Analysis of the S-layer Transporter Mechanism and Smooth Lipopolysaccharide Synthesis in *Caulobacter crescentus*

by

Peter Alan Awram

B.Sc., The University of British Columbia, 1992

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in
THE FACULTY OF GRADUATE STUDIES
(Department of Microbiology and Immunology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
November 1999

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
Vancouver, Canada

Date Dec 1/99

DE-6 (2/88)

Abstract

C. crescentus is a Gram-negative bacterium that possesses an hexagonal array called the S-layer that covers the entire outer surface of the bacterium. This array is composed of an estimated 60 000 copies of the 98 kDa protein RsaA. RsaA secretion is directed by a C-terminal secretion signal located in the last 82 amino acids of the protein. Once RsaA is secreted from the cell, it assembles into the S-layer and attaches to the outer membrane via a specific species of smooth lipopolysaccharide (S-LPS). The mechanisms required for the secretion of RsaA and the synthesis of the S-LPS were examined in this thesis.

Tn5 mutagenesis of wildtype *C. crescentus* demonstrated the presence of two genes, *rsaD* and *rsaE*, 3' of the *rsaA* gene that were required for transport of RsaA. These genes were isolated and are capable of complementing the Tn5 mutations 3' of RsaA *in trans*. The resulting proteins of *rsaD* and *rsaE* belong to the type I secretion family that uses three components: an ATP Binding Cassette-transporter (RsaD), a Membrane Fusion Protein (RsaE) and an outer membrane protein (OMP), to secrete proteins through both membranes of Gram-negative bacteria. The OMP, RsaF, of the Rsa system was found by screening the partial *Caulobacter* genome sequence for sequence identity to other type I OMPs. The gene for RsaF is found 5 kb 3' of *rsaE*. Deletion of the N-terminus or C-terminus of RsaF prevents the Rsa secretion mechanism from functioning.

The secretion of the S-layer subunits in a number of other *Caulobacter* species was also examined. A partial ORF from FWC27 with 44.6% identity to RsaA was isolated. In addition, the ABC-transporter components from FWC6, FWC8 and FWC39 were isolated. These components were >95% identical to

RsaD. These results were used to explore the evolutionary relationships between the different *Caulobacter* species.

Eighteen Tn5 mutations resulting in the inability of the S-layer to attach to the surface of the bacterium were also isolated. Southern blot analysis demonstrated that twelve of these insertions were linked to the Rsa transporters. The Tn5 insertion points were isolated and sequenced allowing identification of several putative genes involved in S-LPS synthesis from the *Caulobacter* genome sequence. A total of twelve open reading frames (ORFs) were found by Tn5 mapping and two more were found 3' of *rsaE*. Six of these putative genes may code for proteins involved in the synthesis of sugar residues including five that make perosamine. Five of the genes appear to be glycosyltransferases involved in forming the linkages between sugar residues in the O-antigen. One of the genes appears to be a repressor, while the remaining genes are unidentified. These data suggest that the major component of the O-antigen is perosamine and that a number of different linkages are made between the perosamine residues.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

ABC	ATP–Binding Cassette
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
C-terminus	carboxy terminus
DNA	deoxyribonucleic acid
EDTA	ethylene diaminetetra-acetic acid
EGTA	ethylene glycol-bis(β-aminoethyl Ether) NNN'N' tetraacetic acid
G+C	guanosine and cytosine content of DNA
FWC	freshwater *Caulobacter*
HCl	hydrochloric acid
KDO	ketodeoxy octulosonic acid
kDa	kilodalton
Km	kanamycin
LPS	lipopolysaccharide
min	minute
MFP	membrane fusion protein
mg	milligram
ml	millilitre
µl	microlitre
µg	microgram
NaCl	sodium chloride
NeuNAc	N-acetyl neuraminic acid (sialic acid)
NAD	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
N-terminus	amino terminus
NTG	1-methyl-3-nitro-1-nitrosoguanidine
O-antigen	antigenic determinant found on the outside of cell consisting of repeating units of oligosaccharides
ORF	open reading frame
OMP	outer membrane protein
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PYE	peptone yeast extract
RNA	ribonucleic acid
S-layer	surface layer
S-LPS	smooth lipopolysaccharide of *C. crescentus*
SDS	sodium dodecyl sulphate
Sm	streptomycin
Tc	tetracycline
T_m	Melting temperature of two strands of DNA
TIGR	The Institute for Genome Research
Tris	Tris (hydroxymethyl) methylamine
UV	ultra violet light

List of Species Abbreviations

B. pertussis	*Bordetella pertussis*
B. melitensis	*Brucella melitensis*
C. crescentus	*Caulobacter crescentus*
C. fetus	*Campylobacter fetus*
C. jejuni	*Campylobacter jejuni*
E. chrysanthemi	*Erwinia chrysanthemi*
E. coli	*Escherichia coli*
P. aeruginosa	*Pseudomonas aeruginosa*
S. enterica	*Salmonella enterica*
S. marcescens	*Serratia marcescens*
V. cholerae	*Vibrio cholerae*
R. meliloti	*Rhizobium meliloti*
R. leguminosarum	*Rhizobium leguminosarum*

Acknowledgements

I would like to acknowledge myself for persevering throughout this process. I would further like to thank my wonderful girlfriend Bianca Kuipers for being supportive during this time. I would like to thank John Nomellini and Stephen Walker for helpful suggestions and thoughtful insights and the occasional gel along the way. I would also like to thank all my 'partners in pain' that started with me.

Chapter 1
Introduction

This thesis focuses on the secretion and attachment of the S-layer of *Caulobacter crescentus*. S-layers are not well understood and have not been studied extensively even though they are found on a wide range of prokaryotes (Messner and Sleytr, 1992; Sleytr *et al.*, 1993; Sleytr and Sara, 1997). Consequently, there is a need for basic research to describe these structures. Despite this lack of study, some research has been done on the commercial aspects of S-layers (Sleytr *et al.*, 1997a). The research presented here is applicable to both of these areas. It is of general interest to know the methods of secretion and attachment of the S-layer and this information can also be applied to the commercial aspects of S-layers.

Evidence is presented that the S-layer subunit of *C. crescentus* is secreted by a type I secretion mechanism and that the S-layer subunits of a number of other *Caulobacter* species are probably secreted by an almost identical type I mechanism. Also presented are several putative proteins involved in the synthesis of the O-antigen that support the predicted composition of the O-antigen as being a polymer of a 4,6-dideoxy-4-amino-hexose with complex linkages (Walker *et al*, 1994; Smit unpublished). Furthermore these data suggest that the 4,6-dideoxy-4-amino-hexose is perosamine and that a number of glycosyltransferases provide complex linkages between the perosamine residues.

The S-layer of *C. crescentus* can be used as a biotechnology vehicle. The S-layer is a 2-dimensional array made from approximately 60 000 copies of the protein, RsaA (Smit *et al.*, 1981). This layer covers the entire outer surface of the bacterium and makes up about 10% of the cell's protein. Therefore, RsaA must be secreted, passing through both membranes, from the Gram-negative cell. An uncleaved C-terminal secretion signal directs this secretion of RsaA (Bingle *et al.*, 1999; Bingle *et al.*, 1996; Bingle *et al.*, 1997b; Bingle and Smit, 1994). Once secreted, the S-layer is attached to the outer membrane via the smooth

lipopolysaccharide (S-LPS) (Walker *et al.*, 1994). If the S-LPS is disrupted or absent the S-layer detaches from the membrane and aggregates into particles that are up to 90% pure RsaA making it easy to collect large amounts of relatively pure protein (Fig. 1-1). It has been found that the N-terminus of RsaA contains the attachment domain and a C-terminus Ca^{2+} binding domain is responsible for aggregation of the protein (Bingle *et al.*, 1997b).

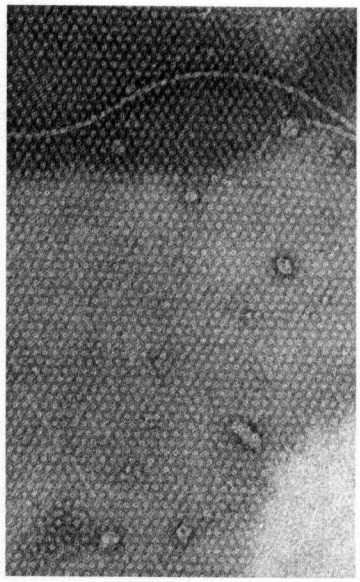

Figure 1-1. Shed S-layer from *C. crescentus*. EM photograph of S-layer shed from a strain with defective S-LPS. (Photo courtesy John Smit)

To produce recombinant proteins it is desirable to produce large quantities that are easily isolated from the rest of the cellular protein. The properties of the *C. crescentus* S-layer and secretion apparatus allow this. The C-terminal secretion signal and Ca^{2+} binding domain can be fused to a desired protein and recombinant proteins can then be secreted from *C. crescentus* by the RsaA secretion signal. The proteins aggregate together in the medium where they can be filtered away from the cells. This process has been shown to be viable and recombinant proteins have been expressed and purified from *C. crescentus* (Bingle *et al.*, 1997a).

S-layers also have other uses such as the expression of epitopes in S-layers to be used for recombinant vaccines. Another aspect that is being examined is to use the regular arrays formed by the S-layer as templates for the deposition of metal or silicon atoms to allow creation of circuitry finer than is allowed by current integrated circuit etching technology. It would also be possible to use the arrays as surface supports to which biologically active molecules could be attached (Sara and Sleytr, 1996a; Sara and Sleytr, 1996b; Sleytr *et al.*, 1997a; Sleytr *et al.*, 1997b; Sleytr and Sara, 1997). Obviously, all these uses could be applied to the S-layer of *C. crescentus.*

To increase the utility of *C. crescentus* S-layers for such applications it is vital to understand how the RsaA protein is secreted and attached to the surface. For example, it is necessary to understand the conformation of the protein when it is

passing through the secretion apparatus. This will determine what kind of foreign proteins or epitopes can be secreted and are capable of forming aggregates using the RsaA secretion pathway. To answer some of these questions this thesis examines the RsaA secretion and S-LPS synthesis pathways.

***C. crescentus* is a Gram-negative, motile eubacterium found in soil and aquatic environments including drinking water.** The non-pathogenic bacterium derives its name from the crescent shape of the cells. *C. crescentus* undergoes a dimorphic developmental life cycle (for reviews see Brun *et al.*, 1994; Gober and Marques, 1995; Poindexter, 1981; Shapiro, 1976; Shapiro and Losick, 1997) during which it switches between a motile (swarmer) phase and a sessile stalked phase (Fig 1-2). In both phases the bacterium is completely covered by the S-layer (Smit *et al.*, 1981). In the swarmer phase the cell expresses a single flagellum, pili and a holdfast (an adhesin) at one pole. When the cell differentiates into the stalked form, it loses the flagellum and a stalk (containing no cytoplasm) grows out from the cell envelope keeping the holdfast on its tip. Stalked cells divide and produce a swarmer cell with the flagellum being created at the pole furthest from the stalked cell. Most of the current research on *C. crescentus* focuses on the developmental process resulting in these two different forms and the development of the flagellum (Brun *et al.*, 1994; Roberts *et al.*, 1996; Shapiro and Losick, 1997).

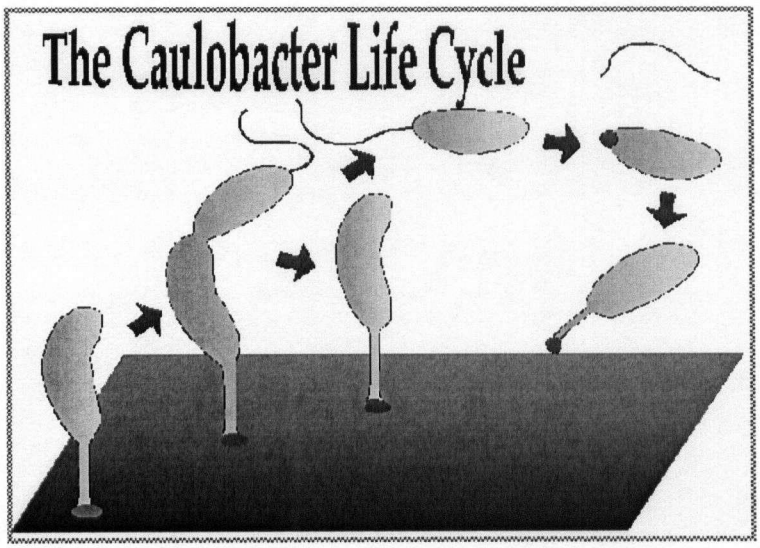

Figure 1-2. Developmental cycle of *C. crescentus*
Sessile cells attached to the surface via the holdfast bud off swarmer cells which move to a new location where they lose their flagellum and grow a stalk to attach to the surface again. (Figure courtesy Ian Bosdet.)

S-layers are two-dimensional arrays that cover the outside surface of many prokaryotes. *C. crescentus* is one of many species of bacteria covered with a crystalline protein surface layer (S-layer) (Boot and Pouwels, 1996; Sleytr and Messner, 1983; Sleytr and Sara, 1997; Smit *et al.*, 1981). Thousands of copies of nearly always a single protein or glycoprotein self-assemble into a crystalline-like lattice (Sleytr and Messner, 1983). The S-layers described so far have subunits ranging in size from 30 to 220 kDa (Messner and Sleytr, 1992). Although a large number of bacteria have been found to have S-layers, enteric bacteria, the most studied, lack them and consequently have not been studied much (Hovmoller *et al.*, 1988; Sleytr and Messner, 1988). For reviews on S-layers see Beveridge *et al.*, 1997; Sleytr, 1992; Sleytr and Messner, 1983.

S-layers typically make up 10% of the protein in a cell and thus represent a large energy expenditure by the cell (Sleytr and Messner, 1983). Many bacteria have been found to lose their S-layers when there is no environmental pressure for maintenance, such as during sub-culturing in the laboratory, showing that S-layers are not essential for growth (Blaser *et al.*, 1985; Borinski and Holt, 1990; Luckevich and Beveridge, 1989; Stewart and Beveridge, 1980). Considering the energy expenditure, the function of the S-layer must be required for survival in the normal environment of the bacterium. It is presumed that most S-layers have a protective barrier role because the pore-like structures formed by the layer likely act as molecular sieves and prevent the entry of molecules, such as proteases and lytic enzymes, larger than the pore (Sleytr and Messner, 1983) as shown by several cases (Koval and Hynes, 1991; Sleytr, 1976). In addition, some infectious bacteria use their S-layers to adhere to and invade the cells of other organisms (Blaser *et al.*, 1988; Messner and Sleytr, 1992; Munn *et al.*, 1982). It has been demonstrated that the S-layer of *C. cresentus* protects it from a *Bdellovibrio*-like organism (Koval and Hynes, 1991), but the S-layer also acts as a receptor for the bacteriophage øCR30 (Edwards and Smit, 1991) showing that the S-layer also allows *C. crescentus* to be infected by a parasite.

S-layers have common features, such as an acidic pI, an absence of cysteine residues and a high number of hydroxylated amino acids. Subunits are held

together and to the surface by noncovalent (hydrophobic, ionic, hydrogen or polar) bonds (Koval and Murray, 1984; Messner and Sleytr, 1992; Sleytr and Messner, 1983). Despite these similarities, there is very little sequence similarity among S-layer proteins (Gilchrist *et al.*, 1992; Messner and Sleytr, 1992), suggesting that S-layers may have arisen by convergent evolution.

The S-layer of *C. crescentus* is composed of the protein RsaA. Six copies of RsaA form a ring-like subunit (Fig. 1-3) that interconnects with other subunits to form a two-dimensional hexagonal array (Smit *et al.*, 1992). The gene for RsaA has been cloned (Smit and Agabian, 1984) and sequenced (Gilchrist *et al.*, 1992). N-terminal protein sequencing of the mature RsaA polypeptide has shown that only the initial N-formyl methionine is cleaved, leaving a mature polypeptide of 1025 residues with a molecular weight of 98 kDa (Fisher *et al.*, 1988; Gilchrist *et al.*, 1992). The S-layer is anchored to the cell surface via a noncovalent interaction between the N-terminus of the protein and a specific smooth LPS in the outer membrane (Walker *et al.*, 1994). Ca^{2+} is required for the proper crystallization of RsaA into the S-layer and its removal using EGTA disrupts S-layer structure (Nomellini *et al.*, 1997; Walker *et al.*, 1994).

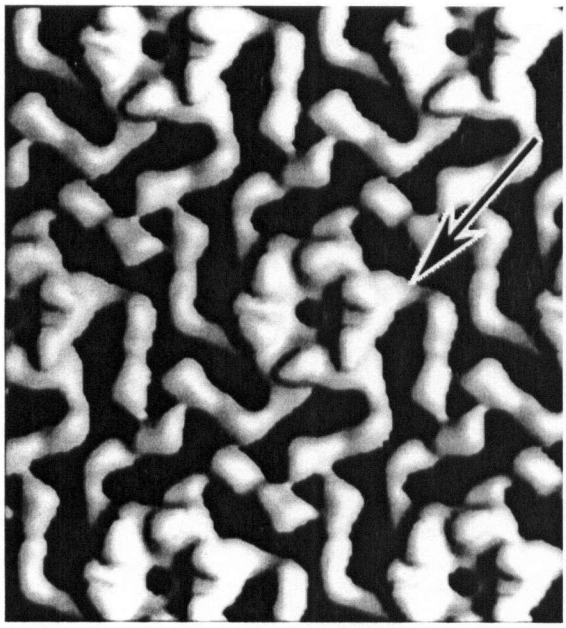

Figure 1-3. 3-Dimensional reconstruction of the S-layer. The arrow indicates a single C-shaped RsaA monomer. (Figure from Smit *et al*, 1992).

RsaA is a true secreted protein. RsaA must pass through both the inner and outer membranes to form the S-layer on the outer surface of the bacterium. As there is a large amount of RsaA (10 to 12% of the cellular protein), an efficient secretion system or a large number of transport complexes are required to secrete the protein during the 105 min generation time. Linker mutagenesis of RsaA has shown that the

extreme N-terminus is required for surface attachment while the C-terminus is required for secretion. Further, deletion and hybrid protein analyses have indicated that secretion of RsaA relies on an uncleaved C-terminal secretion signal located within the last 82 amino acids of the RsaA protein (Bingle *et al.*, 1999; Bingle *et al.*, 1996; Bingle *et al.*, 1997a; Bingle *et al.*, 1997b; Bingle and Smit, 1994). The presence of an uncleaved C-terminal secretion signal usually indicates secretion by a type I system (Binet *et al.*, 1997; Salmond and Reeves, 1993) rather than a type II, III or IV system. Most Gram-positive bacterial S-layers have been shown to use the General Secretion Pathway (GSP) or Sec-dependent pathway (Pugsley, 1993) for export (Messner and Sleytr, 1992; Sleytr and Messner, 1988; Sleytr *et al.*, 1993; Sleytr and Sara, 1997), whereas S-layer proteins in Gram-negative bacteria are secreted using a type II system (Boot and Pouwels, 1996) which also employs the GSP to transport the S-layer subunit across the inner membrane. Recently, it has been shown that the S-layer of *Campylobacter fetus* is secreted by a type I mechanism (Thompson *et al.*, 1998) and an S-layer-like protein in *Serratia marcescens* with significant similarity to RsaA has been shown to use a type I secretion mechanism (Kawai *et al.*, 1998).

In addition to the secretion signal, the C-terminal portion of RsaA also contains repeats of a glycine and aspartate acid-rich region which are thought to bind calcium ions (Gilchrist *et al.*, 1992) and result in the aggregation of free RsaA in the medium. Such Ca^{2+}-binding motifs are found in most proteins secreted by type I systems (Binet *et al.*, 1997) and consist of a glycine/aspartate rich GGXGXD motif that repeats 4-36 times (Welch, 1991). *C. crescentus* has two groups of three repeats separated by 12-16 residues containing this motif. Interestingly, there are no obvious repeat regions in the S-layer of *C. fetus* (Thompson *et al.*, 1998). It has been suggested that these motifs are important for the proper presentation of secretion signal to the ABC transporter (Duong *et al.*, 1996; Létoffé and Wandersman, 1992; Sutton *et al.*, 1996). Thus, in the case of RsaA, the glycine and aspartate rich repeats may function (along with Ca^{2+}) both in maintaining the crystalline structure of the S-layer and in the secretion of the S-layer protein itself.

There are four described Gram-negative bacterial transport systems. These systems have been named type I through type IV. The type I system requires 3 proteins that are thought to form a pore through the inner and outer membranes allowing the protein to be secreted. This is the method by which RsaA is secreted and it is discussed in depth below.

Type II systems use the GSP for export across the inner membrane and then use a complex of 12-14 proteins for secretion to the outside of the bacterium. The secretion substrates contain classical Sec-dependent N-terminal signal sequences that direct transport across the inner membrane by the Sec pathway (Pugsley, 1993). Proteins are transported across the cytoplasmic membrane in an unfolded state and then fold in the periplasm. This folding is necessary as the components for secretion seem to recognize the secondary or tertiary structure of the substrate as no sequence similarity has been found (Lu and Lory, 1996). Both ATP hydrolysis as well as proton motive forces appear to be required for secretion of the substrate (Feng *et al.*, 1997; Letellier *et al.*, 1997). For a review of type II secretion systems see Russel, 1998.

The auto-secreting proteins, such as the IgA proteases, like the type II secreted proteins, use the GSP to cross the inner membrane. These proteins have an N-terminal signal sequence and a C-terminal pro-sequence. They are exported across the cytoplasmic membrane by the Sec dependent pathway in the usual manner with cleavage of the N-terminus signal sequence. The pro-sequence then forms a pore in the outer membrane through which the rest of the protein passes. Once the protein is outside, autocatalytic cleavage of the pro-sequence occurs, releasing the protease from the cell (Pohlner *et al.*, 1987).

Type III secretion has only been found in pathogens and is used to deliver bacterial proteins into the host cytoplasm to alter the host's metabolism to the advantage of the bacterium. Type III systems are the most complex of the secretion systems, involving more than 20 proteins. The proteins form a needle-like structure that spans the inner and outer membrane (Kubori *et al.*, 1998). Before secretion can occur, the bacterium must make contact with the host cell. Secretion seems to be directed by the mRNA. It is thought that the mRNA forms a hairpin loop that obscures the translation start signal until the 5' region of the mRNA interacts with the

secretion apparatus (Anderson and Schneewind, 1997). A signal recognition protein may mediate this process. Therefore, secretion is coupled with translation. ATP hydrolysis appears to be required for secretion, as components of type III systems are capable of hydrolyzing ATP *in vitro* (Eichelberg *et al.*, 1994). The substrate may then pass through the needle structure to the outside of the cell, though this has not been proven. For reviews of type III secretion see Anderson and Schneewind, 1999; Galan and Collmer, 1999

Type IV secretion systems have only recently been discovered and are not well understood. This transport pathway, like the type III, has so far been found exclusively in pathogens. The type IV system seems to have been designed to transport DNA, though the *Bordetella pertussis* Ptl system only transports proteins (Weiss *et al.*, 1993). There are at least 9 proteins involved in the transport process and their functions are not well understood. There are usually two proteins containing nucleotide binding motifs that appear to be the transporting units that hydrolyze ATP to effect transport. It is not known if the substrate is transported in a one step process where the substrate bypasses the periplasm or a two step process where the substrate is first transported to the periplasm and then a second transport process secretes the protein. For a review of type IV secretion see Burns, 1999

RsaA is secreted by a type I mechanism. The goal of this thesis was to elucidate the secretion mechanism of RsaA. Initial indications suggested that it was a type I secretion mechanism (i.e., a C-terminal secretion signal and the presence of glycine/aspartate rich repeats) and data are presented here directly demonstrating that RsaA is secreted by a type I mechanism. Figure 1-4 shows the predicted structure of the *C. crescentus* membrane and also serves as a general model of a type I mechanism.

The best described type I secretion systems are those required for the secretion of *Escherichia coli* α –hemolysin (HlyA), *Erwinia chrysanthemi* metalloproteases (PrtB) and *Pseudomonas aeruginosa* alkaline protease (AprA) (Binet *et al.*, 1997; Salmond and Reeves, 1993). A type I secretion apparatus requires three components (Delepelaire and Wandersman, 1991). One component, the ABC transporter, is embedded in the inner membrane and contains an ATP-

binding cassette (ABC). It has been shown that this component recognizes the C-terminal signal sequence of the substrate protein and hydrolyzes ATP during the transport process (Binet and Wandersman, 1995; Koronakis *et al.*, 1993). Another component, the membrane fusion protein (MFP), is anchored in the inner membrane and appears to span the periplasm (Dinh *et al.*, 1994). The remaining component is an outer membrane protein (OMP) that has been shown to interact with the MFP. It is thought that these three components form a channel that extends from the cytoplasm through the two membranes to the outside of the cell (Akatsuka *et al.*, 1997; Hwang *et al.*, 1997). The substrate may pass through this channel (probably in an unfolded state) to the outside of the cell. In many cases, the genes for all three transport components are found immediately adjacent to the substrate gene(s) (Duong *et al.*, 1992; Létoffé *et al.*, 1990). In other type I systems, only the ABC-transporter and MFP genes are next to the substrate gene (Létoffé *et al.*, 1994b; Mackman *et al.*, 1985). The Rsa genes are organized like the latter and the OMP gene is not adjacent to the ABC-transporter and MFP. Recently, it was determined that the OMP gene is only separated from the MFP gene by five ORFs and a distance of 5 kb in the Rsa system. There are also instances where the substrate gene is separate from the secretion genes (Finnie *et al.*, 1998; Scheu *et al.*, 1992). As shown in Figure 1-4, from analysis of the ABC-transporters it is thought that the protein components work in multimers of at least 2. Some members of the ABC-transporter family, such as P-glycoprotein, contain two

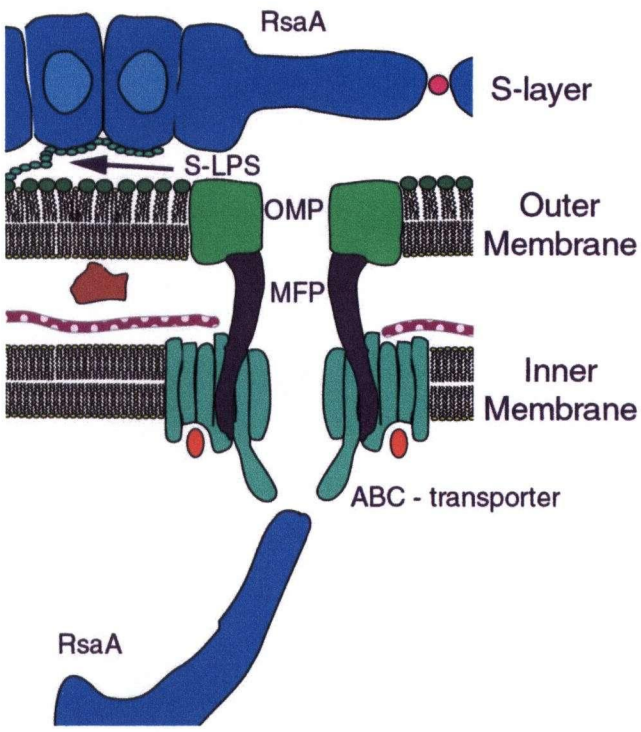

Figure 1–4. Type I secretion system. Diagram of the hypothetical membrane architecture of *C. crescentus* showing the predicted type I secretion mechanism of RsaA

almost identical domains in tandem, each with its own membrane spanning and ABC region (Sheps *et al.*, 1996). Association of two ABC transporters has been shown for monomeric ABC-transporters (Davidson and Nikaido, 1991). The proteins may work in pairs so that one ATP is hydrolyzed for transport and a second ATP is hydrolyzed to return the complex to the resting conformation. It is also possible that the proteins work in tandem and small sequential conformational changes in each separate protein push the proteins along (Welsh, 1998). Recent work indicated that while the ABC-transporters may work as a dimer, the MFP may work as a hexamer and the OMP as a trimer (Holland, 1999; Koronakis *et al.*, 1997).

The ABC-transporter family is very large and the type I secretion systems make up only a small portion. They are found in all forms of life and are sufficient to transport a substrate across a single membrane. There is significant sequence similarity among the ABC-transporters, even between eukaryotic and prokaryotic genes. The eukaryotic P-glycoprotein shares close to 50% conserved amino acids with many of the bacterial ABC-transporters such as HlyB and PrtD over the entire length of the protein (Croop, 1998; Sheps *et al.*, 1996). Mammalian P-glycoproteins actually have more sequence identity to these prokaryotic transporters than to proteins considered to belong to the P-glycoprotein family. ABC-transporters are also involved in the import of substrates such as the Mal transporter where maltose is transported across the inner membrane (for reviews see Boos and Shuman, 1998; Ehrmann *et al.*, 1998; Nikaido, 1994).

The basic monomeric ABC-transporter consists of 2 domains. One domain, usually N-terminal and consisting of six to eight membrane spanning segments, recognizes the substrate and forms the pore through the membrane. The other domain contains the ABC region, which provides the energy for transport from the hydrolysis of ATP. The ABC domain is highly conserved and consists of about 215 amino acids and within this region there are four distinct motifs. Like all ATPases, ABC-transporters contain Walker A or P-loop (consensus GXXGXGK[ST])[1] and Walker B (hhhhD)[1] motifs which interact directly with ATP binding and hydrolysis

[1] X-denotes any amino acid; h-denotes hydrophobic amino acid; brackets indicate alternative amino acids at a single position

(Walker *et al.*, 1984), but they are immediately followed by a specific ABC-transporter motif (LSGGQ[QRK]QR)[1] (Bairoch, 1992; Gorbalenya and Koonin, 1990) which is thought to be involved in energy transduction (Hyde *et al.*, 1990). A fourth motif has recently been identified in a majority of *E. coli* and *Saccharomyces cerevisiae* ABC-transporters (Decottignies and Goffeau, 1997; Linton and Higgins, 1998). This fourth motif is hhhhH[1] followed by a charged residue and is found approximately 30 amino acids C-terminal of the aspartic acid in the Walker B motif. No one has so far been able to make a 3-dimensional crystal of the complete ABC-transporter from which the structure could be determined. However, the ABC domain has been crystallized from two proteins (Armstrong *et al.*, 1999; Hung *et al.*, 1998) showing that the ABC forms an L with 2 arms; arm 1 binds with the ATP and arm 2 interacts with the membrane-spanning domain. It is thought that hydrolysis of ATP causes a conformational change in arm 2 which transfers the energy to the membrane spanning domain, possibly through the ABC-transporter motif found at the end of arm 2, and the conformational change in the membrane spanning domain results in transport of the substrate (Welsh, 1998).

The MFP is characterised by a single hydrophobic transmembrane domain in the N-terminus that sits in the inner membrane. A hydrophilic domain spans the periplasm and the C-terminus consists of beta sheet that may interact with the outer membrane component (Dinh *et al.*, 1994). The MFP family contains the conserved motif [LIVM]XXG[LM]XXX[STGAV]X[LIVMT]X[LIVMT][GE]X[KR]X[LIVMFYW] [LIVMFYW]X[LIVMFYW][LIVMFYW][LIVMFYW][1] (PROSITE:PDOC00469)

The OMP sits in the outer membrane and interacts with the MFP. Of the known OMPs only TolC, from the α-hemolysin transporter, has been studied extensively. It has been found that three smooth LPS synthesis genes are required for secretion of α-hemolysin. It is likely that the smooth LPS is required for proper insertion of TolC in the membrane (Stanley *et al.*, 1993; Wandersman and Létoffé, 1993). Two-dimensional crystals of TolC have been examined using electron microscopy and show that TolC forms a trimer. It also appears that a portion of the C-terminus is located in the periplasm (Koronakis *et al.*, 1997). TolC contains a centrally located sequence of 44 amino acids in the middle of the protein that is highly similar to a sequence in HlyD (the MFP); these sequences are required for

transport and can be interchanged and still allow transport (Schulein *et al.*, 1994). Thus, TolC is thought to provide the essential function of linking the transporter complex to the external environment.

While members of the ABC-transporter family secrete a huge range of substrates ranging from Ca^{2+} ions to cancer drugs to proteins, the type I secretion subfamily has been found to only secrete proteins. The specific features for secretion of a protein by a type I system are not known except that the secretion signal is located in approximately the last 60 amino acids of the C-terminus of the protein (Mackman *et al.*, 1985). As little as 15 amino acids of the C-terminus of the protease, PrtG, from *E. chrysanthemi* still allows secretion, although this is only 1% as effficient. It was found that substrates can be secreted by closely related type I systems (Binet and Wandersman, 1996; Létoffé *et al.*, 1994a; Létoffé *et al.*, 1994b), but only if there is more than 25% amino acid identity between ABC-transporters of the systems (Delepelaire and Wandersman, 1990; Fath *et al.*, 1991). No sequence similarity is found among the secretion signals of the different substrate proteins; however, in the proteases, lipases and NodO a conserved motif of a negatively charged amino acid followed by several hydrophobic amino acids has been found at the end of the C-terminus (Binet *et al.*, 1997). The C-terminal signal sequence of α-hemolysin was extensively mutagenized, but few individual amino acids were found to affect secretion (Kenny *et al.*, 1992). Because of this lack of sequence similarity and identification of important residues it is thought that the secretion signal relies on secondary structure to initiate transport. NMR and circular dichroism studies of the C-terminus of PrtG, HasA (the heme acquisition protein from *Serratia marcescens*), HlyA (the hemolysin from *E. coli*) and LktA (the leukotoxin from *Pasteurella haemolytica*) have shown that there are two α helices in the C-terminus (Wolff *et al.*, 1997; Wolff *et al.*, 1994; Yin *et al.*, 1995). Mutation of these α helical regions in HlyA and LktA showed that the secretion signal appears to bind to a pocket in the ABC-transporter and induce a conformational change that causes transport to occur (Zhang *et al.*, 1998).

Presented in this thesis is evidence that all three components of a type I secretion system have been found in *C. crescentus* and these components are required for the secretion of RsaA. They have greatest similarity to the protease

type I secretion systems from *P. aeruginosa* and *E. chrysanthemi* and the proteases from these systems can be secreted by the Rsa system.

The S-layers subunits from other *Caulobacter* species appear to be secreted by type I systems. Several FWC species with S-layers have been isolated from a wide number of aquatic sources (MacRae and Smit, 1991; Walker *et al.*, 1992). The subunits of these S-layers react with anti-RsaA antibody and their smooth-LPS reacts with antibody raised against the smooth-LPS of NA1000. The S-layer subunits from these FWC species range in size from 100 to 193 kDa and can be removed from the bacterium's surface using low pH or EGTA (Walker *et al.*, 1992). Portions of the genome of the FWC species with S-layers hybridize to the rsaA gene while the genomes of FWC species without S-layers do not (MacRae and Smit, 1991). It is shown in Ch. 5 that the protease, AprA from *P. aeruginosa*, was expressed and secreted in some of these FWC species. These facts suggest that type I secretion mechanisms secrete the S-layer subunits in the FWC species. Since the FWC species secrete S-layer subunits varying widely in size, it is desirable to examine the S-layer subunits and their corresponding secretion systems and examine the differences and similarities to allow one to determine how the mechanisms work, what parts of the protein are essential for secretion and what parts provide specificity. With these goals in mind, procedures are reported here for the characterisation of the S-layer subunit, ABC-transporter and MFP genes from various FWC species.

The S-layer is attached to the surface of *C. crescentus* using a species of smooth LPS. The outer membrane of Gram-negative bacteria contains phospholipids, proteins and LPS (Nikaido and Vaara, 1985). In many cases, including *C. crescentus,* there is also an extracellular polysaccharide (EPS) (Ravenscroft *et al.*, 1991); the S-layer is external to all of these molecules (although the EPS may pass through the S-layer). Smooth LPS is a major component of the outer membrane of Gram-negative bacteria and consists of three regions. The lipid A moiety is the endotoxic part of LPS and is anchored in the outer leaflet of the outer membrane. The core, a branched chain oligosaccharide linked to ketodeoxy

octulosonic acid (KDO), is attached to the lipid A molecule. Extending from the core is the O-antigen which contains a repeating linkage of oligosaccharides (Schnaitman and Klena, 1993). It has been found in *C. crescentus* that the S-LPS anchors the S-layer to the cell surface via a noncovalent interaction with the N-terminus of RsaA. Immunolabelling showed that the S-LPS is completely occluded by the S-layer (Walker *et al.*, 1994). Isolation and characterization of the S-LPS showed that the core sugars and fatty acids are identical to those of the rough LPS and that the O-antigen is of a homogeneous length, unlike the variable length S-LPS found in many enteric bacteria. Previous reports (Walker *et al.*, 1994) indicated that the O-antigen was composed of a 4,6-dideoxy-4-amino-hexose, a 3,6-dideoxy-3-amino-hexose and glycerol, but recent results (Smit, unpublished) indicate that glycerol is a contaminant of the S-LPS isolation procedure, and that the 3,6-dideoxy-3-amino-hexose assignment is likely due to a co-purifying polymer. Therefore, it seems possible that the O-antigen is composed solely of a 4,6-dideoxy-4-amino-hexose. Anomeric traces found by analysis of proton NMR spectra indicate that the linkages between the 4,6-dideoxy-4-amino-hexose are not identical, implying the involvement of a larger number of glycosyltransferases than needed for a simple polymer with only one kind of linkage.

These data correlate with the information presented in this thesis. I have found a number of S-LPS synthesis genes, indicating that *C. crescentus* may make perosamine, a 4,6-dideoxy-4-amino-hexose, and that perosamine is likely a component of the S-LPS. A number of glycosyltransferases were also found as would be expected considering that several transferases would be required to produce the different linkages that result in the different anomeric proton traces found by proton NMR.

Evidence is presented in this thesis demonstrating how RsaA is secreted and how the S-LPS, involved in attachment of the S-layer, is synthesized. Three genes composing the ABC-transporter, MFP and OMP of a type I secretion system required for secretion of RsaA in *C. crescentus* are described. A type I secretion system is also required for secretion of the S-layer subunits of other FWC species. The genes required for the secretion of RsaA and the synthesis of S-LPS are linked

leading to the discovery of a number of putative genes involved in the synthesis of the S-LPS required for S-layer attachment. Additional genes involved in synthesis of the S-LPS were discovered by Tn5 mutagenesis.

Chapter 2
Materials and Methods

Strains, plasmids and growth conditions. All strains, libraries and plasmids used in this study are listed in Table 2-1. Plasmids with NA1000 DNA inserts are listed in Figure 2-1.The *E. coli* strains DH5α JM109 or RB404 were used for all *E. coli* cloning manipulations. *E. coli* was grown at 37°C in Luria broth (1% tryptone, 0.5% NaCl, 0.5% yeast extract), with 1.2% agar for plates. *C. crescentus* strains were grown at 30°C in PYE medium (0.2% peptone, 0.1% yeast extract, 0.1% $CaCl_2$, 0.2% $MgSO_4$, with 1.2% agar for plates). Ampicillin was used at 100 μg/ml, streptomycin at 50 μg/ml, kanamycin at 50 μg/ml in both *C. crescentus* and *E. coli*, and tetracycline was used at 0.5 μg/ml and 10 μg/ml and chloramphenicol was used at 2 μg/ml and 20 μg/ml in *C. crescentus* and *E. coli,* respectively, when appropriate.

Recombinant DNA manipulations. Standard methods of DNA manipulation and isolation were used (Sambrook *et al.*, 1989). Electroporation of *C. crescentus* was performed as previously described (Gilchrist and Smit, 1991). Southern blot hybridizations were done according to the membrane manufacturer's instructions (Amersham Hybond-N). Southern blot analysis allowing up to 30% mismatch between the probe and chromosomal DNA was performed in an identical manner except the hybridization step was performed at 50°C instead of 65°C. Blots were washed: twice for 15 min at room temperature with 2X SSPE (0.18M NaCl, 0.01M $NaPO_4$, 0.001 EDTA pH 8.0), 0.1% SDS; once for 15 min at 50°C with 1X SSPE, , 0.1% SDS. Radiolabelled probes were made by nick translation using the DNase/DNA Pol manufacturer's instructions (GIBCO/BRL). Chromosomal DNA was isolated as previously described (Yun *et al*, 1994).

PCR products were generated using the primers listed in Table 2-2. PCR was performed using Taq polymerase (BRL), following the manufacturer's suggested protocols. Annealing temperatures (T_A) 2°C below the melting temperature T_m of the

Table 2-1. Strains and Plasmids used in this study

	Relevant characteristics	Reference or Source
Bacterial strains		
E. coli		
JM109	*recA1, endA1,gyrA96, thi,hsdR17,supE44,relA1, Δ(lac-proAB),λ⁻*F', [*traD36, proAB lacI^q, lacZΔM15*]	(Yanisch-Perron *et al.*, 1985)
RB404	F-*dam-3,dam-6,metB1, galK2, galT22 lacY1, thi-1, tonA31, tsx-78, mtl-1, supE44*	(Brent and Ptashne, 1980)
DH5α	*recA1, endA1,gyrA96, thi,hsdR17,supE44,relA1, Δ(lacZYA-arfF)U196 λ⁻* (ϕ80*lacZΔM15*)	Life Technologies
C. crescentus		
NA1000	Ap^r, *syn-1000*. Variant of wild-type strain CB15, ATCC 19089, that synchronizes well	
JS1001	S-LPS mutant of NA1000, sheds S-layer into medium	(Edwards and Smit, 1991)
JS1003	NA1000 with *rsaA* interrupted with KSAC Km^r cassette	(Edwards and Smit, 1991)
JS3001	S-LPS mutant of ATCC 15252, sheds S-layer into medium	
JS4000	S-layer negative, derivative of ATCC 15252	
Plasmids		
pBBR1MCS	Cm^r, broad host range vector	(Kovach *et al.*, 1994)
pBBR1AprF	*Eco*R1-*Bam*H1 fragment containing *aprF* from pJUEK72 in pBBR1MCS	this study
pBBR1PrtF	*Hin*dIII-*Pst*I fragment containing *prtF* from pRUWinh4 in pBBR1MCS	this study
pBBR3	Sm^r, broad host range vector	this study
pBBR3AprA	*aprA*⁺, *aprA* cloned into pBBR3 using *Eco*R1 and *Pst*1	this study
pBBR3PrtB	*prtB*⁺, *prtB* cloned into pBBR3 using	this study
pBBR3AprA:pRAT5	*aprA*⁺, *rsaD*⁺, *rsaE*⁺, Ap^r, Sm^r, pBBR3AprA fused with pRAT5 at the *Xba*1 site	this study
pBBR3PrtB:pRAT5	*prtB*⁺, *rsaD*⁺, *rsaE*⁺, Ap^r, Sm^r, pBBR3PrtB fused with pRAT5 at the *Xba*1 site	this study
pBBR3AprA: pCR2.1F11Sal1	*aprA*⁺, *rsaF*⁺, pBBR3AprA fused with pCR2.1F11Sal1 at the *Xba*1 site	this study
pBBR5	Tc^r, broad host range, broad host range	this study
pBSKS+	ColE1 cloning vector, *lacZ*, Ap^r	Stratagene
pBSKS-gcc1984	736 bp PCR product containing valyl tRNA synthetase made using the primers gcc1984-1407 and gcc1984-I2143 and T-tailed into pBSKS	this study
pCR2.1	Km^r, Ap^r, commercial T-tail cloning vector	Invitrogen
pCR2.1F11Sal1	PCR product generated using Tn5 and Tn5Sal1 primers from ligation of F11Tn5 chromosomal DNA cut with Sal1 in pCR2.1	this study
pCR2.1F11Xma1	PCR product generated using Tn5 and Tn5Xma primers from ligation of F11Tn5 chromosomal DNA cut with Xma1 in pCR2.1	this study
pCR2.1rsaF(1984)	2.1kb PCR product generated using primers gcc1984-28 and gcc1984-I2310	this study
pJUEK72	*aprD*⁺, *aprE*⁺, *aprF*⁺, *aprA*⁺, *apr*I⁺	(Guzzo *et al.*, 1990)
pRAT1	*rsa*A⁺, *rsaD*⁺, *rsaE*⁺, Ap^r	this study
pRAT4ΔH	*rsa*A⁺, *rsaD*⁺, *rsaE*⁺, Ap^r, *rsa*A is under control of a *lacZ* promoter	this study
pRAT4ΔH : pBBR5	*rsa*A⁺, *rsaD*⁺, *rsaE*⁺, Ap^r, Tc^r, pBBR5 was fused with pRAT4ΔH at the *Sst*1 site	this study
pRAT5	*rsaD*⁺, *rsaE*⁺, Ap^r	
pRAT5 : pRK415	*rsaD*⁺, *rsaE*⁺, Ap^r, Tc^r, pRK415 was fused with pRAT5 at the *Sst*1 site	this study
pRAT HI (B/E)	*Bam*H1/ *Eco*R1 fragment from pRAT1 cloned into pTZ18U	this study
pRK415	*lacZ*⁺,Tc^r, broad host range	(Keen *et al.*, 1988)
pRK415 *rsaA* ΔPK	*rsa*A under control of *lacZ* promoter in pRK415	this study
pRUW500	*prtB*⁺, Ap^r	(Delepelaire and Wandersman, 1990)
pRUW500 : pRK415	*prtB*⁺, Tc^r pRK415 was fused with pRUW500 at the *Pst*1 site	this study
pSUP2021	carries Tn5, unable to replicate in *C. crescentus*	(Simon *et al.*, 1983)
pTZ18UB:*rsaA*ΔP	The wildtypepromoter of*rsaA* has been replaced with a *lacZ* promoter	(Bingle *et al.*, 1997)
pTZ18R and pTZ18U	Ap^r,ColE1 cloning vector	(Mead *et al.*, 1986)
pTZ19U	A phagemid version of pUC18 or pUC19	
pTZ18U(CHE)	Cm^r, Ap^r gene of pTZ18U replaced with Cm^r gene	this study
pTZ19UΔSSm	Sm^r, Sm^r gene inserted into *Sca*1 site in Ap^r gene of pTZ19U	this study
pTZ18R *aprA*	*aprA*⁺, Ap^r The *Eco*R1-*Bgl*II fragment from pJUEK72 containing *aprA* was inserted into the *Eco*R1-*Bam*H1 sites of pTZ18R	this study
pTZ19UΔSSmΔNΔC-RsaF(973)	internal *Kpn*1-*Pst*1 fragment of *rsaF*(973) in pTZ19UΔSSm	this study
pTZ19UΔSSm973circ	recircularized plasmid isolated from *Bam*H1 digestion of NA1000::pTZ19UΔSSmΔNΔC-RsaF(973)	this study
pTZ18U(CHE)ΔNΔC-RsaF(1984)	internal *Pvu*II-*Stu*1 fragment of *rsaF*(1984) in *Sma*1 site of pTZ18U(CHE)	this study
pUC8	ColE1 cloning vector, *lacZ*, Ap^r	(Vieira and Messing, 1982)
pUC9 *rsaA*ΔNΔC	*rsaA* missing the extreme N-terminus and C-terminus	(Bingle *et al.*, 1996; Bingle and Smit, 1994)
pUC8 *neoR*	*Hin*dIII-*Bam*H1 from Tn5 containing neomycin resistance gene in pUC8	this study
pTZ18R *aprA* : pRK415	*aprA*⁺, Tc^r pRK415 was fused with pTZ18R at the *Bam*H1 site	this study
Libraries		
NA1000 cosmid	1000 cosmids containing 20 - 25 Kb of NA1000 DNA	(Alley *et al.*, 1991)

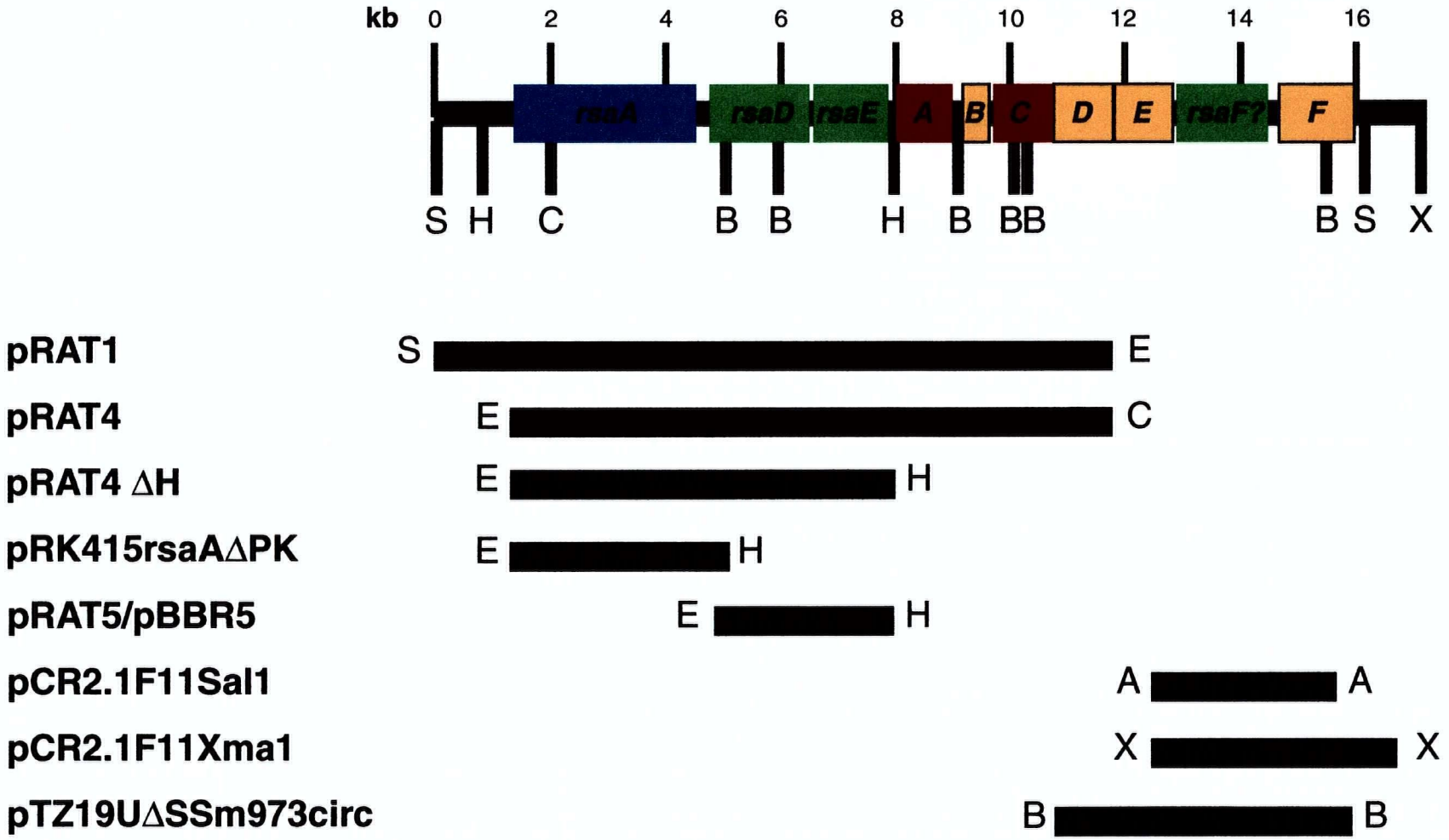

Figure 2–1. Plasmids containing NA1000 chromosomal DNA. E-*Eco*R1, B-*Bam*H1, C-*Cla*1, A-*Sal*1, H-*Hind*III S-*Sst*1

primers were used. Extension times (t_E) were based on 60 sec/1000 bp of DNA. General PCR parameters were 95°C – 30 sec, T_A - 30 sec, 72°C - t_E. The vector pBSKS+ was cut at the *Eco*RV site and T-tailed (Holton and Graham, 1991) and the PCR product was ligated into this vector.

Cloning of chromosomal DNA adjacent to Tn5 insertions: Chromosomal DNA of the Tn5 mutant was cut with *Bam*HI, *Sal*1 or *Xma*1. *Bam*HI fragments were cloned directly into the *Bam*HI site of pTZ18 vectors. A second method that was used for isolating the chromosomal DNA adjacent to the Tn5 insertions involved an inverse PCR method developed by V. Martin (Martin and Mohn, 1999).

PCR product	forward primer name- sequence (5'-3')	reverse primer name- sequence (5'-3')
RAT5	RsaD-A-CGGAATCGCGCTACGCGCTGG	RsaE1-GGGAGCTCGAAGGGTCCTGA
Degenerate primers for RsaF search	F60- (GC)CG(GC)(AGT)(GC)(GTC)(GC)(GC)(GC) (CT)T(CG)CT(CG)CC(CG)CAGCT(CG)G FB110- CT(GC)(CA)(GC)CAG(AC)C(GC)AC)T(GC)T TCGAC	IF340- GCCGCC(CG)(CGT)(TAG)(GA)(TA)A(GC)A (GT)(GC)GG(GC)AG(GC)(TCG)(TA)(CG)T IFB415- CTG(TC)TC(GC)GC(GC)(AT)(CT)(GC)AG(G C)ACGTC
Inverse PCR to obtain chromosomal DNA next to Tn5 insertion	Tn5 universal – GGTTCCGTTCAGGACGGCTAC	Tn5Xma1-AGGCAGCAGCTGAACCAA Tn5Sal1-ATGCCTGCAAGCAATTCG
Degenerate primers for amplification of internal portion of RsaD homologues in FWC species	RD43B- TA(TC)ATGCT(GC)CAGGT(GC)TAT(GC)AC CGIG	IRD477B- C(GC)A(GT)(GC)CGCTG(GC)CGCTGGCC GC
Unsuccessful PCR of *rsaF*(1984)	RsaF140-GCGGTCGAGCAGGGGGTGCT	RsaFIEND-ACGAATCCTTGCGCGCCTTGG
Amplification of pUC type vectors	TZ1920- GAGGCCTAGTACTCTGTCAGACCAAGTTT ACTCATA	TZI1060- GAGGCCTACTCTTCCTTTTTCAATATTATT GAA
Amplification of gcc1984 (numbers correspond to bp in contig)	Gcc1984-28- CGCTCTACACCGGCGGTCGCGCCAGCGC Gcc1984-1407- GCCGGAACCCGAACCTGAACCGGTGTCG	Gcc1984-I1200 – GGAGCTCTGGCGCCCCACCAGGGACGC GTAGAACG Gcc1984-I2143- GTGGTCGGTGCCCGGCAGCCACAGGG
Amplification of gcc973 (numbers correspond to bp in contig)	Gcc973-1600- GGAATCCATGTCACATGGGAAGAGACGG TCCGCCGT	Gcc973-I2310- GCTGGCGCCCCACCAGGGACGCGTAGA ACG

Table 2-2. Primers used for PCR for this report.

Construction of plasmid vectors that replicate in *C. crescentus.*: The plasmid pBBR5 was constructed from the plasmids pBBR1MCS (Kovach *et al.*, 1994) and pHP45Ω-Tc (Fellay *et al.*, 1987). The Ω-Tc fragment from pHP45Ω-Tc was removed using *Hind*III and the ends were blunted using T4 polymerase. A 0.3 kbp portion of the Cm^r gene was removed from pBBR1MCS by cutting with *Dra*I and replaced with the blunted Ω-Tc fragment producing a Tc^r broad host range vector that replicates in *C. crescentus*. The plasmid pBBR3 was constructed in an identical manner except the plasmid pHP45Ω-Sm (Fellay *et al.*, 1987) was used to provide a Sm^r marker. Both these plasmids were constructed by John Nomellini.

Construction of vectors that replicate only in *E. coli*: The vector pTZ18U(CHE) was constructed by amplification of all of pTZ18U except the ap^r gene using the primers TZ1920 and TZI1060 that were designed with *Stu*1 sites. The PCR product was cut with *Stu*1 and a Cm^r gene (Morales *et al.*, 1991) with blunt ends was inserted into the site.

Tn5 mutagenesis. Tn5 mutagenesis was accomplished using the narrow host range (ColE1 replicon) plasmid pSUP2021 (Simon *et al.*, 1983) which is not maintained in *C. crescentus*. The plasmid was introduced by electroporation and 20,000 colonies that were streptomycin and kanamycin resistant were pooled, frozen at -70°C and aliquots were used for subsequent screening.

Southern blot analysis of chromosomal DNA isolated from the Tn5 library was used to assess the randomness of insertions. Hybridization with a Tn5 probe, pUC8neoR, indicated that while there were some hot spots of Tn5 integration, the Tn5 insertions were randomly distributed throughout the chromosome (data not shown).

SDS-PAGE and Western blot analysis. Proteins and S-LPS were isolated from *C. crescentus* as previously described (Walker *et al.*, 1992; Walker *et al.*, 1994). SDS-polyacrylamide gel electrophoresis (PAGE) and Western immunoblot analysis was performed as previously described (Walker *et al.*, 1992). After transfer of proteins to nitrocellulose, the blots were probed with polyclonal antibody and antibody binding

was visualized using goat anti-rabbit serum coupled to horseradish peroxidase and colour-forming reagents (Smit and Agabian, 1984).

