Characterisation of O-antigen biosynthesis genes in *Vibrio anguillarum* and their association with IS*1358*

By

Kathy Eva Daniels (nee Jedani)
B.Sc Hons. Dip. Ed. (Adelaide)

A thesis submitted for the degree of
Doctor of Philosophy

Department of Microbiology and Immunology
University of Adelaide

July, 1999

ABSTRACT

Vibrio anguillarum is an important pathogen that causes the disease vibriosis in feral and cultured fish. To date 23 serotypes of *V. anguillarum* have been identified. Strains from the O1 and O2 serotypes are the most frequent serotypes isolated from disease outbreaks. The pathogenesis of *V. anguillarum* is not well understood nor are the virulence determinants involved in the process of infection. The best characterised virulence determinant is the iron transport system located on the large plasmid, pJM1. More recent studies have focussed on characterising other factors involved in the pathogenic process including: the haemolysin, MOMP and flagella. LPS is an important virulence factor in numerous pathogenic bacteria and has been found to contribute to the virulence of *V. anguillarum*. However at present the genes involved in the biosynthesis of this virulence determinant have not been isolated, it was the objective of this thesis to characterise the genes required for the synthesis of O-antigen in *V. anguillarum* and to gain a better understanding of the mechanisms involved in the rearrangements of O-antigen biosynthesis gene clusters in *Vibrio* species.

Two different antisera, one raised against *V. anguillarum* O1 and the other against the O4 serotype were generated in this study to aid in the investigation of virulence determinants. The anti *V. anguillarum* O4 serum reacted only to a major outer membrane protein (MOMP) of ~36 kDa. MOMP varied in size between the 10 serotypes (O1-O10) tested. Purified MOMP from *V. anguillarum* O1 was N-terminal sequenced and found to be homologous to an immunogenic, highly variable protein P2, in *Haemophilus influenzae* and *H. sommus*. The *V. anguillarum* O1 serum contained antibodies to both the MOMP and LPS as determined by Western blotting. Affinity purification was used to obtain LPS

specific antibodies to further investigate by immunogold electron microscopy and immunofluorescence, LPS expression in *V. anguillarum* O1.

IS*1358*, a recently characterised insertion sequence, has been shown to be linked to the O-antigen biosynthesis regions in numerous *V. cholerae* serotypes. IS*1358* was found to be widely distributed in a number of *V. anguillarum* serotypes including O1 and O2. IS*1358* was found to be present in multiple copies in serotypes O2, O7 and O9 suggesting that this element is a site for recombination, gene duplication or that it may be capable of transposition. The IS*1358* elements from the different serotypes were cloned and sequenced. Analysis of the sequence revealed either the presence of a complete ORF in serotypes containing multiple copies of the element or three small ORFs in those strains that possess only one IS*1358* element. It was found by T7 promoter/polymerase expression assays that serotypes O2, O7 and O9 produce a protein of 42 kDa which corresponds to the predicted molecular weight.

IS*1358* was found to be associated with polysaccharide biosynthesis genes in *V. anguillarum* O1 and O2. In *V. anguillarum* O2, PCR was used to amplify DNA between copies of IS*1358*. These PCR products were cloned and sequenced. Analysis of the sequence indicated the presence of genes homologous to polysaccharide biosynthesis and regulation genes. A chromosomal map was generated showing the organisation of the amplified DNA with respect to the adjacent IS*1358* elements. Inverse PCR and cosmid cloning were used to further map and sequence this region.

In *V. anguillarum* O1, cosmid cloning was used to isolate the DNA flanking the single copy of IS*1358*. Cosmid clones were screened with a DIG-labelled IS*1358* probe, allowing the regions adjacent to be partially cloned and sequenced. Characterisation of the O-antigen biosynthesis region in *V. anguillarum* O1, designated *wbh* has resulted in construction of a physical chromosomal map with 24 open reading frames (ORFs) being identified. Most ORFs showed significant homology to other polysaccharide biosynthesis genes from numerous other bacteria. Genes involved in the biosynthesis of rhamnose were identified in the *wbh* region of *V. anguillarum* O1 based on homology. In addition genes homologous to nucleotide sugar transferases and components of homopolymer O-antigen transport systems were also detected. Transposon insertions using Tn*5phoA* (Cm^R) in the *wbh* region were characterised and shown to result in different O-antigen phenotypes as observed by silver staining of SDS-PAGE. The differences observed in the O-antigen profile of these mutants was dependent on the location of the mutation. Analysis of a mutant with an insertion outside of the sequenced *wbh* region in conjunction with sequencing and subcloning of cosmids downstream of the characterised *wbh* region suggests that additional genes may be required for expression of LPS in *V. anguillarum* O1. However these additional genes were not directly linked to the *wbh* region and Southern mapping indicated that it was located within 40 kb.

Sugar analysis of the LPS from the *V. anguillarum* O1 wildtype and transposon mutants indicates the presence of glucose, rhamnose and glucosamine in varying amounts. Genes present in the *wbh* region account for the rhamnose component of the LPS.

In this study the *wbh* region responsible for O-antigen biosynthesis was isolated and partially characterised. The operon appears to be made up of genes that were acquired from other bacteria. The presence of IS*1358* indicates that it may have played a role in the acquisition or rearrangement of the polysaccharide biosynthesis genes in *V. anguillarum* O1.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University or other tertiary institution, and to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited in the University library being available for loan and photocopying.

Kathy Eva Daniels

To my wonderful husband Craig,
and my Mum and Dad,
all of you will be forever in my heart.

Acknowledgments

"You never know where your going until you get there"

I must start this rather difficult task by thanking Professor Paul Manning for allowing me to undertake this study in his laboratory. Next, I would like to extend a special thank you to Dr. Uwe Stroeher for his friendship, supervision and support to the end.

I am eternally grateful to my "last minute" supervisor Dr. Renato Morona. I would not have finished this Ph.D without your help, knowledge and expertise- wish I could say more but you know how I feel! I hope you know that I have appreciated your extreme effort!

I would like to thank Dr. Debra Milton, who provided me with support from the other side of the world. I must also acknowledge her expertise in working with *V. anguillarum,* and thank her for the many strains, tips and pointers she has given during our acquaintance. I hope that one day we shall meet, so I can thank you in person. Many thanks!

During the years of my Ph.D I met and enjoyed the company of numerous people- both students and staff. To everybody in the lab both past and present, thanks for the past five years. I would like to make special mention of Sarah, a friend and gym partner (when I wasn't lazy); Liz, I thank you for all the small things you did that helped enormously and Luisa- thanks for your help when I really needed things to go right, especially with IFs. I must also thank Chris Cursaro for his photographic talents etc and the "computer boys" for their help at crucial times during the writing of this thesis. I would like to thank Carole Gannon for allowing me to find my true calling in life- you have provided me with an enjoyable outlet and I thank you very much, for that, and your friendship.

