STUDIES ON THE TRANSCRIPTION OF PHOTOSYNTHESIS GENES OF THE PHOTOSYNTHETIC BACTERIUM RHODOBACTER CAPSULATUS

by

Mary Elspet Forrest

B.Sc., University of Calgary, 1982

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES (DEPARTMENT OF MICROBIOLOGY)

We accept this thesis as conforming to the required standards

THE UNIVERSITY OF BRITISH COLUMBIA
AUGUST 1988
© Mary Elspet Forrest, 1988
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Microbiology

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date Oct 12/88
Rhodobacter capsulatus is a Gram negative bacterium that exhibits a variety of growth modes, including chemoheterotrophic growth and photoheterotrophic growth. Upon a shift of cultures from high to low oxygen concentrations the photosynthetic apparatus is synthesized and incorporated into the inner membrane. The puf operon contains genes that encode structural proteins found in the light-harvesting and reaction center complexes. In a preliminary attempt to pinpoint the location of the puf promoter R. capsulatus RNA polymerase was purified by standard techniques and used in in vitro run-off transcription assays. It was found that the polymerase was capable of specific transcription with linearized pUC13 DNA but no specific transcription could be obtained with R. capsulatus DNA. It was concluded that some factor or condition necessary for specific transcription with R. capsulatus DNA was absent from these assays. The location of the puf promoter was subsequently found through a series of deletions and oligonucleotide-directed mutations in the 5' region of the puf operon. Fragments that contained these mutations were placed translationally in-frame with the lacZ gene of Escherichia coli in plasmids that could be conjugated into R. capsulatus. Assays of beta-galactosidase activities under low and high oxygen conditions resulted in localization of the promoter
to a position approximately 540 basepairs upstream of what was previously believed to be the first gene of the operon, the $\text{pufB}$ gene. RNA 5' end-mapping experiments showed that the quantity of RNA transcripts obtained were comparable to the $\text{lacZ}$ activities. The existence of multiple low abundance RNA 5' ends prompted the theory that the primary transcript has a short half-life, and is rapidly processed to yield a more stable transcript with a 5' end that maps just upstream of the $\text{pufB}$ gene. It was found that only the 5' end nearest to the promoter could be capped by guanylyl transferase, and this could only be detected when the putative processing sites were deleted. The DNA sequence between the promoter and the $\text{pufB}$ gene contains a new gene of the $\text{puf}$ operon, the $\text{pufQ}$ gene. Deletion of this gene showed that it plays an essential role in the formation of mature light-harvesting and reaction center complexes.
TABLE OF CONTENTS

ABSTRACT........................................................................................................... ii
LIST OF TABLES.................................................................................................... vii
LIST OF FIGURES.................................................................................................. viii
ABBREVIATIONS AND SYMBOLS....................................................................... x
ACKNOWLEDGEMENTS....................................................................................... xii

INTRODUCTION................................................................................................. 1
MATERIALS AND METHODS.............................................................................. 12

1. Bacterial strains, growth conditions, and beta-galactosidase assays........... 12

    1.1. Bacterial strains............................................................................................. 12

    1.2. Growth of bacterial cultures for RNA polymerase purification.............. 12

    1.3. Growth of bacterial cultures for beta-galactosidase assays....................... 13

2. Purification of R. capsulatus RNA polymerase............................................. 13

    2.1. Purification of RNA polymerase for run-off transcription assays............. 13

    2.2. Further purification of RNA polymerase.................................................... 14

    2.3. RNA polymerase assays............................................................................. 14

3. Run-off transcription assays......................................................................... 15
4. Polyacrylamide gel electrophoresis ................................................................. 16
  4.1. Slab gel electrophoresis of RNA polymerase ............................................... 16
  4.2. Polyacrylamide gel electrophoresis of RNA transcripts ................................ 16

5. Preparation of DNA and cellular RNA ............................................................ 17
  5.1. Preparation of DNA .................................................................................. 17
  5.2. Purification of cellular RNA ..................................................................... 17

6. Plasmid constructions ...................................................................................... 17
  6.1. Construction of the promoter-identification vector pXCA601 ................... 17
  6.2. Construction of the lacZ fusion plasmids .................................................. 19
  6.3. Site-directed mutagenesis ......................................................................... 23
  6.4. Construction of the pufQ gene deletion ................................................... 26

7. S1 nuclease mapping of RNA .......................................................................... 26
  7.1. S1 nuclease mapping of puf operon RNA ends .......................................... 26
  7.2. S1 nuclease protection experiments for detection of RNA from the puc operon .................................................. 27

8. Primer extension experiments ........................................................................... 28

9. Capping experiments ....................................................................................... 28

10. Absorption spectroscopy of various strains of R. capsulatus ....................... 29

RESULTS ................................................................................................................ 30

1. Purification and characterization of RNA polymerase .................................... 30
  1.1. Purification of RNA polymerase ............................................................... 30
1.2. Optimization of the assay conditions..........................................................33
1.3. Subunit composition of RNA polymerase......................................................35
1.4. Gel electrophoresis of run-off transcription products.....................................39

2. Localization of the *puf* promoter..................................................................47
   2.1. Localization of the *puf* promoter through gene fusion experiments................47
   2.2. Absorption spectrophotometric analysis of *puf* promoter mutants..................52
   2.3. Mapping of RNA transcripts from the *puf* operon....................................57
   2.4. Mapping of a capped RNA 5' end.............................................................69

3. Assessment of the role of the *pufQ* gene product.............................................74
   3.1. Absorption spectroscopy of cells containing wild-type or mutant *pufQ* genes..74
   3.2. The effects of *pufQ* on transcription and translation of *puf* or *puc* mRNA..78

DISCUSSION...........................................................................................................81
REFERENCES..........................................................................................................91
APPENDIX...............................................................................................................98
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table I.</th>
<th>Summary of RNA polymerase purification</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table II.</td>
<td>Utilization of different templates by <em>R. capsulatus</em> RNA polymerase</td>
<td>34</td>
</tr>
<tr>
<td>Table III.</td>
<td>Utilization of templates used for run-off transcriptions</td>
<td>43</td>
</tr>
<tr>
<td>Table IV.</td>
<td>Assays of beta-galactosidase activities of cells containing puf 5' regions fused to the <em>lacZ</em> gene of pXCA601</td>
<td>49</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. A diagramatic representation of the photosynthetic apparatus of
Rhodobacter sphaeroides......................................................3

Figure 2. A representation of the puf operon and mapped RNA
transcripts.............................................................................7

Figure 3. Nucleotide sequence of part of the puf promoter region............9

Figure 4. A representation of the promoter identification vector pXCA601...18

Figure 5. Construction of the puf promoter deletion mutants in pXCA601....20

Figure 6. Construction of oligonucleotide-directed mutations in pXCA601...24

Figure 7. Chromatography of RNA polymerase from chemoheterotrophically
grown cells.............................................................................31

Figure 8. SDS polyacrylamide gel electrophoresis of purified RNA
polymerase from chemoheterotrophically and
photoheterotrophically grown R. capsulatus..................................36

Figure 9. SDS polyacrylamide gel electrophoresis of RNA polymerase after
DEAE-Sephadex chromatography.............................................38

Figure 10. SDS polyacrylamide gel electrophoresis of RNA polymerase after
glycerol gradient centrifugation...............................................40
Figure 11. A representation of the construct pJAJ21.................................42
Figure 12. Polyacrylamide gel electrophoresis of run-off transcripts obtained
with photoheterotrophic RNA polymerase............................................44
Figure 13. Polyacrylamide gel electrophoresis of run-off transcripts obtained
with photoheterotrophic RNA polymerase and the purified R. capsulatus promoter fragment as template..............................46
Figure 14. A representation of the vector pTB999......................................54
Figure 15. Absorption spectra of various strains of R. capsulatus..............55
Figure 16. Construction of DNA probes used in RNA end-mapping
experiments.........................................................................................59
Figure 17. 5' end-mapping of B10 RNA with S1 nuclease.........................61
Figure 18. Primer extension titration experiment.....................................62
Figure 19. 3' end-mapping of B10 RNA with S1 nuclease..........................65
Figure 20. 5' end-mapping of RNA from R. capsulatus ΔRC6 containing
puf/lac fusion plasmids by S1 nuclease................................................67
Figure 21. 5' end-mapping of capped RNA with nucleases.....................70
Figure 22. Absorption spectra of R. capsulatus strains B10, ΔRC6, and U43....75
Figure 23. Absorption spectra of various strains of R. capsulatus............77
Figure 24. S1 nuclease protection measurement of puc mRNA from B10 and
ΔRC6.................................................................................................80
Figure 25. Processing of puf mRNA..........................................................89
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ap</td>
<td>ampicillin</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bchl</td>
<td>bacteriochlorophyll a</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>ca</td>
<td>approximately</td>
</tr>
<tr>
<td>Ci</td>
<td>Curie</td>
</tr>
<tr>
<td>CTP</td>
<td>cytidine triphosphate</td>
</tr>
<tr>
<td>cyt c2</td>
<td>cytochrome c2</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dpm</td>
<td>disintegrations per minute</td>
</tr>
<tr>
<td>dut</td>
<td>dUTPase gene of <em>E. coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>GC</td>
<td>guanine/cytosine</td>
</tr>
<tr>
<td>GTA</td>
<td>gene transfer agent</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>lacY</td>
<td><em>E. coli</em> lactose permease gene</td>
</tr>
<tr>
<td>lacZ</td>
<td><em>E. coli</em> beta-galactosidase gene</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>B870</td>
<td>light-harvesting I (LH I) complex</td>
</tr>
<tr>
<td>B800-850</td>
<td>light-harvesting II (LH II) complex</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>ntr</td>
<td><em>E. coli</em> nitrogen regulation genes</td>
</tr>
<tr>
<td>ONPG</td>
<td>ortho-nitrophenyl-beta-D-galactoside</td>
</tr>
<tr>
<td>puc</td>
<td><em>R. capsulatus</em> puc operon, encodes B800-850 structural proteins</td>
</tr>
<tr>
<td>puf</td>
<td><em>R. capsulatus</em> puf operon, encodes B870 (B and A) and reaction center (L and M) structural proteins, plus two open reading frames (Q and X)</td>
</tr>
<tr>
<td>puhA</td>
<td><em>R. capsulatus</em> gene for the H protein of the reaction center</td>
</tr>
<tr>
<td>RF</td>
<td>replicative form</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>SDS-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>ss</td>
<td>single stranded</td>
</tr>
<tr>
<td>Tc</td>
<td>tetracycline</td>
</tr>
<tr>
<td>TCA</td>
<td>trichloroacetic acid</td>
</tr>
<tr>
<td>U</td>
<td>uracil</td>
</tr>
<tr>
<td>ung</td>
<td><em>E. coli</em> uracil glycosylase gene</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine triphosphate</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to thank my supervisor Tom Beatty for his constant support and for teaching me how to become a researcher. I would also like to thank my fellow lab members, especially Cheryl Keen, Tim Lilburn, and Anthony Zucconi, for their friendship and advice. I am very grateful for the support from a University of British Columbia Fellowship, a Natural Sciences and Engineering Research Council Postgraduate Fellowship, and an Izaak Killam Postgraduate Fellowship. This work was supported by a Natural Sciences and Engineering Research Council Grant, A-2796. I would like to acknowledge the members of my committee and all the people in the Department of Microbiology at the University of British Columbia for their warmth and helpfulness. I dedicate this thesis to Gord and to my family for their love and for always believing in me.
INTRODUCTION

My thesis has been directed toward the study of gene expression in the Gram negative photosynthetic bacterium *Rhodobacter capsulatus* (formerly known as *Rhodopseudomonas capsulata* [25]). *R. capsulatus* shares many traits with other photosynthetic bacteria, especially *R. sphaeroides*. Although most of the information presented in this section was obtained from studies of *R. capsulatus*, occasionally some general statements derive from studies of *R. sphaeroides* [27].

*R. capsulatus* regulates the composition and amount of its cell membrane in response to oxygen concentration and light intensity. Under high oxygen tension (high O$_2$) cells grow chemoheterotrophically by respiration and the cell membrane composition is similar to that of non-photosynthetic bacteria [19]. When the oxygen tension is lowered, the photosynthetic apparatus is synthesized and the cytoplasmic membrane becomes an invaginated intracytoplasmic membrane into which the photosynthetic complexes are inserted [19].

There are three distinct pigment-protein complexes contained in the photosynthetic apparatus (seen in Fig. 1). There are two different light-harvesting
complexes that gather the light energy and funnel it toward the third complex, the reaction center, where the light energy is turned into electron flow [20]. Cyclic electron flow results in protons being pumped across the membrane to create a proton gradient that drives the synthesis of ATP [20]. The key pigment involved is bacteriochlorophyll a (bchl) which either traps and transfers light energy in the light-harvesting complexes, or is involved in the excitation of electrons to initiate cyclic electron transport in the reaction centers [19]. Carotenoids are present to shield the cell from harmful byproducts due to photo-oxidation, and also may harvest some light energy at wavelengths outside of the bchl/protein complex absorption range [47].

