby releasing SigK, which can now recruit RNAP to its target genes. Interestingly, the disulfide bond does not alter the conformation of the SigK σ₄ protein domain, suggesting that the disulfide does not contribute to -35 promoter binding; rather the role of the disulfide is restricted to influencing SigK-RskA interactions (25).

The presence of the redox-responsive disulfide within SigK was the first example of a sigma factor having its own sensor, and prompted a search for sigma factor homologs that might also have this feature (25). The search revealed that 212 ECF sigma factors present in many different organisms, and that otherwise share little sequence similarity (>27% amino acid sequence identities), have similarly positioned cysteines that could form disulfide bonds. The majority of these sigma factors are present in Actinobacteria, but they are also present in
Proteobacteria, with relative frequencies of representation in the order $\alpha > \gamma > \beta$ (25). This finding suggests redox responsiveness may be a conserved feature amongst these sigma factors. Importantly, although not identified by Shukla et al., the two conserved cysteines are also present in RSPH17029_3536 (Figure 26). Note that, for SigK (25) and presumably other sigma factors having the di-cysteine motif, the signaling event to which the anti-sigma factor responds is independent of this internal redox mechanism, and so represents another layer of regulation in the system.

**Figure 26.** Amino acid sequence alignment of *M. tuberculosis* SigK and *R. sphaeroides* 2.4.9 ECF sigma factor RSPH17029_3536. The two redox-sensing cysteines, in SigK, C133 and C183 (25), and the corresponding cysteines of RSPH17029_3536, C151 and C201, are indicated (red arrows). Underneath the aligned amino acid sequences, an * (asterisk) indicates residue identity, a : (colon) indicates residues with highly similar properties, and a . (period) indicates residues with weakly similar properties. The alignment was generated using CLUSTALW (16).
Unlike as is the case for other ECF sigma factors that transcribe their own genes, it is not clear whether or not "self transcription" is true of rsph17029_3536. The upstream sequences of the gene do contain two of the AAC motifs of the -35 region of many ECF promoters, as described by Lane et al. (Figure 27; 15); however, there are no appropriately positioned similarities to the putative -10 elements of the hemT and rsph17029_0267 putative promoter sequences, both of which have been shown to be recognized by this sigma factor (Chapter III). It may be that another unknown sigma factor is responsible for transcribing rsph17029_3536, or that one or the other of the AAC motifs does correspond to the sigma factor promoter, despite the lack of similarity of the corresponding -10 region to that of hemT or rsph17029_0267.

![Figure 27. Upstream sequences of rsph17029_3536. Highlighted in green and yellow are two putative ECF-like promoter sequences with the spacing as indicated. In bold and underlined is the translation start site.](image)

**Bioinformatic analysis of ECF sigma factor RSPH17029_3603 and anti-sigma factor RSPH17029_3602.** BLAST searches of ECF sigma factor RSPH17029_3603 identified homologs among other Bacteria belonging mainly in the Proteobacteria group, but none are of known function. Its cognate anti-sigma factor protein RSPH17029_3602 is described as a member of the FecR super-family. According to Staron, *et al.*, the Fecl-like ECF sigma factors
can be divided into six distinct groups, which are restricted to the Proteobacteria group as far as is known at present (27). These different groups of FecI-like proteins have an extended genomic context conservation. Many members are involved in iron acquisition (27). In *E. coli*, which has the best-described FecR-FecI system, FecR interacts with, and negatively regulates the activity of FecI, which transcribes genes in response to ferric-citrate. Under iron-rich conditions, and conditions in which cells are starved for iron in the absence of ferric-citrate, FecR sequesters FecI at the cytoplasmic membrane (5). When ferric citrate is added to iron-starved cells, the periplasmically-located C-terminus of FecR interact with the outer membrane TonB-dependent ferric-citrate transporter FecA, leading to conformational changes in the FecR-FecI complex that results in release of FecI so that it may recruit RNAP to the ferric-citrate transport genes *fecABCDE* (5).