To my dear friends on the outside, Helen and Andrea, thanks for being there when I needed to complain, cry and be disappointed. You will be forever my friends and I wish you all the luck and happiness in the world- I will miss you heaps!!

To all of my family both new and old, I thank you for all understanding that we didn't see you as much as we would have liked-please know that you will never be far from our minds and are forever in our hearts. I would also like to thank you for all your support over the years (and years) of study that I have undertaken.

My final acknowledgment must go to my dearest friend, my husband, Craig- I would not have finished if it weren't for you. There are not enough words in all of the dictionaries to describe the love and gratitude I feel for you. I know that our life starts on a new path with the completion of our theses, and I can't wait to go down it together, and wherever it leads- starting in Canada!! I love you.

Abbreviations

A:	adenine
A_{260}:	absorbance at 260nm
aa:	amino acid
ACP:	acyl carrier protein
ACL:	antigen carrier lipid (bactoprenol)
Ap:	ampicillin
ATP:	adenosine-5'-triphosphate
bp:	base pair
BSA:	bovine serum albumin
C:	cytosine
CIP:	calf intestinal phosphatase
Cm:	chloramphenicol
CTP:	cytosine-5'-triphosphate
C-terminal:	carboxy terrminal
DIG:	digoxigenin
DNA:	deoxyribonucleic acid
DNase:	deoxyribonuclease
dNTP:	deoxyribonucleoside triphosphate
ddNTP:	dideoxyribonucleoside triphosphate
DTT:	dithiothreitol
ECA:	enterobacterial common antigen
EDTA:	ethylene-diamine-tetra-acetic acid
EtBr:	ethidium bromide
G:	guanine
Gm:	gentamycin
GTP:	guanine-5'-triphosphate

HRP:	horse radish peroxidase
IPTG:	isopropyl-β-D-thiogalactopyranoside
IS:	insertion sequence
kb:	kilobase pairs
kDa:	kilodalton
KDO:	keto-3-deoxy-D-manno-octulosonic acid
Km:	kanamycin
LA:	luria agar
LB:	luria broth
LPS:	lipopolysaccharide
LR-LPS:	long range PCR
mg:	milligram
ml:	millilitre
mM:	millimolar
mRNA:	messenger RNA
MOMP:	Major Outer Membrane Protein
MQ:	milli Q water
NA:	nutrient agar
NB:	nutrient broth
nt:	nucleotide
N-terminal:	amino terminal
OD:	optical density
OM:	outer membrane
ORF:	open reading frame
PAGE:	polyacrylamide gel electrophoresis
PCR:	polymerase chain reaction
PEG:	polyethylene glycol -8000
^{R}**:**	resistant
R-LPS:	rough LPS;

RBS:	ribosome binding site
rfb:	O-antigen biosynthesis genes
RNA:	riobonucleic acid
RNase:	ribonuclease
rpm:	revolutions per minute
RT:	room temperature
^{s}:	sensitive
SD:	Shine-Dalgarno
SDS:	sodium dodecyl sulphate
S-LPS:	smooth LPS;
SR-LPS:	semi-rough LPS;
Sm:	streptonycin
Sp:	spectinomycin
sv:	serovar
T:	thymine
TBS:	Tris-buffered saline
Tc:	tetracycline
TEMED:	N,N,N',N'-tetramethyl-ethylene-diamine
Tn:	transposon
Tris:	Tris (hydroxymethyl) aminomethane
TSA:	Tryptic Soy Agar
TSB:	Tryptic Soy Broth
TTBS:	Tris-buffered saline with Tween-20 added
TTP:	thymine-5'-triphosphate
ts:	temperature sensitive
U:	uracil
UV:	ultraviolet
V:	voltage
v/v:	volume per volume

wbh:	*V. anguillarum* O1 O-antigen biosynthesis locus
wc:	whole cells
w/v:	weight per volume
X-gal:	5-bromo-4-chloro-3-indolyl-β-galactopyranoside

TABLE OF CONTENTS

6. CHAPTER SIX: CHARACTERISATION OF THE WBH OPERON BY MUTATIONAL ANALYSIS 122

7. CHAPTER SEVEN: DISCUSSION 141

Chapter One

Introduction

1.1 Introduction

Vibrio anguillarum, a Gram-negative bacterium, belongs to the genus Vibrionaceae and is closely related to the human pathogen *V. cholerae* (Dorsch *et al.*, 1992; Stroeher *et al.*, 1994). *V. anguillarum* is the aetiological agent of vibriosis, a terminal haemorrhagic septicaemia in salmonid and marine fish species. Disease outbreaks have been recorded in both feral and cultured fish populations (Sørenson and Larson, 1986). A significant increase in bacterial infections due to the expansion of the aquaculture industry, coupled with overstocking and poor water quality, has resulted in economic loss.

1.1.1 Serotypes of *Vibrio anguillarum*

Sørenson and Larson (1986) developed a serotyping scheme for *V. anguillarum* based on heat stable somatic O-antigens. It was from this study that *V. anguillarum* was classified into 10 distinct serotypes. Serotypes O1, O2 and O3 cause the majority of infections in fish (Kitao *et al.*, 1983; Sørenson and Larson., 1986; Austin *et al.*,1995). Interestingly all three of the serotypes are found in the environment and are associated with disease outbreaks in cultured salmonids. However, the majority of isolates found in cultured salmonids belong to the O1 serotype while the O2 serotype predominates in diseased feral seawater fish (Sørenson and Larson, 1986). *Vibrio anguillarum* O2 has been subdivided into two serotypes designated O2a and O2b based on specific antigenic epitopes found in the lipopolysaccharide (LPS) (Rassmussen, 1987 a,b). The O3 serotype shows heterogeneity within the type strains and has therefore also been subdivided into two subclasses designated O3A and O3B (Santos *et al.*, 1995). Vibriosis has also been reported in freshwater fish (Mutharia *et al.*, 1993; Ohnishi and Muroga, 1977; Smith, 1988), juvenile molluscs and crustaceans (Bowser *et al.*, 1981; Disalvo *et al.*, 1978; Tubiash *et al.*, 1990). In Japan, *V. anguillarum* is the most frequent and serious bacterial infection of the freshwater cultured fish, the ayu (*Plecoglossus altivelis*) (Aoki *et al.*, 1981; Hirono *et al.*, 1996).