The B800-850 light-harvesting complex is so named because of the maximal wavelengths at which it absorbs light [56]. Each complex contains two alpha and two beta peptides, which bind six molecules of bchl and three carotenoids. A third peptide, of 14 kiloDaltons, co-purifies with the alpha and beta peptides but it does not bind pigments and its function is unknown [56]. This complex is present in the cell in varying amounts relative to the reaction center, depending on the light intensity [41]. As the light intensity decreases the number of B800-850 complexes increases in order to maximize the efficiency of photon capture. The second light-harvesting complex, the B870 complex, absorbs light maximally at 870 nm and also contains two peptides (B870 alpha and beta) that bind two molecules each of bchl and carotenoids [20]. These complexes are
Fig. 1. A diagramatic representation of the photosynthetic apparatus of *R. sphaeroides*, a facultatively photosynthetic bacterium that is functionally similar to *R. capsulatus*. Shown are the B800-850 and B875 (comparable to the B870 of *R. capsulatus*) complexes as well as the reaction center complex (designated RC). Also shown is the cyt c2/bc1 complex with the mobile cyt c2 protein, and the ATP phosphohydrolase complex. The succinate DH and NADH DH complexes are involved in maintainence of the redox state of the quinone pool. The symbols are as follows: Q and QH2, the quinone pool; ADP, adenosine 5'-diphosphate; ATP, adenosine 5'-triphosphate; DH, dehydrogenase; NADH, reduced nicotinamide adenine dinucleotide. Reproduced from Kiley and Kaplan [27] with the permission of the authors and the American Society for Microbiology.
believed to be in a relatively constant ratio with the reaction centers
(approximately 10-20 B870 complexes per reaction center). It is generally believed
that the flow of light energy is from B800-850 to B870 to reaction center [21]. The
reaction center contains three peptides, designated L, M, and H, which orient
four bchl molecules and other cofactors (bacteriopheophytin, iron, and quinones)
in the membrane [54]. The cycle of electron flow that initiates at the reaction
center is completed by the diffusion of a reduced quinone to the cytochrome
c2:oxidoreductase (cyt c2 / bc1) complex, and the transfer of an electron from this
complex back to the reaction center by a mobile, periplasmic cytochrome c2 (cyt
c2) carrier [14]. However, R. capsulatus (in contrast to R. sphaeroides) is also
capable of transferring electrons directly from the bc1 complex to the reaction
center [15].

Genetic analysis of the R. capsulatus genome has been accomplished by
use of promiscuous R plasmids that act to catalyze chromosome transfer [34,51],
and fine-structure mapping has been done by use of gene transfer agent (GTA).
GTA acts like a generalized transducing phage and packages random fragments
of chromosomal DNA, but only host DNA has been found in the GTA particles
[33]. GTA has been used as a vehicle to transfer genes mutated in vitro into the
chromosome by homologous recombination [51]. The analyses of these mutants
have greatly advanced the study of genes needed for photosynthesis.

Almost all of the genes known to be necessary for photosynthesis have
been found in a cluster of about 50 kilobases (kb) on the chromosome [34]. Genetic and physical maps have been aligned and specific genes have been isolated. The genes encoding the two B870 peptides have been designated the B and A genes and are part of an operon that also encodes two of the reaction center peptides (L and M). This was previously designated the rxa operon [54], but is now known as the puf operon. The gene for the third reaction center peptide is called puhA and is separate from the L and M genes [54]. The H peptide has been shown to be present in R. sphaeroides cells grown aerobically and may be involved in assembly of the reaction center upon a shift to low O2 [12]. Other genes found in the cluster are genes involved in the biosynthesis of the bchl and carotenoid pigments.

The genes for the B800-850 structural proteins map just outside of the 50 kb cluster in close proximity to one another, and have been shown to be in one operon, the puc operon [58].

At the time I began my thesis work it was known that after a shift from chemoheterotrophic growth to photoheterotrophic growth there was an 80 fold increase in bchl, and that the number of photosynthetic complexes increased 10-20 fold [40]. Corresponding with these increases was an increase in levels of mRNA encoding bchl biosynthetic enzymes and photosynthetic complex peptides [13]. It had also been found that cells that contained plasmids with photosynthesis peptide gene fusions to the lacZ gene of E. coli had more beta-
galactosidase activity when grown in low O\textsubscript{2} conditions [8]. Since the levels of mRNA and \textit{lacZ} expression roughly corresponded to the levels of photosynthetic complexes found in cells grown under high O\textsubscript{2} and low O\textsubscript{2} conditions, it appeared that expression of these genes was regulated by oxygen and that the regulation was at the level of transcription. The mechanism of regulation by O\textsubscript{2} was not known but had been postulated by various people to involve a redox carrier, a repressor or activator protein, or the intracellular levels of some small effector molecule [27].

I have focussed my work on the localization of the O\textsubscript{2}-regulated promoter for the \textit{puf} operon. Figure 2a shows a representation of the \textit{puf} operon with the four known genes: B and A (genes for the B870 peptides) and L and M (genes for two of the three reaction center peptides), and the two open reading frames Q (see below) and X [54]. The use of S1 nuclease for RNA end-mapping studies of this operon had located a 5' RNA end (end 4 in Fig. 2a) about 100 base pairs upstream from the \textit{pufB} gene [8]. This RNA transcript was found to have three 3' ends that mapped 0.49, 0.5, and 2.7 kb downstream from the 5' end [8]. The molar ratio of the longer:shorter transcripts is about 1:9, and the longer transcript in Figure 2a has been shown to be a precursor transcript which is then processed to yield the shorter transcripts [8]. The \textit{pufX} open reading frame is co-transcribed with the 2.7 kb \textit{puf} transcript but no gene product has yet been found.

A segment of the 5' region of the \textit{puf} operon, from the XhoI to the
Fig. 2. A representation of the *puf* operon and mapped RNA transcripts. 2a represents the *puf* operon with the 6 known genes (Q, B, A, L, M, and X; L and M actually overlap [54]); 2b represents the *puf* promoter region expanded from 2a (1.1 kb from XhoII to SphI). The thin lines represent non-coding regions of DNA and the thick lines represent *puf* genes. The arrows represent RNA transcripts with thinner arrows designating less abundant transcripts. The arrows above the *puf* operon in 2a represent transcripts previously studied [8], and the 5' end of these transcripts has been designated end 4. The thinner arrows below the promoter fragment in 2b represent recently mapped transcripts, with 5' ends 1, 2, and 3, while the thick arrow represents the 5' end of the transcripts seen in 2a. The letters shown below the DNA designate restriction endonuclease recognition sites as follows: X, XhoII; Sp, SphI; A, AccI; S, SalI; E, EcoRI; M, MnII.
SphI restriction sites shown in Figure 2a, is expanded in Figure 2b and can be seen to contain the *pufQ* gene and the 5' end of the *pufB* gene, as well as the site of the 5' end of the RNA transcripts described above. This thesis presents the results of RNA end-mapping studies that have located the 5' ends of RNA molecules that map even further upstream than the previously discovered 5' end, and that are less abundant in the cell. Three of these less abundant transcripts have 5' ends that map upstream of the *pufQ* gene and are shown in Figure 2b, designated ends 1, 2, and 3.

A large portion of the DNA fragment shown in Figure 2b has been sequenced [1,5,54] and part of this sequence is shown in Figure 3. An important feature of this sequence is the presence of an open reading frame that is capable of encoding a 74 amino acid long protein [1,5]. This open reading frame is preceded by a possible Shine-Dalgarno sequence that is similar to other known *R. capsulatus* Shine-Dalgarno sequences [54]. The open reading frame, designated *pufQ*, has been shown to contain a hydrophobic stretch of amino acids [1]. Other *puf* operon gene products have also been shown to contain hydrophobic stretches [54] and a comparison of these showed sequence similarity between *pufQ* and the *pufL* and *pufM* gene products [1]. The segments of L and M that align with the *pufQ* gene product are believed to be transmembrane stretches that bind bacteriochlorophyll pigments [2,54]. The possible role of *pufQ* has been studied and will be discussed later in this thesis. Also shown in Figure 3 are
Fig. 3. Nucleotide sequence of part of the puf promoter region. The sequence begins approximately 115 nucleotides downstream of the XhoII site, shown in Fig. 2, to within the pufB gene. The amino acid sequences of the pufQ and part of the pufB genes are shown above the DNA sequence. The Shine-Dalgarno sequence (SD) of the pufB gene is indicated above the DNA sequence as is a possible Shine-Dalgarno sequence for pufQ. The end-points of the Δ4, Δ14, and Δ24 deletions are also shown above the DNA sequence. The palindromes removed to create the Δ41 and Δ42 deletions are indicated by solid lines beneath the DNA sequence, as are the locations of key restriction enzyme recognition sites. Two • symbols adjacent to the AccI site indicate the possible position of the 5' end of RNA transcript 1 discussed in the text.
sequences that were mutagenized to aid in localization of the puf promoter.

The photosynthesis genes have not yet been expressed in heterologous systems such as E. coli and it is believed that part of this block is at the level of transcription [26]. For this reason, studies of R. capsulatus gene expression have been performed in mutant cells of R. capsulatus through what is sometimes slow and tedious work. Therefore I initially attempted to develop an in vitro transcription assay system using R. capsulatus RNA polymerase in order to create a simple and fast method of assaying for promoter activity. The results of this work are described in the first part of my thesis.

Another approach taken to locate the puf promoter was to compare the in vivo effects of wild-type and mutant promoter fragments on genes carried by conjugative plasmids. This was accomplished with the use of vectors that enabled promoter activity to be monitored by following expression of the E. coli lacZ gene. Since R. capsulatus does not ordinarily use lactose as a carbon source, expression of the lacZ gene in this bacterium proved to be a useful tool for study of gene expression and promoter mutations. The location of the puf promoter was found through the use of deletions and oligonucleotide-directed mutations, as were sequences that may be involved in O2 regulation of the promoter.

The results of the gene fusion work were correlated with measurement of the levels of mRNA found in the cells, and RNA 5' ends were mapped by a number of techniques. My data show that transcription initiates at RNA end 1
(see Fig. 2) and that processing is the probable cause of the presence of the other RNA transcripts, including the most stable transcripts with end 4. This processing of the primary transcript may be another method of control of synthesis of the photosynthetic complexes.
MATERIALS AND METHODS

1. Bacterial strains, growth conditions, and beta-galactosidase assays

1.1 Bacterial strains. The *R. capsulatus* wild-type strain B10 has been described [33], as have the deletion mutant strains ΔRC6 [11], and U43 [55].

1.2. Growth of bacterial cultures for RNA polymerase purification. Cultures of *R. capsulatus* B10 were grown at 34°C in 20 liter batches of RCV medium [49] supplemented with 10 mM potassium phosphate buffer (pH 6.8) and 0.1% yeast extract, in a glass fermentation vessel. Chemoheterotrophic (aerobic) cultures were vigorously aerated with 12 liters per minute of sterile air, with 250 rpm agitation. These cultures were grown only to a density of approximately 2.4 x 10^8 cells/ml to minimize the effects of oxygen limitation. Measurement of bacteriochlorophyll in such cultures confirmed that levels of this pigment were 300-fold lower than in cells grown photosynthetically. Medium for photoheterotrophic (anaerobic) growth of cells was purged of oxygen by vigorous bubbling with 5% CO₂ in N₂ for 0.5 hr prior to inoculation. After inoculation the medium was sparged for 15 min before sealing the fermentor.
Illumination was provided by a bank of nine 60 Watt Lumiline tungsten lamps encircling the vessel. The cells were grown to a density of about $4 \times 10^8$ cells/ml.

The cells were collected by centrifugation through a Sharples continuous flow centrifuge. The cell paste was stored at -80°C immediately after centrifugation.

1.3. Growth of bacterial cultures for beta-galactosidase assays. *R. capsulatus* was grown under low and high $O_2$ conditions as described [8]. The cultures were inoculated at a cell density of $0.8 \times 10^8$ cells/ml and grown to a cell density of $3.2 \times 10^8$ cells/ml. Twenty-five ml of cells were then harvested for assays as described [8].

2. Purification of *R. capsulatus* RNA polymerase

2.1. Purification of RNA polymerase for run-off transcription assays. RNA polymerase was purified using a modified version of a published procedure [45]. The procedure was performed at 0-4°C and was changed as described below. The fractions from the DNA cellulose column containing activity were combined and concentrated as before [45], diluted with buffer B (buffer A + 10% glycerol) to a conductivity of 5 milliohms (equivalent to buffer B containing 0.1 M NaCl), and applied to a 1 x 3 cm heparin-Sepharose column.
The column was washed successively with 3 column volumes each of solutions of buffer B containing 0.1 M, 0.25 M and 0.35 M NaCl, and then the RNA polymerase activity was eluted with buffer B containing 0.6 M NaCl.

2.2. Further purification of RNA polymerase. One milliliter (150 µg) of RNA polymerase from the heparin Sepharose eluate was diluted to a conductivity equivalent to 0.1 M NaCl in buffer B and applied to a 1 x 2 cm column of DEAE Sephadex. The enzyme was eluted with 12 ml of a 0.1 M to 0.6 M NaCl gradient in buffer B. Fractions of 0.5 ml were collected and assayed for activity.

A second milliliter (150 µg) of RNA polymerase from the heparin Sepharose column was applied to a 15 to 30% glycerol gradient made in buffer A with 0.5 M NaCl (in a 9/16 by 3 1/2" tube). After centrifugation for 24 hr at 37,000 rpm in a SW41 rotor at 4°C, the bottom of the tube was pierced and 0.5 ml fractions were collected and assayed for activity.