The *R. sphaeroides* 2.4.9 genome does not include genes corresponding to *fecA, B, C, D* or *E*. However, it does include genes that are currently annotated as *fecI* (*rsph17029_4119*) and *fecR* (*rsph17029_4120*), and whose products are 34% and 29% identical to the *E. coli* proteins with coverage of approximately 90%. But the *rsph17029_4119* and 4120 gene products do not meet criteria for these designations according to the ECFfinder algorithm developed by Staron *et al.* Nevertheless, genes positioned directly downstream of the so-called *fecI* and *R* genes (*rsph17029_4121-4124*) are purported to encode proteins involved in ferrichrome transport, suggesting the purported FecI sigma factor does play a role in iron acquisition.

An examination of genes in the vicinity of *rsph17029_3602* and 3603 reveals that a gene encoding for a putative TonB-dependent siderophore receptor (*rsph17029_3601*) is located directly downstream of *rsph17029_3602*, and this is followed by ABC-like metal transporter genes *rsph17029_3595-rsph17029_3592* (Figure 29, operon E). BLAST searches using the
product of the \textit{rsph17029\_3601} gene as the query indicate that it is most similar to a member of the FhuE superfamily, involved in the transport of ferric-coprogen and ferric-rhodotorulic acid. Both ferric-coprogen and ferric-rhodotorulic acid are hydroxamate-type siderophores which are produced by fungal organisms, such as \textit{Aspergillus} and \textit{Rhodotorula}, and are utilized as sources of iron by bacteria, such as \textit{E. coli}, which can transport the iron from those siderophores but does not itself make them (7, 11). BLAST analysis of the products of genes \textit{rsph17029\_3595-3593} suggest they may be involved in the transport of zinc, as the products of genes \textit{rsph17029\_3594} and \textit{rsph17029\_3593} bear similarities to the zinc transport proteins ZnuC and ZnuB of \textit{E. coli}, respectively. In \textit{E. coli}, ZnuABC is a high affinity zinc transport system (22). ZnuA is the periplasmic binding protein, which delivers zinc to the membrane permease ZnuB; ZnuB transports zinc across the cytoplasmic membrane with the aid of the ATPase activity of the ZnuC transporter protein. Comparison of RSPH17029\_3594 and RSPH17029\_3593 to ZnuABC reveals they are 38% identical (51% similarity), and 38% identical (56% similarity) to ZunC and ZnuB, respectively. While there are no significant matches of RSPH17029\_3595 with ZnuA, the zinc periplasmic binding protein, the \textit{rsph17029\_3595} product does bear similarity to SitA, a Mn$^{2+}$/Fe$^{2+}$ periplasmic solute binding protein in \textit{R. sphaeroides} strain 2.4.1. Although there exists an ECF-like promoter upstream of these genes (Figure 30, operon E), RNA-seq data does support the idea that these genes as being transcribed by this sigma factor.

Within the upstream sequences of \textit{rsph17029\_3603} are sequences that resemble those of the putative \textit{hemT} and \textit{rsph17029\_0267} promoters (Figure 28), suggesting that RSPH17029\_3603 may transcribe its own gene. Interestingly, there is a gap of 118 bp between the stop codon of \textit{rsph17029\_3603} and the start codon of its anti-sigma factor \textit{rsph17029\_3602}, suggesting they may not be co-transcribed, as is true of other ECF sigma factor-anti-sigma factor
gene pairs. Inspection of the upstream sequence of the anti-sigma factor reveals the presence of a similar promoter sequence, but the -35 and -10 putative promoter elements are not appropriately spaced relative to each other (Figure 28).

rsph17029_3603: -35 <-----N19------> -10
   AAAGCCGCTCCGCCCAAGCCGAG
   GGCACGATCGTCTCCCCGATGCCTCGCCAGTCCACGATCGGGGCTGATGGGGGGCCCGCCGATCCCG
   -35 <-----N14------> -10
   AGGCAGTCAMACTATGAGGGCCGGCACGCCACCACGCAGCTCGGTTACGCTACATCTG
   TCGCTGCCGATGGTGGAGGAGCGAAGGTCGCCCAATGAGCCCCAGAATCAAGCGACATTTC
   GGCACAGGCCAGCGTGCCGATGGCCGGCCGCCAGCGACACTGGTTGCTCCAGAGCCATCGCC
   MET
   ACTGTCGGCGAGGATGCGGTCGATG

rsph17029_3602: -35 <-----N13------> -10
   AAATGCTCGAACGAGGGGG
   CTCCGCCTCATTCCGTTGCGATATATCCCTGAGCGTGACCCCGATG
   MET
   ACTGTCGGCGAGGATGCGGTCGATG