Importantly all 10 serotypes have been associated with disease outbreaks the severity of which is dependent upon the serotype involved. Serotypes O8, O9 and O10, although first isolated from diseased fish, have not been involved subsequently in any reported infection and are found only in the environment. These serotypes have been described as true environmental isolates that have low pathogenicity (Sørenson and Larson, 1986). The most common environmental strains belong to serotypes O3, O4, O5 and O6 with serotypes O1 and O2 being the least frequent serotypes isolated. Serotype O7 has not been isolated from the environment and has to date only been isolated from diseased fish (Sørenson and Larson, 1986). Together these data suggest serotypes O1, O2 and O7 have pathogenic potential.

In recent years the serotyping scheme for *V. anguillarum* has been refined, resulting in additional O-serotypes being identified and characterised on the basis of LPS profiles and serological reactions (Grisez and Ollevier, 1995; Pederson *et al.*, 1999). These schemes have increased the known serotypes of *V. anguillarum* from 10 to 23. The additional serotypes have all been isolated from diseased fish (Pederson *et al.*, 1999). It is also evident from these studies that some of the strains isolated did not react with available antisera suggesting the presence of additional O-serotypes for *V. anguillarum* (Pederson *et al.*, 1999).

1.1.2 Vaccines

It has been known that inoculation of fish with vaccines consisting of heat-inactivated, Formalin-treated or phenol-water extracts of *V. anguillarum* LPS protects against subsequent infection (Harrell, 1978; Itami and Kusuda, 1980; Evelyn, 1984; Mutharia *et al.*, 1993). The protection provided is serotype specific. Vaccines of this type have long been used in the aquaculture industry, however the vaccine preparation needs to be injected intraperitoneally for effective protection and therefore the vaccination process is time consuming, costly and requires the fish to be sedated (Horne *et al.*, 1982; Olsson *et al.*, 1992). In addition the protection offered is serotype specific and the geographic origin of the vaccine strain needs to be considered when administering to fish stocks.

1.1.3 Pathogenesis of *V. anguillarum*

Vibriosis induced by *V. anguillarum* was first described in an outbreak of disease in eels from the Baltic Sea in 1909. The disease was characterised by the appearance of bloody lesions in the musculature of the infected fish (Chen and Hanna, 1992). Vibriosis also causes the intestine and rectum to be swollen with a clear viscous fluid (Olsson *et al.*, 1992; Hastein and Holt, 1972).

To date the detail of the pathogenesis of *V. anguillarum* is superficial. It is not known how the pathogen gains entry, although the gastrointestinal tract is thought to be the site of colonisation (Horne and Baxendale, 1983; Olsson *et al.*, 1992). For a pathogen to induce a successful systemic infection it must have properties (ie virulence factors) that allow it to establish in host tissues and have mechanisms to either avoid or suppress host immune responses (Trust *et al.*, 1981). As yet specific factors that allow *V. anguillarum* to invade the fish epithelial layer have not been characterised although the flagellum has been suggested to play a role in this process (McGee *et al.*, 1996; Milton *et al.*, 1996; O'Toole *et al.*, 1996). Comparisons have been made to *Vibrio cholerae* which requires adherence to intestine mucosal cells to induce disease even though it is a non-invasive human pathogen (Pierce *et al.*, 1985). Attachment assays have shown that *V. anguillarum* serotypes O1, O2, O4 and O8 adhere to rainbow trout (*Oncorhynchus mykiss*) fish cells *in vitro*. It was shown that serotypes O1 and O2 have greatest attachment potential which is not surprising since these serotypes are considered to be the most virulent (Chen and Hanna, 1992). It has been proposed that infection requires adherence/attachment and penetration of the epithelial cell layers as invasion is implied because the disease involves haemorrhagic septicaemia (Chen and Hanna, 1994). Recently, Wang and colleagues (1998) demonstrated *in vitro* that some strains of *V. anguillarum* were capable of entering and surviving as intracellular parasites in epithelial cell lines which would provide a mechanism for the bacteria to enter deeper tissues and cause the characteristic bloody lesions in the musculature of the fish host.

1.1.4 Experimental fish model

To investigate the potential virulence of isolates and manipulated laboratory strains, a experimental fish model has been established (Norqvist *et al.*, 1989). The animal used in this model is one of the natural hosts, rainbow trout (*Oncorhynchus mykiss*). The model allows the investigation of at least two steps in the proposed infection process of *V. anguillarum*. When fish are inoculated by the immersion method (bath vaccination), the bacteria require mechanisms to gain entry to the host and to invade the epithelial cell layer.

Intraperitoneal vaccinations (by injection) allow the study of the later stages of the infection, after invasion has occurred.

1.2 Chromosomal loci associated with virulence

Although the virulence determinants of *V. anguillarum* are not well characterised, numerous secreted and surface expressed factors are thought to contribute to the pathogenicity of this bacteria based on comparisons with similar pathogens. *V. anguillarum* produces several toxins including: haemolysin (Hirono *et al.*, 1996; Munn, 1978, 1980), cytotoxin (Toranzo *et al.*, 1983), proteases including a zinc metalloprotease (Milton *et al.*, 1992; Norqvist *et al.*, 1990), haemagglutinin (Toranzo *et al.*, 1983; Trust *et al.*, 1981), and adhesion factors required for attachment (Chen and Hanna, 1992).

1.2.1 Haemolysin

Bacterial haemolysins cause haemorrhagic septicaemia and diarrhoea, and have been associated with virulence in numerous species of Vibrionaceae including: *V. cholerae* (Alm *et al.*, 1988, 1991; Radar and Murphy, 1988) *V. parahaemolyticus* (Nishibuchi and Kaper, 1985), *V. mimicus* (Terai *et al.*, 1990), *V. vulnificus* (Yamamoto *et al.*, 1990), *Aeromonas salmonicida* (Hirono and Aoki, 1993) and *A. hydrophila* (Aoki and Hirono, 1991). The haemolysins from these species have been cloned and are well characterised. Surprisingly, the structures of these haemolysins are quite different from each other with the exception of *V. cholerae*, *V. vulnificus* and the aeromonads (Hirono and Aoki, 1991; Hirono and Aoki, 1993; Hirono *et al.*, 1996) Recently, the haemolysin gene from *V. anguillarum* was cloned and sequenced (Hirono *et al.*, 1996). The haemolysin gene isolated was found to be most homologous to the *V. cholerae* O1 El Tor haemolysin

(57.3%), *A. hydrophilia* AHH1 haemolysin (46.2%) and *A. salmonicida* ASH4 haemolysin (43.0%) (Hirono *et al.*, 1996). Haemolytic assays have shown the haemolysin of *V. anguillarum* to have broad spectrum activity although fish erythrocytes were more sensitive than mammalian erythrocytes. Hybridisation data indicated that the haemolysin gene was widely distributed throughout the *V. anguillarum* serotypes.

