Comparison of S-layer Secretion Genes in Freshwater Caulobacters

by

Mihai Iuga

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
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
October 2000

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 & Immunology

The University of British Columbia
Vancouver, Canada

Date 11/6/00

Abstract

The Smit laboratory strain collection contains numerous freshwater caulobacter strains (FWC) isolated from locales throughout North America. 40 of these strains are morphologically similar to *Caulobacter crescentus* and secrete a surface (S-) layer. Their S-layer proteins have an estimated molecular weight (MW) ranging from 100 kDa to 193 kDa, and react positively to an antibody raised against the *C. crescentus* S-layer protein, RsaA. They can be divided in three size groups based on their MW: the small, medium and large S-layer groups. This thesis investigated whether all FWC strains, independent of the size of their S-layer, secreted the S-layer protein to the surface by a type I secretion system. The transporter genes composing the type I apparatus and the S-layer secretion signal were cloned, sequenced and compared for a subset of FWC strains, two from each S-layer size group. The protein cross-expression ability of the FWC strains and the arrangement of the S-layer secreting genes were also examined.

The transporter genes and the regions encoding the C-termini of S-layer proteins were isolated by screening libraries or by PCR amplification of genomic DNA from a subset of FWC strains. The sequence data showed that all FWC strains studied contained adjacent genes encoding ABC transporters and membrane fusion proteins (MFP) that were highly similar to those of *C. crescentus*. Comparison of the available S-layer C-termini showed low similarity for the small and medium size groups, with the presence of conserved residues and predicted secondary structure features, typical for type I secretion signals.

Protein cross-expression studies showed that FWC strains were capable of recognizing secretion signals from *C. crescentus* and other type I systems (*Pseudomonas aeruginosa* alkaline protease AprA), and that *C. crescentus* secreted the C-terminus of one FWC strain, FWC19. This suggested the presence of functional type I transporters in FWC strains, and that the S-layer subunits are secreted by a type I mechanism. The successful creation, in FWC strains from all size groups, of ABC transporter-inactivated mutants that no longer secrete an S-layer demonstrated that the identified type I transport system was the one that secreted the S-layer subunit.

The arrangement of S-layer secretion genes was also analyzed. The ABC transporter and MFP genes were adjacent in all strains examined, the same as in *C. crescentus*. The S-layer gene was adjacent to the transporter genes in the small S-layer group, located further upstream in the medium S-layer group, and elsewhere on the genome in the large S-layer group. In all FWC strains studied, the OMP gene had the same disposition as in *C. crescentus*, further downstream of the OMP gene.

Interestingly, the S-layer gene location is correlated with the sequence divergence of transporter proteins among size groups, dividing the S-layer secreting FWC strains into at least two groups. The small S-layer group has the same arrangement of S-layer secretion genes as *C. crescentus*, almost identical transporter genes and low similarity of secretion signals. The medium-large S-layer group has highly similar transporter genes and low similarity of secretion signals. The S-layer gene location further divides this group into subgroups: the medium S-layer group (S-layer gene located further upstream of the transporter genes) and the large S-layer group (S-layer gene located elsewhere in the genome).

Overall, these results show that the FWC strains studied secrete their S-layer by a type I secretion system proprietary to the S-layer protein.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

3-D	three-dimensional
aa	amino acid
ABC	ATP–Binding Cassette
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
Cm	chloramphenicol
C-terminus	carboxy terminus
DNA	deoxyribonucleic acid
EGTA	ethylene glycol-bis(β-aminoethyl Ether) N'N'N'N' tetraacetic acid
FWC	freshwater *Caulobacter*
GC	guanosine and cytosine content of DNA
h	hour(s)
h	hydrophobic amino acid
kDa	kilodalton
Km	kanamycin
LPS	lipopolysaccharide
min	minute
MFP	membrane fusion protein
mg	milligram
ml	millilitre
µl	microlitre
µg	microgram
NaCl	sodium chloride
N-terminus	amino terminus
oligo	oligonuclotide
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
SLPS	smooth lipopolysaccharide of *C. crescentus*
SDS	sodium dodecyl sulphate
Sm	streptomycin
Tris	Tris (hydroxymethyl) methylamine
x	any amino acid

List of Species Abbreviations

C. crescentus	*Caulobacter crescentus*
C. fetus	*Campylobacter fetus*
E. chrysanthemi	*Erwinia chrysanthemi*
E. coli	*Escherichia coli*
P. aeruginosa	*Pseudomonas aeruginosa*
S. marcescens	*Serratia marcescens*
R. meliloti	*Rhizobium meliloti*
R. leguminosarum	*Rhizobium leguminosarum*

Acknowledgments

I would like to thank my supervisor, Dr. John Smit, for guiding and supervising my research. I thank Dr. John Nomellini for priceless practical suggestions and contributions, and for being my friend. I thank my committee members Dr. Tom Beatty and Dr. Jim Kronstad for their useful suggestions and support. I thank Dr. Peter Awram for opening the path for my research, and for countless suggestions. I thank the present and past members of the Smit laboratory: Theo Blake, Ying Zhao, Wade Bingle, Kathyrin Clark, Irene Dorocicz, Ian Bodset for their help, suggestions and support. I would also like to thank Ali Tehrani, Yves Leduc and Steve Seredick for believing in me and being such great friends over the past two years. Most of all I would like to thank my wife, Daniela Ginta, for loving and understanding me.

Chapter 1
Introduction

This thesis focuses on the secretion of the S-layer in various freshwater caulobacter (FWC) species. Most of the FWC strains of the Smit lab collection were shown to secrete S-layer proteins with sizes ranging from ~100 kDa to ~190 kDa (Walker et al., 1992), but more research was needed to elucidate the secretion mechanism of S-layer subunits so different in size.

Evidence is presented here that the S-layer of various FWC species are secreted by type I transporter systems proprietary to the S-layer proteins. FWC strains can be divided into two groups based on transporter protein sequence conservation: small size (low MW) S-layers and medium-large size (medium-high MW) S-layers. The arrangement of the S-layer secretion genes further divides the FWC strains into three groups: small, medium and large S-layer-secreting FWC strains.

Caulobacters are non-pathogenic, Gram-negative eubacteria, ubiquitous in aquatic environments, including drinking water. *Caulobacter crescentus*, the best characterized species, derives its name from the crescent shape of the cells. It has a dimorphic life cycle, switching between a flagellated, motile phase (swarmer) and a stalked, sessile phase (for reviews see Brun et al., 1994; Shapiro, 1976; Poindexter, 1981)(Figure 1-1). Swarmer cells express at one pole a flagellum and pili. Differentiation into stalked cells involves loss of the flagellum and pili, and localized growth of the cell envelope at the same pole to form a stalk that has a holdfast on its tip. Stalked cells bud off swarmer cells with the flagellum located at

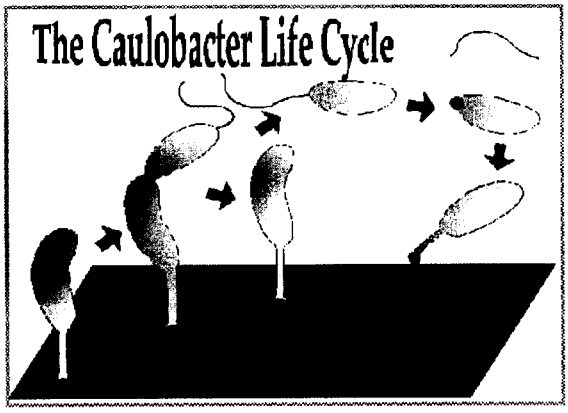

Figure 1-1. Life cycle of *C. crescentus*. Attached stalked cells bud off swarmer cells which move to a new location where they lose the flagellum and grow a stalk to attach to the surface again via the holdfast. (Figure courtesy of Ian Bodset)

the pole furthest from the stalked cell. In both phases the cells are completely enveloped by an S-layer (Smit et al., 1981).

S-layers are two-dimensional protein arrays that cover the surface of many prokaryotes, and are one of the most common surface structures in bacteria (for reviews see Sara and Sleytr, 2000; Sara and Sleytr, 1996b; Messner and Sleytr, 1992; Smit, 1986). They have been identified in over three hundred different species from all phylogenetic groups of bacteria, and are an almost universal feature in Archaea (Sara and Sleytr, 1996b). S-layers are made of thousands of copies of usually a single protein that self-assembles into a paracrystalline lattice onto the cell surface (Messner and Sleytr, 1992).

Since S-layers are ubiquitous and, when present, make up to 10% of the total cell protein and thus represent a large energy expenditure for the cell, they should have a vital function. However, some bacteria have been shown to lose their S-layers during sub-culturing in the laboratory, suggesting that S-layers are not essential for growth (Blaser et al, 1985; Luckevich and Beveridge, 1989). Hence, the function of the S-layers must be required for survival in the natural environment of the bacterium. The most common function attributed to S-layers is that of a protective barrier against parasites and lytic enzymes, due to the impermeability of the layer to molecules or particles larger than its pore-like structures (Koval and Hynes, 1991; Sleytr and Messner, 1983). Also, S-layers might play the role of virulence determinants in some pathogenic bacteria such as *Treponemes*, *Chlamydia*, *Clostridia*, *Rickettsias*, although a direct implication in the pathogenic process has not been shown in every case (Smit, 1986). The fish pathogen *Aeromonas salmonicida* uses its S-layer to adhere to and invade macrophages (Munn et al, 1982), whereas in *Campylobacter fetus* variation of the S-layer confers resistance to the host immune response by evading phagocytosis (Dworkin and Blaser, 1996). The S-layer of *C. crescentus* protects the cells from a *Bdellovibrio*-like organism (Koval and Hynes, 1991), but it also allows a parasite to infect *C. crescentus* by acting as a receptor for the bacteriophage øCR30 (Edwards and Smit, 1991).

S-layers are usually composed of a single protein or glycoprotein with molecular weights of 40 kDa to 200 kDa. Only in a few organisms, such as *Clostridium difficile* and

Bacillus anthracis, are the S-layers made of two different protein subunits (Sara and Sleytr, 2000). S-layers are rather similar in overall amino acid composition, with a high content of acidic and hydrophobic residues and the absence of cysteine residues (Sleytr and Messner, 1983). The S-layer subunits are held together by non-covalent forces including hydrophobic, ionic, hydrogen and polar bonds (Messner and Sleytr, 1992). Despite all the similarities mentioned above, there is very little sequence similarity among S-layer proteins (Messner and Sleytr, 1992). Generally, the N-terminal region is responsible for binding the S-layer to the underlying cell envelope layer and is usually more conserved than the C-terminal region, which comprises the domains involved in secretion and self-assembly (Sara and Sleytr, 2000).

The S-layer of *C. crescentus* is a hexagonal lattice that forms on the surface of the outer membrane, and is composed of the RsaA protein. Six RsaA subunits form a ring-like, circular structure that interconnects with other rings to form a two-dimensional hexagonal array (Smit et al., 1992). Three-dimensional reconstruction showed that the RsaA subunits are approximately rod-shaped, perpendicular to the plane of the array and have a linker arm that accounts for the connection between the rings (Smit et al., 1992)(Figure 1-2). The core region of the rings creates an open pore 3 nm in diameter, predicted to exclude molecules larger than 17 kDa (Smit et al., 1992). It takes approximately 40,000 RsaA subunits to cover the entire cell surface (Smit et al., 1981).

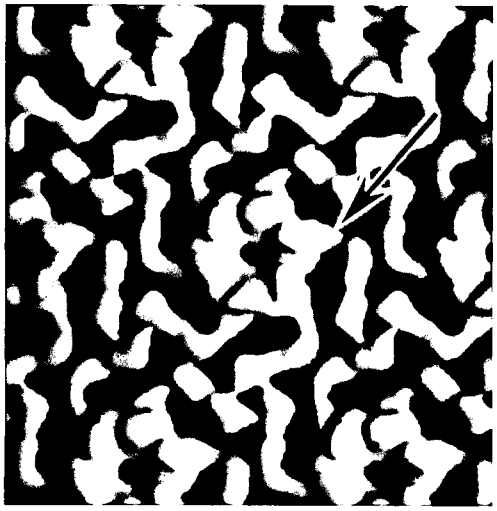

Figure 1-2. 3-D reconstruction of *C. crescentus* S-layer. The arrow indicates a single C-shaped RsaA subunit. (Figure from Smit *et al*, 1992).

The gene for RsaA has been cloned (Smit and Agabian, 1984) and sequenced (Gilchrist et al., 1992). RsaA is synthesized without a leader peptide, and aside from the cleavage of the initial N-formyl methionine, it is not post-translationally processed, leaving a mature polypeptide of 1025 amino acids with a molecular weight of 98 kDa (Fisher et al, 1988; Gilchrist et al, 1992). The RsaA mRNA

has a relatively long half-life and it is constantly transcribed during the cell cycle (Fisher et al., 1988).

RsaA is a secreted protein, so it must pass through both the inner and outer membranes to form the S-layer on the cell surface. Deletion and hybrid protein studies showed that secretion of RsaA relies on an uncleaved C-terminal secretion signal located within the last 82 amino acids (Bingle et al., 2000; Bingle et al., 1997b). Once secreted, the S-layer is attached to the cell surface via a specific smooth lipo-polysaccharide (SLPS) in the outer membrane (Walker et al, 1994). If the SLPS is absent, the S-layer is shed from the cell surface and aggregates into particles that have a two-dimensional array structure (Walker et al., 1994)(Figure 1-3). Ca^{2+} is essential for the proper crystallization of RsaA onto the surface, being perhaps involved in the attachment of the S-layer to the outer membrane and in subunit-subunit interactions, (Smit et al, 1992). If Ca^{2+} is removed with EGTA, the S-layer detaches from the surface and its structure is disrupted (Nomellini et al, 1997; Walker et al, 1994). Linker mutagenesis of RsaA has shown that the extreme N-terminus of RsaA is required for surface attachment (Bingle et al., 1997b). The C-terminus, in addition to the secretion signal, contains 5-6 glycine-rich nonapeptide repeats centered with an aspartate residue (an RTX-like motif, see below), located between amino acids 860 to 920, which are thought to bind Ca^{2+} ions (Gilchrist et al, 1992).

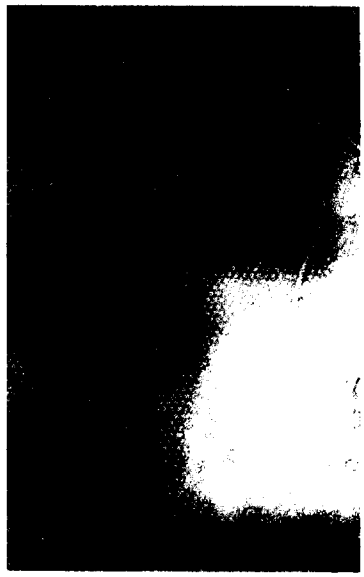

Figure 1-3. Shed S-layer from an SLPS-defective *C. crescentus* mutant - E M photograph (Photo courtesy John Smit)

The S-layer of *C. crescentus* can be used as a biotechnology vehicle. For a protein expression system to be suitable it is desirable to produce large quantities of recombinant protein that can be easily isolated from the rest of the cellular proteins. It was shown that the last 336 amino acids of RsaA (containing the Ca^{2+}-binding domain

and the secretion signal) are sufficient to target for secretion a foreign polypeptide fused to the N-terminus. The secreted recombinant protein, lacking the N-terminal surface attachment domain, is shed into the culture medium and forms aggregates that can be readily filtered away from the cells. The aggregates are typically 90% pure fusion protein, making *C. crescentus* a viable protein expression system (Bingle et al., 1997a). S-layer proteins can also be used for recombinant vaccine production, e.g. as carriers that expose useful epitopes on the cell surface (Sara and Sleytr, 1996a). Potential sites for epitope insertion were identified in the S-layer of *C. crescentus* (Bingle et al., 1997b). Another application for S-layers is molecular nanotechnology, where the crystalline arrays can be used as templates for metal or silicon atom deposition, to allow generation of circuitry finer than permitted by current technology. Also, S-layers can be utilized for production of ultrafiltration membranes with well-defined molecular cutoffs, or as surface supports for binding of monolayers of functional molecules (enzymes, antibodies, epitopes)(Sara and Sleytr, 2000; Sara and Sleytr 1996a; Sleytr et al., 1997). Evidently, all these uses can be applied to the S-layer of *C. crescentus*.

RsaA is secreted by a type I secretion apparatus (Awram and Smit, 1998), which is fairly uncommon among S-layers. Thus far, all S-layers from Gram-positive bacteria are secreted by the Sec-dependent General Secretion Pathway (GSP) (Messner and Sleytr, 1992; Sleytr et al., 1993), while S-layers of Gram-negative bacteria are exported by type II systems (Boot and Pouwels, 1996), which also employ the GSP to transport the protein across the inner membrane. Indeed, out of over 30 S-layer proteins from unrelated bacterial species whose sequences are known (Sara and Sleytr, 2000), the only other S-layers shown to be secreted by a type I system besides *C. crescentus* are the S-layer of *Campylobacter fetus* (Thompson et al., 1998) and an S-layer-like protein in *Serratia marcescens* (Kawai et al., 1998). The *C. fetus* S-layer protein is the most closely related to RsaA of all currently sequenced S-layer proteins: it is hexagonally-packed, anchors to the surface via its N-terminus to a particular LPS (Dvorkin et al., 1995), and has the greatest sequence similarity (24% identity, 40% similarity) to RsaA (Gilchrist et al., 1992).