Figure 28. Upstream sequences of rsph17029_3603 and rsph17029_3602. Highlighted in green and yellow are the putative ECF-like promoter sequences similar to those upstream of hemT and rsph17029_0267 (AAAT-N19-TTGT) with the spacing as indicated. The translation start sites of genes rsph17029_3603 and rsph17029_3602 are 118 bp apart. In bold and underlined is the translation start site.
Bioinformatic analysis of the extended genomic "neighborhood" of RSPH17029_3603 and 3536. There are three ribosomal RNA operons on CII of *R. sphaeroides* wild type strain 2.4.9, but only two on CII of wild type strain 2.4.1, and the approximately 200 kbp region between the two 2.4.9-CII rRNA operons is also absent from strain 2.4.1. This region, which includes the two 2.4.9-specific ECF sigma factors, was surveyed for genes coding for products that potentially modify the extracellular environment and/or are involved in coping with periplasmic stress. It should be noted that the *R. sphaeroides* 2.4.9 genome is considerably less well annotated than the 2.4.1 genome, and the absence of these sequences from strain 2.4.1 makes it impossible to use sequence comparisons between the two strains to assist in gene identification.

According to the current annotation of genes within the 2.4.9-specific region, there are eleven operons between the two ribosomal RNA operons whose genes encode products that are annotated as being involved in transport of extracellular solutes. Four of the operons are annotated as being involved in the import of C₄-dicarboxylic acids, such as succinate, malate, and fumarate, while the other two are annotated as being involved in the import of amino acids (Figure 29). Of the other five operons four are annotated as encoding products involved in transport of unknown compounds, while one encodes products related to the transport of chromate. An inspection of the upstream sequences of these eleven operons for the presence of ECF-like promoter elements (Figure 30; 10, 15) revealed that seven operons have -35 promoter element sequences similar to those upstream of the *hemT* gene (Chapter III), some having more than one ECF promoter like sequences (Figure 30, operon F).

Representation of operons that are possibly connected to the need to transport C₄-dicarboxylic acids and/or amino acids, suggests one or the other, or both of the ECF the sigma
factors might be involved in the transport of these substrates. A possible metabolic relationship between all of these genes is the fact that ALA, whose formation is catalyzed by the *hemT* gene product, is derived from a C₄-dicarboxylic acid (succinyl-CoA) and an amino acid (glycine).
Figure 29. Genomic context of ECF sigma factor RSPH17029_3536 and RSPH17029_3603. Shown is the genomic region surrounding the ECF sigma factor RSPH17029_3536 and RSPH17029_3603 (black). Operons labeled A (rsph17029_3545-3541), C (rsph17029_3571–3573), I (rsph17029_3645-3644) and J (rsph17029_3651-3653) code for proteins involved in the transport of C₄-dicarboxylic acids. Operons labelled B (rsph17029_3551-3555) and D (rsph17029_3587-3584) code for proteins involved in the
transport of amino acids. Operon E (rsph17029_3592-3595) code for proteins involved in the transport of metal, while operon F codes for proteins involved in the import of chromate. Finally, the operons labelled G (rsph17029_3611-3613), H (rsph17029_3636-3639) and K (rsph17029_3665-3667) code for ABC-transporter type proteins. The ribosomal RNA genes (black) define the endpoints of the 200 kbp chromosomal region not present in *R. sphaeroides* strain 2.4.1. The image was generated by the Chromosome Viewer tool available at the Joint Genome Institute (img.jgi.doe.gov).
Operon A

rsph17029_3545: AAAAGGAAACTTATCTTATCGGAGGAGGAGGAGGATCCGGAGGGGAATCGCC

Operon B

rsph17029_3551: AAAAGGATCGCTGCGAGGCGCTTAGGGCCGGCAGTCACGGACGGAGCACTGACCAAGGGAGGAGGACG

Operon C

rsph17029_3571: AAACATTTCGATCTTCAGGCGCTCGGGAGGAGGAGGACGCTAGTAGGGCTGCCGAACGGCAGCTCGCA

Operon D

rsph17029_3587: AAAAGATATGTTTACAAAGGCTCAAGGGCGTCAATCGATGCAACAGCAGCCCCGCAACAGAGAGGGAACGAG

Operon E

rsph17029_3595: CGAACAAGGGTTTTCTCGCCGATCGACCGAGGAGTGGCTCTCCGAGGCGGAGGAGGAGGACCTCTTCTCGCTGAGGCGGCGCTATCT

Operon F

rsph17029_3609: GGCTCTTTCTTGGATCCCAGAAGGCTGGGTTCTCGCCGATCGACCGAGGAGTGGCTCTCCGAGGCGGAGGAGGAGGACCTCTTCTCGCTGAGGCGGCGCTATCT
-35 <------N_{14}------> -10
-35 <------N_{18}------> -10