2.3. RNA polymerase assays. RNA polymerase assays for monitoring polymerase activity through its purification contained: 48 mM Tris-HCl, pH 8.2; 8 mM MgCl$_2$; 40 mM NaCl; 10% glycerol; 1.6 mM ATP, CTP, and GTP; 16 µM UTP; 28 nM $^3$H-UTP (14 µCi/nmole); 5 µg salmon sperm DNA; and various amounts of RNA polymerase in a total volume of 500 µl. The tubes were incubated at
34°C for 10 min and RNA was precipitated by addition of 1.0 ml of 10% trichloroacetic acid. One unit of activity was defined as one nmole of UTP incorporated in 10 min. The amount of protein present was determined by measurement of Coomassie blue dye binding (Bio-Rad), with crystallized bovine serum albumin (BSA) as standard. Specific activity was defined as units of activity per mg of protein.

3. Run-off transcription assays

Run-off transcription assays contained: 60 mM Tris-HCl, pH 8.2; 8 mM MgCl₂; 40 mM KCl; 13% glycerol; 16 mM CTP, UTP, and GTP; 1.6 μM ATP; 66 nM ³²P-ATP (756 μCi/nmole); 1 μg template DNA; and RNA polymerase in a 4:1 molar ratio of RNA polymerase:DNA, in a volume of 200 μl. The tubes were placed at 34°C for 20 min. One tenth of the reaction mixture was diluted into 1.0 ml of 10% trichloroacetic acid for measurement of the amount of radioactivity incorporated into RNA, and the remainder was extracted with an equal volume of phenol:chloroform (1:1), then with chloroform, and then, after the addition of 1/10 volume of 3M sodium acetate, the RNA was precipitated by addition of 0.5 ml 95% ethanol.
4. Polyacrylamide gel electrophoresis

4.1. Slab gel electrophoresis of RNA polymerase. Electrophoresis of RNA polymerase was performed as described by Laemmli [31] with a 14 to 20% gradient of polyacrylamide [35]. Aliquots of the RNA polymerase preparation were diluted in protein loading buffer (20% glycerol; 10% 2-mercaptoethanol; 0.0625 M Tris-HCl, pH 6.9; 0.1% bromphenol blue), and heated at 90°C for 5 min before loading onto the gel. The gels were run at 100V for 4 hr and then stained overnight in 0.1% Coomassie blue stain (in 30% methanol and 10% acetic acid). Destaining was accomplished in the above solution with no dye. \textit{E. coli} RNA polymerase subunits were used as molecular weight markers.

4.2. Polyacrylamide gel electrophoresis of RNA transcripts. Gel electrophoresis was done in 5% polyacrylamide-7 M urea gels, in 0.5X TBE buffer [32]. After a 70% ethanol wash, the precipitated samples were dissolved in 20 μl of 80% formamide; 0.5X TBE; 0.025% xylene cyanol; 0.025% bromophenol blue, and then heated for 5 min at 90°C before loading. Radiolabelled DNA from the single stranded phage M13mp11, digested with \textit{HaeIII}, was used for molecular length markers. The gels were electrophoresed at 100V for 3 hr, dried, and used for autoradiography.
5. Preparation of DNA and cellular RNA

5.1. Preparation of DNA. Salmon sperm DNA was obtained from Sigma. *R. capsulatus* chromosomal DNA and plasmid DNA were purified by CsCl gradient ultracentrifugation [32]. Fragments of plasmid DNA were generated by restriction enzyme digestion, gel electrophoresis, and electroelution [32].

5.2. Purification of cellular RNA. Cellular RNA was purified as described [48] from cells that had been grown under high $O_2$ conditions to 80 Klett units ($3.2 \times 10^8$ cells/ml) and then shifted to low $O_2$ conditions [8] for 45 min prior to cell harvest.

6. Plasmid constructions

6.1. Construction of the promoter-identification vector pXCA601. This vector was used for in-frame gene fusions and was constructed by Cam Adams in Dr. S.N. Cohen's lab at Stanford. The fragment containing the *lacZ* gene was obtained from plasmid pMC1403Plac [24]. The 3' ends of fusion transcripts were stabilized by insertion of the *ompA* terminator [7]. To prevent transcription initiated within the vector from continuing into the DNA inserts being tested for
Fig. 4. A representation of the promoter identification vector pXCA601. The thin line represents vector DNA, the boxes designate sequences of importance as follows: T1, the T4 phage terminator; T2, the \textit{ompA} terminator; and \textit{lacZ}, the \textit{lacZ} gene of \textit{E. coli}. The letters designate restriction endonuclease digestion sites as follows: P-PstI, B-BamHI, S-SalI. The arrow shows the direction of transcription and Tc\textsuperscript{R} implies tetracycline resistance.
promoter activity the T4 phage transcription-translation termination signals [38] were introduced upstream of the insertion site. The replicon for this vector was derived from pTJS133 [39], which can be mobilized into *R. capsulatus*. A diagram of pXCA601 is shown in Figure 4.

6.2. Construction of the *lacZ* fusion plasmids. Figure 5 shows a flowchart of the construction of the *puf* promoter mutants used for this thesis work. The exonuclease III deletions were made as described [23], with *XhoI*-cleaved plasmid pJAJ21 [26] as substrate. After completion of the procedure the DNA was cleaved with *BamHI* to release shortened fragments which were then sized on a 5% polyacrylamide gel. The fragments were recovered by electroelution and religated into M13mp18 that had been cleaved at the *HincII* and *BamHI* sites. The resultant phages were screened to identify those that contained fragments of the desired length, which were then subcloned as *PstI* to *BamHI* fragments into pXCA601 to create pΔ4, pΔ14, and pΔ24.

The 935 construct was made by digestion of pJAJ21 with *PstI* and *BamHI* and subcloning of the insert into pXCA601. The 932 deletion was made by a complete *SalI* digestion of pJAJ21 and purification of the larger fragment, followed by recircularization. The resultant plasmid (pUC932) was cleaved with *PstI* and *BamHI* and the smaller fragment was cloned into pXCA601.

The pAMSP plasmid was made from pUCESp, a derivative of pJAJ21, as
Fig. 5. Construction of the puf promoter deletion mutants in pXCA601. The thin lines represent vector DNA, the thicker lines represent <i>R. capsulatus</i> chromosomal DNA, and the boxes represent the <i>pufQ</i> gene and the 5' end of the <i>pufB</i> gene (see Fig. 11). The letters designate restriction endonuclease digestion sites as follows: P-<i>PstI</i>, S-<i>SalI</i>, X-<i>XhoI</i>, A-<i>AccI</i>, E-<i>EcoRI</i>, M-<i>MnII</i>, B-<i>BamHI</i>. For p935, pA4, pA14, pA24, pAM, pMSP, and p932 the small <i>PstI</i> to <i>BamHI</i> fragments obtained were cloned into pXCA601 (that had been digested with <i>PstI</i> and <i>BamHI</i>) to yield the final constructs shown. The diagrams of the constructs are not to scale. A detailed description of each construction can be found in section 6.2 of the Materials and Methods.
M′E

pJAJ21

Pst I and Bam HI, small fragment

p935

S

Xho I, exonuclease III,

Klenow, Bam HI, small fragments,

M13mp18 (Hind III and Bam HI),

Pst I and Bam HI, small fragments

X

pUCESp

Sal I, large fragment, ligation

M

B

pUCΔMSP

B

P

Pst I, mung bean nuclease,

Bam HI, small fragment

Pst I and Bam HI, small fragment

pUCΔAM′

B

A'/M'

S

XI

pΔMSP

B

P

Pst I and Bam HI, small fragment

p932

pΔAM

A'/M'

S

XI

M

E

S

P

X

I

Sal I, Taq I methylase,

Eco RI, large fragment,

Klenow, ligation

M

pUC13 cut at Hind III and Bam HI

M11 and Bam HI,
c.a. 100 bp fragment,

B

pUC932

M

E

S

P

B

pΔ4, Δ14, Δ24

B

M

E

S

P

Sal I, large fragment,

ligation

B
described below. The plasmid pJAJ21 was digested with SalI, methylated with TaqI methylase enzyme, digested with EcoRI and the large fragment purified. Methylation blocked digestion of the EcoRI sites found 3' of the pufB gene segment in pJAJ21 [26] so that only the EcoRI site shown in Figure 5 was cleaved. The purified fragment was treated with Klenow enzyme to form blunt ends and religated to produce pUCESp. The pUCESp plasmid was digested to completion with MnlI and BamHI, and the 100 base pair fragment containing the 5' end of the pufB gene was purified and subcloned into pUC13 digested with HincII and BamHI to create pUCΔMSP, yielding a PstI site upstream of the pufB ribosome binding site. The approximately 110 base pair PstI to BamHI fragment was then subcloned into pXCA601 to give pΔMSP.

The pΔAM construct was made by digestion of the plasmid pUCΔMSP with PstI, followed by mung-bean nuclease digestion to create a blunt end, and then digested with BamHI. The smaller fragment released was purified and subcloned into pJAJ21 that had been treated with TaqI methylase (to protect the SalI sites from cleavage by the AccI enzyme), digested with AccI, the site filled in with Klenow fragment [32], followed by digestion with BamHI. The larger fragment was purified and ligated with the small fragment from pUCΔMSP. The resultant plasmid, named pUCΔAM, was then digested with BamHI and PstI and the smaller fragment was subcloned into pXCA601. After transformation into E. coli C600 the plasmids were conjugated into strains of R. capsulatus [18].
6.3. Site-directed mutagenesis. Figure 6 shows a flowchart of the process of oligonucleotide-directed mutagenesis. The single-stranded template used was M13mp18Δ4, which contained the Δ4 deletion (shown in Table IV, and Figs. 3 and 5) inserted between the HincII and BamHI sites of M13mp18. The procedure used was adapted from Zoller and Smith [60]. The oligonucleotides were annealed to the templates in a 4:1 molar ratio of oligonucleotide:template. Specific priming was tested for by dideoxy sequencing using the mutagenic oligonucleotide as primer, and completion of second strand synthesis was ascertained by restriction endonuclease digestion. After annealing, extension and ligation the DNA was transformed directly into E. coli TM101. Enrichment for mutants was achieved by use of E. coli RZ1032 (ung, dut) for production of deoxyuracil enriched template [29]. The potential mutants were screened by dot-blot hybridization with the appropriate labelled oligonucleotide as probe with washes of increasing temperature, and candidates were verified by dideoxy sequencing of the PstI to AccI region. The RF forms of the M13 mutants were purified by standard techniques [32] and digested with PstI and AccI to obtain the mutated regions which were used to replace the unmutated PstI to AccI region of pJAJ21. The mutant derivatives of pJAJ21 were then digested with PstI and BamHI to subclone into pXCA601. The Δ42 mutant was created with the oligonucleotide 5'-GAAGATTTATCTAGACGCTTCCT-3', so that sequences
Fig. 6. Construction of oligonucleotide-directed mutations in pXCA601. The thin lines represent vector DNA, the thicker lines represent *R. capsulatus* chromosomal DNA, and the boxes represent the *pufQ* and the 5' end of the *pufB* genes (see Fig. 11). The dashed lines designate uricil-rich DNA, the thick arrow represents primer hybridization, and the thinner arrows represent second-strand synthesis. The mutations were introduced by synthesized oligonucleotides that were used for the primers. The letters designate restriction endonuclease digestion sites as follows: P-PstI, A-AccI, B-BamHI. Although the figure represents the construction of the Δ41 deletion the same procedure was performed for the Δ42 and Δ44 mutations.
124-165 of Figure 3 were replaced with 5'-AGA-3' to generate an XbaI site. The Δ41 construct was created with the oligonucleotide 5'-CTTCCTTCTAGACCCCCCTTCAT-3' so that sequences 176-205 of Figure 3 were replaced with 5'-TAG-3' to create an XbaI site, and the Δ44 construct was made with the oligonucleotide 5'-CATGGGTTGGCTGGGTAGCGTC-3' so that the A residues at 222 and 224 of Figure 3 were replaced by G residues. After transformation into E. coli C600 the plasmids were conjugated into strains of R. capsulatus [18].

6.4. Construction of the pufQ gene deletion. The plasmid pJAJ21 (see Fig. 11) was digested with Ncol, treated with TaqI methylase, digested with EcoRI, and the large fragment was purified and religated to form pUCΔQ after formation of blunt ends with Klenow fragment (this created a new EcoRI site at the ligation point). The construct pUCΔQ was either digested with PstI and BamHI and the small fragment cloned into pJAJ103::lac903 (a derivative of pJAJ103 [26,62]), or digested with HindIII and EcoRI and the small fragment cloned into pTB999 (a derivative of pRCVI [11], see Fig. 14). After transformation into E. coli the constructs were conjugated into strains of R. capsulatus [18].

7. S1 nuclease mapping of RNA

7.1. S1 nuclease mapping of puf operon RNA ends. S1 nuclease
(Bethesda Research Laboratories) experiments were performed as described [48], with hybridization at 55°C for 3 hr and S1 nuclease treatment for 30 min at 37°C. After one ethanol precipitation the samples were denatured in formamide loading dye (see section 4.2) and loaded onto 6% polyacrylamide-urea gels. After electrophoresis and drying the gels were used for autoradiography.