There are four Gram-negative bacterial transport systems described so far, designated type I through IV. The **type I** system is utilized by *C. crescentus* to secrete RsaA and is described in detail below.

Type II systems use the Sec-dependent GSP to transport the substrate across the inner membrane and a complex of 12-14 proteins for secretion across the outer membrane. This system is utilized by the majority of secreted proteins in Gram-negative bacteria (for a review see Russel, 1998). All GSP-secreted proteins (for a review of GSP see Pugsley, 1993) possess an N-terminal signal peptide which is recognized by SecA proteins that deliver the substrate to the translocation machinery, which is a membrane-spanning complex made of Sec proteins. The signal peptide is inserted into the inner membrane and the remainder of the chain is threaded through in an unfolded state. The signal peptide is then cleaved off and the protein is released into the periplasmic space, where it folds into the native state, either spontaneously or aided by molecular chaperones. Folding is necessary for recognition and transport across the outer membrane by type II-specific secretion complexes (Lu and Lory, 1996).

Auto-transporting proteins, like the type II secreted proteins, use the GSP to cross the inner membrane. Examples of proteins secreted by an auto-transporter mechanism are the IgA proteases produced by *Neisseria meningitidis, Neisseria gonorrhoeae*, and *Haemophilus influenzae* (Klauser et al., 1993). They have an N-terminal signal sequence and a C-terminal pro-sequence. They are exported across the cytoplasmic membrane by the Sec dependent pathway, 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 systems have only been described in pathogenic bacteria, and are used to deliver bacterial proteins into the cytoplasm of the host, in order to alter the host's metabolism to the advantage of the bacterium (for a review see Galan and Collmer, 1999). They are present in *Yersinia* spp., *Shigella flexneri*, *Salmonella typhimurium*, enteropathogenic *Escherichia coli*, *Pseudomonas aeruginosa*, and *Chlamydia* spp. (Hueck, 1998). Type III systems consist of more than 20 proteins that form a needle-like

structure that spans both bacterial membranes (Kubori et al., 1998). Secretion occurs only after the bacterium makes contact with the host cell, and seems to be directed by the substrate mRNA. It is thought that the mRNA forms a hairpin loop that conceals 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 seems to be required for secretion, as components of type III systems hydrolyze ATP *in vitro* (Eichelberg *et al.*, 1994). It is believed that the substrate is then injected through the needle structure into the cytoplasm of the host cell (Hueck, 1998).

Type IV systems have only recently been discovered (for a review see Burns, 1999). This transport pathway, like the type III, has so far been found exclusively in pathogens. The prototypical type IV system is the *Agrobacterium tumefaciens* T-DNA transfer machine, which delivers oncogenic nucleoprotein particles to plant cells (Christie and Vogel, 2000). Other pathogens, including *Bordetella pertussis*, *Legionella pneumophila*, *Brucella* spp. and *Helicobacter pylori*, use type IV machines to export effector proteins to the extracellular milieu or the mammalian cell cytosol. There are at least 9 proteins involved in the transport and their roles are not well understood. Usually, there are two proteins containing nucleotide binding motifs that appear to be the transporting units that hydrolyze ATP to effect transport. It is not known whether the substrate is secreted across both membranes in one step, or across one membrane at a time, in two separate translocation events.

Type I secretion systems have three components (Delepelaire and Wandersman, 1991) (Figure 1-4). One component, the ABC (**A**TP-**B**inding **C**assette) transporter, is located in the inner membrane, recognizes the C-terminal secretion signal, and binds and hydrolyzes ATP during transport. (Binet and Wandersman, 1995). The second component, the membrane fusion protein (MFP), is anchored in the inner membrane and spans the periplasm (Dinh et al., 1994). The last component is an outer membrane protein (OMP) that interacts with the MFP. It is believed that the three components form a channel from the cytoplasm through the two membranes to the outside of the cell

(Hwang et al., 1997). ABC transporters have been shown to function as dimers (Davidson and Nikaido, 1991), while recent studies suggested that MFPs work as hexamers and OMPs as trimers (Holland, 1999; Koronakis et al., 2000).

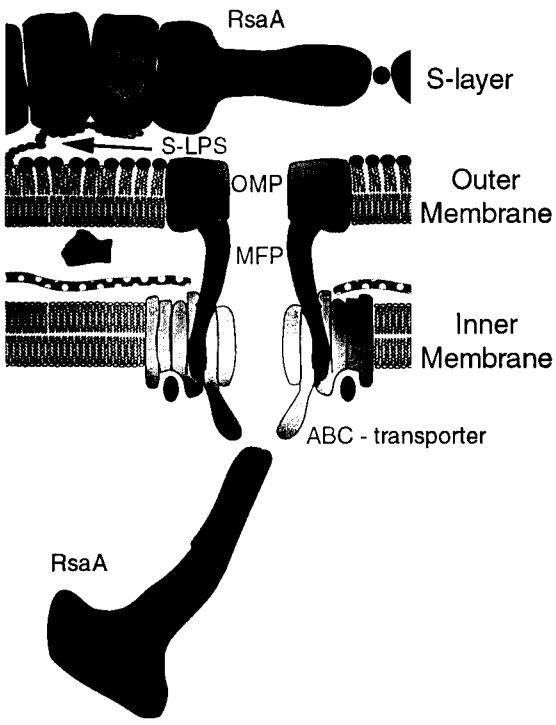

Figure 1-4. Type I secretion system. Diagram of the hypothetical membrane architecture of *C. crescentus* showing the predicted type I secretion mechanism of RsaA (courtesy Peter Awram)

ABC transporters form a very large family, with type I protein secretion systems accounting for only a small portion of them. They are present in all life forms and mediate transmembrane import or export of a wide variety of substrates, such as ions, sugars, amino acids, antibiotics, peptides and proteins (for a review see Higgins, 1992). ABC transporters are remarkably conserved, even between prokaryotes and eukaryotes. For instance, the eukaryotic P-glycoprotein, or the multidrug resistance protein, has more than 50% amino acid sequence identity with many bacterial ABC transporters over the entire length of the protein (Croop, 1998).

The typical ABC transporter is composed of four membrane-associated domains. Two of them, the transmembrane domains, consist of six to eight membrane spanning segments that together form the pore through the membrane, and determine the substrate specificity of the transporter. The other two domains are at the cytoplasmic side of the membrane, and contain the ATP-Binding Cassette (ABC), that binds ATP and provides energy for the transport by hydrolyzing ATP. The individual domains are most often expressed as separate polypeptides, especially in prokaryotes, but they can also be fused into larger, multifunctional polypeptides (for a review see Higgins, 1992).

The ABC domain consists of about 215 amino acids and is remarkably conserved among ABC transporters, with four distinct motifs (Linton and Higgins, 1998). The Walker A motif (consensus GxxGxK[S/T], where x – any residue) and Walker B motif (hhhhD, where h – any hydrophobic residue) are present in all ATP-ases, and are directly involved in the binding and hydrolysis of ATP (Walker et al., 1984). They are immediately followed by the ABC signature (LSGGQ[Q/R/K]QR) which is thought to be involved in energy transduction (Hyde et al., 1990). A fourth motif has recently been found in most of the *E. coli* and *Saccharomyces cerevesiae* ABC transporters (Linton and Higgins, 1998). This motif is a conserved histidine located ~ 30 amino acids downstream of the Walker B motif, usually preceded by four hydrophobic residues (hhhhH), and followed by a charged residue.

The other two components of the type I secretion machinery, the MFP and the OMP, have been considerably less studied than the ABC transporters, probably due to their much lower occurrence, which is restricted to bacterial type I systems.

The **MFP** has two domains: a hydrophobic N-terminal transmembrane domain that sits in the inner membrane, and a hydrophilic C-terminal domain consisting of β-sheets that span the periplasm and may interact with the outer membrane component to form a channel through both membranes (Dinh et al., 1994).

The **OMP** is located in the outer membrane and, by interacting with the MFP, links the transporter complex to the external environment. TolC from the *E.coli* α-hemolysin transporter system is the best studied OMP. Studies of 3-D crystals of TolC showed that three monomers assemble to form a continuous channel that spans both the outer membrane and the periplasmic space (Koronakis et al., 2000). The periplasmic end of the tunnel is sealed by sets of coiled helices, that are thought to be untwisted by an allosteric mechanism, mediated by protein-protein interactions, to open the tunnel (Koronakis et al., 2000).

Although ABC transporters can transport substrates ranging from ions to proteins, type I secretion systems have been found to only secrete proteins. Type I secreted proteins can be grouped into families (Table 1.1). The **RTX toxin family** is named for a glycine-

aspartate rich nonapeptide (consensus [L/I/F]xGGxG[N/D]Dx) called RTX (Repeats in Toxin), repeated 4-36 times (Welch, 1991). It was shown that the RTX motifs bind Ca^{2+}, which is required for the cytolytic activities exerted by most of the toxins (Boehm et al., 1990). RTX motifs are present in a large number of toxin proteins such as *E.coli* α-hemolysin (HlyA), *Bordetella perutssis* cyclolysin (CyaA), *Pasteurella haemolytica* leukotoxin (LktA) *Actinobacillus pleuropneumoniae* hemolysins (ApxIA, ApxIIA), *Neisseria meningitidis* iron-regulating proteins (FrpA,C) (Kuhnert et al., 1997). The **protease family** comprises extracellular proteases and lipases, such as *Pseudomonas aeruginosa* alkaline protease (AprA), *Erwinia chrysanthemi* metalloproteases (PrtA,B,C,G), *Serratia marcescens* metalloproteases (PrtA, SprA) and lipase (LipA), *Rhizobium leguminosarum* nodulation protein (NodO). A recently emerged family is composed of **S-layer proteins**: *Campylobacter fetus* S-layer homologs (SapA,B,A1) (Thompson et al., 1998) *S. marcescens* S-layer-like protein (SlaA) (Kawai et al., 1998) and *C. crescentus* RsaA. The transporters associated with these type I secreted proteins are presented in Table 1-1.

Not all type I secreted proteins have their own secretion system; sometimes two or more related or unrelated proteins are secreted by the same system (e.g., *E chrysanthemi* PrtA,B,C,G are all secreted by the PrtDEF transporter complex, and likewise *S. marcescens* PrtA, LipA and S-layer protein SlaA by LipBCD; see Table 1-1).

There is high sequence similarity among cognate transporter proteins (at least 50% aa sequence similarity) from the same family. Substrate proteins are secreted with high efficiency only by transport systems from the same family. Transport apparatuses from different families allow only low, yet detectable, levels of secretion (Akatsuka et al., 1997; Binet and Wandersman, 1995). Inactivated transporter genes could be complemented to detectable secretion levels only by transporter genes from the same family. In addition, studies on hybrid transporter complexes showed that they function with sufficient efficiency only within the families, and defined the ABC transporter as the component responsible for substrate specificity (Akatsuka et al, 1997; Binet et al., 1997).

Table 1-1. Type I secreted proteins and their transporters

Substrate	Organism	ABC	MFP	OMP/ linked
S-layer protein RsaA	*Caulobacter crescentus*	RsaD	RsaE	RsaF/no
S-layer homologs SapA,B, A1	*Campylobacter fetus*	SapD	SapE	SapF/yes
S-layer protein SlaA	*Serratia marcescens*	LipB	LipC	LipD/yes
Lipase A LipA	*Serratia marcescens*	LipB	LipC	LipD/yes
Major metalloprotease PrtA	*Serratia marcescens*	LipB	LipC	LipD/yes
Heme-binding protein HasA	*Serratia marcescens*	HasD	HasE	HasF/no
Alkaline phosphatase AprA	*Pseudomonas aeruginosa*	AprD	AprE	AprF/yes
Metalloprotease PrtA, B, C, G	*Erwinia chrysanthemi*	PrtD	PrtE	PrtF/yes
Hemolysin HlyA	*Escherichia coli*	HlyB	HlyD	TolC/no
Cyclolysin CyaA	*Bordetella pertussis*	CyaB	CyaD	CyaE/yes
Leukotoxin LktA	*Pasteurella haemolytica*	LktB	LktD	LktE/no
Nodulation protein NodO	*Rhizobium leguminosarum*	NodI	NodJ	NodT/no

C-terminal secretion signal. Type I secreted proteins lack the typical N-terminal signal peptides; rather, their secretion is driven by uncleaved C-terminal signals. The first type I secretion signal was identified in *E.coli* α-hemolysin using deletions and gene fusions, and it is located within the last 60 amino acids (Mackman et al., 1985). This short peptide can be secreted by itself or can promote secretion of foreign polypeptides fused to its N-terminus (Mackman et al, 1987). Similarly, deletion and fusion analysis showed that the last 29 aa of *E.chrysanthemi* protease PrtG ensure efficient secretion, and low but significant secretion can still be promoted by the last 19 aa (Ghigo and Wandersman, 1994). Another deletion study revealed that the 24 C-terminal amino acids are sufficient for secretion of *R.leguminosarum* nodulation protein NodO although

at 40% of the of the wild type levels, whereas the last 50 amino acids ensure 80% secretion levels (Sutton et al., 1996).

Defining the specificity of the secretion signal proved to be an elusive task. Extensive mutagenesis of *E.coli* α-hemolysin C-terminus has defined the last 46 amino acids as the targeting signal for secretion, but few individual amino acids turned out to be essential for export (Kenny et al., 1992; Stanley et al., 1991). This led to the hypothesis that a few critical residues scattered throughout the signal are necessary for substrate recognition by the transporter complex, while the other amino acids can be changed without effect on transport (Kenny et al., 1992).

Comparison of secretion signals of various type I secreted proteins showed no clear consensus sequence, but rather a lack of sequence similarity that makes the heterologous recognition of the export signals somewhat surprising (Koronakis et al., 1989). This lack of sequence similarity and scarcity of residues crucial for secretion suggested that the secretion signals rely on conserved secondary structure features rather than on conserved primary sequence motifs. Secondary structure predictions proposed a helix-linker-helix motif in the secretion signals of *E. coli* hemolysin HlyA and *P. haemolytica* leukotoxin LktA (Zhang et al., 1993). This motif was later confirmed by NMR and circular dichroism studies, and it was also found in *E. chrysanthemi* protease PrtG and *S.marcescens* heme-binding protein HasA (Wolff et al., 1997; Wolff et al., 1994). This conservation of secondary structure among different organisms suggested that the helix-linker-helix motif might be a prerequisite for transport. Its interaction with a binding pocket in the ABC transporter might induce a conformational change that results in the transport of the substrate protein (Zhang et al, 1998).

The "critical residues" model and the "conserved structural" model are two distinct explanations for the recognition of the secretion signal by the transporter complex, but in reality they are overlapping. Several elements near the C-terminus were proposed as being important to the secretion of RTX proteins: a cluster of charged residues, a putative amphipathic helix, and, within the last 10-15 aa, a cluster of hydroxylated residues (S or T) surrounded by hydrophobic aa (Koronakis et al., 1989; Ghigo and Wandersman, 1994). A conserved motif of an acidic residue (D or E) followed by a hydrophobic tail was found at the extreme C-terminus of type I secreted proteases,

lipases and NodO (Binet et al., 1997), and some of them had the helix-linker-helix motif (Wolff et al., 1997). No single type I secreted protein, however, contains all of these components. Random and directed mutagenesis of *E.coli* hemolysin HlyA secretion signal showed that the second helix is not critical for secretion; rather some individual residues were found to be essential irrespective of the secondary structure (Chervaux and Holland, 1995). A recent study, using random mutagenesis of the first helix, linker region and second helix followed by combinatorial analysis of the libraries, showed that the second helix is not required for transport, whereas the first helix is essential (Hui et al., 2000).

In conclusion, both models seem partially correct, and so recognition of the secretion signal perhaps requires a combination of conserved residues and secondary structure features.

RTX motifs. In addition to the C-terminal secretion signal, effective type I secretion of proteins also requires a Ca^{2+}-binding domain located upstream of the C-terminal signal (Lettofe and Wandersman, 1992; Ghigo and Wandersman, 1994). This Ca^{2+}-binding domain is the RTX motif, and it is present not only in all toxins, but also in proteases and lipases secreted by a type I system (Binet et al., 1997). Secretion levels of polypeptides fused to the C-terminus decline as the size of the C-terminal peptide is reduced, and for some passengers it is considerably lower if the RTX motif is not included (Letoffe and Wandersman, 1992; Ghigo and Wandersman, 1994). However, the RTX motifs are not essential for secretion; rather, it is thought that they are important for the proper presentation of the signal to the ABC transporter (Duong et al., 1996).

C. crescentus RsaA has an RTX-like motif between aa 860-920, with 5-6 repeats that are thought to bind Ca^{2+} (Gilchrist et al., 1992). As mentioned before, a protein fusion study showed that the RsaA secretion signal is located within the last 82 aa (Bingle et al, 2000). Nevertheless, the shortest RsaA C-terminal peptide capable of autonomous secretion and formation of macroscopic aggregates contained the last 242 aa, which include the RTX region. Shorter C-terminal peptides are secreted autonomously, but they only form microscopic aggregates. However, all C-terminal peptides produced

macroscopic aggregates when a cellulase domain was fused to them, the amount of protein recovered declining with their size. These observations suggest that the RTX region is not necessary for aggregation, but that it significantly influences the efficiency of secretion and perhaps the degree of aggregation (Bingle et al., 2000).