AAATTTAAGTGTTTTTCTCAGTCGACCCTTGTGGGGCCCGGCTTTTCCGCTGC
TTGATACTACCACCGGAGAAGAAAATCAGGCGGGCCCGAGACCGGGAGGCTCTCCGACTGCCCGCCGCCGCCG

CGGATCGGTCTGCGCCCTCAGGCGAGAGCAGGCTCTCGGGATGAACACCGTGCCAGGGCACACGC

**Operon G**

*rsph17029_3615*:  -35 <------N_{18}------> -10

GAACCTCATGAATATACTATCTTGAAATATGACATACAGTTTTAATACTATCT

**Operon H**

*rsph17029_3636*:  -35 <-----N_{11}-----> -10  -35 <-----N_{12}-----> -10

TTGCAATATTGCAAATTGGGAATGTGGAAACCGGCGATTCCATGTTGACGCTAGGAAGAAAAGACGATAAT

**Operon I**

*rsph17029_3645*:  -35 <-----N_{13}-----> -10

AAATTTAACCTCAAATACACAGGCTATATAATCCTAAAGGGCCCTGTTAAACCAAAATATTATTTTGCTTTTCTGCA

TTGATACTAATTTTATCTGCCATGTATTGGATAGTGACCAGCGCCGACGGAGGCGGCCTCGGCGAGCGACCACATCGC

**Operon ATG**

GCCCGGAGGACCACATACCCGGGGAGGATCCAC
Figure 30. Upstream sequences of operons A – H (Figure 29) within the 200 kbp region encompassing the ECF-sigma factors encoding products involved in the transport of C₄-dicarboxylic acids, amino acids, metals, chromate, or unknown substrates.

Highlighted in green, yellow, or cyan, are ECF sigma factor-like -35 and -10 promoter elements similar to that found upstream of *hemT* (AAAAT-N₁⁹-TTGT) with the spacing as indicated, with the exception of operons E and G, which have ECF-like promoters, however, they do not resemble those sequences identified upstream of *hemT*. In red, underlined and bold are -35 and -10 promoter elements which are most similar to the housekeeping $\sigma^{70}$ factor. Operons J and K are not shown due inability to identify any $\sigma^{70}$, or ECF sigma factor-like promoter sequences upstream. Operons are labeled as in Figure 29 and the upstream sequence of the first gene in each operon are shown. In bold and underlined are the translation initiation start sites.
Investigating the redox responsiveness of RSPH17029_3536. Based on the studies of SigK in *Mycobacteria* (25), the cysteine residues at positions 151 and 201 in *R. sphaeroides* RSPH17029_3536 protein correspond to cysteines that could form a disulfide under oxidizing conditions, and thereby mediate redox-responsiveness of the ECF sigma factor. To examine this, a mutant allele coding for an RSPH17029_3536 protein in which both cysteines are replaced by serines was inserted into plasmid vector pBBR-MCS (Table 4). It was hypothesized that a disulfide formed in the wild type protein in the presence of oxygen (oxidizing conditions) would be absent from the serine-substituted mutant protein, and so transcription that is dependent upon the reduced form of the sigma factor would take place even in an aerobic environment. Since *hemT* transcription depends upon this sigma factor (Chapter III), to test this, the mutant sigma factor protein was introduced in trans into wild type strain 2.4.9 bearing the 2.4.9-*hemT::lacZ* transcription reporter plasmid pNC1 (Table 4), and samples of cells cultured under highly aerobic conditions were assayed for β-galactosidase activities.
Figure 31. β-galactosidase activities in extracts of *R. sphaeroides* strains with the 2.4.9-
*hemT::lacZ* reporter plasmid pNC1. NC360 is 2.4.9(pNC1) with the empty vector pBBR-MCS2,
NC392 is 2.4.9(pNC1) pBBR::rsph17029_3536-C151S,C201S, and NC397 is strain 2.4.9
(pNC1) with pBBR::rsph17029_3536. For further details regarding strains and plasmids see
Table 4. The bacteria were cultured by sparging with 30% oxygen. Values are the average of at
least three replicates and ranges are as indicated.
The results (Figure 3) indicate that, unlike the wild type protein, the mutant sigma factor is able to transcribe hemT under non-reducing conditions. This is consistent with the idea that the two cysteines within the σ₄ domain of RSPH17029_3536 are involved in sensing changes in cellular redox. However, the identity of the inducing signal that is transmitted to this ECF via its anti-sigma protein cannot be determined by these experiments. Therefore, the full range of control of the activity of RSPH17029_3536 requires knowing the nature of the extracytoplasmic signal.