7.2. S1 nuclease protection experiments for detection of RNA from the puc operon. The probe used was an anti-sense RNA molecule which had been transcribed by T7 RNA polymerase from a pT7-2 vector (United States Biochemical Corporation) that contained the 5' end of the puc operon [61]. About 500 ng of linearized DNA template were used to homogeneously label RNA with T7 RNA polymerase as recommended by the supplier (Pharmacia). After phenol extraction and ethanol precipitation the samples were resuspended in DNase buffer and treated with 23 units of DNase I [61], and then phenol extracted and precipitated again. About 5 μg of RNA were obtained with a specific activity of 8 x 10^5 dpm/μg RNA. This labelled anti-sense RNA was hybridized to RNA from B10 or ΔRC6 (purified as described [48]), or to yeast tRNA, and then treated as for an S1 nuclease RNA end-mapping experiment (see section 7.1).
8. Primer extension experiments

Primer extension mapping of 5' RNA ends was performed as described [52]. Ten μg of cellular RNA were mixed with a 5'-32P end-labelled oligodeoxyribonucleotide primer, with the sequence of 5'-GTGAAGCTCAG GTCGTTCTTAT-3' (complementary to mRNA transcribed from bases 932 to 953 of the puf 5' DNA sequence [see Fig. 3]) and hybridized at 55°C for 3 hr. After extension, phenol extraction, and ethanol precipitation the samples were denatured in formamide loading dye (see section 4.2) and run on a 6% polyacrylamide-urea gel.

9. Capping experiments

Cellular RNA was capped essentially as described [36,61]. Forty μg of RNA were mixed with 100 pmoles of alpha 32P-GTP (3000 Ci/mmole) and 12.5 units of guanylyl transferase (Bethesda Research Laboratories, Inc.). After incubation and phenol extraction the RNA was ethanol precipitated three times to remove unincorporated label, and resuspended in RNA storage buffer (20 mM sodium phosphate, pH 6.5/1 mM EDTA). An average of 4 x 10^4 dpm/μg of RNA were obtained in successful capping experiments. Approximately 8 μg of capped RNA were run on a 6% polyacrylamide-urea gel to obtain an autoradiogram for examination to ensure uniform capping.
Capped RNA was hybridized at 55°C for 3 hr to DNA probes as described [36,61] and after S1 nuclease (750U) digestion at 37°C the samples were digested with 2.5 ng of boiled RNase A at room temperature. The products were run on 6% polyacrylamide-urea gels after denaturation in formamide loading dye (see section 4.2).

10. Absorption spectroscopy of various strains of *R. capsulatus*

Cells were grown under high or low O₂ conditions [8] in RCV medium [49] as described. Equal amounts of cells were resuspended in 22.5% BSA and scanned in a Varian DMS 100 spectrophotometer.
RESULTS

1. Purification and characterization of RNA polymerase

1.1. Purification of RNA polymerase

RNA polymerase was purified from both chemoheterotrophically and
photoheterotrophically grown cells as described in Materials and Methods.
Figure 7 shows a typical elution of RNA polymerase activity from the Biogel
A1.5m column, the DNA cellulose column and the heparin Sepharose column.
Although salmon sperm DNA was routinely used as a template to monitor the
polymerase activity throughout the purification, comparable results were
obtained with R. capsulatus chromosomal DNA.

A typical purification of RNA polymerase is summarized in Table I.
The overall yield was usually between 10 and 15%, with an increase in specific
activity of 100-200 fold, which compare favourably with purifications of other
bacterial RNA polymerases [10,50]. The final specific activity obtained was from
50 to 100 units/mg protein. Fractions containing the peak of activity from the
heparin Sepharose column were pooled and then separated into aliquots. The
purified RNA polymerase was either stored at -20°C in the elution buffer (buffer
Fig. 7. Chromatography of RNA polymerase from chemoheterotrophically grown cells. Fig. 1a, elution from BioGel A1.5m; Fig. 1b, elution from DNA cellulose; Fig. 1c, elution from heparin Sepharose. The supernatant fluid applied to the columns were purified as described in Materials and Methods. Fractions were collected and assayed for RNA polymerase activity, (•–•). The optical density of the fractions at 280 nm was also measured, (o–o). The arrows designate the fraction at which the [NaCl] was changed as described in Materials and Methods.
Table I: Summary of RNA polymerase purification

One unit of activity is defined as one n mole of UTP incorporated in 10 min at 34°C.

<table>
<thead>
<tr>
<th>Purification step:</th>
<th>mg protein</th>
<th>units of activity</th>
<th>units/mg protein</th>
<th>% yield</th>
<th>fold purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>crude extract:</td>
<td>2241.0</td>
<td>594</td>
<td>0.26</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0-30% (NH₄)₂SO₄ supernatant:</td>
<td>405.0</td>
<td>986</td>
<td>2.44</td>
<td>166</td>
<td>9</td>
</tr>
<tr>
<td>30-60% (NH₄)₂SO₄ pellet:</td>
<td>294.0</td>
<td>642</td>
<td>2.18</td>
<td>108</td>
<td>8</td>
</tr>
<tr>
<td>Bio-Gel A1.5m eluate:</td>
<td>67.0</td>
<td>495</td>
<td>7.40</td>
<td>83</td>
<td>28</td>
</tr>
<tr>
<td>DNA cellulose eluate:</td>
<td>3.6</td>
<td>160</td>
<td>45.70</td>
<td>27</td>
<td>176</td>
</tr>
<tr>
<td>heparin Sepharose eluate:</td>
<td>1.4</td>
<td>87</td>
<td>60.70</td>
<td>15</td>
<td>234</td>
</tr>
</tbody>
</table>
B containing 0.6 M NaCl) supplemented with 40% glycerol, or at -80°C in the same buffer containing 10% glycerol. Specific activity declined during storage under both conditions over a period of 3-4 months to about 50% of the initial activity, and then stabilized.

Several DNAs were compared as templates for the *R. capsulatus* anaerobic RNA polymerase. (Table II). The greatest specific activity was obtained with phage M13 RF as template and *R. capsulatus* DNA gave the lowest activity. Phage M13 differed from the other templates in that it was supercoiled. Although the *R. capsulatus* DNA was linear, like the other templates tested, it has a much higher GC content (67% as opposed to 41-50% for the other templates [6,43]) and this may have caused its lower activity.

1.2. Optimization of the assay conditions

After the first purification was completed the assay conditions were optimized for the concentration of DNA, nucleotides, and other components, and for temperature and pH. The assay conditions used in the first purification were taken from published procedures for *in vitro* transcriptions [44], but subsequent purifications utilized the conditions given in the Materials and Methods. The optimization increased activity three-fold.

Conditions were optimized with either *R. capsulatus* chromosomal DNA or salmon sperm DNA as template; optimal conditions for the two
Table II: Utilization of different templates by *R. capsulatus* anaerobic RNA polymerase

Specific activity is in units/mg protein; assay conditions were those used for salmon sperm DNA (see Materials and Methods), 10 µg of each DNA template were used.

<table>
<thead>
<tr>
<th>Specific activity</th>
<th>DNA template</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>Salmon sperm</td>
</tr>
<tr>
<td>35</td>
<td>T7</td>
</tr>
<tr>
<td>83</td>
<td>M13mp11</td>
</tr>
<tr>
<td>23</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>9</td>
<td><em>R. capsulatus</em></td>
</tr>
</tbody>
</table>
templates were very similar. The changes made for use of *R. capsulatus* DNA as template were to increase the glycerol content to 13%, increase the Tris-HCl buffer to 60 mM, and to include 40 mM KCl instead of NaCl.

The enzyme had an absolute requirement for MgCl₂, DNA, and nucleotides. The optimal pH was 8.2 for both templates, with a sharp drop in activity below pH 7.0. A concentration of at least 25 mM Tris-HCl buffer was necessary for activity. There was no consistent stimulation of activity when spermidine, EDTA, or 2-mercaptoethanol were added to the reaction mixture.

1.3. Subunit composition of RNA polymerase

The purified polymerase was subjected to SDS polyacrylamide gel electrophoresis (SDS-PAGE) to assess its subunit composition. Figure 8 shows the results of an experiment in which RNA polymerase preparations purified from aerobic and anaerobic cultures of *R. capsulatus* were compared to a preparation of *E. coli* RNA polymerase. The gel contained a number of bands representing proteins that were associated with *R. capsulatus* RNA polymerase activity. Of the 10-11 visible bands, six were designated as RNA polymerase subunits because they consistently co-purified with RNA polymerase activity through two additional purification steps (see below). The designation and molecular mass (estimated by comparison with *E. coli* RNA polymerase subunits as standards) of the subunits are: beta and beta-prime subunits of about 150,000 to 160,000.
Fig. 8. SDS polyacrylamide gel electrophoresis of purified RNA polymerase from chemoheterotrophically and photoheterotrophically grown *R. capsulatus*. Ten μg of protein from the heparin Sepharose eluate, along with ten μg *E. coli* RNA polymerase, were subjected to electrophoresis and stained as described in Materials and Methods. The subunit designations for *R. capsulatus* RNA polymerase are given on the left, and the subunits of *E. coli* RNA polymerase are labelled on the right. Lane 1: RNA polymerase from chemoheterotrophically grown cells. Lane 2: RNA polymerase from photoheterotrophically grown cells. Lane 3: *E. coli* RNA polymerase.
Daltons; an alpha subunit of about 45,000 Daltons; and a sigma subunit of 70,000 Daltons. Two smaller proteins, tentatively designated as omega factors, were also consistently observed. The pattern of bands in the gel was identical for aerobic and anaerobic RNA polymerase preparations. However there was a much larger amount of one protein, which migrated to a position between the putative alpha and sigma subunits, associated with the polymerase activity purified from cells grown aerobically.

Because of the large number of proteins present in the preparations analyzed in Figure 8, aliquots of the polymerase were put through two additional purifications to see if some of the proteins could be removed without loss of RNA polymerase activity. The methods used were centrifugation through a glycerol gradient and ion-exchange chromatography over a DEAE-Sephadex column.

A single peak of activity was observed with chromatography over DEAE-Sephadex, and SDS-PAGE revealed about 7-8 bands in the three fractions with greatest activity (Fig. 9). The specific activity of the peak fraction (lanes 3, Fig. 9) in the preparation from aerobic cells increased 1.5-fold, whereas the specific activity of the peak fraction of RNA polymerase from anaerobic cells remained the same.

The results from the glycerol gradient were comparable. The three fractions containing the greatest activity were analyzed by SDS-PAGE as shown
Fig. 9. SDS polyacrylamide gel electrophoresis of RNA polymerase after DEAE-Sephadex chromatography. The subunit designations for *R. capsulatus* RNA polymerase are given on the left of each gel. A, RNA polymerase from chemoheterotrophically grown cells. B, RNA polymerase from photoheterotrophically grown cells. Lanes 1: 10 μg of RNA polymerase from the heparin Sepharose eluate. Lanes 2 to 4: peak fractions from the DEAE Sephadex column (equal volumes of each loaded with lane 3 containing 10 μg of protein). Lanes 5: 10 μg of *E. coli* RNA polymerase.
in Figure 10. The number of bands present was reduced to 7 or 8, including the 6 that have been designated as RNA polymerase components. The specific activity of the peak fraction in the preparation from aerobic cells increased 3-fold, whereas the specific activity of the polymerase from anaerobic cells increased only slightly.

Only the gel bands that I have designated as being representative of *R. capsulatus* RNA polymerase subunits were visible in both preparations after both of the additional purification steps. On occasion extra bands were seen to co-purify with the RNA polymerase. These bands were not present in every purification performed so it is unlikely that they were components of the RNA polymerase that were essential for activity.

1.4. Gel electrophoresis of run-off transcription products

The RNA polymerase preparations used for run-off transcriptions were obtained by step elution from heparin Sepharose with 0.6 M NaCl (see Figs. 7 and 8). A restriction endonuclease site map of a recombinant plasmid, designated pJAJ21 [26], which contains the *R. capsulatus* *puf* transcriptional regulatory region, is depicted in Figure 11. The fragment of *R. capsulatus* DNA inserted into this plasmid has been shown to contain an O$_2$-regulated promoter [26], and is shown in Figure 2b.

Preliminary *in vitro* transcription assays were performed to determine
Fig. 10. SDS polyacrylamide gel electrophoresis of RNA polymerase after glycerol gradient centrifugation. The subunit designations for *R. capsulatus* RNA polymerase are given on the left of each gel. A, RNA polymerase from chemoheterotrophically grown cells. B, RNA polymerase from photoheterotrophically grown cells. Lanes 1: 10 μg of RNA polymerase from the heparin Sepharose eluate. Lanes 2 to 4: peak fractions of activity from the glycerol gradient (equal volumes were loaded with lane 3 containing 10 μg protein). Lanes 5: 10 μg *E. coli* RNA polymerase.
if the presence of *R. capsulatus* DNA in the pUC13 vector stimulated activity. Table III shows the results from both aerobic and anaerobic RNA polymerase. The linearized pUC13 and pJAJ21 templates were present in equimolar amounts. The presence of *R. capsulatus* DNA caused an increase of 20-30% in activity with both the forms of RNA polymerase.

Run-off transcription assays were performed as described in Materials and Methods. An example of an autoradiogram of a polyacrylamide-urea gel of RNA transcripts produced by a preparation of RNA polymerase from anaerobically grown cells is shown in Figure 12. The template consisted of 1 μg of either the plasmid pUC13 or pJAJ21, digested with BamHI and HindIII, and various amounts of heparin were added to the standard reaction assay. It had been found previously that inclusion of heparin reduced a background that could obscure bands due to specific transcripts. About 3-4 distinct RNA species were visible over the entire range of heparin concentrations. Especially notable was an approximately 104 nucleotide transcript that was probably RNA I (a predominant transcript involved in plasmid replication and copy number [42]). The additional transcripts seen ranged from approximately 500 to about 2500 nucleotides in length. A longer exposure of the autoradiogram revealed no additional transcripts, but the resolution of the bands at the top of the gel was lost. It can be seen that the pattern of bands in the autoradiogram is the same when either pUC13 or pJAJ21 were used as template. It was concluded
Fig. 11. A representation of the construct pJAJ21 [26]. The thin line represents pUC13 vector DNA, the thicker line represents the puf promoter region of R. capsulatus, and the boxes represent the genes present on the R. capsulatus insert. The arrow designates the direction of transcription, and the letters represent restriction endonuclease digestion sites as follows: H-HindIII, P-PstI, X-XhoII, N-NcoI, S-SalI (not unique), E-EcoRI (not unique), B-BamHI. The ApR designation implies ampicillin resistance.
Table III: Utilization of linear templates used for run-off transcriptions

Specific activity is in units/mg protein; assay conditions were those used for *R. capsulatus* DNA (see Materials and Methods), equimolar amounts of each template were used.