Other type I secreted S-layers have recently been identified in *Campylobacter fetus* (Thompson et al., 1998) and in *Serratia marcescens* (Kawai et al., 1998). Interestingly, there are no obvious RTX repeats in the *C. fetus* S-layer protein (Sap) (Thompson et al., 1998), whereas *S. marcescens* SlaA has six scattered RTX-like motifs (Kawai et al., 1998). The S-layer of *C. fetus* confers resistance to the host immune system by antigenic variation, acquired by inversion of a transporter gene-containing DNA fragment flanked by sap homologs (Thompson et al., 1998). Alignment of C-termini of four *C. fetus* Sap homologs and the C-terminus of RsaA revealed little primary sequence homology except for a few conserved residues, such as GDGS(T/G), GxTYV, V(V/I)D, DxxIKLxG (Thompson et al., 1998). Present in all strains were one of the motifs identified in toxins, 1-4 hydroxylated residues surrounded by hydrophobic aa within the last 10 aa, and an altered protease motif, the extreme C-terminus acidic residue followed by hydrophobic aa and one or two hydroxylated residues. Secondary structure analysis predicted a helix-linker-helix motif in all proteins examined, with the linker usually being a β-sheet (Thompson et al., 1998). These findings suggest that the secretion signals of *C. fetus* Sap homologs and RsaA contain elements that are specific to both families of type I secreted proteins, therefore placing the type I secreted S-layers somewhere in between the RTX toxins and proteases.

Arrangement of S-layer secretion genes. Usually, the genes encoding the ABC transporter and the MFP are linked to the gene for the substrate protein (Binet et al., 1997). A notable exception is the lipBCD system in *S.marcescens*, with the gene of one substrate (S-layer protein SlaA) directly linked to the transporter genes, and the genes of two other substrates (lipase LipA and protease PrtA) located somewhere else in the genome (Kawai et al., 1998). In addition, the *C.fetus* S-layer transporter genes (sapDEF) are located on an invertible DNA element flanked by two similar but not

identical S-layer genes, whose inversion is responsible for the antigenic variation of this bacterium (Thompson et al., 1998).

The *C. crescentus* S-layer transporter genes have been cloned and sequenced (Awram and Smit, 1998). They were designated *rsaD*, *rsaE* and *rsaF* (ABC transporter, MFP and OMP, respectively) (Awram and Smit, 1998). *rsaA*, *D* and *E* are adjacent and consecutive on the chromosome and are followed downstream by genes involved in SLPS synthesis (Awram and Smit, 1998) (Figure 1-5). The gene for the third component, *rsaF*, is located approximately 5 kb downstream of *rsaE* (Awram and Smit, 2000). RsaD, E and F are essential for the S-layer subunit secretion, their disruption abolishing the secretion of the S-layer on the cell surface (Awram and Smit, 1998, Awram and Smit, 2000). This indicates that the RsaDEF transporters are proprietary to the S-layer protein, and that secretion of RsaA is not shared by other type I systems that might or might not be present in *C. crescentus*.

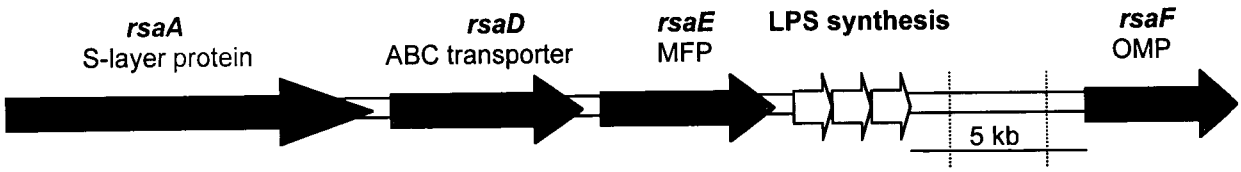

Figure 1-5. Diagramof *C. crescentus* S-layer secretion genes (genes not drawn to scale)

RsaD, E and F have high sequence similarity to other type I transporters from the protease family, the closest sequence similarity being to those for the alkaline protease (AprA) of *P. aeruginosa*, AprDEF, and the metalloproteases (PrtA,B,C) of *E. chrysanthemi*, PrtDEF (Awram and Smit, 1998). Both AprA and PrtB can be secreted in active form by the *C. crescentus* RsaA transport apparatus (Awram and Smit, 1998).

Other caulobacters. The bacterium first identified as *Caulobacter* was isolated from freshwater, and the majority of strains isolated thereafter came from freshwaters such as river water, pond water, well water, waste waters, and tap water (Poindexter, 1981).

Caulobacters are not restricted to freshwaters, and they also occur in seawater, soil, and the intestinal tract of some millipedes (Poindexter, 1981).

The Smit laboratory strain collection contains over 50 strains isolated from a variety of freshwaters and waste waters throughout North America, designated freshwater caulobacters (FWC) (MacRae and Smit, 1991). The majority of them were morphologically similar to *C. crescentus*, but were distinguishable by protein band profiles and gross colony morphology (MacRae and Smit, 1991). Subsequently, it was shown that about 40 of these FWC strains had an S-layer (Walker et al., 1992). A recent study, based on 16S rDNA sequencing, rDNA interspacer regions profiling, lipid analysis, immunological profiling and salt tolerance characteristics, reclassified the strains without S-layers in the genus *Brevundimonas* (Abraham et al., 1999). All S-layer FWC strains clustered within the same phylogenetic group, *Caulobacter sensu strictu* (Abraham et al., 1999).

The S-layer subunits of FWC strains range in size from 100 kDa (*C. crescentus*-like) to 193 kDa, and can be removed from the cell's surface using low pH or EGTA (Walker et al., 1992). All S-layer proteins react with an anti-RsaA antibody to some degree, and most of them have a SLPS that reacts to antibody against the *C. crescentus* SLPS (Walker et al., 1992). A probe containing the *rsaA* gene hybridizes to the genome of most S-layer-producing FWC strains under moderate stringency conditions (MacRae and Smit, 1991).

The purpose of this thesis was to elucidate several key aspects concerning the S-layer secreting FWC strains from the vast collection available in our laboratory. Early studies characterized the size of the S-layer proteins and their reactivity to an anti-RsaA antibody (Walker et al., 1992), but they were little studied since. The first question I asked was whether FWC strains that secrete S-layers so various in size did so by using, like *C. crescentus*, a type I secretion mechanism. Examination of the S-layer secretion signals, the changes in transporters required to accommodate the differently sized subunits of various FWC strains, and the arrangement of S-layer secretion genes are also important matters when characterizing a collection of related strains, so they were addresed in the present study.

This thesis demonstrates that the S-layers of FWC strains, independent of their size, are secreted by proprietary type I transporter systems, that are highly conserved. The transporter genes have the same arrangement as in *C. crescentus*, with the ABC transporter and the MFP genes adjacent to each other and in the same orientation, and the OMP gene further downstream on the chromosome. The S-layer gene location correlates with the sequence divergence in transporter protein among size groups, dividing the S-layer secreting FWCs into at least two groups. The small S-layer group has the same arrangement of S-layer secretion genes as *C. crescentus*, almost identical transporter genes and low similarity of secretion signals. The medium-large S-layer group has highly similar transporter genes and low similarity of secretion signals. The S-layer gene location further divides this group into subgroups: the medium-size group, with the S-layer gene further upstream of transporter genes, and the large S-layer group, with the S-layer gene elsewhere in the genome.

Chapter 2
Materials and Methods

Bacterial strains, plasmids and growth conditions. All strains and plasmids used in this study are listed in Table 2-1. The *E. coli* strain DH5α was used for all 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. Ampicillin, streptomycin and kanamycin were all used at 50 μg/ml in *E. coli*, and chloramphenicol at 20 μg/ml.

A subset of six FWC strains, two from each size group, was chosen so that the size of the S-layer subunit spans the entire range (Table 2-1). The locale where each of these strains was isolated, together with a list of all FWC strains in the Smit laboratory strain collection, are shown in Table 2-2. Most of the work was performed on these six strains and the results, if corroborative, were extrapolated to other FWCs in the same size group.

All caulobacter strains were grown at 30° C in PYE medium (0.2% peptone, 0.1% yeast extract, 0.01% $CaCl_2$, 0.02% $MgCl_2$), with 1.2% agar for plates. For some experiments, to obtain higher cell density, the medium was further supplemented with glucose to 0.1% and ammonium chloride to 0.05%. Streptomycin and kanamycin were used at 50 μg/ml, and chloramphenicol at 2 μg/ml.

Recombinant DNA manipulations. Standard methods of DNA manipulation and isolation were used (Sambrook et al., 1989). Plasmid DNA was extracted from *E. coli* using Qiagen (Qiagen Inc., Chatsworth, Calif.) or Invitrogen (Invitrogen Inc., Carlsbad, Calif.) miniprep kits. Chromosomal DNA was isolated as previously described (Yun et al., 1994). DNA-modifying enzymes were purchased from Life Technologies (Burlington, Ontario, Canada) or New England BioLabs (Mississauga, Ontario, Canada). *E. coli* and caulobacter strains were made electrocompetent, and electroporated as described previously (Gilchrist and Smit, 1991). DNA fragments were purified from agarose gels with a QIAEXII (Qiagen Inc., Chatsworth, Calif.) kit.

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

Strain or plasmid	Relevant characteristics	Reference or source
E. coli		
DH5α	recA1, endA	Life Technologies
Caulobacter sp.		
CB15A	Ap^r syn-1000 Variant of *C.crescentus* wild-type strain (ATCC 19089) that synchronizes well	
CB2A	Spontaneous Rsa- mutant of *C. crescentus* strain CB2 (ATCC15252)	
FWC1	freshwater strain, predicted S-layer Molecular Weight (MW) 100 kDa	Walker et al., 1992
FWC19	freshwater strain, predicted S-layer MW 108 kDa	Walker et al., 1992
FWC9	freshwater strain, predicted S-layer MW 133 kDa	Walker et al., 1992
FWC16	freshwater strain, predicted S-layer MW 151 kDa	Walker et al., 1992
FWC27	freshwater strain, predicted S-layer MW 145 kDa	Walker et al., 1992
FWC39	freshwater strain, predicted S-layer MW 193 kDa	Walker et al., 1992
FWC42	freshwater strain, predicted S-layer MW 181 kDa	Walker et al., 1992
Plasmids		
pBBR3	Sm^R, broad host range vector	Awram, 1999
pBBR3AprA	Sm^R, *aprA*$^+$; *aprA* cloned into pBBR3 as *Eco*RI/*Pst*1	Awram, 1999
pBSKS+	Ap^R, *lacZ*; ColE1 cloning vector	Stratagene
pTZ18U, pTZ19U	Ap^R, *lacZ*; ColE1 cloning vectors, phagemid versions of pUC18 or pUC19	Mead et al., 1986
pTZ18U(CHE)	Cm^R; Ap^R gene of pTZ18U replaced with Cm^R gene	Awram, 1999
pTZ18U(Km)	Km^R; Ap^R gene of pTZ18U replaced with Km^R gene	Awram, 1999
pTZ19UΔSSm	Sm^R; Sm^R gene inserted into *Sca*1 site in Ap^R gene of pTZ19U	Awram, 1999
pRAT6	Cm^R, *rsaD*$^+$, *rsaE*$^+$; *rsaD* and *rsaE* genes cloned into pBBR1	Awram, 1999
pRAT9	Sm^R, *rsaD*$^+$, *rsaE*$^+$; *rsaD* and *rsaE* genes cloned into pBBR3	Awram, 1999
pBBR1CE	Cm^R, *lacZ*, pBSKS multiple cloning site, broad host range vector	Smit lab creation
pBBR1CE(0-690)	Cm^R; C-terminus of *rsaA* (last 336 aa) cloned in frame to *lacZ* of pBBR1CE as *EcoRI*/*Pst*I	Smit lab creation

Southern blot hybridizations with radiolabelled probes were done according to the membrane manufacturer's instructions (Amersham Hybond-N+, Amersham Life Science, Mississauga, Ontario, Canada) or, for DIG-labelled probes, following the DIG labelling kit manual. Probes were DNA fragments or PCR products labeled with ^{32}P using a RediPrime (Amersham Pharmacia Biotech Inc., Piscataway, NJ) kit, or with digoxigenin (DIG) using a DIG-labelling (Boehringer Mannheim GmbH) kit. After hybridization, blots were washed twice for 5 min at room temperature with 2xSSC (0.03M Na Citrate, 0.3 NaCl pH 7.0), 0.1% SDS, and twice for 15 min at: 65° with 0.1xSSC, 0.1% SDS for high-stringency conditions, 65° with 1xSSC, 0.1% SDS for moderate stringency, and room temperature with 1xSSC, 0.1%SDS for low-stringency. Following washing, blots were exposed to film for radioactive probes, or detected according to the manual for the DIG-labelled probes.

DNA colony hybridizations were performed as follows: the colonies were adsorbed to filter paper discs (Whatman 541), lysed by soaking the disks in 0.5M NaOH for 5 min,

then neutralized in 1M Tris-HCl (pH 7.0) for 5 min twice. The discs were then soaked in 0.5M Tris-HCl (pH 7.0), 1.5 M NaCl for 5 min, washed with 70% ethanol and baked for 2 h at 80°. The discs were hybridized with radiolabeled PCR products using the high-stringency Southern blot procedure described above. To exclude false positive results, positive colonies were transferred to new plates and subjected to a second colony hybridization with the same probe.

Table 2-2. List of S-layer producing FWC strains in the Smit laboratory collection.
The subset of 6 strains used in this study are shown in bold letters, and their source is indicated.

FWC size group	S-layer MW (kDA)	Source
Small S-layers		
C. crescentus	100	Pond water
NA1000 or **CB15A**		
FWC 1	100	Freshwater slough next to Lake Washington, Seattle, USA
FWC11	108	
FWC15	110	
FWC17	106	
FWC19	108	Secondary treatment, activated sludge process, UBC
FWC20	108	
FWC22	107	
FWC25	105	
FWC28	106	
FWC31	106	
FWC33	110	
FWC34	110	
FWC35	102	
FWC44	106	
FWC46	110	
FWC50	115	
FWC51	105	
Medium S-layers		
FWC2	132	
FWC8	122	Secondary treatment, activated sludge process, UBC
FWC9	133	
FWC12	133	
FWC16	151	Primary treatment facility for Greater Vancouver, Iona Island, B.C.
FWC18	131	
FWC24	147	
FWC26	140	
FWC27	145	
FWC29	124	
FWC32	133	
FWC37	149	
FWC41	137	
FWC45	140	
Large S-layers		
FWC6	181	
FWC7	177	
FWC39	193	Secondary treatment facility, activated sludge, Kelowna, B.C.
FWC42	181	Secondary treatment facility, activated sludge, Calgary, Alberta

Cloning of FWC transporter genes and C-terminal secretion signal. The following approaches were used:

Screening of FWC genomic libraries. Genomic libraries were made by cloning FWC chromosomal DNA digested with EcoRI and HindIII into the cloning vector pTZ19U. The libraries were screened for transporter genes by colony blot hybridization using an 800-bp radiolabelled PCR product derived from the *rsaD* gene of *C. crescentus* (primers RD43-IRD477, Table 2-3). Screening for S-layer protein genes was done with a PCR product spanning the *rsaA* gene (primers RAT1330-17[0-1], Table 2-3). The positive clones obtained, containing portions of FWC transporter genes, were subcloned and sequenced using the pUC universal primers.

Chromosomal integration using transporter gene fragments already cloned as described above. The strategy is depicted in Figure 2-1. Briefly, a plasmid that doesn't replicate in caulobacters (pTZ19UΔSSm, Table 2-1) containing transporter gene fragments was introduced into the corresponding FWC. The single cross-over integration events were selected for on Sm. Integration was checked by PCR with two primers that amplify an 800 bp internal fragment of the suicide vector (TZ1920-TZI1060, Table 2-3). Chromosomal DNA was then extracted, cut at restriction sites outside the integration region, ligated and electroporated into DH5α. Transformants containing DNA adjacent to the integration region were subcloned and sequenced with the pUC universal primers.

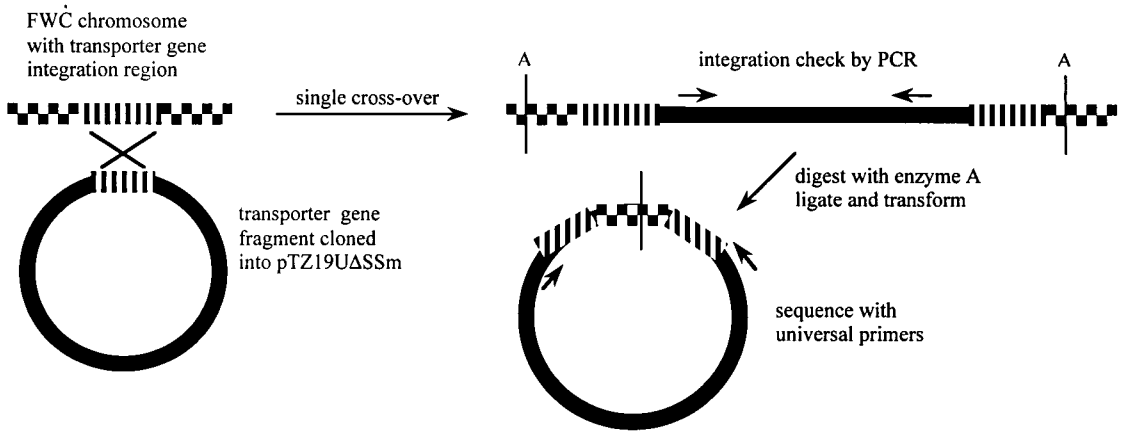

Figure 2-1. Chromosomal integration strategy for cloning DNA adjacent to FWC transporter genes. See text for details.