**Investigating the role of C₄-dicarboxylic acid with respect to the transcriptional activities of the ECF sigma factors.** There are two lines of evidence that suggest C₄-dicarboxylic acids play a role in controlling RSPH17029_3536 and 3602 activities. First, both of these ECF sigma factors transcribe the rsph17029_0267 gene (Chapter III), which is the first gene in an operon whose products encode a tripartite ATP-independent periplasmic (TRAP) transporter of C₄-dicarboxylic acids. Second, no fewer than four other operons (A, C, I and J) within the chromosomal region that is unique to strain 2.4.9 are also annotated as coding for proteins involved C₄-dicarboxylic acid transport; three of those have putative ECF-like -35 and/or -10 sequence elements (Figure 30). While analysis of RNA-seq data profiling the transcriptome belonging to these sigma factors is still ongoing, other than the potential role of the rsph17029-0267 product in C₄-dicarboxylic acid transport, no other guidance as to the nature of the extracytoplasmic signal to which anti-sigma factors RSPH17029_3537 or RSPH17029_3602 are responding has been forthcoming; nor has the bioinformatic analyses provided any additional clues.

To evaluate the possibility that C₄-dicarboxylic acid is a signaling molecule for one or the other, or both of the anti-sigma factors, β-galactosidase activity was assayed in extracts of wild
type 2.4.9, as well as the 2.4.9 sigma factor mutant strains JZ5321 and JZ5323, all having the
hemT::lacZ transcription reporter plasmid pNC1 (Table 4). The bacteria were cultured in
Sistrom's minimal medium with pyruvate, or with the C₄-dicarboxylic acid succinate (Figure 32).
Figure 32. β-galactosidase activities of *hemT* in *R. sphaeroides* wild type 2.4.9 and the 2.4.9 sigma factor mutant strains grown in pyruvate *versus* succinate media, under (A) aerobic and (B) anaerobic-dark conditions with DMSO as the alternate electron acceptor and supplemented with 0.1% peptone. In all cases, the reporter plasmid was 2.4.9-*hemT::lacZ* and the strains are as indicated. Cultures were grown with 34 mM pyruvate or 34 mM succinate to determine the effect C₄-dicarboxylic acids on *hemT* expression through the sigma factors of the individual
sigma factors RSPH17029_3536 and RSPH17029_3603. Values are the average of at least three replicates and ranges are as indicated.

The results (Figure 32) are complex. However, the following is apparent: (i) Regardless of the presence or absence of the C₄-dicarboxylic acid succinate, in wild type cells the most important factor with respect to hemT transcription is the growth condition; expression is higher when cells are grown anaerobically in the dark with DMSO than when they are grown aerobically. (ii) Transcription of hemT that is relying upon sigma factor RSPH17029_3603 alone (i.e. mutant strain JZ5323) is unresponsive to the presence or absence of succinate, but remains dependent upon growth conditions; as is true of hemT transcription in the wild type strain, expression is higher when cells are grown anaerobically in the dark with DMSO than when grown aerobically. (iii) Transcription of hemT that is relying upon sigma factor RSPH17029_3536 alone (i.e. mutant strain JZ5321) is responsive to the presence or absence of succinate, and this is independent of growth conditions.