Earlier studies on the haemolysin showed that the protein was a thermolabile toxin as haemolytic activity was abolished if the haemolysin was heated to 50°C-100°C (Munn, 1978). This not an unusual property of haemolysins as other thermolabile haemolysins have been reported in *V. parahaemolyticus* (Sakurai *et al.*, 1974). Due to the nature of the *V. anguillarum*-mediated vibriosis, characterised by terminal haemorrhagic septicaemia, it should be expected that the haemolysin would play an important role in the disease process (Hirono *et al.*, 1996). However, numerous studies on the haemolysin of *V. anguillarum* have yet to display a clear link between its presence and virulence (Munn, 1978, 1980). It has been shown that both pathogenic and non-pathogenic strains produce a haemolysin (Toranzo *et al.*, 1983). Thus, it is possible that the haemolysin is only one of several toxins that determine virulence.

1.2.2 Zinc metalloprotease

Proteases have been implicated as important virulence factors in numerous pathogen systems including: *V. cholerae* (Finkelstein *et al.*, 1992; Ogierman *et al.*, 1997; Ingole *et al.*, 1998) *Pseudomonas aeruginosa* (Pavlovskis *et al.*, 1985), *V. vulnificus* (Kothary *et al.*, 1985) and *Legionella pneumophila* (Keen and Hoffman, 1989). These proteases are often associated with entry (ie. initial stages of the infection) although the actual role of these proteins in the virulence processes of these bacteria is not well understood.

A gene encoding a zinc metalloprotease from *V. anguillarum* has been cloned and sequenced (Milton *et al.*, 1992). Initial studies using a rifampicin resistant mutant indicated that the protease had a role in promoting invasion as the mutant displayed a decrease in virulence when immersion infections were used as compared to intraperitoneal infections (Norqvist *et al.*, 1990). However, subsequent mutational analysis using a defined chromosomal mutant showed that virulence in the fish experimental model was not significantly effected, suggesting that the metalloprotease does not have a role in entry (Milton *et al.*, 1992). When the defined mutant was analysed on SDS-PAGE, the protease was no longer present but two new proteases of 75 and 30kDa were detected in the mutant strain that could not be found in the wild type. This suggests that these proteases may compensate for the loss of the zinc metalloprotease and hence *V. anguillarum* may have a multiplicity of proteases (Milton *et al.*, 1992). Multiple proteases have been implicated in virulence in other bacterial systems including: *Pseudomonas aeruginosa*, *V. vulnificus* and *A. salmonicida* (Howe and Iglewski, 1984; Smith and Merkel, 1982; Sheeran *et al.*, 1983).

The zinc metalloprotease open reading frame (ORF) encodes a 611 amino acid polypeptide which is considerably larger than the protein seen on SDS-PAGE gels where the protein is approximately 330 amino acids. Therefore the 611 amino acid polypeptide may be an unprocessed pre-protein (Milton *et al.*, 1992), which has also been described for other bacterial zinc metalloproteases (Kothary *et al.*, 1987; Häse and Finkelstein, 1991; David *et al.*, 1992).

The protein sequence was compared to other homologous proteases. Proteases of this type show good homology in the zinc binding sites, active site and the substrate binding site. The most variable site is the substrate binding site which is not that surprisingly as different bacterial species may have specific substrate requirements. The *V. anguillarum* zinc metalloprotease is very similar to the *V. cholerae* haemagglutinin/protease as the amino acids in the substrate binding site are identical suggesting that evolution has conserved the protein between these species (Milton *et al.*, 1992).

The HA/protease of *V. cholerae* has been postulated to have an indirect role in virulence, if any role at all (Finkelstein *et al.*, 1992). Mutants in HA/protease do not effect virulence in the infant rabbit model and therefore this protein has been suggested to be a 'detachase' which allows *V. cholerae* to detach from host epithelium to facilitate spread to another host (Finkelstein *et al.*, 1992). This is postulated to occur by the digestion of several putative receptors for *V. cholerae* adhesins by the HA/protease (Finkelstein *et al.*, 1992).

To date there is no conclusive evidence for the role the zinc metalloprotease may have in the virulence of *V. anguillarum*.

1.2.3 *virC*

virC is located on the chromosome and was originally isolated by transposon mutagenesis and screening for the loss of virulence (Norqvist and Wolf-Watz, 1993). The transposon mutant isolated had two insertions and did not express lipopolysaccharide (LPS). Plasmid insertion mutants were constructed at each locus effected by a transposon. One mutant (*virC*) had a 10^4-fold increase in LD_{50} and the second mutant showed a wild type phenotype. Both mutants expressed LPS suggesting either a double mutation was required to lose LPS expression or the presence of additional transposon insertions in the original mutant. *virC* is located within a potential transcriptional unit consisting of three putative ORFs. The first ORF shows 30% identity to the *Escherichia coli* and *Salmonella cysG* gene, with the size of the encoded protein being 34.8 kDa. The second and third genes do not show any significant homology to any proteins lodged in GenBank. The second gene, *virC,* which was found to be defective in virulence assays is essential for virulence, however its function is unknown (Milton *et al.,* 1995).

1.2.4 Quorum sensing

In recent years it has been discovered that certain Gram-negative pathogens control virulence gene expression via cell to cell communication through small diffusible signal molecules termed autoinducers. This type of communication is called quorum sensing.

V. anguillarum is another pathogen that has genes required for this type of communication. Milton *et al* (1997) discovered the autoinducer of *V. anguillarum* to be *N*-(3-oxodecanoyl)-L-homoserine lactone (ODHL) and the gene responsible for its synthesis (*vanI*). *vanI* was cloned and sequenced, and its product was found to be homologous to the LuxI family of putative N-acyl homoserine lactone (AHL) autoinducers. A second gene, *vanR* is located further upstream of *vanI* and it product is 38% identical to the LuxR family of positive transcriptional activators. Between *vanI* and *vanR* is a *lux* box-like sequence that is required in these regulatory circuits (Milton *et al.*, 1997). The *lux* box is a 20 bp region of dyad symmetry that binds LuxR (Devine *et al.*, 1989).

As it was apparent that *V. anguillarum* used AHLs as a mechanism for gene expression, Milton *et al* (1997) searched for virulence genes that may have a *lux*-like box for regulation by *vanR*. The zinc metalloprotease gene, *empA,* that has been implicated in virulence was found to contain this sequence upstream of the translational start codon. However, mutations in both *vanI* and *vanR* that abolish ODHL synthesis did not effect the production of the metalloprotease nor was virulence affected in the experimental fish model (Milton *et al.*, 1997). Therefore further investigation of other virulence determinants and their mechanisms of regulation is required to determine if the *vanI/vanR* system is involved in the pathogenicity and/or expression of virulence factors.