PCR amplification of FWC genomic DNA and cloning of PCR products. Primers were designed to bind to regions that are conserved among S-layer genes and transporter genes of various type I systems (Table 2-3). Assuming that these regions are also conserved in FWC strains, genomic DNA was amplified by PCR with *TaqI* polymerase (Life Technologies), initially at annealing temperatures 5-10° below the melting point of the primers. Melting temperature for primers was calculated by adding 2° for each A or T and 3° for each G or C. If products of the same size as those generated from *C. crescentus* were obtained, the annealing temperature was increased step-wise as much as possible, to eliminate mispriming but to keep the desired product. General PCR parameters were denaturing at 95° for 30 sec, annealing for 30 sec and extension at 72°. Extension times were based on 60 sec/1000 bp of DNA. PCR products generated from FWCs always contained contaminating products of different size due to some degree of mispriming, which precluded direct cloning into a T-tailed vector. Rather, the desired product was separated by agarose electrophoresis, gel purified then blunted, since *TaqI* polymerase adds a 5' adenosine to ~ 50% of the strands. Blunting was performed following the manufacturer's instructions, with *Vent* polymerase (New England BioLabs) or Klenow fragment of DNA polymerase I (Life Technologies). Next, the PCR product was cloned into a blunt-ended vector (pBSKS cut with *EcoR*V), then subcloned and sequenced with the pUC universal primers. If only products of a different size than those from *C. crescentus* were obtained even after adjusting the annealing temperature, they were still cloned and sequenced.

Amplification with *Vent* and *Pfx* polymerases was also tried, because it generates PCR products with blunt ends that, after gel purification, can be cloned directly, and consequently with better yields, into a blunt-ended vector.

Primers for transporter genes (Table 2-3). Sequence alignments of RsaD and RsaE with closely related type I transport systems from other species (*P. aeruginosa* alkaline phosphatase AprA, *E. chrysanthemi* metalloprotease PrtB, *S. marcescens* lipase LipA and heme-binding protein HasA) were used to design pairs of primers in the most conserved regions (data not shown).

Table 2-3. Primers used for PCR in this study

Primer - sequence	Gene/region	Direction	Fragment amplified in *C. crescentus*/Comments
RD43 - TA(TC)ATGCT(GC)CAGGT(GC)TAT(GC)ACCGIG	*rsaD*/aa 43	downstream	~800 bp of *rsaD*
IRD477 - C(GC)A(GT)(GC)CGCTG(GC)CGCTGGCCGC	*rsaD*/aa 477	upstream	
RTXL - ACGATCCGCGGCGGCGCTGGCGCCG	*rsaA*/aa 860 *rsaD*/aa 44	downstream	~800 bp - C-terminus *rsaA* and front end of *rsaD*;
IRD44 – CAGCACGCGGTCATAGACCTGCAGC		upstream	works only if *rsaA* and D are adjacent
RD476 – CCGGCGGCCAGCGCCAGCGCCTGGCC	*rsaD*/aa 476 *rsaE*/aa 438	downstream	~1,800 bp - C-terminus *rsaD* and almost all of *rsaE*;
IRE – GGTACTGCAGCACGGTGCGCTCGCC		upstream	works only if *rsaD* and *E* are adjacent
ICT – TAGGTGTC(GA)CC(GA)CC(GA)TACTGGAACCA	*rsaA*/aa1,000	upstream	together with RTXL amplifies *rsaA* C-terminus ~400 bp
RTXR – ATCCGCGCCCGTGCCACCCGTGAAGG	*rsaA*/aa 910	upstream	together with RTXL amplifies the RTX-like region ~ 150 bp
F39AN - GTIACITT(GC)GCICAGCT(GC)AC(GC)(AT)(GC)ITA(CT)TTCAC(GC)AA	FWC39 *rsaA* /aa 1-13	downstream	with ICT or RTXR might amplify FWC39 S-layer gene
KOD – ATCGC(GC)GCCATGGT(GC)TTCAGCTTCTTC	FWC19,16, 39 *rsaD*/aa 24	downstream	with pUC universal primer to confirm integration in ABC transporter gene
RAT1330 – GAATTTGCTGTACCGGTTAGAAAAATGCT	*rsaA* upstream of ATG 100bp	downstream	~ 3.2 kb – full-length *rsaA*
17[0-1] - CAGGACTTCGGTGGCGA	*rsaA* aa 1021	upstream	
FWCD140 – ATCGCG(GT)T(CT)TG(CT)GACGCGCC(GC)TGGAC	FWC strains *rsaD*/ aa 140	downstream	sequencing primer
EATG – ATGAAGCCCCCCAAGATCCAGCGTCCG	FWC strains *rsaE*/ aa 1	downstream	used primarily as a sequencing primer; together with RD476 amplifies 1.2 kb of *rsaE*
EMID – GGCAG(GC)CG(AC)ACCTCGGTGACCATGCCC	FWC strains *rsaE*/ aa 341	upstream	sequencing primer
RsaF973 -GAATTCCGGGGTCGGCTACAGTCTAGGAAATGC	rsaF uptream of ATG	downstream	~ 1.5 kb - full length rsaF
IRsaF973B - GAGCTCCCTAGAGCGTTCTCCGATCCGTGCG	rsaF down-stream of stop	upstream	
TZ1920 - GAGGCCTAGTACTCTGTCAGACCAAGTTTACTCATA	pTZ position 1920	dowstream	used to check for chromosomal integration, together they amplify a 800 bp internal fragment of pTZ19U
TZI1060 - GAGGCCTACTCTTCCTTTTTCAATATTATTGAA	pTZ position 1060	upstream	

Primers for the C-terminus of RsaA (Table 2-3). RTXL and RTXR were designed to bind at both ends of the RTX-like region of RsaA (amino acids 860-910, Gilchrist et al., 1992), assuming that RTX-like motifs are present in the S-layer proteins of FWC strains. ICT, intended for cloning the C-terminus of large S-layer genes, was designed after we obtained S-layer C-termini sequence for several FWC strains, in the most conserved region among *C. crescentus*, FWC1 and FWC19.
F39AN was designed from the N-terminal protein sequence of the FWC39 S-layer subunit. This sequence was obtained by N-terminal sequencing of EGTA extracted FWC39 S-layer protein that was transferred onto a nylon membrane.

Nucleotide sequencing and sequence analysis. Sequencing was performed by the dideoxy termination method on a DNA sequencer (Applied Biosystems model 373). After use of universal primers, additional sequence was obtained by subcloning, or by using primers based on the acquired sequence, designed to bind to regions conserved among the subset of FWC strains (FWCD140, EATG, EMID). Nucleotide and amino acid sequence data were analyzed using MacVector software (Oxford Molecular Group) and the NCBI BLAST e-mail server using the BLAST algorithm (Altschul *et al.*, 1990). Protein alignments were generated using the ClustalW algorithm as implemented by the MacVector software.

Isolation of cell-associated and cell-free S-layer proteins. Cell-associated S-layer protein was extracted from the cell surface by EGTA treatment (10 mM EGTA in 10mM HEPES, pH 7.5), as previously decribed (Walker et al., 1992), and directly analyzed by SDS-PAGE. When the secreted protein fails to attach to the surface, it forms insoluble macroscopic particulate material in the culture fluid (Walker et al., 1994). The formation of aggregates is promoted by a large culture volume relative to flask volume, slow shaking speeds and growth to stationary phase (Bingle et al., 1996). Aggregated protein from such cultures was recovered by gravity filtration through a fine mesh, washed with distilled water, solubilized in 8M urea, 100mM Tris-HCl, pH 8.5, and used directly for SDS-PAGE and Western analysis.

SDS-polyacrylamide electrophoresis (SDS-PAGE) and Western blot analysis. SDS-PAGE and Western immunoblot analysis were performed according to standard procedures (Sambrook et al., 1989). After SDS-PAGE, proteins were visualized by Coomassie staining. For Western analysis, proteins were transferred to a nitrocellulose membrane, blots were probed with anti RsaA C-terminus polyclonal antibody, and antibody binding was visualized with goat anti-rabbit serum coupled to horseradish peroxidase, and color-forming reagents (Smit and Agabian, 1984).

Protein cross-expression in FWC strains.

Recognition of the C-terminal secretion signal of *C. crescentus*. FWC strains were made electrocompetent and transformed with a construct containing the coding region for the last 336 amino acids of RsaA on a compatible plasmid (pBBR1CE(0-690), Cm^R), and plated on PYE+Cm plates. The protein aggregates secreted in the medium (if any) were isolated as described above and subjected to SDS-PAGE to check for the size of the secreted protein. A duplicate of the same gel was subjected to Western blotting with polyclonal antibody raised against the last 336 aa of RsaA.

Recognition of FWC19 C-terminal secretion signal by *C. crescentus*. Once available, the C-terminus of FWC19 was cloned in frame to the first 5 aa of lacZ from pBBR1CE(0-690), introduced into CB2A, the S-layer defective *C. crescentus* mutant, and the secreted protein was analyzed as described above.

Secretion of *P. aeruginosa* AprA. Electrocompetent FWCs were transformed with the full-length *aprA* gene cloned on a compatible plasmid (pBBR3 aprA, Sm^R) and plated on skim milk-PYE plates (PYE/Sm+0.1% skim milk). Secretion of alkaline protease outside the cells was detected by halo formation around the colonies, due to casein degradation (Awram and Smit, 1998).

Inactivation of FWC transporter genes. For one strain from each size group (FWC19, 16, 39), a fragment of the ABC transporter gene was subcloned into a vector that is non-replicative in caulobacters (pTZ19UΔSSm, pTZ18U(CHE) or pTZ18U(Km)) and introduced into the corresponding FWC. Cells in which the vector had integrated inside the ABC transporter gene by homologous recombination were selected as described

above for chromosomal integration. In order for the gene to be inactivated, it was necessary that neither the N- nor the C- terminus of the gene be present on the fragment used for recombination. Transformants were grown in PYE and the corresponding antibiotic and checked for integration by PCR with two primers: KOD (designed in a region N-terminal of the integration fragment, that was conserved among FWC19, 16, 39 ABC transporters, Table 2-3) and a pUC universal primer (forward or reverse, depending on the orientation of the fragment in the multiple cloning site). A PCR product of correct size is generated *only* if the integration occurred in the desired transporter gene.

S-layer secretion in the transporter-inactivated strains (clones that were positive for integration) was checked by performing an EGTA extraction on the cells, and subjecting the extract to SDS-PAGE, alongside the wild type extract.

Complementation of the transporter-inactivated strains was attempted, by introducing the full-length *C. crescentus rsaD* and *E* genes under a *lacZ* promoter on plasmids carrying a different antibiotic than the one used for integration (pRAT6, Cm^R, pRAT9, Sm^R) in the knocked-out strains. Transformants were selected on PYE plates containing both antibiotics and S-layer secretion was assessed by EGTA extraction and SDS-PAGE.

Mapping of FWC S-layer secretion genes.

The OMP gene. Genomic DNA of the subset of FWC strains, digested with various single restriction enzymes, was subjected to high-stringency Southern blot hybridization using as a probe a PCR product amplified from the *C. crescentus rsaE* gene (primers EATG-RD476, Table 2-3). Only the enzymes that generated fragments larger than ~5 kb were used further. The procedure was repeated, then the blots were stripped and probed under the same conditions with a probe derived from the *rsaF* gene (primers rsaF973-IrsaF973, Table 2-3). If the OMP gene were located in the vicinity of the MFP gene in FWCs, the same fragment should react to both probes. To confirm that the bands observed with the two probes were identical, the blots were stripped again and probed with *rsaE* and *rsaF* at the same time. If only one band was detected, it was assumed that a single DNA fragment reacted to both probes.

The S-layer gene. Genomic DNA of FWC39 and FWC 42 was subjected to the same mapping procedure as described above for the OMP gene, but with probes derived from *rsaD* (primers RD43-IRD477) and *rsaA* (RAT1330-17[0-1]), and using a low stringency hybridization with the *rsaA* probe (1x SSC room temperature washes), due to the lesser sequence similarity of FWC S-layer genes to *C. crescentus rsaA* (MacRae and Smit, 1991).

Chapter 3

Cloning and sequencing of transporter genes and S-layer C-terminal secretion signals from FWC strains

The main purpose of this thesis was to assess whether all FWC strains secrete their S-layers by a type I mechanism, despite the significant difference in the size of their S-layer subunits (100 kDa to 193 kDa, Walker et al., 1992). In a preliminary study in our laboratory (Awram and Smit, unpublished data) DNA fragments with sequence similarity to *C. crescentus rsaD* were isolated from a few FWC strains, indicating the presence of ABC transporters in these strains. However, it was not known if the ABC transporters identified in this preliminary study were solitary, or whether they were part of a type I system. Even if the latter were the case, that particular type I system was not necessarily the one that secreted the S-layer subunits. To clarify these ambiguities, cloning and sequencing of the ABC transporter genes and the DNA upstream and downstream was required. This would allow comparison of the transporter genes among FWC strains with various S-layer sizes, hopefully providing insight about the sequence changes required to accommodate subunits so different in size. In addition, cloning of the FWC S-layer C-termini was attempted, to verify the presence of type I secretion signals, compare their sequences, and look for a correlation between them and the transporters at the sequence level.

Results

Screening of FWC genomic libraries with a *rsaD*-derived probe was the first approach used to isolate transporter gene fragments. The method is described in detail in Chapter 2, and the fragments cloned in this manner are represented in Table 3-1 and Figure 3-1. All fragments contained DNA regions with high sequence similarity to *rsaD*, and some of them to *rsaE*, with the same arrangement as in *C. crescentus*. This latter finding suggests that the ABC transporters identified in the preliminary study are not isolated in the genome, but rather are part of a type I system similar to that of *C. crescentus*.

Table 3-1. Transporter gene fragments cloned by screening of FWC genomic libraries

Strain	Size cloned (EcoRI-HindIII)
FWC19	2.8 kb
FWC42	6.0 kb
FWC39	2.3 kb

Chromosomal integration. This method (see Chapter 2) was applied successfully only to FWC19, yielding an extra 800 bp upstream and 1 kb downstream of the initial clone (Figure 3-1). The success of this strategy is limited by the availability of restriction sites located at distances short enough so as to still allow efficient ligation (usually less than 15 kb, Sambrook et al., 1989). This is probably the reason this method failed when applied to other FWC strains for which transporter genes fragments were isolated (Table 3-1).

PCR amplification of FWC genomic DNA and cloning of DNA products. This third method was adopted because it allows an immediate, target-directed isolation of PCR products that span the FWC genes of interest, instead of random restriction fragments generated by the screening method. The PCR parameters and the choice of primers are described in Chapter 2. The PCR products obtained by this method are presented in Figure 3-1 and Table 3-2.

This part of the project was notably arduous, because the PCR products generated were usually accompanied by numerous other products that prevented a one-step cloning into a T-tailed vector. Rather, they required gel extraction followed by Vent polymerase or Klenow blunting, which drastically lowered the cloning yields especially for large products. Amplification with Vent or Pfx polymerase, which both produce blunt-ended fragments, generated significantly different product patterns compared to Taq, and sequence analysis of the predominant PCR product produced no match to the expected genes at either end. For this reason, despite the low yields, amplification with Taq followed by gel extraction and blunting was preferred, although occasionally it led to misprimed products. Repeated annealing temperature adjustments and sometimes a second round of amplification (using as template the agarose plug containing the desired band) were also required.