To detect *C. crescentus* whole cells synthesizing an S-layer, a colony blot assay was used (Bingle *et al.*, 1997a). Briefly, cell material was transferred to nitrocellulose by pressing the membrane onto the surface of an agar plate. The membrane was air dried for 10 to 15 min, washed in a blocking solution (3% skim milk powder, 20 mM Tris (pH 8.0), 0.9% NaCl) with vigorous agitation on a rotary shaker and then processed in the standard fashion (Bingle *et al.*, 1997a).

Surface protein from *C. crescentus* cells was extracted using pH 2.0 HEPES buffer as shown by Walker (Walker *et al.*, 1992). To compare the amounts of surface protein extracted from different mutants equal amounts of cells growing at log phase were harvested and equal amounts of the protein extract were loaded on the protein gel. SDS-PAGE and Western blotting were performed according to standard procedures (Sambrook *et al.*, 1989).

Isolation of cosmids containing *rsaA*, *rsaD* and *rsaE*. The NA1000 and JS4000 cosmid libraries were probed with radiolabelled *rsaA*, using the plasmid pUC9 *rsaA*ΔNΔC. 5 cosmids from the NA1000 library were isolated and 4 cosmids from the JS4000 library. Southern blot analysis of the cosmids hybridizing to the probe was used to determine which cosmids contained DNA 3' of *rsaA*. An 11.7 kb *SstI-Eco*RI fragment containing *rsaA* plus 7.3 kb of 3' DNA was isolated from one of the NA1000 cosmids and cloned into the *SstI-Eco*RI site of pBSKS+; the resulting plasmid was named pRAT1. The 3' end of the cloned fragment consisted of 15 bp of pLAFR5 DNA containing *Sau3*A, *Sma*I and *Eco*RI sites. *Bam*HI fragments from the NA1000 cosmid were subcloned into the *Bam*HI site of vector pTZ18R for sequencing. The 3' end fragment was subcloned using *Bam*HI and *Eco*RI into pTZ18R. The 5' end fragment was subcloned using *SstI-Hind*III into pTZ18R. A cosmid containing the *rsaA, rsaD and rsaE* genes was isolated from the JS4000 cosmid library and pieces were subcloned as *BamHI* fragments in pTZ18U for sequencing. *Hind*III/*Bam*HI fragments containing the *rsaA* gene were cloned directly from the genome of JS4000 and JS3001 by isolating bands of the correct size from an agarose gel and ligating to pUC8. Colonies were probed with *rsaA* from NA1000 for plasmids

containing the correct insert. These clones were subcloned in three pieces as *Hind*III/*Cla*I, *Cla*I/*Eco*RV and *Eco*RV/*Bam*HI fragments into pUC type vectors. *Cla*I sites for cloning were generated in the vector by cutting with *Bam*HI and filling in the 5' overhangs with Klenow fragment. Ligation of the blunt ends then produces a *Cla*I site.

Isolation of FWC S-layer subunit genes. FWC27 chromosomal DNA was digested with *Bam*HI and *Pst*I. The digested DNA was ligated to a pTZ19U vector also digested with *Bam*HI and *Pst*I. A portion of the ligation mixture was electroporated in to *E. coli* JM109 and allowed to incubate at 37°C for 1 hour in 1 ml of Luria broth. The mixture was divided evenly and spread on 10 agarose plates and incubated overnight. The colonies were adsorbed to sterile filter paper (Whatman 541). The colonies were then lysed by soaking the filter paper in 0.5M NaOH for 5 min. The filter paper was neutralized by soaking the filter paper in 1M Tris-HCl (pH 7.0) for 5 min twice. A filter was then soaked in 0.5M Tris-HCl (pH 7.0), 1.5M NaCl. Then, the filter was washed with 70% EtOH and baked at 80°C for 2 hours. The filters were then probed with pUC8neoR using the Southern blot hybridization procedure allowing 30% mismatch (see above).

Nucleotide sequencing and sequence analysis. Sequencing was performed on a DNA sequencer (Applied Biosystems model 373). After use of universal primers, additional sequence was obtained by "walking along" the DNA using 15-20 bp primers based on the acquired sequence. DNA was sequenced in both directions for all original sequence, thereafter DNA was only sequenced in both directions when ambiguities were found. Nucleotide and amino acid sequence data were analyzed using Geneworks and MacVector software (Oxford Molecular Group) and the NCBI BLAST e-mail server using the BLAST algorithm (Altschul *et al.*, 1990). Primers were designed with the help MacVector and Amplify 1.2 (Engles, 1993) Protein alignments were generated using the ClustalW algorithm as implemented by the MacVector software using the default settings. The sequences for NA1000 *rsaADEF* and *lpsABCDEF* were submitted to Genbank and can be accessed as AF06235. The sequences for JS3001 *rsaA* and JS4000 *rsaADE* can be accessed

using the accession numbers AF193063 and AF193064. Preliminary sequence data of the *C. crescentus* genome was obtained from The Institute for Genomic Research through the website at http://www.tigr.org. Signal peptides predictions were made using the SignalP web server (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen *et al.*, 1997).

Chapter 3
Secretion of RsaA

Introduction

The major purpose of my thesis was to elucidate the transport pathway of RsaA. The strain NA1000 was chosen for these studies because *rsaA* had originally been isolated from NA1000 and it is this gene that has been sequenced and used for all recombinant manipulations in the Smit Lab. In addition a number of useful mutants, with and without S-layers have been derived from NA1000. The lack of a cleaved secretion signal, the presence of calcium repeats, no periplasmic intermediate and a C-terminal secretion signal, indicated that RsaA was probably transported using a type I secretion system (Bingle *et al.*, 1999; Bingle *et al.*, 1996; Bingle *et al.*, 1997a; Bingle *et al.*, 1997b; Bingle and Smit, 1994) in which case other proteins would be required for secretion.

Results and Discussion

***C. crescentus* was screened for genes involved in the secretion of the S-layer subunit, RsaA.** Since a type I secretion system uses 3 main proteins to form the transport mechanism, it was necessary to devise a method for finding the genes coding for the components by screening for the loss of RsaA secretion. Unfortunately, there is no easy method to detect the presence of RsaA on the exterior of a colony, as found for α-hemolysin or the metalloproteases which can be detected using blood or skim milk plates (Mackman *et al*, 1985; Wandersman *et al*, 1987).

Previous research had shown that the lytic phage φCR30 could only infect *C. crescentus* when an S-layer was present (Edwards and Smit, 1991). This phage was isolated using the strain CB15BE, a derivative of ATCC 19089, as is NA1000. When the phage was used to lyse NA1000 cells with an S-layer using an moi of 10^4, it was found that spontaneous mutants occurred at a high frequency of approximately 10^{-5}. When these mutants were examined, it was found that approximately 15% had lost their S-layer while the remaining 85% still retained their S-layer and were susceptible to re-infection. Obviously, the phage was not lysing all

the bacteria with an S-layer, since these bacteria still behaved like the wildtype strain. Of the bacteria that no longer had an S-layer, RsaA secretion was restored if a plasmid carrying the *rsaA* gene was expressed inside the bacterium (data not shown). It seems that the *rsaA* gene is a more likely target for mutation when selection pressure against the S-layer is applied. This is in agreement with the observation that many bacteria lose their S-layers during sub-culturing in the laboratory environment. This method was discarded in favour of a colony immunoblot assay which was much more labour intensive, but did not have a high background.

For the colony immunoblot assay, two polyclonal primary antibodies were used: α-RsaA (Walker *et al.*, 1992) and α-S-LPS (Walker *et al.*, 1994). α-RsaA reacts to RsaA and α-S-LPS reacts to the smooth LPS required for the anchoring of the S-layer to the surface of the bacterium (Walker *et al.*, 1994). When α-RsaA was used, colonies with an S-layer reacted with the antibody and appeared as a spot on the blot (Fig. 3-1). It was also found that a 'halo' could be detected around colonies when the S-layer could not anchor to the cells (e.g., cells with a defective S-LPS). The halo occurs when shed S-layer diffused away from the colony and was detected by α-RsaA as a ring around the colony (Fig. 3-1). When α-S-LPS was used, the antibody reacted to exposed S-LPS only when the cells of a colony lacked an S-layer; S-layer blocks the binding of α-S-LPS. RsaA appears to be completely degraded when it is not secreted (Bingle *et al.*, 1996; Bingle and Smit, 1994), therefore cell lysis during this procedure and release of unsecreted RsaA was not a concern. Using this method, it was possible to differentiate between cells secreting RsaA, cells secreting and shedding S-layer and cells without an S-layer.

Figure 3-1. Colony Immunoblot. Example of an immunoblot using α-RsaA against colonies demonstrating the different phenotypes exhibited.

Identification of Tn5 mutants lacking an S-layer. A pooled NA1000 Tn5 library was screened for S-layer negative mutants using the Western colony immunoblot

assay. In total, 9,000 colonies from the pooled Tn5 mutant library were screened using α-S-LPS antibody and 22,000 colonies were screened using α-RsaA. Eighteen Tn5 S-layer negative mutants were found. SDS-PAGE and Western blot analysis of whole cell lysates and culture supernatants confirmed that no S-layer was found in or on the cells or in the culture supernatant of these mutants (data not shown). One mutant, B12, on further examination was found to have an S-layer and was kept for use as a random Tn5 mutation control. Twenty-six Tn5 mutants with a shedding phenotype were also isolated during the screening and are described in Ch. 6.

Identification of Tn5 mutants defective in RsaA secretion. Several possible Tn5 insertion events, in addition to those in secretion genes could result in an S-layer negative phenotype. To eliminate Tn5 insertions in the *rsaA* gene, Southern blot analysis was performed on the S-layer negative mutants. Eleven of the mutants contained Tn5 insertions in *rsaA* and were not further characterised. Five mutants, B5, B9, B13, B15 and B17, contained insertions in the DNA immediately 3' of *rsaA* and one mutant, B2, had a Tn5 integration elsewhere on the chromosome (Fig. 3-2). These six mutants represented possible RsaA translocator mutants.

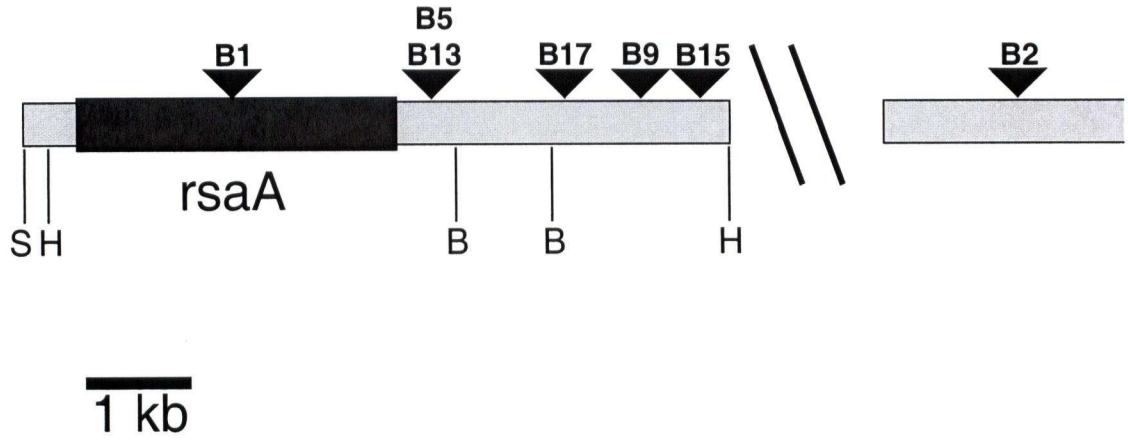

Figure 3-2. S-layer negative Tn5 insertions. Graphical representation of the positions of Tn5 insertions from mutants that no longer secreted RsaA. B = *Bam*HI, H = *Hind*III, S = *Sst*I. Triangles indicate Tn5 insertion points.

To determine whether the loss of S-layer was caused by a mutation affecting regulation of the gene, *rsaA* was expressed in the mutants under the control of a *lacZ* promoter, using the plasmid pRK415*rsaA*ΔPK. This construct restored RsaA production in JS1003 and B1, mutants with an interrupted *rsaA* gene, although wildtype RsaA expression levels were not reached. No S-layer was found on any of the five mutants with a Tn5 insertion in the DNA immediately 3' of *rsaA* secreted RsaA when *rsaA* was expressed *in trans* in this manner (Fig. 3-3).

In addition, the one mutant (B2) where the Tn5 insertion was not adjacent to the *rsaA* gene also produced an S-layer when complemented with the plasmid pRK415*rsaA*ΔPK. This indicates that the B2 insertion was not in a gene involved in RsaA secretion. B2 may have an interruption in a gene responsible for regulation of

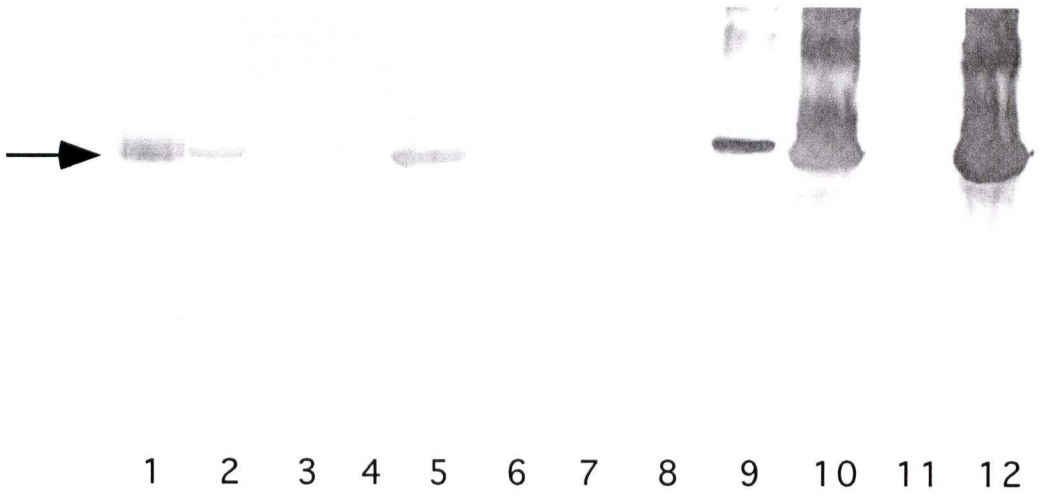

Figure 3-3. Complementation of Tn5 mutants with *rsaA*. Protein was extracted from the surface of the Tn5 mutants and JS1003 carrying the plasmid pRK415 *rsaA*ΔPK which expresses RsaA under control of the *lac* promoter and wildtype and *rsaA* knockout mutants that did not contain any plasmid to demonstrate differences in expression. Equal amounts of surface extracts were loaded on the gel and a Western performed using polyclonal antibody against RsaA. The lanes are as follows: Lanes 2 through 10 are surface extractions from cells containing the plasmid pRK415 *rsaA*ΔPK indicated by (ΔPK). 1, purified RsaA; 2, JS1003(ΔPK); 3, B9(ΔPK); 4, B13(ΔPK); 5, B1(ΔPK) (a Tn5 insertion in *rsaA*); 6, B5(ΔPK); 7, B15(ΔPK); 8, B17(ΔPK); 9, B2(ΔPK); 10, B12(ΔPK) (a random Tn5 insertion); 11, JS1003 (*rsaA*$^-$); 12, NA1000 (wildtype). The arrow indicaties wildtype RsaA.

RsaA production or, possibly, the Tn5 insertion mutation does not eliminate secretion and a second mutation in *rsaA* was responsible for the loss of secretion.

Isolation and sequencing of DNA near rsaA. A previously constructed cosmid library was used to isolate an 11.8 kb DNA fragment containing *rsaA* plus 7.3 kb of 3' DNA. This fragment was cloned into pBSKS+ forming the plasmid, pRAT1, and sequenced to search for translocator genes. An open-reading frame (ORF) was found 5' of *rsaA,* confirming earlier results (Fisher *et al.*, 1988) and 5 ORFs were found 3' of *rsaA* (Fig. 3-4).

A search of sequence databases showed that there were two ORFs immediately 3' of *rsaA* that encoded proteins with significant similarity to the ABC transporter and membrane fusion proteins (MFP) of two type I secretion systems: the alkaline protease transport system of *P. aeruginosa* (Guzzo *et al.*, 1990) and metalloprotease transport system of *E. chrysanthemi* (Létoffé and Wandersman, 1992) (Figs. 3-5, 3-6).

The first ORF was 1734 bp long and started 246 bp after the termination codon of *rsaA.* This ORF was predicted to code for a 578 amino acid protein with a predicted molecular weight of 62.0 kDa and pI of 9.02. Alignments of the predicted amino acid sequence show that the putative protein is 46% identical and 69% similar

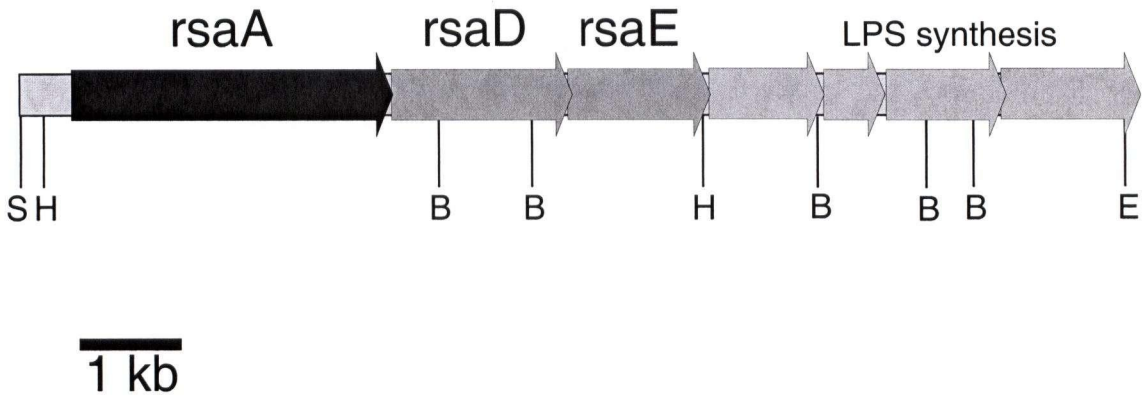

Figure 3-4. Genes 3' of *rsaA*. Graphic showing the ORFs found after sequencing the plasmid pRAT1. B = BamHI, E = EcoRI, H = HindIII, S = SstI.

to AprD from *P. aeruginosa* and 33% identical and 62% similar to PrtD from *E. chrysanthemi*.. The gene was designated *rsaD* because of this similarity (Fig. 3-5). RsaD exhibits several N-terminal hydrophobic domains that may be transmembrane regions and a possible ATP binding site in the C-terminal half of the protein. The predicted protein contains Walker A, Walker B, and ABC signature motifs as well as the newly discovered *E. coli* motif (hhhhH). These motifs are highlighted in Fig. 3-5.

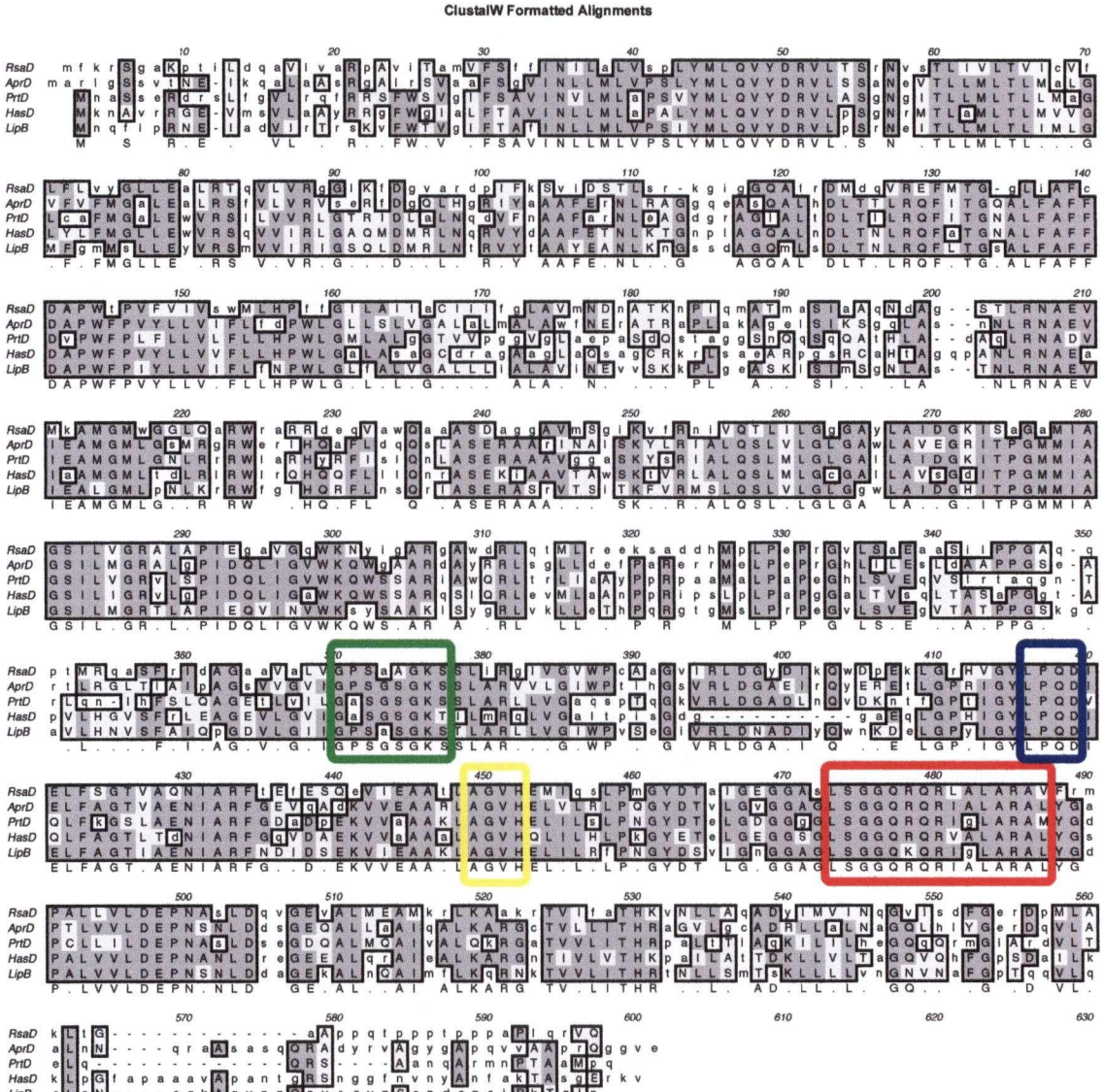

Figure 3–5. ClustalW alignment of ABC-transporters. Alignment of RsaD with AprD (Accession number CAA05795), PrtD (AAB03671), HasD (CAA57069) and LipB (BAA08631) which are the most closely related ABC transporters. The green box surrounds the Walker A motif, the blue box surrounds the Walker B motif, the red box surrounds the ABC motif and the yellow box surrounds the fourth ABC transporter motif recently discovered in most *E. coli* ABC transporters.

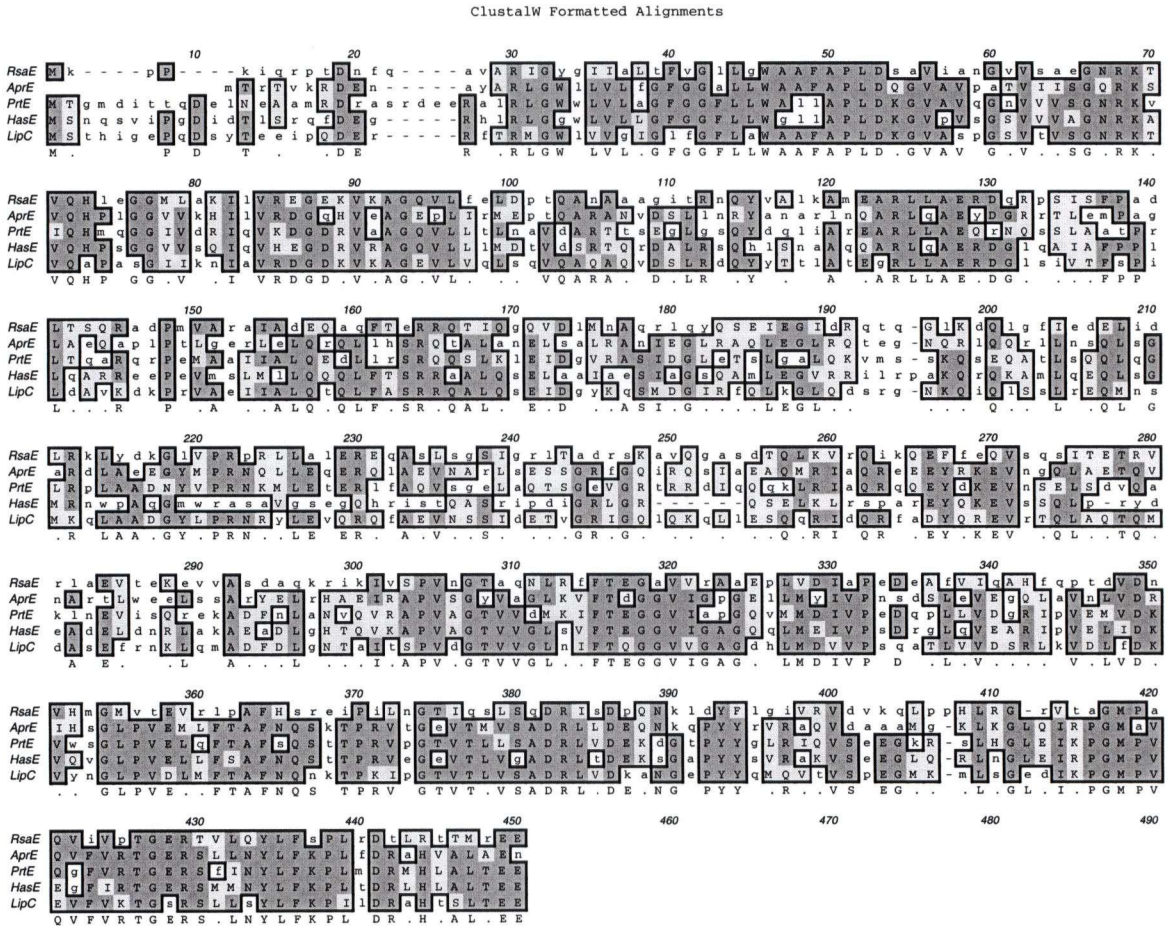

Figure 3-6. ClustalW alignment of MFPs. Alignment of RsaE with AprE (Accession number CAA45856), PrtE (CAA37343), HasE (CAA57067) and LipC (BAA08632) which are the most closely related ABC transporters.

RsaD was predicted to have a insertion signal sequence consistent with insertion of the RsaD protein in the cytoplasmic membrane.

The second ORF started 68 bp after *rsaD*, contained 1308 bp and encoded a protein of 436 residues with a predicted molecular weight of 48.4 kDa and pI of 6.59. Alignment of the predicted protein shows that the sequence is 28% identical and 50% similar to AprE from *P. aeruginosa* and 29% identical and 52% identical to PrtE from *E. chrysanthemi*. The gene was designated *rsaE* because of this similarity (Fig. 3-6). The deduced protein sequence of *rsaE* was predicted to have a typical N-terminal insertion signal sequence that would direct it to the inner membrane.

Possible ribosome binding sites were found 7 bp and 8 bp upstream of the ATG initiation codon for *rsaD* and *rsaE*, respectively. There was no indication of a promoter immediately 5' of either *rsaD* or *rsaE*, but there was a putative rho-independent terminator immediately after the stop codon of *rsaE* suggesting that they may be part of a polycistron which includes *rsaA*. It has been found in the type I secretion systems secreting *E. coli* α -hemolysin and *E. chrysanthemi* metalloprotease that the genes are part of an operon consisting of the substrate and the transport genes. It seems likely that transcription of the Rsa genes is similar.

Three more ORFs were found 3' of *rsaE*. None of these ORFs encoded proteins similar to the third component of type I secretion systems. Instead, these ORFs encoded proteins similar to those involved in synthesis of perosamine, a dideoxyaminohexose (see Ch. 6).

The chromosomal DNA near B1, B2, B5, B9, B13, B15 and B17 Tn5 insertions was isolated and sequenced to determine the Tn5 insertion point. It was found that the B1 Tn5 interrupts *rsaA*, as expected from the Sourthern blot analysis. B5 and B13 are identical insertions interrupting the N-terminus of RsaD while B17 is located 22 amino acids from the C-terminus. B9 and B13 are Tn5 insertions in *rsaE*. The sequence interrupted by the B2 Tn5 insertion has no sequence similarity to any known proteins.

Complementation of the secretion-defective Tn5 mutants. To demonstrate that the Tn5 insertions were responsible for the secretion defect the mutations were complemented *in trans*. First, the cosmid, 17A7, containing the entire Rsa locus, was introduced into the mutants. All attempts at complementation using this cosmid were unsuccessful, including an attempt to restore RsaA production in JS1003 (which contains an inactivated *rsaA* gene). Since RsaA production in JS1003 can be restored with other plasmids containing *rsaA*, it is believed that expression of the genes was too low for complementation.

A PCR product containing the genes *rsaD* and *rsaE* was generated and cloned into a suitable expression vector; the result was named pRAT5:PRK415 (see Ch. 2). This plasmid was introduced into the Tn5 mutants B15 and B17. With this plasmid, mutant B17 secreted RsaA while the B15 mutant did not (Fig. 3-7A).

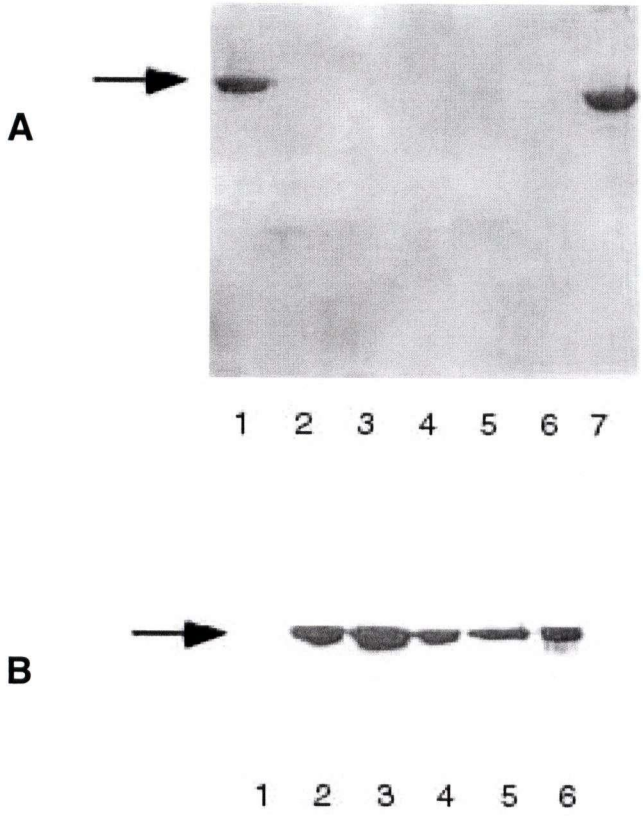

Figure 3-7. Complementation of transport deficient mutants using *rsaD* and *rsaE*.
Westerns of surface extracted protein using anti-S antibody. A) Lanes are as follows: 1, B17 (DE); 2, B15 (DE);3, B1(DE); 4, B17 (17A7); 5, B15 (17A7); 6, JS1003; 7, NA1000. (DE) indicates that the cells carried the plasmid pRAT5:pRK415 containing the genes *rsaD* and *rsaE*. (17A7) indicates that the cells carry the cosmid 17A7 containing the entire RSA operon. Equal amounts of surface extract were loaded in all lanes. The arrow indicates full length RsaA. B) Lanes are as follows: 1, B1 (DE); 2, B5 (DE); 3, B9 (DE); 4, B15 (DE); 5, B17 (DE); 6; NA1000. DE indicates that the cells carry the plasmid pRAT5:pBBR5 expressing the genes *rsaD* and *rsaE*. Equal amounts of surface extract were loaded in all lanes except (6) where there was only one quarter of the amount loaded in the other lanes. The arrow indicates full length RsaA.

To address the problems with B15 complementation, a new tetracycline-resistant (Tc^r) broad host range vector, pBBR5, was constructed. It was hoped that this vector would have a higher copy number and expression of the Rsa genes that would alleviate the problems encountered when using pRK415 or pLAFR5 (the cosmid vector). In the resulting constructs a *lac* promoter is used for transcription of the *rsaD* and *rsaE* genes in pRAT5: PBBR5 and the *rsaA*, *rsaD* and *rsaE* in pRAT4ΔH: PBBR5. When pRAT5: PBBR5 was introduced into the mutants B1, B5, B9, B15 and B17, Western blot analysis showed that the mutants with defective *rsaD* or *rsaE* genes expressed RsaA on the surface while the *rsaA* mutant B1 did not (Fig. 3-7B). When pRAT4ΔH:pBBR5 was expressed in the same mutants, RsaA was only found on the surface of the B1 and B17 mutants (data not shown). The ability to complement the

Tn5 insertions in *rsaD* and *rsaE* using pRAT5:pBBR5 expressing *rsaD* and *rsaE in trans* indicates that these genes are responsible for the secretion of RsaA.

The lack of complementation in some cases was probably the result of lower expression of the Rsa genes. It was necessary to use Tc to maintain the vectors as Tn5 confers kanamycin and streptomycin resistance, but *C. crescentus* does not tolerate Tc well. When cells carry the Tc resistance marker are exposed to even low levels of Tc (0.5 μg/ml), they appear anomalous by microscopy. The cells are often severely elongated and there are few motile cells. It was difficult to grow cultures carrying Tc^r plasmids with the Rsa genes to densities high enough to extract sufficient protein to be seen on the Western blot. It seems probable that the Tc was causing membrane abnormalities and that these factors contributed to lower expression of the Rsa genes with all the plasmids.

The cosmid, 17A7, only has 1-2 copies per cell and similarly, pRAT5:pRK415 would be maintained at 2-3 plasmids per cell (Keen *et al*, 1988). Preliminary experiments with pBBR5 suggest that it has a much higher copy number than either pLAFR5 or pRK415 based vectors which would result in higher expression of any genes that pBBR5 carries (data not shown).

Expression levels would also be affected by the promoter transcribing the genes. The *lac* promoter transcribes at higher levels than the wildtype *rsaA* promoter (Yap *et al.*, 1994). In addition, in the cosmid and pRAT4ΔH:pBBR5, *rsaD* and *rsaE* are either transcribed by their wildtype promoter or as part of the *rsaA* transcript as described above. In either case, a lesser amount of transcript would be produced than from the lacZ promoter of pRAT5:pBBR5.

These data suggest why the complementation occurred only in some cases. The plasmid pRAT5:pBBR5 (strong promoter and high copy number) produced the highest levels of RsaD and RsaE allowing full complementation of all the transport mutants while the cosmid, 17A7, (weaker promoter and low copy number) produced the lowest levels and could not complement any of the mutants. The plasmids pRAT5:pRK415 (strong promoter and low copy number) and pRAT4ΔH:pBBR5 (weak promoter and high copy number) probably make an intermediate amount of protein that is only enough to complement the mutant B17. This mutant may differ from the others because the Tn5 insertion is only 22 amino acids from the C-

interrupted (Fig. 3-8). Smaller zones of clearing are seen around the wildtype strain, NA1000, and the S-layer producing B12 (representing a random Tn5 insertion unrelated to secretion), as compared to JS1003 or B1, where the *rsaA* gene has been interrupted, suggesting that there was competition between RsaA and PrtB for the secretion machinery, further supporting the supposition that RsaD and RsaE are parts of a type I secretion mechanism. Identical results were found when *aprA* was expressed in the Tn5 mutants (data not shown).

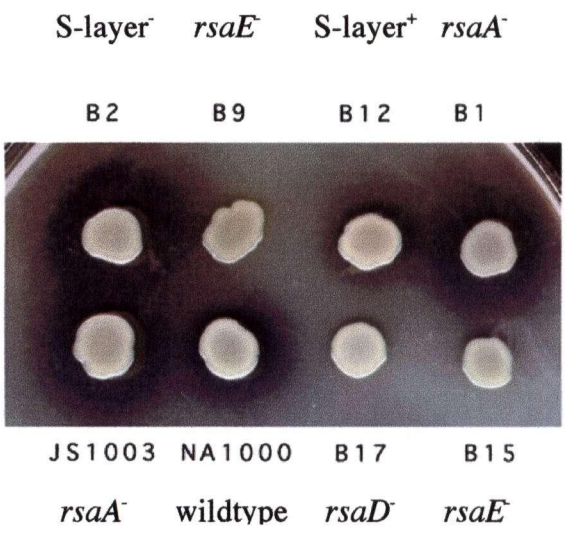

Figure 3-8. Expression of *prtB* in *C. crescentus*. PrtB was expressed in all the colonies shown using the plasmid pRK415: pRUW500. The cells were spotted on to PYE plates containing 1% skim milk. Halos around colonies indicate that active PrtB is being secreted. Note that NA1000 and B12 cells are producing RsaA as well as PrtB and the halos surrounding these colonies are smaller. B12 represents a random Tn5 mutant control.

Summary

Analysis of the region 3' of *rsaA* revealed the presence of two genes (*rsaD* and *rsaE*) encoding proteins with significant sequence similarity to components of the type I secretion systems used by *P. aeruginosa* and *E. chrysanthemi* to secrete two different extracellular proteases (Duong *et al.*, 1992; Wandersman *et al.*, 1990). Because interruption of *rsaD* and *rsaE* eliminated secretion of RsaA and the defects could be restored by complementation, it was apparent that their gene products make up part of the RsaA translocator machinery.

When these results were reported (Awram and Smit, 1998), it was the first example of an S-layer that is secreted using a type I secretion system. Before then, S-layers had only been found to be secreted by a type II system (Messner and Sleytr, 1992; Sleytr *et al.*, 1993). It is now known that a protein with amino acid

unrelated to secretion), as compared to JS1003 or B1, where the *rsaA* gene has been interrupted, suggesting that there was competition between RsaA and PrtB for the secretion machinery, further supporting the supposition that RsaD and RsaE are parts of a type I secretion mechanism. Identical results were found when *aprA* was expressed in the Tn5 mutants (data not shown).

Summary

Analysis of the region 3' of *rsaA* revealed the presence of two genes (*rsaD* and *rsaE*) encoding proteins with significant sequence similarity to components of the type I secretion systems used by *P. aeruginosa* and *E. chrysanthemi* to secrete two different extracellular proteases (Duong *et al.*, 1992; Wandersman *et al.*, 1990). Because interruption of *rsaD* and *rsaE* eliminated secretion of RsaA and the defects could be restored by complementation, it was apparent that their gene products make up part of the RsaA translocator machinery.

When these results were reported (Awram and Smit, 1998), it was the first example of an S-layer that is secreted using a type I secretion system. Before then, S-layers had only been found to be secreted by a type II system (Messner and Sleytr, 1992; Sleytr *et al.*, 1993). It is now known that a protein with amino acid sequence similarity to RsaA is secreted by the *S. marcescens* type I secretion system (Kawai *et al.*, 1998). In addition, the *C. fetus* S-layer protein is secreted by a type I secretion mechanism. The *C. fetus* S-layer shares several features in common with that of *C. crescentus.* It is produced by a free-living Gram-negative bacterium, is hexagonally-packed, anchors to the cell surface via its N-terminus to a particular species of LPS (Bingle *et al.*, 1997b; Dworkin *et al.*, 1995; Walker *et al.*, 1992) and so far has the greatest similarity of any S-layer protein to RsaA (Gilchrist *et al.*, 1992).

The genes for the ABC transporter and the MFP components of type I secretion systems are generally found in an operon that includes the transported protein (Binet *et al.*, 1997; Salmond and Reeves, 1993). In this respect then, the organization of the *rsaA*, *rsaD* and *rsaE* genes was not surprising. In contrast, the gene encoding the outer membrane protein component of type I secretion systems may or may not be closely linked to the other secretion genes. The third component of the Rsa transporter has now been found 5 kb 3' of *rsaE* and is described in Ch. 4.

A potential Rho-independent terminator sequence is located after the *rsaA* coding region (Gilchrist *et al.*, 1992). This predicted terminator results in a predicted transcript that matched closely the size of a transcript found using Northern blot analysis (Fisher *et al.*, 1988). In this study, no obvious indications of a promoter were found immediately 5' of either the *rsaD* or *rsaE* genes suggesting that transcription of *rsaD* and *rsaE* is similar to transcription of the *hlyA*, *hlyB* and *hlyD* genes of *E. coli*, where a similar Rho-independent terminator is found after the *hlyA* gene and terminates most transcripts at this point. An anti-terminator, RfaH, prevents termination and when it does, a larger transcript including the *hlyB* and *hlyD* genes is made (Leeds and Welch, 1996). This transcript is difficult to detect because it has a short half-life and an analogous transcript in *C. crescentus* may have been missed in the northern blot analysis. Transcription of the *E. chrysanthemi* protease secretion genes appears to be accomplished by a similar method (Létoffé *et al.*, 1990) and it is postulated that the same is true for the Rsa operon. A transcription pattern like this may account for the reduced expression found in the JS1003 and B1 mutants when they are complemented with *rsaA*. The kanamycin fragment interrupting *rsaA* in JS1003 does not have a transcription terminator and transcription may continue through to the end of *rsaE*, resulting in a transcript 1.5 kb longer than the wildtype, which would likely be more unstable and result in fewer transport complexes. In B1, it is likely that *rsaD* and *rsaE* are transcribed off one of the Tn5 promoters resulting in decreased amounts of transcript and, in turn, transport complexes.

Type I secretion systems can be grouped into families. The RTX toxins, such as α-hemolysin (*E. coli*) and leukotoxin (*P. hemolytica*), comprise one family while extracellular proteases (e.g. AprA, PrtB) and lipase from *S. marcescens* constitute

another (Binet *et al.*, 1997). Within the families there is high sequence similarity and functional secretion mechanisms can be constructed from using components from the different members without a dramatic drop in protein transport. Because it has been demonstrated that AprA and PrtB proteins can be secreted from *C. crescentus* in active form and there is higher sequence similarity between these proteins than with RTX toxins, presumably, RsaA can be grouped with the protease family of type I secretion systems.

Chapter 4

Identification of the Outer Membrane Protein Component of the RsaA Transport Complex

Introduction

The gene encoding the OMP component of the RsaA secretion machinery proved difficult to isolate since it was not found immediately 3' of the MFP, as in many other type I systems. This difficulty has also been found with most of the other type I secretion systems where the OMP is separated from the rest of the transporter complex. In fact, the OMP has only been found in 2 other cases of this type: TolC, required for transport of α-hemolysin in *E. coli* (Wandersman and Delepelaire, 1990) and HasF, part of the heme transporter in *S. marcescens* (Binet and Wandersman, 1996). In both of these cases the experimenters had simple, efficient screens to look for mutants.

Several different strategies were considered to find the OMP component. As none of the original S-layer negative Tn5 mutants interrupted the OMP and considering the number of mutants screened, it was believed that the NA1000 Tn5 library did not contain the mutant. The Tn5 library may not have been complete or a Tn5 insertion in the OMP may have been lethal. If a Tn5 insertion was lethal there was no further point in screening another Tn5 library. It seemed possible that a point mutant with reduced secretion, but not having a lethal phenotype could be constructed. Since a UV/NTG point mutant library had been previously made by others, it was decided that this library could be screened for an OMP mutant.

Alternatively, a functional type I system could be reconstructed as was done in *E. coli* using *hasDE,* the ABC-transporter and MFP genes, from *S. marcescens* and the OMP gene, *tolC* (Binet and Wandersman, 1996). This secretion apparatus was capable of secreting the *S. marcescens* heme-acquisition protein, HasA, as well as AprA and PrtB. The *S. marcescens* OMP gene, *hasF*, was then isolated by expressing a protease along with *hasDE* in an *E. coli tolC*$^-$ mutant along with a plasmid library of *S. marcescens* chromosomal DNA, and screening for the presence

of protease secretion on skim milk plates. It was hoped that a similar method would be capable of identifying the Rsa OMP gene.

A third option for finding the OMP was to screen by similarity to OMP components from other bacteria. There are two ways to approach this. One method is to search the genome of *C. crescentus* for DNA fragments hybridizing to the genes from OMP components. The other is to compare the sequences of different OMP components to find regions of similarity and design primers with degenerate sequences for PCR amplification of a portion of the OMP DNA sequence that can be used to isolate the complete gene by hybridization.

All of these approaches were attempted and are summarized below, but none worked. The OMP gene was eventually found using the partial *C. crescentus* genome sequence provided by The Institute for Genome Research (TIGR). Two partial ORFs with similarity to OMP components from other bacteria were found in this sequence data and this information was used to devise strategies to clone the complete sequence and to test which of the two ORFs was a legitimate OMP gene involved in the secretion of RsaA.

Results and Discussion

Screening libraries for OMP mutants defective in secretion. Since the original immunoblot assay was very labour intensive, attempts were made to develop a new screening method for finding secretion deficient mutants. The proteases, AprA and PrtB, are secreted by type I transporters and can be secreted by the Rsa secretion machinery, allowing skim milk plates to be used for rapid screening. Therefore, vectors carrying these genes were designed for screening the libraries. The plasmid pBBR3AprA:pRAT5 was constructed and consists of the *aprA* gene and the *rsaDE* genes under the control of separate *lacZ* promoters. The plasmid pBBR3PrtB:pRAT5 is identical to pBBR3AprA:pRAT5 except the *aprA* gene is replaced with *prtB*. When these plasmids were introduced into the UV/NTG mutant library, no secretion of AprA or PrtB was observed. The *rsaDE* genes had originally been included in the plasmid to exclude *rsaDE* mutants from being found during the

screening process, but since the plasmid did not work the approach was dropped. When the plasmids pBBR3AprA and pBBR3PrtB were used to express their respective proteases in the UV/NTG mutant library a large number of colonies failed to show secretion of the proteases. When some of these colonies were examined, it was found that they were still capable of secretion of RsaA. This was an unexpected result as expression of the proteases in NA1000 results in protease secretion from >99.9% of colonies. It was concluded that these proteases are not tolerated well by *C. crescentus* and could not be used as a screen. In agreement with this was the observation that *C. crescentus* colonies expressing the proteases could not be sub-cultured after growing for 5 days while normally *C. crescentus* can be sub-cultured even after several weeks. It appeared that the proteases were killing the bacteria. (see Ch. 5 for further discussion about protease expression in *Caulobacter* species).

Without a rapid screening method, it was decided to drop screening of mutant libraries in favour of the other approaches.

Searching for the OMP using complementation systems. If a complementation system was going to succeed in finding the OMP component, it was necessary to determine if a functional system could be constructed using the *C. crescentus* transporter components. In many other type I systems the components can be interchanged with components from other bacterial systems and allow heterologous secretion. To determine if the Rsa system would work in a similar manner plasmids expressing RsaD and RsaE were expressed in bacterial hosts along with OMP components from several different bacterial systems.

The plasmids pBBR3AprA:pRAT5, pBBR3PrtB:pRAT5 and pRAT4ΔH were constructed and express either a protease or *rsaA* along with *rsaD* and *rsaE*. These plasmids were introduced into *E. coli tolC*$^{+}$ alone or with either of the plasmids pBBR1AprF and pBBR1PrtF which express OMP components from the Apr and Prt systems. None of these strains secreted either the protease or RsaA (data not shown). Since *E. coli* is an enteric microorganism and *C. crescentus* is a free-living groundwater bacterium, their outer membranes are quite different. It is possible that the Rsa transport complex was unable to assemble in the membrane of *E. coli*.

Rhizobium meliloti and *Rhizobium leguminosarum* are ground water bacteria living in environments similar to *C. crescentus* and likely have a membrane resembling that if *C. crescentus*. In addition, the type I secretion systems, Nod and Prs, with similarity to the Rsa secretion machinery have been found in *R. leguminosarum* (Finnie *et al.*, 1998; Scheu *et al.*, 1992). In *R. leguminosarum*, as in the Rsa system, the OMP gene of the Prs secretion system has not been found close to the other transport genes and is expected to be elsewhere on the chromosome and could possibly complement the Rsa machinery. With this in mind, pBBR3AprA:pRAT5, pBBR3PrtB:pRAT5 and pRAT4ΔH were expressed in *R. meliloti* and *R. leguminosarum.* Again, none of the constructs expressed the proteases or RsaA. Further experiments were tried by introducing pBBR3AprA:pRAT5, pBBR3PrtB:pRAT5 and pRAT4ΔH along with pBBR1AprF and pBBR1PrtF, in various combinations in the *Rhizobium* species. In no case was secretion of RsaA or the protease found (data not shown).

Sequence similarity to other OMP genes was used to search for the Rsa OMP gene. Southern blots of *C. crescentus* chromosomal DNA were probed with the OMP genes, *aprF* and *prtF* under conditions allowing 30% mismatch. No hybridization of these probes to *C. crescentus* DNA was found (data not shown) demonstrating that this method could not be used.

A sequence alignment of OMP components revealed areas of sequence identity among the different proteins. The protein sequences of the OMPs from a number of closely related type I transport systems (with OMP genes that are both linked and unlinked to the other transporter genes) were aligned (Fig 4-1). The OMP, HasF, was given the highest priority in the comparison because it is from the type I system with an unlinked OMP gene most closely related to the Rsa system. Areas of significant homology were examined for the purpose of designing degenerate primers to amplify a portion of the OMP gene using PCR. Four areas, shown in Fig 4-1, were chosen for making primers. The primers were designed by taking the consensus amino acid sequence and using the codon preferences of *C.*

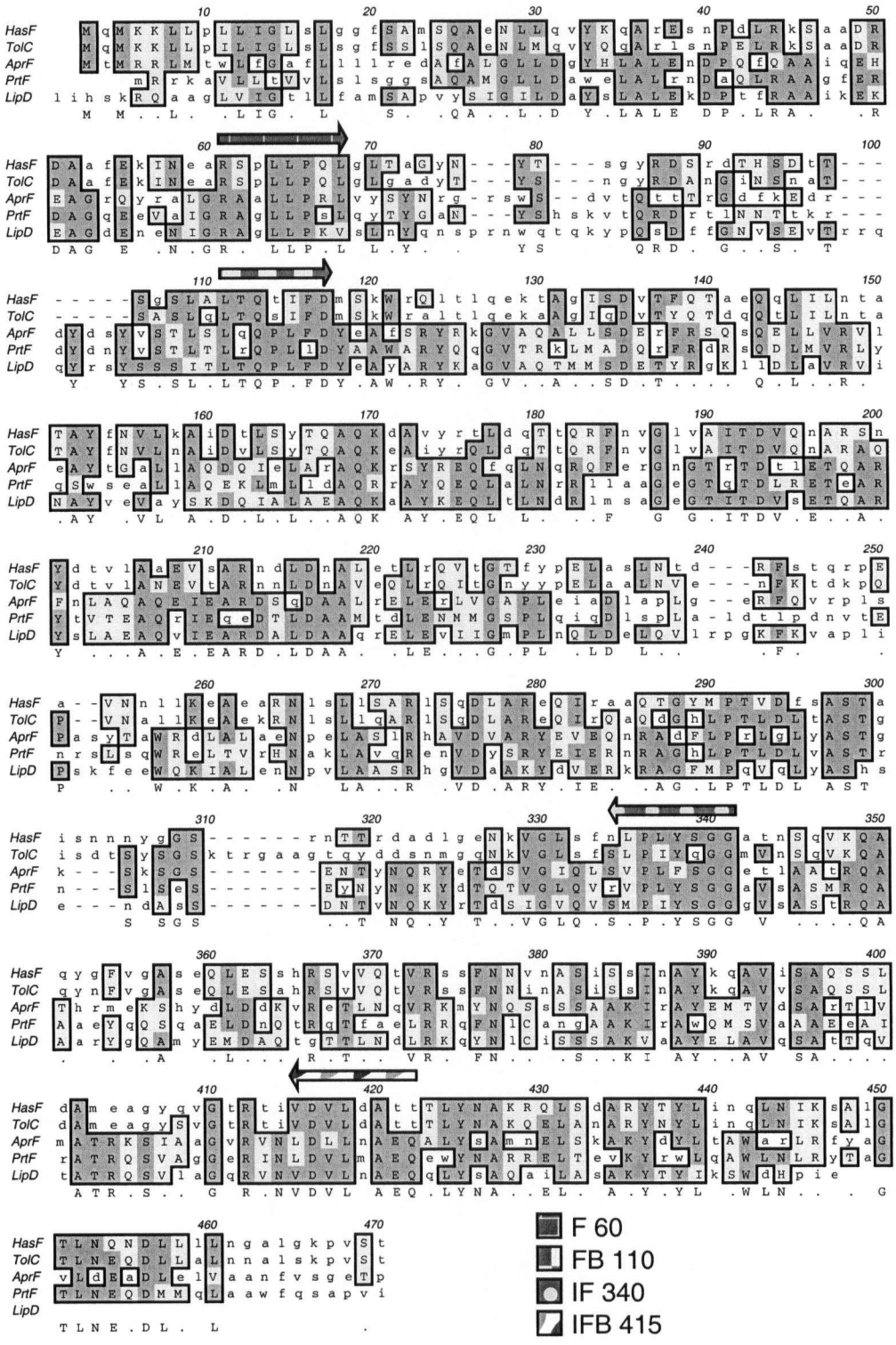

Figure 4-1. Alignment of OMP components. Arrows are placed above regions of similarity that were used to design degenerate primers. The arrows are colour coded according the primer they were used to create (see legend)

crescentus to determine the DNA sequence. The design process was governed by the suggestions in Colnaghi *et al.*, 1996; Maser and Kaminsky, 1998; and Tobin *et al.*, 1997. A variety of conditions, as well as different combinations of the primers, were used to amplify fragments from NA1000 chromosomal DNA (see Ch. 2). When the PCR conditions resulted in a product, multiple bands were always seen. Three DNA fragments of the expected size were gel purified and cloned. Sequencing of these products revealed similarity to 23S RNA, poly (3-hydroxybutyrate) biosynthesis genes and NADH dehydrogenase genes. The primers appeared to be amplifying undesired DNA sequences and as a result these experiments were abandoned.

Two candidates for the Rsa OMP gene were identified in the preliminary *Caulobacter* genome data. As all other attempts had failed to identify the OMP gene, contact was made with The Institute for Genome Research (TIGR) who provided preliminary sequence data from the *Caulobacter* genome. FASTA searches (Pearson *et al.*, 1997) of this database produced two contigs with similarity to known OMP components. Contig gcc_973 contains an ORF coding for the first 225 amino acids of a possible OMP component with a G+C content of 65.3%. Examination of the DNA 5' of this ORF revealed that this ORF is 5 kb 3' of the *rsaE* gene and there are 5 intervening ORFs that likely code for S-LPS synthesis proteins (Fig. 4-2). This ORF has been designated *rsaF*(973). The deduced amino acid

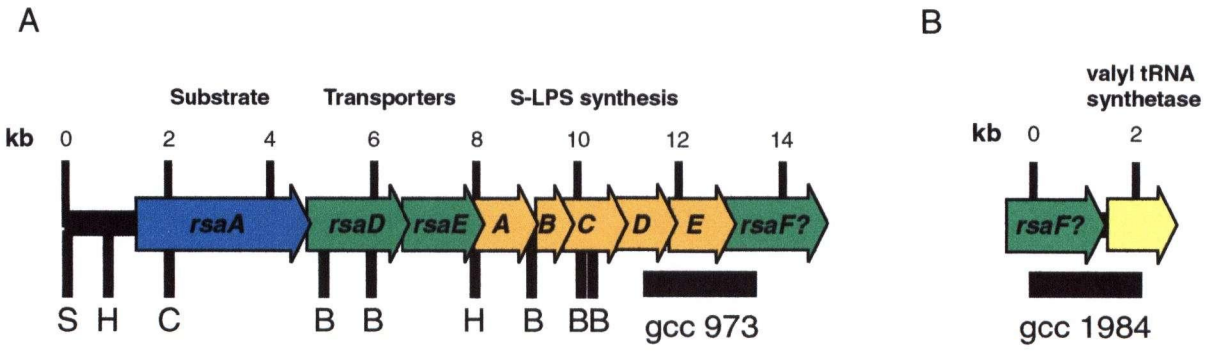

Figure 4-2. The two possible OMPs, *rsaF*(973) and *rsaF*(1984). A) The figure shows that *rsaF*(973) is located 5 kb downstream of *rsaE*. B) *rsaF*(1984) is located adjacent to a gene coding for valyl tRNA synthetase. The location of the gcc contigs is shown with black bars. B-*Bam*HI, C-*ClaI*, S-*SstI*,

sequence of *rsaF*(973) had greatest similarity to TolC with 26.1% identity and 52.2% similarity over the 225 amino acids coded by gcc_973 (Fig. 4-3). Contig gcc_1984 has a G+C content of 67% and contains an ORF coding for the last 384 amino acids of a possible OMP. This ORF has been designated *rsaF*(1984). 3' of *rsaF*(1984) is an ORF coding for valyl tRNA synthetase (Fig 4-2). The coding sequence of *rsaF*(1984) had greatest similarity to the HasF OMP with 26.8% identity and 48.5% similarity (Fig. 4-3). The G+C content of these two ORFs is comparable to *C. crescentus's* 67%, suggesting that neither is a recent genetic acquisition. These two contigs overlap with 59.6% identity over a region of 344 bp indicating that they are not part of the same ORF, but suggest that one arose by gene duplication of the other (Fig. 4-3).