1.2.5 Adhesins and flagella

Bacterial adhesion to the host tissues is of importance during the initial stages of infection (Beachey, 1981; Krovacek *et al.*, 1987). The mechanisms that are involved in adhesion of *V. anguillarum* have not been extensively studied. It has been shown that *V. anguillarum* adhere to cultured fish cells (Krovacek *et al.*, 1987; Chen and Hanna, 1992).

Haemagglutinating activity of bacteria has been correlated with their adherence and is thought to play a role in virulence. Toranzo *et al.* (1983) showed a correlation between the production of haemagglutinin and pathogenicity in strains from the North Atlantic. However, this contradicted an earlier study by Trust *et al.* (1981) who were unable to show a correlation between strain virulence and the ability to agglutinate fish erythrocytes.

The haemagglutinins produced by some strains of *V. anguillarum* were found to be inhibited by D-mannose suggesting this sugar maybe apart of the erythrocyte receptor (Toranzo *et al.*, 1983). Further work is required to determine the precise role of the haemagglutinin in the pathogenesis of *V. anguillarum*, and as yet the gene encoding this protein has not been cloned.

Flagellae have been recognised as potential virulence factors, either as a motility organelle or as a factor involved in adhesion (McGee *et al.*, 1996). The flagellum is the motility organelle of many bacteria. It comprises a complex membrane-associated structure that consists of a basal body, an external hook, helical filament and a filament cap. The flagellum enables the bacterium to move by a utilising the membrane associated-flagellar motor that results in propulsion of the bacteria (Macnab., 1992). The flagellae of *E. coli* and *Salmonella* are the best characterised and are composed of multiple copies of a single flagellin subunit (Nei, 1987). Numerous other bacteria have been found to have a more complicated flagellum which is composed of more than one flagellin protein. These flagella are associated with the bacteria: *Campylobacter jejuni* (Guerry *et al.*, 1991; Guerry *et al.*, 1992; Yao *et al.*, 1994), *Helicobacter* (Kostrzynska *et al.*, 1991), *Aeromonas salmonicida* (Umelo and Trust, 1997) and *V. anguillarum* (McGee *et al.*, 1996).

V. anguillarum possesses a single polar flagellum that is sheathed with LPS (Norqvist and Wolf-Watz, 1993; O'Toole *et al.*, 1996). The flagella structure is complex and believed to consist of four flagellin proteins (Milton *et al.*, 1996). These flagellin proteins are encoded by the genes *flaA* (Milton *et al.*, 1996), *flaB,C* and *D* (McGee *et al.*, 1996). Studies involving the analysis of the role of the flagellum in pathogenesis have indicated that chemotactic motility is required for *V. anguillarum* to invade fish (rainbow trout) when they are immersed in infected sea water but not required for virulence once the epithelial cell layer is crossed (McGee *et al.*, 1996; O'Toole *et al.*, 1996).

Mutations of the structural genes of the flagellum have indicated that only *flaA* is essential for virulence (Milton *et al.*, 1996). Mutations in *flaB, C* and *D* result in only a slight decrease in motility and do not effect invasion of the host (McGee *at al.*, 1996).

FlaA is thought to have two roles in the pathogenesis of *V. anguillarum*. Firstly, it appears that motility is essential for efficient initial invasion of the host. FlaA may also have a function once the fish integument is crossed as a FlaA mutant with partial motility was avirulent by intraperitioneal injection. It is not known how FlaA is involved at this stage of the infection process but McGee *et al* (1996) have suggested it may involve suppression of other virulence determinants due to the inability of C-terminal deleted FlaA mutants to secrete the truncated proteins out of the cytoplasm (Milton *et al.*, 1996).

1.2.6 Outer membrane porins

The outer membrane of Gram-negative bacteria contain numerous pore-forming proteins (or porins) that act as channels, which either facilitate the transport of specific or non-specific molecules. Porin expression can depend on the culture conditions and environmental factors.

The outer membrane profile of *V. anguillarum* has been examined in numerous studies (Buckley *et al.*, 1981; Pazos *et al.*, 1993) with the best characterised outer membrane proteins being those involved in iron uptake (Actis *et al.*, 1995; Tolmasky *et al.*, 1995).

The porins which have been analysed in *V. anguillarum* are the 40 kDa major outer membrane protein (MOMP) of the O1 strains (Simon *et al.*, 1996) and a 35 kDa porin-like protein, Omp35La (Suzuki *et al.*, 1996). It is not known whether these proteins are similar proteins or if in fact they constitute a single porin (Davey *et al.*, 1998).

The MOMP is common to all 10 serotypes of *V. anguillarum* with the size of the protein varying between 35 to 42 kDa. Immunological studies have shown cross-reactivity between MOMPs of all the serotypes (Suzuki *et al.*, 1994; Simon *et al.*, 1996).

The MOMP from *V. anguillarum* O1 has been isolated and the two dimensional structure determined by electron microscopy and image processing. The structure determined was consistent with that of other bacterial porins. The MOMP can be characterised as a general diffusion porin with weak cation selectivity and moderate surface charge. MOMP is similar to OmpF from *E. coli* (Simon *et al.*, 1996). The gene which expresses this MOMP has not been cloned.

Further analysis of the 40 kDa MOMP of *V. anguillarum* O2 has indicated the amount of the protein expressed is influenced by osmolarity, salt concentrations and the chelator EDTA in growth media (Davey *et al.*, 1998). OmpF of *E. coli* is also regulated by these culture conditions (Mizano and Mizushima, 1990). In addition when *V. anguillarum* O2 was grown at elevated temperatures (37°C) and in EDTA-containing media, novel proteins of 60 and 19 kDa, respectively, were observed in cell lysates (Davey *et al.*, 1998).

It is unclear how the porins of *V. anguillarum* are regulated and further studies at the molecular level are necessary for a better understanding of both regulation and the potential role of the porin in either pathogenicity or survival of the bacteria in the marine environment.

1.3 Iron sequestering system associated with a large plasmid, pJM1

Bacteria require iron for growth as it is an essential nutrient that is involved in numerous biological processes including electron transport chains and as cofactor of enzymes of intermediatory metabolism (Neilands, 1981).

Bacteria and other organisms have systems that enable them to obtain iron from mammalian storage systems (transferrin, lactoferrin and ferritin). In addition, bacteria have specialised iron transport systems that enable then to scavenge iron in limiting conditions imposed by either the host or environment (Bullen, 1981; Crosa, 1984; Griffiths, 1987 a, b; Sussman, 1974; Crosa, 1989).