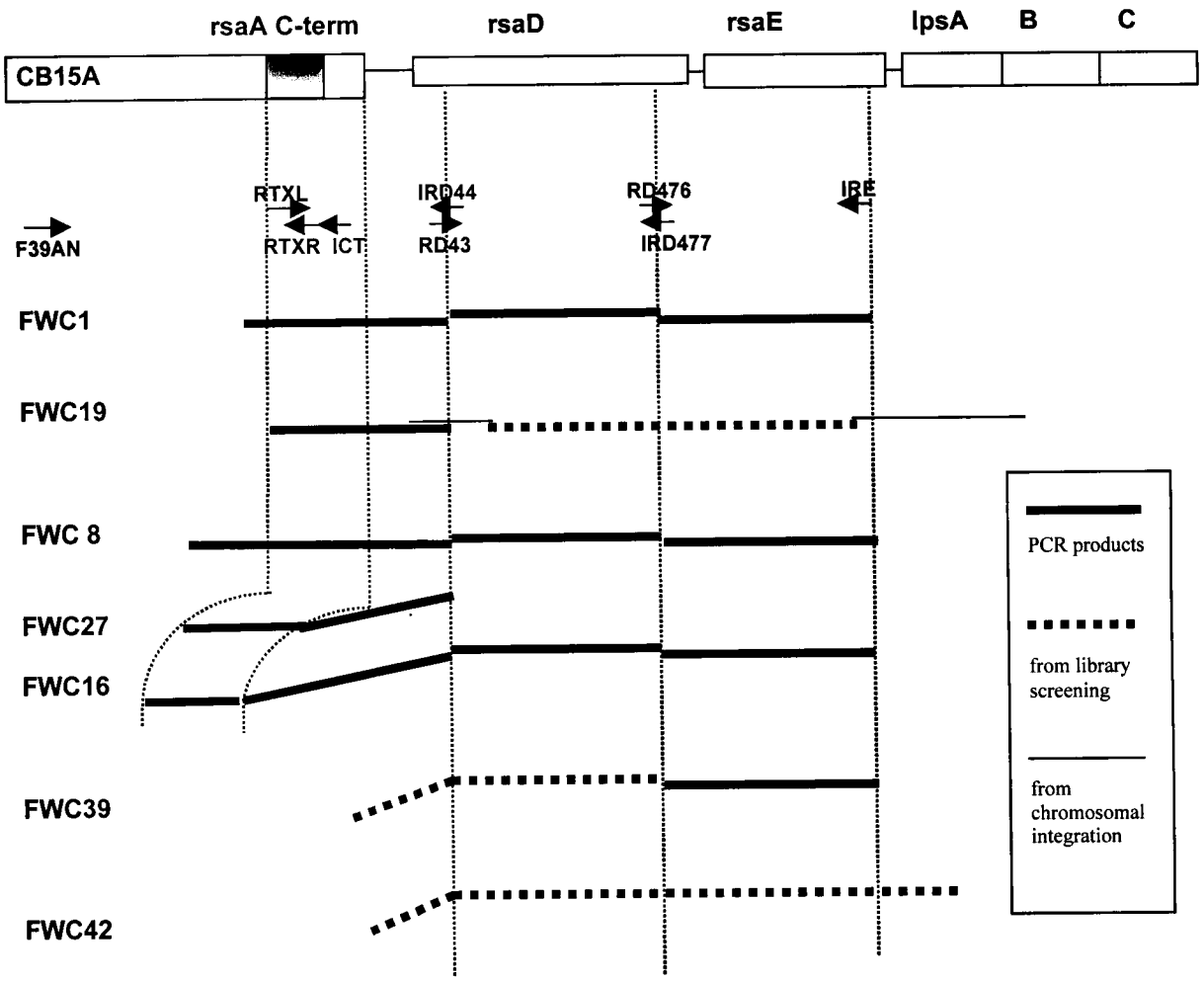

Figure 3-1. FWC transporter genes and S-layer C-terminus-encoding regions isolated in this study. Lines corresponding to a FWC strain are shown as discontinuous and in different styles to illustrate the individual clones. Horizontal lines denote sequenece similarity of the FWC gene to the corresponding *C. crescentus* gene, and their styles indicate: thick solid - cloned PCR products; thick dashed – fragments from library screening; thin solid – fragments cloned by chromosomal integration. Vertical lines delineate the PCR products generated by various primers. Sloping lines indicate FWC DNA regions with no sequence similarity to the corresponding regions in *C. crescentus*.

Table 3-2. PCR products cloned in this study

Pair of primers/ strain	PCR product	Comments
RD43-IRD477		
FWC 1	~ 800 bp	same size as *C. crescentus*; sequence similarity to *rsaD*
FWC16/FWC8	~ 800 bp	
RD476-IRE		
FWC1	~ 1,800 bp	**same** size as *C. crescentus*; sequence similarity to *rsaD* at one end and *rsaE*
FWC16/FWC8	~ 1,800 bp	at the other → ABC and MFP genes are adjacent
RTXL-IRD44		
FWC1/FWC19	~900bp/ ~800 bp	• **FWC1** product slightly bigger than *C. crescentus* (800 bp); sequence similarity to *rsaA* at one end and *rsaD* at the other
FWC16	~ 800 bp	• ***C. crescentus*** size; *rsaD* sequence similarity at one end and nothing at the other → S-layer gene elsewhere
	~ 2.2 kb	• **significantly** bigger than *C. crescentus*; sequence similarity to *rsaA* at one end and *rsaD* at the other → S-layer gene ~ 2 kb upstream of ABC transporter gene
FWC8	~1.2 kb	• **FWC8** product bigger than *C.crescesntus*; sequence similarity to *rsaA* at one end and *rsaD* at the other
FWC27	~1.7 kb	• sequence similarity to *rsaA* at one end and *rsaD* at the other
FWC39/FWC42	no product	• **lowering** annealing temperature did not help
RTXL-RTXR		
FWC39/FWC42	~500 bp	*C. crescentus* product is 160 bp; sequencing revealed no sequence similarity to *rsaA* → mispriming
RTXL-ICT		
FWC39/FWC42	~700 bp	*C. crescentus* product is 500 bp; sequencing revealed no sequence similarity to *rsaA* → mispriming
FWC39AN		
F39AN-RTXR	~ 2 kb	**no product** in *C. crescentus* since F39AN sequence doesn't match *rsaA*; sequencing revealed no sequence similarity to *rsaA* → mispriming
F39AN-ICT	~ 3.5 kb ~1.9 kb	

The search for the large S-layer genes. All approaches used to isolate the S-layer genes of the high MW group of FWC strains were unsuccessful , as shown in Figure 3-1 and Table 3-2. The first approach, sequencing of FWC39 and FWC42 clones isolated from library screening with an *rsaD*-derived probe revealed that the DNA upstream of the ABC transporter gene did not have sequence similarity to *rsaA* but rather to a *Sinorhizobium meliloti* galactosyl transferase, ExpC (Becker et al., 1997). This finding

indicated that the S-layer gene of the high MW group wasn't located immediately upstream of the transporter genes. In fact, as shown in Chapter 5, the S-layer gene was not located within ~6 to 8 kb upstream or downstream of the ABC gene in either FWC39 or FWC42. This finding precluded the successful application of the chromosomal integration and inverse PCR approaches.

The next approach, screening of libraries with a *rsaA*-derived probe, was unsuccesful due to the low stringency required to get detectable bands by Southern hybridization (room temperature, 1x SSC washes, see Chapter 2). Nevertheless, the *rsaA*-derived probe generated signals which, although weak, were still detectable. This indicated that large S-layer genes have low sequence similarity to *rsaA*, yet enough to yield distinct bands with a *rsaA* probe under low stringency conditions. Applying the same stringency to colony blots would have resulted in practically every colony reacting to the probe. An identical predicament was encountered when trying to screen with the C-termini of small or medium sized S-layers.

Probing of FWC39 genomic DNA with the end-labeled F39AN oligo yielded hybrids only under low-stringency conditions, perhaps due to the high degeneracy of the oligo and its small size. These hybridization conditions rendered the method unusable for library screening.

The amplification of genomic DNA with various pairs of primers also failed to provide the correct clones, as shown in Table 3-2. RTXL-IRD44 primers could have amplified the C-terminus of the S-layer gene and N-terminus of ABC transporter, RTXL-R the RTX region, RTXL-ICT the C-terminus of the S-layer gene and F39AN-RTXR/ICT almost the entire S-layer gene. Amplification of FWC39 or FWC 42 with the primers RTXL-IRD44 yielded no product, even after lowering the annealing temperature or increasing the extension time, probably because the S-layer gene was not located upstream of the transporter genes. The pairs of primers showed in Table 3-2 usually yielded many different products, and cloning of the predominant ones revealed no sequence similarity to RsaA at either end. Further subcloning of some products was also performed to sequence the internal regions, but no *rsaA* sequence similarity was found.

Sequence analysis. All fragments and PCR products isolated from FWC strains were first sequenced with the universal primers, and then analyzed further only if they contained DNA with sequence similarity to *rsaA*, *rsaD* or *rsaE* at least at one end. Additional sequence was obtained by subcloning, or by designing primers based on the acquired sequence (see Chapter 2). Nucleotide and amino acid sequence data were analyzed as described in Chapter 2. At the DNA level, the codon usage of the transporter genes and the S-layer C-termini was similar to that of *C. crescentus*, with a strong bias towards the use of G and C at position 3, and an overall G+C content of >60% (Gilchrist et al., 1991). The GC content of the FWC strains analyzed, based on the sequence acquired, was as follows: FWC1 – 65.2%, FWC19 – 64.7%, FWC8 – 64.5%, FWC16 – 64.7%, FWC39 – 61.3%, FWC42 – 62.6%.

The ABC transporters exhibited the closest sequence similarity, with an overall 80-85% identity and 90-95% similarity at the protein level (Figure 3-2). All ABC transporter-specific motifs were identical in all strains studied (Figure 3-2). Moreover, based on sequence conservation, FWC strains could be divided into two groups: the small S-layer group (FWC1, FWC19 and *C. crescentus*) and medium-large S-layer group (FWC8, FWC16, FWC39, FWC42), within which the sequence similarity is even higher. The variations in protein sequence were mostly changes of isolated aa or stretches of 2-4 aa, usually conservative, and usually identical within the groups. They suggested a group divergence of shape or conformation between the small and medium-large S-layer subunits and were likely to reflect the adaptation of the transporters to the group differences. The non-conservative substitutions were scattered and mostly isolated, being probably required to accommodate the three-dimensional characteristics of the individual S-layer subunits.

MFPs showed less sequence conservation than the ABC transporters, though still significant with an overall of 70-75% identity and 80-85% similarity at the protein level (Figure 3-3). The clustering of FWCs in two groups and the sequence variations were similar to the ABC transporters, with the remark that the changes in sequence are less conservative and more frequent, without extending over more than 4-5 aa.

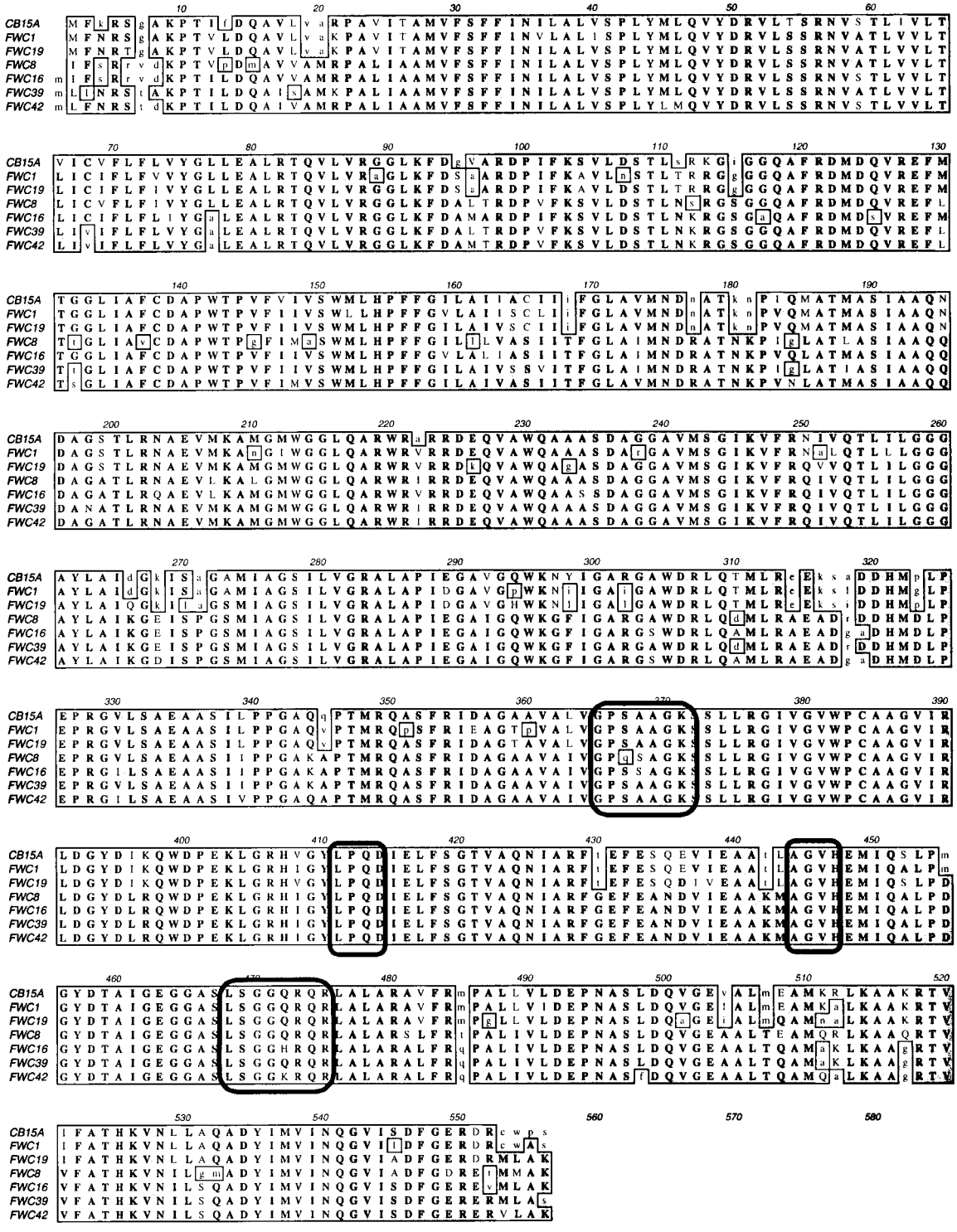

Figure 3-2. Amino acid sequence alignment of RsaD and FWC ABC transporters (dark gray - identity; light gray – similarity). The boxes indicate, in order, the Walker A, Walker B, the fourth motif and the ABC signature (Linton and Higgins, 1998)

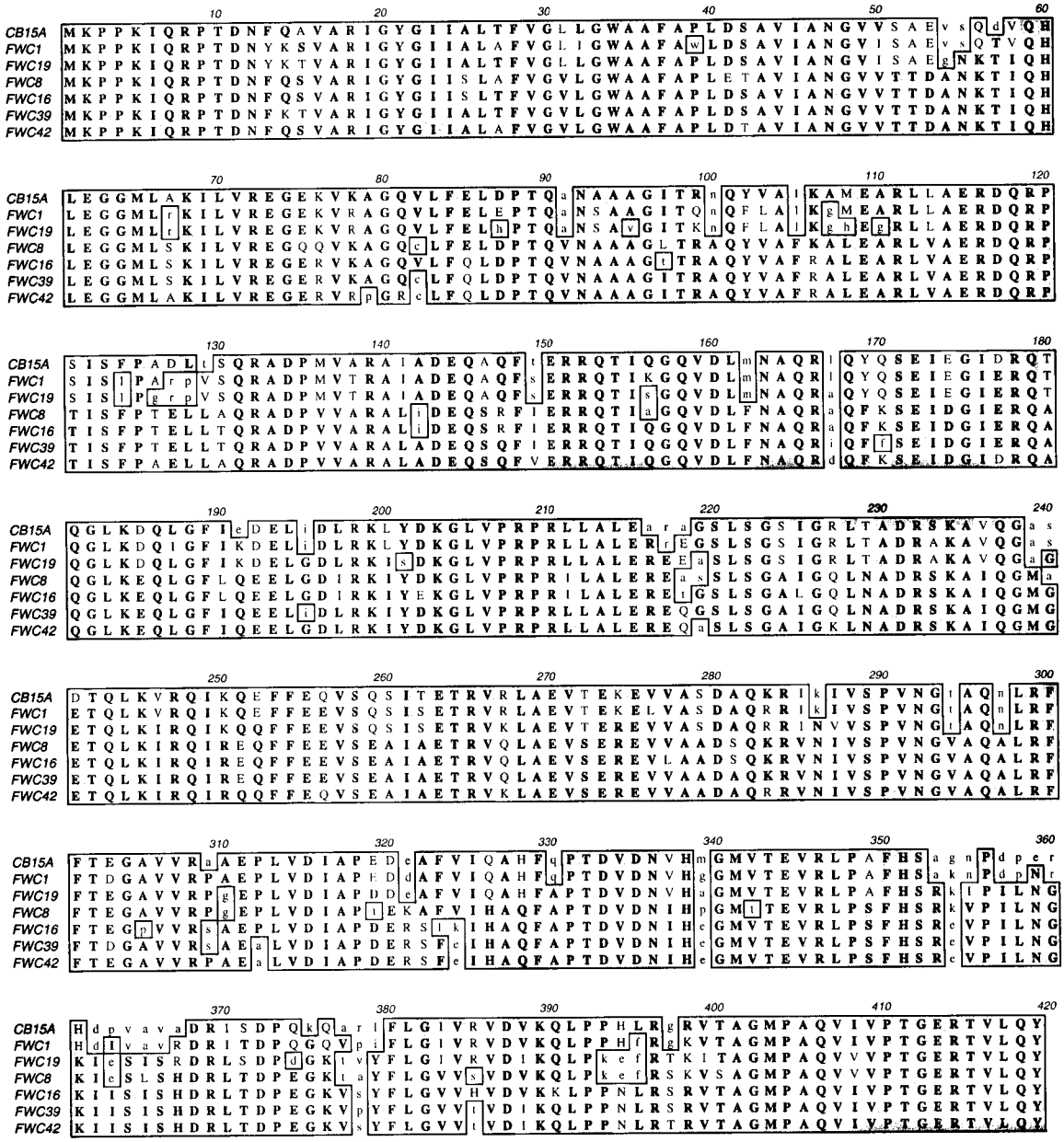

Figure 3-3. Amino acid sequence alignment of RsaE and FWC MFPs (amino acids 1-419 of each protein were aligned; dark gray - identity; light gray – similarity).