<table>
<thead>
<tr>
<th>RNA polymerase</th>
<th>DNA template</th>
<th>Specific activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>aerobic</td>
<td>pUC13</td>
<td>18</td>
</tr>
<tr>
<td>aerobic</td>
<td>pJAJ21</td>
<td>24</td>
</tr>
<tr>
<td>anaerobic</td>
<td>pUC13</td>
<td>23</td>
</tr>
<tr>
<td>anaerobic</td>
<td>pJAJ21</td>
<td>28</td>
</tr>
</tbody>
</table>
Fig. 12. Polyacrylamide gel electrophoresis of run-off transcripts obtained with photoheterotrophic (anaerobic) RNA polymerase. The templates used were: A, pUC13, and B, pJAJ21, both digested with BamHI and HindIII. Reaction conditions and gel electrophoresis were as described in Materials and Methods, except that the following amounts of heparin were added to the reactions: lanes 1, 0.1 μg; 2, 0.3 μg; 3, 1.0 μg; 4, 3.0 μg; 5, 10.0 μg.
that under these conditions the amount of transcription originating from the *R. capsulatus* promoter present in pJAJ21 was undetectably low compared to vector-derived activity.

In other experiments the cloned fragment of *R. capsulatus* DNA was purified and used as a template for run-off transcription. Figure 13 shows that transcription of the insert alone yielded a large number of transcripts of various sizes, and that an increase in the concentration of heparin in the reaction decreased the intensity of all bands equally.

Because the pattern of transcripts seen with the purified fragment could have resulted from initiation of transcription at sites other than promoters, the concentrations of various components of the transcription assay were titrated in an attempt to increase specificity of transcription. For example, the concentrations of glycerol and KCl were varied and titration of spermidine concentration was also attempted. These conditions did not reduce the number of bands seen in autoradiograms. Although this preparation of *R. capsulatus* RNA polymerase could be used to initiate transcription at promoters present on the plasmid pUC13, additional factors appear to be necessary for efficient recognition of the *puf* promoter. However, inclusion of cruder fractions from earlier stages in the purification did not reduce the number of bands either.
Fig. 13. Polyacrylamide gel electrophoresis of run-off transcripts obtained with photoheterotrophic (anaerobic) RNA polymerase and the purified *R. capsulatus* promoter fragment as template. The fragment was purified by digestion of pJAJ21 with BamHI and HindIII, agarose gel electrophoresis, and electroelution [32]. Reaction conditions and gel electrophoresis were as described in Materials and Methods, except that the following amounts of heparin were added: lane A, 0.01 μg; B, 0.03 μg; C, 0.1 μg; D, 0.3 μg; E, 1.0 μg; F, 3.0 μg; G, 10.0 μg.
2. Localization of the \textit{puf} promoter

2.1. Localization of the \textit{puf} promoter through gene fusion experiments

Initial studies were performed with the use of a vector called pJAJ103::lac903, a derivative of pJAJ103 [26,62], that contained the \textit{lacZ} gene (see Appendix). Unique \textbf{PstI} and \textbf{BamHI} sites upstream of the \textit{lacZ} gene allowed for the cloning of \textit{puf} operon promoter fragments from pJAJ21 into the vector as operon fusions. The search for the promoter at one time focussed on regions immediately upstream of the position where the 5' end of the predominant RNA transcript mapped (end 4, Fig. 2). This region initially appeared to have promoter activity, but it was later concluded that this was due to read-through transcription from vector sequences (see Appendix). When this was discovered a new vector was obtained that eliminated read-through transcription with the T4 terminator sequences, and allowed for gene fusions with \textit{lacZ}. This vector was designated pXCA601 (see Materials and Methods).

A series of fragments from the \textit{puf} operon promoter region was cloned into this vector and tested for promoter activity. Construction of all of these plasmids is described in detail in Materials and Methods. The constructs were mated into \textit{R. capsulatus} and beta-galactosidase assays were performed on cells grown under low and high O$_2$ conditions, enabling both activity and oxygen
regulation of the promoter to be studied. All of the plasmids were mated into the ΔRC6 strain of *R. capsulatus* which has been deleted for the *puf* operon [11]. This strain was used so that results obtained could be directly compared to results obtained from RNA end-mapping experiments (described further on in this thesis). Equivalent results were obtained with the use of the wild-type strain B10 (J. T. Beatty and C. W. Adams, personal communication).

Table IV shows representations of these fragments and gives the beta-galactosidase activities obtained. The MSP construct contained a fragment of the *puf* operon that extended from the *MnlI* site, 6 base pairs upstream of the Shine-Dalgarno sequence of the *pufB* gene, to a *BamHI* site introduced after the 20th codon of the *pufB* gene [26]. The activities obtained with this fragment were considered to be background level, and showed the absence of read-through transcription from the vector.

The fragment designated 935 is the fragment described previously that was cloned into pUC13 to create pJAJ21 (see Figs. 2 and 11) [26]. This fragment in the pXCA601 vector gave strong expression of the *lacZ* gene as seen by the high activities obtained. More than 6-fold higher activity was obtained under low O₂ conditions as opposed to high O₂ conditions, supporting the concept that the fragment contains the O₂-regulated promoter for the *puf* operon (Table IV).

Removal of about 180 base pairs from the 5' end of 935 to yield the Δ4 construct had a minor effect on the expression of *lacZ*, but removal of another
Table IV. Assays of beta-galactosidase activities of cells containing puf 5' regions fused to the lacZ gene of pXCA601

1 Representation of the constructs used for puf promoter mapping. Restriction sites are indicated by vertical lines and are labelled in the top construct as follows: X-XhoI, A-Accl, M-MnlI. The structural genes are designated by thick lines and are labelled in the 935 representation. The dashed line represents the sequences that were removed in the ΔAM construct. The hillocks represent the palindromic sequences that were studied and the x's designate sites of oligonucleotide-directed deletions or mutations.

2 Activities are expressed as nmoles ONPG/min/mg protein. The values in brackets are standard deviations of 3 to 7 assays that were performed on each construct.

3 The ratio of activities obtained under low O2 versus high O2.
<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>REPRESENTATION</th>
<th>1</th>
<th>LOW O$_2$</th>
<th>2</th>
<th>HIGH O$_2$</th>
<th>2</th>
<th>L / H</th>
</tr>
</thead>
<tbody>
<tr>
<td>935</td>
<td><img src="935" alt="Diagram" /></td>
<td></td>
<td>3181 (190)</td>
<td>494 (36)</td>
<td></td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>ΔMSP</td>
<td><img src="%CE%94MSP" alt="Diagram" /></td>
<td>12</td>
<td>6 (9)</td>
<td>2</td>
<td></td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Δ4</td>
<td><img src="%CE%944" alt="Diagram" /></td>
<td></td>
<td>2092 (556)</td>
<td>327 (83)</td>
<td></td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>Δ14</td>
<td><img src="%CE%9414" alt="Diagram" /></td>
<td></td>
<td>52 (5)</td>
<td>28 (2)</td>
<td></td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Δ24</td>
<td><img src="%CE%9424" alt="Diagram" /></td>
<td></td>
<td>64 (10)</td>
<td>29 (3)</td>
<td></td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>932</td>
<td><img src="932" alt="Diagram" /></td>
<td></td>
<td>29 (5)</td>
<td>16 (2)</td>
<td></td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>ΔAM</td>
<td><img src="%CE%94AM" alt="Diagram" /></td>
<td></td>
<td>3550 (771)</td>
<td>806 (209)</td>
<td></td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Δ41</td>
<td><img src="%CE%9441" alt="Diagram" /></td>
<td></td>
<td>37 (8)</td>
<td>23 (4)</td>
<td></td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Δ42</td>
<td><img src="%CE%9442" alt="Diagram" /></td>
<td></td>
<td>1983 (338)</td>
<td>431 (110)</td>
<td></td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>Δ44</td>
<td><img src="%CE%9444" alt="Diagram" /></td>
<td></td>
<td>111 (10)</td>
<td>33 (3)</td>
<td></td>
<td>3.4</td>
<td></td>
</tr>
</tbody>
</table>
160 base pairs to create the Δ14 construct decreased expression by 90-95%. The Δ14 level of expression is almost as low as the MSP expression and the ratio of low to high O₂ activities seems to show the loss of O₂ regulation. Further deletions to give the Δ24 construct and the 932 construct had no additional effect on activity (Table IV).

The Δ14 construct showed that sequences upstream of this deletion were essential for promoter activity, so an internal deletion was constructed between the AccI site and the MnlI site (see Figs. 2 and 3) and designated ΔAM. This construct, shown in Table IV, retained both full activity as well as O₂ regulation. Thus it can be concluded that sequences between the 5' end of the Δ4 construct and the AccI site are necessary and sufficient for O₂-regulated initiation of transcription.

The sequence within this transcriptional control region was examined and two palindromes (inverted repeats) were discovered (see Fig. 3). Inverted repeats have been associated with transcriptional regulation in other systems [37] so these were each deleted separately by oligonucleotide-directed mutagenesis (see Materials and Methods). The more upstream of the two sequences, from nucleotides 124 to 165 in Figure 3, was replaced by 5'-AGA-3' to yield the construct Δ42. As can be seen in Table IV the beta-galactosidase activities of cells containing this construct were comparable to the Δ4 construct from which the deletion came. However, replacement of the other palindrome, from nucleotides
176 to 205, with the sequence 5'-TAG-3' (to yield the construct Δ41) resulted in a
dramatic reduction in lacZ expression to about the levels seen with the Δ14 construct.

Transcriptional regulation of the *R. capsulatus* puc operon, which encodes structural proteins for the B800-850 light-harvesting complexes, has also been studied [61]. Two initiation sites have been mapped near a direct repeat of 7 base pairs with the sequence 5'-ACACTTG-3', found just upstream of the puc structural genes. The region immediately upstream of the AccI site of the puf operon contains a sequence similar to this, 5'-ACATGGG-3', near nucleotide 225 (Fig. 3). This sequence was changed to 5'-GCGTGGG-3' by oligonucleotide-directed mutagenesis of the Δ4 construct to yield the Δ44 construct. Table IV shows that mutation of two base pairs within this region reduced expression 90-95% compared to the Δ4 construct.

2.2. Absorption spectrophotometric analysis of puf promoter mutants

In order to determine if results obtained from fusions of the puf promoter region and various mutants of this region with the lacZ gene were comparable to results from fusions of these fragments with puf operon genes, a vector was obtained from J. T. Beatty that contained all of the puf operon. This vector, designated pTB999, is shown in Figure 14 and is a derivative of pRCVI [11]. The effects of mutations in the regulatory region of the puf operon can be
observed by replacement of the \textit{Hind}III to \textit{EcoRI} fragment of the regulatory region with various mutated fragments. These constructs were mated into a strain of \textit{R. capsulatus} called U43. This strain has had chromosomal \textit{puf} sequences from an \textit{Apal} site in the \textit{pufQ} gene to an \textit{Apal} site just beyond the \textit{pufX} gene replaced with a DNA fragment that determines spectinomycin resistance [55]. This replacement of \textit{puf} operon sequences was done in a strain that previously had been isolated as a mutant deficient in B800-850 complexes, and the result is that no light-harvesting or reaction center complexes are detected in U43 cells [55]. This strain is useful for analysis of plasmid-borne mutant copies of \textit{puf} operon sequences because the absorption spectrum of B870 complexes is not masked by B800-850 complex absorption (see Introduction).

The deletions tested with pTB999 in U43 were the Δ4, Δ41, and Δ42 constructs (see Table IV), to compare the effects of loss of either of the two palindromic sequences. Figure 15 shows absorption spectra of U43 cells that contain these constructs grown under low and high O$_2$ conditions. The presence of B870 spectral peaks seen only under low O$_2$ conditions confirms that the O$_2$-regulated \textit{puf} promoter is present on the plasmid. Removal of the upstream palindrome in Δ42 yields a spectrum that is virtually identical to the parental Δ4, but removal of the other palindrome to yield Δ41 causes a reduction in the amount of absorbancy at 870 nm, under low O$_2$, presumably due to loss of B870 complexes. Although this reduction is noticeable it is not as dramatic as the loss
Fig. 14. A representation of the vector pTB999. The thin line represents vector DNA, the thicker line represents *R. capsulatus* chromosomal DNA, and the thickest lines represent the structural genes of the *puf* operon. The arrow shows the direction of transcription, and the letters outside the plasmid designate restriction endonuclease digestion sites as follows: H-*HindIII*, E-*EcoRI*. The letters inside the plasmid indicate *puf* operon genes, and the Tc\(^R\) designates tetracycline resistance.
Fig. 15. Absorption spectra of various strains of *R. capsulatus*. The abscissa represents wavelength of light, and the ordinate represents absorbancy units. The strains are designated as follows: Δ4, ΔRC6 (pTBΔ4); Δ41, ΔRC6 (pTBΔ41); Δ42, ΔRC6 (pTBΔ42). The representations of the constructs shown at the top of the page are from Table IV. The cells were grown under low or high O₂ conditions as described in Materials and Methods. The zoom scans enlarge the region of the spectrum where the B870 complexes absorb maximally.
of beta-galactosidase activity seen with the pXCA601 constructs in the ΔRC6 strain (see Table IV). The vector used in the study of the absorption spectra (pTB999) is similar to the pJAJ103::lac903 vector that showed the existence of read-through activity (see Appendix) when the initial lacZ fusion assays were performed. Thus the presence of B870 complexes in pTBA41, under low O₂ conditions, is most likely due to read-through from upstream on the pTB999 vector. Although a quantitative comparison can not be made between the pTB999 constructs in U43 and the pXCA601 constructs in ΔRC6 the results obtained do show that removal of the downstream palindrome results in a decrease in B870 gene expression, in keeping with the hypothesis that this palindrome is important for promoter activity.