Pathogenic strains of *V. anguillarum* possess a 65 kb plasmid pJM1 that encodes an iron transport system which is composed of the siderophore anguibactin and an energy-dependent transport system (Crosa *et al.*, 1977; Actis *et al.*, 1988; Crosa, 1989). This plasmid-mediated iron-uptake system is an important component of the pathogenic processes of these bacteria (Crosa, 1980; Crosa *et al.*, 1980). The iron-uptake system is located on a 25 kb segment of the plasmid pJM1 (Tolmasky and Crosa, 1984; Tolmasky *et al.*, 1988). Chromosome-mediated regulation, essential for repression of the iron transport genes, (Waldbeser *et al.*, 1993; Tolmasky *et al.*, 1994; Waldbeser *et al.*, 1995) is mediated by the Fur repressor protein (Chen and Crosa, 1996).

The genes *fatDCB* and *A*, involved in iron transport are transcribed as a polycistronic mRNA (Waldbeser *et al.*, 1993; Actis *et al.*, 1995). FatA is the receptor for ferric anguibactin complexes (Actis *et al.*, 1985), FatB is a membrane located lipoprotein (Actis *et al.*, 1985), and Fat C and FatD are cytoplasmic integral membrane proteins (Koster *et al.*, 1991) (Fig. 1.1). These genes as well as the genes required for anguibactin are expressed under iron-limiting conditions. Anguibactin and the positive transcriptional activators, AngR and TAF (transacting factor) are essential for synthesis of the siderophore and regulation of the iron transport genes. The action of AngR depends on the presence of TAF which is encoded by pJM1 in a region non-contiguous to the other genes involved in iron transport (Chen *et al.*, 1996) (Fig. 1.1).

The AngR protein is 110 kDa in size and possess typical features of a DNA binding protein including two helix-turn-helix motifs, one at the N-terminus and the other at the C-terminus (Chen *et al.*, 1996). The two helix-turn-helix domains are also proceeded by a leucine zipper motif (Tolmasky *et al.*, 1993) which in other systems has been demonstrated to be involved in protein-protein and DNA-protein interactions (Agre *et al.*, 1989; Turner and Tjian, 1989; Vinson *et al.*, 1989; Pu and Struhl, 1991; Chen *et al.*, 1996).

The structure and composition of the anguibactin has been determined (Actis *et al.*, 1986; Jalal *et al.*, 1989). It has a molecular weight of 348 and belongs to the phenolate family of siderophores, although its structure is unique when compared to other known bacterial iron chelators (Fig. 1.2). Anguibactin has been identified as ω-*N*-hydroxy-ω-[[2'-(2",3"-dihydroxyphenyl)thiazolin-4'-yl}-carboxy]histamine by crystal X-ray diffraction studies, nuclear magnetic resonance spectroscopy, fast atom bombardment, mass spectroscopy and chemical degradation (Jalal *et al.*, 1989). These analyses revealed a 1:1 complex of iron to anguibactin (Jalal *et al.*, 1989).

Figure 1.1: The iron sequestering system of *V. anguillarum*

Organisation of the pJM1 iron uptake region. FatDCBA are transcribed on a single mRNA. Fat A is the receptor for ferric anguibactin complexes; FatB is the membrane located lipoprotein; FatC and FatD are cytoplasmic membrane proteins. In low iron conditions AngR and Taf act synergistically to positively regulate (+ve) FatB, FatA and anguibactin biosynthesis gene expression. In high iron conditions the iron transport and anguibactin biosynthesis genes are down regulated by the pJM1-derived antisense RNAα and chromsomally encoded Fur repressor protein. Diagram adapted from Crosa (1989).

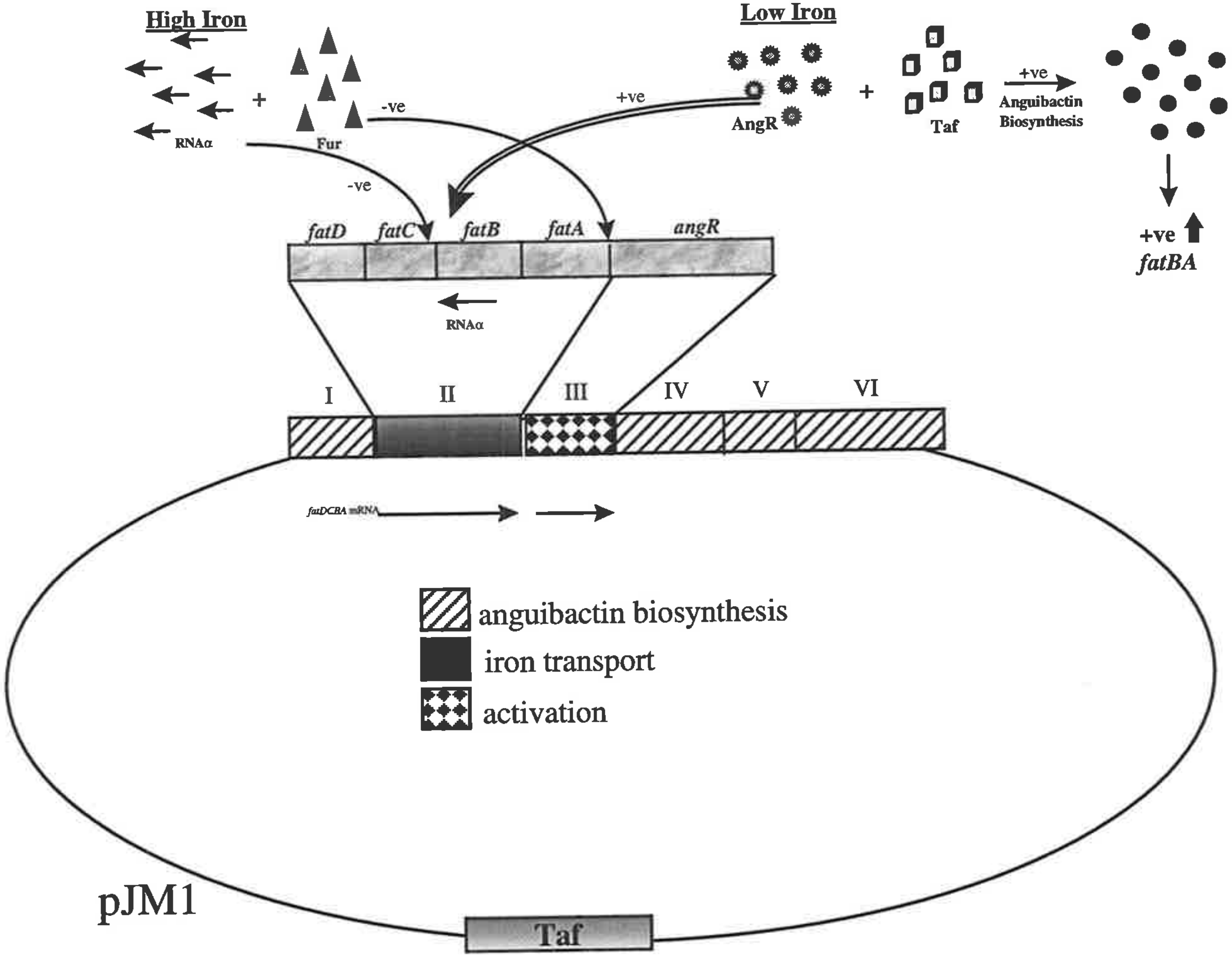
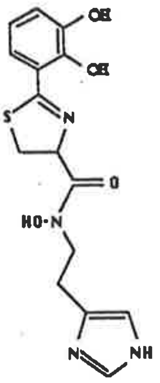