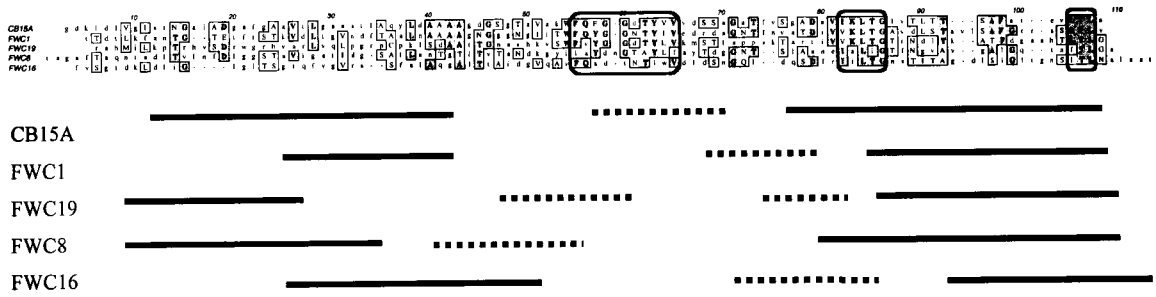

Figure 3-4. Amino acid sequence alignment of the last 100 aa of RsaA and FWC S-layer proteins (dark grey – identity; light grey – similarity). Predicted secondary structure features are shown below the alignment (solid lines - α-helices; dotted lines - β-strands). Boxes indicate conserved residues.

S-layer protein C-termini sequence comparisons showed a wider divergence than the transporter genes. There was a closer sequence similarity among the last 100 aa of the S-layer protein C-termini that were isolated (Figure 3-4, Table 3-3), whereas further upstream the sequence diverges drastically. This corroborated with the finding that the C-terminal secretion signal was localized within the last 82 amino acids of RsaA (Bingle et al., 2000). Again, the same division into two size groups was observed, with more primary sequence similarity within the same size group (Table 3-3).

Table 3-3. Sequence comparison of the last 100 amino acids of FWC S-layer C-termini

FWC strains compared	Sequence comparison of the last 100 aa of S-layer protein
C. crescentus, FWC 1, FWC19 (small size)	29% identical 44% similar
FWC8 vs. FWC16 (medium size)	23% identical 44% similar
FWC19 vs. FWC8	14% identical, 29% similar
FWC1 vs. FWC16	15% identical, 28% similar

The protein alignment of the last 100 amino acids of the FWC S-layer proteins (Figure 3-3) showed low primary sequence similarity except for several short stretches and isolated amino acids that are almost identical in all of the strains compared, such as (L/I)TLh at the extreme C-terminus, (V/I)KLTG at position –20 and FQY(G/D)GxTYhV at –45 (x – any residue, h – hydrophobic residue), with more divergence further

upstream. Of the primary structure motifs mentioned in Chapter 1, only a C-terminal hydroxylated residue followed by a hydrophobic tail is present in all strains studied. Secondary structure analysis of the C-termini sequenced predicted a helix-β-strand-helix motif in all strains, consistent with the proposal of such a conserved structural motif as a prerequisite for transport (Zhang et al., 1993; Zhang et al., 1998).

RTX repeats weren't found in any of the C-termini isolated, but that doesn't exclude their presence, since all C-termini were isolated by amplifying FWC genomic DNA with the RTXL primer, designed in the RTX-like region of RsaA. At least the region where the primer annealed is similar to an RTX nonapeptide, since the annealing temperature used for RTXL was 68°, only 4° below the calculated melting temperature. More RTX repeats may be located upstream of the isolated fragments.

In summary, the ABC transporters identified in FWC in the preliminary study were indeed part of a type I system, as revealed by the isolation and sequencing of adjacent genes encoding ABC transporters and MFPs, highly similar to those of *C. crescentus*. Moreover, in strains from the small and medium S-layer size group (FWC1, FWC19, FWC8, FWC16), the S-layer gene was located upstream of the ABC transporter gene, within 1.5 kb or less. The finding that the transporters were linked to the S-layer gene on the chromosome was an indication that they might secrete the S-layer subunits. Large S-layer genes were not linked to the transporter genes (not within ~6 to 8 kb upstream or downstream of the ABC gene, as shown in Chapter 5), and, despite intensive efforts, could not be isolated.

A sequence divergence was observed among the size groups, dividing the FWC strains into at least two groups. The small S-layer group had almost identical transporter genes and low similarity of secretion signals, and the medium-large S-layer group had very similar transporter genes and low similarity of secretion signals.

Sequence analysis of the FWC C-terminal secretion signals agreed with the division in size groups observed for transporter genes, and were consistent with the attempted definition of the secretion signal as a combination of conserved residues and predicted secondary structure features.

Chapter 4

Protein cross-expression and transporter mutants in FWC strains

Isolation and sequencing of FWC transporter genes, as described in Chapter 3, suggest that they form a type I system that secretes the S-layer subunits. It was previously shown that type I systems are capable of recognizing and secreting other type I secreted proteins with detectable efficiency (Binet and Wandersman, 1995). Hence, it would be expected that FWC transporters are able to recognize the RsaA secretion signal, given the sequence similarity among their secretion signals. *P. aeruginosa* alkaline protease, AprA was secreted in active form by the *C. crescentus* apparatus (Awram and Smit, 1998). Given the high sequence similarity between RsaD and E and the FWC transporters, the latter would be expected to also secrete AprA. However, recognition of the RsaA secretion signal and AprA secretion were not previously tested in FWC strains, so these matters were addressed in the present study. In addition, recognition of the secretion signal of one strain, FWC19, by the *C. crescentus* type I apparatus was examined.

Generally, not all type I substrate genes are linked to their transporter genes (Binet et al., 1997). Other transporters present on the chromosome might be in fact responsible for the secretion of the S-layer, even if the type I transporters identified in FWC strains generally have the same gene arrangement as *C. crescentus*. This matter was investigated by creating mutants of the isolated FWC transporter genes and checking if they still secreted the S-layer to the surface.

Results

Protein cross-expression in FWC strains.

Recognition of the C-terminal secretion signal of *C. crescentus*. FWC strains from the subset were all able to recognize the secretion signal contained within the last 336 aa of RsaA when introduced in the cells on a compatible plasmid (see Chapter 2). All strains secreted in the extracellular medium a polypeptide in the form of macroscopic aggregates, which had the correct size when analyzed by SDS-PAGE, and which reacted with the anti-RsaA C-terminus antibody (Figure 4-1).

The various strains produced equivalent amounts of protein, but the amounts used for SDS-PAGE and Western blotting (Figure 4-1) were not normalized, since this experiment was intended as qualitative rather than quantitative.

Recognition of the C-terminal secretion signal of FWC19 by *C. crescentus*. The S-layer defective *C. crescentus* mutant, CB2A, recognized the secretion signal contained within the last 216 aa of the FWC19 S-layer, and secreted a protein which formed small aggregates. The failure to form macroscopic aggregates was perhaps due to the absence of the Ca^{2+}-binding domain from the cloned C-terminus. SDS-PAGE analysis of the collected protein showed that the predominant fraction had the expected size (Figure 4-3). To produce a detectable reaction, the full-size S-layer of FWC19 requires a high concentration of anti-RsaA C-terminus primary antibody, because of the primary/secondary structure differences between the two S-layers. Consequently, Western blotting with the anti-RsaA C-terminus antibody was inconclusive, because at the high antibody concentration required, non-specific binding was predominant.

Secretion of *P. aeruginosa* AprA. All FWC strains from the subset secreted AprA, as shown by halo formation around the colonies (Awram and Smit, 1998; Figure 4-2). While in small S-layer producing strains (*C. crescentus*, FWC1, FWC19) all colonies on the plate secreted AprA, in strains that produce medium and large S-layers some colonies didn't have halos. This was attributed to the fairly high rate of spontaneous Sm resistance (the selective marker on the plasmid carrying *aprA)* mutations in caulobacters, as observed on negative control plates (water instead of plasmid DNA electroporated). The proportion of spontaneous mutants varied with each experiment, sometimes reaching 60%.

In conclusion, protein cross-expression studies showed that FWC strains were capable of recognizing and secreting type I secretion signals of both homologous (different strain) and heterelogous (different species) proteins. Moreover, the C-terminus of a FWC was recognized and secreted by *C. crescentus*.

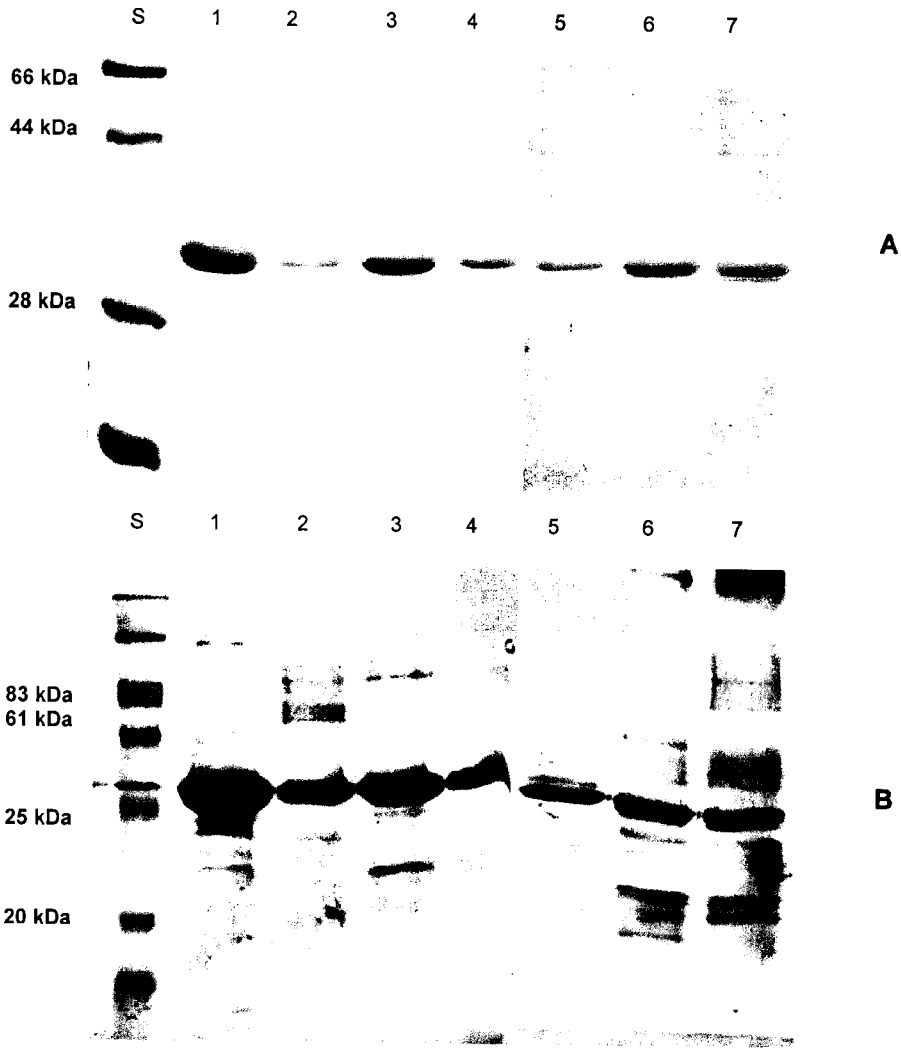

Figure 4-1. Secretion of the last 336 amino acids of RsaA by: 1. CB2A, 2. FWC8, 3. FWC39, 4. FWC42, 5. FWC1, 6. FWC19, 7. FWC16, S size standard. **A**. SDS-PAGE, Coomassie staining, **B**. Western blot with anti-RsaA C-terminus antibody.

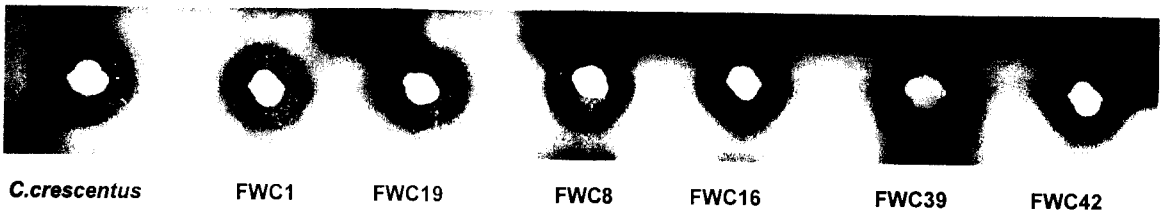

Figure 4-2. AprA secretion by FWC strains as shown by formation of a clear halo around colonies on skim milk PYE plates.

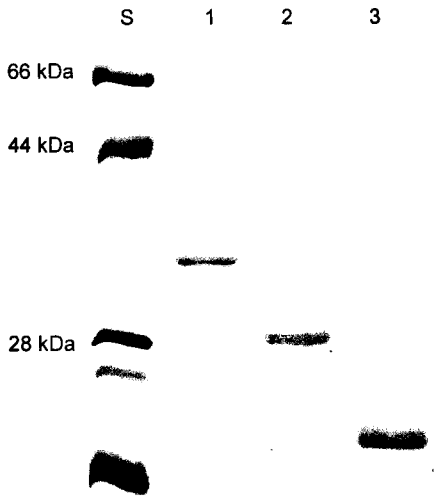

Figure 4-3. Secretion of FWC19 S-layer C-terminus by *C.crescentus* S-layer defective mutant. SDS-PAGE, Coomassie staining. Lanes are as follows: 1. last 336 aa of RsaA, 2. last 224 aa of RsaA, 3. FWC19 S-layer last 219 aa, **S** size standard

Inactivation of FWC transporter genes. One FWC strain from each size group: FWC19 (small), FWC16 (medium) and FWC39 (large), was subjected to chromosomal integration to inactivate the ABC transporters using pTZ19UΔSSm, pTZ18U(CHE) and respectively pTZ18U(Km), as described in Chapter 2. Mutant strains that were positive for integration (data not shown) and lacked the S-layer were isolated for all strains tested (Figure 4-4). The absence of S-layer on the surface of these transporter-inactivated strains was confirmed by EM (courtesy of John Smit).

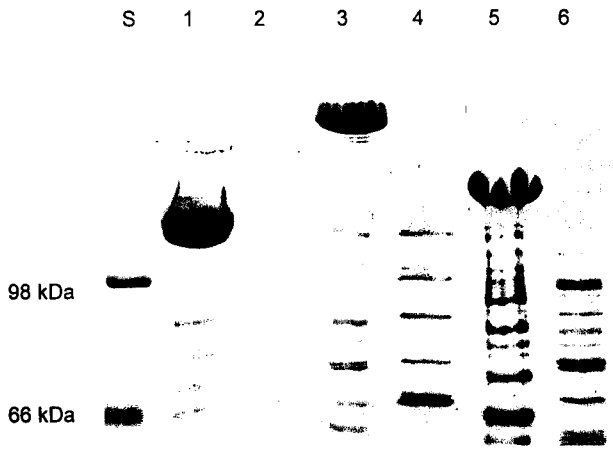

Figure 4-4. Inactivation of ABC transporters in FWC strains. SDS-PAGE, Coomassie staining of EGTA extracts. Lanes are as follows: 1. FWC19, 2. FWC19 ABC transporter-inactivated, 3. FWC39, 4. FWC39 ABC transporter-inactivated, 5. FWC16, 6. FWC16 ABC transporter-inactivated, S. size standard

Complementation of the ABC transporter defective strains was tried as described in Chapter 2 (transporter-inactivated FWC19, FWC16 and FWC39 were transformed with pRAT6, pRAT9 and pRAT6, respectively). Although colonies were obtained after

transformation for all the transporter-inactivated strains, none of them secreted an S-layer to the S-layer on the surface, as shown by EGTA extraction followed by SDS-PAGE (data not shown). Failure of complementation does not mean that the mutant strains were not as true transporter-inactivated mutants, because the lack of S-layer on their surface and generation of a PCR product of correct size is sufficient proof that integration occurred in the desired gene.

The successful creation of ABC transporter mutants that didn't secrete an S-layer on the surface in all size groups demonstrated that there was only one type I transport system that secretes the S-layer subunit, the one identified in Chapter 3. This correlated with the observation that Southern blotting with *rsaD* or *rsaE* probes showed the presence of only one copy of each transporter gene in the genome of all strains studied (data not shown).

Chapter 5

Organization of S-layer secretion genes in FWC strains

The gene arrangement of type I secreted proteins and transporters varies widely, as discussed in Chapter 1. Generally, the ABC transporter and MFP genes are adjacent, usually but not always linked to the gene for the substrate protein, and the OMP gene may or may not be linked to the other transporter genes (Binet et al, 1997). Cloning and sequencing of transporter genes and C-termini of FWC S-layer proteins (see Chapter 3) indicated that the ABC transporter and MFP genes were adjacent in all strains examined. The S-layer gene was adjacent to the transporter genes in the small S-layer group, located further upstream (as far as 1.6 kb in FWC16) in the medium S-layer group, and somewhere else on the genome in the large group. No information was available on the OMP gene position. Therefore, in order to define better the S-layer secretion gene organization in FWC strains, the location of the third component of the transporter apparatus, the OMP, and the position of the S-layer gene in the large size group were still required.