2.3. Mapping of RNA transcripts from the puf operon

Cam Adams (in S.N. Cohen's lab at Stanford) performed RNA 5' end mapping experiments with the use of the 935 fragment (see Table IV, and Fig. 2b) as probe and obtained the predominant RNA 5' end (designated end 4 in Fig. 2) previously reported [8], plus three or more less abundant RNA 5' ends that mapped further upstream. The three most upstream 5' ends are designated 1, 2, and 3, and are shown in Figure 2b. An interesting point to note is that the most 5' of these ends maps just downstream from the sequences shown by lacZ fusion experiments to be involved in transcription of the puf operon. It had previously
been believed that these ends might have been artifacts in the S1 nuclease procedure because of their low intensity in comparison with the predominant RNA end 4. I performed a series of RNA end-mapping experiments to confirm these data and to better correlate these 5' ends with the $\text{lacZ}$ fusion results. The DNA probes used in these experiments are described in Figure 16.

Figure 17 shows an autoradiogram of an S1 nuclease mapping experiment in which pJAJ21 had been 5' end labelled at the EcoRI site (see Fig. 16a). This site was chosen in order to maximize hybridization to the less abundant transcripts that map upstream of the EcoRI site. The probe was hybridized to RNA from $R. \text{capsulatus}$ B10 cells (a wild type strain). Digestion with S1 nuclease yielded three RNA 5' ends that mapped to the positions of the three most upstream 5' ends shown in Figure 2 (ends 1, 2, and 3).

A primer extension experiment was performed to look for the presence of these 5' ends through the use of a different technique. Initial primer extension experiments yielded many more 5' ends than had been seen in the S1 protection experiments, so a primer titration was performed in order to minimize non-specific priming. Figure 18 shows the results of a primer titration experiment in which a primer, complementary to bases 932 to 953 of the sequence shown in Figure 3, was hybridized to $R. \text{capsulatus}$ B10 RNA. Hybridization of the highest amount of primer to the RNA yielded bands corresponding to a multiple number of 5' ends. Ends 1, 2, and 3 can be seen as well as end 4, and many other
Fig. 16. Construction of DNA probes used in RNA end-mapping experiments. The arrow inside the plasmid pJA]21 (shown in greater detail in Fig. 11, [26]) depicts the direction of transcription. The thin lines represent vector DNA, the thicker lines represent *R. capsulatus* chromosomal DNA, and the boxes represent the *pufQ* gene and the 5' end of the *pufB* gene. The letters designate restriction endonuclease digestion sites as follows: P-PstI, A-AccI, E-EcoRI, M-MnlI, B-BamHI, N-NaeI. Some of the DNA modifying enzymes are also abbreviated as follows: CIP, calf intestine alkaline phosphatase; PNK, T4 polynucleotide kinase. The asterisk designates the site of radioactive labelling. The use of each probe is described in the text and the results obtained shown in the following figures.
Taq I methylase, EcoRI, CIP, gamma 32 P-ATP and PNK, Bam HI, large fragment

Bam HI, CIP, gamma 32 P-ATP and Klenow, Bam HI, small fragment

c: Bam HI, CIP, gamma 32 P-ATP and PNK, Pst I, small fragment
d: same as c, except no labelling step

e: Bam HI, CIP, gamma 32 P-ATP and PNK, Pst I, small fragment

Pst I and Bam HI and Nae I, small fragments

g, h
Fig. 17. 5' end-mapping of B10 RNA with S1 nuclease. The 5' end-mapping of RNA transcripts from the wild-type strain B10 to the puf operon promoter region was as described in Materials and Methods. The DNA probe is described in Fig. 16a. Lanes A and B contain 40 ng of DNA probe hybridized to 10 μg of yeast tRNA, and lanes C to E contain 40 ng of DNA probe hybridized to 10 μg of B10 RNA. Varying amounts of S1 nuclease were added to the reactions as follows: lanes A and C, 500 units; lanes B and E, 1500 units; and lane D, 1000 units. Lane F contains M13mp11 ssDNA cut with HaeIII for molecular length markers. The arrows designate the bands that correspond to RNA ends 1, 2, and 3.
Fig. 18. Primer extension titration experiment. The primer extension experiments were performed as described in Materials and Methods. Lane A contains the results of an experiment in which 5 picomoles of primer were hybridized to 10 µg of yeast tRNA. Lanes B to E contain the results of primer hybridized to 10µg of B10 RNA in the following amounts: B, 5.0 picomoles; C, 1.0 picomoles; D, 0.2 picomoles; and E, 0.04 picomoles. Molecular length markers based on the digestion of M13mp11 ssDNA with HaeIII are shown in lane F. The arrows indicate the bands that correspond to ends 1, 2, 3, and 4.
ends immediately upstream and downstream of end 4 are seen as well. As the primer was titrated out the RNA bands disappeared until only end 4 remained. Since the three most upstream of the less abundant ends (ends 1, 2, and 3) showed up consistently in experiments based on different principles, it was more likely that they were genuine RNA ends and not artifacts of one type of experimental procedure.

The presence of some of the less abundant ends that mapped close to the predominant end were likely due to premature termination of the reverse transcriptase enzyme. This interpretation is based on unpublished results obtained by others in this laboratory, and the results seen in Figure 20 which will be discussed.

Since all of the sequences necessary and sufficient for O2-regulated initiation of transcription of the puf operon are present upstream of the AccI site it is possible that the most 5' of the RNA ends, end 1, is part of the primary transcript which is processed to yield the shorter transcripts. The processing previously found with other segments of the puf transcripts (shown in Fig. 2a [8]) strengthens this theory.

A 3' end mapping experiment was performed in an attempt to find 3' RNA ends that would correspond to the processed 5' ends seen, and confirm the processing theory. Figure 19 shows the result of an experiment in which pJAJ21 was 3'-end labelled at the AccI site (see Fig. 16b) and hybridized to R. capsulatus
B10 RNA. Many bands were seen in lanes that contained the results of experiments in which the probe was hybridized to yeast tRNA (Fig. 19, lanes A and B), or when the probe was hybridized to B10 RNA (Fig. 19, lanes C to E). No B10 RNA-specific bands were obtained. In many cases when mRNA processing occurs there is rapid 3'-5' degradation of the processed products [3]. If this is the case visualization of the 3' ends of the putative processing products may not be possible.

The differences in the level of expression of the lacZ gene with the various fragments cloned into the pXCA601 vector localized sequences important for transcription of the puf operon to a position just upstream of the AccI site shown in Figures 2 and 3 (see Table IV). It was important to correlate the beta-galactosidase activities from the deletion experiments with the results of RNA end-mapping experiments to better understand which sequences are involved in initiation of transcription. In addition, it was important to compare the pattern of plasmid-derived puf mRNA 5' ends with the ends of the chromosomal-derived transcripts. Therefore, the RNA from a number of the strains containing lacZ fusion plasmids with the inserts described in Table IV were purified and hybridized to complementary puf DNA probes. The results of 5' end mapping experiments with S1 nuclease are shown in Figure 20. The Δ4 construct (see lanes E and F), which contains all the sequences necessary for O2-regulated initiation of transcription, gave bands representing 5' ends of a length
Fig. 19. 3' end-mapping of B10 RNA with S1 nuclease. The 3' end-mapping of RNA transcripts from wild-type strain B10 to the puf operon promoter region was done as described in Materials and Methods. The DNA probe used is shown in Fig. 16b. Lanes A and B contain the results of an experiment in which 200 ng of DNA were hybridized to 10 μg of yeast tRNA, the results in lanes C to E are from experiments in which 200 ng of DNA were hybridized to 10 μg of B10 RNA. Varying amounts of S1 nuclease were added as follows: lanes A and C, 500 units; lanes B and E, 1500 units; and lane D, 1000 units. The positions of molecular length markers, based on M13mp11 ssDNA cut with HaeIII, are shown on the right.
comparable to RNA molecules with the three less abundant 5' ends 1, 2, and 3, as well as the predominant end 4 previously discussed (Fig. 2). Other 5' ends were also obtained that mapped close to end 4, but that did not show up in the photograph of the autoradiogram. A comparison of Figures 18 and 20 showed that the 5' ends obtained in lanes E and F of Figure 20 were the only ones that showed up in both techniques, thus these ends are likely to be genuine RNA 5' ends. Further examination of Figure 20 shows that when RNA from the strain containing the Δ14 construct (see lanes G and H) was hybridized to the same probe used for Δ4 (lanes E and F) no RNA 5' ends were detected. This corresponds to the loss of 95% of beta-galactosidase activity found with the analogous lacZ fusions (Table IV). A similar result was obtained with the Δ41 construct (lanes I and J) in which the downstream palindrome was removed (see Table IV). Hybridization of RNA from the strain containing the Δ44 construct (Table IV) with the same probe yielded a low level of the predominant 5' RNA end 4 (not visible in the photograph), and no other RNA protected bands were visible (see lanes K and L). This is consistent with the low levels of beta-galactosidase activity found with the corresponding lacZ fusions (Table IV).

Also included in Figure 20 is the result of an S1 nuclease mapping of RNA from the ΔAM construct (lanes O and P) with the PstI to BamHI fragment of pUCΔAM as probe (see Fig. 16e). One protected end that mapped to the position of the RNA end 1 was detected.
Fig. 20. 5' end-mapping of RNA from R. capsulatus ΔARC6 containing puf/lac fusion plasmids by S1 nuclease. The 5' end-mapping experiments were performed as described in Materials and Methods. The DNA used for probe in lanes C to L is described in Fig. 16c (25 ng was used per reaction), and the DNA probe used in lanes M to P is shown in Fig. 16e (5 ng was used per reaction). Lanes C, E, G, I, K, M, and O are the results of 500 units of S1 nuclease added to the reactions, and lanes D, F, H, J, L, N, and P are the results of 1000 units of S1 nuclease added to the reactions. Lanes C, D, M, and N contain probes hybridized to 10 μg of yeast tRNA. The other lanes contain probes hybridized to 10 μg of RNA from strains of R. capsulatus ΔARC6 containing the following plasmids: lanes E and F, pΔ4; lanes G and H, pΔ14; lanes I and J, pΔ41; lanes K and L, pΔ44, lanes O and P, pΔAM. The longest band in each lane is probe hybridized to its complementary strand, the arrows on the left indicate the positions of the three most upstream less abundant RNA transcripts (ends 1, 2, and 3 of Fig. 2) and the predominant transcript (end 4) seen in lanes E and F. Lanes A, B, and Q contain M13mp11 ssDNA cut with HaeIII for molecular length markers.
2.4. Mapping of a capped RNA 5' end

Although RNA 5' end 1 maps immediately downstream of the sequences shown to be necessary and sufficient for transcription of the \textit{puf} operon, it is conceivable that this end may be processed. To determine if this end is the first nucleotide of the primary transcript, capping experiments were performed using guanylyl transferase to end label di- and triphosphate ends of 5' RNA from \textit{R. capsulatus}. Initial experiments were performed with capped RNA from the wild type strain B10, with the 935 fragment (see Fig. 16d) as probe. The results of this experiment can be seen in Figure 21a. No protected capped end could be detected with the 935 probe (lane C), although a positive control with a \textit{puc} operon probe [61] detected capped RNA (Fig. 21a, lane A). It is possible that the capped end from the \textit{puf} operon was present in amounts too low to be detected by this technique.