Once sequence was available it was assumed that it would be relatively simple to obtain both complete genes. This did not prove to be the case. Using these sequences, primers were designed to amplify portions of *rsaF*(973) and *rsaF*(1984) that could then be used as probes to isolate the complete genes. These primers had melting temperatures (T_m) between 58°C and 62°C and did not appear to have any hairpin loops or secondary priming sites when analyzed using primer analysis and design programs. Primers of this size and T_m have been used routinely for PCR amplification of *C. crescentus* DNA with excellent results. These primers produced products of the expected size, but when cloned and sequenced the products were identical to the *C. crescentus* DNA gyrase and glutamate permease genes.

Suspecting that there may be something peculiar about the structure of the DNA around the *rsaF* genes it was decided to attempt to isolate the DNA of the adjacent regions. Since the start of *rsaF*(973) is found in the genome 1.5 kb 3' of sequences cloned into pRAT1, a 2 kb *Bam*HI-*Eco*RI fragment was sub-cloned from pRAT1 and designated pRAT HI (B/E). To amplify a fragment of DNA close to the *rsaF*(1984) gene, new primers were made to amplify a 736 bp region 3' of *rsaF*(1984). These primers were designed with T_m of 70°C and were 26-28 bp long.

A. **BlastX comparison of gcc_973**

	Sequences producing High-scoring Segment Pairs:	High Score	Smallest Probability P(N)
1.	gi|72556 outer membrane protein tolC E.coli	92	4.0e-11
2.	gi|3080540 (D49826) LipD [Serratia marcescens]	115	7.4e-07
3.	gi|4826418 (Y19002) PrtF protein [Erwinia amylovora]	115	1.0e-06
4.	gi|281563 agglutination protein - Pseudomonas putida	61	3.4e-05

B. **BlastX comparison of gcc_1984**

	Sequences producing High-scoring Segment Pairs:	High Score	Smallest Probability P(N)
1.	gi|1405817 (X98513) HasF ABC exporter outer membrane .	154	1.0e-23
2.	gi|135980 OUTER MEMBRANE PROTEIN TOLC PRECURSOR E.coli	159	1.2e-23
3.	gi|3080540 (D49826) LipD [Serratia marcescens]	126	8.3e-23
4.	gi|4826418 (Y19002) PrtF protein [Erwinia amylovora]	111	4.2e-21

C. Overlap of gcc_973 and gcc_1984.

```
gcc_973  CAGACCTCGACCCTCTCTCTGAGCCAGAGCCTCTACACCAACGGTCGTTTCTCGGCCCGC
                                        :::::::::  ::::::   :   ::  ::
gcc_1984                               CGCTCTACACCGGCGGTCGCGCCAGCGCGGGC

gcc_973  CTGGCGGGTGTCGAGGCGCAGATCAAGGCCGCGCGCGAGAACCTGCGCCGCATCGAGATG
          :        :  :: ::  :  :   : : ::::: ::    :: :::    :::::  :
gcc_1984 GTCAGCCCCGCTGAAGCCGACGTGCTGTCTGCGCGGGAAGGTCTTCGCGCGGTCGAGCAG

gcc_973  GACCTGCTGGTCCGCGTGACCAACGCCTATATCTCGGTGCGCCGCGACCGCGAGATCCTG
         :   :::::::: :::::   : : :::::: ::   :::::::: :::::  ::   ::::
gcc_1984 GGGGTGCTGGTCAGCGTCGTCCAGGCCTATGTCGACGTGCGCCGAGACCAGGAACGCCTG

gcc_973  CGGATCAGCCAAGG-CGGTGAAGCCTGGCTGCAGAAGCAATTGAAGGACACCGAGGACAA
         :: ::: :::::::     :  ::    ::::::   ::  :  : ::  :  : :
gcc_1984 CGCATC-GCCAAGGAAAACGTCGCGGTTCTGCAGCGCCAGCTCGAAGAATCGAACGCTCG

gcc_973  GTACAGCGTCCGTCAGGTGACCTTGACCGACGTGCAGCAGGCCAAGGCCCGCCTGGCGTC
          : :  :::  :: :: : :::  ::: ::::::    ::: :  :::: ::: :::: ::
gcc_1984 CTTCGACGTGGGTGAGATCACCCGGACGGACGTCGCCCAGTCTCAGGCGCGCTTGGCTTC

gcc_973  GGCCAGCACTCAGGTGGCGAACGCCCAGGCGCAGCTGAATGTCAGCGTAGCGTTCTACGC
         :::::   :    :: ::  ::::::::: ::: :: : ::::::   :: : ::::::
gcc_1984 GGCCAAGGCCAGCCTGTCGGGCGCCCAGGCCCAGTTGGAAGTCAGCCGCGCCTCCTACGC

gcc_973  GTCCCTGGTGGGGCGCCAGCCGGAGAC
           :  :::: :: :
gcc_1984 TGCGGTGGTCGGTCAAACGCCCGGCGAACTGGCTCCCGAGCCGAGCTTGGCCGGACTGCT
```

Figure 4-3. Comparison of possible Rsa OMP components. A) Closest similar proteins to the ORF from gcc_973. B) Closest similar proteins to the ORF from gcc_1984. C) comparison of gcc_973 to gcc_1984. Note that the P(N) numbers are higher for gcc_1984 than gcc_973 because the gcc_1984 contig has a larger portion of the ORF.

PCR using these primers produced a product of the expected size that was successfully cloned and the resulting plasmid was called pBSKS-gcc1984. When sequenced, the product proved to be the correct fragment.

The NA1000 cosmid library was probed with pRAT HI (B/E) and pBSKS-gcc1984. A number of cosmids hybridized to pRAT HI (B/E), but all proved to contain only DNA 5' of *rsaF*(973) and it was concluded that *rsaF*(973) was not located within the NA1000 cosmid library. The cosmid, 7A22, hybridized to pBSKS-gcc1984. Southern blots of the cosmid showed that pBSKS-gcc1984 hybridized to a 5.5 kb *Bam*HI band. Several attempts were made to subclone this fragment and while the surrounding fragments could be cloned, it was not possible to subclone the fragment containing *rsaF*(1984).

Yet another approach was taken to isolate the *rsaF* genes. The plasmids pRAT HI (B/E) and pBSKS-gcc1984 will not replicate in *C. crescentus* and could be forced to integrate into the genome by homologous recombination. The plasmid pBSKS-gcc1984 was not successfully integrated into the chromosome, but pRAT HI (B/E) was, giving NA1000::pRAT HI (B/E). Chromosomal DNA from NA1000::pRAT HI (B/E) was partially digested with *Bam*HI and ligated under conditions promoting the circularization of the DNA fragments. The ligation mix was electroporated into *E. coli* and plated on selective medium which allowed only the growth of cells carrying the plasmid pRAT HI (B/E) and chromosomal DNA adjacent to the integration points that had circularized during the ligation. The 14 kb plasmid, pTZ19UΔSSm973Bcirc, was isolated in this manner. Restriction mapping and Southern blotting of this plasmid showed that insert consisted of DNA from 2.5 kb of 5' to 5.5 kb 3' of *rsaF*(973). Fragments of this plasmid were sub-cloned and sequenced, including a fragment containing the N-terminal of RsaF(973), but it proved impossible to subclone and sequence the entire *rsaF*(973) from this plasmid. This is not the first example of DNA from *C. crescentus* that has proved impossible to subclone. A 6.6 kb fragment, containing the holdfast genes involved in *C. crescentus* attachment, has proven resistant to the subcloning efforts of several graduate students and postdoctoral fellows (Smit, unpublished).

Fortuitously, one of the shedder Tn5 mutants, F11 (see Ch. 6), contains a Tn5 insertion 400 bp 5' of the *rsaF*(973) ORF. Using primers that hybridize to the

Tn5 it was possible to use an inverse PCR method (Martin and Mohn, 1999) to isolate and clone two fragments of DNA containing *rsaF*(973). Plasmid pCR2.1F11SalI contains the DNA from the F11 Tn5 insertion to the *Sal*I site 1.1 kb 3' of *rsaF*(973). The other, pCR2.1F11XmaI, contains the DNA from the F11 Tn5 insertion to the *Xma*I site 2.0 kb 3' of *rsaF*(973). Again, both of these clones proved difficult to isolate. Large amounts of PCR product were obtained from the PCR reaction, but cloning of these fragments only produced one clone of pCR2.1F11SalI and two clones of pCR2.1F11XmaI. Usually when cloning products in this manner a minimum of 50 clones and as many as 300 clones can be expected. *E. coli* carrying these plasmids grow slowly and appear distended and malformed when observed by phase contrast light microscopy. It is possible that the inserts in these plasmids are not identical to wildtype NA1000 chromosomal DNA sequences, but contain mutations generated by inaccuracies in the Taq polymerase amplification. It may be that the majority of PCR product is lethal when introduced into *E. coli*, but some of the PCR product containing mutations in *rsaF*(973) making the product less toxic could be cloned in *E. coli*. The sequence of the insert from pCR2.1F11SalI assembled together with sequence from the plasmid pTZ19UΔSSm973Bcirc and the TIGR genome (Fig. 4-4, Appendix I). The RsaF (973) sequence from pCR2.1F11SalI, showed considerable similarity to other OMPs. The highest degree of sequence similarity was to *E. coli* TolC with 25.2% identity and 48.6% similar amino acids. The OMPs AprF and PrtF from *P. aeruginosa* and *E. chrysanthemi* were not as similar (Fig. 4-5). Analysis of the sequence of RsaF(973) revealed the presence of a predicted signal sequence encompassing the first 32 amino acids and the presence of β-strands capable of forming a β-barrel structure typical of outer membrane proteins.

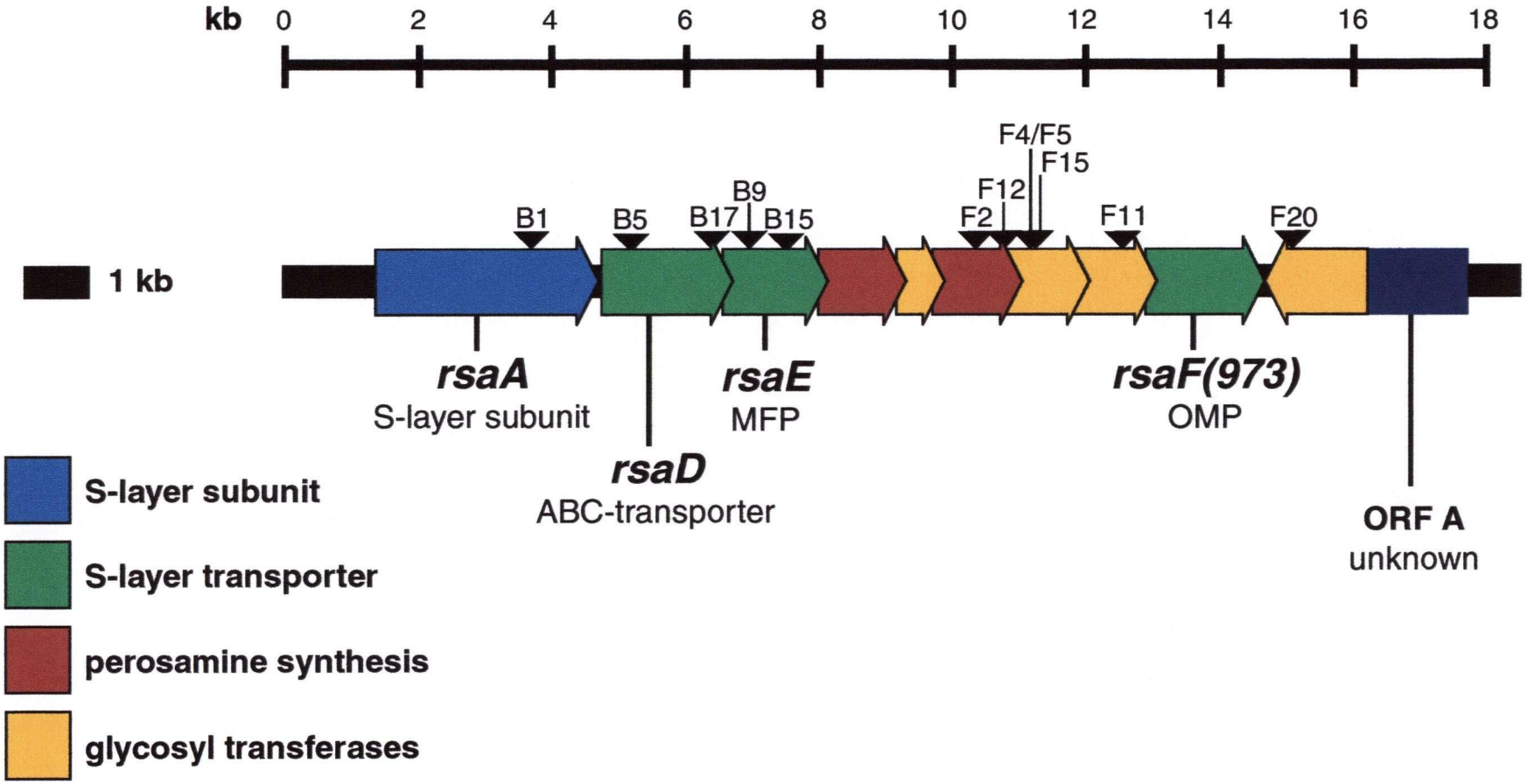

Figure 4–4. **DNA surrounding *rsaA*.** Boxes represent genes or ORFs. Triangles indicate Tn5 insertions. Numbers above the triangles indicate the designation of the Tn5 insertion (B for S-layer negative mutants and F for shedding mutants).

Comparison of RsaF(973) to the protein databases

	Document ID	Accession	Protein	Species	High Score	Smallest Probability P(N)
1.	gi|3860786	(AJ235270)	TolC	Rickettsia prowazekii	160	5.7e-23
2.	gi|882565	(U28377)	n/a	Escherichia coli	103	5.1e-17
3.	gi|135980	(X54049)	TolC	Escherichia coli	103	6.9e-17
4.	gi|3080540	(D49826)	LipD	Serratia marcescens	115	1.9e-16
5.	gi|2495191	(U25178)	TolC	Salmonella enteritidis	90	1.4e-14
6.	gi|4826418	(Y19002)	PrtF	Erwinia amylovora	115	3.2e-13
7.	gi|281563	(M64540)	n/a	Pseudomonas putida	99	1.3e-11
8.	gi|72556	(X00016)	TolC	Escherichia coli (partial)	92	1.4e-11
9.	gi|1405817	(X98513)	HasF	Serratia marcescens	90	3.4e-11
10.	gi|4838370	(AF121772)	NatC	Neisseria meningitidis	111	3.2e-10
11.	gi|4115627	(AB015053)	PrtF	Pseudomonas fluorescens.	92	1.0e-09
12.	gi|117799	(X14199)	CyaE	Bordetella pertussis	87	1.9e-09
13.	gi|3493599	(AF064762)	ZapD	Proteus mirabilis	94	5.9e-09
14.	gi|4063019	(AF083061)	TliF	Pseudomonas fluorescens	85	1.1e-08
15.	gi|2983554	(AE000721)	n/a	Aquifex aeolicus	108	1.6e-08
16.	gi|416635	(X64558)	aprF	Pseudomonas aeruginosa	86	5.3e-08
17.	gi|5759289	(AF175720)	n/a	Porphyromonas gingivalis	66	6.7e-06
18.	gi|5759287	(AF175719)	n/a	Porphyromonas gingivalis	83	0.00017
19.	gi|1653357	(D90913)	n/a	Synechocystis sp.	70	0.00018
20.	gi|3646415	(AJ007827)	EprF	Pseudomonas tolaasii..	78	0.00024
21.	gi|3184190	(AB011381)	OprM	Pseudomonas aeruginosa	74	0.00035
22.	gi|5091481	(AF031417)	TtgC	Pseudomonas putida	66	0.00043
23.	gi|3914250	(L23839)	OprK	Pseudomonas aeruginosa	74	0.0011
24.	gi|95600	(S12527)	PrtF	Erwinia chrysanthemi	80	0.0015

Figure 4-5. BLASTX search showing OMPs similar to RsaF(973). Lines 1 and 2 are predicted from ORF found in genome sequences. OMP from type I systems with the greatest similarity to RsaD and RsaE are underlined. The P(N) value gives the probability of the match arising by chance.

Was either of RsaF(973) or RsaF(1984) the OMP component involved in secretion of RsaA? Sequence similarity was not enough to show that either or both of the genes coded for the OMP. One approach to determine this, was to construct knockout mutants of these ORFs and determine if this prevented secretion. The plasmids pTZ19UΔSSmΔNΔC-RsaF(973) and pTZ18U(CHE)ΔNΔC-RsaF(1984) were constructed to perform the required integration events. Both plasmids consisted of internal portions of the respective genes without the N-terminal and C-terminal. These constructs required only a single recombination event to accomplish the knockout. A single cross-over would produce two copies of the gene, one with an N-terminal deletion and one with a C-terminal deletion, neither of which would be expected to function. To make the pTZ18U(CHE)ΔNΔC-

RsaF(1984), it was still necessary to generate a PCR product containing the coding sequence of *rsaF*(1984). New primers were created using the primer selection methods provided by the MacVector software. The resulting primers were 26 and 28 bp long and had T_m of 71-73°C. Once again the PCR process proved difficult. A PCR product could not be generated at any annealing temperature higher that 55°C, considerably lower than the predicted T_m. When a product was generated, contaminating bands were always present and could not be eliminated by changes in the PCR reaction conditions. Instead, the band of the expected size was gel purified and cloned, giving the plasmid pCR2.1rsaF(1984) which was then used for constructing the deletion clone pTZ18U(CHE)ΔNΔC-RsaF(1984).

The plasmids pTZ19UΔSSmΔNΔC-RsaF(973) and pTZ18U(CHE)ΔNΔC-RsaF(1984) were electroporated into the strains NA1000, and JS4000. JS4000 is a strain of *C. crescentus* that cannot make RsaA, but has functional *rsaDE* genes virtually identical to that of NA1000 (see Ch. 5). Knockouts were only obtained in the strain JS4000 and not NA1000, resulting in the mutants JS4000*rsaF*(973) and JS4000*rsaF*(1984). When AprA was expressed in these mutants, AprA was not secreted by JS4000*rsaF*(973), but was by JS4000*rsaF*(1984) (Fig. 4-6). From these data it was concluded that RsaF(973) is the OMP of the RsaA secretion system.

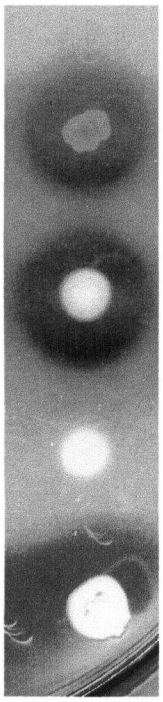

Figure 4-6. AprA secretion from *C. crescentus.* AprA was expressed in all bacteria using pBBR3AprA on skim milk plates. Zones of clearing around the colonies indicate secretion of AprA. Deletion of *rsaF*(973) interrupts secretion of AprA while interruption of *rsaF*(1984) does not interrupt secretion.

To confirm that RsaF(973) was required for secretion, the clone pBBR3AprA:pCR2.1F11Sal1, expressing AprA and RsaF(973) was created. This construct could not be made in *E. coli.* This may be because both of the separate plasmids were toxic, but sublethal. Together the toxic effects may be lethal. The plasmid was obtained by introducing

the ligation mix directly into the knockout strain of RsaF(973). No AprA is secreted from this construct as the plasmid pBBR3AprA:pCR2.1F11Sal1 was unable to complement the knockout. Despite this, it is still believed that RsaF(973) is the OMP of the RsaA secretion system.

Summary

This portion of the project was exceptionally arduous because the *rsaF* genes appeared to be toxic in *E. coli*. This would explain much of the difficulty encountered, such as why the NA1000 cosmid library did not contain *rsaF*(973), why the TIGR genome sequence database does not contain a complete *rsaF* gene sequence, and why it proved difficult to isolate the genes. The lack of colonies resulting from the cloning of the *rsaF*(973) PCR products also suggests a toxic effect. All other attempts to isolate the *rsaF* genes on a fragment of DNA smaller than 7 kb failed, presumably because the smaller inserts were lethal. This suggests that the *rsaF* genes are lethal to *E. coli* and the clones obtained contain mutations that make the insert less toxic.

As mentioned above, this presumed toxicity may explain why the partial TIGR genome sequence contained only partial ORFs of the *rsaF* genes. Other analysis of the TIGR sequence suggests that greater than 80% of the *C. crescentus* genome is represented (see Ch. 6). Given that, the sequence reported here for *rsaF*(973) may differ from the wildtype sequence. Such a mutant *rsaF*(973) gene in the plasmid PCR2.1F11SalI may not produce a protein that functions correctly. This would explain why this plasmid was tolerated in *E. coli* while other constructs appeared to be lethal and would explain why the plasmid pBBR3AprA:pCR2.1F11SalI failed to complement the RsaF(973) knockout. It is unlikely that the phenotype of the *rsaF*(973) knockout is caused by a polar mutation because the gene 3' of *rsaF*(973) is transcribed in the opposite orientation. Even given the failure to complement the knockout, the results presented here indicate that RsaF(973) is the OMP required for secretion of RsaA.

The function *rsaF*(1984) is not known. The entire ORF was never cloned and sequenced so it was not possible to determine if an entire ORF coding for an OMP exists. The sequence identity between the two *rsaF* ORFs suggests that one may

be a gene duplication of the other and that *rsaF*(1984) is no longer functional. Another possibility is that there is a second type I secretion system in *C. crescentus* (though it is not known what it might transport) that uses RsaF(1984) as the OMP component. Determining the function of *rsaF*(1984) represents a future project.

Chapter 5

Identification of the S-layer subunit and transporter genes in Freshwater *Caulobacter* species

Introduction

The Smit laboratory strain culture collection contains numerous strains that have been isolated from locales around the world and are designated FWC (freshwater *Caulobacter*) species (MacRae and Smit, 1991). Analysis of these FWC species showed that not all have an S-layer (Walker *et al.*, 1992). There seems to be a geographical as well as evolutionary distinction between these species (Abraham *et al.*, 1999; MacRae and Smit, 1991). No FWC with an S-layer has been found in Europe, though admittedly, only a small fraction of the FWC species were isolated from European sources while FWC species with and without S-layers were found in North America.

The evolutionary relationships between the different FWC species have recently been examined by 16S rDNA sequencing, profiling of restriction fragments of 16S-23S rDNA interspacer regions, lipid analysis, immunological profiling and salt tolerance characteristics to organize the taxonomy of 76 different strains (Abraham *et al.*, 1999). It was demonstrated that all of the FWC species with S-layers are much more closely related to one another than to the species without S-layers, and the non-S-layer FWC species have been reclassified as the genus *Brevundimonas* instead of *Caulobacter.* Therefore S-layers are a characteristic of *Caulobacter* species.

The S-layers of the *Caulobacter* species have been previously examined. The S-layer subunits range in size from 100 kDa (comparable to NA1000) to 193 kDa and can be removed by a low pH or EGTA extraction method. All the putative S-layer proteins react with antibody raised against RsaA (though most often to a lesser extent) and most also produce a polysaccharide that reacts to antibody against the S-LPS responsible for attachment of the S-layer in NA1000 (Walker *et al.*, 1992). It was also shown that these FWC species will hybridize with an *rsaA* probe under conditions that would allow up to 30% mismatch (MacRae and Smit,

1991). This suggests that the S-layer subunits on these other FWC species are similar to RsaA and may also be secreted by a type I secretion mechanism.

Two strains have been used predominantly for the examination of the S-layer in *C. crescentus*. NA1000 is a variant of the ATCC 19089 strain, whose genome is being sequenced by TIGR. It is from NA1000 that the *rsaA* gene and *rsaD* and *rsaE*, genes responsible for secretion of RsaA, were isolated (see Ch. 3). The second strain used in the Smit lab is JS4000, a lab variant of the ATCC 15252 strain that spontaneously lost its S-layer during culturing, and is being used for expression of recombinant proteins secreted using the NA1000 *rsaA* gene. The S-layer gene from JS4000 has been cloned and expressed in *E. coli* where it produces a 40,000 molecular weight protein in inclusion bodies (Bingle *et al.*, 1999). ATC15252 has an S-layer gene that appears to be identical to RsaA as determined by size and antibody reactivity, yet other characteristics of the bacterium (i.e. cell appearance, growth rates), 16S rRNA sequencing (Stahl *et al.*, 1992) and RFLP mapping of the genome (B. Ely, pers. comm.) showed that it is different from NA1000.

Preliminary investigations of these S-layers that were begun in order to determine the differences between the S-layer subunits and their associated transport systems are presented here and have now been taken over by Mihai Iuga. It is hoped that analysis of these other S-layer systems will provide insight into the transport mechanisms by showing what changes in the transporters are required to transport the different sized subunits.

Results and Discussion

The S-layer subunit, ABC-transporter and Membrane Forming Unit proteins of JS4000 and NA1000 *Caulobacter* species are virtually identical. The S-layer genes from both JS4000 and JS3001, a shedding derivative of ATCC 15252, were cloned and sequenced (see Ch. 2) and have few differences when compared to the sequence of the NA1000 *rsaA*. In a few places the guanosine (G) and cytosine (C) residues are reversed (i.e., GC instead of CG), but these are in regions of high G+C content and appear to be errors in the original sequencing of *rsaA* (Gilchrist *et al.*, 1992) as the partial *Caulobacter* genome sequence from TIGR

supports my sequencing results. The sequence for NA1000 was amended accordingly. The error in the JS4000 sequence that truncates the S-layer protein consists of a guanosine base that has been deleted from codon 357 which causes a termination codon to be read at codon 359. These differences are listed in Table 5-1.

The *rsaD* and *rsaE* genes from JS4000 have been isolated from a cosmid library (see Ch.2) and were sequenced. These genes are almost identical to the NA1000 genes. The differences between the strains are summarized in Table 5-1.

		ATCC 19089	ATCC 15252	
		NA1000	JS4000	JS3001
RsaA	aa 358-359-360	Gln-Asn-Leu	Gln-Thr-None	Gln-Asn-Leu
	aa 475	Val	Ile	Val
	aa 860	Thr	Ser	Thr
RsaD	aa 298	Asn	Thr	ND
RsaE	aa 131-132	Ser-Gln	Arg-Leu	ND

Table 5-1. Differences between the Rsa genes found in lab strains. Deduced amino acid sequence differences between the RsaA, RsaD and RsaE proteins of three common lab strains of *C. crescentus.* ND- not determined

The S-layers of FWC species are probably transported by a type I secretion system. The alkaline protease gene, AprA, from *P. aeruginosa* is secreted by the RsaA secretion machinery (see Ch. 3). AprA was successfully secreted in selected strains covering the range of S-layer subunit sizes, demonstrating that these strains also had type I secretion mechanisms (Table 5-2). AprA secretion was varied in the differing FWC species. While in NA1000 all the colonies containing the *aprA* gene secreted AprA, not all FWC colonies did. While some species (i.e., FWC 19)

Species	AprA secretion	Penetrance* (%)	Subunit size
NA1000	++	>99.9	98 kDa
JS4000	++	>99.9	98 kDa
FWC 8	++	80	122 kDa
FWC 9	+	>99.9	133 kDa
FWC 17	+	78	106 kDa
FWC 19	+	>99.9	108 kDa
FWC 28	+	45	106 kDa
FWC 32	+	10	133 kDa
FWC 39	+	80	193 kDa
FWC 42	+	10	181 kDa

Table 5-2. FWC species secreting alkaline protease.
++ represents 70-100% of the NA1000 secretion level,
+ represents 20-69% of the NA1000 secretion level
* penetrance was the number of colonies expressing AprA

showed full penetrance (all colonies expressed AprA), in other FWC species as few as 10% of the colonies secreted AprA when the *aprA* gene was expressed (i.e., FWC 32). It is not known why only some colonies secreted AprA. *P. aeruginosa* also expresses an inhibitor that binds to the AprA and prevents proteolytic activity inside the cell. As the inhibitor is not expressed with *aprA* in the FWC species, AprA may have a toxic effect on *Caulobacter* cells and there may be selective pressure to eliminate it from the cells. Cells not secreting AprA, may have found a way to prevent expression of the gene. NA1000 and some of the FWC species may be better able to tolerate the toxicity than other species.

FWC species with similar subunit sizes have similar Southern blot banding patterns. To further characterise the FWC species, Southern blot analysis was performed using probes to *rsaA* and *rsaDE.* These blots were performed under conditions that would allow up 30% mismatch. The results are summarized in Table 5-3.

Caulobacter species	Subunit size (kDa)	Fragment size when probed with *rsaD* and *rsaE* (enzyme[1])	Fragment size when probed with *rsaA* (enzyme[1])
NA1000 JS3000	98	>20kb(*Eco*RI),7.1kb(*Hind*III)	7.1kb (*Hind*III)
FWC 17	106	3.5 kb (*Eco*RI), 5kb (*Hind*III)	4.3 kb (*Eco*RI)
FWC 18	131	ND[2]	7.0 kb (*Bam*HI)
FWC 19	108	3.5 kb (*Eco*RI)	4.4 kb (*Eco*RI)
FWC 28	106	3.5 kb (*Eco*RI)	4.3 kb (*Eco*RI)
FWC 31	106	3.5 kb (*Eco*RI)	4.3 kb (*Eco*RI)
FWC 42	181	10 kb (*Eco*RI)	8.0 kb (*Eco*RI)

Table 5-3. Comparison of Southern Blot banding patterns of different FWC species. Chromosomal digests with the enzyme specified were probed with either *rsaA* or *rsaDE*.
[1] Enzyme that chromosomal DNA was cut with for Southern blot analysis
[2] Not Determined

Analysis of the Southern blot data suggests that the S-layer subunits and transporters can be grouped according to size. All of the FWC species with subunits ranging from 106-108kDa have identical Southern banding patterns, while all the other FWC species with different subunit sizes have different banding patterns. The ability of the *rsaDE* genes to hybridize to the chromosome of the differing FWC species suggests that the S-layer subunit is being secreted by a type I transporter. With this in mind, methods were devised for isolating the genes involved.

The ABC-transporter subunits were isolated from several different FWC species. The sequence identity between ABC transporter among different type I systems is the most significant of the 3 transporter components. Using the sequence identity between the ABC-transporters *aprD* (*P. aeruginosa*), *prtD* (*E. chrysanthemi*) and *rsaD* (NA1000), degenerate primers were designed to amplify a central portion of the ABC transporter using PCR. Using these primers it was possible to amplify, clone and sequence fragments of the ABC transporter from FWC6, FWC8 and FWC39. PCR products were not successfully generated from FWC17, FWC26,

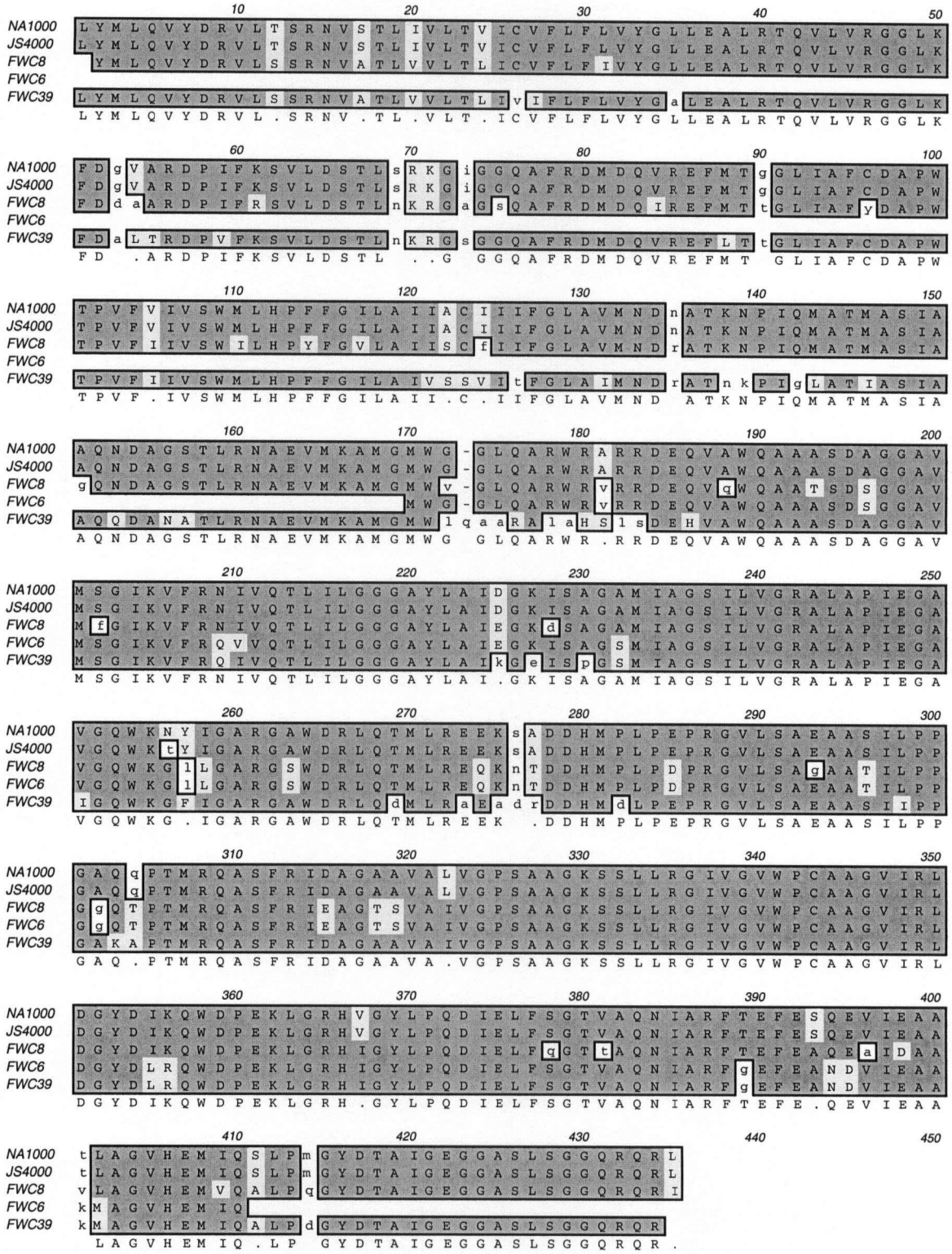

Figure 5-1. **ClustalW alignment of partial RsaD genes from FWC species.** Identical residues have dark shading. Similar residues are shaded lightly. The line underneath the alignment is the consensus sequence.

FWC28, FWC29 and FWC41. Multiple bands were generated from FWC27 and FWC42, but I was unable to clone any of the fragments.

Obviously, the PCR strategy selects for ABC-transporters most closely related to the NA1000 gene. This suggests that even though the subunit of FWC6 is 181kDa and that of FWC39 is 193 kDa, the transporters are still closely related to FWC8 with a subunit of 122 and NA1000 with a subunit of 98 kDa and this was confirmed by sequencing (Fig 5-1). Curiously, FWC species with small subunit sizes close to that of NA1000 failed to generate PCR products suggesting that the sequences of their ABC-transporters have diverged more from the NA1000 sequence. Analysis of the sequence showed little division between the FWC species according to size. In some places along the deduced protein sequence, the transporters of smaller subunits are more similar to one another than to the transporters of larger subunits while in others, the sequences of transporters of differing sizes are more similar to one another (Fig 5-1).

A method for screening the chromosomes of FWC species for the S-layer subunit and S-layer transport genes was devised (see Ch. 2). Using this method, part of the S-layer subunit gene for FWC 27 was isolated. FWC27 has an S-layer subunit size of 145 kDa. Comparison of the sequence to NA1000 reveals that there is a considerable difference in the sequence of these proteins (Fig. 5-2). A BLAST alignment of the RsaA and FWC27 sequences (Altschul *et al.*, 1990) shows that the proteins are 44.6% identical and 61.5% similar over 130 amino acids.

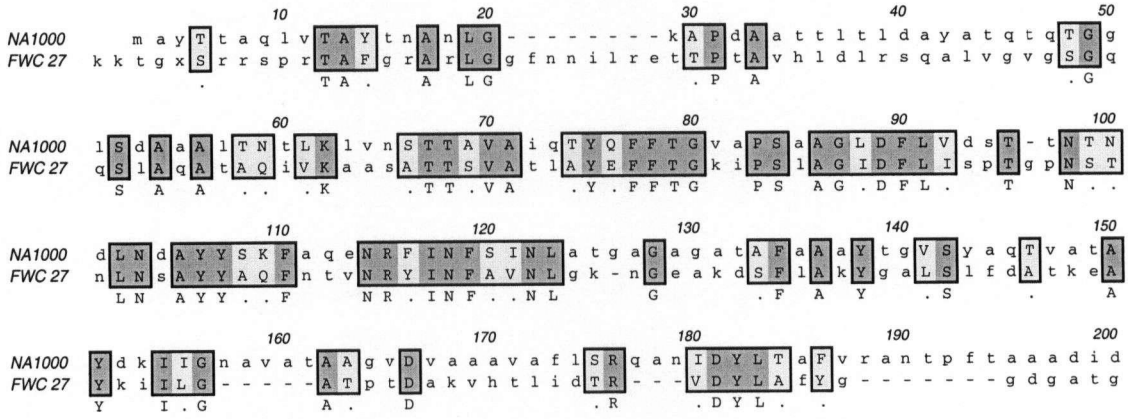

Figure 5-2. ClustalW alignment of FWC 27 with the first 200 amino acids of RsaA. Identical residues have dark shading. Similar residues have light shading. Identical and similar residues are boxed. The line underneath the sequences is the consensus sequence.

The sequence of RsaA contains repeating amino acid sequence elements. Sequence analysis of RsaA has revealed that portions of the sequence exhibit considerable sequence similarity to other portions of the molecule. Table 5-4 shows the similarity of the Ca^{2+} binding domain of RsaA to sequences closer to the N-terminal. These similar units do not appear to be uniform in size and appear to consist of 60 to 90 amino acid segments, but the exact sizes have not been determined. These segments may represent a complete structural domain (i.e., α-helix or β-strand) that is replicated along the length of the protein, but further analysis is required to confirm this.

As Table 5-4A shows, the alignments of RsaA along different portions of itself can result in as much as 28% identical amino acids. Furthermore, the Expect numbers, representing the possibility of the match occurring by chance in a random sequence database of the current size, are very small. Table 5-4B shows the other hits in the database to the same portion of RsaA. The Sap proteins from *C. fetus* are S-layer proteins with the greatest identity to RsaA. HlyA from *Aquifex aeolicus* and the hypothetical protein from *Rhodobacter capsulatus* both contain the calcium binding motifs found in proteins secreted by type I systems, leading to higher identity. As the Expect numbers show, the identity to RsaA along itself is greater than what would be found by chance in the sequence database. This repetitive

nature is also seen at the DNA level (data not shown). It must be taken into account that the nature of the RsaA composition (26% threonine and serine) leads to a higher number of repetitive sequences occurring than would be expected by chance. This explains why a low Expect number occurs with alignments to a membrane glycoprotein from Equine herpesvirus which also contains a high number of threonine and serine residues. It is only at Expect numbers of 1.8e-08, much higher than the best expect number of 6e-14 of RsaA to itself, that random proteins begin to show identity. Overall, the repetitive nature found here is higher than could be expected by chance and suggests that RsaA evolved by duplicating structural portions of the molecule to form a larger protein.

Table 5-4. BLAST alignment of RsaA with itself.

A

```
pir||A48995 paracrystalline surface layer protein RsaA - Caulobacter crescentus
           Length = 1026

 Score =  573 bits (1461), Expect = e-163
 Identities = 300/300 (100%), Positives = 300/300 (100%)

Query: 1    QLGATAGATTFTNVAVNVGLTVLAAPTGTTTVTLANATGTSDVFNLTLSSSAALAAGTVA 60
            QLGATAGATTFTNVAVNVGLTVLAAPTGTTTVTLANATGTSDVFNLTLSSSAALAAGTVA
Sbjct: 721  QLGATAGATTFTNVAVNVGLTVLAAPTGTTTVTLANATGTSDVFNLTLSSSAALAAGTVA 780

Query: 61   LAGVETVNIAATDTNTTAHVDTLTLQATSAKSTVVTGNAGLNLTNTGNTAVTSFDASAVT 120
            LAGVETVNIAATDTNTTAHVDTLTLQATSAKSTVVTGNAGLNLTNTGNTAVTSFDASAVT
Sbjct: 781  LAGVETVNIAATDTNTTAHVDTLTLQATSAKSTVVTGNAGLNLTNTGNTAVTSFDASAVT 840

Query: 121  GTGSAVTFVSANTTVGEVVTIRGGAGADSLTGSATANDTTIGGAGADTLVYTGGTDTFTG 180
            GTGSAVTFVSANTTVGEVVTIRGGAGADSLTGSATANDTTIGGAGADTLVYTGGTDTFTG
Sbjct: 841  GTGSAVTFVSANTTVGEVVTIRGGAGADSLTGSATANDTTIGGAGADTLVYTGGTDTFTG 900

Query: 181  GTGADIFDINAIGTSTAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQ 240
            GTGADIFDINAIGTSTAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQ
Sbjct: 901  GTGADIFDINAIGTSTAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQ 960

Query: 241  YLDAAAAGDGSGTSVAKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFAT 300
            YLDAAAAGDGSGTSVAKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFAT
Sbjct: 961  YLDAAAAGDGSGTSVAKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFAT 1020

 Score = 78.4 bits (190), Expect = 6e-14
 Identities = 85/318 (26%), Positives = 133/318 (41%), Gaps = 37/318 (11%)

Query: 2    LGATAGATTFTNVAVNVGLTVLAAPTGTT--TVTLANATGTSDVFNLTLSSSAALAAGTV 59
            L AT  A    NVAV+ G  V  A TG T  T T+   +  S   ++++++S+    G +
Sbjct: 360  LTATTAAQAANNVAVDGGANVTVASTGVTSGTTTVGANSAASGTVSVSVANSSTTTTGAI 419

Query: 60   ALAGVETVNIAATDTNTTAHVDTLTLQATSAKSTVVTGNAGLNLTNTGNTAVTSFDASAV 119
            A+ G   V +A T  N    V+T   QA     + VTGN+          TA  +  A+
Sbjct: 420  AVTGGTAVTVAQTAGNA---VNTTLTQA----DVTVTGNSSTTAVTVTQTAAATAGATVA 472

Query: 120  TGTGSAVTF----VSANTTVGEVVTIR-GGAGADSLTGSATANDTTIG-----GAGADTL 169
                 AVT      ++ TT G++ T+  G  GA ++  SA     + G     G G   L
Sbjct: 473  GRVNGAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALTTVNLSGTGTSLGIGRGAL 532

Query: 170  VYTGGTDTFTGGTGADIFDINAIGTSTAFVTITDAAVGD---KLDLVGISTNGAIADGAF 226
              T   +T T        ++N + T+T  +T ++AA  D    +++ G + +  TA
Sbjct: 533  TATPTANTLT-------LNVNGL-TTTGAITDSEAAADDGFTTTINIAGSTASSTIASLVA 584

Query: 227  GAAVTLG----AAATLAQYLDAAAAGDGSGTSVAKWFQFGGDTYVVVDSSAGATFV---S 279
              A TL     A  T+  +  AA  G     SV          T +V    AGA  +   +
Sbjct: 585  ADATTLNISGDARVTITSHTAAALTGITVTNSVGATLGAELATGLVFTGGAGADSILLGA 644

Query: 280  GADAVIKLTGLVTLTTSA 297
              A++   G  T+T S+
Sbjct: 645  TTKAIVMGAGDDTVTVSS 662
```

Table 5-4 continued

```
 Score = 66.3 bits (159), Expect = 3e-10
 Identities = 94/361 (26%), Positives = 143/361 (39%), Gaps = 80/361 (22%)

Query: 4   ATAGATTFTNVAVNVGLTVLAAPTG----TTTVTLANAT--GTSDVFNLTLSSSAALAAG 57
           A +  TT   +AV  G  V  A T      TT+T A+ T  G S    +T++ +AA  AG
Sbjct: 409 ANSSTTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTGNSSTTAVTVTQTAAATAG 468

Query: 58  -TVA--LAGVETVNIAATDINTTA-HVDTLTLQATSAKSIVVTGNAGLNLTNTGNTAVTS 113
            TVA  + G  T+  +A  + TTA  + T+TL +  A +I  +     +NL+ TG +
Sbjct: 469 ATVAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALTTVNLSGTGTSLGIG 528

Query: 114 FDASAVTGTGSAVTF-VSANTTVGEV-------------VTIRGGAGADSLTGSATANDT 159
             A   T T + +T  V+  TT G +              + I G   + ++     A+ T
Sbjct: 529 RGALTATPTANTLTLNVNGLTTTGAITDSEAAADDGFTTINIAGSTASSTIASLVAADAT 588

Query: 160 IIGGAGADTLVYTGGTD-----------------------TFTGGTGADIFDINA---- 191
           +  +G   +  T  T                          FTGG GAD   + A
Sbjct: 589 TLNISGDARVTITSHTAAALTGITVINSVGATLGAELATGLVFTGGAGADSILLGATTKA 648

Query: 192 --IGTSTAFVTITDAAV--------GDKLDLVGISTNGA--IADGAFGAAVTLGAAATLA 239
             +G     VT++ A +        GD  D++  + NG+   AD AFG   TL
Sbjct: 649 IVMGAGDDTVTVSSATLGAGGSVNGGDGTDVLVANVNGSSFSADPAFGGFETLRV----- 703

Query: 240 QYLDAAAAGDGSGTSVAKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFA 299
              A AA  GS  +        G T + + ++AGAT  +     + LT L   T +
Sbjct: 704 ----AGAAQGSHNA-------NGFTALQLGATAGATTFTNVAVNVGLTVLAAPIGTTTV 752

Score = 66.0 bits (158), Expect = 3e-10
 Identities = 85/301 (28%), Positives = 121/301 (39%), Gaps = 46/301 (15%)

Query: 2   LGATAGATTFTNVAVNVGLTVLAAPIGTTTVTLANATGTSDVFNLTLSSSAALAAGTVAL 61
           L A   A T    A ++ L V AA  GT    + NA         T+S     A  T A+
Sbjct: 172 LTAFVRANTPFTAAADIDLAVKAALIGT----ILNAA--------TVSGIGGYATATAAM 219

Query: 62  AGVETVNIAATDINTTAHVDTLTLQATSAKSIVVTGNAGLNLTNTGNTAVTSFDASAVTG 121
             +  ++  A  T+  A V+  T   +S  S               G+T   +      +TG
Sbjct: 220 --INDLSDGALSTDNAAGVNLFTAYPSSGVS--------------GSTLSLTTGTDTLTG 263

Query: 122 TGSAVTFVSANTTVGEVVTIRGGAGADSLTGSATANDTIIGGAGADTLVYTGGTDTFTGG 181
           T +  TFV+      GEV      AGA +LT      DT+ GGAG D L +
Sbjct: 264 TANNDTFVA-----GEV------AGAATLT----VGDTLSGGAGTDVLNWVQAAAVTALP 308

Query: 182 TGADIFDINAIG-TSTAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQ 240
           TG  I  I  +  TS A +T+  ++     L  +  +T+GA       GA   L  A T AQ
Sbjct: 309 TGVTISGIEIMNVTSGAAITLNTSSGVTGLTALNTNTSGAAQTVTAGAGQNL-TATTAAQ 367

Query: 241 YLDAAAAGDGSGTSVAKWFQFGGDTYVVVDSSA-GATFVSGADAVIKLTGLVTLTTSAFA 299
             +  A   G+  +VA      G T V  +S+A G   VS A++     TG + +T
Sbjct: 368 AANNVAVDGGANVTVASTGVTSGTTTVGANSAASGTVSVSVANSSTTTTGAIAVTGGTAV 427
```

Table 5-4 continued

```
 Score = 62.8 bits (150), Expect = 3e-09
 Identities = 77/293 (26%), Positives = 125/293 (42%), Gaps = 38/293 (12%)

Query: 12  TNVAVNVGLTVLAAPIGITTVTLANATGTSDVFNLTLSSSAALAAGTVALAGVEIVNIAA 71
           T+ A  V L      +G +  TL+  TGT  +     +++     AG VA A   TV
Sbjct: 230 TDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGT--ANNDTFVAGEVAGAATLTVGDTL 287

Query: 72  TDINTTAHVDTLTLQATSAKSIVVTGNAGLNLTNTGNTAVTSFDASAVTGTGSAVTFVSA 131
           +    T  ++ +   A +A    VT +    +  T   A+T   +S VTG    +T ++
Sbjct: 288 SGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSGVTG----LTALNT 343

Query: 132 NTTVGEVVTIRGGAGAD--SLTGSATANDTIIGGAGADTLVYTGGTDTFTGGTGADIFDI 189
           NT+ G   T+  GAG +  + T +  AN+  + G    T+  TG T      +G
Sbjct: 344 NTS-GAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVASTGVT------SGTTTVGA 396

Query: 190 NAIGTSTAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQYLDAAAAGD 249
           N+  + T  V++ +++           +T GAIA       VT G A T+AQ       AG+
Sbjct: 397 NSAASGTVSVSVANSST---------TTTGAIA-------VTGGTAVTVAQ-----TAGN 435

Query: 250 GSGTSV--AKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFAT 300
              T++  A     G  +   V  +  A   +GA     ++ G VT+T  SA A+
Sbjct: 436 AVNTTLTQADVIVIGNSSTTAVTVIQTAAATAGATVAGRVNGAVTITDSAAAS 488
```

B

Sequences producing High-scoring Segment Pairs:		High Score	Smallest Probability P(N)
1. gi|477427	RsaA - Caulobacter crescentus	1461	1.1e-187
2. gi|2120535	SapB - Campylobacter fetus	154	9.5e-17
3. gi|2120536	SapA - Campylobacter fetus	108	1.1e-11
4. gi|2114323	membrane glycoprotein Equine herpesvirus 1	153	1.5e-11
5. gi|94640	SapA - Campylobacter fetus	100	1.4e-10
6. gi|2983562	HlyA - Aquifex aeolicus-hemolysin protein	132	9.9e-09
7. gi|2114321	membrane glycoprotein Equine herpesvirus 1	130	1.8e-08
8. gi|3128319	hypothetical protein-Rhodobacter capsulatus	98	4.3e-08
9. gi|2606019	envelope glycoprotein - Equine herpesvirus 4	127	4.7e-08
10. gi|4063042	glycoprotein - Cryptosporidium parvum	125	8.7e-08
11. gi|790694	epimerase -Azotobacter vinelandii	111	4.1e-07
12. gi|3128317	hypothetical protein-Rhodobacter capsulatus	102	1.4e-06
13. gi|790692	epimerase -Azotobacter vinelandii	109	1.4e-06

Table 5-4. BLAST alignment of RsaA with itself. A) Portions of the sequence of RsaA exhibit considerable sequence similarity to other portions of the molecule. Query represents the 300 amino acid segment of RsaA from 721-1020. Sbjct represents the entire sequence of RsaA. Numbers alongside the sequence indicate amino acid positions. The line between the Query and Sbjct lines indicates identical amino acids with the appropriate letter code and similar amino acids with a '+'. Identities refers to the number of identical amino acids shared between the sequences. Positives refers to the combined number of identical and similar amino acids shared between the sequences. Expect gives the possibility of the sequence alignment occurring by chance considering the current size of the sequence databases. B) Result of BLAST search showing the closest matches to the amino acids 721-1020. P(N) numbers are almost identical to Expect numbers for Expect numbers<0.001 (Altschul *et al.*, 1990).

Phylogenetic analysis of the FWC species has shown that the FWC species can be divided into five branches. Analysis of the phylogenetic study Abraham *et al.*, 1999 shows that there are two branches, B and D, of the *Caulobacter* phylogenetic tree that contain species with only small, 100-108, kDa S-layers (Fig 5-3). FWC19, FWC28 and FWC31 belong to one of these branches and FWC 17 belongs to the other. These are the four strains with identical Southern blot banding patterns (Table 5-3) suggesting that the S-layers and associated transporters of these two branches are more closely related to each other than to the other three branches. The three other branches show no correlation between subunit size and evolutionary distance as they have S-layer subunit sizes ranging from small (102 kDa) to large (193 kDa). In addition to this, the species FWC6, FWC8 and FWC39, that proved easiest to amplify the ABC-transporter by degenerate PCR, all belong to different branches. This may simply reflect the conserved nature of the ABC-transporter. It may be that the larger S-layers evolved separately from one another and the similarities between ABC-transporters transporting large subunits (but not found in ABC-transporters transporting small subunits) may represent convergent evolution required to accommodate secretion of a larger subunit.

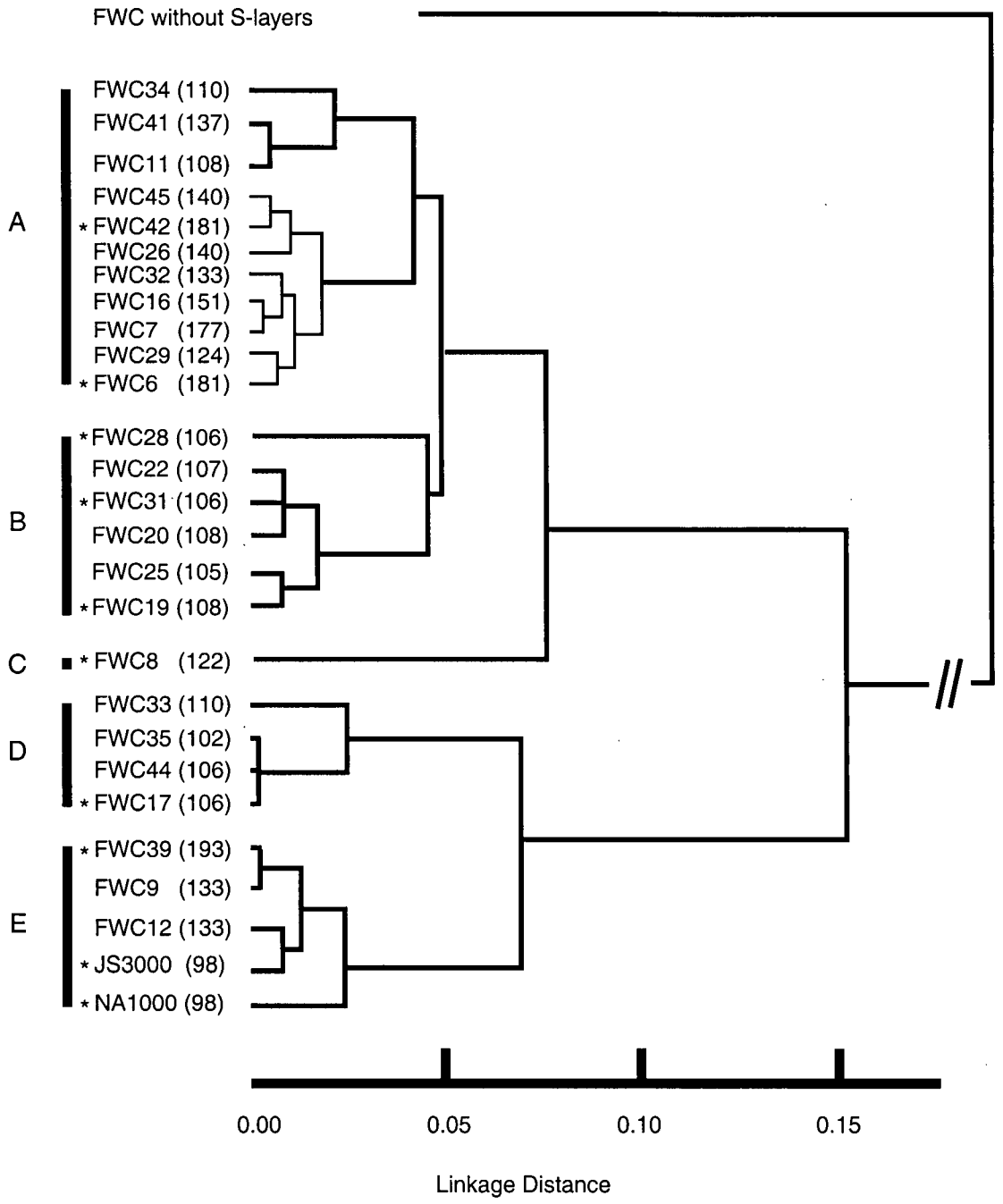

Figure 5-3. Dendrogram derived from *Caulobacter* glycolipid content (Adapted from Abraham *et al*, 1999). The FWC species have been organized into 5 groups with a linkage difference of more than 0.05. * species examined in this study. Numbers in brackets refer to the size of the S-layer subunit in kDa.