Results

Mapping of S-layer genes.

The small S-layer FWC group. As inferred from the cloning and sequencing data, the S-layer gene was located immediately upstream of the ABC transporter gene in both FWC strains examined from this group, with the same arrangement as C. crescentus.

The medium-sized S-layer group had the S-layer gene located further upstream of the ABC transporter gene, as shown by cloning and sequencing of the PCR products obtained with the RTXL-IRD pair of primers (see Chapter 3, Figure 3-1 and Table 3-2). The primers were designed as follows: RTXL in the RTX-like region of C. crescentus (C-terminus of RsaA, aa position 860), and IRD44 at the N-terminus of RsaD (aa position 44). FWC8 yielded an ~1.2 kb PCR product, which upon cloning and sequencing revealed sequence similarity to RsaA C-terminus at one end and RsaD at the other end. The product was bigger than the one obtained for C. crescentus (800 bp), with the S-layer and ABC transporter genes further apart, 500 bp versus 250 bp. FWC16 yielded

an ~2.2 kb product, with sequence similarity to *rsaA*, respectively *rsaD* at the extremities. This product was significantly bigger than the C. crescentus one, with the genes separated by 1.6 kb. BLAST analysis (see Chapter 2) of the spacing DNA in FWC8 and FWC16 showed only a few database hits over portions shorter than 100 bp. A few open reading frames (ORF), each shorter than 150 bp were identified (data not shown), and none of them superimposed with the database hits. Consequently, it was concluded that the S-layer gene and the ABC transporter gene were not separated by any genes. Genomic DNA of FWC27, FWC29 and FWC32 (all from the medium size group) was amplified by PCR with the RTXL-IRD44 primers, using the same conditions as for FWC8 and FWC16, in order to examine the location of their S-layer gene and establish a pattern for the medium size group. PCR products of 1.7 kb, 1 kb and respectively 1 kb were obtained. The FWC27 ~1.7 kb PCR product was cloned and sequenced (see Figure 3-1 and Table 3-2), and it showed sequence similarity to *rsaA* at one end and *rsaD* at the other. Hence, we proposed that in the medium size group, the S-layer gene is located upstream of the ABC transporter gene, at a distance of 500 bp to 1.6 kb.

The large S-layer group. Preliminary sequencing data revealed that ~ 1kb of DNA upstream of the ABC transporter gene in FWC39 and FWC42 did not have sequence similarity to *rsaA*, but rather to the gene for a *Sinorhizobium meliloti* galactosyl transferase, *expC* (Becker et al., 1997), suggesting that the S-layer subunit gene of the high MW group was located elsewhere than immediately upstream of the transporter genes. Also, amplification of genomic DNA with the RTXL-IRD44 primers (see Table 2) yielded no PCR product.

To identify the location of the S-layer genes, genomic DNA of FWC39 and FWC42 was subjected to Southern blotting with probes derived from *rsaD* and *rsaA*, following the mapping procedure described in Chapter 2 (Figure 5-1). Strong hybridization signals were obtained with the *rsaD* probe at high stringency, as expected from the high similarity of FWC39 and FWC42 ABC transporter genes to *rsaD*. The *rsaA* derived probe generated only weak, but detectable signals at low stringency washes. This

indicated that large S-layer genes have low sequence similarity to *rsaA*, yet enough to yield distinct bands with a *rsaA* probe under low stringency conditions.

The enzymes *EcoR*I and *Hind*III were chosen to digest the genomic DNA because they produced fragments larger than ~6.5 kb that reacted with the *rsaD* probe, and because of the position of their restriction sites within the ABC transporter gene. From the restriction maps available for FWC39 and FWC42, it was deduced that the *EcoR*I-*EcoR*I fragments contained ~1.4 kb, respectively 1.2 kb of the ABC transporter gene at the 5' end (Figure 5-2). The *Hind*III-*Hind*III fragments had the ABC transporter gene (which has a size of 1.6 kb) at ~1 kb, respectively 1.5 kb from the 3' end (Figure 5-2).

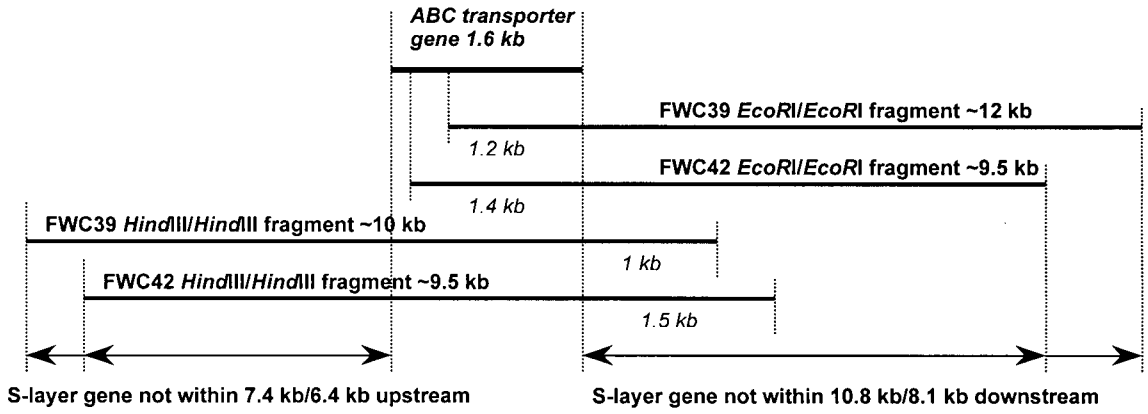

Figure 5-1. **Diagram of the FWC39 and FWC42 restriction fragments that reacted to the *rsaD* probe**. The relative position of the ABC transporter genes (sizes in italic type) on the *EcoR*I/*EcoR*I and *Hind*III/*Hind*III restriction fragments (bold type), derived from restriction maps, is depicted. The arrows at the bottom show the DNA stretch, upstream and downstream of the ABC transporter genes, whereon the S-layer gene was not situated (slash separates sizes for FWC39, respectively FWC42).

None of the fragments that reacted to the *rsaD* probe was identical to those that reacted to the *rsaA* probe, as shown in Figure 5-2. From the relative position of the ABC transporter gene on the restriction fragments, it was calculated that the S-layer gene of FWC39/FWC42 was not located within 7.4 kb/6.4 kb upstream, or 10.8 kb/8.1 kb downstream, respectively (Figure 5-2). Hence, the S-layer gene was not linked to the transporter genes, distinguishing the group of large S-layer strains from the others.

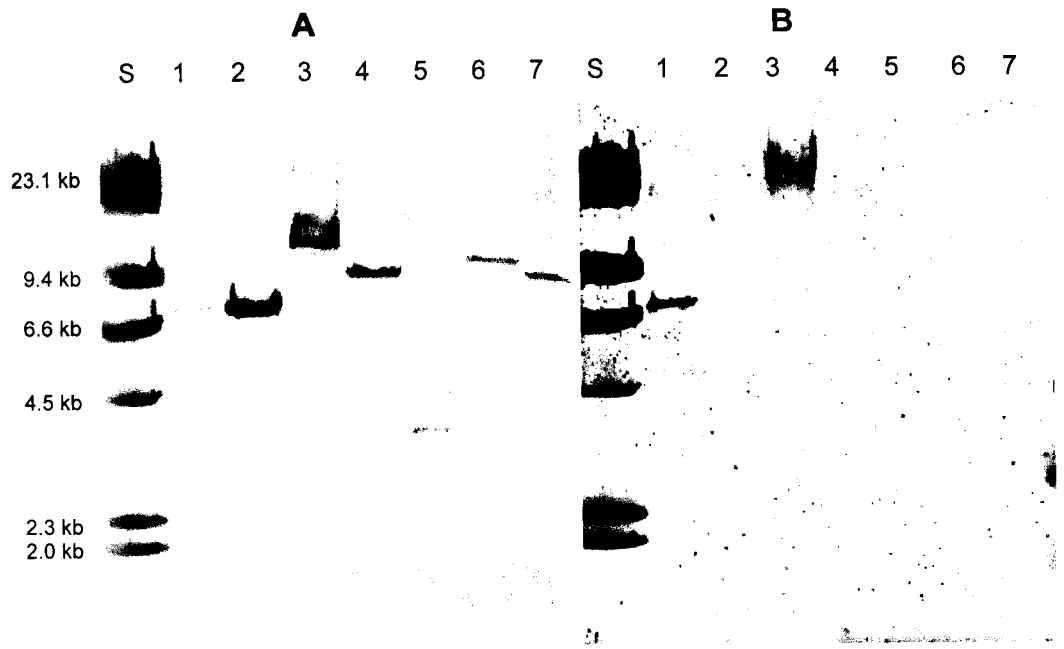

Figure 5-2. Large S-layer gene mapping. Southern blots of FWC genomic DNA probed with A. rsaD and B. rsaA. S HindIII digested λ-DNA, 1. *C.crescentus* HindIII, 2-4 FWC39 digested with BamHI, EcoRI, HindIII, 5-7 FWC42 digested with BamHI, EcoRI, HindIII. rsaD-reactive fragments are different from the ones that react with rsaA.

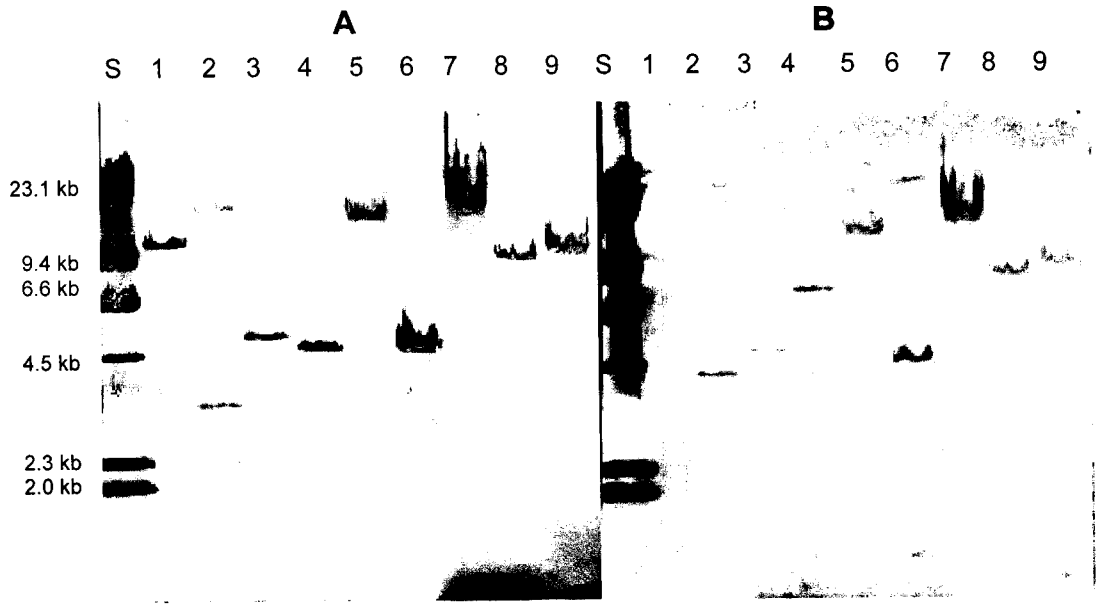

Figure 5-3. OMP gene mapping. Southern blots of FWC genomic DNA probed with A. rsaE and B. rsaF. Lanes are as follows (genomic DNA/restriction enzyme): S. size standard - λ-DNA/HindIII, 1. FWC1/HindIII, 2-3. FWC19/EcoRV, /SmaI, 4-5. FWC8/BamHI, /EcoRI, 6-7. FWC16 /BamHI, /EcoRI, 8. FWC39/HindIII, 9. FWC42/EcoRI. Fragments in lanes 1,3,5, 6,7,8,9 are identical in both A. and B.

The OMP gene. To identify the location of the OMP genes, genomic DNA of FWC strains was subjected to Southern hybridization with probes derived from *rsaE* and *rsaF*, following the mapping procedure described in Chapter 2 (Figure 5-3). Strong hybridization signals were obtained with the *rsaF* probe after high stringency washes, suggesting high similarity of FWC OMP genes to *rsaF*.

The OMP gene was found to be located downstream (as deduced from restriction maps of MFP genes, data not shown) of the MFP gene within as close as 5 kb (FWC16) and as far as 20 kb (FWC8) (Figure 5-3). There is no correlation between the S-layer size and the distance of the OMP from the MFP gene. In conclusion, the OMP gene of FWC strains had the same disposition as in *C. crescentus*, downstream of the OMP gene.

Chapter 6
Conclusions and Discussion

The purpose of this thesis was to answer several questions about the large collection of caulobacter strains available in our laboratory. Early studies showed that most of them secreted an S-layer on the surface (Walker et al., 1992), but they have not been studied since. The first question to ask about 40 related strains that secrete S-layers so varying in size (ranging from 100 kDa to 193 kDa) was whether they all did so by using a type I secretion mechanism, like *C. crescentus*. Other questions we addressed was how similar were their transporter genes, and what changes at the sequence level were required to accommodate the differently sized S-layer subunits of various FWC strains. Since none of the type I systems reported in the literature were found in a large number of related strains, the availability of 40 S-layer-secreting related FWC strains was a unique opportunity to compare their secretion signals, and their ability to recognize and secrete homologous or heterologous type I proteins. We also wanted to investigate the likelihood of a lock-and-key correspondence between the secretion signal and the transporters at the sequence level. If such a correlation existed, it would be possible to predict which C-terminal signal would be recognized by the transporters of which FWC strain, just by analyzing their sequence.

Conclusions

The major conclusions derived from the results of this study are summarized below, and then discussed in further detail.

Cloning and sequencing (Chapter 3). The sequence data proved that all FWC strains studied contained adjacent genes encoding ABC transporters and MFPs, highly similar to those of *C. crescentus*. Comparison of the S-layer C-termini aa sequence showed low sequence similarity for the small and medium size groups, with the presence of conserved residues and predicted secondary structure features, typical for type I secretion signals.

Protein cross-expression (Chapter 4) data showed that the FWC strains were capable of recognizing secretion signals from *C. crescentus* and one other type I secreted

protein (*P. aeruginosa* AprA), and that *C. crescentus* secreted the S-layer protein C-terminus of one FWC strain, FWC19. This suggested the presence of functional type I transporters in the FWC strains, and that the S-layer subunits are secreted by a type I mechanism.

Inactivation of FWC transporter genes (Chapter 4). The successful creation, in all size groups, of ABC transporter mutants that no longer secrete an S-layer demonstrated that the identified type I transport system was the one that secreted the S-layer subunit. All these data show that the studied FWC strains secrete their S-layer by a type I secretion system that is proprietary too the S-layer protein.

Organization of S-layer secretion genes in FWC strains (Chapter 5). The ABC transporter and MFP genes were adjacent in all strains examined, as in *C. crescentus*. The S-layer gene was adjacent to the transporter genes in the small S-layer group, located further upstream in the medium S-layer group, and elsewhere on the genome in the large S-layer group. In all FWC strains studied, the OMP gene had the same disposition as in *C. crescentus*, downstream of the OMP gene.

Discussion and future considerations

As much of a resource as the FWC strain collection of our laboratory might be, it was very challenging to study and compare a large number of strains, especially when the primary basis of comparison was the sequencing data. We only examined two strains from each size group, but the total length of DNA sequenced amounted to roughly 35 kb, making this study a considerable project. The conclusions inferred for the subset of FWC strains could only tentatively be extrapolated to other members of the same size group, although they were randomly selected from within that group. The transition between the different groups is not a steep one, but rather a gradual one, given that the delineation of the size groups was based on estimated rather than measured protein size study (Walker et al., 1992). Hence, a given FWC strain that was ascribed to a certain size group might, upon detailed analysis, behave as belonging to a different size group, and there also might be a more-or-less unbroken continuum of S-layer sizes if more strains were studied.

Transporter protein sequence comparisons revealed almost identical ABC transporters and highly similar MFPs (see Chapter 3). The lesser degree of conservation among MFPs suggests that they are influenced more than the ABC transporters by the size/shape/conformation of their substrate. Based on transporter sequence variations, FWC strains can be divided into two groups: **small S-layer** strains (FWC1, FWC19 and *C. crescentus*) and **medium-large S-layer** strains (FWC8, FWC16, FWC39, FWC42). The variations in protein sequence were mostly changes of single aa or stretches of 2-4 aa, usually conservative, and usually identical within the groups. This suggest a group divergence of shape or conformation between the small and medium-large S-layer subunits, perhaps reflecting the adaptation of the transporters to the group differences. The non-conservative substitutions were scattered and mostly isolated, being perhaps required to accommodate the three-dimensional characteristics of the individual S-layer subunits.