In an attempt to bypass this problem the experiment was repeated using RNA from the \textit{ΔRC6} strain that contained the \textit{ΔAM} plasmid (see Table IV). The \textit{ΔAM} construct does not contain the putative processing sites (see Fig. 2), so RNA transcripts from the initiation site might be present in much greater amounts. This hypothesis was supported by the observation that the intensity of the band representing end 1 in the 5' end-mapping experiment with RNA from the \textit{ΔAM} construct (Fig. 20, lanes O and P) was similar to that of the predominant
Fig. 21. 5' end-mapping of capped RNA with nucleases. The capping experiments were performed as described in Materials and Methods. For Fig. 21a 10 µg of capped RNA from B10 was hybridized to either 25 ng of a puc promoter region probe [61] (lane A) or 50 ng of the PstI to BamHI fragment seen in Fig. 16d (lane C). Lane B contains the results of capped RNA hybridized to 50 ng of non-specific pUC13 DNA. Fig. 21b, lane A contained 10 µg of capped RNA from ΔRC6 (pΔAM) treated in the absence of DNA probe. Lane B contains the results of 10 µg of capped RNA from ΔRC6 (pΔAM) hybridized to 35 ng of the small PstI to BamHI fragment from pUCΔAM (see Fig. 16f). Fig. 21c contains the results of 10 µg of capped RNA from ΔRC6 (pΔAM) hybridized to either 40 ng of the small PstI to NaeI fragment seen in Fig. 16g (lane B) or 20 ng of the small NaeI to BamHI fragment in Fig. 16h (lane C). Lane A contains the results of capped RNA treated in the absence of DNA probe. Molecular length markers based on the digestion of M13mp11 ssDNA with HaeIII are shown on the right of Fig. 21a. The arrows indicate protected capped RNA bands discussed in the text.
band (representing end 4) seen in the Δ4 hybridization (Fig. 20, lanes E and F). It should be noted that two different probes were used in Figure 20, 34 femtomoles of one probe (see Fig. 16c), with a specific activity of $1.3 \times 10^4$ dpm/femtomole, were hybridized to 10 μg of Δ4 RNA to yield the results seen in lanes E and F, while 16 femtomoles of another probe (see Fig. 16e), with a specific activity of $2.0 \times 10^4$ dpm/femtomole, were hybridized to 10 μg of ΔAM RNA to give the results seen in lanes O and P.

Figure 21b shows the results of a capping experiment in which RNA from ΔRC6 (pΔAM) was hybridized to the ΔAM probe shown in Figure 16f. This experiment yielded a band representative of a capped, protected RNA that mapped to a position corresponding to RNA end 1. Another band also appeared further up on the gel that appeared to represent a protected RNA segment of about 290 nucleotides in length. This other band did not appear in the S1 nuclease mapping experiment (see Fig. 20, lanes O and P), so it was either an experimental artifact, or resulted from a transcript that initiated further upstream and terminated before the AccI site, or else was from the opposite strand. An experiment was performed in which the ΔAM probe was digested at the NaeI site seen in Figure 3 to yield two probe fragments. The putative longer capped RNA should have hybridized to one of these fragments. If this 290 nucleotide band was a true protected capped transcript from the same strand as the puf mRNA it would have then appeared to be 190 nucleotides long.
However, if it were from the other strand it would be 130 to 200 nucleotides in length. In any event the NaeI to BamHI fragment should still have protected RNA end 1. Figure 21c shows the results of this experiment. Lane B contained the results of the hybridization of ΔAM RNA to the fragment upstream of the NaeI site (Fig. 16g) and no clear band was visible. In contrast, hybridization of the DNA downstream of the NaeI site (Fig. 16h) to capped ΔAM RNA once again yielded a protected fragment (lane C) with 5' end 1 of the same length as previously seen in Figure 21b. The approximate position of this 5' end in the DNA sequence is shown in Figure 3 at nucleotide 228 (•).

Upon comparison of the length of protected DNA that appears in the S1 nuclease mapping experiment of ΔAM (Fig. 20, lanes O and P) with the length of protected RNA seen in the capping experiment shown in Figure 21, it was found that they appear to be 4-6 nucleotides different in length. This may be due to the differences in the mobility between a single-stranded DNA molecule and its complementary RNA, or the presence of the GMP molecule added in the capping experiment may cause alterations in mobility on the gel. The S1 protected fragment seen in Figure 20, lanes O and P, has a 5' end that maps approximately to nucleotide 233 (•) in Figure 3.

The presence of an RNA 5' end that is capable of being capped, and that maps immediately downstream of sequences shown to be necessary and sufficient for correct initiation of transcription of the *puf* operon, supports the
conclusion that the O$_2$-regulated promoter for the puf operon has been located.

3. Assessment of the role of the pufQ gene product

3.1. Absorption spectroscopy of cells containing wild-type or mutant pufQ genes. As described in the Introduction, an open reading frame was found upstream of the pufB gene that was capable of encoding a protein of 74 amino acids, and that was preceded by a Shine-Dalgarno like sequence [1,5]. Preliminary studies were performed on this open reading frame in order to deduce a possible function for the gene product.

One approach taken in the study of the role of pufQ was to compare the light absorption spectra of strains of R. capsulatus that were deleted for pufQ to the spectra of strains that contained pufQ.

The spectrum shown in Figure 22A is of the wild-type strain B10. The B800-850 light-harvesting absorption peaks are easily seen, whereas the less abundant B870 and reaction center peaks are hidden by the absorption at 850 nm.

The result obtained with the ΔRC6 strain, which has a chromosomal deletion that begins at the SalI site within pufQ (see Fig. 2) and extends to an XhoII site beyond the pufX gene, is shown in Figure 22B. It can be seen that although the genes for the B800-850 peptides (found in the puc operon) are still present in this strain, very little absorption due to B800-850 complexes was seen.
Fig. 22. Absorption spectra of *R. capsulatus* strains B10 (A), ΔRC6 (B), and U43 (C). The abscissa represents wavelength and ranges from 350 nm to 900 nm. The ordinate represents absorbance units and ranges from 0.0 to 0.6.
The second strain used was U43 (described in section 2.2), a strain from which the \textit{puf} operon has been deleted, and which is mutated in the \textit{puc} operon. Thus U43 contains no B800-850, B870, or reaction center peptides, as well as no \textit{pufQ} gene product. The absorption spectrum for this strain is shown in Figure 22C.

Figure 23A shows the absorption spectrum of a strain of \textit{\textDelta RC6} that contains the plasmid pJAJ935. The plasmid pJAJ935 has cloned into it the \textit{XhoII-SphI} region of the \textit{puf} operon shown in Figure 2b, thus it contains a complete copy of the \textit{pufQ} gene. The absorption scan shows a dramatic increase in absorbancy at 800 and 850 nm in comparison with \textit{\textDelta RC6} (Fig. 22B). To ensure that the formation of the B800-850 complexes was due to the presence of the \textit{pufQ} gene, and not the result of the presence of some other component of the insert or vector, the \textit{pufQ} gene was deleted from the plasmid pJAJ935 by removal of the \textit{NcoI-EcoRI} segment shown in Figure 11, and the resultant plasmid (pJAJ\textDelta Q) was tested in cells of \textit{\textDelta RC6}. As shown in Figure 23B, the removal of \textit{pufQ} from pJAJ935 resulted in an absorption scan similar to that of \textit{\textDelta RC6}. Thus it appeared that the \textit{pufQ} gene product was necessary for the formation of B800-850 complexes.

To observe the effects of \textit{pufQ} on B870 complex formation a strain of U43 that contained the plasmid pTB999 (shown in Fig. 14) was used. The use of
Fig. 23. Absorption spectra of various strains of *R. capsulatus*. The abscissa represents wavelength and ranges from 350 nm to 900 nm. The ordinate represents absorbancy units and ranges from 0.0 to 0.6.
U43 allows for B870 complexes to be seen as they are no longer masked by the B800-850 complexes (which are not formed in U43). Figure 23C shows an absorption scan of U43 (pTB999) in which there is absorbance at 870 nm due to the presence of B870 complexes. The pufQ gene was deleted in pTB999 as described in Materials and Methods and, as shown in Figure 23D, the removal of pufQ results in the loss of B870 complex absorption.

3.2. The effects of pufQ on transcription and translation of puf or puc mRNA. The pufQ gene product could be involved in light-harvesting complex formation in several ways. For example, it may be needed for initiation of puf or puc operon transcription or mRNA translation, or it could be necessary for posttranslational assembly of mature complexes.

The results that I have obtained while studying the promoter of the puf operon show that pufQ is not needed for transcription of the puf operon. This was shown by the high lacZ activities obtained with the ΔAM construct (see Table IV) in which pufQ has been deleted, and by the amount of RNA that was obtained from ARC6 containing pΔAM (shown in Fig. 20, lanes O and P). Both the lacZ activities and the quantity of RNA obtained are similar to those of the Δ4 construct, which contains pufQ (see Table IV and Fig. 20, lanes E and F).

The effects of pufQ on transcription of puc genes were tested by Anthony Zucconi in our lab by measurements of beta-galactosidase activities in
cells of B10 or ΔRC6 that contained plasmids in which the \textit{puc} promoter was fused to the \textit{E. coli lacZ} gene. One plasmid construct (pAZIII) contained the entire \textit{lacZ} gene and translation initiation was controlled by the native \textit{lacZ} Shine-Dalgarno and associated translational regulatory sequences (operon fusion) [62]. The specific activities of beta-galactosidase extracts of B10 and ΔRC6 cells were very similar [62]. This indicates that deletion of \textit{pufQ} has no affect on initiation of transcription at the \textit{puc} promoter.

The second plasmid used by Zucconi (pAZV) contained the same segment of the \textit{puc} operon, except that there was an in-frame fusion between the twenty-fifth codon of the \textit{pucB} gene and the eighth codon of \textit{lacZ} (gene fusion) [62]. The activities obtained with this plasmid in cells of B10 or ΔRC6 showed that there are no significant differences regardless of the presence or absence of \textit{pufQ} [62]. Therefore, \textit{pufQ} does not seem to affect the frequency of translation initiation at \textit{puc} translational regulatory sequences.

In order to evaluate the effect of \textit{pufQ} deletion on \textit{puc} mRNA levels, I performed S1 nuclease protection measurements of \textit{puc} mRNA from B10 or ΔRC6 cells. The results are shown in Figure 24. The hybridization was performed with the probe concentration in excess of that of the \textit{puc} mRNA [61], and the levels of RNA detected from B10 and ΔRC6 were similar. This result confirms that \textit{puc} operon transcription occurs normally in strain ΔRC6, and also shows that the stability of \textit{puc} mRNA is equivalent in the two strains.
Fig. 24. S1 nuclease protection measurement of puc mRNA from B10 and ΔRC6. The S1 experiments were performed as described in Materials and Methods. Lane A, 500 ng of puc probe hybridized to 10 μg yeast tRNA; lane B, 500 ng of puc probe hybridized to 10 μg B10 RNA; lane C, 500 ng puc probe hybridized to 10 μg ΔRC6 RNA. Molecular length markers based on M13mp11 ssDNA digested with HaeIII are shown on the right. The arrow indicates the bands that correspond to puc mRNA.
DISCUSSION

In the first part of this thesis I described the purification of RNA polymerase from *R. capsulatus* cells grown either chemoheterotrophically or photoheterotrophically. The subunit composition of the enzyme was similar to that of other eubacterial RNA polymerases [10,50]. The putative beta and beta-prime subunits were very close in molecular weight to those of *E. coli*, whereas the putative alpha subunit was larger than the 40,000 Dalton *E. coli* alpha subunit. Although the polymerase that was obtained from the heparin Sepharose column had several components that could have been the sigma subunit(s), the two further purifications performed removed most of these additional protein bands and prompted my provisional designation of sigma. This component appeared to be the same for RNA polymerase purified from chemoheterotrophically and photoheterotrophically grown *R. capsulatus*.

The RNA polymerase purified from the heparin Sepharose column was the preparation used for *in vitro* transcription. Radioactive RNA transcripts made from linearized pUC13 or pJAJ21 templates showed that the preparation of RNA polymerase can initiate transcription efficiently at some promoters in the absence of accessory factors. Although the *R. capsulatus* region present in pJAJ21
promotes regulated transcription of adjacent genes in vivo [26], no transcripts other than those found with the linearized vector (pUC13) alone were detected in vitro. When the cloned R. capsulatus DNA fragment was used alone as template a much lower level of transcription was obtained than when pUC13 DNA was present, and this activity was much more sensitive to heparin inhibition. It was also observed that a quasi-random size distribution of RNA transcripts was obtained. This result is similar to results obtained with core RNA polymerase preparations [30,46], and I conclude that some factor or condition necessary for specific transcription from the puf promoter is absent in these assays. For example, transcription of the puf promoter may be regulated by the extent of supercoiling of the DNA, so that the use of linear templates may have prevented the initiation of transcription. The control of transcription by the extent of supercoiling has been described for other promoters [9,53], and has recently been implicated in the oxygen regulation of the photosynthesis genes of R. capsulatus [59]. Alternatively, it may be that a protein factor, lost during RNA polymerase purification, is necessary for specific initiation of transcription from this promoter. Although tests of cruder extracts showed no significant decrease in the number of run-off transcripts, the salts or polyethylene-glycol present at earlier stages of purification could have inhibited such a factor. Another possibility is that some or all of the "non-specific" transcripts arose from premature termination after initiation at a specific site. It also cannot be ruled
out that the genuine sigma factor was lost during the purification.

The purification and preliminary characterization of *R. capsulatus* RNA polymerase has shown that the enzyme is similar to *E. coli* RNA polymerase in basic subunit structure. Both enzymes were capable of specific initiation of run-off transcription with linearized pUC13 DNA as template, but neither enzyme gave specific transcripts in a run-off assay with the *puf* DNA fragment (*E. coli* data not shown). These results, along with the observation that the *puf* promoter does not function in *E. coli* [26] are consistent with the possibility that an accessory factor which facilitates transcription of photosynthesis genes is present in cells of *R. capsulatus*.

The second part of my thesis was based on localization of the *puf* promoter through the use of *in vitro* mutagenesis and evaluation of the mutagenesis effects on the functional activity of the promoter as measured with fusions of the *puf* promoter region to the *E. coli lacZ* gene. This approach was complemented by the use of RNA end-mapping and capping techniques. The existence of a predominant RNA transcript with a 5' end (end 4) that mapped just downstream of the *EcoRI* site seen in Figure 2 led to an initial assumption that the promoter lies just upstream of the *EcoRI* site. However, the results of large scale deletions of the *puf* promoter region showed that the promoter lies upstream of the *AccI* site, which is more than 540 base pairs upstream of the *EcoRI* site (see Table IV and Fig. 2).
Exonuclease deletion experiments localized the *puf* operon promoter between the 5' end of the Δ4 deletion shown in Table IV and the AccI site. RNA end-mapping experiments revealed the existence of less abundant RNA transcripts whose 5' ends mapped further upstream from the predominant RNA 5' end (see Fig. 2; ends 1, 2, and 3). The most upstream of these transcripts maps just before the AccI site.