The question raised by the β-galactosidase assay results for mutant strain JZ5321 with the reporter plasmid pNC1 is whether or not the presence of C₄-dicarboxylic acid is acting as a positive signal for the anti-sigma factor of RSPH17029_3536. The alternative possibility would be that the cytoplasm of cells grown in pyruvate is less reducing than the cytoplasm of cells grown in succinate, thereby effectuating changes in the status of the putative disulfide in the sigma factor itself. To distinguish between these two possibilities, hemT transcription was examined in anaerobically dark-grown wild type cells having the hemT::lacZ transcriptional
reporter pNC1, and also either the empty plasmid vector pBBR1-MCS2, plasmid pBBR::rsph17029_3536, or pBBR::rsph17029_3536-C151S,C201.

Figure 3. β-galactosidase activities of hemT in response to C₄-dicarboxylic acids with respect to the redox signaling component of sigma factor RSPH17029_3536. NC360 is 2.4.9(pNC1) with the empty vector pBBR-MCS2, NC392 is 2.4.9(pNC1) pBBR::rsph17029_3536-C151S,C201S, and NC397 is strain 2.4.9 (pNC1) with pBBR::rsph17029_3536. For further details regarding strains and plasmids see Table 4. Cultures were grown under anaerobic-dark conditions supplemented with 0.1% peptone and DMSO as the alternate electron acceptor. Values are the average of at least three replicates with ranges indicated.
The results of the β-galactosidase activity assays (Figure 3) show that the level of \textit{hemT} transcription was higher in the presence of succinate vs. pyruvate in all strains. This suggests that the C₄-dicarboxylic acid is working through the anti-sigma factor for RSPH17029\_3536. Had it been otherwise, there should have been no difference in activities for cells grown in either pyruvate or succinate since the sigma factor disulfide would be reduced regardless of carbon source under anaerobic-dark with DMSO conditions.

**Discussion**

The two sigma factors examined here are also present in other bacteria. This information did not add to an understanding of their role in \textit{R. sphaeroides} 2.4.9, as it is not known for any organism. Analyses of the genomic contexts of each sigma factor were also limited in their usefulness toward determining the role either RSPH17029\_3536 or 3603 may play in \textit{R. sphaeroides}. However, as this chromosomal region is present in 2.4.9 and absent in strain 2.4.1, it raises the possibility that strain 2.4.9 has unique metabolic features that are conferred by these additional sequences.

The results presented here show that RSPH17029\_3536 activity is responsive to changes in cellular redox due to the presence of cysteines that can form a disulfide bond within the σ₄-domain of the sigma factor. They also show that the presence of C₄-dicarboxylic acids such as succinate serves as an activating signal for the anti-sigma factor RSPH17029\_3537, which controls the activity of the ECF sigma factor RSPH17029\_3536 in \textit{R. sphaeroides} 2.4.9.

While the activating signal for the RSPH17029\_3603 anti-sigma factor is not yet known, it is clear that that the transcriptional activity of this sigma factor depends upon growth
conditions; anaerobic-dark with DMSO conditions are inducing. It is also clear that this sigma factor is unresponsive to the presence of succinate. Whether or not it is responsive to other C₄-dicarboxylic acids is not yet known. That a total of four operons annotated as encoding products involved in transport of such compounds are present in the 2.4.9-specific region of CII seems important, and it may be that these transporters are dedicated to transporting different C₄-dicarboxylic acids. That a C₄-dicarboxylic acid other than succinate might signal activation of RSPH17029_3603 remains a possibility.

The pattern of hemT transcription in wild type strain 2.4.9 by and large reflects the combined contributions of both sigma factors. However, the situation is more complex in that hemT transcription is relatively higher in the sigma factor mutant strain JZ5321 in which RSPH17029_3603 is absent, but this is only true under aerobic conditions. There are many possible explanations for this, several of which implicate an additional regulatory component. For example, a negative regulator of hemT that itself is transcribed by RSPH17029_3603 may be keeping hemT expression minimal until necessary; i.e. during the transition from aerobic to anaerobic-dark growth on DMSO. Presumably, further analysis of the transcriptomic data for *R. sphaeroides* 2.4.9 and the two sigma mutant strains JZ5321 and JZ5323 will be useful in identifying this missing component.
References