S-layer secretion signal analysis. Sequence comparisons of the available FWC S-layer C-termini showed a much wider divergence than observed for the transporter genes. In fact, only the last 100 amino acids had >30% sequence similarity; further upstream the sequence diverged drastically. The low sequence conservation among the C-termini was not unexpected, since comparison of secretion signals of various type I secreted proteins showed no clear consensus sequence, but rather a significant lack of sequence similarity (Koronakis et al., 1989). Indeed, the RsaA C-terminus has only moderate sequence similarity to C-termini of type I secreted proteins from the closest related systems (such as *C. fetus* S-layer subunit - 24% identity, 40% similarity; *E. chrysantemi* protease PrtB - 32% identity; *S. marcescens* protease PrtA - 30.5% identity)(Gilchrist et al, 1992). Studies of type I C-terminal secretion signals proposed that recognition of the signal by the transporter complex rely on both conserved residues and conserved secondary structure features (Ghigo and Wandersman, 1994). A number of elements were proposed as being important for type I signal recognition. Putative primary structure elements for RTX toxins were: one or several hydroxylated amino acids (S or T) surrounded by hydrophobic residues within the last 10 amino acids, and a cluster of charged residues (Stanley et al., 1991), and for proteases, an

acidic residue (D or E) followed by a hydrophobic tail at the extreme C-terminus (Ghigo and Wandersman, 1994). The proposed secondary structure feature was the helix-linker-helix motif, with the linker usually being a β-strand (Zhang et al., 1993; Zhang et al., 1998). The importance of these candidate elements was supported by mutational analyses (Stanley et al., 1991; Chervaux et Holland, 1995; Hui et al., 2000). No single type I secreted protein, however, contains all components.

A study of *C. fetus* S-layer protein homologs (Thompson et al., 1998) compared the C-termini of four *C. fetus* Sap homologs and the C-terminus of RsaA. Alignment of the last 70 aa of each protein showed little primary sequence similarity except for a few conserved residues, such as GDGS(T/G), GxTYV, V(V/I)D, DxxIKLxG (Thompson et al., 1998). A toxin-specific motif (1-4 hydroxylated residues within the last 10 aa) and a somewhat altered protease-specific motif (the extreme C-terminus acidic residue followed by several hydrophobic aa and one or two hydroxylated residues) were present all strains. Secondary structure analysis predicted a helix-linker-helix motif in all proteins examined, with the linker being usually a β-sheet (Thompson et al., 1998). These data showed that the secretion signals of *C.fetus* Sap homologs and RsaA contained elements that are specific to both families of type I secreted proteins. In fact, their extreme C-termini had a novel motif that was a combination of the toxin- and protease-specific motifs: an acidic residue followed by several hydrophobic and 1-2 hydroxylated residues. Hence, a classification of type I secreted S-layers with one of the two existing families, the proteases and or the RTX-toxins, was not possible based on this analysis.

We tried to extend this analysis by comparing the C-termini available for FWC strains, and those of *S.marcescens* S-layer protein, SlaA (Kawai et al., 1998), and two *C.fetus* Sap homologs, SapA and SapA1.

The protein alignment of the last 100 aa of FWC S-layers and Rsa (Figure 3-3) showed little sequence conservation except for several residues that were conserved in almost all of the strains compared, such as (L/I)TLh at the extreme C-terminus, (V/I)KLTG at position –20 and FQY(G/D)GxTYhV at –45 from the C-terminus (x – any residue, h – hydrophobic residue).

Of the primary structure motifs, only one RTX-toxin motif, the C-terminal 2-3 hydroxylated residues surrounded by hydrophobic aa, was present in all strains studied.

The modified protease motif mentioned before was only found in *C. crescentus* and FWC19.

Computer secondary structure analysis predicted a helix-β-strand-helix motif within the last 100 aa of all S-layer C-termini sequenced, consistent with the proposal of such a conserved structural motif as a requisite for transport (Zhang et al., 1993; Zhang et al., 1998).

The residues conserved among the FWC S-layer C-termini were also present in the *C.fetus* Sap homologs and *S.marcescens* SlaA C-termini, as shown in the protein alignment in Figure 6-1. The C-terminal hydroxylated residues (toxin motif) were encountered in all strains, whereas the altered protease motif was found in all but three FWC strains (FWC1, FWC8, FWC16). All C-termini had a predicted helix-beta-helix motif. In conclusion, the type I secreted S-layers could not be assigned to any of the type I protein families, since their secretion signals contained elements that belonged to both protease and toxin families. Rather, the presence of the novel motif at the extreme C-termini, and of residues conserved among all S-layers analyzed, argue for the existence of a third family of type I secreted proteins, the S-layer family.

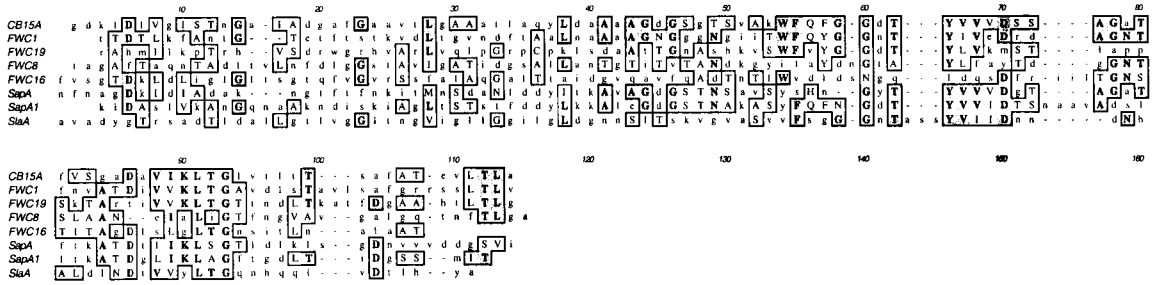

Figure 6-1. Amino acid sequence alignment of the last 100 amino acids of RsaA, FWC S-layer proteins, *C. fetus* Sap homologs and *S. marcescens* SlaA (dark grey– identity; light grey – similarity). The acidic residues of the novel motif can be observed within the last 10 aa (conserved D residues in FWC16, SapA, SapA1, SlaA and an E in RsaA). The hydroxylated residues (S or T) surrounded by hydrophobic aa can also be seen within the last 15 aa of each strain.

Any attempts at defining a lock-and-key correlation between the primary sequence of an FWC S-layer C-terminus and the cognate transporters would be purely speculative,

because no accurate correspondence could be found based only on the available sequence data. The lesser sequence similarity of the C-terminus of medium-sized S-layer proteins to RsaA correlated with the protein sequence divergence of the transporters of medium and large-sized S-layer strains from those of small S-layer strains. One could speculate then that if a given FWC C-terminal signal had less sequence similarity to the C-terminus of RsaA, it would be likely to be recognized by the transporter apparatus of a medium or large-sized S-layer FWC, rather than a small one. In order to establish a precise correlation, a detailed mutational analysis of the C-termini combined with a three dimensional study of the secretion signal and transporter conformation would perhaps be required.

The search for the large S-layer gene in FWC39 and FWC42 was unsuccessful for several reasons. Firstly, the low sequence similarity between large and small S-layer genes rendered inapplicable a library screening with probes derived from *rsaA* or FWC S-layer genes. Presumably for the same reason, PCR amplification of genomic DNA with *rsaA* based primers also failed. Secondly, because the large S-layer genes were not linked to the transporter genes, inverse PCR or the chromosomal integration strategy could not be used.

One method that would lead to the isolation of the large S-layer genes is Tn5 mutagenesis of a large S-layer FWC strain, followed by screening of the library for loss-of-function mutants. This could be done by Western colony immunoblotting with either an anti-RsaA antibody, which would show the absence of the S-layer from the surface, or with an anti-SLPS antibody, which would detect the exposed underlying SLPS. The colonies negative for the anti-RsaA and/or positive for the anti-SLPS antibody reaction either would be defective in the S-layer secretion, or would have an interrupted S-layer gene. The S-layer secretion defective mutants could have any of the ABC, MFP or OMP genes interrupted, hence Tn5 insertions in at least 4 genes (including the S-layer gene) would lead to the same phenotype. This would require the screening and subsequent analysis of numerous isolates in order to identify the insertions in the S-layer gene, and would make such a method a project in itself, exceeding the scope of the current investigation.

Protein cross-expression. The type I transporters of all FWC strains tested were capable of recognizing and secreting the secretion signal of RsaA, but secretion of larger proteins was not tested. The large S-layer FWC strains have already the intrinsic ability to secrete the large subunits to the surface, therefore it would be of general interest to investigate whether they can secrete large proteins. This could have useful biotechnological applications, such as the utilization of large S-layer strains for the expression of recombinant proteins too large for the *C. crescentus* type I apparatus to secrete. In addition, quantification of the amount of certain polypeptides secreted by FWC strains may lead to the identification of a strain with higher secretion efficiency than *C. crescentus.*

The C-terminus of the FWC19 S-layer was secreted by *C. crescentus*, but it failed to form macroscopic aggregates. A possible cause was perhaps the absence of the Ca^{2+}-binding domain from the cloned C-terminus. In order to locate the Ca^{2+}-binding domain on the C-terminus, cloning of the full-length S-layer gene followed by studies of N-terminal deletions expression would be required. Examination of the capability of a small S-layer FWC strain, even *C. crescentus*, to secrete a large S-layer, would shed more light upon the significance of the transporter sequence variations among size groups, but this would require the cloning of a large S-layer gene.

Because all FWC strains tested secreted AprA in active form, despite the potential competition for the transporters with the S-layer subunits, the FWC strains might be grouped with the protease family of type I secreted proteins, as suggested for *C. crescentus* (Awram and Smit, 1998). On the other hand, comparison of the secretion signals argued for the classification of type I secreted S-layer as an independent family. Studies the ability of FWC strains to secrete an RTX toxin would further clarify the grouping of S-layers within the type I secreted proteins.

Transporter gene mutants. Only one FWC strain from each size group was transporter-inactivated, as it was assumed that other strains from the same size group would behave similarly, due to the high sequence similarity of the transporters. Creation of an ABC transporter mutant in FWC39 (a large S-layer strain) that didn't secrete an S-

layer on the surface was essential in order to minimally demonstrate the very fact that the large S-layer proteins were secreted by a type I system, since none of the large S-layer genes was isolated.

All transporter-inactivated strains grew poorly under antibiotic selection, perhaps because there is only one copy of the antibiotic resistance gene per cell, instead of the multiple copies carried by plasmids. Yet another reason might be that the S-layer subunits produced could not be secreted and therefore they had to be degraded intracellularly, a useless energy expenditure which negatively impacts the cell growth and division.

The unsuccessful complementation of the ABC transporter-inactivated strains by the *rsaD* and *rsaE* genes does not mean that they were not true mutants. One possible reason that might have caused the failure would be that the *C. crescentus* transporter genes were transcribed inefficiently due to the absence of the native promoter on the plasmid carrying the transporter genes.

Organization of S-layer secretion genes. Generally for type I systems, the ABC transporter and MFP genes are adjacent and usually linked to the gene for the substrate protein (Binet et al., 1997), although exceptions have been noted (the S-layer protein gene of *S. marcescens*, *slaA*; Kawai et al, 1998). The S-layer genes of the FWC small and medium size groups were linked to the transporter genes, whereas the large S-layer genes were not. Future cloning of the large S-layer genes might reveal their location relative to the transporter genes.

Interestingly, the S-layer gene location is correlated with the sequence divergence among size groups, dividing the S-layer secreting caulobacters into at least two groups. The small S-layer group has the same arrangement of S-layer secretion genes, almost identical transporter genes and low similarity of secretion signals. The medium-large S-layer group has highly similar transporter genes and low similarity of secretion signals. The S-layer gene location further divides this group into subgroups: the medium-sized S-layer group (where the S-layer gene is as far as 1.6 kb upstream of the ABC transporter gene), and the large S-layer group (where the S-layer gene is somewhere else in the genome).

The OMP gene is either linked to the transporter genes, or located elsewhere on the genome, as shown in Table 1-1, Chapter 1. For instance, in *C. crescentus*, rsaF is not immediately linked to *rsaD* and *E* (Awram and Smit, 2000), whereas in *C.fetus* the OMP gene is linked to the other transporter genes (Thompson et al., 1998). The RsaF protein was only recently characterized in *C. crescentus* (Awram and Smit, 2000), so a detailed analysis of the OMP gene in FWC strains was not attempted. Instead, its position on the chromosome, relative to the other two transporter genes, was examined. The gene for the OMP was not directly linked to the other transporter genes, but was located within 15 kb downstream of the MFP genes in all FWC strains studied. Further analysis, such as cloning and sequencing of the OMP genes would provide valuable information about these least studied components of the type I systems.

Summary

FWC strains secrete their S-layer proteins by type I mechanisms. They contain type I transporters that are highly conserved and are proprietary to the S-layer proteins. The organization of their S-layer secretion genes was also examined.

The C-termini of several FWC S-layer proteins were analyzed and they all contain conserved residues and predicted secondary structures that are candidates for secretion signals.

FWC strains were capable of secreting homologous and heterologous type I secreted proteins.

FWC strains can be divided in two groups based on transporter sequence comparisons: small, and medium-large S-layer-secreting strains. S-layer gene position further divides the latter group into medium and large S-layer-secreting FWC strains.

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.

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*. *Science*, 278, 1140-1143.

Awram, P. and Smit, J. (2000) Identification of lipopolysaccharide O-antigen synthesis genes required for attachment of the S-layer of *Caulobacter crescentus*. Manuscript in preparation.

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.

Becker A., Ruberg S., Kuster H., Roxlau A.A., Keller M., Ivashina T., Cheng H.P., Walker G.C., Puhler A. (1997) The 32 kilobase exp gene cluster of Rhizobium meliloti directing the biosynthesis of galactoglucan: genetic organization and properties of the encoded gene products. *J. Bacteriol.* 179, 1375-1384

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.

Bingle, W.H.,.Nomellini, J.F, and Smit, J. (2000) Secretion of the *Caulobacter crescentus* S-layer protein: further localization of the C-terminal secretion signal and its use for secretion of recombinant proteins. *J. Bact.* submitted.

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.

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.

Boehm, D.F., Welch, R.A., and Snyder, I.S. (1990) Domeins of *Escherichia coli* hemolysin (HlyA) involved in binding of calcium and erythrocyte membranes. *Infect. Immun.* 58, 1959-64

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

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.

Chervaux, C. and Holland, I.B. (1995) Random and directed mutagenesis to elucidate the functional importance of helix II and F-989 in the C-terminal secretion signal of *Escherichia coli* hemolysin. *J. Bact.*, 178, 1232-36

Christie P.J. and Vogel J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.*, 8, 354-60.

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

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.

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.

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.

Dworkin, J., and Blaser, J. (1996) Generation of *Campylobacter fetus* S-layer protein diversity utilizes a single promoter on an invertible DNA segment. *Mol. Microbiol.* 19, 1241-1253

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.

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.

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

Ghigo, J.M. and Wandersman, C. (1994) A carboxyl-terminal four amino acid motif is required for secretion of the metalloprotease PrtG through the *Erwinia chrysantemi* protease secretion pathway. *J. Biol. Chem.* 269, 8979-85

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.

Higgins, C. (1992) ABC Transporters: from microorganisms to man. *Annu. Rev. Cell Biol.*, 8, 67-113

Hueck, C.J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev.*, 62, 379-433.

Hui, D., Morden, C., Zhang, F. and Ling V. (2000) Combinatorial analysis of the structural requirements of the *Escherichia coli* hemolysin signal sequence. *J. Biol. Chem.*, 275, 2713-20

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.

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.

Klauser, T., Pohlner, J. and Meyer, T.F. (1993) The secretion pathway of IgA protease-type proteins in gram-negative bacteria. *Bioessays*, 15, 799-805.

Koronakis V., Koronakis E., Hughes C. (1989) Isolation and analysis of the C-terminal signal directing export of Escherichia coli hemolysin protein across both bacterial membranes. *EMBO J.* 8,595-605.

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.

Koronakis V., Sharff A., Koronakis E., Luisi B., Hughes C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*, 405, 914-9.

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

Kuhnert, P., Heyberger-Meyer, B., Burnens, A.P., Nicolet, J. and Frey, J. (1997) Detection of RTX toxin genes in gram-negative bacteria with a set of specific probes. *Appl. Environ. Microbiol.* 63, 2258-2265

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. 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.

Mackman N., Baker K., Gray L., Haigh R., Nicaud, J.M. and Holland I.B. (1987) Release of a chimeric protein into the medium from *Escherichia coli* using the C-terminal secretion signal of haemolysin. *EMBO J.* 6, 2835-41

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

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.

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.

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.

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.

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 biomimetics 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.

Sara, M. and Sleytr, U.B. (2000) S-Layer proteins. *J. Bacteriol.* 182, 859-68.

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

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

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. (1997) Advances in S-layer nanotechnology and biomimetics. *Adv. Biophys.*, 34, 71-79.

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. (1986) Protein surface layers of bacteria, p.343-376 Bacterial outer membranes as model systems, John Wiley & Sons, Inc., N.Y.

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.

Staley, J.T., Konopka, A.E. and Dalmasso, J.P. (1987) Spatial and temporal distribution of *Caulobacter* spp. in two mesotrophic lakes. *FEMS Microbiol Lett*, 25, 1-6

Stanley, P., Koronakis, V. and Hughes, C. (1991) Mutational analysis supports a role for multiple structural features in the C-terminal secretion signal of *Escherichia coli* hemolysin. *Mol.Microbiol.* 5, 2391-2403

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.

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.

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

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.

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., Greig, D.I. and Ling V. (1993) Functional replacement of the hemolysin A transport signal by a different primary sequence. *Proc. Natl. Acad. Sci.*, 90, 4211-15

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