High Iron
RNAα
+
Fur
-ve
-ve
Low Iron
+ve
AngR
+
Taf
+ve
Anguibactin Biosynthesis
+ve
fatBA
fatD
fatC
fatB
fatA
angR
RNAα
I
II
III
IV
V
VI
fatDCBA mRNA
anguibactin biosynthesis
iron transport
activation
pJM1
Taf

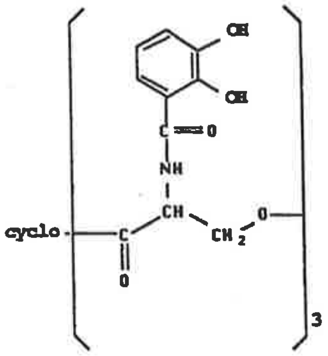
Figure 1.2: Iron chelator molecules

The structures of siderophores anguibactin, enterobactin, pyochelin and aerobactin.

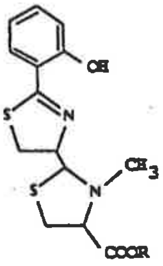
Diagram from Crosa, (1989).



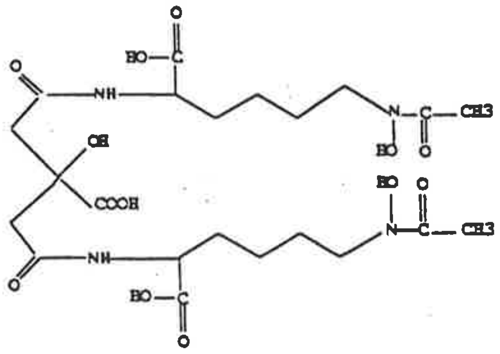
Anguibactin



Enterobactin



Pyochalin



Aerobactin

The plasmid pJM1 also contains a negative regulator known as RNAα. RNAα is a 650 nucleotide RNA which is encoded by the *fatB* gene in the complimentary strand and is expressed during iron-rich conditions. RNAα reduces the expression of FatA and FatB as it binds to the polycistronic mRNA which encodes these genes causing a change in the secondary structure and prevents translation of the message (Waldbeser *et al.*, 1993; Waldbeser *et al.*, 1995). The Fur repressor protein is required for expression of RNAα (Chen and Crosa, 1996) (Fig. 1.1).

1.4 *V. anguillarum* Lipopolysaccharide

LPS is a known virulence factor in many Gram negative pathogens and has been shown to be important in the pathogenicity of *V. anguillarum.* Mutants in LPS biosynthesis genes render the bacteria avirulent (Norqvist and Wolf-Watz, 1993) and a recent study showed the importance of LPS (in particular the O-antigen) to the resistance of complement-mediated killing (Boesen *et al.*, 1999).

Although the importance of LPS in *V. anguillarum* as a virulence determinant has been established, little is known about the genetics, composition and biosynthesis of this surface molecule.

1.4.1 Lipopolysaccharide biosynthesis in *V. anguillarum* O1

V. anguillarum O1 is the most significant serotype in disease outbreaks. To date little is known about the genetics or exact composition of the LPS molecule in this serotype.

Two genes *virA* and *virB* were isolated by insertional mutagenesis and subsequent screening for avirulence in a fish experimental model. Originally these genes were thought to be involved in the production of a flagellar sheath antigen (Norqvist and Wolf-Watz, 1993), however subsequent analysis of these genes showed that the antigen was in fact LPS and that it was expressed over the entire surface of the bacteria (Milton *et al.*, 1996). These findings suggest the genes, *virA* and *virB* are involved in the biosynthesis of the O-antigen and that LPS is essential for the virulence of *V. anguillarum* O1.

The *virA* and *virB* genes were cloned and sequenced (Norqvist and Wolf-Watz, 1993). The sequences were analysed with the programs TFASTA and FASTA, however no significant homologues to any sequences in the database were found at the time.

1.4.2 Heterogeneity of lipopolysaccharide in *V. anguillarum* O2

V. anguillarum O2 is the most prominent serotype involved in disease outbreaks of vibriosis in feral fish stocks (Amor and Mutharia, 1995; Sørenson and Larsen, 1986). In the past there has been some confusion in typing strains belonging to the O2 serotype due to cross-reactive and specific epitopes found on the LPS molecule as determined by monoclonal antibodies (Mutharia and Amor, 1994). The confusion has been increased due to the presence of a closely related fish pathogen, *V. ordalii,* which is indistinguishable from *V. anguillarum* O2 by Western immunoblotting (Chart and Trust, 1984; Mutharia *et al.*, 1993). However, it is now widely accepted that *V. ordalii* is in fact *V. anguillarum* O2 (Sørenson and Larsen, 1986; Mutharia *et al.*, 1993; Amor and Mutharia, 1995) but it expresses serotype specific epitopes (Salati and Kusuda, 1986). Interestingly, when fish are vaccinated with *V. ordalii* and challenged with *V. anguillarum* O2, the fish are susceptible to infection which suggests that the specific epitopes expressed are important for an effective infection (Mutharia *et al.*, 1993). Due to these antigenic differences, Rasmussen (1987a, b) subdivided *V. anguillarum* O2 into O2a and O2b based on immunoelectrophoretic patterns of heat-stable polysaccharide antigens obtained with rabbit polyclonal serum. An alternative subdivision was also described by Bolinches *et al.* (1990) where the O2 serotype was subdivided on the basis of the reactions of rabbit polyclonal antiserum raised against bacterial antigens. The subdivisions were reported as O2α and O2β. *V. ordalii* was not included in these subdivisions. All these data leads one to suspect that there is heterogeneity in LPS (in particular O-antigen) within the O2 serotype and that further classification is required to distinguish the strains within this group.