Summary

The evolutionary relationships of the S-layer subunits and associated transporters of the different FWC species have been examined here. These results are still preliminary and more work needs to be done to substantiate these conclusions. While keeping this in mind, I will hypothesize on the evolutionary relationships that the data presented here suggest.

The repetitive nature of RsaA suggests how the different sizes of S-layers could have arisen among the different FWC species. The larger S-layer subunits from such strains as FWC39 and FWC41 may consist of an even more repetitive nature to account for the greater bulk. Larger S-layer subunits might arise from a duplication of DNA within the gene for the subunit.

The phylogenetic analysis of the FWC species by Abraham and collegues shows little evolutionary relatedness with regard to S-layer subunit size (Fig. 5-3). While groups B and D contain only smaller S-layer subunits other groups contain a range of sizes. The most pronounced difference in subunit size is found in group E between the species with the largest (FWC 39) and the smallest (NA1000/JS3000) subunits, yet the bacteria are very closely related according to glycolipid content. Thus, it seems that the large S-layer subunits arose independently. The identical amino acid changes seen in the ABC-transporters with large S-layer subunits suggest that these amino acids may be required changes for transporting a subunit of a large size. Further work on analyzing these differences is required before anything conclusive can be determined, and is of great interest since this information would help determine the factors that must be considered when designing recombinant proteins for secretion.

In reviewing all current data, I hypothesize that the progenitor of the six branches of FWC species had a small (106-108 kDa) S-layer subunit and the two branches consisting solely of small S-layer subunits represent FWC that are most closely related to the progenitor. The S-layer subunits of the FWC species in the other four branches may have altered their sizes more recently. The repetitive nature of the S-layer sequence may have assisted in the duplication of sequence segments by allowing slippage during gene replication to create larger S-layer

subunits. Smaller subunits such as the 98 kDa NA1000 subunit may have resulted from deletion of repeated units. It may be that to accommodate the different sized subunits, the ABC-transporter components must be changed at specific residues to allow secretion of larger subunits. If convergent evolution resulted in the similarities found between the large subunit transporters here, then these similarities will indicate what portions of the protein are involved in transport of the larger subunit. I believe that the analysis of the S-layer subunits and transporters in this manner will allow a much greater understanding of the type I secretion systems.

Chapter 6
Identification of genes involved in the synthesis of the O-Antigen of *C. crescentus*

Introduction

The S-LPS of *C. crescentus* is responsible for attachment of the S-layer to the surface of the bacterium. Disruption of proper O-antigen formation in the S-LPS causes the RsaA molecules to slough off or 'shed' from the surface and assemble into sheets (Fig. 6-1). The S-LPS has been isolated and analyzed from S-layer negative NA1000 mutants (Walker *et al.*, 1994) and has the same core and lipid composition as the rough LPS (Ravenscroft *et al.*, 1992). Further analysis of the O-antigen (Smit, unpublished) has revealed that the O-antigen of the S-LPS appears to be composed of a homopolymer of a 4,6-dideoxy-4-amino-hexose. Mass spectrometry indicates that the O-antigen has a mass consistant with of forty of these hexose units. This homopolymer is unusual in that a number of different anomeric proton signals can be found when it is analyzed by proton NMR suggesting that the individual sugar units may not all be linked in the same manner. Presented in this report is evidence that this 4,6-dideoxy-4-amino-hexose is, most likely, the sugar perosamine. Perosamine is not commonly found in the O-antigen and only a few species, including *Vibrio cholerae*, *Brucella melitensis* and *E. coli* O157, contain perosamine residues (Stroeher *et al.*, 1995; Wang and Reeves, 1998). In addition, a number of glycosyltransferases have been found which may be

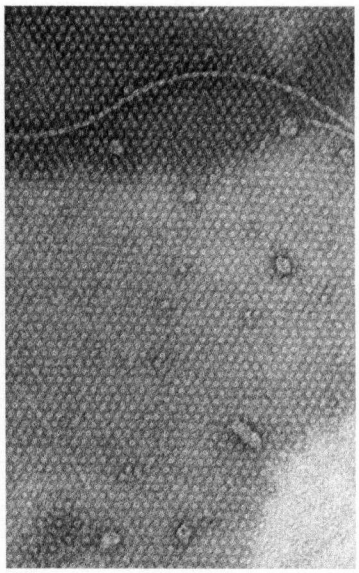

Figure 6-1. Shed S-layer from *C. crescentus*. EM photo-graph of S-layer shed from a strain with defective S-LPS. (Photo courtesy John Smit)

Figure 6-2. Colony Immunoblot. Example of an immunoblot demonstrating the different phenotypes exhibited by mutants.

the basis for the different linkages making up the homopolymeric O-antigen.

Results and Discussion

Several Tn5 mutants producing altered S-LPS were found. The screen used to detect transport deficient mutants also detected S-LPS mutants in the NA1000 Tn5 library. On plates, these mutants exhibit a 'halo' of RsaA protein diffusing out from the colonies that can be easily distinguished with an immunoblot from bacterial colonies not shedding the S-layer (Fig. 6-2). This method was used to isolate a total of 26 'shedders' from the NA1000 Tn5 library with altered S-LPS (Fig 6-3).

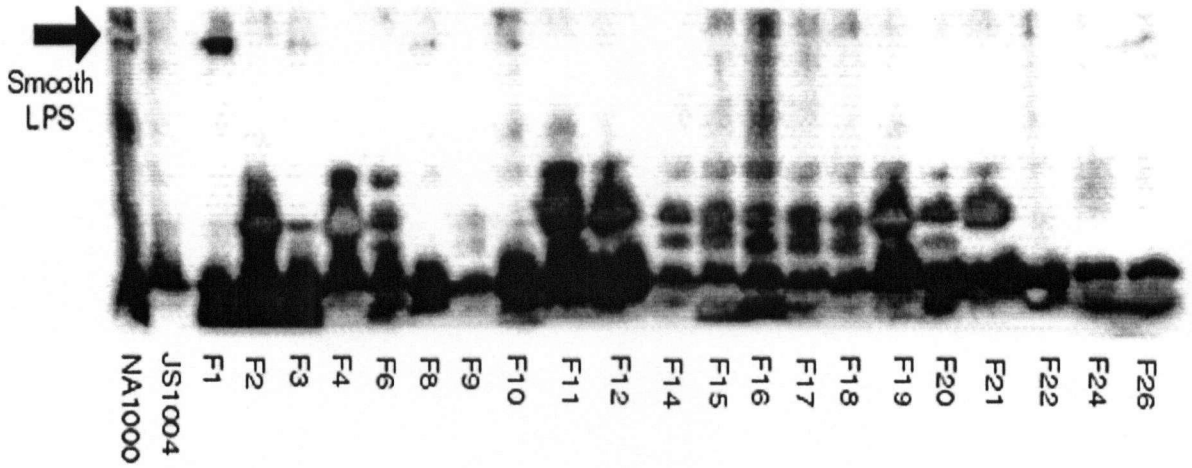

Figure 6-3. S-LPS of shedding Tn5 mutants. Silver stained polyacrylamide gel of S-LPS extracts from representative NA1000 shedder Tn5 mutants. NA1000 shows the wildtype form of S-LPS. JS100 is a spontaneous shedder mutant with a defective S-LPS. The large dark band at the bottom is the rough LPS.

Southern blot analysis of these mutants has shown that mutants F1-F22 consisted of 16 different Tn5 insertions (data not shown). Further Southern blot characterisation of the mutants showed that F8 was not a proper Tn5 insertion since the banding pattern was incorrect when probed with Tn5. Southern blots probed with the coding sequence of *rsaA* showed that the *rsaA* band in the mutant F21 was not the same as wildtype. This suggested that the Tn5 mutation did not result in the shedding phenotype, but instead a second mutation resulting in a deletion of the

rsaA gene was responsible (data not shown). To further characterise these mutants, Southern blot analysis using *Eco*RI and *SstI* was performed on the chromosomal DNA of these mutants. Both of these enzymes do not cut Tn5 and as a result can be used to determine if the Tn5 insertions are linked. The Southern blots were probed with a portion of the Tn5 and the banding patterns have been summarized in Tables 6-1 and 6-2. The results showed that the majority of these mutants have identical banding patterns (groups C and I) and are linked. Of the remaining mutants: F10 and F22 appear to be linked, while F3 and F9 are not linked to any of the others (Tables 6-1 and 6-2). Four of these mutants were isolated at a later date and were not characterised by Southern (F23-F26).

Southern blot analysis of chromosomal DNA digested using *Eco*RI

Mutant	Group A 8.1 kb	Group B 15 kb	Group C 23 kb	Group D 30 kb	Group E 35 kb
F1			x		
F2			x		
F3		x			
F4			x		
F6	x				
F9					x
F10				x	
F11			x		
F12			x		
F14			x		
F15			x		
F19			x		
F20			x		
F22				x	

Table 6-1. Compilation of Southern blot data from *Eco*RI digestion of shedder mutant chromosomal DNA. *Eco*RI does not cut Tn5. The Southern blots were probed with a fragment of Tn5. Mutants are grouped according the band size seen on the Southern blots.

Southern blot analysis of chromosomal DNA digested using *SstI*

Mutant	Group F 9.3 kb	Group G 14 kb	Group H 18 kb	Group I 20 kb	Group J 21 kb	Group K 23 kb
F1				x		
F2				x		
F3	x					
F4				x		
F6				x		
F9		x				
F10						x
F11				x		
F12				x		
F14			x			
F15				x		
F19				x		
F20					x	
F22						x

Table 6-2. Compilation of Southern blot data from *SstI* digestion of shedder mutant chromosomal DNA. *SstI* does not cut Tn5. The Southern blots were probed with a fragment of Tn5. Mutants are grouped according the band size seen on the Southern blot.

Half of the Tn5 and associated chromosomal DNA from a representative of each of these 16 groups and F23-F26 was cloned by one of two methods. The majority of Tn5 insertions were cloned by cutting the chromosomal DNA with *Bam*HI. This cuts the Tn5 in half, but leaves the kanamycin resistance gene intact. This DNA was ligated into a pUC-based vector and selected on kanamycin. This gives an insert with Tn5 sequences on one side and chromosomal DNA on the other. A few mutants proved resistant to this technique and were cloned using an inverse PCR method, developed by V. Martin (Martin and Mohn, 1999). Sequencing off the end of the Tn5 revealed the insertion site of the Tn5 and this sequence was used to search the partial TIGR *C. crescentus* genome library for the DNA surrounding the Tn5 insertion site. All of the Tn5 insertion sites were found in the partial genome

sequence. Open reading frames (ORFs) were determined using the sequence from the partial genome and analyzed for *C. crescentus* codon preference. These ORFs were used to search the known protein databases for similar proteins using the BLAST algorithm (Altschul *et al.*, 1990). The genes[2] interrupted by the Tn5 insertions were characterised using this data (Table 6-3).

Tn5 mutant group	Similarity to known proteins	Location*	ORF designation
F1, F7	regulator and transcription repressor LacI	gcc 433	*lpsI*
F2	perosamine synthetase, RfbE – *V. cholerae*	RAT1	*lpsC*
F3	nucleotide sugar epimerase/dehydratase	gcc 1444	*lpsK*
F4, F5	similarity to mannosyl transferase WbaZ - *E. coli*	RAT1	*lpsD*
F6	methyl-accepting chemotaxis receptor	gcc 648	orf1
F9, F13, F17	Phosphomannomutase, RfbB – *V. cholerae*	gcc 227	*lpsG*
F10	none-downstream of kpsT-like ORF (O-antigen transporter)	gcc 279	orf2
F11	similarity to mannosyl transferase (rfb region)	gcc 973	*lpsE*
F12	similarity to mannosyl transferase WbaZ from *E. coli*	RAT1	*lpsD*
F14, F16	mannose–6-phosphate isomerase	gcc 506	*lpsH*
F15, F18	similarity to mannosyl transferase WbaZ from *E. coli*	RAT1	*lpsD*
F19	similarity to mannosyl transferase WbaZ from *E. coli*	RAT1	*lpsD*
F20	similarity to mannosyl transferases	gcc 395	*lpsF*
F22	none-downstream of kpsT-like ORF (O-antigen transporter)	gcc 1290	orf2
F23	Phosphomannomutase	gcc 227	*lpsG*
F24	galactosyl-1-phosphate transferase, WlaH *C. jejuni*	gcc 2537	*lpsJ*
F25	mannose–6-phosphate isomerase	gcc 506	*lpsH*
F26	Rhamnosyl transferase	gcc 2218	*lpsL*

Table 6-3. List of shedder mutants. ORFs with similarity to sugar modification enzymes have been given an *lps* designation. * Location gives either the contig (gcc) found in the partial *Caulobacter* genome or shows that the gene was found in the RAT1 fragment 3' of *rsaE* and had been sequenced while looking for the third translocator protein, RsaF.

[2] For clarity the ORFs will be referred to as genes and the corresponding deduced protein sequences as proteins even though it is acknowledged that neither assumption has been proven.

The S-LPS synthesis genes are genetically linked to the RsaA transport genes. Analysis of the DNA sequence around the *rsaA* transporter complex (see Ch. 3 and Ch. 4) revealed 5 ORFs with coding sequences having significant similarity to S-LPS synthesis enzymes between *rsaE* and *rsaF*(973) and one ORF 3' of *rsaF*(973) was found. The first ORF encoded a protein with similarity to GDP-D-mannose dehydratase (Currie *et al.*, 1995; Stroeher *et al.*, 1995), the second ORF encoded a protein with similarity to UDP-N-acetylglucosamine acyltransferases (Canter Cremers *et al.*, 1989; Vuorio *et al.*, 1994) and the third protein had similarity to perosamine synthetase (Bik *et al.*, 1996; Stroeher *et al.*, 1995). The fourth and fifth proteins have similarities to mannosyltransferases (Drummelsmith and Whitfield, 1999; Rocchetta *et al.*, 1998). These five ORFs have been designated *lpsA, lpsB, lpsC, lpsD* and *lpsE* (Fig. 6-4). Another ORF, *lpsF*, was found 3' of *rsaF*(973), and also had similarity to glycosyl transferases (Kido *et al.*, 1998).

Since the S-LPS is required for attachment of the S-layer, it is not that surprising that some of the genes involved in S-LPS synthesis are physically near *rsaA* and the transport genes. Smooth LPS genes have also been implicated in the proper formation of the transport complex in some type I secretion signals (Wandersman and Létoffé, 1993). It is thought that smooth LPS is required for proper insertion of the OMP into the outer membrane. Sequencing of the Tn5 insertions in the shedders has shown that F2 is located within *lpsC* and the four different insertions F4, F12, F15, and F19 are located within *lpsD*. The presence of four different Tn5 mutations in *lpsD* suggests that the Tn5 mutations are the cause of the shedding phenotype and this gene plays a role in S-LPS synthesis. In addition, F11 is found in *lpsE* and F20 is found in *lpsF*.

Most of the remaining Tn5 insertions are also in genes that have similarity to smooth LPS synthesis genes (Fig. 6-5, Table 6-4). Two of these insertions interrupt genes with similarity to glycosyltransferases. Four Tn5 insertions are found in genes that have been implicated in pathways for the production of GDP-4-keto-6-D-deoxymannose, a precursor of GDP-L-fucose and GDP-perosamine. One insertion appears in a gene with similarity to transcription regulators. Two other insertions are in unknown genes.

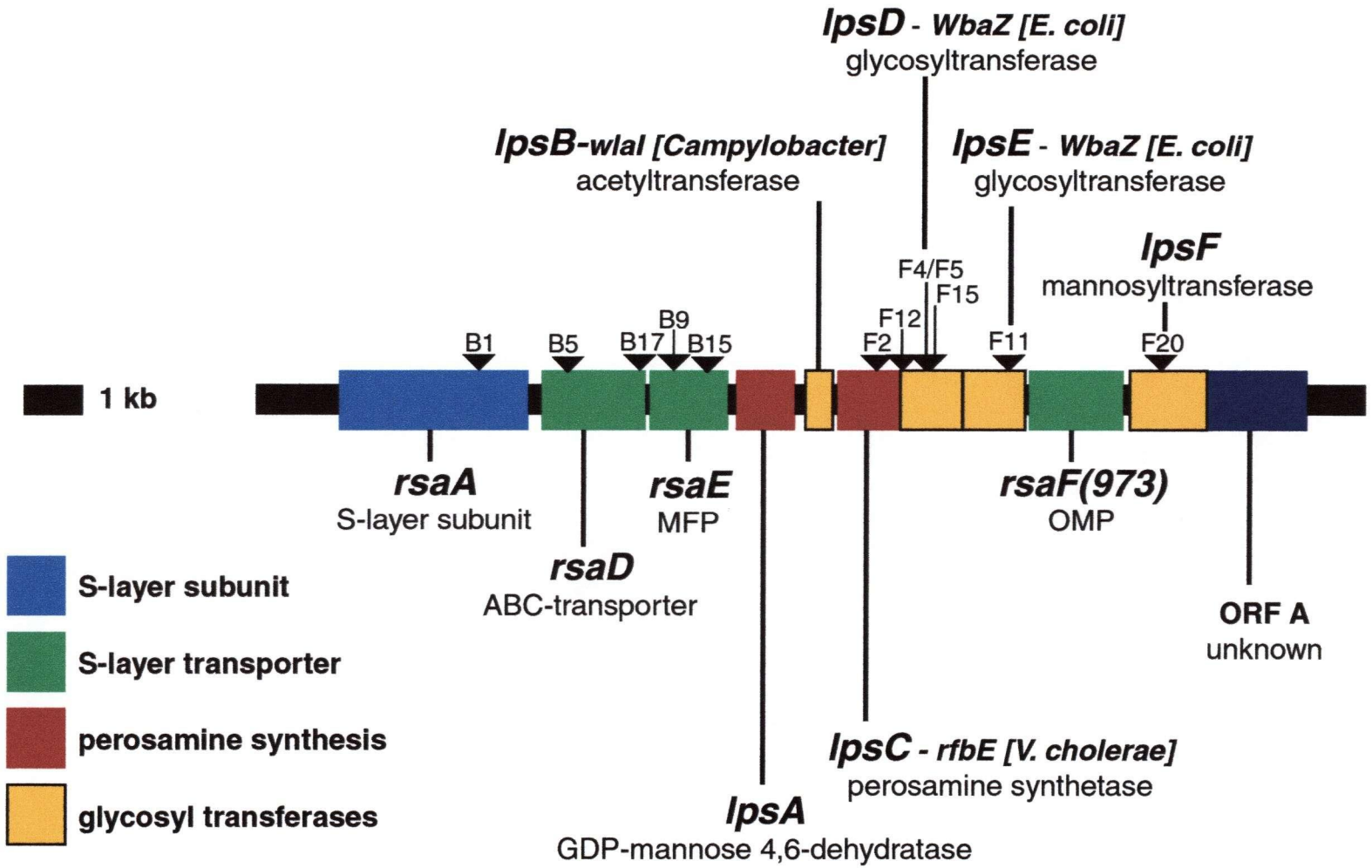

Figure 6–4. **S-LPS synthesis genes linked to *rsaA*.** Boxes represent genes or ORFs. Protein functions have been assigned on the basis of sequence similarity. Triangles indicate Tn5 insertions. Numbers above the triangles indicate the designation of the Tn5 insertion (B for S-layer negative mutants and F for shedding mutants).

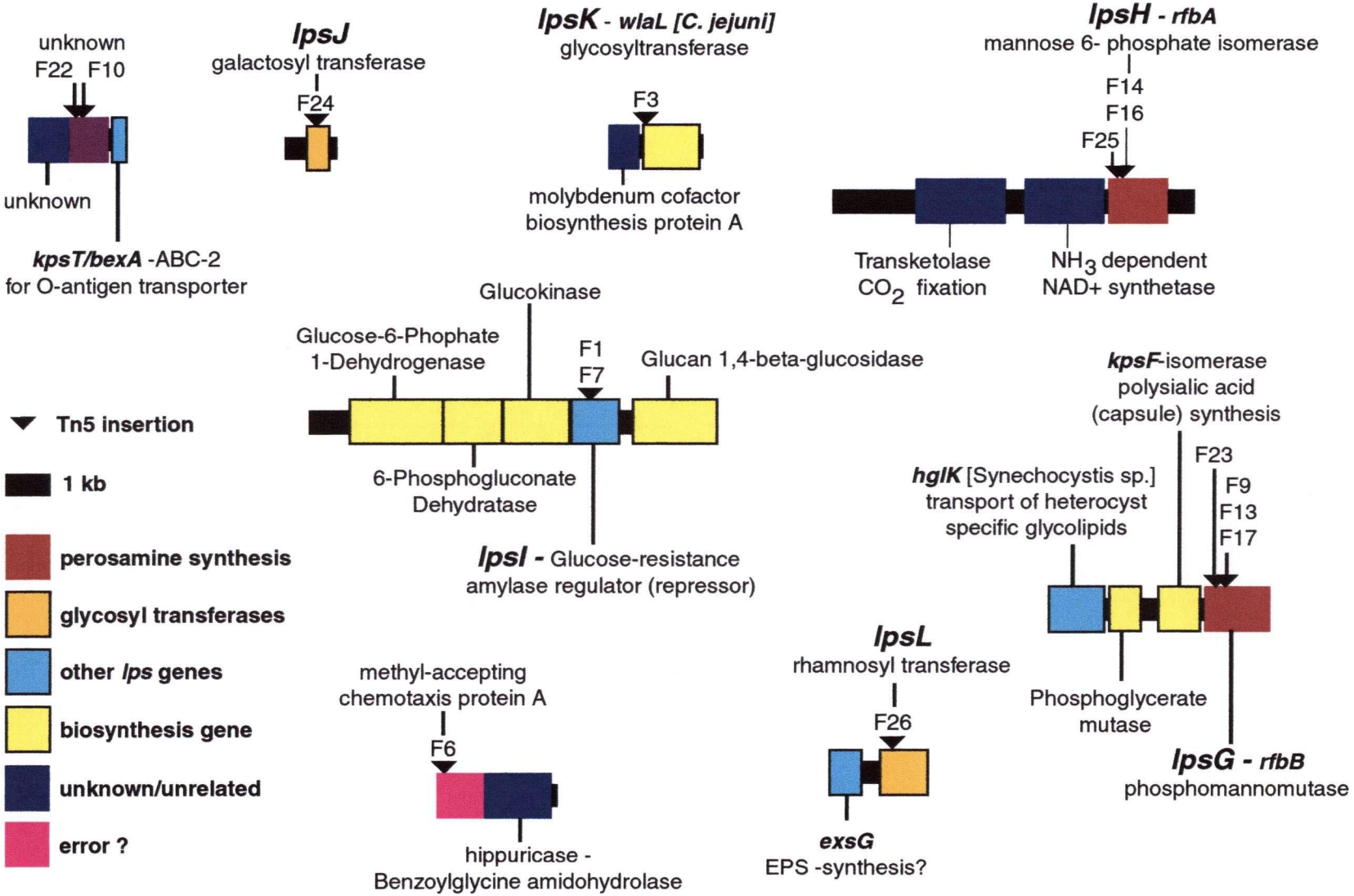

Figure 6-5. Genes interrupted by Tn5 insertions in shedder mutants. Protein functions have been assigned on the basis of sequence similarity. Triangles indicate Tn5 insertion sites causing shedding phenotype. Numbers above the triangles indicate the designation of the Tn5 insertion.

Caulobacter protein	Similar Proteins	Organism	Function	Identity/% Similarity	Accession
LpsA	GCA	*Pseudomonas aeruginosa*	GDP-mannose dehydratase	65.2/88.6	Q51366
	RfbB	*Synechocytis species*	GDP-mannose dehydratase	55.2/83.4	P72586
	GMD	*Escherichia coli*	GDP-mannose dehydratase	55.7/85.0	P32054
	GMD	*Escherichia coli* O157	GDP-mannose dehydratase	55.7/84.9	O85339
LpsB	YvfD	*Bacillus subtilis*	Serine O-acetyltransferase	47.2/83.1	P71063
	WlaI	*Campylobacter jejuni*	Serine O-acetyltransferase	37.9/83.4	O86157
	NeuD	*Escherichia coli*	acetyltransferase	32.4/77.2	Q46674
	WbdR	*Escherichia coli* O157	N-acetyltransferase	30.3/72.2	O85344
LpsC	SpsC	*Synechocytis species*	Spore coat polysaccharide synthesis	50.0/86.1	P73981
	Mth334	*Methanobactium thermoautotropicum*	Perosamine synthetase	46.4/82.4	O26434
	RfbE	*Escherichia coli* O157	Perosamine synthetase	45.4/82.4	O07894
	RfbE	*Vibrio cholerae*	Perosamine synthetase	42.3/80.1	Q06953
LpsD	WbaZ-1	*Archaeoblubus fulgidus*	Mannosyl transferase	24.3/69.8	O30192
	Mth332	*Methanobactium thermoautotropicum*	LPS biosynthesis	24.5/68.6	O26432
	ORF18.9	*Salmonella enterica*	Mannosyl transferase	19.6/62.0	Q00483
	ExpE4	*Sinorhizobium meliloti*		25.0/40.7	P96434
LpsE	ORF18.9	*Salmonella enterica*	Mannosyl transferase	26.5/89.7	Q00483
	WbaZ-2	*Archaeoblubus fulgidus*	Mannosyl transferase	24.5/64.6	O29649
	WbaZ-1	*Methanobactium thermoautotropicum*	Mannosyl transferase	24.2/66.5	O30192
LpsF	WbdA	*Escherichia coli*	Mannosyl transferase	19.4/66.2	O66234
	AF0617	*Archaeoblubus fulgidus*	LPS biosynthesis protein	24.8/69.9	O29638
	Mth370	*Methanobactium thermoautotropicum*	LPS biosynthesis protein, RfbU -like	29.0/65.7	O26470
LpsG	AlgC	*Pseudomonas aeruginosa*	phosphomannomutase	36.0/57.4	P26276
	PGM	*Neisseria gonorrhoeae*	phosphomannomutase	32.9/50.6	P40390
	PmmA	*Mycobacterium*	phosphomannomutase	38.0/54.2	O86374
	PGM	*Neisseria meningitidis*	phosphomannomutase	35.0/53.5	P40391
LpsH	XanB	*Xanthomonas campestris*	Phosphomannose isomerase	38.3/71.7	P29956
	ManC	*Yersinia enterocolitica*	Mannose-1-phosphate guanyltransferase	33.2/64.0	Q56874
	RfbM	*Escherichia coli*	Mannose-1-phosphate guanyltransferase	32.6/65.9	Q59427
LpsI	CcpA	*Bacillus megaterium*	Catabolite control protein	34.9/74.2	P46828
	CcpA	*Bacillus subtilis*	Catabolite control protein	33.1/74.5	P25144
	DegA	*Bacillus subtilis*	Degradation activator	33.1/74.9	P37947
	LacI	*Bacillus subtilis*	LacI repressor like protein	30.0/72.9	O34396
LpsJ	LpsB1	*Rhizobium etli*	galactosyltransferase	59.7/71.0	O34301
	CapM	*Staphylococcus aureus*	unknown	45.7/79.6	P95706
	RfbW	*Vibrio cholerae*	galactosyltransferase	47.2/79.8	Q56624
	PssA	*Rhizobium leguminosarum*	galactosyltransferase	34.6/69.2	Q52856
LpsK*	WlaL	*Campylobacter jejuni*	amino sugar epimerase	43.8/79.6	O86159
	BplL	*Bordetella pertussis*	LPS biosynthesis	31.0/64.4	Q45387
	LpsB2	*Rhizobium etli*	dTDP-glucose 4,6, dehydratase	25.9/39.4	O34302
	CAPD	*Bacillus subtilis*	unknown	26.5/69.9	P72370
LpsL	CPS23FV	*Streptococcus pneumoniae*	Rhamnosyltransferase	29.8/51.7	O86159
	CPS23FI	*Streptococcus pneumoniae*	LPS biosynthesis	29.8/51.7	AAC69532
	ORF51x5	*Vibrio anguillarum*	unknown	26.7/45.0	O31012

Table 6-4. Deduced proteins involved in O-antigen synthesis and their homologues. BLAST and FASTA alignments were used to determine identity and similarity. Percentage similarity represents identical amino acids and conserved substitutions.
* incomplete ORF

As shown by Southern blotting, the Tn5 insertions, F1, F2, F4, F6, F11, F12, F14, F15, F19 and F20 are linked. Figure 6-4 shows that the Tn5 insertions F2, F4, F11, F12, F15 and F20 are linked to the RsaA transporter genes. F1, F6, and F14 must be linked as well, but it was not possible to construct the DNA sequence of this linkage. In addition, of the four mutants not characterised by Southern analysis, F23 is in the same ORF as F9, and F25 is in the same ORF as F14. The other two mutants, F24 and F26, were not obviously linked to any of the other insertions.

Analysis and proposed function of individual proteins involved in S-LPS synthesis.

A total of 14 ORFs associated with the formation of the S-LPS were found (Table 6-4). Four of these ORFs are incomplete. A summary of the characteristics of these ORFs is listed in Table 6-5. All of the ORFs start with an ATG codon except *lpsH* which starts with a TTG. Sequence similarity and codon preference indicate that the TTG is the most probable start codon for *lpsH*. Using the *C. crescentus* promoter consensus for biosynthetic genes (Malakooti *et al.*, 1995), possible promoters were found 31 bp and 99 bp 5' of *lpsG*, 52 bp 5' of *lpsH*, 204 bp 5' of *lpsI*, 154 bp 5' of *lpsJ* and 63 bp 5' of *lpsK*. In some clusters of smooth LPS genes the G+C content of the individual clusters varies with respect to the G+C content of the bacterium suggesting recent acquisition of the genes (Fallarino *et al.*, 1997; Fry *et al.*, 1998; Stroeher *et al.*, 1995). The G+C content of these ORFs is consistent with the average *C. crescentus* content of 67%.

ORF	Translation start	Size (aa)	Predicted mass (kDa)	pI	G+C %
lpsA	TGTTACTGGAGTCAGCGGATACGC**ATG**	325	36.3	6.2	65.1
lpsB	CATCGCGCGGCGCTCGCGCAACTG**ATG**	215	21.4	8.5	69.3
lpsC	GAACGTGACTATGTACTCGAATGC**ATG**	346	37.8	5.9	63.1
lpsD	CTCGATCAGGTGTTGGTCTAGCCG**ATG**	346	39.1	5.7	65.2
lpsE	GCCTGACCTCATGAGAACGCCCGC**ATG**	345	38.2	5.8	65.8
lpsF	GCGTCTCGCCCGCCTGCATCGCCC**ATG**	430	47.0	7.5	69.1
lpsG	CATCTCAACTGAAGCGAGCCTTCA**ATG**	>469	ND	5.0	65.8
lpsH	CCTAAGACTGTGTGGGGACAAGAC**TTG**	434	45.5	4.6	67.4
lpsI	CGGGCTCGCCATGACAGCCTTGTC**ATG**	356	38.7	6.3	65.4
lpsJ	TCTGGCCTAGGCCGAGCCGGCTGA**ATG**	187	20.5	10.5	66.0
lpsK	TTCACCGCTTCAGAGGTTCGTTTC**ATG**	>459	ND	10.4	69.8
*lpsL**	ND	>336	ND	5.5	68.7
orf1*	ND	>352	ND	6.4	73.8
orf2	GGCCTACCGCGAAACCCAGGCCGC**ATG**	316	34.1	10.2	72.5

Table 6-5. Characteristics of the putative S-LPS synthesis genes. Start codons are in bold. Putative Shine-Dalgarno sequences are underlined. * - incomplete ORF. ND – not determined because ORF is incomplete.

LpsA resembles GDP-mannose 4,6-dehydratases. The start codon for *lpsA* is 143 bp 3' of *rsaE*. No promoter matching the consensus sequence was found upstream of *lpsA*, as would be expected if there is a terminator after *rsaE* (see Ch.3). The LpsA sequence has up to 65.2% identity and 88.6% similarity over its entire length to GDP-mannose 4,6-dehydratases from *P. aeruginosa* and *E. coli*. (Table 6-4). These enzymes convert GDP-mannose to GDP-4-keto-6-deoxymannose (Stevenson *et al.*, 1996) as part of biosynthetic pathways polysaccharides. One example of this is the synthesis of perosamine in *V. cholerae* and *E. coli* O157. The significant similarity to GDP-mannose 4,6-dehydratases suggests that this is also the function of LpsA, although no Tn5 insertion was found in the gene.

LpsB is similar to N-acetyltransferases. The gene *lpsB* follows *lpsA* by 2 bp suggesting that these genes are transcriptionally coupled. The protein encoded by the gene shows significant similarity to WlaI from *C. jejuni* and NeuD from *E. coli* (Table 6-4). WlaI is involved in the synthesis of the O-antigen (Fry *et al.*, 1998) while

the function of NeuD is not clear, but is thought to be involved in NeuNAc transfer (Annunziato *et al.*, 1995). These proteins also show some similarity to the LpxA genes from *E. coli* and *S. enterica*. The LpxA proteins are UDP-N-acetylglucosamine O-acetyltransferases that are involved in the first step of Lipid A biosynthesis and have 24 to 26 unique hexapeptide motifs starting with an isoleucine, leucine or valine residue often followed by a glycine (Vaara, 1992; Vuorio *et al.*, 1994). LpsB, WlaI and NeuD contain several of these hexapeptide repeats (Fig. 6-6). The protein WbdR from *E. coli* O157 also contains these hexapeptide repeats and has 72.2% sequence similarity to LpsB. WbdR is thought to encode an N-acetyltransferase which converts GDP-perosamine to GDP-N-acetyl perosamine (Wang and Reeves, 1998). Since the data in this chapter suggest that the genes involved in perosamine synthesis in *E. coli* O157 are also present in *C. crescentus* LpsB may acetylate GDP-perosamine like WbdR.

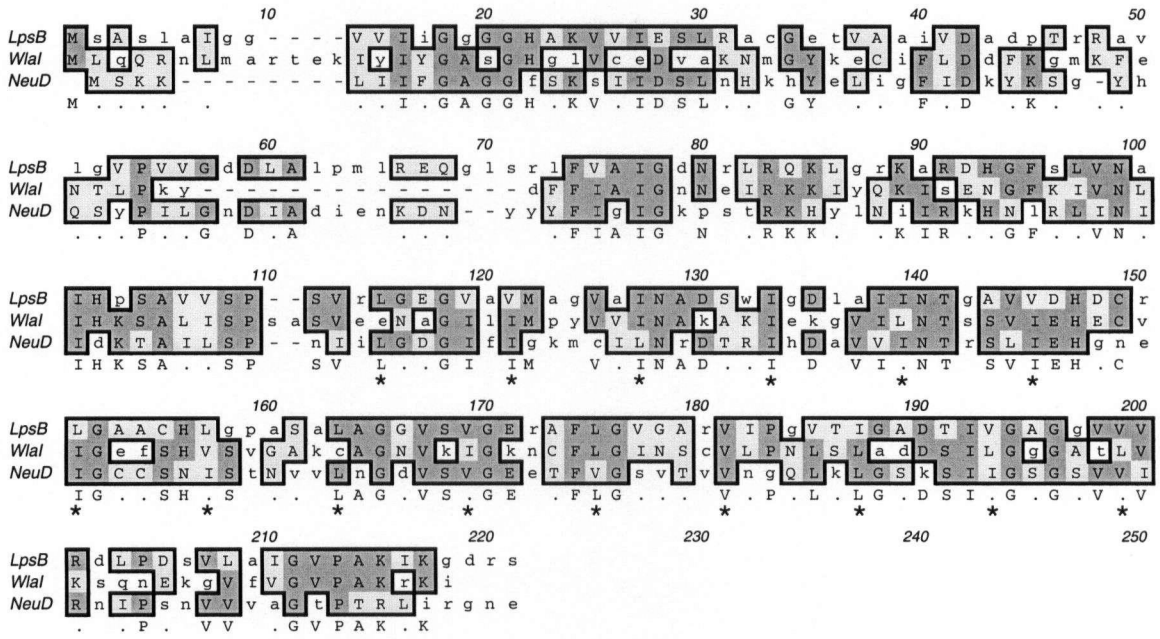

Figure 6-6. ClustalW alignment LpsB. Alignment of LpsB with WlaI from *C. jejuni* (Accession CAA72358) and NeuD from *E. coli* (ACC43301). Asterisks mark the hexapeptide motifs found in glycosyl transferase. Identical and similar residues are boxed.

LpsC appears to be a perosamine synthetase. The gene encoding LpsC starts 74 bp 3' of *lpsB*, but no promoter sequence was found between *lpsB* and *lpsC*. LpsC has considerable identity over its entire length to the *rfbE* and *per* gene products that are thought to synthesize perosamine (Table 6-4). These proteins likely catalyze the conversion of GDP-4-keto-6-D-deoxymannose to GDP-perosamine (4-amino-4,6-dideoxymannose) in *V. cholerae* and *E. coli* O157 (Stroeher *et al.*, 1995; Wang and Reeves, 1998) and show similarity to two classes of pyridoxal-binding proteins involved in the synthesis of amino sugars similar to perosamine. The perosamine synthetic pathway has not been proven chemically, but the proteins suspected in the synthesis of perosamine are the only highly similar proteins involved in O-antigen synthesis found in common between *Vibrio cholerae*, and *E. coli* O157 supporting these predictions (Wang and Reeves, 1998). Based on the similarity to these genes, it is likely that LpsC is a perosamine synthetase.

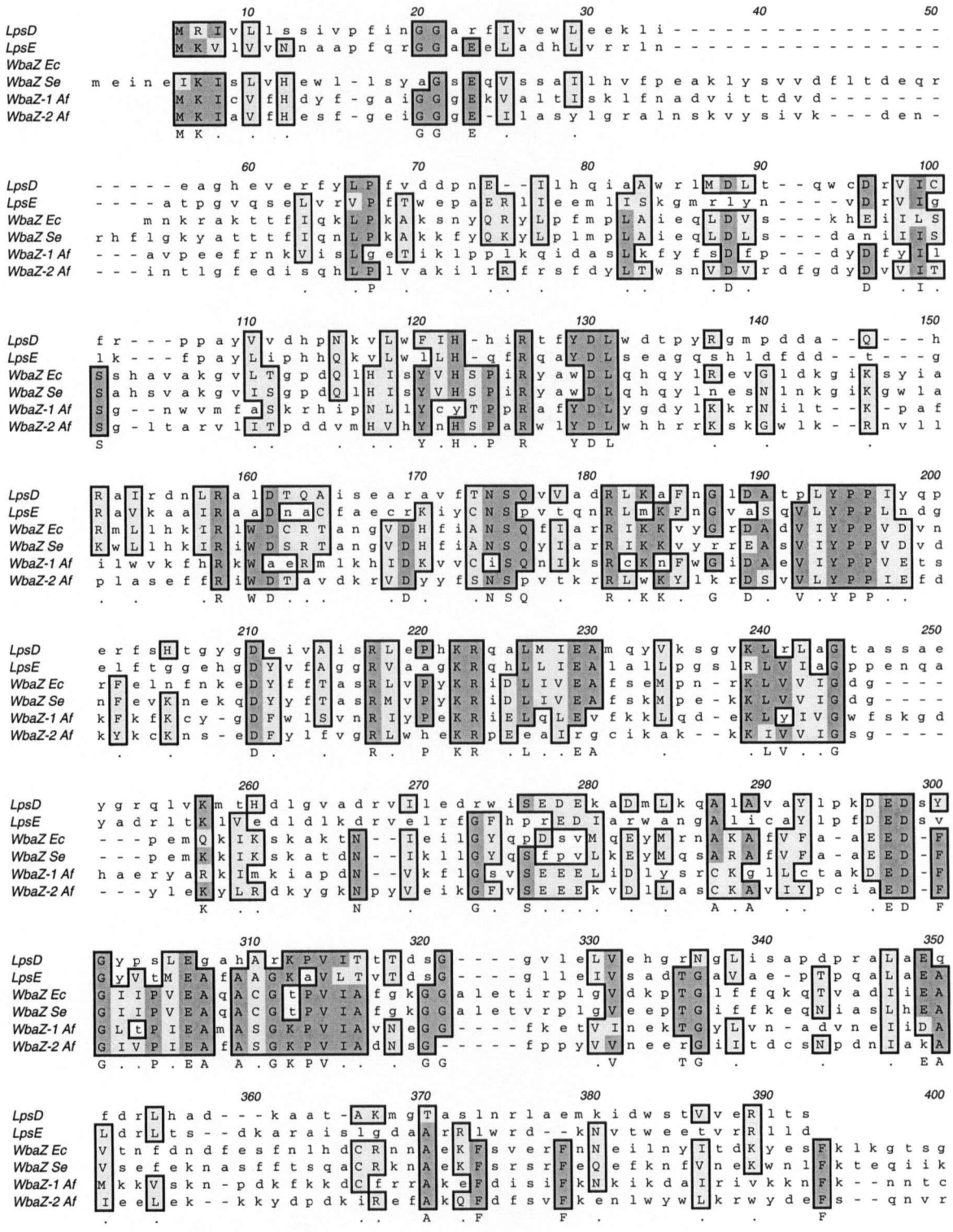

Figure 6-7. ClustalW Alignment of LpsD and LpsE with WbaZ genes from *E. coli* (Accession AAD21571) and *S. enterica* (X61917) and WbaZ homologues from *A. fulgidus* (AAB91187). Identical and similar residues are boxed. Identical residues have dark shading. Similar residues have light shading. The consensus sequence is located below the alignment.

LpsD and LpsE resemble glycosyltransferases. The gene for LpsD follows *lpsC* by 6 bp and the gene for LpsE follows *lpsD* by 13 bp, suggesting that all three genes are part of a polycistron. Both LpsD and LpsE have significant similarity to the WbaZ proteins (Fig 6-7). These proteins also have similarity to the RfbU related proteins, but size and amino acid similarity indicates that the WbaZ-like protein are a separate family. WbaZ is a known mannosyltransferase in *S. enterica* (Liu *et al.*, 1993). It seems likely that LpsD and LpsE function to link perosamine monomers to the O-antigen with each providing a different form of linkage.

LpsF is similar to perosamine transferases. The gene for LpsF is separated from *lpsABCDE* by *rsaF* and is transcribed in the opposite orientation. LpsF, like LpsD and LpsE, appears to be a mannosyltransferase, but has greater similarity to the RfbU family. The similarity to mannosyltransferases is much less than that seen with LpsD and LpsE, but it does have significant similarity to the C-terminal of *E. coli* mannosyltransferases, WbdB and WbdA (Kido *et al.*, 1998; Sugiyama *et al.*, 1998) and RfbU, from *V. cholerae* (Wang and Reeves, 1998). RfbU, from *V. cholerae,* is known to transfer a perosamine residue onto the growing O-antigen chain. These proteins contain a signature motif that is also found in LpsF (Fig 6-8). This motif consists of the sequence EX[XF]GXXXXE[AG] with a serine preceding the motif by 3 to 5 residues (Geremia *et al.*, 1996; Rocchetta *et al.*, 1998). Again, it seems likely that LpsF acts to add perosamine residues onto the O-antigen.

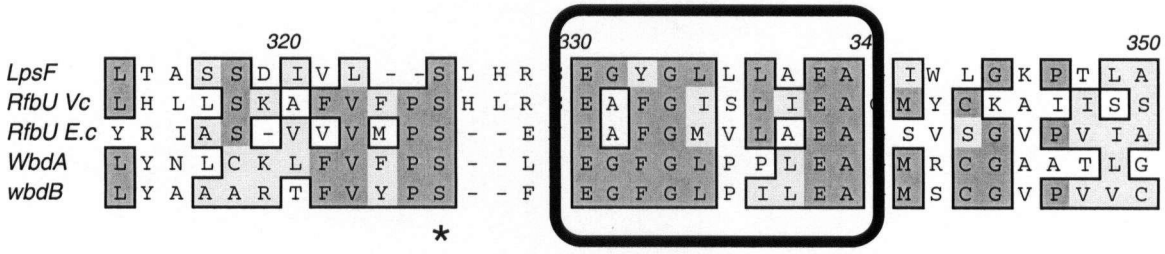

Figure 6-8. ClustalW alignment of LpsF with a number of known mannosyl transferases. The mannosyl transferase motif is boxed. The conserved serine is marked with *. RfbU -*Vibrio cholerae* (Accession Y07788), RfbU – *E. coli* (BAA31838), WbdA, WbdB – *E. coli* (D43637). Identical and similar residues are boxed. Identical residues have dark shading. Similar residues have light shading. The consensus sequence is located below the alignment.

LpsG is similar to phosphomannomutases. Two Tn5 insertions mutants had interrupted LpsG genes. The LpsG gene does not appear to be linked to any of the other *lps* genes (Table 6-1 and Table 6-2). This protein has very high identity along its entire length to a number of phosphomannomutase enzymes suggesting that this is the function of LpsG (Table 6-4). Phosphomannomutase converts mannose-6-phosphate to mannose-1-phosphate and is one of the enzymes implicated in perosamine synthesis (Stroeher *et al.*, 1995; Wang and Reeves, 1998).

LpsH may have a dual function as a phosphomannoisomerase and mannose-1-phosphate guanyltransferase. Two shedder mutants have Tn5 insertions within *lpsH* that result in loss of proper O-antigen production. It was not possible to link this gene with the RsaA transport genes using the TIGR *Caulobacter* genome sequence, but Southern analysis showed that *lpsH* is linked (Table 6-1 and Table 6-2). LpsH has significant identity over its entire length to a large family of enzymes that have dual functions as a phosphomannoisomerase and mannose-1-phosphate guanyltransferase (Table 6-4) Both functions are required for the synthesis of perosamine (Stroeher *et al.*, 1995) and are probably also performed by LpsH in *C. crescentus.* These functions are split up in *E. coli* O157 into the *manA* and *manC* genes (Wang and Reeves, 1998).

LpsI has similarity to the LacI repressor family. The Tn5 insertion in mutant F1 interrupts *lpsI.* Southern blot analysis indicated that this insertion is linked to the Rsa locus. This insertion has a different phenotype than every other shedder Tn5 insertion. Analysis of the O-antigen by SDS-PAGE and silver staining reveals that a lower amount of O-antigen is produced by this mutant. Analysis of LpsI indicates that the highest degree of identity is with CcpA, the catabolite control protein in *Bacillus subtilis.* CcpA represses carbohydrate utilization enzymes such as α-amylase and acetyl coenzyme A synthetase and has a positive regulatory affect on excess carbon excretion proteins such as acetate kinase (Henkin *et al.*, 1991). Lower sequence identity is found to a number of LacI repressor-like proteins (Table 6-4). Analysis of the genes adjacent to *lpsI* revealed the presence of analogues of

glucokinase, 6-phosphogluconate dehydratase and glucose-6-phosphate-1-dehydrogenase enzymes involved in basic metabolic pathways. This positioning suggests that LpsI may regulate the transcription of these genes. If LpsI has a repressor effect on these enzymes it could slow the production of O-antigen as glucose-6-phosphate would tend not be shunted into the perosamine synthetic pathway. Instead, it would be used for energy production in central metabolism.

LpsJ is similar to galactosyl transferases. The Tn5 insertion F24 interrupts a gene with sequence similarity to several galactosyl transferases (Fig. 6-9). These enzymes appear to transfer the first sugar residue (usually a galactose) to undecaprenol phosphate, the lipid precursor. RfbW is one of these enzymes and its

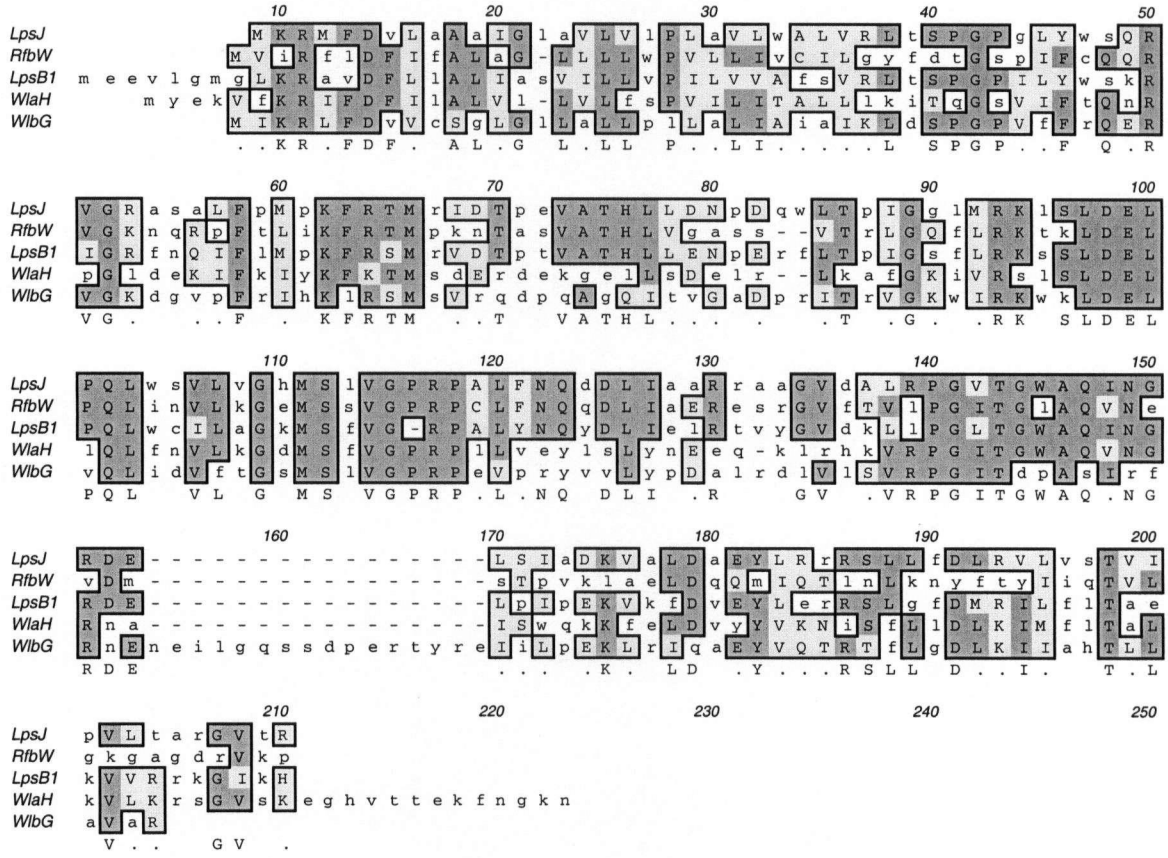

Figure 6-9. ClustalW alignment of LpsJ with putative galactosyltransferases. RfbW-*V. cholerae* (Accession Y07788), LpsB1-*R.etli* (U56723), WlaH-*C. jejuni* (CAA72357), WblG - *Bordetella pertussis* (X90711). Identical and similar residues are boxed. Identical residues have dark shading. Similar residues have light shading. The consensus sequence is located below the alignment.

sequence is 47.2% identical and 79.8% similar to LpsJ over 144 amino acids. RfbW is involved in the synthesis of the perosamine homopolymer making up the O-antigen of *V. cholerae* O1 (Fallarino *et al.*, 1997) suggesting that RfbW may transfer the first perosamine to the lipid precursor. In *C. crescentus*, LpsJ may initiate the formation of the O-antigen by attaching the first sugar residue (probably a perosamine) to the undecaprenol phosphate.

LpsK has sequence similarity to amino sugar synthesis enzymes. The mutant, F3, has an interruption in *lpsK*. It was only possible to determine the sequence for the 5' end of *lpsK* from the TIGR genome. The partial sequence of LpsK is similar to a number of large proteins, usually consisting of over 600 amino acids, suggesting that approximately 150 amino acids are missing from the C-terminal of the LpsK coding sequence (Fig 6-10). There is still considerable similarity, especially in the middle of the protein, to WlaL, RfbV and WlbL from *C. jejuni, V. cholerae* O1 and *B. pertussis*. These proteins contain 5 hydrophobic, predicted transmembrane domains in the N-terminus. The central portion contains an NAD-binding site and is homologous to UDP-glucose-4-epimerases. Two motifs have been implicated in binding of NAD in these proteins, GXGXXG and GAGGSIG (Fallarino *et al.*, 1997). As seen in Fig 6-10, the second motif is found in all the proteins, but the first only occurs in RfbV and WlbL suggesting that not all members of this family contain this motif. The C-terminal 300 amino acids of these proteins have identity with dTDP-glucose 4,6-hydratases (Bechthold *et al.*, 1995; Linton *et al.*, 1995). These proteins are usually associated with synthesizing amino 6-deoxy and dideoxy sugars involved in LPS synthesis or extracellular polysaccharides and probably perform multiple functions to account for the 3 domains. LpsK was not found linked to the other O-antigen synthesis genes. This may indicate that LpsK is involved in the synthesis of a core sugar, possibly the terminal core sugar. Interruption of this gene may prevent attachment of the O-antigen to the core, resulting in the observed shedding phenotype.

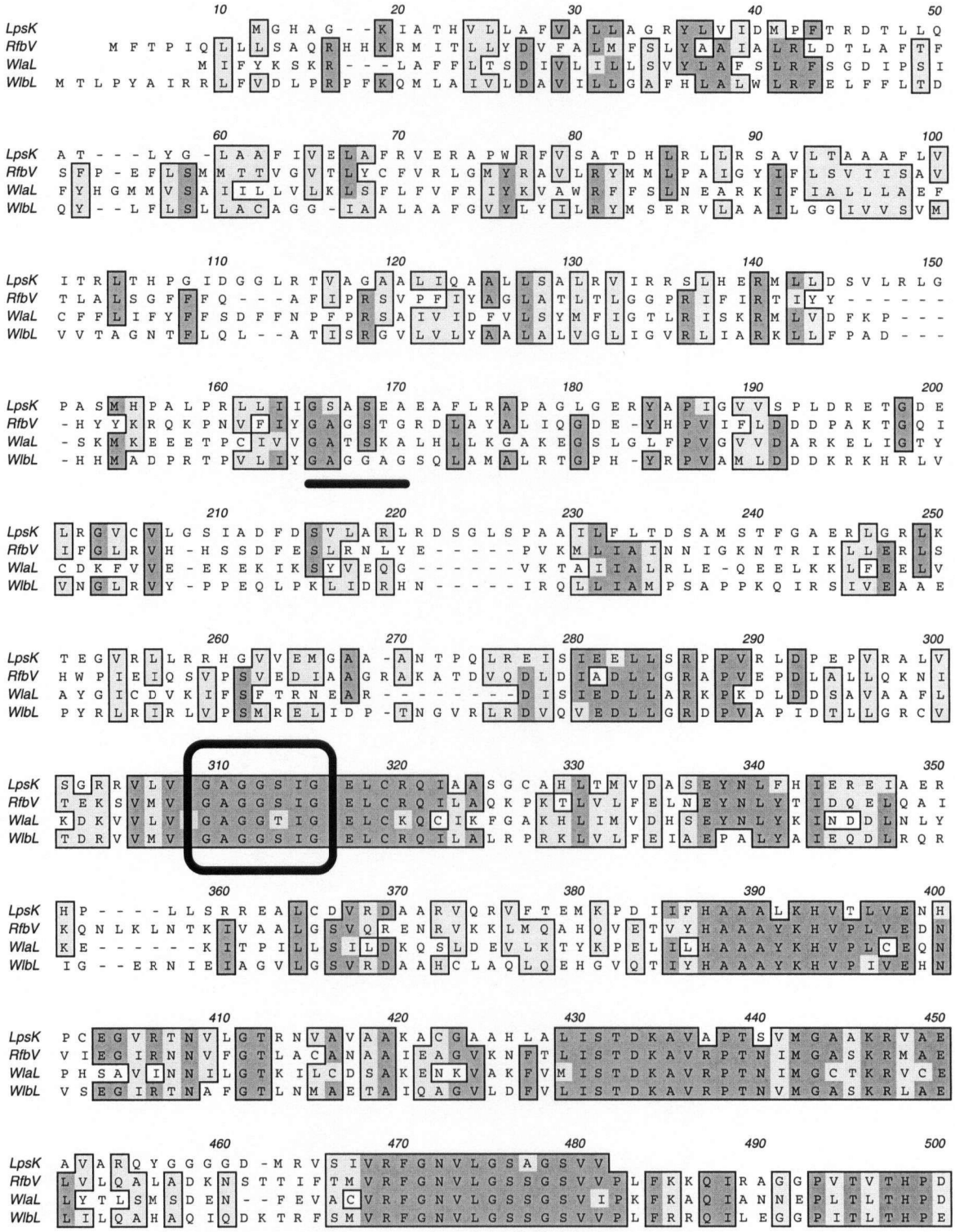

Figure 6–10. ClustalW alignment of LpsK. The first NAD binding motif is underlined. The second NAD motif is boxed. Only RfbV and WlbL contain the first motif. Only a partial sequence of LpsK has been deduced and the alignment is truncated after the LpsK sequence. RfbV - *V. cholerae* (Accession Y07788), WlaL – *C. jejuni* (CAA72360), WlbL – *B. pertussis* (X90711)

LpsL may be a glycosyltransferase. The mutant F26 has an insertion in *lpsL*. This gene is 5' to an ORF with similarity to *exsG* which was implicated in extracellular polysaccharide synthesis (Becker *et al.*, 1995). The LpsL amino acid sequence is 29.8% identical and 51.7% similar over a range of 87 amino acids to a putative rhamnosyl transferase in *Streptococcus pneumoniae* (Table 6-4). Rhamnose is a 6-deoxy derivative of mannose, as is perosamine, suggesting that LpsL may be another perosamine transferase.