The promoter was located more precisely by oligonucleotide-directed mutagenesis. An inverted repeat between nucleotides 176-205 (see Fig. 3) was shown to be necessary for transcription initiation (see Δ41, Table IV). Just downstream of this inverted repeat is a sequence of DNA similar to sequences of the *puc* promoter region implicated in transcription initiation [61]. Since a change of only two nucleotides in the *puf* sequence decreased activity drastically (Δ44, Table IV), I propose that the AT-rich region around these two nucleotides, 5'-TTACAT-3' (see Fig. 3), is the "-10" region of the promoter. This would place the "-35" region within the inverted repeat that was deleted in the Δ41 construct. The finding that the Δ41 and Δ14 deletions eliminated promoter activity (see Table IV and Fig. 3) strengthens this suggestion.

It may be that the cis-active component involved in O2 regulation of promoter activity also lies within the "-35" inverted repeat. Upon comparison of the ratio of low to high O2 activities shown in Table IV it can be seen that the Δ41 construct (in which the "-35" inverted repeat has been deleted) has a lower ratio
than the Δ44 construct, in which the inverted repeat and putative "-35" are present but the "-10" has been mutated. The Δ44 construct has three-fold higher activity under low O₂ conditions than the Δ41 construct. Although the values are low, due to removal of essential promoter sequences, this difference in ratios may indicate that the inverted repeat is involved in cis-active regulation by O₂.

A promoter, termed P1, was located recently in the region in which I have located the puf promoter [5]. A second, constitutive promoter, P2, with 3% of the activity of P1 was located about 190 base pairs downstream of P1 [5]. P2 lies at about position 420 in Figure 3, which is between the 5' ends of the Δ24 and 932 constructs shown in Table IV. The data I obtained, shown in Table IV, showed a decrease in promoter activity of about 50% when that region was deleted. I believe that the residual activity seen in the Δ24 and 932 constructs is due to non-specific initiation of transcription because a further deletion to the MnII site (see ΔMSP) resulted in a further decrease in activity of approximately 50%, supporting the theory of non-specific initiation over the existence of a weak constitutive promoter. Although there is evidence for multiple promoters in R. shaeroides (which is closely related to R. capsulatus), one upstream of pufQ [27] and two upstream of pufB [57], I interpret my data as suggesting that these two species use different methods of expression of the puf operon.

It was also suggested by Bauer et al. that the sequence at their P1 had homology with the ntrA sigma factor consensus recognition sequence and that
the puf promoter may be transcribed by an RNA polymerase that contains an
ntrA-like sigma factor [5]. This seems unlikely as the ntrA sigma subunit exists in
E. coli and it has been shown that puf promoter does not function in E. coli [26].
However, the lack of transcription may be due to incompatibility of other
components of the E. coli RNA polymerase.

The wild-type strain B10 was used as a source of RNA for end-mapping
experiments that were performed in order to test the validity of the gene fusion
data. The existence of the less abundant RNA transcripts was confirmed by 5'
end-mapping with S1 nuclease (see Fig. 17) and primer extension (Fig. 18). The
use of these two different techniques to show the existence of RNA 5' end 1
supports the conclusion that transcription initiates upstream of the AccI site
shown in Figure 2. The additional 5' ends detected by both procedures (2, 3, and
4) prompted the hypothesis that the most 5' of the less abundant ends (end 1, Fig.
2) maps to the site of transcription initiation, and the presence of the other 5'
ends is due to processing of the primary transcript. A 3' end-mapping experiment
was attempted to determine if stable 3' ends could be detected as a product of this
putative processing, but no 3' ends could be detected. The existence of active 3' to
5' RNases in the cell may make the detection of these processed 3' ends difficult.

RNA preparations from the puf deletion strain ΔRC6 that contained a
number of the plasmids with mutant puf 5' sequences (shown in Table IV) were
studied to see if the effects on lacZ activity presented in Table IV corresponded to
the levels of RNA obtained from the cells. The results displayed in Figure 20 showed that RNA levels matched the levels of beta-galactosidase found.

These RNA end-mapping experiments also gave support to the processing theory. The ΔAM construct has all of the putative processing sites removed (see Fig. 2) so it was hypothesized that RNA transcripts that initiated from the site just upstream of the AccI site might not be processed in this strain and would be present in the cells in an abundance comparable to the predominant transcript described in Figure 2 (end 4), and seen in lanes E and F of Figure 20. Comparison of the intensity of the band seen in lanes O and P of Figure 20 with the predominant band in lanes E and F shows that the intensity is similar. Although the amount of probe (Fig. 16e) added to the ΔAM RNA to yield the results seen in lanes O and P was only about 50% of the amount of the other probe (Fig. 16c) added to the same amount of Δ4 RNA to yield the results seen in lanes E and F, the former probe had about twice the specific activity. If the probes were in excess, and processing was not occurring then the band seen in lanes O and P would appear to be about twice as intense as the band representing end 1 in lanes E and F. The band in lanes O and P is many times more intense than that, and is similar in intensity to the band representing end 4 in lanes E and F. Therefore it was concluded that ends 2, 3, and 4 arise by processing of the primary transcript (which initiates with end 1), and that the low abundance of this segment of the puf transcript is due to this processing
The presence of an RNA 5' end that could be capped, and which mapped immediately downstream of the sequences shown to be essential for promoter activity confirmed that the \textit{puf} promoter had been located. Although this capped end could only be found in the ΔAM construct, the extreme instability of the wild-type primary transcript could account for the inability to detect it. Shown in Figure 25 is a model for the initiation of \textit{puf} mRNA transcription and subsequent processing to yield the multiple transcripts that have been found. The B and A gene products are present in the cell in much greater amounts than the L and M gene products, and differential mRNA stability has been suggested as one method of control of production of these components of the photosynthetic apparatus [8]. Along the same lines, because of the low amount of the \textit{pufQ} segment of the \textit{puf} transcript it may be that the \textit{pufQ} gene product is required in much lower amounts than the other \textit{puf} gene products.

The presence of an open reading frame located between the sequences found to be involved in promoter activity and what was previously believed to be the first gene of the operon, \textit{pufB}, helped to explain the distance between the two regions. The predicted amino acid sequence of the product of the new gene was similar to a segment of both the L and M gene products of \textit{puf}, specifically the region where the peptides had been shown to bind to bacteriochlorophyll [1]. The low level of \textit{pufQ} segments of \textit{puf} mRNA suggests that the \textit{pufQ} gene product is present in relatively low amounts in the cell and has a catalytic rather
Fig. 25. Processing of puf mRNA. The top line represents *R. capsulatus* chromosomal DNA, the boxes represent puf operon genes. The horizontal arrows designate puf RNA transcripts, with the primary transcript indicated with an asterisk. The thicker the arrows the more abundant the RNA molecules represented.
than a structural role (see Fig. 2). I have shown that $pufQ$ is necessary for formation of mature photosynthetic pigment/protein complexes, but this role of $pufQ$ is not at the level of transcription or translation of the peptide genes (Figs. 23 and 24). Whether Q is involved in assembly of complexes or in bacteriochlorophyll synthesis is as yet unknown. Deletion of $pufQ$ has no effect on transcription of the $bchC$ gene, which encodes a bchl biosynthetic enzyme (C. Keen, personal communication). It is known that synthesis of both peptides and bacteriochlorophyll is necessary for assembly of the complexes [28]. No peptides are found in $bch^{-}$ mutants [17] and no free bchl is seen in strains that contain mutations in photosynthetic complex peptide genes (as in U43) [55]. Yet, when all of the bchl biosynthetic enzyme genes and B800-850 peptide genes are present in a strain that is $pufQ^{+}$ (as in ΔRC6) no complexes can be seen. Thus it appears that $pufQ$ may be necessary for assembly of mature complexes, or bchl biosynthesis [5].

My localization of the $puf$ promoter and determination of the sequences involved in initiation of transcription will aid in localization of other photosynthesis promoters of $R. capsulatus$, and will enable further studies to be performed on this and other promoters to achieve a better understanding of gene expression in general.
REFERENCES


49. Weaver, P. F., J. D. Wall, and H. Gest. 1975. Characterization of


APPENDIX

The vector pJAJ103::lac903 [62] is a derivative of pJAJ103 [26], which in turn is a derivative of the broad host range vector pRK404 [18] into which was placed a fragment that contained the lacZ gene [Casadaban, etal. 1980. J. Bacteriol. 143: 971-980]. No attempt was made to place transcriptional or translational terminators into this vector. The first six constructs seen in Table V show a gradual decrease in activity as 5' sequences are removed. This trend continued as more and more sequences were removed between the 5' ends of the 932 construct and the ΔMSP construct (see ESP and A2). Removal of promoter sequences should result in a dramatic loss of activity rather than a gradual decrease so the results of these experiments led to the belief that read-through transcription from upstream sequences on the vector was contributing to the activity seen.

The vector pTB999 (section 2.2 of the results) is a derivative of pRCVI [11] which was derived from pTJS133 [39]. No attempt was made to place terminators into this vector either. Since both pRK404 and pTJS133 are derived from pRK2, and share sequences upstream of the insertion site for fragments, this commonality in the two vectors led me to assume that read-through was a likely possibility in the pTB999 vector as well. This would explain the presence of B870 complexes in U43 (pTBA41).
Table V. Assays of beta-galactosidase activities of cells containing *puf* 5' region operon fusions to the *lacZ* gene of pJAJ103::lac903

1 Representation of the constructs used for *puf* promoter mapping. Restriction sites are indicated by vertical lines and are labelled in the top construct as follows: X-XhoII, A-AccI, M-MnlII. The structural genes are designated by thick lines and are labelled in the 935 representation. The dashed line represents the sequences that were removed in the ΔAM construct. The hillocks represent the palindromic sequences that were studied and the x's designate sites of oligonucleotide-directed deletions or mutations.

2 Activities are expressed as nmoles ONPG/min/mg protein. The values in brackets are standard deviations of 3 to 7 assays that were performed on each construct.

3 The ratio of activities obtained under low O₂ versus high O₂.
<table>
<thead>
<tr>
<th>CONSTRUCT</th>
<th>REPRESENTATION</th>
<th>LOW O₂</th>
<th>HIGH O₂</th>
<th>L / H</th>
</tr>
</thead>
<tbody>
<tr>
<td>935</td>
<td>X Q M B'</td>
<td>742 (206)</td>
<td>164 (40)</td>
<td>4.5</td>
</tr>
<tr>
<td>ΔMSP</td>
<td></td>
<td>70 (18)</td>
<td>43 (20)</td>
<td>1.6</td>
</tr>
<tr>
<td>Δ4</td>
<td>Ω Ω</td>
<td>531 (63)</td>
<td>163 (43)</td>
<td>3.3</td>
</tr>
<tr>
<td>Δ14</td>
<td></td>
<td>409 (53)</td>
<td>272 (50)</td>
<td>1.5</td>
</tr>
<tr>
<td>Δ24</td>
<td></td>
<td>384 (58)</td>
<td>311 (50)</td>
<td>1.2</td>
</tr>
<tr>
<td>932</td>
<td></td>
<td>327 (43)</td>
<td>249 (28)</td>
<td>1.3</td>
</tr>
<tr>
<td>ESP</td>
<td></td>
<td>219 (22)</td>
<td>158 (16)</td>
<td>1.4</td>
</tr>
<tr>
<td>Δ2</td>
<td></td>
<td>149 (2)</td>
<td>110 (7)</td>
<td>1.4</td>
</tr>
<tr>
<td>ΔAM</td>
<td>Ω Ω</td>
<td>611 (84)</td>
<td>272 (25)</td>
<td>2.2</td>
</tr>
<tr>
<td>Δ41</td>
<td>Ω X</td>
<td>144 (10)</td>
<td>108 (6)</td>
<td>1.3</td>
</tr>
<tr>
<td>Δ42</td>
<td>Ω Ω</td>
<td>416 (13)</td>
<td>142 (11)</td>
<td>2.9</td>
</tr>
<tr>
<td>Δ44</td>
<td>Ω Ω</td>
<td>128 (5)</td>
<td>107 (6)</td>
<td>1.2</td>
</tr>
</tbody>
</table>
PUBLICATIONS

Forrest, M.E., and J. T. Beatty
Purification of Rhodobacter capsulatus RNA Polymerase and Its use for in vitro Transcription

Adams, C. W., M. E. Forrest, S. N. Cohen, and J. T. Beatty
Transcriptional Control of the Rhodobacter capsulatus puf Operon: A Structural and Functional Analysis.
Submitted to J. Bacteriol.

Forrest, M. E., A. J. Zucconi, and J. T. Beatty
The pufQ Gene Product of Rhodobacter capsulatus is Essential for Formation of Mature B800-850 Light-harvesting Complexes.
Submitted to FEMS Microbiology Letters

GRADUATE AWARDS

1983-1987 The University of British Columbia Post-graduate Fellowship
1985-1987 Natural Sciences and Engineering Research Council Pre-doctoral Fellowship
1985-1987 Izaak Killam Pre-doctoral Research Fellowship