CHAPTER V: SUMMARY AND CONCLUSIONS

*R. sphaeroides* is a metabolically versatile microorganism able to grow under a multitude of conditions that involve different means of obtaining energy such as, but not limited to, aerobic respiration, anoxygenic phototrophy, and anaerobic respirations that involve appropriate alternate electron acceptors. To optimize growth, the cell must regulate the biosynthesis of components necessary to perform these various energy metabolisms. This regulation extends to the production of metabolites such as tetapyrroles, which support those energy gathering processes. The production of the three major tetapyrroles, vitamin B$_{12}$, heme and bacteriochlorophyll (bchl), are dependent upon various environmental stimuli, such as light intensity and oxygen availability, and are a part of a branched biosynthetic pathway which begins with the common precursor 5-aminolevulinic acid (ALA). Towards understanding how *R. sphaeroides* achieves the orderly production of the kinds and amounts of tetapyrroles to meet its metabolic needs, this study focused on the regulation of ALA production, which is the precursor for all tetapyrroles.

*R. sphaeroides* encodes two ALA synthase genes, *hemA* and *hemT*, and the inability to determine the significance of having two ALA synthases have been hampered in past years, as in the most intensely studied strain, *R. sphaeroides* 2.4.1, which has both *hemA* and *hemT*, the expression of *hemT* has never been detected in wild type cells (2, 3). Recently, a study reported that *hemT* transcription is responsive to nitrosative stress in that strain (2). However, this could not be confirmed by using transcriptional reporter plasmids involving *hemT* upstream sequences in two strains of *R. sphaeroides*, 2.4.1 and 2.4.9. It has since been found that, in the absence of any stressor and unlike in strain 2.4.1, *hemT* is expressed in strain 2.4.9 and its expression is
induced upon reducing oxygen tensions as cell transition from aerobic to anaerobic-dark growth with DMSO.

Strains of *R. sphaeroides* differ in their ability to transition from aerobic growth to either anaerobic-dark growth with DMSO, or anoxygenic photosynthesis. Wild type strain 2.4.1, in which photosynthetic complexes are completely absent from aerobically grown cells, are able to transition to anaerobic-dark growth with DMSO with virtually no lag, while a lag of 11-12 hours is observed as they transition to phototrophic growth; the former has not been recorded previously, but the latter has already been well-documented, and it has been accepted in the field as being necessary for the bacteria to undergo morphogenesis of the membrane to ICM with accompanying incorporation of photosynthesis pigment-protein complexes comprising the photosynthetic apparatus. For strains such as 2.4.9, in which the presence of oxygen does not curtail the production of photo-complexes to the same extent as in wild type strain 2.4.1, rapid transitioning to phototrophic growth may be due to the fact that they can immediately begin to use photonic energy for growth, despite the limited transcriptional responsiveness of *hemA* to the absence of oxygen. On the other hand, the presence of photosynthetic membranes in aerobic 2.4.9 cells in no way facilitates the transition from aerobic to anaerobic-dark growth; evidently, the presence of *hemT* is critical for that process in the absence of strong upregulation of *hemA* that could overcome heme-mediated feedback inhibition of the HemA enzyme.

The prolonged lag in transitioning to phototrophic growth for strain 2.4.3 compared to strain 2.4.1 may be due to the weak transcriptional response of the *hemA* gene to the absence of oxygen. But there are other possibilities, such as differences in photosynthesis gene expression that rely upon the PrrBAC system, as is suggested by amino acid sequence comparisons to these same proteins in strains 2.4.1 and 2.4.9.
While further investigations are necessary to confirm which among these possibilities is correct, they do provide evidence that the strains examined here exemplify broad adaptations to different environmental niches. At least some among these adaptations appear to be reflected in differences in representation and expression of the ALA synthase genes.

Through efforts to determine the factors governing hemT expression in strain 2.4.9, two unique extracytoplasmic function (ECF) sigma factors were discovered, which are absent from strain 2.4.1, and are necessary for hemT expression. Subsequent to their discovery it was determined that both ECF sigma factors RSPH17029_3536 and RSPH17029_3603, directly bind to and transcribe hemT. Apparently, this involves transcription from overlapping promoter sequences.

The ECF sigma factors identified in strain 2.4.9 could not be assigned to any of the previously characterized ECF sub-groups described by Staron et al. (5). However, sequence comparisons suggested RSPH17029_3536 belongs to a set of ECF sigma factors that are redox responsive by virtue of having two cysteines that form a disulfide under oxidizing conditions (6). This was confirmed by comparing the transcription activity of the wild type protein to a mutant in which the cysteines have been replaced by serines.