The functions of some of the Tn5-interrupted genes are still unidentified. The Tn5 insertions F22 and F10 interrupt an ORF with no identity to any known protein. But 5' of this ORF is an ORF corresponding to an ABC-2 transporter. These transporters are known to transport extra-cellular polysaccharides and O-antigens through the cytoplasmic membranes (Whitfield, 1995). Unlike the ABC transporters of the type I secretion systems, the ABC and transmembrane domains consist of separate proteins. It is possible that the ORF interrupted by F10 and F22 represents the transmembrane protein part of the ABC-2 transporter, but hydropathy analysis does not suggest that this protein contains transmembrane segments. The ABC-2 transporters are often found adjacent to genes involved in polysaccharide synthesis, therefore it may be that the ORF interrupted by the F10 and F22 mutants is also involved in polysaccharide synthesis.

The Tn5 insertion F6 interrupts orf1 which has similarity to a chemotaxis receptor (Ward *et al.*, 1995). CheY, a chemotaxis regulator, is found linked to a number of O-antigen synthesis genes with similarity to *lpsJ, lpsB, lpsC and lpsK* in *C. jejuni.* It may be that the genes involved in chemotaxis are found close to the O-antigen synthesis genes in *C. crescentus* and that the F6 insertion has a polar effect on downstream S-LPS genes. It is also possible that this ORF has nothing to do with LPS synthesis and the Tn5 insertion may not cause the shedding phenotype. Instead, a second mutation may cause the altered phenotype.

Summary

As stated at the beginning of the chapter, it seems likely that the S-LPS of *C. crescentus* is a fixed length homopolymer of approximately forty 4,6-dideoxy-4-amino-hexose residues. Proton NMR anomeric traces suggest that the linkages between the hexose residues may not all be identical. Several of the genes discussed in this chapter are similar to genes found in the synthesis of perosamine in *V. cholerae* and *E. coli* O157 (Stroeher *et al.*, 1995; Wang and Reeves, 1998) and as perosamine is a 4,6-dideoxy-4-amino-hexose, it seems likely that the O-antigen of *C. crescentus* consists of perosamine residues. All of the enzymes responsible for perosamine synthesis can be found in the *lps* genes listed above. Four enzymes are involved in converting fructose-6-phospate to perosamine (Fig. 6-11). The first enzyme in the pathway described by Stroeher *et al* (1995) is a phosphomannoisomerase, RfbA. Mutants F25 and F14 are located in *LpsH* which has significant similarity to RfbA. The second step in the pathway is performed by the enzyme RfbB, a phosphomannomutase. Two Tn5 mutants, F9 and F23, are in the gene for LpsG, an enzyme with considerable similarity to RfbB. The third step in the pathway is catalyzed by RfbA. RfbD, a GDP-mannose 4,6-dehydratase, catalyses the fourth reaction. No Tn5 insertion has been found in a gene with similarity to RfbD, but the coding sequence of *C. crescentus* gene immediately 3' of *rsaE*, *lpsA*, shows considerable similarity to RfbD. The last step of the process requires RfbE, the perosamine synthetase. LpsC presumably fulfills this role in *C. crescentus*, and the shedding mutant F2 has a Tn5 insertion in the LpsC gene.

Two more genes need to be considered as part of the perosamine pathway in *C. crescentus* (Fig. 6-11). Bacteria using the Embden-Meyerhof-Parnas pathway require phosphoglucoisomerase as part of the pathway leading into the bottom half of glycolysis, but *C. crescentus* uses the Entner-Doudoroff glycolytic pathway instead (Riley and Kolodziej, 1976) and as such would not be expected to normally have the enzyme phosphoglucoisomerase for converting glucose-6-phosphate to fructose-6-phosphate. But *C. crescentus* requires phosphoglucoisomerase if it makes perosmaine by the pathway described here (Fig 6-11). None of the Tn5 hits were found in such a gene, so the TIGR *Caulobacter* genome was searched for a phosphoglucoisomerase analogue and one was found in contig gcc_2205. A

Perosamine Biosynthetic Pathway

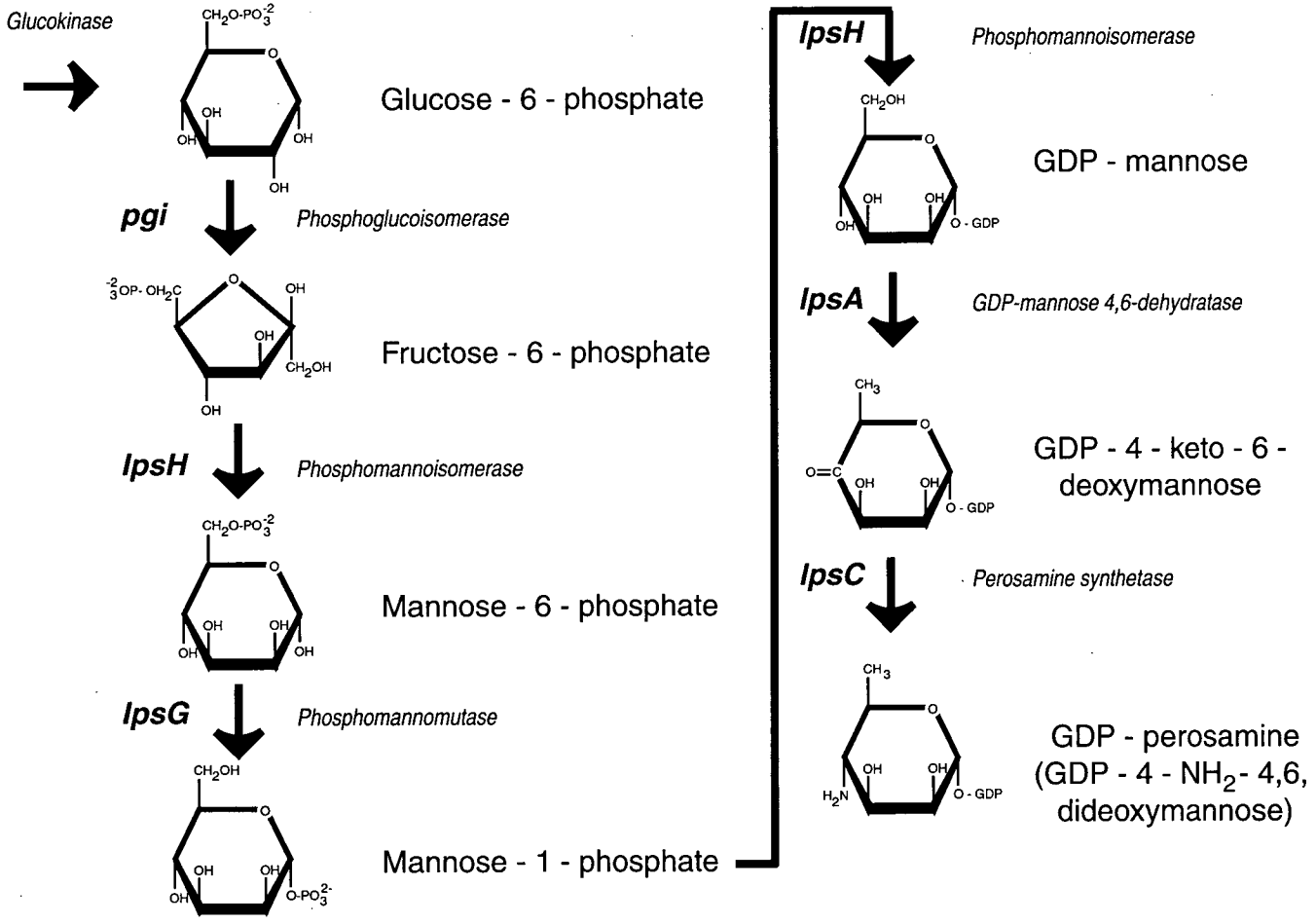

Figure 6-11. Perosamine synthesis pathway [Adapted from Stroeher et al. 1995]. The enzyme pgi is required for this pathway but is not normally found in species using the Entner-Doudoroff pathway. pgi has been found in the TIGR *Caulobacter* genome. Glucokinase is also required by this pathway and its gene is found adjacent to the F1 Tn5 insertion which may act as a repressor on its synthesis.

second enzyme, glucokinase, is required for converting glucose to glucose-6-phophate. A glucokinase analogue was found next to the F1 Tn5 insertion in the potential repressor *lpsI*. From the position of *lpsI* may be deduced that LpsI has a regulatory effect on the synthesis of glucokinase. Interruption of LpsI by the F1 insertion may alter the expression of glucokinase, which in turn would affect perosamine synthesis, resulting in the phenotype seen in the F1 mutant (less O-antigen). These data suggest that *C. crescentus* contains all the genes necessary for the synthesis of perosamine. Furthermore, 5 separate Tn5 insertions in 3 of the ORFs cause loss of O-antigen synthesis, strengthening the argument that perosamine makes up the O-antigen of the S-LPS.

Six of the Tn5 insertions appear to be in glycosyltransferases (*lpsD, lpsE, lpsF*, *lpsJ*, and *lpsL*) (Fig. 6-4). This is would be expected since proton NMR suggests there are a number of different linkages between the sugars in the O-antigen. The similarities of LpsJ to galactosyltransferases, which transfer the initial sugar to the lipid precursor, suggest that LpsJ may initiate the first addition of a sugar to the undecaprenol phosphate. The S-LPS chemical composition suggests that this first sugar is a perosamine, but it is possible that it is galactose. Galactose is found in the core and it is possible that traces found during analysis of the O-antigen would be attributed to contamination from the core.

LpsK may be involved in the synthesis of a sugar residue. As all the enzymes for the synthesis of perosamine are accounted for in the other *lps* genes, LpsK may synthesize an unidentified sugar in the O-antigen (possibly an initial galactose linked by LpsJ) or a sugar in the LPS core.

O-antigens are elongated at either the reducing terminus or the non-reducing terminus. If the O-antigen elongates at the reducing terminus, individual sugars are 'flipped' across the cytoplasmic membrane by a flippase enzyme and the O-antigen is assembled in the periplasm. If synthesis of the O-antigen occurs at the non-reducing terminus, the chain elongates in the cytoplasm and an ABC-2 transporter is required to transport the O-antigen chain across the cytoplasmic membrane (Whitfield, 1995). If the ABC-2 transporter upstream of the F10 and F22 insertions is involved in the transport of the O-antigen, it suggests that the O-antigen is elongated by polymerization at the non-reducing terminus. The O-antigen would then be

transported through the cytoplasmic membrane by the ABC-2 transporter where it would then be transferred to the lipid-A core.

While it has not been proven that any of the ORFs listed here are required for O-antigen synthesis, the presence of multiple Tn5 insertions in some of the ORFs confirms that the Tn5 is responsible for causing the defective S-LPS phenotype and the interrupted ORF is very likely a gene involved in making the S-LPS.

Chapter 7
Conclusions and Future Considerations

The attachment and secretion of the S-layer appear to be linked, although RsaA can be secreted even when the S-LPS is defective and the S-layer cannot attach to the surface. While searching for the secretion components, genes involved in the synthesis and assembly of the S-LPS were found linked to the transport complex. In prokaryotes, genetic linkage often implies linkage of the function. In this case, the most obvious link is that the S-LPS is required for attachment of the S-layer. Since *C. crescentus* is a non-pathogenic bacterium, the only apparent function for the S-LPS is to allow attachment of the S-layer to the outer membrane. As such, it seems likely that the bacterium coordinates production of the S-layer and S-LPS and that clustering of the genes allows better control. Similar linkages between the S-LPS and S-layer translocation have been found in *Acinetobacter* sp. and *Aeromonas salmonicida* (Belland and Trust, 1985; Thorne *et al.*, 1976). A linkage between type I secretion systems and S-LPS has also been found. Three genes involved in the synthesis of the smooth LPS have also been implicated in the secretion of α-hemolysin from *E. coli* (Stanley *et al.*, 1993; Wandersman and Létoffé, 1993). It is suspected that these genes are required for the proper insertion of the OMP component in the outer membrane.

RsaA is secreted by a type I secretion mechanism. All three main components of this system have been found and all are linked to the *rsaA* gene although the OMP gene is separated from the others by 5 kb. These genes are similar to a number of other type I secretion mechanisms. The highest similarity was found to systems secreting proteases and lipases from *P. aeruginosa, E. chrysanthemi* and *S. marcescens.* The identity between these systems is high enough that the proteases, AprA and PrtB, were successfully secreted by the RsaA transport machinery. The genetic arrangement of the RsaA transporter genes is unusual. Typically, either all three genes are on either side of the substrate gene or the OMP gene is unlinked to the rest of the genes. In the RsaA transport system, 5 genes are found between the MFP and the OMP, an arrangement that has not been found

before. These 5 genes appear to be required for the synthesis of the O-antigen.

Another unusual finding was the presence of a homologous ORF of the OMP component found elsewhere in the genome. This homologue has 60% identity to *rsaF*, but is not required for the secretion of RsaA. The function of this homologue remains to be discovered or even if the gene produces a functional protein.

RsaA accounts for a large portion of the cellular protein (10 to 12%). As far as can be determined, the RsaA secretion machinery secretes a larger fraction of total cell protein than any other known type I secretion mechanism. This high level of protein production is apparently necessary to keep the cell completely covered with S-layer at all times and is similar to the levels noted for other bacterial S-layer proteins (Messner and Sleytr, 1992). This means that the RsaA secretion machinery is either more efficient than that of other type I secretion systems or that a larger number of transport complexes exist in the membranes or a combination of both factors. This question is an important one to answer from a fundamental research perspective, to address such things as what makes a secretion apparatus more efficient. It is also important because some current research is engaged in evaluating the potential of the S-layer protein secretion system for the secretion of heterologous proteins and peptides in a biotechnological context (Bingle *et al.*, 1997a; Bingle *et al.*, 1997b), where increased levels of secretion has obvious utility.

Now that the genes involved in the transport of RsaA have been discovered, it will be possible to address such issues. For example, gene duplications of the transporter genes can be made to see if more copies of the transporter components increase secretion. In addition, with the genes in hand it will be possible to produce and isolate the individual components and make antisera against them. Antibodies can then be used to assess the amount of protein present in the cell.

Most of the genes involved in O-antigen synthesis are linked to the transporter genes. In addition to the O-antigen synthesis genes mentioned above, a number of other genes involved in O-antigen synthesis have been found by Tn5 mutagenesis. While the linkage pattern of these genes was not as obvious, Southern blot analysis showed that the majority of the Tn5 insertions found were linked to the transporter genes as well. However, it was not demonstrated that all of the Tn5 insertions were

as closely linked to the transporters. As the Southern analysis of the Tn5 insertions only used two restriction enzymes, further analysis may prove that these other genes are also linked. Usually, all the genes involved in the synthesis of the O-antigen are linked on a 20-30kb fragment of DNA. Sequencing further, past *lpsF*, should reveal other genes involved in O-antigen synthesis, possibly including genes not found here by Tn5 mutagenesis.

Perosamine appears to be the major component of the O-antigen. Analysis of the O-antigen showed that it is composed of a 4,6, dideoxy-4-amino-hexose, of which perosamine is an example. It was shown in this report that all the genes required for the synthesis of perosamine are found in the genome of *C. crescentus*. Furthermore, three of these genes were disrupted by transposon mutagenesis leading to an altered O-antigen. It is reasonable to conclude from these data that perosamine is the 4,6, dideoxy-4-amino-hexose seen in the chemical analysis of the O-antigen.

Several glycosyltransferases are involved in the synthesis of the O-antigen. NMR analysis of the O-antigen revealed a number of different anomeric proton signals, suggesting that there are several different linkages between the sugar residues. This implies the presence of multiple glycosyltransferases to produce these linkages. A number of Tn5 insertions altering the O-antigen were found in genes with similarity to mannosyltransferases. Since perosamine is a derivative of mannose the transferases are probably highly similar and this has been found with the perosamine transporter, RfbV from *E. coli* O157 (see Ch. 6). One Tn5 insertion interrupts a gene with similarity to galactosyltransferases that transfer the first sugar residue to the lipid precursor of the O-antigen. It may be that this enzyme, LpsJ, transfers a galactose to the lipid precursor as a first step in the growing O-antigen. Alternatively, since perosamine is an isomer of galactose, a perosamine may be the first residue of the O-antigen chain. Galactose may have been missed in the analysis of the O-antigen since it is also found in the core and a slightly increased level, relative to other core sugars, would have gone unnoticed.

Several other genes involved in the proper formation of the smooth LPS have also been found. One, *lpsK*, may be involved in synthesis of a core or O-antigen sugar. Another, *lpsI*, appears to code for a transciption repressor that affects smooth LPS production. Tn5 insertions interrupting O-antigen synthesis were found in two ORFs with no similarity to any known proteins. Two of these insertions are 3' of an ORF coding for an ABC-2 transporter. ABC-2 transporters export O-antigens and extracellular polysaccharides. If this is the ABC-2 transporter that exports the O-antigen, it suggests that the O-antigen is synthesized in the cytoplasm by addition of sugar residues to the non-reducing terminus.

The information provided here should assist in determining the correct structure of the S-LPS and may also allow the attachment site(s) of the O-antigen to RsaA to be determined. A number of possibilities present themselves for future steps in analysis of the S-LPS. The first obvious step is to isolate the DNA containing the genes *lpsGHIJKL* and determine how closely they are linked. Sequencing of this DNA may reveal other genes involved in O-antigen synthesis and possibly synthesis of the core (for example LpsK may be involved in synthesis of a core sugar and the DNA surrounding it may contain the remaining synthesis genes). The other obvious experiment is to knock-out LpsA and LpsB and confirm that they are involved in the synthesis of the O-antigen.

There may be more genes involved in the synthesis of the O-antigen that were not found when screening the Tn5 library. For example, interruption of O-antigen synthesis genes that did not result in complete detachment of the S-layer may have been missed by the screen. An example of this might be enzymes involved in the transfer of the sugar residues that are not involved in the attachment of process.

The S-layer lies very close to the outer membrane of the bacterium as seen in electron micrographs (Smit *et al*, 1981, Smit *et al*, 1984). If the O-antigen consisted of a single chain, it would be 40 residues long; long enough to span the distance between the S-layer and outer membrane numerous times. This suggests that the S-layer either attaches to several points along the chain (Fig 1-4) or the O-antigen has multiple branches. Selective mutation of the various transferases, or by using

the Tn5 mutants, should allow one to determine which of these possibilities is correct by analyzing the different sized O-antigens that are produced.

Summary

RsaA, the S-layer subunit of *C. crescentus*, is transported by a type I secretion system involving three proteins, an ABC-transporter, a periplasmic spanning Membrane Forming Protein and an outer membrane protein.

It was shown that a number of other FWC species also contain type I secretion systems that probably secrete the S-layer subunit. The evolutionary relationships of these type I secretion systems and the S-layer subunit genes was examined.

A number of genes involved in the synthesis of the smooth LPS were found. Some of these genes code for enzymes involved in the synthesis of perosamine, the likely major component of the O-antigen. Other genes code for the glycosyltransferases that link the sugar residues of the O-antigen to each other.

Bibliography

Abraham, W., Strömpl, C., Meyer, H., Lindholst, S., Moore, E.R.B., Christ, R., Vancanneyt, M., Tindall, B.J., Bennasar, A., Smit, J. and Tesar, M. (1999) Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species and emended description of the genera *Brevundimonas* and *Caulobacter*. *Intl J. of System. Bacteriol.*, **49**, 1053-1073.

Akatsuka, H., Binet, R., Kawai, E., Wandersman, C. and Omori, K. (1997) Lipase secretion by bacterial hybrid ATP-binding cassette exporters: molecular recognition of the LipBCD, PrtDEF, and HasDEF exporters. *J. Bacteriol.*, **179**, 4754-4760.

Alley, M.R., Gomes, S.L., Alexander, W. and Shapiro, L. (1991) Genetic analysis of a temporally transcribed chemotaxis gene cluster in *Caulobacter crescentus*. *Genetics*, **129**, 333-341.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403-410.

Anderson, D.M. and Schneewind, O. (1997) A mRNA signal for the type III secretion of Yop proteins by *Yersinia enterocolitica* [see comments]. *Science*, **278**, 1140-1143.

Anderson, D.M. and Schneewind, O. (1999) Type III machines of Gram-negative pathogens: injecting virulence factors into host cells and more. *Curr. Opin. Microbiol.*, **2**, 18-24.

Annunziato, P.W., Wright, L.F., Vann, W.F. and Silver, R.P. (1995) Nucleotide sequence and genetic analysis of the *neuD* and *neuB* genes in region 2 of the polysialic acid gene cluster of *Escherichia coli* K1. *J. Bacteriol.*, **177**, 312-319.

Armstrong, S., Zhang, H., Tabernero, L., Hermodson, M. and Stauffacher, C. (1999) Powering the ABC transporter: The crystallographic structure of the ATP-binding cassette, RbsA. *ATP-Binding Cassette Transporters: From Multidrug Resistance to Genetic Disease*. FEBS Advanced Lecture Course, Gosau, Austria, p. 3.

Awram, P. and Smit, J. (1998) The *Caulobacter crescentus* paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus. *J. Bacteriol.*, **180**, 3062-3069.

Bairoch, A. (1992) PROSITE: a dictionary of sites and patterns in proteins. *Nuc. Acids Res*, **20 Suppl**, 2013-2018.

Bechthold, A., Sohng, J.K., Smith, T.M., Chu, X. and Floss, H.G. (1995) Identification of *Streptomyces violaceoruber* Tu22 genes involved in the biosynthesis of granaticin. *Mol. Gen. Genet.*, **248**, 610-620.

Becker, A., Kuster, H., Niehaus, K. and Puhler, A. (1995) Extension of the *Rhizobium meliloti* succinoglycan biosynthesis gene cluster: identification of the exsA gene encoding an ABC transporter protein, and the *exsB* gene which probably codes for a regulator of succinoglycan biosynthesis. *Mol Gen Genet*, **249**, 487-497.

Belland, R.J. and Trust, T.J. (1985) Synthesis, export, and assembly of *Aeromonas salmonicida* A-layer analyzed by transposon mutagenesis. *J. Bacteriol.*, **163**, 877-881.

Beveridge, T.J., Pouwels, P.H., Sara, M., Kotiranta, A., Lounatmaa, K., Kari, K., Kerosuo, E., Haapasalo, M., Egelseer, E.M., Schocher, I., Sleytr, U.B., Morelli, L., Callegari, M.L., Nomellini, J.F., Bingle, W.H., Smit, J., Leibovitz, E., Lemaire, M., Miras, I., Salamitou, S., Beguin, P., Ohayon, H., Gounon, P., Matuschek, M. and Koval, S.F. (1997) Functions of S-layers. *FEMS Microbiol. Rev.*, **20**, 99-149.

Bik, E.M., Bunschoten, A.E., Willems, R.J., Chang, A.C. and Mooi, F.R. (1996) Genetic organization and functional analysis of the *otn* DNA essential for cell-wall polysaccharide synthesis in *Vibrio cholerae* O139. *Mol. Microbiol.*, **20**, 799-811.

Binet, R., Létoffé, S., Ghigo, J.M., Delepelaire, P. and Wandersman, C. (1997) Protein secretion by Gram-negative bacterial ABC exporters - a review. *Gene*, **192**, 7-11.

Binet, R. and Wandersman, C. (1995) Protein secretion by hybrid bacterial ABC-transporters: specific functions of the membrane ATPase and the membrane fusion protein. *EMBO J.*, **14**, 2298-2306.

Binet, R. and Wandersman, C. (1996) Cloning of the *Serratia marcescens has*F gene encoding the Has ABC exporter outer membrane component: a TolC analogue. *Mol. Microbiol.*, **22**, 265-273.

Bingle, W.H., Awram, P., Nomellini, J.F. and Smit, J. (1999) The Secretion Signal of *C. crescentus* S-layer Protein is Located Within the C-Terminal 82 Amino Acids of the Molecule. *submitted J. Bacteriol.*

Bingle, W.H., Le, K.D. and Smit, J. (1996) The extreme N-terminus of the *Caulobacter crescentus* surface-layer protein directs export of passenger proteins from the cytoplasm but is not required for secretion of the native protein. *Can. J. Microbiol.*, **42**, 672-684.

Bingle, W.H., Nomellini, J.F. and Smit, J. (1997a) Cell-surface display of a *Pseudomonas aeruginosa* strain K pilin peptide within the paracrystalline S-layer of *Caulobacter crescentus. Mol. Microbiol.*, **26**, 277-288.

Bingle, W.H., Nomellini, J.F. and Smit, J. (1997b) Linker mutagenesis of the *Caulobacter crescentus* S-layer protein: toward a definition of an N-terminal anchoring region and a C-terminal secretion signal and the potential for heterologous protein secretion. *J. Bacteriol.*, **179**, 601-611.

Bingle, W.H. and Smit, J. (1994) Alkaline phosphatase and a cellulase reporter protein are not exported from the cytoplasm when fused to large N-terminal portions of the *Caulobacter crescentus* surface (S)-layer protein. *Can J Microbiol*, **40**, 777-782.

Blaser, M.J., Smith, P.F. and Kohler, P.F. (1985) Susceptibility of Campylobacter isolates to the bactericidal activity of human serum. *J. Infect. Dis.*, **151**, 227-235.

Blaser, M.J., Smith, P.F., Repine, J.E. and Joiner, K.A. (1988) Pathogenesis of *Campylobacter fetus* infections. Failure of encapsulated *Campylobacter fetus* to bind C3b explains serum and phagocytosis resistance. *J. Clin. Invest.*, **81**, 1434-1444.

Boos, W. and Shuman, H. (1998) Maltose/maltodextrin system of *Escherichia coli:* transport, metabolism, and regulation. *Microbiol. Mol. Biol. Rev.*, **62**, 204-229.

Boot, H.J. and Pouwels, P.H. (1996) Expression, secretion and antigenic variation of bacterial S-layer proteins. *Mol. Microbiol.*, **21**, 1117-1123.

Borinski, R. and Holt, S.C. (1990) Surface characteristics of *Wolinella recta* ATCC 33238 and human clinical isolates: correlation of structure with function. *Infect. Immun.*, **58**, 2770-2776.

Brent, R. and Ptashne, M. (1980) The *lexA* gene product represses its own promoter. *Proc. Natl. Acad. Sci. U. S. A.*, **77**, 1932-1936.

Brun, Y.V., Marczynski, G. and Shapiro, L. (1994) The expression of asymmetry during Caulobacter cell differentiation. *Annu. Rev. Biochem.*, **63**, 419-450.

Burns, D.L. (1999) Biochemistry of type IV secretion. *Curr. Opin. Microbiol.*, **2**, 25-29.

Canter Cremers, H., Spaink, H.P., Wijfjes, A.H., Pees, E., Wijffelman, C.A., Okker, R.J. and Lugtenberg, B.J. (1989) Additional nodulation genes on the Sym plasmid of *Rhizobium leguminosarum* biovar *viciae*. *Plant Mol. Biol.*, **13**, 163-174.

Colnaghi, R., Pagani, S., Kennedy, C. and Drummond, M. (1996) Cloning, sequence analysis and overexpression of the rhodanese gene of *Azotobacter vinelandii*. *Eur. J. Biochem.*, **236**, 240-248.

Croop, J.M. (1998) Evolutionary relationships among ABC transporters. *Methods Enzymol.*, **292**, 101-116.

Currie, H.L., Lightfoot, J. and Lam, J.S. (1995) Prevalence of *gca*, a gene involved in synthesis of A-band common antigen polysaccharide in *Pseudomonas aeruginosa*. *Clinical and Diagnostic Lab. Immun*, **2**, 554-562.

Davidson, A.L. and Nikaido, H. (1991) Purification and characterization of the membrane-associated components of the maltose transport system from *Escherichia coli*. *J. Biol. Chem.*, **266**, 8946-8951.

Decottignies, A. and Goffeau, A. (1997) Complete inventory of the yeast ABC proteins. *Nature Genet.*, **15**, 137-145.

Delepelaire, P. and Wandersman, C. (1990) Protein secretion in gram-negative bacteria. The extracellular metalloprotease B from *Erwinia chrysanthemi* contains a C-terminal secretion signal analogous to that of *Escherichia coli* alpha-hemolysin. *J Biol. Chem.*, **265**, 17118-17125.

Delepelaire, P. and Wandersman, C. (1991) Characterization, localization and transmembrane organization of the three proteins PrtD, PrtE and PrtF necessary for protease secretion by the gram-negative bacterium *Erwinia chrysanthemi*. *Mol. Microbiol.*, **5**, 2427-2434.

Dinh, T., Paulsen, I.T. and Saier, M.H., Jr. (1994) A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. *J. Bacteriol.*, **176**, 3825-3831.

Drummelsmith, J. and Whitfield, C. (1999) Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). *Mol. Microbiol.*, **31**, 1321-1332.

Duong, F., Lazdunski, A., Cami, B. and Murgier, M. (1992) Sequence of a cluster of genes controlling synthesis and secretion of alkaline protease in *Pseudomonas aeruginosa*: relationships to other secretory pathways. *Gene*, **121**, 47-54.

Duong, F., Lazdunski, A. and Murgier, M. (1996) Protein secretion by heterologous bacterial ABC-transporters: the C-terminal secretion signal of the secreted protein confers high recognition specificity. *Mol. Microbiol.*, **21**, 459-470.

Dworkin, J., Tummuru, M.K.R. and Blaser, M.J. (1995) A lipopolysaccharide-binding domain of the *Campylobacter fetus* S-layer protein resides within the conserved N-terminus of a family of silent and divergent homologs. *J. Bacteriol.*, **177**, 1734-1741.

Edwards, P. and Smit, J. (1991) A transducing bacteriophage for *Caulobacter crescentus* uses the paracrystalline surface layer protein as a receptor. *J. Bacteriol.*, **173**, 5568-5572.

Ehrmann, M., Ehrle, R., Hofmann, E., Boos, W. and Schlosser, A. (1998) The ABC maltose transporter. *Mol. Microbiol.*, **29**, 685-694.

Eichelberg, K., Ginocchio, C.C. and Galan, J.E. (1994) Molecular and functional characterization of the *Salmonella typhimurium* invasion genes invB and invC: homology of InvC to the F0F1 ATPase family of proteins. *J. Bacteriol.*, **176**, 4501-4510.

Fallarino, A., Mavrangelos, C., Stroeher, U.H. and Manning, P.A. (1997) Identification of additional genes required for O-antigen biosynthesis in *Vibrio cholerae* O1. *J. Bacteriol.*, **179**, 2147-2153.

Fath, M.J., Skvirsky, R.C. and Kolter, R. (1991) Functional complementation between bacterial MDR-like export systems: colicin V, alpha-hemolysin, and *Erwinia* protease. *J. Bacteriol.*, **173**, 7549-7556.

Fellay, R., Frey, J. and Krisch, H. (1987) Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vitro insertional mutagenesis of gram-negative bacteria. *Gene*, **52**, 147-154.

Feng, J.N., Russel, M. and Model, P. (1997) A permeabilized cell system that assembles filamentous bacteriophage. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 4068-4073.

Finnie, C., Zorreguieta, A., Hartley, N.M. and Downie, J.A. (1998) Characterization of *Rhizobium leguminosarum* exopolysaccharide glycanases that are secreted via a type I exporter and have a novel heptapeptide repeat motif. *J. Bacteriol.*, **180**, 1691-1699.

Fisher, J.A., Smit, J. and Agabian, N. (1988) Transcriptional analysis of the major surface array gene of *Caulobacter crescentus*. *J. Bacteriol.*, **170**, 4706-4713.

Fry, B.N., Korolik, V., ten Brinke, J.A., Pennings, M.T., Zalm, R., Teunis, B.J., Coloe, P.J. and van der Zeijst, B.A. (1998) The lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116. *Microbiolology*, **144**, 2049-2061.

Galan, J.E. and Collmer, A. (1999) Type III secretion machines: bacterial devices for protein delivery into host cells. *Science*, **284**, 1322-1328.

Geremia, R.A., Petroni, E.A., Ielpi, L. and Henrissat, B. (1996) Towards a classification of glycosyltransferases based on amino acid sequence similarities: prokaryotic alpha-mannosyltransferases. *Biochem. J.*, **318**, 133-138.

Gilchrist, A., Fisher, J.A. and Smit, J. (1992) Nucleotide sequence analysis of the gene encoding the *Caulobacter crescentus* paracrystalline surface layer protein. *Can. J. Microbiol.*, **38**, 193-202.

Gilchrist, A. and Smit, J. (1991) Transformation of freshwater and marine caulobacters by electroporation. *J. Bacteriol.*, **173**, 921-925.

Gober, J.W. and Marques, M.V. (1995) Regulation of cellular differentiation in *Caulobacter crescentus*. *Microbiol. Rev.*, **59**, 31-47.

Gorbalenya, A.E. and Koonin, E.V. (1990) Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. *J Mol Biol*, **213**, 583-591.

Guzzo, J., Murgier, M., Filloux, A. and Lazdunski, A. (1990) Cloning of the *Pseudomonas aeruginosa* alkaline protease gene and secretion of the protease into the medium by *Escherichia coli*. *J. Bacteriol.*, **172**, 942-948.

Henkin, T.M., Grundy, F.J., Nicholson, W.L. and Chambliss, G.H. (1991) Catabolite repression of alpha-amylase gene expression in Bacillus subtilis involves a trans-

acting gene product homologous to the *Escherichia coli lacI* and *galR* repressors. *Mol. Microbiol.*, **5**, 575-584.

Holland, I.B. (1999) personal communication.

Holton, T.A. and Graham, M.W. (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nuc. Acids Res.*, **19**, 1156.

Hovmoller, S., Sjogren, A. and Wang, D.N. (1988) The structure of crystalline bacterial surface layers. *Prog. Biophys. Mol. Biol.*, **51**, 131-163.

Hung, L.W., Wang, I.X., Nikaido, K., Liu, P.Q., Ames, G.F. and Kim, S.H. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter [see comments]. *Nature*, **396**, 703-707.

Hwang, J., Zhong, X. and Tai, P.C. (1997) Interactions of dedicated export membrane proteins of the colicin V secretion system: CvaA, a member of the membrane fusion protein family, interacts with CvaB and TolC. *J. Bacteriol.*, **179**, 6264-6270.

Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., Gallagher, M.P., Gill, D.R., Hubbard, R.E. and Higgins, C.F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport [see comments]. *Nature*, **346**, 362-365.

Kawai, E., Akatsuka, H., Idei, A., Shibatani, T. and Omori, K. (1998) *Serratia marcescens* S-layer protein is secreted extracellularly via an ATP-binding cassette exporter, the Lip system. *Mol. Microbiol.*, **27**, 941-952.

Keen, N.T., Tamaki, S., Kobayashi, D. and Trollinger, D. (1988) Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. *Gene*, **70**, 191-197.

Kenny, B., Taylor, S. and Holland, I.B. (1992) Identification of individual amino acids required for secretion within the haemolysin (HlyA) C-terminal targeting region. *Mol Microbiol*, **6**, 1477-1489.

Kido, N., Sugiyama, T., Yokochi, T., Kobayashi, H. and Okawa, Y. (1998) Synthesis of *Escherichia coli* O9a polysaccharide requires the participation of two domains of WbdA, a mannosyltransferase encoded within the wb* gene cluster. *Mol Microbiol*, **27**, 1213-1221.

Koronakis, V., Hughes, C. and Koronakis, E. (1993) ATPase activity and ATP/ADP-induces conformational change in the soluble domain of the bacterial protein translocator HlyB. *Mol. Microbiol.*, **8**, 1163-1175.

Koronakis, V., Li, J., Koronakis, E. and Stauffer, K. (1997) Structure of TolC, the outer membrane component of the bacterial type I efflux system, derived from two-dimensional crystals. *Mol. Microbiol.*, **23**, 617-626.

Kovach, M.E., Phillips, R.W., Elzer, P.H., Roop, R.M. and Peterson, K.M. (1994) pBBR1MCS: a broad-host-range cloning vector. *Biotechniques*, **16**, 800-802.

Koval, S.F. and Hynes, S.H. (1991) Effect of paracrystalline protein surface layers on predation by *Bdellovibrio bacteriovorus*. *J. Bacteriol.*, **173**, 2244-2249.

Koval, S.F. and Murray, R.G. (1984) The isolation of surface array proteins from bacteria. *Can J Biochem Cell Biol*, **62**, 1181-1189.

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J.E. and Aizawa, S.I. (1998) Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science*, **280**, 602-605.

Leeds, J.A. and Welch, R.A. (1996) RfaH enhances elongation of *Escherichia coli hly*CABD mRNA. *J. Bacteriol.*, **178**, 1850-1857.

Letellier, L., Howard, S.P. and Buckley, J.T. (1997) Studies on the energetics of proaerolysin secretion across the outer membrane of *Aeromonas* species. Evidence for a requirement for both the protonmotive force and ATP. *J. Biol. Chem.*, **272**, 11109-11113.

Létoffé, S., Delepelaire, P. and Wandersman, C. (1990) Protease secretion by *Erwinia chrysanthemi*: The specific secretion functions are analogous to those of *Escherichia coli* alpha-haemolysin. *EMBO J*, **9**, 1375-1382.

Létoffé, S., Ghigo, J.M. and Wandersman, C. (1994a) Iron acquisition from heme and hemoglobin by a *Serratia marcescens* extracellular protein. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 9876-9880.

Létoffé, S., Ghigo, J.M. and Wandersman, C. (1994b) Secretion of the *Serratia marcescens* HasA protein by an ABC transporter. *J. Bacteriol.*, **176**, 5372-5377.

Létoffé, S. and Wandersman, C. (1992) Secretion of CyaA-PrtB and HlyA-PrtB fusion proteins in *Escherichia coli*: Involvement of the glycine-rich repeat domain of *Erwinia chrysanthemi* protease B. *J. Bacteriol.*, **174**, 4920-4927.

Linton, K.J. and Higgins, C.F. (1998) The *Escherichia coli* ATP-binding cassette (ABC) proteins. *Mol. Microbiol.*, **28**, 5-13.

Linton, K.J., Jarvis, B.W. and Hutchinson, C.R. (1995) Cloning of the genes encoding thymidine diphosphoglucose 4,6- dehydratase and thymidine diphospho-4-keto-6-deoxyglucose 3,5-epimerase from the erythromycin-producing *Saccharopolyspora erythraea. Gene*, **153**, 33-40.

Liu, D., Haase, A.M., Lindqvist, L., Lindberg, A.A. and Reeves, P.R. (1993) Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterization of transferase genes of groups B, C2, and E1. *J. Bacteriol.*, **175**, 3408-3413.

Lu, H.M. and Lory, S. (1996) A specific targeting domain in mature exotoxin A is required for its extracellular secretion from *Pseudomonas aeruginosa. EMBO J.*, **15**, 429-436.

Luckevich, M.D. and Beveridge, T.J. (1989) Characterization of a dynamic S layer on *Bacillus thuringiensis. J. Bacteriol.*, **171**, 6656-6667.

Mackman, N., Nicaud, J.M., Gray, L. and Holland, I.B. (1985) Identification of polypeptides required for the export of haemolysin 2001 from *E. coli. Mol. Gen. Genet.* **201**, 529-536.

MacRae, J.D. and Smit, J. (1991) Characterization of *caulobacters* isolated from wastewater treatment systems. *Appl. Environ. Microbiol.*, **57**, 751-758.

Malakooti, J., Wang, S.P. and Ely, B. (1995) A consensus promoter sequence for *Caulobacter crescentus* genes involved in biosynthetic and housekeeping functions. *J. Bacteriol.*, **177**, 4372-4376.

Martin, V.J. and Mohn, W.W. (1999) An alternative inverse PCR (IPCR) method to amplify DNA sequences flanking Tn5 transposon insertions [In Process Citation]. *J. Microbiol. Methods*, **35**, 163-166.

Maser, P. and Kaminsky, R. (1998) Identification of three ABC transporter genes in *Trypanosoma brucei* spp. *Parasitol. Res.*, **84**, 106-111.

Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Eng.*, **1**, 67-74.

Messner, P. and Sleytr, U.B. (1992) Crystalline bacterial cell-surface layers. In Rose, A.H. and Tempest, D.W. (eds.), *Advances in Microbial Physiology*. Academic Press, London, Vol. 33, pp. 213-275.

Morales, V.M., Backman, A. and Bagdasarian, M. (1991) A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene*, **97**, 39-47.

Munn, C.B., Ishiguro, E.E., Kay, W.W. and Trust, T.J. (1982) Role of surface components in serum resistance of virulent *Aeromonas salmonicida. Infect. Immun.*, **36**, 1069-1075.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.*, **10**, 1-6.

Nikaido, H. (1994) Maltose transport system of *Escherichia coli*: an ABC-type transporter. *FEBS Lett.*, **346**, 55-58.

Nikaido, H. and Vaara, M. (1985) Molecular basis of bacterial outer membrane permeability. *Microbiol. Rev.*, **49**, 1-32.

Nomellini, J.F., Kupcu, S., Sleytr, U.B. and Smit, J. (1997) Factors controlling in vitro recrystallization of the *Caulobacter crescentus* paracrystalline S-layer. *J. Bacteriol.*, **179**, 6349-6354.

Pearson, W.R., Wood, T., Zhang, Z. and Miller, W. (1997) Comparison of DNA sequences with protein sequences. *Genomics*, **46**, 24-36.

Pohlner, J., Halter, R., Beyreuther, K. and Meyer, T.F. (1987) Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature*, **325**, 458-462.

Poindexter, J.S. (1981) The caulobacters: ubiquitous unusual bacteria. *Microbiol. Rev.*, **45**, 123-179.

Pugsley, A.P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.*, **57**, 50-108.

Ravenscroft, N., Walker, S.G., Dutton, G.G. and Smit, J. (1991) Identification, isolation, and structural studies of extracellular polysaccharides produced by *Caulobacter crescentus. J. Bacteriol.*, **173**, 5677-5684.

Ravenscroft, N., Walker, S.G., Dutton, G.S. and Smit, J.K. (1992) Identification, isolation, and structural studies of the outer membrane lipopolysaccharide of *Caulobacter crescentus*. *J. Bacteriol.*, **174**, 7595-7605.

Riley, R.G. and Kolodziej, B.J. (1976) Pathway of glucose catabolism in *Caulobacter crescentus*. *Microbios*, **16**, 219-226.

Roberts, R.C., Mohr, C.D. and Shapiro, L. (1996) Developmental programs in bacteria. *Curr. Top. Dev. Biol.*, **34**, 207-257.

Rocchetta, H.L., Burrows, L.L., Pacan, J.C. and Lam, J.S. (1998) Three rhamnosyltransferases responsible for assembly of the A-band D- rhamnan polysaccharide in *Pseudomonas aeruginosa*: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis [published erratum appears in Mol Microbiol 1998 Dec;30(5):1131]. *Mol. Microbiol.*, **28**, 1103-1119.

Russel, M. (1998) Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *J. Mol. Biol.*, **279**, 485-499.

Salmond, G.P. and Reeves, P.J. (1993) Membrane traffic wardens and protein secretion in gram-negative bacteria. *Trends in Biochem. Sci.*, **18**, 7-12.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.

Sara, M. and Sleytr, U.B. (1996a) Biotechnology and biomimetic with crystalline bacterial cell surface layers (S-layers). *Micron*, **27**, 141-156.

Sara, M. and Sleytr, U.B. (1996b) Crystalline bacterial cell surface layers (S-layers): from cell structure to biomimetics. *Prog. Biophys. Mol. Biol.*, **65**, 83-111.

Scheu, A.K., Economou, A., Hong, G.F., Ghelani, S., Johnston, A.W. and Downie, J.A. (1992) Secretion of the *Rhizobium leguminosarum* nodulation protein NodO by haemolysin-type systems. *Mol. Microbiol.*, **6**, 231-238.

Schnaitman, C.A. and Klena, J.D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.*, **57**, 655-682.

Schulein, R., Gentschev, I., Schlor, S., Gross, R. and Goebel, W. (1994) Identification and characterization of two functional domains of the hemolysin translocator protein HlyD. *Mol. Gen. Genet.*, **245**, 203-211.

Shapiro, L. (1976) Differentiation in the Caulobacter cell cycle. *Annu. Rev. Microbiol.*, **30**, 377-407.

Shapiro, L. and Losick, R. (1997) Protein localization and cell fate in bacteria. *Science*, **276**, 712-718.

Sheps, J.A., Zhang, F. and Ling, V. (1996) Phylogenetic Analysis of Members of the ABC transporter superfamily. In Rothman, S.R. (ed.) *Membrane Protein Transport*. JAI Press, Greenwich, Conneticut, Vol. 3, p. 81.

Simon, R., Priefer, U. and Puhler, A. (1983) A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram negative bacteria. *Bio/technology*, **1**, 784-790.

Sleytr, U.B. (1976) Self-assembly of the hexagonally and tetragonally arranged subunits of bacterial surface layers and their reattachment to cell walls. *J. Ultrastruct. Res.*, **55**, 360-377.

Sleytr, U.B., Bayley, H., Sara, M., Breitwieser, A., Kupcu, S., Mader, C., Weigert, S., Unger, F.M., Messner, P., Jahn-Schmid, B., Schuster, B., Pum, D., Douglas, K., Clark, N.A., Moore, J.T., Winningham, T.A., Levy, S., Frithsen, I., Pankovc, J., Beale, P., Gillis, H.P., Choutov, D.A. and Martin, K.P. (1997a) Applications of S-layers. *FEMS Microbiol. Rev.*, **20**, 151-175.

Sleytr, U.B. and Messner, P. (1983) Crystalline surface layers on bacteria. *Annu. Rev. Microbiol.*, **37**, 311-339.

Sleytr, U.B. and Messner, P. (1988) Crystalline surface layers in procaryotes. *J. Bacteriol.*, **170**, 2891-2897.

Sleytr, U.B., Messner, P., Pum, D. and Sara, M. (1993) Crystalline bacterial cell surface layers. *Mol. Microbiol.*, **10**, 911-916.

Sleytr, U.B., Pum, D. and Sara, M. (1997b) Advances in S-layer nanotechnology and biomimetics. *Adv. Biophys.*, **34**, 71-79.

Sleytr, U.B. and Sara, M. (1997) Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications. *Trends Biotechnol.*, **15**, 20-26.

Smit, J. and Agabian, N. (1984) Cloning of the major protein of the *Caulobacter crescentus* periodic surface layer: detection and characterization of the cloned peptide by protein expression assays. *J. Bacteriol.*, **160**, 1137-1145.

Smit, J., Engelhardt, H., Volker, S., Smith, S.H. and Baumeister, W. (1992) The S-layer of *Caulobacter crescentus*: three-dimensional image reconstruction and structure analysis by electron microscopy. *J. Bacteriol.*, **174**, 6527-6538.

Smit, J., Grano, D.A., Glaeser, R.M. and Agabian, N. (1981) Periodic surface array in *Caulobacter crescentus*: fine structure and chemical analysis. *J. Bacteriol.*, **146**, 1135-1150.

Stahl, D.A., Key, R., Flesher, B. and Smit, J. (1992) The phylogeny of marine and freshwater caulobacters reflects their habitat. *J. Bacteriol.*, **174**, 2193-2198.

Stanley, P.L., Diaz, P., Bailey, M.J., Gygi, D., Juarez, A. and Hughes, C. (1993) Loss of activity in the secreted form of *Escherichia coli* haemolysin caused by an *rfaP* lesion in core lipopolysaccharide assembly. *Mol. Microbiol.*, **10**, 781-787.

Stevenson, G., Andrianopoulos, K., Hobbs, M. and Reeves, P.R. (1996) Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J. Bacteriol.*, **178**, 4885-4893.

Stewart, M. and Beveridge, T.J. (1980) Structure of the regular surface layer of *Sporosarcina ureae. J. Bacteriol.*, **142**, 302-309.

Stroeher, U.H., Karageorgos, L.E., Brown, M.H., Morona, R. and Manning, P.A. (1995) A putative pathway for perosamine biosynthesis is the first function encoded within the *rfb* region of *Vibrio cholerae* O1. *Gene*, **166**, 33-42.

Sugiyama, T., Kido, N., Kato, Y., Koide, N., Yoshida, T. and Yokochi, T. (1998) Generation of *Escherichia coli* O9a serotype, a subtype of *E. coli* O9, by transfer of the wb* gene cluster of Klebsiella O3 into E. coli via recombination. *J. Bacteriol.*, **180**, 2775-2778.

Sutton, J.M., Peart, J., Dean, G. and Downie, J.A. (1996) Analysis of the C-terminal secretion signal of the *Rhizobium leguminosarum* nodulation protein NodO; a potential system for the secretion of heterologous proteins during nodule invasion. *Mol. Plant Microbe Interact.*, **9**, 671-680.

Thompson, S.A., Shedd, O.L., Ray, K.C., Beins, M.H., Jorgensen, J.P. and Blaser, M.J. (1998) *Campylobacter fetus* surface layer proteins are transported by a type I secretion system. *J. Bacteriol.*, **180**, 6450-6458.

Thorne, K.J., Oliver, R.C. and Glauert, A.M. (1976) Synthesis and turnover of the regularly arranged surface protein of *Acinetobacter* sp. relative to the other components of the cell envelope. *J. Bacteriol.*, **127**, 440-450.

Tobin, M.B., Peery, R.B. and Skatrud, P.L. (1997) Genes encoding multiple drug resistance-like proteins in *Aspergillus fumigatus* and *Aspergillus flavus*. *Gene*, **200**, 11-23.

Vaara, M. (1992) Eight bacterial proteins, including UDP-N-acetylglucosamine acyltransferase (LpxA) and three other transferases of *Escherichia coli*, consist of a six-residue periodicity theme. *FEMS Microbiol. Lett.*, **76**, 249-254.

Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, **19**, 259-268.

Vuorio, R., Harkonen, T., Tolvanen, M. and Vaara, M. (1994) The novel hexapeptide motif found in the acyltransferases LpxA and LpxD of lipid A biosynthesis is conserved in various bacteria. *FEBS Letters*, **337**, 289-292.

Walker, J.E., Saraste, M. and Gay, N.J. (1984) The unc operon. Nucleotide sequence, regulation and structure of ATP- synthase. *Biochim. Biophys. Acta.*, **768**, 164-200.

Walker, S.G., Karunaratne, D.N., Ravenscroft, N. and Smit, J. (1994) Characterization of mutants of *Caulobacter crescentus* defective in surface attachment of the paracrystalline surface layer. *J. Bacteriol.*, **176**, 6312-6323.

Walker, S.G., Smith, S.H. and Smit, J. (1992) Isolation and comparison of the paracrystalline surface layer proteins of freshwater caulobacters. *J. Bacteriol.*, **174**, 1783-1792.

Wandersman, C., Delepelaire, P. and Létoffé, S. (1990) Secretion processing and activation of *Erwinia chrysanthemi* proteases. *Biochimie*, **72**, 143-146.

Wandersman, C. and Létoffé, S. (1993) Involvement of lipopolysaccharide in the secretion of *Escherichia coli* alpha-haemolysin and *Erwinia chrysanthemi* proteases. *Mol. Microbiol.*, **7**, 141-150.

Wang, L. and Reeves, P.R. (1998) Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect. Immun.*, **66**, 3545-3551.

Ward, M.J., Bell, A.W., Hamblin, P.A., Packer, H.L. and Armitage, J.P. (1995) Identification of a chemotaxis operon with two cheY genes in *Rhodobacter sphaeroides. Mol. Microbiol.*, **17**, 357-366.

Weiss, A.A., Johnson, F.D. and Burns, D.L. (1993) Molecular characterization of an operon required for pertussis toxin secretion. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 2970-2974.

Welch, R.A. (1991) Pore-forming cytolysins of gram-negative bacteria. *Mol. Microbiol.*, **5**, 521-528.

Welsh, M.J. (1998) The ABC of a versatile engine. *Nature*, **396**, 623-624.

Whitfield, C. (1995) Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol.*, **3**, 178-185.

Wolff, N., Delepelaire, P., Ghigo, J.M. and Delepierre, M. (1997) Spectroscopic studies of the C-terminal secretion signal of the *Serratia marcescens* haem acquisition protein (HasA) in various membrane-mimetic environments. *Eur. J. of Biochem.*, **243**, 400-407.

Wolff, N., Ghigo, J.M., Delepelaire, P., Wandersman, C. and Delepierre, M. (1994) C-terminal secretion signal of an *Erwinia chrysanthemi* protease secreted by a signal peptide-independent pathway: proton NMR and CD conformational studies in membrane-mimetic environments. *Biochemistry*, **33**, 6792-6801.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-119.

Yap, W.H., Thanabalu, T. and Porter, A.G. (1994) Influence of transcriptional and translational control sequences on the expression of foreign genes in *Caulobacter crescentus*. *J Bacteriol*, **176**, 2603-2610.

Yin, Y., Zhang, F., Ling, V. and Arrowsmith, C.H. (1995) Structural analysis and comparison of the C-terminal transport signal domains of hemolysin A and leukotoxin A. *FEBS Lett.*, **366**, 1-5.

Yun, C., Ely, B. and Smit, J. (1994) Identification of genes affecting production of the adhesive holdfast of a marine caulobacter. *J Bacteriol.*, **176,** 796-803

Zhang, F., Sheps, J.A. and Ling, V. (1998) Structure-function analysis of hemolysin B. *Methods Enzymol.*, **292**, 51-66.

Appendix 1

RAT fragment-*rsaADE, lpsABCDE, rsaF, lpsF*

```
LOCUS       NA1000RATX  16458 bp    DNA             BCT      07-OCT-1999
DEFINITION  Caulobacter crescentus sst1, S-layer subunit (rsaA),
            ABC-transporter (rsaD), Membrane Forming Unit (rsaE), putative
            GDP-mannose-4,6-dehydratase (LpsA), putative acetyltransferase
            (LpsB), putative perosamine synthetase (LpsC), putative
            mannosyltransferase (LpsD), putative mannosyltransferase (LpsE),
            Outer membrane protein (rsaF), and putative perosamine transferase
            (LpsE) genes, complete cds.
ACCESSION   NA1000RATX
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1230 to 2387)
  AUTHORS   Fisher,J.A., Smit,J. and Agabian,N.
  TITLE     Transcriptional analysis of the major surface array gene of
            Caulobacter crescentus
  JOURNAL   J. Bacteriol. 170 (10), 4706-4713 (1988)
  MEDLINE   89008089
REFERENCE   2  (bases 1336 to 4645)
  AUTHORS   Gilchrist,A., Fisher,J.A. and Smit,J.
  TITLE     Nucleotide sequence analysis of the gene encoding the Caulobacter
            crescentus paracrystalline surface layer protein
  JOURNAL   Can. J. Microbiol. 38 (3), 193-202 (1992)
  MEDLINE   93007489
REFERENCE   3  (bases 1 to 16458)
  AUTHORS   Awram,P. and Smit,J.
  TITLE     The Caulobacter crescentus paracrystalline S-layer protein is
            secreted by an ABC transporter (type I) secretion apparatus
  JOURNAL   J. Bacteriol. 180 (12), 3062-3069 (1998)
  MEDLINE   98292737
REFERENCE   4  (bases 1 to 16458)
  AUTHORS   Awram,P.A. and Smit,J.K.
  TITLE     Identification of Genes involved in the Synthesis of the Smooth
            Lipopolysaccharide
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 16458)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Microbiology and Immunology, University of
            British Columbia, 300-6174 University Blvd, Vancouver, BC V6T 1Z3,
            Canada
FEATURES             Location/Qualifiers
     source          1..16458
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(227..799)
                     /gene="sst1"
```

```
                /note="unknown"
CDS             complement(227..799)
                /gene="sst1"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /product="Sst1"
                /translation="MAAQVLSFFQRSPRYAPQPADWSQQELAEFYRVESALIRAGIRV
                GTDRGLSDENEPWFVFYRADDGEVVIHFARIDGEYLIAGPAYEEIARGFDFTSLVRNL
                VARHPLIRRSDSGSNLSVHPAALLVAVVGTAFFKTGEARAAETGQSNATSGHNRPVLL
                SSSSNASLNDRCRAGRLPAARLCLGATAGQ"
gene            1443..4523
                /gene="rsaA"
CDS             1443..4523
                /gene="rsaA"
                /citation=[1]
                /citation=[2]
                /codon_start=1
                /transl_table=11
                /product="S-layer subunit"
                /translation="MAYTTAQLVTAYTNANLGKAPDAATTLTLDAYATQTQTGGLSDA
                AALTNTLKLVNSTTAVAIQTYQFFTGVAPSAAGLDFLVDSTTNTNDLNDAYYSKFAQE
                NRFINFSINLATGAGAGATAFAAAYTGVSYAQTVATAYDKIIGNAVATAAGVDVAAAV
                AFLSRQANIDYLTAFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAA
                MINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVA
                GAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSG
                VTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVASTGVTSGTT
                TVGANSAASGTVSVSVANSSTTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTG
                NSSTTAVTVTQTAAATAGATVAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATI
                DSSALTTVNLSGTGTSLGIGRGALTATPTANTLTLNVNGLTTTGAITDSEAAADDGFT
                TINIAGSTASSTIASLVAADATTLNISGDARVTITSHTAAALTGITVTNSVGATLGAE
                LATGLVFTGGAGADSILLGATTKAIVMGAGDDTVTVSSATLGAGGSVNGGDGTDVLVA
                NVNGSSFSADPAFGGFETLRVAGAAAQGSHNANGFTALQLGATAGATTFTNVAVNVGL
                TVLAAPTGTTTVTLANATGTSDVFNLTLSSSAALAAGTVALAGVETVNIAATDTNTTA
                HVDTLTLQATSAKSIVVTGNAGLNLTNTGNTAVTSFDASAVTGTGSAVTFVSANTTVG
                EVVTIRGGAGADSLTGSATANDTIIGGAGADTLVYTGGTDTFTGGTGADIFDINAIGT
                STAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQYLDAAAAGDGSG
                TSVAKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFATEVLTLA"
gene            4766..6502
                /gene="rsaD"
CDS             4766..6502
                /gene="rsaD"
                /note="ABC-transporter of RsaA type I secretion system"
                /codon_start=1
                /transl_table=11
                /product="ABC-transporter"
                /translation="MFKRSGAKPTILDQAVLVARPAVITAMVFSFFINILALVSPLYM
                LQVYDRVLTSRNVSTLIVLTVICVFLFLVYGLLEALRTQVLVRGGLKFDGVARDPIFK
                SVLDSTLSRKGIGGQAFRDMDQVREFMTGGLIAFCDAPWTPVFVIVSWMLHPFFGILA
                IIACIIIFGLAVMNDNATKNPIQMATMASIAAQNDAGSTLRNAEVMKAMGMWGGLQAR
                WRARRDEQVAWQAAASDAGGAVMSGIKVFRNIVQTLILGGGAYLAIDGKISAGAMIAG
                SILVGRALAPIEGAVGQWKNYIGARGAWDRLQTMLREEKSADDHMPLPEPRGVLSAEA
                ASILPPGAQQPTMRQASFRIDAGAAVALVGPSAAGKSSLLRGIVGVWPCAAGVIRLDG
                YDIKQWDPEKLGRHVGYLPQDIELFSGTVAQNIARFTEFESQEVIEAATLAGVHEMIQ
                SLPMGYDTAIGEGGASLSGGQRQRLALARAVFRMPALLVLDEPNASLDQVGEVALMEA
                MKRLKAAKRTVIFATHKVNLLAQADYIMVINQGVISDFGERDPMLAKLTGAAPPQTPP
                PTPPPAPLQRVQ"
```

```
gene            6570..7880
                /gene="rsaE"
CDS             6570..7880
                /gene="rsaE"
                /note="MFP of RsaA type I secretion system"
                /codon_start=1
                /transl_table=11
                /product="Membrane Forming Unit"
                /translation="MKPPKIQRPTDNFQAVARIGYGIIALTFVGLLGWAAFAPLDSAV
                IANGVVSAEGNRKTVQHLEGGMLAKILVREGEKVKAGQVLFELDPTQANAAAGITRNQ
                YVALKAMEARLLAERDQRPSISFPADLTSQRADPMVARAIADEQAQFTERRQTIQGQV
                DLMNAQRLQYQSEIEGIDRQTQGLKDQLGFIEDELIDLRKLYDKGLVPRPRLLALERE
                QASLSGSIGRLTADRSKAVQGASDTQLKVRQIKQEFFEQVSQSITETRVRLAEVTEKE
                VVASDAQKRIKIVSPVNGTAQNLRFFTEGAVVRAAEPLVDIAPEDEAFVIQAHFQPTD
                VDNVHMGMVTEVRLPAFHSREIPILNGTIQSLSQDRISDPQNKLDYFLGIVRVDVKQL
                PPHLRGRVTAGMPAQVIVPTGERTVLQYLFSPLRDTLRTTMREE"
gene            8020..8997
                /gene="LpsA"
CDS             8020..8997
                /gene="LpsA"
                /codon_start=1
                /transl_table=11
                /product="putative GDP-mannose-4,6-dehydratase"
                /translation="MAKTALITGVTGQDGAYLAKLLLEKGYTVHGMLRRSASADVIGD
                RLRWIGVYDDIQFELGDLLDEGGLARLMRRLQPDEVYNLAAQSFVGASWDQPHLTGSV
                TGLGTTNMLEAVRLECPQARFYQASSSEMYGLVQHPIQSETTPFYPRSPYAVAKLYAH
                WMTVNYRESFGLHASAGILFNHESPLRGIEFVTRKVTDAVAAIKLGQQKTVDLGNLDA
                KRDWGHAKDYVEAMWLMLQQETPDDYVVATGKTWTVRQMCEVAFAHVGLNYQDHVTIN
                PKFLRPAEVDLLLGDPAKAKAKLGWEPKTTMQQMIAEMVDADIARRSRN"
gene            8997..9644
                /gene="LpsB"
CDS             8997..9644
                /gene="LpsB"
                /codon_start=1
                /transl_table=11
                /product="putative acetyltransferase"
                /translation="MSASLAIGGVVIIGGGGHAKVVIESLRACGETVAAIVDADPTRR
                AVLGVPVVGDDLALPMLREQGLSRLFVAIGDNRLRQKLGRKARDHGFSLVNAIHPSAV
                VSPSVRLGEGVAVMAGVAINADSWIGDLAIINTGAVVDHDCRLGAACHLGPASALAGG
                VSVGERAFLGVGARVIPGVTIGADTIVGAGGVVVRDLPDSVLAIGVPAKIKGDRS"
gene            9716..10756
                /gene="LpsC"
CDS             9716..10756
                /gene="LpsC"
                /codon_start=1
                /transl_table=11
                /product="putative perosamine synthetase"
                /translation="MDTTWISSVGRFIVEFEKAFADYCGVKHAIACNNGTTALHLALV
                AMGIGPGDEVIVPSLTYIASANSVTYCGATPVLVDNDPRTFNLDAAKLEALITPRTKA
                IMPVHLYGQICDMDPILEVARRHNLLVIEDAAEAVGATYRGKKSGSLGDCATFSFFGN
                KIITTGEGGMITTNDDDLAAKMRLLRGQGMDPNRRYWFPIVGFNYRMTNIQAAIGLAQ
                LERVDEHLAARERVVGWYEQKLARLGNRVTKPHVALTGRHVFWMYTVRLGEGLSTTRD
                QVIKDLDALGIESRPVFHPMHIMPPYAHLATDDLKIAEACGVDGLNLPTHAGLTEADI
                DRVIAALDQVLV"
gene            10760..11797
                /gene="LpsD"
CDS             10760..11797
```

```
                     /gene="LpsD"
                     /codon_start=1
                     /transl_table=11
                     /product="putative mannosyltransferase"
                     /translation="MRIVLLSSIVPFINGGARFIVEWLEEKLIEAGHEVERFYLPFVD
                     DPNEILHQIAAWRLMDLTQWCDRVICFRPPAYVVDHPNKVLWFIHHIRTFYDLWDTPY
                     RGMPDDAQHRAIRDNLRALDTQAISEARAVFTNSQVVADRLKAFNGLDATPLYPPIYQ
                     PERFSHTGYGDEIVAISRLEPHKRQALMIEAMQYVKSGVKLRLAGTASSAEYGRQLVK
                     MTHDLGVADRVILEDRWISEDEKADMLKQALAVAYLPKDEDSYGYPSLEGAHARKPVI
                     TTTDSGGVLELVEHGRNGLISAPDPRALAEQFDRLHADKAATAKMGTASLNRLAEMKI
                     DWSTVVERLTS"
gene                 11808..12845
                     /gene="LpsE"
CDS                  11808..12845
                     /gene="LpsE"
                     /codon_start=1
                     /transl_table=11
                     /product="putative mannosyltransferase"
                     /translation="MKVLVVNNAAPFQRGGAEELADHLVRRLNATPGVQSELVRVPFT
                     WEPAERLIEEMLISKGMRLYNVDRVIGLKFPAYLIPHHQKVLWLLHQFRQAYDLSEAG
                     QSHLDFDDTGRAVKAAIRAADNACFAECRKIYCNSPVTQNRLMKFNGVASQVLYPPLN
                     DGELFTGGEHGDYVFAGGRVAAGKRQHLLIEALALLPGSLRLVIAGPPENQAYADRLT
                     KLVEDLDLKDRVELRFGFHPREDIARWANGALICAYLPFDEDSVGYVTMEAFAAGKAV
                     LTVTDSGGLLEIVSADTGAVAEPTPQALAEALDRLTSDKARAISLGDAARRLWRDKNV
                     TWEETVRRLLD"
gene                 12902..14485
                     /gene="rsaF"
CDS                  12902..14485
                     /gene="rsaF"
                     /note="OMP of RsaA type I secretion system"
                     /codon_start=1
                     /transl_table=11
                     /product="Outer membrane protein"
                     /translation="MRVLSKVLSVRTSLIALAMAMAVVGRADLAHAETLAEAITAAYQ
                     SNPNIQAQRAAMRALDENYTQARSAYGLQASASVAEVYGWSKGVNAKNGVEAASQTST
                     LSLSQSLYTNGRFSARLAGVEAQIKAARENLRRIEMDLLVRVTNAYISVRRDREILRI
                     SQGGEAWLQKQLKDTEDKYSVRQVTLTDVQQAKARLASASTQVANAQAQLNVSVAFYA
                     SLVGRQPETLKPEPDIDGLPTTLDEAFNQAEQANPVLLAAGYTEKASRAGVAEARAQR
                     LFSVGARADYRNGSSTPYYARGGLREDTVNASITLTQPLFTSGQLNASVRQSIEENNR
                     DKLLMEDARRSMVLSVSQYWDSLVAARKSLVSLEEEMKANTIAFYGVREEERFALRST
                     IEVLNAQAELQNAQINFVRGRANEYVGRLHLLAQVGTLEVGNLAPGVQPYDPERNFRK
                     VRYRGALPTELIIGTFDKIALPLEPKKPAPGDTSPIRPPSSELPARPVSADKVTPPAS
                     MNDLPALTDDTPVQTAPRN"
gene                 complement(14591..15880)
                     /gene="LpsE"
CDS                  complement(14591..15880)
                     /gene="LpsE"
                     /codon_start=1
                     /transl_table=11
                     /product="putative perosamine transferase"
                     /translation="MTSRLLEIWRRLPTPIRRSAHVVAGAPRAALEALDKALAEHRHR
                     SAERTALARARRRAGPRGLSPTLPVTVIGFHSAVHGLGEGARMLARGFGDMGLGVRAL
                     DLSASVGFAAEIAPAYSSPDPDERGVTISHINPPELLRWARETEGRFLEGRRHIGYWA
                     WELEEVPSDWLPAFDFVDEVWTPSAFAADAIRRVAPRGVKVTPVPYPLYLNPRPQADR
                     QRFGLQDDRVVVLMAFDLRSTAQRKNPDAALRAFRDATVKATRPATLVCKVVGADLYP
                     ETFQALAAEVADDPSIRLLTDNLSAQDMAALTASSDIVLSLHRSEGYGLLLAEAIWLG
                     KPTLATGWSSNVEFMDPASSQFVDYRLVPVEGDGVIYRAGRWADADVGDAAEKLARMI
```

SDDAWRNTLAAATARNGHVSFNRDAWVAMTSARLPLT"

BASE COUNT 2845 a 5542 c 5354 g 2717 t

ORIGIN

```
       1 gagctcaccg cgtgaaccgg cgtgttgtcg acggtctgag tcggcggggg gaggcgcgct
      61 ggccccgccg ccaaaaccac tgatgccggc cggcaaggcc gaggcgtcgc tgaagtcgag
     121 agccgcagcg gcggcggtca ggcccgagcg cacggccggg tccaggctcg cggcgtccac
     181 acggaagtcc gaggccagca gggcggcggc cagaaccagc accgcctcat tggccggcag
     241 ttgcaccgag gcataggcgg gctgctggaa gccgaccggc gcgacaccgg tcgttcaggc
     301 tggcgttcga gctgctgctg agcaggacgg ggcgattgtg gcccgaggtc gcattgctct
     361 ggccggtctc ggcggcgcgc gcttcgccgg tcttgaagaa ggcggtgccg accacggcca
     421 ccagcagggc ggcggggtga accgacaggt tgctgccgct gtcggagcgg cggatcagcg
     481 gatggcgggc gacgaggttg cgcaccaggc tggtgaagtc gaagccgcgc gcaatctctt
     541 cataggcggg gccggcgatc agatactcgc cgtcgatacg cgcaaagtgg atcaccacct
     601 cgccgtcgtc cgctcggtag aagacgaacc agggttcgtt ctcgtcgctc aggccgcgat
     661 cagtgccgac gcggatgccc gcgcggatca gggcgctctc gacacggtag aactcggcca
     721 gctcctgctg gctccagtcc gccggttgcg gcgcgtagcg cggcgagcgc tgaaagaacg
     781 acaggacctg tgcggccata cggcggaagc ttccccaagc ctaggtgaaa agccgacccc
     841 ccgtcggccc aaacacgcta gcagacgaca cgataaccga actagtcttc gctgaacagg
     901 atctcgtagg tgatcggatc atagaagcgg agcagttcgc gcacgaacgt cttctccgag
     961 acctccaggg ccttggccca gtcgcggtaa cggtccggcg gaatgcggcc gcgacccgtc
    1021 tccagctgag agatgaaggt gtaataatca gcgccgacct tagcggccag ctggcgttgc
    1081 gacaggccgg cggcctcgcg catctccttc agccagcggc caccttcgcg gcggaggtct
    1141 tgcacctccg aggcgctgcg gcgttgcggg ttaccataca ttataaagcc tcgcgcgttg
    1201 accgagggca ggagcgcggg cgcgctcact cacccgccag gtgaacagtc tttatatata
    1261 gcgcttttcg gcgggggtga caaggaacgc tatataggaa tttgctgtac cggttagaaa
    1321 aatgctgtac ccctgaaatt cggctattgt cgacgtatga cgtttgctct atagccatcg
    1381 ctgctcccat gcgcgccact cggtcgcagg gggtgtggga ttttttttgg gagacaatcc
    1441 tcatggccta tacgacggcc cagttggtga ctgcgtacac caacgccaac ctcggcaagg
    1501 cgcctgacgc cgccaccacg ctgacgctcg acgcgtacgc gactcaaacc cagacgggcg
    1561 gcctctcgga cgccgctgcg ctgaccaaca ccctgaagct ggtcaacagc acgacggctg
    1621 ttgccatcca gacctaccag ttcttcaccg gcgttgcccc gtcggccgct ggtctggact
    1681 tcctggtcga ctcgaccacc aacaccaacg acctgaacga cgcgtactac tcgaagttcg
    1741 ctcaggaaaa ccgcttcatc aacttctcga tcaacctggc cacgggcgcc ggcgccggcg
    1801 cgacggcttt cgccgccgcc tacacgggcg tttcgtacgc ccagacggtc gccaccgcct
    1861 atgacaagat catcggcaac gccgtcgcga ccgccgctgg cgtcgacgtc gcggccgccg
    1921 tggctttcct gagccgccag gccaacatcg actacctgac cgccttcgtg cgcgccaaca
    1981 cgccgttcac ggccgctgcc gacatcgatc tggccgtcaa ggccgccctg atcggcacca
    2041 tcctgaacgc cgccacggtg tcgggcatcg gtggttacgc gaccgccacg gccgcgatga
    2101 tcaacgacct gtcggacggc gccctgtcga ccgacaacgc ggctggcgtg aacctgttca
    2161 ccgcctatcc gtcgtcgggc gtgtcgggtt cgaccctctc gctgaccacc ggcaccgaca
    2221 ccctgacggg caccgccaac aacgacacgt tcgttgcggg tgaagtcgcc ggcgctgcga
    2281 ccctgaccgt tggcgacacc ctgagcggcg gtgctggcac cgacgtcctg aactgggtgc
    2341 aagctgctgc ggttacggct ctgccgaccg gcgtgacgat ctcgggcatc gaaacgatga
    2401 acgtgacgtc gggcgctgcg atcaccctga acacgtcttc gggcgtgacg ggtctgaccg
    2461 ccctgaacac caacaccagc ggcgcggctc aaaccgtcac cgccggcgct ggccagaacc
    2521 tgaccgccac gaccgccgct caagccgcga acaacgtcgc cgtcgacggc ggcgccaacg
    2581 tcaccgtcgc ctcgacgggc gtgacctcgg gcacgaccac ggtcggcgcc aactcggccg
    2641 cttcgggcac cgtgtcggtg agcgtcgcga actcgagcac gaccaccacg ggcgctatcg
    2701 ccgtgaccgg tggtacggcc gtgaccgtgg ctcaaacggc cggcaacgcc gtgaacacca
    2761 cgttgacgca agccgacgtg accgtgaccg gtaactccag caccacggcc gtgacggtca
    2821 cccaaaccgc cgccgccacc gccggcgcta cggtcgccgg tcgcgtcaac ggcgctgtga
    2881 cgatcaccga ctctgccgcc gcctcggcca cgaccgccgg caagatcgcc acggtcaccc
    2941 tgggcagctt cggcgccgcc acgatcgact cgagcgctct gacgaccgtc aacctgtcgg
    3001 gcacgggcac ctcgctcggc atcggccgcg gcgctctgac cgccacgccg accgccaaca
    3061 ccctgaccct gaacgtcaat ggtctgacga cgaccggcgc gatcacggac tcggaagcgg
    3121 ctgctgacga tggtttcacc accatcaaca tcgctggttc gaccgcctct tcgacgatcg
    3181 ccagcctggt ggccgccgac gcgacgaccc tgaacatctc gggcgacgct cgcgtcacga
```

3241 tcacctcgca caccgctgcc gccctgacgg gcatcacggt gaccaacagc gttggtgcga
3301 ccctcggcgc cgaactggcg accggtctgg tcttcacggg cggcgctggc gctgactcga
3361 tcctgctggg cgccacgacc aaggcgatcg tcatgggcgc cggcgacgac accgtcaccg
3421 tcagctcggc gaccctgggc gctggtggtt cggtcaacgg cggcgacggc accgacgttc
3481 tggtggccaa cgtcaacggt tcgtcgttca gcgctgaccc ggccttcggc ggcttcgaaa
3541 ccctccgcgt cgctggcgcg gcggctcaag gctcgcacaa cgccaacggc ttcacggctc
3601 tgcaactggg cgcgacggcg ggtgcgacga ccttcaccaa cgttgcggtg aatgtcggcc
3661 tgaccgttct ggcggctccg accggtacga cgaccgtgac cctggccaac gccacgggca
3721 cctcggacgt gttcaacctg accctgtcgt cctcggccgc tctggccgct ggtacggttg
3781 cgctggctgg cgtcgagacg gtgaacatcg ccgccaccga caccaacacg accgctcacg
3841 tcgacacgct gacgctgcaa gccacctcgg ccaagtcgat cgtggtgacg ggcaacgccg
3901 gtctgaacct gaccaacacc ggcaacacgg ctgtcaccag cttcgacgcc agcgccgtca
3961 ccggcacggg ctcggctgtg accttcgtgt cggccaacac cacggtgggt gaagtcgtca
4021 cgatccgcgg cggcgctggc gccgactcgc tgaccggttc ggccaccgcc aatgacacca
4081 tcatcggtgg cgctggcgct gacaccctgg tctacaccgg cggtacggac accttcacgg
4141 gtggcacggg cgcggatatc ttcgatatca acgctatcgg cacctcgacc gctttcgtga
4201 cgatcaccga cgccgctgtc ggcgacaagc tcgacctcgt cggcatctcg acgaacggcg
4261 ctatcgctga cggcgccttc ggcgctgcgg tcaccctggg cgctgctgcg accctggctc
4321 agtacctgga cgctgctgct gccggcgacg gcagcggcac ctcggttgcc aagtggttcc
4381 agttcggcgg cgacacctat gtcgtcgttg acagctcggc tggcgcgacc ttcgtcagcg
4441 gcgctgacgc ggtgatcaag ctgaccggtc tggtcacgct gaccacctcg gccttcgcca
4501 ccgaagtcct gacgctcgcc taagcgaacg tctgatcctc gcctaggcga ggatcgctag
4561 actaagagac cccgtcttcc gaaagggagg cggggtcttt cttatgggcg ctacgcgctg
4621 gccggccttg cctagttccg gtggctatga tttagcggga ctgggggct tgctcacttt
4681 ccgccacaat ttcgtggtcg agacggcgcc ttagttgtta ctgtacatgg ccgcgtcggt
4741 tcgcgcggcg tcctgaaggc tcacaatgtt caagcgcagc ggcgcgaagc cgacgatcct
4801 cgaccaggcc gtgctggtcg cccgcccggc ggtgatcacc gccatggtct tcagcttctt
4861 catcaacatt ctggccctgg tcagcccgct gtacatgctg caggtctatg accgcgtgct
4921 gaccagccgc aacgtttcga ccctgatcgt gttgacggtc atctgcgtct tcctgttcct
4981 ggtctacggc ctgctcgagg cgctgcgcac ccaggtgctg gtgcgcggcg gtctgaagtt
5041 cgacggcgtg gcccgggatc cgatcttcaa gtcggtgctg gactccacgc tcagccgcaa
5101 gggcatcggc ggccaggcgt tccgcgacat ggaccaggtc cgagagttca tgaccggcgg
5161 cctgatcgcc ttctgcgatg cgccctggac gccggtgttc gtcatcgtct cgtggatgct
5221 gcacccgttc ttcggcatcc tggcgatcat cgcctgtatc atcatcttcg gcctggccgt
5281 gatgaacgac aacgccacca agaacccgat ccagatggcc accatggcct cgatcgccgc
5341 ccagaacgac gccggttcca ccctgcgcaa cgccgaggtc atgaaggcca tgggcatgtg
5401 gggcggcctg caagcccgct ggcgcgcgcg ccgcgacgag caggtggcct ggcaggccgc
5461 cgccagcgac gccggcggcg cggtgatgtc gggcatcaag gtgttccgca acatcgtcca
5521 gaccctgatc ctgggcggcg gcgcctatct ggccatcgac ggcaagatct cggccggcgc
5581 gatgatcgcc ggctcgatcc tggtcggccg cgccctggcg cccatcgagg gcgccgtggg
5641 tcagtggaag aattatatcg gcgcgcgcgg cgcctgggat cgcctgcaga ccatgctgcg
5701 cgaggaaaag agcgccgacg accacatgcc gctgcccgag ccgcgcggcg tgctgtcggc
5761 cgaagccgcc tcgatcctgc cgccgggcgc gcaacagccg accatgcgcc aggccagctt
5821 ccgtatcgac gccggcgccg cggtggccct tgtcggtccc agcgcggcgg gcaagtcctc
5881 gctgctgcgc ggtatcgtcg gcgtctggcc ctgcgcggcg ggcgtcatcc gcctcgacgg
5941 ctacgacatc aagcagtggg atcccgagaa gctgggtcgc cacgtcggct acctgccgca
6001 ggacatcgag ctgttctcgg gcaccgtcgc ccagaacatc gcccgcttca ccgagttcga
6061 atcgcaggaa gtcatcgagg ccgcgaccct ggcgggcgtg cacgagatga tccagagcct
6121 gccgatgggc tatgacacgg cgatcggcga gggcggcgcc tcgctgtccg gcggccagcg
6181 ccagcgcctg gccctggcgc gtgcggtgtt ccgcatgccg gccctgctgg tgctggacga
6241 gccgaacgcc agcctcgacc aggtgggcga agtggcgctg atggaagcga tgaagcggct
6301 taaggccgct aagcgcacgg tgatcttcgc cacccacaag gtgaacctgt tggcccaggc
6361 cgactacatc atggtgatca accagggtgt gatcagcgac tttggcgaac gcgacccgat
6421 gctggccaag ctgaccgggg ctgcgccgcc ccagacgccg ccgccgacgc cgccgcccgc
6481 gccgttgcag cgcgtccagt aagcgccttc gtcagtccgc ctctcccttc ctggccgttt
6541 cagaacgcgc ccatcaggct tgaatctcaa tgaagccccc caagatccag cgtccgacgg
6601 acaacttcca ggctgtggcc cgtatcggct acggcatcat cgccctgacc tttgtcggtc

```
 6661 tgttgggctg ggccgcgttc gccccgctcg acagcgcggt gatcgccaac ggcgtcgtct
 6721 ccgccgaggg taatcgcaag accgtgcagc acctcgaagg cggcatgctg gccaagatcc
 6781 tggtccgcga aggcgagaag gtgaaggccg gccaggtgct gttcgagctg gacccgaccc
 6841 aggccaacgc cgccgccggc atcacccgca accagtacgt ggctttgaag gccatggaag
 6901 cgcgcctgct ggccgagcgc gaccagcgtc cgtccatcag cttccccgcc gacctgacca
 6961 gccagcgcgc cgatccgatg gtcgcccgcg ccatcgccga cgaacaggcc cagttcactg
 7021 agcgtcgcca gacgatccag ggccaggtcg acctgatgaa cgcccagcgt ttgcagtatc
 7081 agagcgagat cgagggcatc gaccgtcaga cccagggcct gaaggaccaa ctcggcttca
 7141 tcgaggacga gctgatcgac ctgcgtaagc tctatgacaa gggcctggtg ccccggccgc
 7201 gtctgctggc cctggagcgc gagcaggcct cgctgtcggg ctcgatcggc cgtctgaccg
 7261 cagaccgctc caaggccgtc cagggcgcct ctgacaccca gctcaaggtt cgccagatca
 7321 agcaggagtt cttcgagcag gtcagccaga gcatcaccga gacccgggtt cgcctggccg
 7381 aggtgaccga gaaggaggtc gtcgcctccg acgcccagaa gcggatcaag atcgtgtcgc
 7441 ccgtcaacgg cacggcgcag aacctgcgct tcttcaccga gggcgctgtc gttcgcgccg
 7501 ccgagccgct ggtcgacatc gcgcccgagg acgaggcctt cgtgatccag gcgcatttcc
 7561 agccgaccga tgtggacaat gtccatatgg gcatggtcac cgaagttcgg ctgccggcct
 7621 tccactcgcg ggaaatcccg atcctgaacg gcacgatcca gtcgctgtcg caggaccgca
 7681 tttccgatcc gcagaacaag ctcgactact tcctcgggat cgtgcgcgtg gacgtcaagc
 7741 agctgccgcc gcatctgcgc ggcagggtca ccgccggcat gccggcccag gtgatcgtgc
 7801 cgaccggcga gcgcaccgtg ctgcagtacc tgttctcgcc gctgcgagac accctgcgca
 7861 ccacgatgcg cgaggagtag ggcaaaggtt tcaagggcct gatttccaaa gcttttcgga
 7921 tgggcggcgc gggcgagtta aattcgccgg cgctgctttc cattcgcggg caatagtgta
 7981 gtcaggaccc ttcgttgtta ctggagtcag cggatacgca tggcgaaaac ggctttgatc
 8041 accggtgtga ccggtcagga cggggcgtac ctcgccaagc tgctgctgga gaagggttac
 8101 accgtccacg gcatgctgcg tcgctcggcc tcggccgatg tgatcggcga ccgcctgcgc
 8161 tggatcggcg tctatgacga catccagttc gagctgggcg acctcttgga cgagggcggt
 8221 ctggcgcgcc tgatgcggcg cctgcagccg gacgaggtct acaacctggc ggcccagagc
 8281 ttcgtcggcg cctcgtggga ccagccgcac ctgacgggct cggtgacggg cctgggcacg
 8341 accaacatgc tcgaagccgt gcgtctggaa tgcccgcagg cgcggttcta tcaggcctcg
 8401 tcgtccgaaa tgtacggtct ggtgcagcac ccgatccagt cggagacgac gccgttctat
 8461 ccccgctcgc cctatgcggt ggccaagctc tacgcccact ggatgacggt gaactatcgc
 8521 gagagctttg gcctgcacgc ctcggccggc atcctgttca accacgagag cccgctgcgc
 8581 ggcatcgagt tcgtgacccg caaggtcacc gacgcggtgg cggccatcaa gctgggtcag
 8641 caaaagaccg tcgatctggg caatctcgac gccaagcgcg actggggtca cgccaaggac
 8701 tatgtcgagg ccatgtggct gatgctgcag caggagacgc cggacgacta tgtggtcgcg
 8761 accggcaaga cctggaccgt gcgccagatg tgcgaagtgg ccttcgccca tgtcggcctg
 8821 aactatcagg accacgtgac gatcaatccg aagttcctgc gtccggcgga agtggacctg
 8881 ctgctgggcg atccggccaa ggccaaggcc aagctcggct gggaacccaa gacgaccatg
 8941 caacagatga tcgccgaaat ggtcgacgcc gacatcgcgc ggcgctcgcg caactgatga
 9001 gcgcttccct cgccatcggg ggcgtcgtca tcatcggcgg cggcggccac gccaaggtgg
 9061 tcatcgagag cctgcgggcc tgcggtgaga cggtggcggc cattgtcgat gcggatccga
 9121 cgcggcgcgc ggtgttgggc gttccggtag tgggcgatga cctggcgctg ccgatgcttc
 9181 gcgagcaggg gctgtccaga ctgttcgtgg cgatcggcga caaccggctg cgccagaagc
 9241 tgggccgcaa ggcgcgcgac cacggctttt cgctggtcaa cgccatccat ccctctgccg
 9301 tcgtttcgcc tagcgtacgt ctgggcgagg gggttgcggt gatggccggc gtcgcgatca
 9361 acgctgacag ctggatcggc gacctggcga tcatcaacac cggcgctgtt gtcgaccatg
 9421 actgccgcct gggcgcggcc tgccacctgg gacccgcctc ggccctggcc ggcggcgtat
 9481 ccgtgggaga gcgggctttt ctcggtgtcg gcgcccgggt catacctggc gtcacgatcg
 9541 gcgccgacac gatcgtcggc gccggcggtg tcgtcgtgcg cgaccttccg gactcggtcc
 9601 ttgcgatcgg cgtgccggcc aagatcaaag gagaccgttc gtgagtgacc tgccgcgcat
 9661 ttccgtcgcc gcgccgcgcc tcgacggcaa cgaacgtgac tatgtactcg aatgcatgga
 9721 cacgacctgg atctcgtcgg tcggacgctt catcgttgag ttcgaaaagg ccttcgccga
 9781 ctactgtggc gtcaagcacg cgatcgcctg caacaacggt acgaccgcct tgcacctggc
 9841 cctggtggcg atggggatcg gacccggcga cgaggtgatc gttccgagcc tgacctatat
 9901 cgcctcggcc aattcagtca cctattgcgg cgcgacgcct gtgctggtcg acaacgatcc
 9961 gcggaccttc aacctggacg ccgcgaagtt ggaggcgctg ataacgccgc gcacgaaggc
10021 gatcatgccc gtgcacctct acggtcagat ttgcgacatg gatccgatcc tcgaagttgc
```

10081 tcgcaggcat aacctgctcg tgatcgagga tgcggccgag gcggtgggcg cgacctaccg
10141 gggcaagaag tcaggctcgc tgggcgactg cgccaccttc agcttcttcg gcaacaagat
10201 catcaccacc ggcgagggcg ggatgatcac caccaatgat gatgacctgg cggccaagat
10261 gcgcttgctg cgaggccagg gcatggatcc caaccgccgc tactggtttc cgatcgtcgg
10321 cttcaattac cggatgacca acatccaggc cgcgatcggt ctggcgcagc tggagcgggt
10381 cgacgaacac ctggccgcgc gcgaaagggt cgtgggctgg tacgagcaga agctggcgcg
10441 cctgggcaat cgggtcacca agccccatgt cgcgctgacc ggtcgccacg tgttctggat
10501 gtacactgtg cgcctgggcg agggcctttc caccacgcgc gatcaggtga tcaaggatct
10561 cgacgcgttg ggcattgaga gccgtccggt gttccacccg atgcacatca tgccgcccta
10621 tgcgcatctg gccacggatg atctgaagat cgccgaagcc tgcggggtcg acggcttgaa
10681 cctgccgacc cacgcggggc tgactgaagc cgatatcgac cgtgtcatcg cggcgctcga
10741 tcaggtgttg gtctagccga tgcgcatcgt cctgctgtcc tcgatcgtgc cgttcatcaa
10801 cggcggcgcg cgcttcatcg tcgagtggct cgaggagaag ctgatcgagg ccggccacga
10861 ggtcgagcgg ttctacctgc cgtttgtcga cgatccgaac gagatcctgc accagatcgc
10921 cgcctggcgg ctgatggacc tgacccagtg gtgcgaccgg gtgatctgct tccggccgcc
10981 ggcctatgtg gtggaccatc cgaacaaggt cttgtggttc atccaccaca tccgcacctt
11041 ctacgacctg tgggacacgc cctatcgcgg catgcctgac gacgcgcagc accgggccat
11101 ccgcgacaat ctccgcgcgc tcgacaccca ggcgatttcg gaagcccgcg cggtgttcac
11161 caactcccag gtggtggccg accgcttgaa ggcgttcaac ggcctggacg ccacgccgct
11221 gtatccgccg atctatcagc ccgagcgctt ttcccatacc ggctatggcg acgagatcgt
11281 ggccatctcg cggctggagc cgcacaagcg tcaggccctg atgatcgagg ccatgcagta
11341 cgtgaagagc ggcgtgaagc tgcgcctggc gggcacggcg tccagcgccg agtatggtcg
11401 acagctggtc aagatgaccc acgacctggg cgtcgccgac cgggtcattc tcgaggatcg
11461 ctggatcagc gaggacgaga aggccgatat gctgaaacag gctctggccg tggcctatct
11521 gcccaaggac gaggacagct atggctatcc ttcgctggag ggcgctcacg cccgcaagcc
11581 ggtgatcacc acgaccgact ccggcggggt gctggaactg gtcgagcatg gccgcaacgg
11641 cttgatcagc gccccggacc cgcgcgcgct ggccgagcag ttcgaccgcc tgcacgctga
11701 caaggctgcg acagccaaga tggggaccgc ctcgctgaac cgtctggccg agatgaagat
11761 cgactggagc accgtcgtgg agcgcctgac ctcatgagaa cgcccgcatg aaggttctgg
11821 tcgtcaacaa cgccgcgccg ttccaacgcg gcggcgccga ggagctggcc gaccatctgg
11881 tccgccgcct gaacgccacg cccggcgtcc agtccgagct ggtgcgcgtg cccttcacct
11941 gggagccggc cgagcgtctg atcgaggaga tgctgatctc caaggggatg cggctctaca
12001 atgtggaccg ggtcattggc ctcaaatttc cggcctatct gatcccgcat caccaaaagg
12061 tgctgtggct gctgcaccag ttccgtcagg cctacgacct gtccgaagcg ggccagagcc
12121 atctggactt cgacgacacg ggcagggcgg tgaaggcggc gatccgcgcg gccgacaacg
12181 cctgcttcgc cgagtgccgc aagatctact gcaactcgcc cgtcacccag aaccgcctga
12241 tgaagttcaa cggggtcgcc agccaggtgc tctatccgcc gctgaacgac ggtgagctgt
12301 tcaccggcgg cgagcatggc gactatgtct tcgcgggcgg ccgggtcgcg gcgggcaagc
12361 gccagcacct gttgattgag gccctagcct tgctgcccgg cagtctgcgg ctggtgatcg
12421 ccggaccgcc ggagaaccag gcctatgccg accgcctgac caagctggtc gaggatctgg
12481 atctgaagga tcgcgtcgag ctgcggtttg gcttccatcc gcgcgaggac atcgcccgtt
12541 gggccaacgg ggccctgatc tgcgcctatc tgccctttga cgaggatagt gtaggttacg
12601 tcacgatgga ggccttcgcc gcaggcaagg ccgtgctgac cgtgaccgac tccggcggcc
12661 tgctggagat cgtcagcgcg gataccggtg cggtcgccga gcccacgccg caagccctgg
12721 ccgaggcgct tgatcgtttg acctcggaca aggcgcgggc gatatcgctg ggcgacgcgg
12781 cgcgcaggct atggcgcgac aagaatgtca catgggaaga gacggtccgc cgtcttcttg
12841 attaagccac aaacattggg ttgaagacca cgttaagacg gggtcggcta cagtctagga
12901 aatgcgagtg ctgtcgaaag ttctgtccgt gcgaacgtct ctgatcgcct tggccatggc
12961 catggcggtc gtcggtcgcg ctgatctcgc ccacgccgag accttggccg aggcgatcac
13021 cgcagcctat cagagcaatc cgaatattca ggcccaacgc gccgccatgc gcgcgctgga
13081 cgagaactac acccaggccc gttcggccta tgggctgcaa gccagcgcct cggtcgctga
13141 ggtctatggc tggtccaagg gcgtcaacgc caagaacggc gtcgaggccg ccagccagac
13201 ctcgaccctc tctctgagcc agagcctcta caccaacggt cgtttctcgg cccgcctggc
13261 gggtgtcgag gcgcagatca aggccgcgcg cgagaacctg cgccgcatcg agatggacct
13321 gctggtccgc gtgaccaacg cctatatctc ggtgcgccgc gaccgcgaga tcctgcggat
13381 cagccaaggc ggtgaagcct ggctgcagaa gcaattgaag gacaccgagg acaagtacag
13441 cgtccgtcag gtgaccttga ccgacgtgca gcaggccaag gcccgcctgg cgtcggccag

```
13501 cactcaggtg gcgaacgccc aggcgcagct gaatgtcagc gtagcgttct acgcgtccct
13561 ggtggggcgc cagccggaga cgctgaagcc tgaacccgat attgacggcc tgcctacaac
13621 cctcgacgag gcgttcaatc aggccgaaca agccaatccg gtcctgctgg cggcgggcta
13681 caccgagaag gcctctcgcg ccggcgtcgc cgaggcgcgg gcccagcgcc tgttctcggt
13741 cggcgcgcgc gcggactatc gcaatggctc cagcacgccg tactacgcgc gtggcggtct
13801 gcgcgaggac accgtcaacg cctcgatcac cctgacccag ccgctgttca ccagcggtca
13861 gctgaacgcc tcggtgcggc agtcgatcga ggagaacaac cgcgacaagc tgctgatgga
13921 agacgcacgt cgcagcatgg tcctgagcgt ctcgcagtac tgggacagcc tggtggccgc
13981 gcggaagtcg ctggtcagcc tcgaagagga aatgaaggcc aacacgatcg ccttctatgg
14041 ggtgcgcgaa gaagagcgtt tcgcgcttcg cagcacgatc gaagtgctga acgcccaagc
14101 cgaattgcag aacgcccaga tcaatttcgt ccgcgggcgc gccaacgagt atgtcggtcg
14161 gctccatctt ctggcgcagg tcggcacgct tgaggtcggc aatctcgctc ccggcgtcca
14221 gccctacgat cctgagcgta acttcaggaa ggtccggtac cgcggcgctt tgccgacgga
14281 gctgatcatc ggaaccttcg acaagatcgc cttgccgctc gagcccaaga agccggcgcc
14341 gggggacacc tcgccgatcc ggccgccgtc gagcgaactg ccggccaggc ctgtttcggc
14401 cgacaaggtg acgccgccgg cgtcgatgaa cgatctgccc gccctgaccg acgacacgcc
14461 ggtccagacc gcgccccgca actagagccc tttccgatcg gatcgcctca atccgatcgg
14521 atcaaagggg ttctatttca gaagtttaga acgttcttcg gtcaccgttt cgcacggatc
14581 ggagaacgct ctaggtcagg ggcagccgcg cgctcgtcat cgccacccag gcgtcgcggt
14641 tgaacgagac gtgaccgttc cgcgccgtcg cggcggccaa ggtgttgcgc caggcgtcgt
14701 cggagatcat ccgcgccagc ttttcggcgg cgtcgcccac atcagcgtcg gcccagcgcc
14761 cggcgcgata gatgacgcca tcgccctcca cgggaaccag ccgatagtcg acgaactggc
14821 tggacgccgg gtccatgaac tccacgttcg atgaccagcc tgtcgccagg gtgggcttgc
14881 cgagccatat cgcttcggcg agcagcaggc cgtaaccctc ggaccggtgc agggacagca
14941 cgatgtcgct gctggcggtg agggcggcca tgtcctgcgc cgacaggttg tccgtcagaa
15001 gccggatgga cggatcgtcg gccacctcgg cggcgagcgc ttgaaaggtc tcgggataga
15061 ggtccgcgcc gacgacttta cagacaaggg ttgcgggacg tgtcgccttg actgtggcgt
15121 cccgaaaggc gcgcagcgcc gcgtcgggat tcttgcgctg ggcggtcgaa cgcagatcga
15181 aggccatcag cacgacgacc cggtcgtcct gcaggccaaa gcgctggcga tcggcctgcg
15241 ggcgcggatt gagatagagc gggtagggga ccggcgtaac ctttacgccc cgcggcgcca
15301 cgcgacggat agcgtcagcg gcgaaggcgg agggggtcca gacctcgtcg acgaaatcga
15361 aggcggggag ccagtcggac gggacctctt ccagctccca ggcccagtag ccgatgtgcc
15421 ggcggccctc caggaaacgg ccttcggtct cccgcgccca tcgcagaagc tctgggggat
15481 tgatgtgcga gatcgtgacc ccgcgttcgt cgggatccgg tgaagaatag gccggggcga
15541 tctcggcggc gaagccgacg gacgccgaca ggtccaaggc gcggacccca aggcccatgt
15601 cgccaaaccc gcgcgccagc atccgtgcgc cttcgcccag accgtgcacc gcgctatgga
15661 agccgatgac cgtcaccggg agcgtgggtg acaaacccct gggaccggcg cgtctgcgcg
15721 cgcgccaa ggcggttctt tccgcactcc ggtgtcggtg ctcggccaag gccttgtcca
15781 gcgcctccag agcggcgcgc ggggcgccgg cgacgacatg cgccgagcgc cggatgggcg
15841 tgggcaggcg gcgccatatc tcaagcaggc gtgaggtcat gggcgatgca ggcgggcgag
15901 acgcatgggc gaccgtatag ccgttcaggc cgagccgaca ccagcaaagc tgcgcggggc
15961 cgcgtcaggc gatcttgtca tgcgaaaggt ttgcatgcag caatggcggc ccactgcgcc
16021 gaggggggaa tacgtgtcga ctgagtggag cgccggctat gtcacggacg tcaactatac
16081 gtttggctat tatggcgagt tgaacccgct ccgctgccgc cttccgctcc tgacggtcgg
16141 ccgccacgct cccaagatcg agaacgcctg cgagctcggg tttggccagg gtctttccgt
16201 ctccatccac gccgcagccc agccggggat caactggtac ggaaccgact tcaacccctc
16261 acaggcggcc ttcgccgcag agatggtccg tctgtcgggc gcggaggcca agctctacga
16321 cgaggccttc gccgagttct gtaatcgcaa ggacctgccg gacttcgact tcatcgggct
16381 ccacggcatc tggacctgga tctccgacca aaatcgccac gtgctggtcg acttcattcg
16441 tcgcaagctc cgcccggg
//
```