Other investigations set out to glean information about the role of these sigma factors from RNA-seq data that compared the levels of transcripts present in the wild type strain to those present in the sigma factor mutant strains. The data identified a second gene, rsph17029_0267, whose transcripts are statistically significantly decreased in the absence of either sigma factor. The product of rsph17029_0267 is a periplasmic binding protein that delivers C4-dicarboxylic acids to its cognate TRAP transporter proteins. EMSAs involving the upstream sequences of rsph17029_0267 revealed this gene is also directly transcribed by each sigma factor indicating
the role of either, or both, sigma factors may be to the presence of C₄-dicarboxylic acids such as succinate in the environment.

The responsiveness of either RSPH17029_3536 or RSPH17029_3603 to the presence of the C₄-dicarboxylic acid succinate was examined, and found to be limited to RSPH17029_3536. However, this does not exclude the possibility that RSPH17029_3603 may be responsive to C₄-dicarboxylic acids other than succinate.

That factors in addition to the ECF sigma factors are involved in hemT transcription seems possible, and might well explain certain features of hemT expression that are not yet understood. Thus, it is not known why aerobic transcription of hemT is elevated in the absence of RSPH17029_3603, but it may involve a negative-acting regulatory protein that is transcribed by this sigma factor. Wild type strain KD131, which only encodes RSPH17029_3603 and not RSPH17029_3536, may be useful in identifying this additional component. This and other investigations might further benefit from knowledge as to representation of these sigma factors among other strains of R. sphaeroides that have already been studied with respect to their ALA synthase gene distribution (4). Toward that end, a preliminary survey was conducted by determining whether amplicons using primers corresponding to rsph17029_3536 and rsph17029_3603 sequences were generated by PCR of genomic DNA isolated from several of those unsequenced wild type strains (Table 6). As might be predicted, representation was variable, and it does not consistently correlate with the presence or absence of a hemT gene - as is already shown here for strain 2.4.1, the presence or absence of hemT does not provide information as to whether it is transcribed or not. Knowing the status of hemA and hemT expression, in combination with growth profiles of these other strains as they transition from one "lifestyle" to another is predicted to provide further insights into the significance of having more
than one ALA synthase gene in these, and perhaps other purple non-sulfur bacteria, while also expanding knowledge as to the role of the ECF sigma factors.

Table 6. Summary of hemA, hemT, and ECF sigma factor gene representation among wild type strains of R. sphaeroides

<table>
<thead>
<tr>
<th>Wild type strain</th>
<th>Presence (+) or absence (-) of gene or amplicon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hemA(^1)</td>
</tr>
<tr>
<td>2.4.9(^3)</td>
<td>+</td>
</tr>
<tr>
<td>SH5</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 21455</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 35053</td>
<td>+</td>
</tr>
<tr>
<td>KD131(^3)</td>
<td>+</td>
</tr>
<tr>
<td>NCIB 8253</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 35054</td>
<td>+</td>
</tr>
<tr>
<td>2.4.1(^3)</td>
<td>+</td>
</tr>
<tr>
<td>WS8N(^3)</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 21286</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 17027</td>
<td>+</td>
</tr>
<tr>
<td>2.4.3(^3)</td>
<td>+</td>
</tr>
<tr>
<td>ATCC 17024</td>
<td>+</td>
</tr>
</tbody>
</table>

\(^1\)Identified by Southern hybridization (4).

\(^2\)Identified by amplification using genomic DNA isolated with the Gene Elute Kit (Sigma Aldrich) and with primers designed according to rsph17029_3536 and rsph17029_3603.
sequences as follows: For rsph17029_3603, primers Sigma3603-UP and Sigma3603-DOWN were used; for rsph17029_3536 primers Sigma3536-UP and Sigma3536-DOWN2 (Table 5) were used. (For primer sequences, see Table 3, with the exception of primer Sigma3536-DOWN2, which are provided in Table 5). The PCR program used to amplify both rsph17029_3603 and rsph17029_3536 was as follows: 34 cycles of 95°C, 1 min.; 50°C 15 sec.; 72°C 45 sec.

3Sequenced genomes.
References