Appendix 2

ATC15252 S-layer subunit and transporter genes

```
LOCUS       JS3001A19    4255 bp    mRNA            BCT       07-OCT-1999
DEFINITION  Caulobacter crescentus S-layer subunit (rsaA) and ABC-transporter
            (rsaD(partial)) mRNAs, complete cds.
ACCESSION   JS3001A19
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 4255)
  AUTHORS   Bingle,W.H., Awram,P.A., Nomellini,J.F. and Smit,J.K.
  TITLE     The Secretion Signal of C. crescentus S-layer Protein is Located in
            the C-terminal 82 Amino Acids of the Molecule
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4255)
  AUTHORS   Bingle,W.H., Awram,P.A., Nomellini,J.F. and Smit,J.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Microbiology and Immunology, University of
            British Columbia, 300-6174 University Blvd,. Vancouver, BC V6T 1Z3,
            Canada
FEATURES             Location/Qualifiers
     source          1..4255
                     /organism="Caulobacter crescentus"
                     /strain="JS3001"
     gene            637..3717
                     /gene="rsaA"
     CDS             637..3717
                     /gene="rsaA"
                     /codon_start=1
                     /transl_table=11
                     /product="S-layer subunit"
                     /translation="MAYTTAQLVTAYTNANLGKAPDAATTLTLDAYATQTQTGGLSDA
                     AALTNTLKLVNSTTAVAIQTYQFFTGVAPSAAGLDFLVDSTTNTNDLNDAYYSKFAQE
                     NRFINFSINLATGAGAGATAFAAAYTGVSYAQTVATAYDKIIGNAVATAAGVDVAAAV
                     AFLSRQANIDYLTAFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAA
                     MINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVA
                     GAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSG
                     VTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVASTGVTSGTT
                     TVGANSAASGTVSVSVANSSTTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTG
                     NSSTTAVTVTQTAAATAGATVAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATI
                     DSSALTTVNLSGTGTSLGIGRGALTATPTANTLTLNVNGLTTTGAITDSEAAADDGFT
                     TINIAGSTASSTIASLVAADATTLNISGDARVTITSHTAAALTGITVTNSVGATLGAE
                     LATGLVFTGGAGADSILLGATTKAIVMGAGDDTVTVSSATLGAGGSVNGGDGTDVLVA
                     NVNGSSFSADPAFGGFETLRVAGAAAQGSHNANGFTALQLGATAGATTFTNVAVNVGL
                     TVLAAPTGTTTVTLANATGTSDVFNLTLSSSAALAAGTVALAGVETVNIAATDTNTTA
                     HVDTLTLQATSAKSIVVTGNAGLNLTNTGNTAVTSFDASAVTGTGSAVTFVSANTTVG
                     EVVTIRGGAGADSLTGSATANDTIIGGAGADTLVYTGGTDTFTGGTGADIFDINAIGT
                     STAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAATLAQYLDAAAAGDGSG
                     TSVAKWFQFGGDTYVVVDSSAGATFVSGADAVIKLTGLVTLTTSAFATEVLTLA"
```

```
     gene            3960..4253
                     /gene="rsaD(partial)"
     CDS             3960..4253
                     /gene="rsaD(partial)"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC-transporter"
                     /translation="MFKRSGAKPTILDQAVLVARPAVITAMVFSFFINILALVSPLYM
                     LQVYDRVLTSRNVSTLIVLTVICVFLFLVYGLLEALRTQVLVRGGLKFDGVARD"
BASE COUNT      722 a   1512 c   1296 g    725 t
ORIGIN
        1 aagcttcccc aagcctaggt gaaaagccga ccccccgtcg gcccaaacac gctagcagac
       61 gacacgataa ccgaactagt cttcgctgaa caggatctcg taggtgatcg gatcatagaa
      121 gcggagcagt tcgcgcacga acgtcttctc cgagacctcc agggccttgg cccagtcgcg
      181 gtaacggtcc ggcggaatgc ggccgcgacc cgtctccagc tgagagatga aggtgtaata
      241 atcagcgccg accttagcgg ccagctggcg ttgcgacagg ccggcggcct cgcgcatctc
      301 cttcagccag cggccacctt cgcggcggag gtcttgcacc tccgaggcgc tgcggcgttg
      361 cgggttacca tacattataa agcctcgcgc gttgaccgag ggcaggagcg cgggcgcgct
      421 cactcacccg ccaggtgaac agtctttata tatagcgctt ttcggcgggg ggtacaagga
      481 acgctatata ggaatttgct gtaccggtta gaaaaatgct gtacccctga aattcggcta
      541 ttgtcgacgt atgacgtttg ctctatagcc atcgctgctc ccatgcgcgc cactcggtcg
      601 caggggggtgt gggatttttt ttgggagaca atcctcatgg cctatacgac ggcccagttg
      661 gtgactgcgt acaccaacgc caacctcggc aaggcgcctg acgccgccac cacgctgacg
      721 ctcgacgcgt acgcgactca aacccagacg ggcggcctct cggacgccgc tgcgctgacc
      781 aacaccctga agctggtcaa cagcacgacg gctgttgcca tccagaccta ccagttcttc
      841 accggcgttg ccccgtcggc cgctggtctg gacttcctgg tcgactcgac caccaacacc
      901 aacgacctga acgacgcgta ctactcgaag ttcgctcagg aaaaccgctt catcaacttc
      961 tcgatcaacc tggccacggg cgccggcgcc ggcgcgacgg ctttcgccgc cgcctacacg
     1021 ggcgtttcgt acgcccagac ggtcgccacc gcctatgaca agatcatcgg caacgccgtc
     1081 gcgaccgccg ctggcgtcga cgtcgcggcc gccgtggctt tcctgagccg ccaggccaac
     1141 atcgactacc tgaccgcctt cgtgcgcgcc aacacgccgt tcacggccgc tgccgacatc
     1201 gatctggccg tcaaggccgc cctgatcggc accatcctga acgccgccac ggtgtcgggc
     1261 atcggtggtt acgcgaccgc cacggccgcg atgatcaacg acctgtcgga cggcgccctg
     1321 tcgaccgaca acgcggctgg cgtgaacctg ttcaccgcct atccgtcgtc gggcgtgtcg
     1381 ggttcgaccc tctcgctgac caccggcacc gacaccctga cgggcaccgc caacaacgac
     1441 acgttcgttg cgggtgaagt cgccggcgct gcgaccctga ccgttggcga caccctgagc
     1501 ggcggtgctg gcaccgacgt cctgaactgg gtgcaagctg ctgcggttac ggctctgccg
     1561 accggcgtga cgatctcggg catcgaaacg atgaacgtga cgtcgggcgc tgcgatcacc
     1621 ctgaacacgt cttcgggcgt gacgggtctg accgccctga acaccaacac cagcggcgcg
     1681 gctcaaaccg tcaccgccgg cgctggccag aacctgaccg ccacgaccgc cgctcaagcc
     1741 gcgaacaacg tcgccgtcga cggcggcgcc aacgtcaccg tcgcctcgac gggcgtgacc
     1801 tcgggcacga ccacggtcgg cgccaactcg gccgcttcgg gcaccgtgtc ggtgagcgtc
     1861 gcgaactcga gcacgaccac cacgggcgct atcgccgtga ccggtggtac ggccgtgacc
     1921 gtggctcaaa cggccggcaa cgccgtgaac accacgttga cgcaagccga cgtgaccgtg
     1981 accggtaact ccagcaccac ggccgtgacg gtcacccaaa ccgccgccgc caccgccggc
     2041 gctacggtcg ccggtcgcgt caacggcgct gtgacgatca ccgactctgc cgccgcctcg
     2101 gccacgaccg ccggcaagat cgccacggtc accctgggca gcttcggcgc cgccacgatc
     2161 gactcgagcg ctctgacgac cgtcaacctg tcgggcacgg gcacctcgct cggcatcggc
     2221 cgcggcgctc tgaccgccac gccgaccgcc aacaccctga ccctgaacgt caatggtctg
     2281 acgacgaccg gcgcgatcac ggactcggaa gcggctgctg acgatggttt caccaccatc
     2341 aacatcgctg gttcgaccgc ctcttcgacg atcgccagcc tggtggccgc cgacgcgacg
     2401 accctgaaca tctcgggcga cgctcgcgtc acgatcacct cgcacaccgc tgccgccctg
     2461 acgggcatca cggtgaccaa cagcgttggt gcgaccctcg gcgccgaact ggcgaccggt
     2521 ctggtcttca cgggcggcgc tggcgctgac tcgatcctgc tgggcgccac gaccaaggcg
     2581 atcgtcatgg gcgccggcga cgacaccgtc accgtcagct cggcgaccct gggcgctggt
     2641 ggttcggtca acggcggcga cggcaccgac gttctggtgg ccaacgtcaa cggttcgtcg
     2701 ttcagcgctg acccggcctt cggcggcttc gaaaccctcc gcgtcgctgg cgcggcggct
```

```
     2761 caaggctcgc acaacgccaa cggcttcacg gctctgcaac tgggcgcgac ggcgggtgcg
     2821 acgaccttca ccaacgttgc ggtgaatgtc ggcctgaccg ttctggcggc tccgaccggt
     2881 acgacgaccg tgaccctggc caacgccacg ggcacctcgg acgtgttcaa cctgaccctg
     2941 tcgtcctcgg ccgctctggc cgctggtacg gttgcgctgg ctggcgtcga gacggtgaac
     3001 atcgccgcca ccgacaccaa cacgaccgct cacgtcgaca cgctgacgct gcaagccacc
     3061 tcggccaagt cgatcgtggt gacgggcaac gccggtctga acctgaccaa caccggcaac
     3121 acggctgtca ccagcttcga cgccagcgcc gtcaccggca cgggctcggc tgtgaccttc
     3181 gtgtcggcca acaccacggt gggtgaagtc gtcacgatcc gcggcggcgc tggcgccgac
     3241 tcgctgaccg gttcggccac cgccaatgac accatcatcg gtggcgctgg cgctgacacc
     3301 ctggtctaca ccggcggtac ggacaccttc acgggtggca cgggcgcgga tatcttcgat
     3361 atcaacgcta tcggcacctc gaccgctttc gtgacgatca ccgacgccgc tgtcggcgac
     3421 aagctcgacc tcgtcggcat ctcgacgaac ggcgctatcg ctgacggcgc cttcggcgct
     3481 gcggtcaccc tgggcgctgc tgcgaccctg gctcagtacc tggacgctgc tgctgccggc
     3541 gacggcagcg gcacctcggt tgccaagtgg ttccagttcg gcggcgacac ctatgtcgtc
     3601 gttgacagct cggctggcgc gaccttcgtc agcggcgctg acgcggtgat caagctgacc
     3661 ggtctggtca cgctgaccac ctcggccttc gccaccgaag tcctgacgct cgcctaagcg
     3721 aacgtctgat cctcgcctag gcgaggatcg ctagactaag agaccccgtc ttccgaaagg
     3781 gaggcggggt ctttcttatg ggcgctacgc gctggccggc cttgcctagt tccggtggct
     3841 atgatttagc gggactgggg ggcttgctca ctttccgcca caatttcgtg gtcgagacgg
     3901 cgccttagtt gttactgtac atggccgcgt cggttcgcgc ggcgtcctga aggctcacaa
     3961 tgttcaagcg cagcggcgcg aagccgacga tcctcgacca ggccgtgctg gtcgcccgcc
     4021 cggcggtgat caccgccatg gtcttcagct tcttcatcaa cattctggcc ctggtcagcc
     4081 cgctgtacat gctgcaggtc tatgaccgcg tgctgaccag ccgcaacgtt tcgaccctga
     4141 tcgtgttgac ggtcatctgc gtcttcctgt tcctggtcta cggcctgctc gaggcgctgc
     4201 gcacccaggt gctggtgcgc ggcggtctga agttcgacgg cgtggcccgg gatcc
//

LOCUS       JS4000RAT1   7493 bp    DNA             BCT       07-OCT-1999
DEFINITION  Caulobacter crescentus S-layer subunit (rsaA(truncated)),
            ABC-transporter (rsaD), and Membrane Forming Unit (rsaE) genes,
            complete cds.
ACCESSION   JS4000RAT1
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 7493)
  AUTHORS   Bingle,W.H., Awram,P.A., Nomellini,J.F. and Smit,J.K.
  TITLE     The Secretion Signal of the C. crescentus S-layer protein is
            Located within the C-Terminal 82 Amino Acids of the Molecule
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7493)
  AUTHORS   Bingle,W.H., Awram,P.A., Nomellini,J.F. and Smit,J.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Microbiology and Immunology, University of
            British Columbia, 300-6174 University Blvd, Vancouver, BC V6T 1Z3,
            Canada
FEATURES             Location/Qualifiers
     source          1..7493
                     /organism="Caulobacter crescentus"
                     /strain="JS4000"
     gene            637..1716
                     /gene="rsaA(truncated)"
     CDS             637..1716
```

```
                     /gene="rsaA(truncated)"
                     /note=" The RsaA protein is truncated because of a deleted
                     G basepair.  A stop codon results after translation of 359
                     amino acids."
                     /codon_start=1
                     /transl_table=11
                     /product="S-layer subunit"
                     /translation="MAYTTAQLVTAYTNANLGKAPDAATTLTLDAYATQTQTGGLSDA
                     AALTNTLKLVNSTTAVAIQTYQFFTGVAPSAAGLDFLVDSTTNTNDLNDAYYSKFAQE
                     NRFINFSINLATGAGAGATAFAAAYTGVSYAQTVATAYDKIIGNAVATAAGVDVAAAV
                     AFLSRQANIDYLTAFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAA
                     MINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVA
                     GAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIETMNVTSGAAITLNTSSG
                     VTGLTALNTNTSGAAQTVTAGAGQT"
     gene            3959..5695
                     /gene="rsaD"
     CDS             3959..5695
                     /gene="rsaD"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC-transporter"
                     /translation="MFKRSGAKPTILDQAVLVARPAVITAMVFSFFINILALVSPLYM
                     LQVYDRVLTSRNVSTLIVLTVICVFLFLVYGLLEALRTQVLVRGGLKFDGVARDPIFK
                     SVLDSTLSRKGIGGQAFRDMDQVREFMTGGLIAFCDAPWTPVFVIVSWMLHPFFGILA
                     IIACIIIFGLAVMNDNATKNPIQMATMASIAAQNDAGSTLRNAEVMKAMGMWGGLQAR
                     WRARRDEQVAWQAAASDAGGAVMSGIKVFRNIVQTLILGGGAYLAIDGKISAGAMIAG
                     SILVGRALAPIEGAVGQWKTYIGARGAWDRLQTMLREEKSADDHMPLPEPRGVLSAEA
                     ASILPPGAQQPTMRQASFRIDAGAAVALVGPSAAGKSSLLRGIVGVWPCAAGVIRLDG
                     YDIKQWDPEKLGRHVGYLPQDIELFSGTVAQNIARFTEFESQEVIEAATLAGVHEMIQ
                     SLPMGYDTAIGEGGASLSGGQRQRLALARAVFRMPALLVLDEPNASLDQVGEVALMEA
                     MKRLKAAKRTVIFATHKVNLLAQADYIMVINQGVISDFGERDPMLAKLTGAAPPQTPP
                     PTPPPAPLQRVQ"
     gene            5763..7073
                     /gene="rsaE"
     CDS             5763..7073
                     /gene="rsaE"
                     /codon_start=1
                     /transl_table=11
                     /product="Membrane Forming Unit"
                     /translation="MKPPKIQRPTDNFQAVARIGYGIIALTFVGLLGWAAFAPLDSAV
                     IANGVVSAEGNRKTVQHLEGGMLAKILVREGEKVKAGQVLFELDPTQANAAAGITRNQ
                     YVALKAMEARLLAERDQRPSISFPADLTRLRADPMVARAIADEQAQFTERRQTIQGQV
                     DLMNAQRLQYQSEIEGIDRQTQGLKDQLGFIEDELIDLRKLYDKGLVPRPRLLALERE
                     QASLSGSIGRLTADRSKAVQGASDTQLKVRQIKQEFFEQVSQSITETRVRLAEVTEKE
                     VVASDAQKRIKIVSPVNGTAQNLRFFTEGAVVRAAEPLVDIAPEDEAFVIQAHFQPTD
                     VDNVHMGMVTEVRLPAFHSREIPILNGTIQSLSQDRISDPQNKLDYFLGIVRVDVKQL
                     PPHLRGRVTAGMPAQVIVPTGERTVLQYLFSPLRDTLRTTMREE"
BASE COUNT     1261 a   2627 c   2358 g   1247 t
ORIGIN
        1 aagcttcccc aagcctaggt gaaaagccga ccccccgtcg gcccaaacac gctagcagac
       61 gacacgataa ccgaactagt cttcgctgaa caggatctcg taggtgatcg gatcatagaa
      121 gcggagcagt tcgcgcacga acgtcttctc cgagacctcc agggccttgg cccagtcgcg
      181 gtaacggtcc ggcggaatgc ggccgcgacc cgtctccagc tgagagatga aggtgtaata
      241 atcagcgccg accttagcgg ccagctggcg ttgcgacagg ccggcggcct cgcgcatctc
      301 cttcagccag cggccacctt cgcggcggag gtcttgcacc tccgaggcgc tgcggcgttg
      361 cgggttacca tacattataa agcctcgcgc gttgaccgag ggcaggagcg cgggcgcgct
      421 cactcacccg ccaggtgaac agtctttata tatagcgctt ttcggcgggg ggtacaagga
```

```
 481 acgctatata ggaatttgct gtaccggtta gaaaaatgct gtacccctga aattcggcta
 541 ttgtcgacgt atgacgtttg ctctatagcc atcgctgctc ccatgcgcgc cactcggtcg
 601 caggggtgt gggatttttt ttgggagaca atcctcatgg cctatacgac ggcccagttg
 661 gtgactgcgt acaccaacgc caacctcggc aaggcgcctg acgccgccac cacgctgacg
 721 ctcgacgcgt acgcgactca aacccagacg ggcggcctct cggacgccgc tgcgctgacc
 781 aacaccctga agctggtcaa cagcacgacg gctgttgcca tccagaccta ccagttcttc
 841 accggcgttg ccccgtcggc cgctggtctg gacttcctgg tcgactcgac caccaacacc
 901 aacgacctga acgacgcgta ctactcgaag ttcgctcagg aaaaccgctt catcaacttc
 961 tcgatcaacc tggccacggg cgccggcgcc ggcgcgacgg ctttcgccgc cgcctacacg
1021 ggcgtttcgt acgcccagac ggtcgccacc gcctatgaca agatcatcgg caacgccgtc
1081 gcgaccgccg ctggcgtcga cgtcgcggcc gccgtggctt tcctgagccg ccaggccaac
1141 atcgactacc tgaccgcctt cgtgcgcgcc aacacgccgt tcacggccgc tgccgacatc
1201 gatctggccg tcaaggccgc cctgatcggc accatcctga acgccgccac ggtgtcgggc
1261 atcggtggtt acgcgaccgc cacggccgcg atgatcaacg acctgtcgga cggcgccctg
1321 tcgaccgaca acgcggctgg cgtgaacctg ttcaccgcct atccgtcgtc gggcgtgtcg
1381 ggttcgaccc tctcgctgac caccggcacc gacaccctga cgggcaccgc caacaacgac
1441 acgttcgttg cgggtgaagt cgccggcgct gcgaccctga ccgttggcga caccctgagc
1501 ggcggtgctg gcaccgacgt cctgaactgg gtgcaagctg ctgcggttac ggctctgccg
1561 accggcgtga cgatctcggg catcgaaacg atgaacgtga cgtcgggcgc tgcgatcacc
1621 ctgaacacgt cttcgggcgt gacgggtctg accgccctga acaccaacac cagcggcgcg
1681 gctcaaaccg tcaccgccgg cgctggccaa acctgaccgc cacgaccgcc gctcaagccg
1741 cgaacaacgt cgccgtcgac ggcggcgcca acgtcaccgt cgcctcgacg ggcgtgacct
1801 cgggcacgac cacggtcggc gccaactcgg ccgcttcggg caccgtgtcg gtgagcgtcg
1861 cgaactcgag cacgaccacc acgggcgcta tcgccgtgac cggtggtacg gccgtgaccg
1921 tggctcaaac ggccggcaac gccgtgaaca ccacgttgac gcaagccgac gtgaccgtga
1981 ccggtaactc cagcaccacg gccgtgacgg tcacccaaac cgccgccgcc accgccggcg
2041 ctacggtcgc cggtcgcgtc aacggcgctg tgacgatcac cgactctgcc gccgcctcgg
2101 ccacgaccgc cggcaagatc gccacggtca ccctgggcag cttcggcgcc gccacgatcg
2161 actcgagcgc tctgacgacc gtcaacctgt cgggcacggg cacctcgctc ggcatcggcc
2221 gcggcgctct gaccgccacg ccgaccgcca acaccctgac cctgaacgtc aatggtctga
2281 cgacgaccgg cgcgatcacg gactcggaag cggctgctga cgatggtttc accaccatca
2341 acatcgctgg ttcgaccgcc tcttcgacga tcgccagcct ggtggccgcc gacgcgacga
2401 ccctgaacat ctcgggcgac gctcgcgtca cgatcacctc gcacaccgct gccgccctga
2461 cgggcatcac ggtgaccaac agcgttggtg cgaccctcgg cgccgaactg gcgaccggtc
2521 tggtcttcac gggcggcgct ggcgctgact cgatcctgct gggcgccacg accaaggcga
2581 tcgtcatggg cgccggcgac gacaccgtca ccgtcagctc ggcgaccctg ggcgctggtg
2641 gttcggtcaa cggcggcgac ggcaccgacg ttctggtggc caacgtcaac ggttcgtcgt
2701 tcagcgctga cccggccttc ggcggcttcg aaaccctccg cgtcgctggc gcggcggctc
2761 aaggctcgca caacgccaac ggcttcacgg ctctgcaact gggcgcgacg gcgggtgcga
2821 cgaccttcac caacgttgcg gtgaatgtcg gcctgaccgt tctggcggct ccgaccggta
2881 cgacgaccgt gaccctggcc aacgccacgg gcacctcgga cgtgttcaac ctgaccctgt
2941 cgtcctcggc cgctctggcc gctggtacgg ttgcgctggc tggcgtcgag acggtgaaca
3001 tcgccgccac cgacaccaac acgaccgctc acgtcgacac gctgacgctg caagccacct
3061 cggccaagtc gatcgtggtg acgggcaacg ccggtctgaa cctgaccaac accggcaaca
3121 cggctgtcac cagcttcgac gccagcgccg tcaccggcac gggctcggct gtgaccttcg
3181 tgtcggccaa caccacggtg ggtgaagtcg tcacgatccg cggcggcgct ggcgccgact
3241 cgctgaccgg ttcggccacc gccaatgaca ccatcatcgg tggcgctggc gctgacaccc
3301 tggtctacac cggcggtacg gacaccttca cgggtggcac gggcgcggat atcttcgata
3361 tcaacgctat cggcacctcg accgctttcg tgacgatcac cgacgccgct gtcggcgaca
3421 agctcgacct cgtcggcatc tcgacgaacg gcgctatcgc tgacggcgcc ttcggcgctg
3481 cggtcaccct gggcgctgct gcgaccctgg ctcagtacct ggacgctgct gctgccggcg
3541 acggcagcgg cacctcggtt gccaagtggt tccagttcgg cggcgacacc tatgtcgtcg
3601 ttgacagctc ggctggcgcg accttcgtca gcggcgctga cgcggtgatc aagctgaccg
3661 gtctggtcac gctgaccacc tcggccttcg ccaccgaagt cctgacgctc gcctaagcga
3721 acgtctgatc ctcgcctagg cgaggatcgc tagactaaga gaccccgtct tccgaaaggg
3781 aggcggggtc tttcttatgg gcgctacgcg ctggccggcc ttgcctagtt ccggtggcta
3841 tgatttagcg ggactggggg gcttgctcac tttccgccac aatttcgtgg tcgagacggc
```

3901 gccttagttg ttactgtaca tggccgcgtc ggttcgcgcg gcgtcctgaa ggctcacaat
3961 gttcaagcgc agcggcgcga agccgacgat cctcgaccag gccgtgctgg tcgcccgccc
4021 ggcggtgatc accgccatgg tcttcagctt cttcatcaac attctggccc tggtcagccc
4081 gctgtacatg ctgcaggtct atgaccgcgt gctgaccagc cgcaacgttt cgaccctgat
4141 cgtgttgacg gtcatctgcg tcttcctgtt cctggtctac ggcctgctcg aggcgctgcg
4201 cacccaggtg ctggtgcgcg gcggtctgaa gttcgacggc gtggcccggg atccgatctt
4261 caagtcggtg ctggactcca cgctcagccg caagggcatc ggcggccagg cgttccgcga
4321 catggaccag gtccgagagt tcatgaccgg cggcctgatc gccttctgcg atgcgccctg
4381 gacgccggtg ttcgtcatcg tctcgtggat gctgcacccg ttcttcggca tcctggcgat
4441 catcgcctgc attatcatct tcggcctggc cgtgatgaac gacaacgcca ccaagaaccc
4501 gatccagatg gccaccatgg cctcgatcgc cgcccagaac gacgccggtt ccaccctgcg
4561 caacgccgag gtcatgaagg ccatgggcat gtggggcggc ctgcaagccc gctggcgcgc
4621 gcgccgcgac gagcaggtgg cctggcaggc cgccgccagc gacgccggcg gcgcggtgat
4681 gtcgggcatc aaggtgttcc gcaacatcgt ccagaccctg atcctgggcg gcggcgccta
4741 tctggccatc gacggcaaga tctcggccgg cgcgatgatc gccggctcga tcctggtcgg
4801 ccgcgccctg gcgcccatcg agggcgcggt gggccagtgg aagacctata tcggcgcgcg
4861 cggcgcctgg gatcgtctgc agaccatgct gcgcgaggaa aagagcgccg acgaccacat
4921 gccgctgccc gagccgcgcg gcgtgctgtc ggccgaagcc gcctcgatcc tgccgccggg
4981 cgcgcaacag ccgaccatgc gccaggccag cttccgcatc gacgccggcg ccgcggtggc
5041 ccttgtcggt cccagcgcgg cgggcaagtc ctcgctgctg cgcggcatcg tcggcgtctg
5101 gccctgcgcg gccggcgtca tccgcctcga cggctacgac atcaagcagt gggatcccga
5161 gaagctgggt cgccacgtcg gctacctgcc gcaggacatc gagctgttct cgggcaccgt
5221 cgcccagaac atcgcccgct tcaccgagtt cgagtcgcag gaagtcatcg aggccgcgac
5281 cctggcgggc gtgcacgaga tgatccagag cctgccgatg ggctatgata cggcgatcgg
5341 cgagggcggc gcctcgctgt ccggcggcca gcgccagcgc ctggccctgg cccgcgcggt
5401 gttccgcatg ccggccctgc tggtgctgga cgagccgaac gccagcctcg accaggtggg
5461 cgaagtggcg ctgatggaag cgatgaagcg gctcaaggcc gccaagcgca cggtgatctt
5521 cgccacccac aaggtgaacc tgttggccca ggccgactac atcatggtga tcaaccaggg
5581 tgtgatcagc gactttggcg aacgcgaccc gatgctggcc aagctgaccg gggctgcgcc
5641 gccccagacg ccgccgccga cgccgccgcc cgcgccgttg cagcgcgtcc agtaagcgcc
5701 ttcgtctgtc cgcctctccc ttcctggccg tttcagaacg cgcccatcag gcttgaatct
5761 caatgaagcc ccccaagatc cagcgtccga cggacaactt ccaggctgtg gcccgtatcg
5821 gctacggcat catcgccctg acctttgtcg gtctgttggg ctgggccgcg ttcgccccgc
5881 tcgacagcgc ggtgatcgcc aacggcgtcg tctccgccga gggtaatcgc aagaccgtgc
5941 agcacctcga aggcggcatg ctggccaaga tcctggtccg cgaaggcgag aaggtgaagg
6001 ccggccaggt gctgttcgag ctggacccga cccaggccaa cgccgccgcc ggcatcaccc
6061 gcaaccagta tgtggcgttg aaggccatgg aagcgcgcct gctggccgag cgcgaccagc
6121 gtccgtccat cagcttcccc gccgacctga cccgcctgcg cgccgatccg atggtcgccc
6181 gcgccatcgc cgacgaacag gcccagttca ctgagcgtcg ccagacgatc cagggccagg
6241 tcgacctgat gaacgcccag cgtttgcagt atcagagcga gatcgagggc atcgaccgtc
6301 agacccaggg cctgaaggac caactcggct tcatcgagga cgagctgatc gacctgcgta
6361 agctctatga caagggcctg gtgccccggc cgcgtctgct ggccctggag cgcgagcagg
6421 cctcgctgtc gggctcgatc ggccgtctga ccgcagaccg ctccaaggcc gtccagggcg
6481 cctctgacac ccagctcaag gttcgccaga tcaagcagga gttcttcgag caggtcagcc
6541 agagcatcac cgagacccgg gttcgcctgg ccgaggtgac cgagaaggag gtcgtcgcct
6601 ccgacgccca gaagcggatc aagatcgtgt cgccggtcaa cgggacggcg cagaacctgc
6661 gcttcttcac cgagggcgct gtcgttcgcg ccgccgagcc gctggtcgac atcgcgcccg
6721 aggacgaggc cttcgtgatc caggcgcact tccagccgac cgatgtggac aatgtccata
6781 tgggcatggt caccgaagtt cggctgccgg ccttccactc gcgggaaatc ccgatcctga
6841 acggcacgat ccagtctctg tcgcaggacc gcatttccga tccgcagaac aagctcgact
6901 acttcctcgg gatcgtgcgc gtggacgtca agcagctgcc gccgcatctg cgcggcaggg
6961 tcaccgccgg catgccggcc caggtgatcg tgccgaccgg cgagcgcacc gtgctgcagt
7021 acctgttctc gccgctgcga gacaccctgc gcaccacgat gcgcgaggag taggtcaaag
7081 gtttcaaggg cctgatttcc aaagcttttc ggatgggcgg cgcgggcgag ctaaattcgc
7141 cggcgctgct ttccattcgc gggcaatagt gtagtcagga cccttcgttg ttactggagt
7201 cagcggatac gcatggcgaa aacggctttg atcaccggtg tgaccggtca ggacggggcg
7261 tacctcgcca agctgctgct ggagaagggt tacaccgtcc acggcatgct gcgtcgctcg

```
     7321 gcctcggccg atgtgatcgg cgaccgcctg cgctggatcg gcgtctatga cgacatccag
     7381 ttcgagctgg gcgacctctt ggacgagggc ggtctggcgc gcctgatgcg gcgcctgcag
     7441 ccggatgagg tctacaacct ggcggcccag agcttcgtcg gcgcctcgtg gga
//
```

Appendix 3

Sequences of *lpsGHIJK*, orf1 and orf2

```
LOCUS       gcc227       4883 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc227.
ACCESSION   gcc227
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 4883)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4883)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..4883
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(1..1242)
                     /gene="orf3"
     CDS             complement(1..1242)
                     /gene="orf3"
                     /codon_start=1
                     /product="putative glycolipid transporter"
                     /translation="MSAAASTPQEYKRLTQYEVDVICAKHDRLWSARMGGARAVFAFC
                     DLSGLSVPGRNLCDADFTGAILVGCDLRKAKLDNANFYGADLQGADLTDASLRRADLR
                     GSSLRGANLTGADMFEADLREGTIAAADRKEGYRVIEPTQREAFAAGANLSGANLERS
                     RLSGIVATKADFSDAILKDAKLVRANLKQANFNGANLAGADLSGANLAGADLRNAVLV
                     GAKTLSWNVNDTNMDGALTDKPSGTSVSDLPYEQMIADHARWIETGGGEGKPSVFDKA
                     DLRNLRSVRGFNLTALSAKGSVFYGLDMEGVQMQGAQLDGADLRACNLRRADLRGARL
                     KGAKLTGADLRDAQLGPLLIAADRLLPVDLTGAILTNADLARADLRQARMAGADVSRA
                     NFTGAQLRDLDLTGAIRLAARG"
     gene            complement(1335..2048)
                     /gene="orf4"
     CDS             complement(1335..2048)
                     /gene="orf4"
                     /codon_start=1
                     /product="putative phosphoglycerate mutase"
                     /translation="MPTLVLLRHGQSQWNLENRFTGWVDVDLTAEGEAQARKGGELIA
                     AAGIEIDRLFTSVQTRAIRTGNLALDAAKQSFVPVTKDWRLNERHYGGLTGLNKAETA
                     EKHGVEQVTIWRRSYDIPPPELAPGGEYDFSKDRRYKGASLPSTESLATTLVRVLPYW
                     ESDIAPHLKAGETVLIAAHGNSLRAIVKHLFNVPDDQIVGVEIPTGNPLVIDLDAALK
                     PTGARYLDDSRAEALPKVG"
     gene            2421..3377
                     /gene="orf5"
     CDS             2421..3377
```

```
                     /gene="orf5"
                     /codon_start=1
                     /product="putative sugar phosphate isomerase  KpsF-like"
                     /translation="MSAFNAVQVGRRVLAVEADALRVLADSLGEAFANAVETIFNAKG
                     RVVCTGMGKSGHVARKIAATLASTGTQAMFVHPAEASHGDLGMIGPDDVVLALSKSGA
                     GRELADTLAYAKRFSIPLIAMTAVADSPLGQAGDILLLLPDAPEGTAEVNAPTTSTTL
                     QIALGDAIAVALLERRGFTASDFRVFHPGGKLGAMLRTVGDLMHGADELPLVAADAAM
                     PDALLVMSEKRFGAVGVVDNAGHLAGLITXGDLRRHMDGLLTHTAGEVMTHAPLTIGP
                     GALAAEALKVMNERRITVLFVVERERPVGILHVHDLLRAGVI"
     gene            3477..4883
                     /gene="lpsg"
     CDS             3477..4883
                     /gene="lpsg"
                     /codon_start=1
                     /product="phosphomannomutase"
                     /translation="MFSSPRADLVPNTAAYENEALVKATGFREYDARWLFGPEINLLG
                     VQALGLGLGTYIHELGQSKIVVGHDFRSYSTSIKNALILGLISAGCEVHDIGLALSPT
                     AYFAQFDLDIPCVAMVTASHNENGWTGVKMGAQKPLTFGPDEMSRLKAIVLNAEFVER
                     DGGKLIRVQGEAQRYIDDVAKRASVTRPLKVIAACGNGTAGAFVVEALQKMGVAEVVP
                     MDTDLDFTFPKYNPNPEDAEMLHAMADAVRETGADLAFGFDGDGDRCGVVDDEGEEIF
                     ADKIGLMLARDLAPLHPGAXFVVXVKSTGLYATDPILAQHGCKVIYWKTGHSYIKRKS
                     AELGALAGFEKSGHFFMNGELGYGYDCGLTAAAAILAMLDRNPGVKLSDMRKALPVAF
                     TSLTMSPHCGDEVKYGVVADVVKEYEDLFAAGGSILGRKITEVITVNGVRVHLEDGSW
                     VLVRASSNKPEVVVVVESS"
BASE COUNT      807 a   1647 c   1587 g    837 t      5 others
ORIGIN
        1 gccgcgcgcc gccagcctga tggcgccggt cagatccagg tcgcgtagct gcgcgccggt
       61 aaagttggcg cgcgagacat ccgcgccggc cattcgggcc tgccgcaggt cggcgcgcgc
      121 gagatcggcg ttggtaagaa tcgctcccgt cagatcgacc ggcagcaagc ggtcggctgc
      181 gatgagcagc ggccccagct gtgcgtcacg caggtccgct cccgtcagct tggcgccctt
      241 caaccgcgcg cctcgcaggt cggcgcggcg caggttgcag gctcgcagat cggcgccatc
      301 caactgcgcg ccctgcatct gcacgccttc catgtcgagg ccgtagaaca ccgacccctt
      361 cgccgacagg gccgtgagat tgaagcctcg gacagatcgc agattccgca gatcggcctt
      421 gtcgaacacc gagggcttgc cctcgccgcc gccggtctcg atccatcgtg cgtggtcggc
      481 gatcatctgc tcatacggca ggtcagagac gcttgtgccc gacggcttgt cggtcaacgc
      541 gccgtccata ttggtgtcgt tgacgttcca ggacagggtc ttggccccaa ccagcacggc
      601 gttgcgcaga tcagcgccgg cgaggttagc gcccgacaga tcggcccccg ccaggttggc
      661 gccgttgaaa ttggcctgtt tgaggttggc ccgaaccagc ttggcgtcct tcaggatagc
      721 gtcactgaag tccgccttcg tcgcgacgat gcccgacagg cgcgaacgct cgaggttggc
      781 gcccgacagg ttcgcgccgg ccgcgaaggc ttcgcgttgg gtcggttcga tgacgcgata
      841 accttccttg cggtcggcgg cggcgatcgt gccctcccgc agatcggcct cgaacatgtc
      901 ggcgccggtc aggttggcgc cgcgcagact ggatccgcgc aggtcagccc gccgcaacga
      961 ggcgtcggtc aagtcggcgc cttgcagatc cgcgccatag aagttggcgt tgtccagctt
     1021 ggcctttcgc aggtcacagc cgacgagaat ggcgccggtg aaatcggcgt cgcagagatt
     1081 gcgccccggc acggagaggc cagacaggtc gcagaacgcg aaaaccgcgc gcgcgccccc
     1141 catccgcgcc gaccacagac ggtcatgctt ggcgcagata acgtccactt cgtactgcgt
     1201 caggcgcttg tattcttgtg gcgtgctggc ggcggcgctc atggctcgct gggtatcccg
     1261 agttaacggg acaccttgcg cagagcaaaa ttaacgccgt gtttccaaca ccttccgctt
     1321 aagctttcgg cgttttaacc aactttggga agcgcctcag cccgactgtc gtccaggtag
     1381 cgcgcgccgg tcggcttcaa cgcagcatcc aggtcgatca ccagcggatt gccggtcgga
     1441 atctccacac cgacgatctg gtcgtccggc acgttgaaca gatgcttgac gatggcgcgc
     1501 agcgagttgc cgtgggcggc gatcagcaca gtctcacccg ccttgagatg cggagcgatg
     1561 tcgctttccc agtagggcag aacgcgaacc agcgtcgtcg ccaggctttc cgtgctgggc
     1621 aggcttgcgc ccttgtagcg gcgatccttg ctgaagtcgt actcgccgcc cggcgccagc
     1681 tccggcggcg ggatatcgta cgaacggcgc cagatggtca cctgctcgac gccgtgcttt
     1741 tcggcggtct cagccttgtt caggccggtc agcccgccat agtggcgctc gttcaggcgc
     1801 cagtccttgg tcacggggac gaagctctgc ttggcggcgt ccagcgccag attgcctgtg
```

```
1861 cggatggcgc gggtctgaac cgaggtgaac aagcggtcga tctcgatgcc ggccgcagcg
1921 atcagctcgc cgcccttccg ggcctgagct tcgccctcag cggtgaggtc cacatcaacc
1981 caaccggtga agcggttttc caggttccac tggctttggc catggcgcag caggacgagc
2041 gtcggcatcg ggcttccttc ggagatcagg gaaatgtcag ggcgggctaa ggccagcccg
2101 cttcagcgtc aagcgccgaa gcgaccaagg acccgatcgg cccggtgcgc gccccctccc
2161 ccaggccgcc gcgcgtgcta taggcgagcc atgcctgatc gcatcttcat gcctctgatg
2221 ggactcatcg ccgccgcgct gatcgccctc tctctcgtct ggccccaggg ccagggcgac
2281 cgctcctggg gccccttcgg ccacacgccc gttcagcaga ccccggagat gaaggccaag
2341 attcagcggg aaaaggactc ggcgcgccgt cgggacgagg cggcgaaaaa ggctgtcgag
2401 gccatggatc agctttcgca atgagcgcct tcaacgctgt tcaggtgggc cgccgcgtcc
2461 tcgccgtaga agccgatgcg ctgcgcgtcc tcgcagactc gctgggcgag gccttcgcca
2521 atgcggtcga gacgatcttc aacgccaagg gtcgcgtcgt ctgcacgggc atgggcaagt
2581 cggggcatgt ggcgcggaag atcgccgcca ccctcgcttc gaccggcacc caggcgatgt
2641 tcgtccaccc cgccgaagcc tcgcacggcg acctgggcat gatcgggcca gacgacgtgg
2701 ttctggcgct gtccaagtcg ggcgccggcc gcgaactggc cgacaccctg gcctacgcta
2761 agcgcttctc gatcccgctg atcgccatga ccgccgtggc cgacagcccg ctcggtcagg
2821 cgggcgacat cctgctgctg ctccccgacg cgcccgaggg gacggcggaa gtgaacgccc
2881 cgaccacgtc gaccaccctg cagatcgcgc tgggcgacgc gatcgctgtg gctctgctgg
2941 agcggcgcgg cttcaccgcc agcgacttcc gcgtcttcca ccccggcggc aagctcggcg
3001 ctatgctgcg cacggtcggc gacctgatgc acggcgccga tgagcttccc ctggtcgccg
3061 ccgacgccgc catgcccgac gctttactgg tcatgagcga aaagcgtttc ggcgcggtcg
3121 gcgtcgttga taacgcgggt cacctggccg gcttgatcac gkacggtgat ctgcgtcgac
3181 acatggatgg gctgctgacc cacaccgccg gcgaggtcat gacgcacgct cccctgacca
3241 tcggccccgg cgccctggcg gctgaagcgc tgaaggttat gaacgagcgg cggatcaccg
3301 tgcttttcgt cgtcgagcgc gaacgccccg tcggcattct acatgtgcac gacctgcttc
3361 gcgcgggtgt gatctaggtc acatcgaaac tttgcaaaac cttgtcatgc gaacgcgcta
3421 gtcgtgccgc gtccgcacgg ctaaagccat accatctcaa ctgaagcgag ccttcaatgt
3481 tctcctcgcc ccgcgccgat ctggttccga atacggccgc ctacgaaaac gaagccctgg
3541 tcaaggcgac gggctttcgc gagtacgacg cgcgctggct gtttgggccg gagatcaatc
3601 tcctgggcgt gcaggccctg ggcctgggtc tgggaaccta tatccacgaa ctgggccaat
3661 cgaagatcgt ggtcggccat gacttccgct cgtattcgac ctcgatcaag aacgccctga
3721 tcctggggct gatcagcgcc ggctgcgagg tgcacgacat tggcctggcc ctgtcgccca
3781 ccgcctattt cgcccagttc gacctcgaca tcccgtgcgt ggccatggtc acggccagcc
3841 acaacgaaaa cggctggacc ggcgtgaaga tgggcgccca gaagccgctg accttcggcc
3901 ccgacgagat gagccgcctc aaggccatcg tgctgaacgc cgagttcgtc gagcgcgatg
3961 gcggcaagct gatccgcgtg cagggcgagg cccagcgcta tatcgacgac gtggccaagc
4021 gcgccagcgt cacccgtccc ctgaaggtga tcgccgcctg cggcaacggc acggccggcg
4081 ccttcgtggt cgaggccctg cagaagatgg gtgtcgctga ggtcgtgccg atggacaccg
4141 acctcgactt caccttcccc aagtacaatc ccaaccccga agacgccgag atgctgcacg
4201 cgatggctga cgctgtccgt gagacgggcg cggacctggc gttcggcttc gacggcgacg
4261 gcgaccgctg cggtgtggtc gatgacgagg gcgaggagat cttcgccgac aagatcggcc
4321 tgatgctggc gcgcgacctg gccccgctac atccgggcgc grcyttcgtc gtgratgtga
4381 agtcgacggg cctmtacgcc accgatccga tcctggccca gcacggctgc aaggtgatct
4441 actggaagac cggccacagc tacatcaagc gcaagagcgc ggagctgggc gccctggccg
4501 gcttcgaaaa gagcggccac ttcttcatga acggcgagct gggctatggc tacgactgcg
4561 gcctgaccgc cgccgcggcc atcctggcca tgctggaccg caatcccggc gtgaagctgt
4621 cggacatgcg caaggccctg ccggtggcct tcacctccct gaccatgagc ccgcactgcg
4681 gtgacgaggt gaagtacggc gtcgtcgccg acgtggtgaa ggaatacgag gacctgttcg
4741 ccgccggcgg ttcgatcctg ggccgcaaga tcaccgaggt gatcacggtc aacggcgtgc
4801 gggtgcacct ggaggacggc tcctgggtcc tggtccgcgc ctcgtccaac aagcccgagg
4861 tcgtcgtcgt ggtcgaaagc agc
//

LOCUS       gcc506       8012 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc506.
```

```
ACCESSION   gcc506
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 8012)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 8012)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..8012
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(1845..3860)
                     /gene="orf6"
     CDS             complement(1845..3860)
                     /gene="orf6"
                     /codon_start=1
                     /product="putative transketolase"
                     /translation="MRVRPSRSPAKHIKTEAPMPVSPIKMADAIRVLSMDAVHKAKSG
                     HQGMPMGMADVATVLWGKFLKFDASKPDWADRDRFVLSAGHGSMLLYSLLHLTGFKAM
                     TMKEIENFRQWGALTPGHPEVHHTPGVETTTGPLGQGLATAVGMAMAEAHLAARYGSD
                     LVDHRTWVIAGDGCLMEGVSHEAISIAGRLKLSKLTVLFDDNNTTIDGVATIAETGDQ
                     VARFKAAGWAVKVVDGHDHGKIAAALRWATKQDRPTMIACKTLISKGAGPKEGDPHSH
                     GYTLFDNEIAASRVAMGWDAAPFTVPDDIAKAWKSVGRRGAKVRKAWEAKLAASPKGA
                     DFTRAMKGELPANAFEALDAHIAKALETKPVNATRVHSGSALEHLIPAIPEMIGGSAD
                     LTGSNNTLVKGMGAFDAPGYEGRYVHYGVREFGMAAAMNGMALHGGIIPYSGTFLAFA
                     DYSRAAIRLGALMEARVVHVMTHDSIGLGEDGPTHQPVEHVASLRAIPNLLVFRPADA
                     VEAAECWKAALQHQRTPSVMTLSRQKTPHVRTQGGDLSAKGAYELLAAEGGEAQVTIF
                     ASGTEVGVAVAARDILQAKGKPTRVVSTPCWELFDQQPAAYQAAVIGKAPVRVAVEAG
                     VKMGWERFIGENGKFIGMKGFGASAPFERLYKEFGITAEAVAEAALA"
     gene            4281..6041
                     /gene="orf7"
     CDS             4281..6041
                     /gene="orf7"
                     /codon_start=1
                     /product="putative NH(3)-dependent NAD(+) synthetase"
                     /translation="MIVVGGPLRDAGRLYNTAIVIQGGKVLGVVPKSFLPNYREFYER
                     RWFTPGAGLTGKTLTLAGQTVPFGTDILFRGEGVAPFTVGVEICEDVWTPTPPSTAQA
                     LAGAEILLNLSASNITIGKSETRRLLCASQSSRMIAAYVYSAAGAGESSTDLAWDGHV
                     DIHEMGALLAETPRFSTGPAWTFADVDVQRLRQERMRVGSFGDAMALSPASTPFRIVP
                     FAFDAPEGDLALARPIERFPFTPSDPARLRENCYEAYNIQVQGLARRLEASGLKKLVI
                     GISGGLDSTQALLVAAKAMDQLGLPRSNILAYTLPGFATSDRTKSNAWALMKAMAVTA
                     AELDIRPAATQMLKDLDHPFGRGEAVYDVTFENVQAGLRTDYLFRLANHNAALVVGTG
                     DLSELALGWCTYGVGDHMSHYNPNCGAPKTLIQHLIRFVAHSGDVGAETTALLDDILA
                     TEISPELVPGEAVQATESFVGPYALQDFNLYYMTRYGMAPSKIAFLAWSAWHDADQGG
                     WPVGLPDNARRAYDLPEIKRWLELFLKRFFANQFKRSAVPNGPKISSGGALSPRGDWR
                     MPSDATADAWLAELRTNAPI"
     gene            6121..7446
                     /gene="lpsH"
```

```
     CDS             6121..7446
                     /gene="lpsH"
                     /codon_start=1
                     /product="putative mannose-6-phosphate isomerase"
                     /translation="VWGQDLAAIYPVILCGGSGTRLWPASRSDHPKQFLKLVSDRSSF
                     QETVLRVKDIPGVAEVVVVTGEAMVGFVSEQTAEIGAWATILVEPEARDSAPAVAAAA
                     AYVEAQDPAGVVLMLAADHHIAQPEIFQQAALTATKAAEQGYIVTFGVQPTVPATGFG
                     YIRPGAPLLDGSVREVAAFVEKPDQATAERYLLEGYLWNSGNFAFQAATLLGEFETFE
                     PSVAAAAKACVAGLQLEAGIGRLDREAFAQAKKISLDYAIMERTQKAAVAPAAFAWSD
                     LGAWDAIWEASTRDGDGNAQTGDVDLHGSSNVLVRSTGPYVGVIGVNDIVVVAEPDAV
                     LVCHRKDSQAVKTLVDGLKAKGRSIASRKSASPNGTETLVSTDGFDVELRRVPAGETL
                     MLPVSTLQVLEGVIEMDGDVYAAGAIIALDDSVQARAIGAATLLVTKPR"
BASE COUNT     1292 a   2635 c   2741 g   1342 t      2 others
ORIGIN
        1 gcttgaagct tgcgcgaggc gaccgccacg gccgcgatga gcaggagcgc gccaaaccac
       61 agggcgttgg ccaggggtgc atcgggattg gcctgccggg ccacgatggg gatgaccagg
      121 agatagccga cgtacggcgt cgctgcggcg atggtcgctg agcgaaggcc gccgtggacg
      181 atgacgacgt tcatcagcgc cccggccagc agcatgccgc ccaaggtcgg cgcgaaccgg
      241 gcgtcggacg cgaacagggg aatcgccagg aatccgaaga cggcggaggt tagaaagggc
      301 agggcgatga cggagaagcg gcgccagcga tcagggctgc tctgcggtcg cttgatgagc
      361 tgaagcgcca gccaggtctc caggaactgg atcgcggtgt agagcgctcc ccagatccaa
      421 acccagcgca agccaacgac gaggtgcagg gtcgcgcaga tcgcaaaggt cgagctgagg
      481 cgcagatagg tattgacgcg tcgcgaggcg acagcgcgtt ccaggtcaat atccttcgtc
      541 tcagcgctca tgggcgctcc gccccccccc aaggagatcg cagagcaatc ccttacccaa
      601 taaccctagt cacgaggggc taagaggaaa tgaaagacca gcgtttcgtc aacggcttgg
      661 gtgtccggcg cgcaccggaa gatttttgcg cgttactgcc attggggcag gttttggggc
      721 ataaaatccc aataatgccg acccgcgcta gttcggcggc ccgaacgcgc gtgcgacatc
      781 ggctgggtcc acgccctcca gttcgagacg ctttagccgc gcggtttcgc cagccgccag
      841 tcgtacggcc gagcraggaa tcttcagggt ctttgccaga aaggcgatga gggccgcatt
      901 cgccgcgccc tcgacgggcg gactggcgac cctcaccttc agatagaggc ggccgtcggc
      961 gtcaagcgcc caaccctcgg cggcgtctcg cccgcccctc ggggtcaggc gcaccacgag
     1021 cgtcaccgcc acgccgtcag ccaagaagag cgatcagcgt gccttgaagc gcaggcagca
     1081 gatagttctg cacgcccgaa atgatcagca gaacaacgat cgggctgatg tcgacgccgc
     1141 cgagcgacgg gatcaggcgt tggaacgggc gaagcaccgg tccggtcacg cgatccagga
     1201 aatccagcac ctggtacaca gcggtgttgc ggcggttgat cacgtcgaac gcgaccagcc
     1261 agctgaggat cgccgagatg acgatggccc accaaaggag gctgagcagg ccgccgagga
     1321 tgaagaaaac gaattgaatg atggcggtca tgggtttccc gtccggactt tccgcttctt
     1381 gccgcgccac gcgcgcgctg gcaagcaagg agccgcttga cgagcgcgct ttggggccgt
     1441 atctccgccg ctccttgggg ccgtagctca gatggtagag cgcctcgttc gcaatgagga
     1501 ggtcaggggt tcgattcccc tcggctccac caaggacctc ccttagcctc gatcacgacc
     1561 cgatcggatc gcccatcgga ctccgtccgc attttcttgc ctaacgtcgc gtgcgccccc
     1621 gcgacgactc gccacagcct gggttcgagg aacagcggcg cctcgccggc aagatcggcc
     1681 agggtcccca tcgctgtggg gcaggcatgg ggccagggtc aggcgcgagc ctgctctgta
     1741 ggcgaccatc aggtgtttct gtttcggacc gcgctatcga tccgccctgg cccttattct
     1801 gcggtcgggg gcggagctcc ggactccacc cccgcgcagc gatttcaggc cagagccgct
     1861 tcggccacgg cttcagcggt gatgccgaac tctttataca ggcgctcgaa cggagccgag
     1921 gcgccgaagc ccttcatgcc gatgaacttg ccgttctcgc caatgaagcg ctcccagccc
     1981 atcttgacgc ccgcttcgac ggcgacgcgc accggggcct tgccgatgac ggcggcctgg
     2041 taggcggcgg gctgctgatc gaacagctcc cagcagggcg tggagaccac gcgggtcggc
     2101 ttgcccttgg cctgcaggat gtcgcgcgcg gcgacggcga cgccgacctc ggtgcccgag
     2161 gcgaagatcg tcacctgcgc ctcgccgccc tcggccgcca gcagctcgta ggcgcccttg
     2221 gccgacaagt cgccgccctg ggtgcggacg tgcggggtct tctggcgcga cagggtcatc
     2281 accgacggcg tgcgttgatg ttgcagggcc gccttccagc actcggcggc ctcgacggca
     2341 tcggccggac ggaagaccag caggttcgga atggcgcgca agctggcaac gtgctcgacc
     2401 ggctggtggg tgggaccgtc ttcgccgaga ccgatggagt cgtgggtcat cacgtggacg
     2461 acgcgggcct ccatcagggc gcccaggcgg atggccgcgc ggctgtagtc ggcgaaggcc
     2521 aggaaggtgc ccgaataggg gatgatcccg ccgtgcaggg ccatgccgtt catggccgcg
```

2581 gccatgccga actcacgcac gccatagtgg acatagcggc cttcgtagcc gggcgcgtcg
2641 aacgcgccca tgcccttgac cagggtgttg ttcgagccgg tcaggtcggc cgagccgccg
2701 atcatctcgg ggatcgccgg gatcaggtgc tccagggccg agccggagtg gacgcgggtg
2761 gcgttgaccg gcttggtctc cagggccttg gcgatgtggg cgtccagcgc ctcgaaggcg
2821 ttcgccggca gctcgccctt catggcgcgg gtgaagtcgg ccccccttggg cgaggcggcc
2881 agcttggcct cccaggcctt gcggaccttg gcgccgcgac ggccgacgct cttccaggcc
2941 ttggcgatgt cgtcgggcac ggtgaagggc gcagcgtccc agcccatggc cacgcgcgag
3001 gcggcgatct cgttgtcgaa cagggtgtag ccgtggctgt gggggtcgcc ttccttgggg
3061 cccgcgccct tcgagatcag cgtcttgcac gcgatcatgg tcgggcggtc ctgcttggtg
3121 gcccagcgca gggccgcagc gatcttgccg tggtcgtggc cgtcgacgac cttgaccgcc
3181 cagccggcgg ccttgaagcg cgcgacctgg tcgccggtct cggcgatggt ggccaccccg
3241 tcgatggtgg tgttgttgtc gtcgaagagg accgtcagct tcgagagctt caggcggccg
3301 gcgatgctga tcgcctcatg gctgacgccc tccatcaggc atccgtcgcc ggcgatcacc
3361 caggtgcggt ggtcgacgag gtcagagccg tagcgggcgg ccaggtgcgc ctcggccatg
3421 gccatgccga cggcggtggc caggccctgg cccagcggac cggtcgtggt ctcgacgccg
3481 ggcgtgtgat gcacttccgg gtggcccggg gtcagcgccc cccactgacg gaagttctcg
3541 atctccttca tcgtcatggc cttgaagccg gtcagatgca gcagggaata gagcagcatc
3601 gagccgtgac cggccgacag cacgaagcgg tcgcggtcgg cccagtcagg cttagacgcg
3661 tcgaatttca ggaacttgcc ccataggacc gtcgccacgt cggccatgcc catcggcatg
3721 ccctggtggc cggacttcgc cttgtgcacg gcgtccatgg agaggacgcg gatcgcgtcg
3781 gccatcttga tgggcgaaac gggcatgggg gcttccgtct ttatatgttt tgcagggctg
3841 cgcgaggggc gcactcgcat gggaaccccc gggggtcaac ccgcgcaggg cggctaaggc
3901 cgtgcggaac gcctgtcgtg gacgctatgc aacatcgtcc gtcggcgtta taggtggagg
3961 cagacttatg ctcagtttca gcataacccg gaaaggccgc tcccttgggt agtccgtcgt
4021 tcttctcgcc ctaccgtcac ggtttcgtcc gggtcgcgac cgccgttccg aaggtcaagc
4081 tggcggatcc cgccgccaat gctcagaacg tcgtggctct ggcccgcgag gcccatgcgg
4141 agggcgtggc tgtggtcgtg ttcccggaac tggggctgac gggctacacg atcgacgacc
4201 ttctgcagca agaggcgttg ctggacgcgg ttgaggccgc gatcgccacc ctgaccgagg
4261 ccagcgcagg cctggcgccg atgatcgtgg tcggaggtcc gctgcgcgac gcaggccgcc
4321 tctacaatac cgcgatcgtc atccagggcg gcaaggtgct gggcgtggtc ccgaaaagct
4381 tcctgcccaa ctatcgcgag ttctacgagc gtcgctggtt cacgccgggc gccggcttga
4441 caggcaagac cctgaccctg gccggccaga ccgttccgtt cgggaccgac attctgttcc
4501 ggggcgaggg cgtcgccccg ttcacggtgg gcgtcgagat ctgcgaggat gtctggaccc
4561 cgaccccgcc cagcaccgcc caggccttgg cgggggccga gatcctgctg aacctgtcgg
4621 ccagcaacat caccatcggc aagtccgaaa cgcggcgtct gctctgcgcc agccagtcgt
4681 cgcggatgat cgcggcctat gtctattcgg cggccggcgc gggcgagagc tcgaccgacc
4741 tggcctggga cggccatgtc gatattcacg agatgggcgc gctgctcgcc gagaccccgc
4801 ggttttcgac gggcccggcc tggaccttcg ccgatgtgga cgtccagcgc cttcggcagg
4861 agcggatgcg cgtcggcagc ttcggcgacg ccatggcgtt atcgccggcc tcgaccccgt
4921 tccggatcgt tccgttcgcc tttgacgcgc ccgagggcga cctggcgctg gcccggccga
4981 tcgaacgctt tcccttcacg ccgtccgacc cagccaggct gcgcgagaac tgctacgagg
5041 cctacaacat ccaggtccag ggcctggcgc ggcgcctcga ggcttcgggt ctcaagaagc
5101 tcgtcatcgg tatttccggc gggctcgact ccacccaggc tctgctggtg gcggccaagg
5161 ccatggacca gctgggcctg ccgcgcagca acatcctggc ctacactctg ccgggctttg
5221 cgacgtccga tcgcaccaag tccaacgcct gggcgctgat gaaggcgatg gccgtcaccg
5281 ccgccgagct cgatatccgg cccgcagcga cccagatgct caaggacctc gaccacccgt
5341 tcgggcgcgg cgaggcggtc tatgacgtca ccttcgagaa tgtgcaggcc ggcctgcgaa
5401 ccgactatct gttccgtctg gccaaccaca acgccgccct ggtcgtcggc acgggggacc
5461 tgtcggagct ggcgctgggc tggtgcacct acggcgtcgg cgaccacatg agccactaca
5521 accccaactg cggtgcgccc aagacgctga tccagcacct gatccgcttc gtggcccatt
5581 cgggtgacgt cggcgccgag accacggctc tgctggacga catcctcgcg accgagatct
5641 cgccggagct ggtgcccggc gaggcggttc aggcgaccga gagcttcgtc ggcccctacg
5701 ccctgcagga cttcaatctc tactacatga cccgctacgg catggcgccg tccaagatcg
5761 ccttcctggc ctggagcgcc tggcatgacg ccgaccaggg cggctggccc gtcggcctgc
5821 ccgacaacgc tcgccgcgcc tacgacctgc ctgagatcaa gcgctggctg gagctgttcc
5881 tgaagcggtt cttcgccaac cagttcaagc gctcggctgt acccaacggg ccgaaaatct
5941 cgtcgggcgg cgcgttgtcg ccgcgggggg actggcgcat gccgtcggat gcgacagccg

```
     6001 atgcctggct ggcggaactg cgcacaaatg cgccgatttg aggaaaactc ttcgttacag
     6061 ctctgtttgc ctttaagcag tcgacgcaat agcacccgat aaggggcgaa gactaagact
     6121 gtgtggggac aagacttggc tgcgatctat ccggtaatcc tgtgtggcgg ctcgggcacc
     6181 cgcctctggc ccgcatcgcg gagcgaccat cccaaacagt tccttaaact cgtgagcgat
     6241 cggtcctcct tccaggagac tgtcctgcgg gtgaaggata ttccgggtgt ggccgaggtg
     6301 gtcgtcgtga ccggcgaggc gatggtcggg tttgtgtccg agcagaccgc cgagatcggc
     6361 gcctgggcca caatcctggt cgaacccgag gctcgcgaca gcgcgccggc cgtggcggcg
     6421 gcggcggcct atgtcgaggc ccaggatccg gccggcgtcg tgttgatgct ggccgccgac
     6481 caccacatcg cccagcccga aatcttccag caggccgccc tcaccgccac taaggcggcc
     6541 gagcagggct atatcgtcac gttcggggtt cagccgacgg tcccggcgac cggctttggt
     6601 tatatccgcc ctggcgcgcc gcttctggat ggttcggtgc gtgaggtcgc cgccttcgtc
     6661 gagaagcccg accaggcgac cgccgagcgc tatcttctgg aaggctatct ctggaacagc
     6721 ggcaatttcg cgttccaggc ggcgaccttg ctgggcgagt tcgagacctt tgaaccgtcg
     6781 gtcgccgccg ccgccaaggc gtgcgtggcc ggcctgcagc tggaggccgg catcggccgc
     6841 ctggatcgcg aggccttcgc ccaggccaag aagatctcgc tcgactacgc catcatggag
     6901 cgcacccaga aggccgctgt cgcccctgcg gcgttcgcct ggtcggacct tggggcctgg
     6961 gacgcgatct gggaggcctc cacccgcgac ggcgacggta acgcccagac gggcgacgtc
     7021 gacttgcacg gctcgtccaa tgttctggtg cgctcgacgg gtccctatgt cggcgtgatc
     7081 ggggtcaacg acatcgtcgt cgtggccgag cccgacgcgg tgctggtctg ccatcgcaag
     7141 gacagccagg cggtgaagac cctggtcgat ggcctgaagg ccaagggccg ctccatcgcc
     7201 tcgcgcaaga gcgcctcgcc gaacgggacc gagaccctgg tctcgaccga cggcttcgat
     7261 gtggagttgc gtcgcgtacc ggcgggagag accttgatgc tgccggtatc gacgcttcag
     7321 gtgctggaag gcgtgatcga gatggacggc gacgtctatg ctgcgggcgc gatcatcgcc
     7381 ttggacgact cggttcaggc tcgggcgatc ggggcggcga ccttgctggt cacgaagccg
     7441 cgttgatcac ccggtccatc tccaggatcg ccaaggcgat atggtagaac gagctggccg
     7501 ggacgggctc ttcgataaag gtttcgtcgg gctggtactt gtcgcgccag aggccgggga
     7561 tcgccgtgcg cagataggcc atcaggccct cagcggcggc ggcggccatg tcccagtagc
     7621 gcgcttggcc ggtgatctcc gccgccagca cggccgcctt gatccgctcg gtttggggcc
     7681 acagtcgggc gccgtcgtcg tgggtcgaga aatcatcgag cagggcgttg atcgccacgc
     7741 ctcgcgatag gtcaacgccg tgggtctcgg cgtcatcgat catgcgcaag gccgccgccg
     7801 tagcgtcggc gcggccggca agctggcccc agcgcatcag cagccagccc cattcgaact
     7861 gatgccccgg ctcgcagatc cgacccgcga cgcctggcgc ggggttccag tcgaggtcga
     7921 agaactcacg gatctggccg ctgggcgcgt gaatgaacct ggagagcgcc agttcggcga
     7981 tctcgtcggn cagggtgcgc cagatcgggt cc
//


LOCUS       gcc433       9041 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc433.
ACCESSION   gcc433
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 9041)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9041)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..9041
```

```
                /organism="Caulobacter crescentus"
                /strain="NA1000"
gene            913..2295
                /gene="orf8"
CDS             913..2295
                /gene="orf8"
                /codon_start=1
                /product="putative Glucose-6-Phosphate 1-Dehydrogenase"
                /translation="MLLPSLYFLELDRLLPHDLRIIGVARADHDAASYKALVREQLGK
                RATVEEAVWNRLAARLDYVPANITSEEDTKKLAERIGAHGTLVIFFSLSPSLYGPACQ
                ALQAAGLTGPNTRLILEKPLGRDLESSKATNAAVAAVVDESQVFRIDHYLGKETVQNL
                TALRFANVLFEPLWDRSTIDHVQITIAETEKVGDRWPYYDEYGALRDMVQNHMLQLLC
                LVAMEAPSGFDPDAVRDEKVKVLRSLRPFTKETVAHDTVRGQYVAGVVEGGARAGYVE
                EVGKPTKTETFVAMKVAIDNWRWDGVPFFLRTGKNLPDRRTQIVVQFKPLPHNIFGPA
                TDGELCANRLVIDLQPDEDISLTIMNKRPGLSDEGMRLQSLPLSLSFGQTGGRRRIAY
                EKLFVDAFRGDRTLFVRRDEVEQAWRFIDGVSAAWEEASIEPAHYAAGTWGPQSAQGL
                ISPGGRAWKA"
gene            2298..2996
                /gene="orf9"
CDS             2298..2996
                /gene="orf9"
                /codon_start=1
                /product="putative 6-phosphogluconolactonase"
                /translation="MPFTPIKLEAFGSREDLYDAAASVLVGALTTAVARHGRVGFAAT
                GGTTPAPVYDRMATMTAPWDKVTVTLTDERFVPATDASSNEGLVRRHLLVGEAAKASF
                APLFFDGVSHDESARKAEAGVNAATPFGVVLLGVGPDGHFASLFPGNPMLDQGLDLAT
                DRSVLAVPPSDPAPDLPRLSLTLAALTRTDLIVLLVTGAAKKALLDGDVDPALPVAAI
                LKQDRAKVRILWAE"
gene            2997..4811
                /gene="orf10"
CDS             2997..4811
                /gene="orf10"
                /codon_start=1
                /product="putative phosphogluconate dehydratase"
                /translation="IAMSLNPVIADVTARIVARSKDSRAAYLANMDRAIENQPGRAKL
                SCANWAHAFAASPGVDKLRALDPNAPNIGIVSAYNDMLSAHQPLEAYPALIKDAARDV
                GATAQFAGGVPAMCDGVTQGRPGMELSLFSRDVIAMATAVALTHDAFDSALYLGVCDK
                IVPGLVIGALTFSHLPALFVPAGPMTSGLPNSEKARIRALYAEGKVGREELLAAESAS
                YHGPGTCTFYGTANTNQMLMELMGFHLPGSAFVHPNTPLREALVKESARRVAAVTNKG
                NEFIPVGRMIDEKSFVNGVVGLMATGGSTNLALHIIAMAAAAGVQLTLEDLDDISKAT
                PLLARVYPNGSADVNHFQAAGGMAFVIRELLKAGLVHEDVQTIAGAGLSLYAKEPVLE
                DGMLTWRDGAHESLDPAIVRPVSDPFSKEGGLRLMAGNLGRGVMKISAVKPEHHVIEA
                PCAVFQEQEDFIAAFKRGELDRDVVVVVRFQGPSANGMPELHNLSPSISVLLDRGHKV
                ALVTDGRMSGASGKTPAAIHVTPEAAKGGPLAYVQDGDVIRVNAETGELKIMVDEATL
                LARTPANVPASKPGFGRELFGWMRSGVGAADAGASVFA"
gene            5856..6926
                /gene="lpsI"
CDS             5856..6926
                /gene="lpsI"
                /codon_start=1
                /product="putative repressor similar to LacI"
                /translation="MAKYSPKRANRTGEGRKLSAKVTIHDVARESGVSIKTVSRVLNR
                EPNVKADTRDRVQAAVAALHYRPNISARSLAGAKAYLIGVFFDNPSPGYVTDVQLGAI
                ARCRQEGFHLIVEPIDSTADVEDQVAPMLTTLRMDGVILTPPLSDHPVVLAALEREGV
                AYVRIAPGDDFDRAPWVSMDDRLAAYEMTKHLVDLGHKDIAFIVGHPDHGASHRRHQG
                FLDAMRDSGLRVRDDRVAQGWFSFRSGFEAAEKLLGGADRPTAIFASNDDMALGVMAV
                ANRLRLDVPTQLSVAGFDDTPGAKITWPQLTTVRQPIHAMAGAAADMLMQGVEREEGA
```

```
                     PPPSRLLDFELVVRESTGPASH"
     gene            7224..9041
                     /gene="orf11"
     CDS             7224..9041
                     /gene="orf11"
                     /codon_start=1
                     /product="putative 1,4-B-D-glucan  glucohydrolase"
                     /translation="MLPRRFAFASALALTIACGSAGVVLAQTPPNATANPAVWPMSAS
                     PAAITDAKTEAFIAQLMSRMTVEEKVAQTIQADGASITPEELKKYRLGSVLVGGNSAP
                     DGNDRASPQRWIEWIRAFRAAALDKRGDRQEIPIIFGVDAVHGHNNVVGATIFPHNVG
                     LGAAHEPDLIRRIGEVTAKEMAATGADWTFGPTVAVPRDSRWGRAYEGYGENPEIVKA
                     YSGPMTLGLQGALEAGKPLAAGRVAGSAKHFLADGGTENGRDQGDAKISEADLVRLHN
                     AGYPPAIEAGILSVMVSFSSWNGVKHTGNKSLLTDVLKERMGFEGFVVGDWNAHGQVE
                     GCSNTSCAQAYNAGMDMMMAPDSWKGLYDNTLAQVKAGQIPMARIDDAVRRILRVKVK
                     AGLFEDKRPLEGKLELLGAPEHRAVAREAVRKSLVLLKNEGVLPLKSSARVLVAGDGA
                     DDIGKASGGWTLTWQGTGNKNSDFPHGQSIYAGVAEAVKAGGGSAELSVSGDFKQKPD
                     VAIVVFGENPYAEFQGDITSIEYQAGDKRDLALLKKLKAAGIPVVSVFLSGRPLWTNP
                     ELNASDAFVAAWLPGSEGGGVADVLVGDKAGKPRHDFQGK"
BASE COUNT     1492 a   3016 c   3052 g   1480 t      1 others
ORIGIN
        1 ggacgcgctg aacaccatca acgcgacccg cacgacgatc ctgccgcccg ccctgaacgc
       61 cgagcgtcgt ctggctaccg cgcgcgccct gatgggtctg ggacgctatg acgcggcgct
      121 ggatctggtc gaaaccgaca ccagccgtga cgggcaggag atccgcggcg agatcgcttg
      181 gaagcaacgc gcctggccgg ccgccggcgc gctctatgag cggtccctgg gcgaccgctt
      241 caggacaggc ggggcgctca gcgcggccga ggaggcgagg ttgctgcgcg cctcggtggc
      301 ctacagtctc gccgatgacg acgcggcgct cggtcgcctg cgcgcccgct ggtcgggctt
      361 catcgagaca gccagcaacc ctgaaggtct tcgcgtagcg ctacagggca tgtcgatggg
      421 cgcggtgtcc gcgtccgact ttggacgggt ttccgccgac aacgaggcct tcaatggctg
      481 gatcggtcgc ctcaaggaac gtttccggac aggacaaccg gccgggtcac ccgcccgcgc
      541 gggcggctag cgctcgctgc ggtgatcgca ggttcgcgcg cgttagaagg gcggcagggc
      601 aaggttcgcc tgaaagcgcg ctcaagaggc tttatggaag gcagggcgtt tccgaatggc
      661 cccgggaatc cctcaaacag ttcctctgaa accggattgg tcttatggcg ggcgccgggc
      721 gtttggccta gttagagccc cgccgggcgc cgatccgagg gcggcgaagc gcggtcgctt
      781 gcctttcccg cgcaggcatg tcatgacaac gttgtcatag ggtggacgac attggctaag
      841 aacaacgacg tcggcgagaa cggccgcgaa gtcttggtgc tgctgggcgg agcgggcgat
      901 ttggccctcc ggatgctgct gccttctctg tatttcctgg agctcgaccg actgctgccg
      961 cacgatctgc ggatcattgg cgtcgcccgg gccgaccatg acgcggccag ctacaaggcg
     1021 ctggtccgcg agcaactggg caagcgcgcg acagtggagg aggcggtttg gaatcgcctc
     1081 gccgcgcgcc tcgactacgt gcctgcgaac atcaccagtg aggaagacac caagaagctg
     1141 gccgaacgga tcggtgcgca tggcacgctg gtcatcttct tctcgctgtc gcccagcctc
     1201 tacggcccgg cttgccaggc tttgcaggcc gccggcctga cggggcccaa cacgcgcttg
     1261 atcctcgaaa agccgcttgg ccgcgatctc gaaagctcca aggccaccaa cgccgccgtc
     1321 gccgctgtgg tcgacgagag ccaagtgttc cgcatcgacc actatctggg caaggaaacc
     1381 gtccagaacc tgacggccct gcgcttcgcc aacgtgctgt tcgagcccct gtgggatcgc
     1441 agcacgatcg accatgtgca gatcaccatc gccgagaccg aaaaggtcgg cgaccgctgg
     1501 ccctactacg acgaatacgg cgcgctgcgg gacatggtgc agaaccacat gctgcaactg
     1561 ctgtgtctgg tcgccatgga agcgccctca ggcttcgatc ccgatgcggt gcgcgacgag
     1621 aaggtcaagg tgctgcgctc cctgcggccc ttcaccaagg agaccgtggc ccacgacacc
     1681 gtgcgtggcc agtacgtcgc cggtgtggtc gagggcggcg cgcgcgctgg ctatgtcgag
     1741 gaagtgggca agcccaccaa gaccgagact ttcgtggcca tgaaggtcgc gatcgacaac
     1801 tggcgttggg acggcgtgcc gttcttcctg cgcaccggca agaacctgcc ggaccgccgc
     1861 acccagatcg tcgtccagtt caagcctttg ccgcacaaca tcttcggtcc ggcgaccgat
     1921 ggcgagctgt gcgccaaccg cctagtcatc gacctgcagc cggacgaaga catctcgctg
     1981 acgatcatga acaagcgtcc gggtctctcg gacgagggca tgcgactgca gtcgctgccg
     2041 ctgtcgctgt cgtttggcca gaccggcggg cgccgtcgca tcgcttacga aaagctgttc
     2101 gtcgacgcct tccgcggcga ccgtacgctg ttcgtgcgtc gcgatgaggt cgagcaggcc
     2161 tggcgcttca tcgacggcgt ctcggcggcc tgggaagagg ccagtatcga accggcgcac
```

2221 tatgcggcgg gcacctgggg accgcagtcc gcccagggcc tgatctcgcc cggcggccga
2281 gcctggaagg cctgagcatg cccttcacgc ccatcaagct cgaagcattt gggtcccgcg
2341 aggacctcta tgacgcggcc gcctcggttc tggtcggcgc tttgacgacg gcggtcgctc
2401 gtcacggcag ggtcggcttc gccgccaccg gcggcacgac gccggcgccg gtctatgacc
2461 gcatggcgac catgaccgcc ccctgggaca aggtcacggt cacgctcacc gacgagcgct
2521 ttgttcccgc caccgacgcc agcagcaatg agggtctggt gcgtcgccac ctgctcgtgg
2581 gcgaggcggc caaggcctcg ttcgcgccgc tgttcttcga cggcgtgagc cacgacgaga
2641 gcgcgcgcaa ggccgaggcg ggcgtcaatg ccgccacccc gttcggcgtc gttctcctgg
2701 gcgtggggcc ggatgggcat ttcgcttcgc tgtttccggg caatccgatg ctggatcagg
2761 gtctggacct cgccaccgac cgttcggtgc tggccgtgcc gcccagcgat cccgcgccgg
2821 acctcccacg cctgagcctg accctggccg ccctgacccg caccgacctg atcgtgctgc
2881 tggtcaccgg cgcggccaag aaagctttgt tggacggcga cgttgatccg gccctgccgg
2941 tcgccgccat tctgaaacag gaccgcgcca aggtccgcat cctctgggcg gagtagatcg
3001 ccatgagcct gaatcccgtc atcgccgacg tcaccgcccg gatcgtggcg cgcagcaagg
3061 acagccgcgc ggcctatctc gccaacatgg atcgggcgat cgagaaccag ccggggcgcg
3121 ccaagctgtc ctgcgccaac tgggcccacg ccttcgccgc ctcgccgggc gtcgacaagc
3181 tccgtgctct ggatccgaac gcgccgaaca tcggcatcgt ctcggcctat aatgacatgc
3241 tgtcagccca ccagccgctg gaagcctatc ccgcgctgat caaggacgcc gcccgggacg
3301 tgggcgcgac cgcccagttc gccggcgggg tgccggccat gtgcgacggt gtcacccagg
3361 gccgtcccgg catggagctg tcgctgttct cgcgcgacgt gatcgccatg gcgaccgccg
3421 tggccctgac ccatgacgcc ttcgactcgg cgctgtatct gggcgtctgc gacaagatcg
3481 tgccgggcct ggtgatcggc gcactgacct tcagccatct gcccgccctg ttcgtgcccg
3541 ccggcccgat gacctcgggc ctgcccaaca gcgagaaggc ccgcatccgc gcgctctacg
3601 ccgagggcaa ggtcggtcgt gaggaactgc tggcggccga gagcgccagc tatcatggcc
3661 cgggcacctg caccttctat ggcacggcca acaccaacca gatgctgatg gagctgatgg
3721 gcttccattt gcctggctcg gccttcgtcc atcccaacac gccgctgcgt gaggccctgg
3781 tcaaggaatc cgcccgccgc gtggctgcgg tgaccaacaa gggcaatgaa ttcatcccgg
3841 tcggccggat gatcgacgag aagtcgttcg tcaacggcgt ggtcgggttg atggcgaccg
3901 gcggctcgac caacctggcg ctgcacatca tcgccatggc cgccgctgcg ggcgtgcaac
3961 tgaccctcga agacctggac gatatctcca aggccacgcc gctgctggcg cgcgtctatc
4021 cgaacggttc ggccgacgtg aaccacttcc aggccgccgg cggcatggct ttcgtgatcc
4081 gtgagctgct gaaggcgggt ctagtgcacg aagacgtcca gacgatcgcg ggcgccggcc
4141 tgtcgctgta cgcgaaggaa ccggtgctcg aggacggcat gctgacctgg cgtgacggcg
4201 ctcacgagag tctggatccc gccatcgtgc ggccggtctc cgacccgttc agcaaggaag
4261 gcggcctgcg cctgatggcg ggcaatctgg gccgcggcgt gatgaagatc tcggccgtga
4321 agcccgagca ccacgtgatc gaggcgccgt gcgccgtgtt ccaggaacag gaagacttca
4381 tcgccgcttt caagcgcggc gagctggatc gcgacgtggt cgtggtggtc cgcttccagg
4441 ggccgtccgc caacggcatg cctgaactgc ataacctgtc gccgtcgatc tcggtgttgc
4501 tggatcgcgg tcacaaggtg gccctggtca ccgacggccg catgtccggc gcctctggca
4561 agacgcccgc cgccatccac gtgacgccgg aagcggccaa gggcgggccg ctggcctatg
4621 tccaggacgg cgatgtgatc cgcgtcaatg ccgagaccgg ggaactgaag atcatggtgg
4681 acgaggcgac cctgctcgcc cggacccccg cgaacgtccc ggcgtccaag ccgggctttg
4741 gccgggaact gtttggatgg atgcggtcgg gggtcggcgc ggccgacgcc ggcgcctccg
4801 tctttgcttg aggaagcgct aggtcatgga cggcaatcac agcggcgggc tcggcctcgt
4861 cggcgacatc ggcggtacga acgcccgctt cgccctggtc gagttcgacg gtcaggaccc
4921 gcgcctgatc gagccgacgg cctatagggg cgaggactac ggcacggccg aggacgccat
4981 cgaggagtat ctccgcaagg tcggtgtcaa gcatcctgac caggcggtgg tcgctgtggc
5041 cgggccgatc gaccacggtc aggtccacat gaccaatctg gactggcgga tctccgagga
5101 cggcctgcgc cgcgcaggcg gttttcggaa cgccaagctg atcaacgact tcaccgccca
5161 ggcgctggcc gcgccgcgcg ttggccctaa ggacctgcgc cagatcggcg aattgccgac
5221 gtcgggggag ggcgatctgg cgatcctggg tccaggcacc ggcttcggcg tcgcaggcct
5281 tgtccgtcgc catggccagg agatcccgct ggccaccgag ggtggtcacg tcgccttcgc
5341 gccggtcgat gacgtcgaga tcgaggtgct ccgcgccctg acccggcgcc tggacggcgg
5401 tcgggtgtcg gtcgagcgga tcctgtcggg tcccggcatg gaggacctcc atgtggatct
5461 ggcggccgct gaagggcgcg gtgtcgaggc gctgaccgcc aagcagatca ccgagcgggc
5521 cgtagagggc tgcgccgaca gcctggcgac ggtgaaccgt ttctgcgcca tcctgggctc
5581 aacggcgggc gacatcgctc tgaccttggg cgcacgcggc ggtgttttca tcgccggcgg

5641 catcgcacca cgcatcatcg acattctgga gaagagcccg ttccgcgagc gcttcgacag
5701 caaggggcgt ctgtccggct tcacccgttc gatcccgacg cacgtgatcc tgcatccgca
5761 caccgccctg atcggtgcgg cggtggcgct cacgccggag ggccgtgcgg cggtgtcgta
5821 gtcgttttct tggggctcgc catgacagcc ttgtcatggc taagtactcg ccgaagcgag
5881 cgaatcggac cggcgaggga cggaaattga gcgccaaagt cacgatccac gacgtggccc
5941 gcgagagcgg agtctcgatc aagaccgtct cgcgtgtcct gaatcgcgag cccaacgtca
6001 aggctgacac ccgtgatcgc gtgcaggccg cggtagctgc gctgcactat cgccccaata
6061 tctcggcccg tagcctggcg ggggcgaagg cctatctgat cggcgttttc ttcgacaacc
6121 ccagccccgg ctacgtcacc gatgtgcagc tcggcgccat cgcccgttgc cggcaggaag
6181 ggttccatct gatcgtcgag ccgatcgact cgaccgccga tgtcgaggat caggtcgcgc
6241 cgatgctgac gacgttgcgc atggacggcg tgatcctgac cccgccgctc agcgatcatc
6301 cggtcgttct ggcggcgctt gagcgggaag gggtggccta tgtgcgcatc gccccaggcg
6361 acgatttcga ccgtgcgccg tgggtcagca tggatgatcg gctggccgcc tacgagatga
6421 ccaagcatct ggtcgatctg ggccacaagg acatcgcctt tattgtaggg caccccgacc
6481 acggcgcttc gcaccggcgt catcaggggt tcctcgatgc aatgcgcgac agcggcctgc
6541 gtgttcgtga tgatcgtgtg gcgcagggct ggttttcgtt tcgctcgggc ttcgaggcgg
6601 ccgagaagct gctgggcggc gcggatcgac cgacggcgat cttcgcctcg aacgatgaca
6661 tggcgctggg cgtcatggcg gtcgccaatc gcttgcggct tgacgttcct actcaactgt
6721 cagtcgccgg cttcgacgac acgccgggag cgaagataac ctggcctcag ctcaccacgg
6781 ttcgccaacc gatccacgcc atggccggag cggccgccga catgctgatg caaggcgtcg
6841 agcgggaaga gggcgcgccc ccgccatcgc ggctcctgga cttcgaactc gtcgtgcggg
6901 agtccaccgg cccagcgtcg cactgacgag tgcgcgattg gcaaggtggt accgcaatca
6961 tattcctctt gacagcgctg tccgatcgac gtgagatcgc atagatcaag acagtcgcca
7021 cgaaaaggcg gcgtcaggag gaaacgccat gacctcgcct ggccagtcca gcggctgtcg
7081 cgccctgtgc gcacgatcac tcccctccaa gcggacctga agccgcttat cttcgctcgg
7141 cgcggacacc ccgcaagcgc cgcgccgagc gcctttcctt taaacaccgc tccgcccgac
7201 ggacgccccc tcaaggatcg cccatgctgc cgcgccgttt cgccttcgct tccgccctgg
7261 ccctgacgat cgcctgtgga tcggccggcg tggtcctggc ccagacgccg ccgaacgcca
7321 ctgcaaaccc cgccgtttgg ccgatgtcgg ctagtccagc cgccatcacc gacgccaaga
7381 ccgaggcctt catcgcccag ctgatgagcc ggatgaccgt cgaggagaag gtcgcccaga
7441 ccatccaggc cgatggcgcc tcgatcacgc ccgaggaact gaagaagtac cggttgggat
7501 cggtgctggt cggcggcaac tcagcgccgg acggcaatga ccgcgccagc ccgcagcgct
7561 ggatcgaatg gatccgcgcc ttccgcgcgg ccgcgctgga caagcgcggc gaccggcagg
7621 aaatcccgat catcttcggc gtcgacgccg tgcatggtca caacaacgtc gtgggcgcca
7681 cgatcttccc gcacaatgtc ggcctgggcg cagcgcacga gcccgacctg atccgtcgta
7741 tcggcgaggt gaccgctaag gaaatggccg ccaccggggc ggactggacc tttggtccga
7801 cggtcgccgt gcctcgcgat tcacgctggg gccgcgccta tgagggctat ggcgagaatc
7861 cggagatcgt gaaggcctat tcgggcccga tgaccctggg gctgcagggg gcgctggaag
7921 ccggcaagcc gctggcggcc ggccgcgtgg cgggctcggc caagcacttc ctcgccgatg
7981 gtggcaccga gaatggccgc gaccagggcg acgcgaagat ctccgaggcc gatctggtgc
8041 gtctgcacaa cgccggctac ccgccggcga ttgaagccgg catcctgtcg gtgatggtct
8101 cgttctccag ctggaacggg gtcaagcaca ccggcaacaa aagcctgctg accgacgtgc
8161 tgaaggagcg catgggcttt gagggcttcg tcgtcggcga ctggaacgcc cacggccagg
8221 tcgagggctg cagcaacacc agctgcgccc aggcttataa cgccggcatg gacatgatga
8281 tggctcccga cagctggaag ggcctytacg acaacacctt ggcgcaggtg aaggccgggc
8341 agatccccat ggcgcggatc gacgatgccg ttcgccgcat cctgcgagtc aaggtcaagg
8401 ccggcctgtt cgaggacaag cggcctttgg agggcaagct ggagctcctc ggcgcgcctg
8461 agcaccgggc cgtggcgcgc gaggcggtgc gcaaatcgct ggtgctgctg aagaacgaag
8521 gcgtgctgcc gctgaagagc tcggctcgtg tgctggtcgc cggagacggc gccgacgaca
8581 ttggcaaggc ctcgggcggt tggaccctga cctggcaggg caccggcaac aagaacagcg
8641 acttcccgca cggccagtcg atctatgcag gcgtcgcgga ggccgtgaaa gccggcggcg
8701 gcagcgcgga actgtcggtt tcgggcgatt tcaagcagaa gcccgacgtg gcgatcgttg
8761 tgttcggcga gaacccctac gccgagttcc agggcgacat caccagcatc gagtatcagg
8821 ctggcgacaa gcgtgacctg gcgctgctga agaagctcaa ggctgcgggc attccggtgg
8881 tgtcggtgtt cctgagcggc cggcccctgt ggaccaaccc cgaactcaac gcgtccgacg
8941 cctttgtcgc ggcgtggctg cccggctcgg agggcggcgg cgtggccgac gttctggtcg
9001 gcgacaaggc gggtaagccg cgccacgact tccagggcaa g

```
//

LOCUS       gcc2537      1177 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc2537.
ACCESSION   gcc2537
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 1177)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1177)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..1177
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(1..1177)
                     /gene="lpsj"
     CDS             complement(441..998)
                     /gene="lpsj"
                     /codon_start=1
                     /product="putative galactosyl-1-phosphate transferase"
                     /translation="TGDAPCGQHRDDGADQDPQIEQKRAPAQIFGVKGDLVGDRQLVS
                     PIDLRPPRHAGTQGVNAGCTARGDQVILIEQSRPGSDQAHVTDEHAPELGQLIETELA
                     HQAADRRQPLIRIVEKVGGHLGRIDAHGAKLRHRKQRRGAPHALRPVETRPRRSQPHK
                     RPQHGQRQDQDRQPDRCSQHIKHTLH"
BASE COUNT      169 a    387 c    416 g    205 t
ORIGIN
        1 gccgaagccg acggcaattg gagttggcgt ggtctcctga cgggcatcga gagcagctcg
       61 ccgggcgatg cggggctggg tgagaagctc gccgcagaga tcaccgccat gggcatcgac
      121 ccccacgctc tgttgccgcg cggacggatc gacgagatcg ccgccgcgct ccagacgcgt
      181 gacggagagg gcgcgcgaga ggtggtcaag cgcctggcgc cggccgcgat ccgtcgcttg
      241 gtccgtcgcc tgttctccga cgcgacgctg aagggcaaca ccgagcgctt cctcaagcgc
      301 tatgcgggca tgatcgacga ggcggccggc caggatcgcg aaggcttcct gctcgccgcc
      361 ctgctgtcct ccgacgctgg gcgggcctat ctgctgctcg acgcggcgag cggcgatctg
      421 gcctaggccg agccggctga atgaagcgta tgtttgatgt gctggctgca gcgatcgggc
      481 tggcggtcct ggtcctgccg ctggccgtgc tgtgggcgct tgtgcggctg acttcgccgg
      541 ggcctggtct ctactggtcg cagcgcgtgg ggcgcgcctc ggcgctgttt ccgatgccga
      601 agtttcgcac catgcgcatc gatacgcccg aggtggccac ccaccttctc gacaatcctg
      661 atcaatggct gacgccgatc ggcggcctga tgcgcaagct cagtctcgat gagttgcccc
      721 agctctggag cgtgctcgtc ggtcacatga gcctggtcgg accccggccg gctctgttca
      781 atcaggatga cttgatcgcc gcgcgccgtg cagccggcgt tgacgccttg cgtcccggcg
      841 tgacggggtg ggcgcagatc aatgggcgag acgagttgtc gatcgccgac aaggtcgccc
      901 ttgacgccga atatctgcgc cggcgctcgc ttctgttcga tctgcgggtc ctggtcagca
      961 ccgtcatccc ggtgctgacc gcacggggcg tcacccgtta gccgagatcg gcgtaggcct
     1021 tttccagcgc cgaaacgaag cggggatcgc cggccagcga cgccgggaac accgagtcca
     1081 gggccaggaa ggcggcgacg tctgcgctgg cctctcccgt ggccgccagg cccagagcca
     1141 ggaggcgatc ttgcagcggg tcgacgatgg cttcacc
```

```
//

LOCUS       gcc1444      2031 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc1444.
ACCESSION   gcc1444
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 2031)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2031)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..2031
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            3..569
                     /gene="orf15"
     CDS             3..569
                     /gene="orf15"
                     /codon_start=1
                     /product="putative molybdenum cofactor biosynthesis
                     protein"
                     /translation="GEAIRLSPQGDDAQAIASAVSPAPVDVIVTIGGASVGDHDLVKP
                     ALRTLGLALSVETVAVRPGKPTWSGRLPDGRRVVGLPGNPASALVCAELFLRPLLAAL
                     TGAAPDIRLIPAGLAAPLPAGGPREHWMRAALSTDPDGRVVATPFPDQDSSLVSVFAR
                     ADALLRRRPGAPPAATGEVVDVLPLRRG"
     gene            658..2031
                     /gene="lpsK"
     CDS             658..2031
                     /gene="lpsK"
                     /codon_start=1
                     /product="putative nucleotide sugar epimerase/dehydratase
                     protein"
                     /translation="MGHAGKIATHVLLAFVALLAGRYLVIDMPFTRDTLLQATLYGLA
                     AFIVELAFRVERAPWRFVSATDHLRLLRSAVLTAAAFLVITRLTHPGIDGGLRTVAGA
                     ALIQAALLSALRVIRRSLHERMLLDSVLRLGPASMHPALPRLLIIGSASEAEAFLRAP
                     AGLGERYAPIGVVSPLDRETGDELRGVCVLGSIADFDSVLARLRDSGLSPAAILFLTD
                     SAMSTFGAERLGRLKTEGVRLLRRHGVVEMGAAANTPQLREISIEELLSRPPVRLDPE
                     PVRALVSGRRVLVTGAGGSIGSELCRQIAASGCAHLTMVDASEYNLFHIEREIAERHP
                     LLSRREALCDVRDAARVQRVFTEMKPDIIFHAAALKHVTLVENHPCEGVRTNVLGTRN
                     VAVAAKACGAAHLALISTDKAVAPTSVMGAAKRVAEAVARQYGGGGDMRVSIVRFGNV
                     LGSAGSVV"
BASE COUNT      255 a    722 c    703 g    350 t      1 others
ORIGIN
        1 gcggtgaggc gatccgactt tccccgcagg gcgacgacgc ccaggcgatc gccagcgccg
       61 tttcgcccgc gcccgtcgac gtcatcgtca cgatcggcgg cgcctcggtc ggcgaccatg
      121 acctggtcaa acccgcactc cgaacgctgg gccttgcgct ttcggtcgag acggtcgccg
```

```
      181 tgcgccccgg caagccgacc tggagcgggc ggttgccgga cggtcgccgc gtggtgggtc
      241 tgccaggaaa cccggcctcg gcgctggtgt gcgcggaact cttcctgcgg cctctgctgg
      301 cggctctcac gggcgcggcg ccggatatcc gcctcattcc cgcgggcttg gccgccccgc
      361 ttccggcggg cggaccgcgg gagcattgga tgcgcgccgc gctgtcgacg gatccggacg
      421 ggcgagtcgt cgcgacaccc ttccccgatc aggattcctc tctggtcagc gtgttcgcgc
      481 gcgccgatgc tctgctacgg cgacggcctg gcgcgccccc tgcggcgacg ggcgaggttg
      541 tcgatgttct gccgctccgg cgcggctgaa accgcgacgg catagaattg acgtgctaag
      601 cccggatttg agttcgccgg gcgtgacccg accttcaccg cttcagaggt tcgtttcatg
      661 gggcatgcag gaaagatcgc gacccacgtt ctgctggcct tcgtggccct gctggccggt
      721 cgctatctcg tcatcgacat gccgttcacg cgggacacgc tgcttcaggc gaccctgtac
      781 ggcctcgcag cattcatcgt ggagttggct ttccgggtgg agcgggcccc gtggcgcttc
      841 gtctcggcca ccgaccacct gcgacttctc cgctcggccg tcctgacggc ggcggcgttc
      901 ctggtcatta cccgcctgac ccatccaggc atcgacggtg gcctgcgcac cgtggccggc
      961 gcggccctga tccaggcggc gctgctgtcg gcgctgcggg tgatccggcg gagcctgcat
     1021 gagcgaatgc tgctcgattc ggtgctgcgc cttggscccg cctcgatgca tccggcgctg
     1081 ccgcgcctgc tgatcatcgg ctcggcctcc gaggccgaag ccttcctgcg cgcgccggcc
     1141 gggcttggcg aacgttacgc cccgatcggc gtggtctcgc cgctcgaccg cgagaccggc
     1201 gatgaactgc gcggcgtctg cgttctgggc tcgatcgccg atttcgacag cgtgctggcc
     1261 cgtctgcgcg acagcggcct gtcgccggcc gcgatcctgt tcctcaccga cagcgcgatg
     1321 agcaccttcg gcgccgagcg tctgggccgc ttgaagacgg aaggcgtgcg cctgctgcgc
     1381 cgccacggcg tggtcgagat gggcgcggcg gccaacaccc cccagctgcg cgagatcagc
     1441 atcgaggaac tcttgagccg gccgcctgtc cgactggatc cagagccggt tcgcgcgctg
     1501 gtgtccggtc gacgggtgct ggtgacaggc gcgggggggca gcatcggttc cgagctctgc
     1561 cgtcagatcg ccgccagcgg ctgcgcccat ctgaccatgg tcgacgcctc cgaatacaac
     1621 ctgttccaca tcgagcgcga gatcgccgag cggcacccgc tcctctcgcg tcgtgaggcg
     1681 ctctgcgacg tccgcgacgc cgcccgcgtc cagcgtgtct tcacggagat gaagccggac
     1741 atcatcttcc acgctgcggc gctgaagcat gtcacgctgg tggagaacca cccctgcgag
     1801 ggcgtccgca ccaatgtgct gggcacccgc aacgtggccg tcgccgccaa ggcctgcggc
     1861 gcggcgcatc tcgccttgat ctcgacggac aaggccgtcg cgccgaccag cgtgatgggc
     1921 gcggccaagc gtgtcgccga ggccgtggcg cgtcagtacg gcggcggcgg cgacatgcgg
     1981 gtcagcatcg tgcgctttgg caatgtgctg ggctcggccg gatcggtcgt a
//

LOCUS       gcc2218      2142 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc2218.
ACCESSION   gcc2218
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 2142)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2142)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..2142
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(3..719)
```

```
                     /gene="orf14"
     CDS             complement(3..719)
                     /gene="orf14"
                     /codon_start=1
                     /product="unknown exsG-like"
                     /translation="SCGQAHAFGERRAQREDQARGREVEHRLAAEVPRQALMHHLGAE
                     PVARGRPRQGGPALFAPDQGQKPRSGRLVDVPFDRDPALGGREGAMARGVGDQLVDGH
                     VHRHRRLGAEGDGRALDLEPSRNVVGEGRQGALQNLLQLRTSPGAAGQQLVRLRERQD
                     PALEDVCKGLGRRGRAQGLAGDRLHDRQGVLHAVIQLAQXEVTVLERGGEVVIETPAL
                     QGRGRGARDHLQLAQHLGRRI"
     gene            1134..2138
                     /gene="lpsL"
     CDS             1134..2138
                     /gene="lpsL"
                     /codon_start=1
                     /product="putative glycosyltransferase"
                     /translation="EPEFRRIDGEGGVVPARHGLARLGPVHGDRQLLAELHGVDGERL
                     VGLGFDRIEPGLAVGGQDVADVVAAEVADPRLHLHQPVQRHHLAALRREQGVVQARAE
                     LAGHQTRPREETRLLVPAQGLGAHQGEAGLLGQALDLGLVFELAEGLGRAHAAPADEQ
                     HVGVTALGEAQGLALKVPAQEVVEHIVAVQHVRQQGRADLDGFVARVVGDDLLVGQHH
                     PHDARHIAGGGPGQQQRRGDPRRVQGADHEQLRVHLGQPVTPDEVEVGDHEGPVEPVR
                     HRRVEIPGLARNVVLVPVLDVIGIGALGVDELVVQLQVRAGAFGHHAFGHQQVDIGRV
                     E"
BASE COUNT      324 a    723 c    757 g    337 t      1 others
ORIGIN
        1 atgatccgcc gcccaaggtg ctgcgcaagc tgcaggtggt cgcgcgcgcc acggcctcgg
       61 ccctggagag cgggcgtctc aatgactacc tcgccgcctc gctcgagcac ggtcacttcm
      121 tgctgcgcga gctggatcac cgcgtgaaga acaccctggc gatcgtgcag tcgatcaccc
      181 gccagaccct gcgcacgacc ccgtcgcccg aggcctttgc agacgtcttc gagagccgga
      241 tcctggcgct ctcgcaggcg cacgagttgc tgacccgccg cgcctgggga cgtccgcaac
      301 tggaggagat tctgcaacgc accctggcga ccttcgccga cgacgttgcg ggacggttcg
      361 aggtccaggg cccggccgtc gccttcagcg ccgagacggc ggtgtcggtg cacatggcca
      421 tccacgagct gatcgccaac gccgcgcgcc atggcgccct ctcgaccccc caaggccggg
      481 tcgcgatcga atggaacgtc gacaaggcga ccgctccggg gcttctgacc ctgatctggc
      541 gcgaacaggg cgggcccgcc ctgtcggggc cgcccacgcg caacgggttc ggctccaaga
      601 tggtgcatca gggcctggcg cgggacctcg gcggccaggc ggtgctcgac ttcgcgccca
      661 cgggcctggt cttcacgctg cgcgcgccgc tctccgaacg catgagcctg gccgcatgaa
      721 gaccgcccgg gtcatgatcg tcgaggatga ggccctggtg gccatgatgg tcgaggacat
      781 gctcggcgac atggggtgtg aggtggccgg ctcgttcggc gccgtcgacg ccgccctggc
      841 ctggctgcgc gatcatccct cgcccgacgg cgcggtgctg gacgtcaata tcggcggcga
      901 gatggtgttt ccggtcgccg aacgcctgcg cgagcagggc gtgccgttcg tcttcgccac
      961 cggctatggc gacctgccgc gcgcgggctt cgagtcggtg caggtgctgg ccaagccgat
     1021 caacgccggc gcgctgcgcc tggccgtcga gcgcttccgg atcggctaga acccttggc
     1081 ctgggatggg atcatccaaa gcgttcaaaa gggcgctgga tgtctgatct ttagagcctg
     1141 agttccggcg tatcgacggg gaaggggcgg tcgtcccagc ccgccatggc cttgcgcgcc
     1201 tgggcccggt ccacggcgat cgccagcttc tggcggaact ccacggggtc gatggtgaac
     1261 ggctggttgg cctggggttc gatcgaatag agccgggcct ggccgtcggc gggcaggatg
     1321 tagccgacgt cgtggcggcg gaagttgccg atccgcgact gcacctccac cagccagttc
     1381 agcgacacca cctggcagcc ctgcggcgag agcagggcgt tgtgcaggcc cgagccgaac
     1441 tcgccggcca ccagacgcgc ccgcgagaag agacgcgcct gctggtccca gcccagggtc
     1501 tcggggcgca ccagggtgag gccggcctcc tgggccaggc cctcgatctc ggcctcgttt
     1561 tcgagctcgc ggaaggactg ggccgtgcgc acgccgcccc ggctgacgaa cagcatgtcg
     1621 gcgtcaccgc ccttggtgaa gctcaaggcc tcgcgctcaa ggttccagcg caggaagtcg
     1681 tggaacacat agttgcggtt cagcatgtgc ggcagcaggg ccgcgccgat ctcgacggct
     1741 tcgtcgcgcg ggtcgtaggt gacgatctcc tggtcgggca gcatcatccg cacgatgcgc
     1801 gccacatagc cgggggcggt ccggggcagc agcagcggcg cggtgatccc cgccgcgttc
     1861 agggcgcgga tcacgaacag cttagggtac atctcggtca gccagtgacc ccagacgaag
```

```
     1921 ttgaagtggg tgatcacgaa ggcccggtcg agccggtgcg tcaccggcgc gtcgaaatcc
     1981 caggcctggc gcgcaatgtc gtgctggtac cagtgctgga cgtaataggg ataggcgccc
     2041 tcggcgtaga cgagctggtc gtccagcttc aggtgcgcgc cggagccttc ggccaccacg
     2101 cgtttggcca ccagcaggtc gacatcgggc gcgtggagat cg
//


LOCUS       gcc648       2699 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc 648.
ACCESSION   gcc648
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 2699)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2699)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..2699
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            1..1056
                     /gene="orf1"
     CDS             1..1056
                     /gene="orf1"
                     /codon_start=1
                     /product="putative chemotaxis receptor protein"
                     /translation="RPVIAPGRTDDQDQVITVLSEQFKALAAGDLTARVDVVFSERYG
                     HVRDEFNAAMTKLGQVMDEISMAAGGLGESSDEVARVSQHLSRGAGRQALDLHGARAA
                     LQKVGAAAGRGVDGLRRVTEAAAGLRIDAASARRSVREAVGSIAEVEQSALRISQAAA
                     LFDEVAQQANVLSLIADVEGARGGEGXGPFQAVAADKMRVLAERASGAAREIKGVTAA
                     NSAQVSRCARLMDAASASFGGMASRITQIDGLVSGLAKSAQEQAHGLRAVDEAVDRAD
                     DIAQTHADQVDEAAAVTGRLIEEAESLIQAASPFRAHVVSRPASRPEPARAGHHAPAG
                     NAVARAHARIAAYARPR"
     gene            1060..2577
                     /gene="orf1"
     CDS             1060..2577
                     /gene="orf1"
                     /codon_start=1
                     /product="putative hippurate hydrolase protein"
                     /translation="MLCHPGKRVALVRDPGAASAALPQSLGPGSTPGFRRGSAGMTQD
                     ISVRGGGGGEHVRRSCDSRNPRPSMKSLFAASALALLIATAAQAGPLNVPATQKVISA
                     QLDRDYPALEALYKDIHAHPELGFQEVETAKKLAAQMRALGFTVTEGVGKTGVVAVLK
                     NGEGPKVLIRTELDGLPMQEKSGLAWASQATATWNGEKVFVAHACGHDIHMAAWVGAA
                     RQLVAMKAKWKGTLVFVAQPSEETVRGARAMLDDGLWDKIGGKPDYGFALHVGSGPXG
                     EVYYKAGVLTSTSDGLDITFNGRGGHGSMPSATIDPVLMAARFTVDVQSVISREKDPS
                     AFGVVTVGSIQAGSAGNIIPDKARVRGTIRTQDNAVREKILDGVRRTVKAVTDMAGAP
                     PADLKLTPGGKMVVNDAALTDRTAVVFKAAFGARAVAQDKPGSASEDYSEFVLAGVPS
                     VYFAIGGSDPAELAKAKAEGREPPVNHSPYFAPVAEPTIRTGVEAMTLAVLNVLK"
```

```
BASE COUNT      396 a    935 c    984 g    381 t      3 others
ORIGIN
        1 cgccccgtga tcgcgccggg ccgcaccgac gatcaggatc aggtgatcac cgtgctgtcc
       61 gagcagttca aggccctggc ggccggcgac ctgaccgccc gcgtcgatgt ggtgttcagc
      121 gagcgctatg gccacgtccg cgacgagttc aacgcggcga tgaccaagct gggccaggtc
      181 atggacgaga tctccatggc ggctggcggg ctgggcgagt cttcggacga ggtggcgcgc
      241 gtctcgcagc atctgtcgcg cggcgcgggg cgtcaggcct tggatctgca cggtgcgcgg
      301 gcggcgctgc agaaggtggg cgcggccgcc gggcggggcg tggacgggct gcgccgcgtc
      361 accgaagccg ccgccggcct gcgcatcgac gccgccagcg cccgccgttc ggtgcgtgag
      421 gcggtggggt cgatcgcgga ggtcgagcag agcgccttgc gcatcagcca ggccgccgcc
      481 ctgttcgacg aggtggctca gcaggccaat gtcctgtcct tgatcgccga cgtcgagggc
      541 gcgcggggcg grgagggcgm ggggcccttc caggccgtcg ccgctgacaa gatgcgcgtc
      601 ctggccgagc gggcctcggg cgcggcgcga gagatcaagg gcgtgacggc cgccaattcg
      661 gcgcaggtct cgcggtgcgc gcggctgatg gacgccgcct cggcctcgtt cggcggcatg
      721 gcgtccagga tcacccagat cgacggtctg gtgtcgggcc tggccaagtc cgcccaggag
      781 caggcccatg gcctgcgcgc cgtcgacgag gcggtggacc gggccgacga tatcgcccag
      841 acccatgccg accaggtcga cgaggccgcg gcggtcaccg gccgcttgat cgaggaggcc
      901 gagagcctga tccaggccgc cagtcctttc cgcgcccatg tggtttcgcg cccggcgtcg
      961 cggcccgaac cagcccgcgc cggccatcac gcgcccgccg gcaacgccgt ggcccgcgcc
     1021 cacgcccgca tcgccgccta tgcgcgaccc cgctaggggа tgctgtgtca tcccggaaag
     1081 cgtgtagcgc ttgtccggga cccaggggcg gcaagcgcgg cgctcccgca gtccctgggt
     1141 cccggctcta cccccggctt tcgccggggt tcggccggga tgacacagga tatttctgtc
     1201 aggggtggcg gaggtggcga gcatgtgcgt cgatcctgcg actcacgaaa cccgagaccc
     1261 tccatgaagt ccctgttcgc cgcctcggct ctcgccctgc tgatcgccac cgccgcccag
     1321 gccgggccgt tgaacgtgcc cgccacgcag aaggtgatca gcgcccagct cgaccgcgac
     1381 tatccggcgc tggaggcgct gtacaaggac atccacgccc accccgagct cggcttccag
     1441 gaggtcgaga ccgccaagaa gctggccgcg cagatgcggg cgctgggctt caccgtcacc
     1501 gagggcgtcg gcaagaccgg cgtggtggcg gtgctgaaga acggcgaggg ccccaaggtg
     1561 ctgatccgca ccgagctgga cggcctgccg atgcaggaaa agtcgggcct ggcctgggcc
     1621 agtcaggcga ccgccacctg gaacggcgag aaggtcttcg tcgcccatgc ctgcggccac
     1681 gacatccaca tggccgcctg ggtgggtgcg gcccgccagc tggtggcgat gaaggccaaa
     1741 tggaagggca cgctggtttt cgtggcccag ccctcggagg agacggttcg cggggcccgc
     1801 gccatgctgg acgacggtct gtgggacaag atcggcggca agcccgacta cggctttgcg
     1861 ctgcacgtcg gttcgggtcc gkccggcgag gtctattaca aggccggcgt cctgacctcg
     1921 acctcggatg gcctggacat caccttcaac ggccggggcg ggcacggctc gatgccctcg
     1981 gccaccatcg acccggtgct gatggccgcc cgcttcaccg tcgacgtgca gagcgtgatc
     2041 agccgcgaga aggacccgtc ggccttcggc gtggtgacgg tcggctcgat ccaggcgggc
     2101 agcgccggta acatcatccc cgacaaggcc cgggtgcgcg gcacgatccg cacccaggac
     2161 aacgccgtgc gcgagaagat cctcgacggc gtgcgccgca cggtgaaggc ggtgaccgac
     2221 atggccggcg ccccgcccgc cgacctgaaa ctgaccccgg gcggcaagat ggtggtcaat
     2281 gatgcggccc tgaccgatcg cacggcggtg gtgttcaagg ccgccttcgg ggcccgcgcc
     2341 gtggcgcagg acaagccggg ctcggcgtcc gaggactatt cggaattcgt gctggccggc
     2401 gtgccgtcgg tctacttcgc catcggtggc tcggaccccg ccgagctcgc caaggccaag
     2461 gccgaaggcc gtgagccgcc ggtcaaccac tcgccgtact tcgcgcccgt ggccgagccg
     2521 acgatccgca cgggggtgga ggcgatgacc ctggcggtgc tgaatgtgtt gaagtgaccc
     2581 ttctcccctt gcgggagaag gtgtcgccgg aggcgacgga tgaggggttt ctcggcctcg
     2641 ccgcgcgacc cctcaaccga cccgctacgc gggccaccct ctcccgcaaa gggagaagg
//

LOCUS       gcc1290      2109 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc1290.
ACCESSION   gcc1290
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
```

```
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 2109)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2109)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..2109
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(822..1769)
                     /gene="orf12"
     CDS             complement(822..1769)
                     /gene="orf12"
                     /codon_start=1
                     /product="Unknown interrupts O-antigen synthesis"
                     /translation="MSRLPPGLKTGRDVSVTGVDAAGRTVLTARDGDPQMVWTPTREE
                     RKALRAAKAVRIDVKLEAVEGKLVGPALYADWGDGFSEDSYARLKAGPDGWFASLPAR
                     SFQLNGVRLDPSEGACAFTVEALTVTRIGDLGRDPRGLRGAAIQALKPMLGPLRGPAG
                     AAWRRGRALLAKGRVARPAGRDEGAVGATYAHAIAVSRNLRSPHYAAPIAAPITLPAE
                     APKVVAFYLPQFHPFPENDTWWGKGFTEWTNVSKAQPQFLGHYQPRLPADLGFYDLVS
                     ARCWPSRWTWPRARASTPSASTTTGSPESAFWNGRWICS"
     gene            complement(1766..2107)
                     /gene="orf13"
     CDS             complement(1766..2107)
                     /gene="orf13"
                     /codon_start=1
                     /product="polysaccharide transporter kpsT-like"
                     /translation="PAELSGLGDFLHLPVRTYSAGMLARLMFTVATVFEADILVLDEW
                     LSAGDAAFVQKAAQRMHRMVEDAKIVVMATHDHDLVQRVCNRVCELQGGKIXFLGSXE
                     DWLAYRETQAA"
BASE COUNT      310 a    719 c    767 g    311 t      2 others
ORIGIN
        1 gtcggcgcgc ttggcgaact ggctctgtgc ctcggcgatg atcggatggg cgttggtcag
       61 gcgcggcagc caggcgctga gcgcggtgcg ggtggcgtgc agatagccgt ggccgaacca
      121 ccggtcgggc tcgagatagg cgccttcgcc ccactcgttc caggcgttga cgaacaccag
      181 cgcctcgccc ttgggatgcc gggcctcggc gtgcttcagc gcccccgaca gccagccgaa
      241 atagctttcc ggatcggcgt tgtggaacgc caccccggcc cagggcttgc gggcctggtt
      301 gtcccagccg ggcatgacgc cgggcacgaa ggcggccgga acctgctcca gctcatccag
      361 cttgtggcgg gccacggcgg gatagtcata gaccttgccg gtgaagccgg cgtgcagcgg
      421 cgtgacccgg ttggtgatct cgccctcgac aatggcgtgc ggcgggaagt cgacgatgcc
      481 gtcgaagccg tggccggcat agtcctggaa gccgaaggcg gtggtgcaca gcaggtgcag
      541 ctcgcccagc cccatggccc gcgcctgatc gcgccagcgc tcagtggtgg ccttggcgtc
      601 gggcaggatc tcggggcggt acagcaatag gagcggcttg ccgctcaccc gcagatagcg
      661 cgggtcgcgc atgtagcgcg ccaggtcctc gaacaccgcg cggtcgtcct gcggcgagtg
      721 ctcctggccc atcaggatgt cgctctcgtc gccgtcccag cggcgggtcc agttctcatt
      781 ggcccagcag agggcgaagg gcaggtccag gctcggatcg ttcaggaaca gatccagcgg
      841 ccgttccaga aggcgctttc cggcgaacca gtagtagtgg aagcagaagg cgtggacgcc
      901 cgcgcccttg gccaggtcca cctgctgggc cagcacctcg cgctgaccag gtcgtagaag
      961 cccagatccg ccggcaggcg cggctggtag tgacccagga actgcggctg ggccttggag
     1021 acgttggtcc actcggtgaa gcccttgccc caccaggtgt cattctccgg gaacggatga
     1081 aactgcggca ggtagaaggc caccaccttg ggcgcttcgg ccggcagggt gatcggggcg
```

```
     1141 gcgatcgggg cggcgtagtg cgggctgcgc aggttgcggg agaccgcgat cgcgtgggcg
     1201 taggtcgcgc cgaccgcgcc ctcgtcgcgt cctgccggac gcgccacgcg tcccttggcc
     1261 agcagcgccc gcccgcgtcg ccaggccgcg ccggcggggc cgcgcaaggg gcccagcatc
     1321 ggcttcagcg cctggatcgc cgcgccgcgc aggccgcgcg gatcgcggcc gagatccccg
     1381 atacgggtga ccgtgagggc ctcgaccgtg aaggcgcagg cgccctcgga cgggtccagc
     1441 cgcacgccgt tcagttggaa actgcgcgcc ggcagcgagg cgaaccagcc gtccggaccg
     1501 gccttcaggc gggcgtagga atcctcggaa aagccgtcgc cccagtcggc gtagagcgcg
     1561 gggccgacca gcttgccctc gaccgcctca agcttgacgt cgatccgcac cgccttggcc
     1621 gcccgcagcg ccttgcgctc ttcgcgggtg ggcgtccaga ccatctgcgg gtcgccgtcg
     1681 cgggcggtca ggaccgtgcg cccggccgca tcgacgccgg tgaccgacac gtcgcggccg
     1741 gtcttcaggc cgggcggcag gcggctcatg cggcctgggt ttcgcggtag gccagccagt
     1801 cctcgktcga gccgaggaag gsgatctttc cgccctgcag ttcgcagacg cggttgcaga
     1861 cccgctggac caggtcatgg tcgtgggtgg ccatcaccac gatcttggcg tcctcgacca
     1921 tccggtgcat ccgctgggcg gccttctgca cgaaggcggc gtcgccggcg ctgagccact
     1981 cgtccagcac caggatgtcg gcctcgaaca cggtggccac cgtgaacatc aggcgcgcca
     2041 gcataccggc cgaataggtg cgcaccggca ggtgcagaaa gtcgcccagg cccgataact
     2101 cggcgggggg
//

LOCUS       gcc 2205     2365 bp    mRNA            BCT       15-OCT-1999
DEFINITION  gcc 2205.
ACCESSION   gcc 2205
VERSION
KEYWORDS    .
SOURCE      Caulobacter crescentus.
  ORGANISM  Caulobacter crescentus
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 2365)
  AUTHORS   Awram,P.A.
  TITLE     Analysis of the S-layer Transporter Mechanism and Smooth
            Lipopolysaccharide Synthesis in Caulobacter crescentus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2365)
  AUTHORS   Awram,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) UBC
FEATURES             Location/Qualifiers
     source          1..2365
                     /organism="Caulobacter crescentus"
                     /strain="NA1000"
     gene            complement(2..550)
                     /gene="orf16"
     CDS             complement(2..550)
                     /gene="orf16"
                     /codon_start=1
                     /product="putative HOMODA hydrolase protein"
                     /translation="MRGLTISGVFAVLVLTASLAQAGEVTVDGRKVAYREWGGGERTL
                     VMVSGLGDGAETFETVGPRLAQGWRVIAYDRAGYGGSADDPRVHDAERAEAELKGLLA
                     ALKVRKPVLLGHSLGGVFAAHFAARNPGEVTGLVLEETRPTGFTAACKAKRMRGCAFP
                     PLLKYAFPPGGRREVETLDRIER"
     gene            complement(559..2178)
                     /gene="pgi"
     CDS             complement(559..2178)
                     /gene="pgi"
                     /codon_start=1
```

/product="putative phosphoglucoisomerase"
/translation="MADLDAAWTRLEAAAKAAGDKRIVEFFDAEPGRLDALTLDVAGL
HLDLSKQAWDEAGLEAALDLAHAADVEGARARMFDGEAINSSEGRAVLHTXLRAPAGA
DVKALGQPVMAEVDAVRQRMKAFAQXVRSGAIKGATGKPFKAILHIGIGGSDLGPRLL
WDALRPVKPSIDLRFVANVDGAEFALTTADMDPEETLVMVVSKTFTTQETMANAGAAR
AWLVAALGEQGANQHLAAISTALDKTAAFGVPDDRVFGFWDWVGGRYSLWSSVSLSVA
VAAGWDAFQGFLDGGAAMDEHFRTAPLEQNAPVLVALAQIFNRNGLDRRARSVVPYSH
RLRRLAAFLQQLEMESNGKSVGPDGQPAKRGTATVVFGDEGTNVQHAYFQCMHQGTDI
TPMELIGVAKSDEGPAGMHEKLLSNLLAQAEAFMVGRTTDDVVAELTAKGVSDAEIAT
LAPQRTFAGNRPSTLVLLDRLTPQTFGALIALYEHKTFVEGVIWGINSFDQWGVELGK
VMANRILPELESGASGQHDPSTAGLIQRLKR"

BASE COUNT 367 a 849 c 783 g 364 t 2 others

ORIGIN

```
       1 gccgctcaat acggtccagc gtttcgacct cgcggcgccc gcccggcgga aacgcgtatt
      61 tgagcagcgg cgggaacgcg cacccgcgca tacgcttagc cttacaggcc gcagtgaagc
     121 cggtcggccg ggtttcctcc agcacgaggc ccgtgacctc tcccggattg cgggccgcga
     181 agtgggcggc gaacacgccg cccagcgaat gccccagaag cacgggcttt cgcaccttca
     241 acgccgccag cagccccttc agctcagcct ccgcccgctc ggcgtcgtgc acacgcggat
     301 catcggcgct gccgccatag cccgcccggt cataggcgat gacgcgccag ccctgggcca
     361 gccgggggcc gaccgtctcg aacgtctcgg ccccgtcgcc aagaccgctg accatcacca
     421 gggtccgctc gccaccgccc cattcgcgat aagcgacctt gcgcccgtcg accgtcacct
     481 cgccggcctg cgccaacgag gccgtcagca ccaggacggc gaaaacgccg ctgatcgtca
     541 gccctcgcat gggatcgcct agcgcttcag gcgctggatc aaccctgcgg tcgaagggtc
     601 atgctggccc gaagcgccgc tctccagctc cggcaggatg cggttcgcca tcaccttgcc
     661 cagctcgacg ccccactggt cgaagctgtt gatcccccag atcacgccct cgacgaaggt
     721 cttgtgctca tagagggcga tcagggcgcc gaaggtctgg ggcgtcaggc ggtcgaggag
     781 caccagggtc gagggccggt tgccggcgaa agttcgctgc ggggccaggg tggcgatttc
     841 ggcgtcagag acgcccttgg ccgtgagctc ggccacgaca tcgtccgtgg tccgcccgac
     901 catgaaggcc tcggcctggg ccaagaggtt cgagagcagc ttctcgtgca tgccggccgg
     961 gccttcgtcc gacttggcga cgccgatcag ctccatcggc gtgatgtcgg tcccctggtg
    1021 catgcactgg aaataggcgt gctgaacatt ggtgccttcg tcgccgaaca ccaccgtggc
    1081 cgtgccgcgc ttggccggct gcccgtcggg gccgaccgac ttgccgttgc tctccatctc
    1141 cagctgctgg aggaaggcgg ccaggcggcg caggcggtgc gagtacggca cgaccgagcg
    1201 ggcccggcgg tccaggccgt tgcgattgaa gatctgggcc agggccacca gcaccggcgc
    1261 attctgctcc agcggggcgg tgcggaagtg ctcatccatg gccgcgccgc cgtccaggaa
    1321 accctggaac gcgtcccagc ccgcggcgac ggccaccgaa aggctgaccg acgaccacag
    1381 cgaatagcgg ccgccgaccc agtcccagaa cccgaacacg cgatcgtccg gcacgccgaa
    1441 ggcggcggtc ttgtccagcg cggtcgagat ggcggccaga tgctgattgg ccccctgctc
    1501 gcctagggcc gccaccagcc aggcccgcgc cgcgccggcg ttggccatgg tctcctgggt
    1561 cgtgaaggtc ttggagacca ccatgaccag ggtctcttcc gggtccatgt cggcggtggt
    1621 cagggcgaac tcggcgccgt cgacattggc gacgaagcgc aggtcgatcg acggcttcac
    1681 cggtcgcagg gcgtcccaca gcaggcgtgg gcccaggtcg ctgccgccga tgccgatgtg
    1741 caggatcgcc ttgaacggct tgccggtcgc gcccttgatc gcccccgaac gcacggmctg
    1801 cgcgaaagcc ttcatccgct ggcggacggc atcgacctcg gccatgaccg gctggcccag
    1861 ggccttgacg tccgctcccg ccggagcgcg caggnccgta tgcagcacag cccggccttc
    1921 ggacgaattg atcgcctcgc cgtcgaacat gcgggcccgg gcgccctcga catcggccgc
    1981 gtgggccaga tcgagcgcgg cctcgagacc cgcctcatcc caggcctgct tggagagatc
    2041 caggtgcagg ccggcgacgt ccagggtcag agcgtcaagg cgtcccggct cggcgtcgaa
    2101 gaactcgacg atacgcttgt cgcccgcagc cttggcggcg gcttccaggc gggtccaggc
    2161 ggcgtcgaga tcggccatgt cgtcctcaca ggtttggtaa atcgctgttt acggacccgg
    2221 gcttatcaaa caggcgccgc cgcgtcatgg cagaccaatg acgtttttg cgggagcccc
    2281 cgatgtcctg tgacgccgtc gccttttccg ccatgctctg gatggcgtcg ttcaatccgg
    2341 agcagacgac cggccccgcc ctcgc
```

//