1.4.2.1 Structure of the O-antigen and capsule from *V. anguillarum* O2

It has been discovered that when *V. anguillarum* O2a is grown in the presence of rainbow trout blood, an acidic capsule is produced (Amor and Mutharia, 1995). This is the only serotype that has been reported to have a capsule. Interestingly, this capsule is antigenically identical to the O-antigen as monoclonal antibodies specific for the O-polysaccharide recognise the capsule material (Sadovskaya *et al.*, 1996; Amor and Mutharia., 1995). This observation is not uncommon in *Vibrio* species. The recent epidemic strain of *V. cholerae* designated O139 also expresses a capsule that is antigenically identical to the O-antigen of the LPS (Waldor *et al.*, 1994). This has also been observed in *E. coli* strains with group I capsules. The K-antigen expressed by these bacteria can either be linked to lipid A-core oligosaccharide as O-antigen or be exported as high molecular weight capsular K antigen (MacLachlan *et al.*, 1993; Whitfield and Roberts, 1999).

The O-antigen and capsule of *V. anguillarum* O2 are composed of linear tetrasaccharide repeat units of diamino uronic acid and a N-L-alanylformamido group (Sadovskaya *et al.*, 1996; Amor and Mutharia, 1995) (Fig. 1.3). This is an unusual

Figure 1.3: Comparison of the capsule and O-antigen polysaccharides from *V. anguillarum* O2 and *V. ordalii* O2.

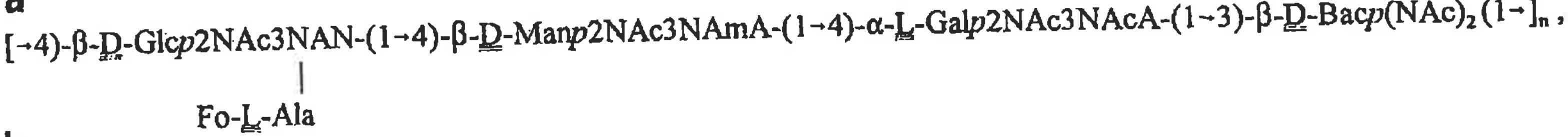
A) The structural formulae of the capsule/O-antigen molecules of *V. anguillarum* O2 (a) and *V. ordalii* (b).

B) Structure of the repeating unit of the O-specific and capsule polysaccharides *of V. ordallii*.

Diagram was adapted from Sadovskaya *et al.*, (1998).

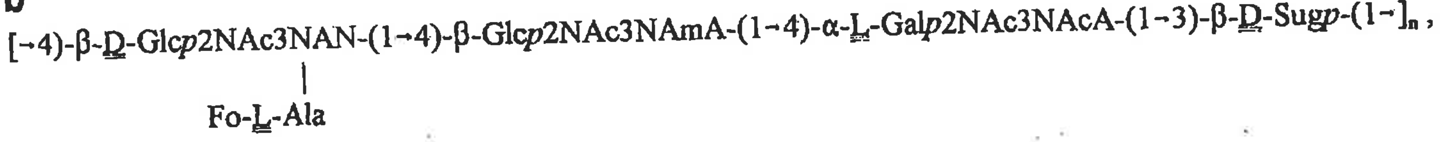
A

a



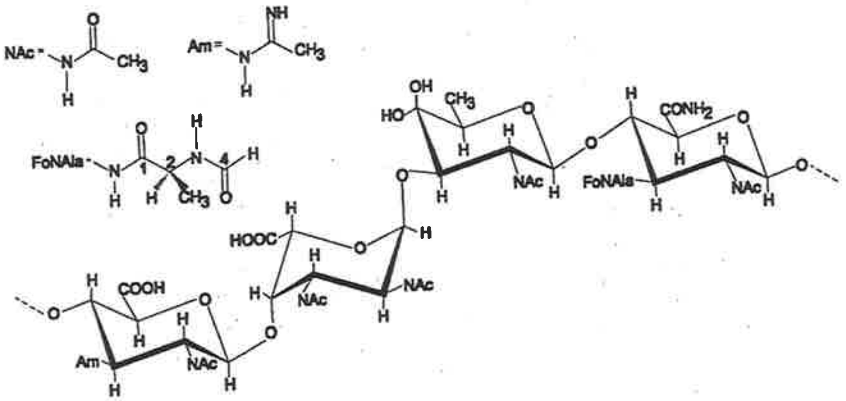
$$[\rightarrow 4)\text{-}\beta\text{-}\underline{\mathrm{D}}\text{-Glc}p\text{2NAc3}\underset{\substack{|\\ \text{Fo-}\underline{\mathrm{L}}\text{-Ala}}}{\text{N}}\text{AN-}(1\rightarrow 4)\text{-}\beta\text{-}\underline{\mathrm{D}}\text{-Man}p\text{2NAc3NAmA-}(1\rightarrow 4)\text{-}\alpha\text{-}\underline{\mathrm{L}}\text{-Gal}p\text{2NAc3NAcA-}(1\rightarrow 3)\text{-}\beta\text{-}\underline{\mathrm{D}}\text{-Bac}p\text{(NAc)}_2(1\rightarrow]_n\,,$$

b



$$[\rightarrow 4)\text{-}\beta\text{-}\underline{\mathrm{D}}\text{-Glc}p\text{2NAc3}\underset{\substack{|\\ \text{Fo-}\underline{\mathrm{L}}\text{-Ala}}}{\text{N}}\text{AN-}(1\rightarrow 4)\text{-}\beta\text{-Glc}p\text{2NAc3NAmA-}(1\rightarrow 4)\text{-}\alpha\text{-}\underline{\mathrm{L}}\text{-Gal}p\text{2NAc3NAcA-}(1\rightarrow 3)\text{-}\beta\text{-}\underline{\mathrm{D}}\text{-Sug}p\text{-}(1\rightarrow]_n\,,$$

B



composition and has not been identified previously in bacterial polysaccharides. The composition of *V. ordalii* polysaccharide (O-antigen and capsule) has also been determined and shares two monosaccharides (2,3-di-acetoamide-2,3-dideoxy-α-L-galacturonic acid and 2-aceto-3-(N-formyl-L-alanyl)-amino-2,3-dideoxy-β-D-glucuronamide) with the O-antigen/capsule of *V .anguillarum* O2a (Sadovskaya *et al.*, 1998) (Fig. 1.3). The cross-reactivity observed between *V. ordalii* and *V. anguillarum* O2 can be explained by these similarities in polysaccharide structure. Specific epitopes on the O-antigen, as previously determined by monoclonal antibodies is probably conferred by the presence of different sugars at the end of the repeat unit (Sadovskaya *et al.*, 1998) (Fig 1.3). To date, the composition of the O-antigen and capsule of *V. anguillarum* O2b are unknown (Amor and Mutharia., 1995).

Genetic studies of *V. anguillarum* O2 LPS biosynthesis is limited. Amor and Mutharia (1995) cloned and expressed the genes involved in O-antigen (and capsule?) biosynthesis of a type strain (ATCC 19264) of *V. anguillarum* O2. The region has yet to be sequenced and so the genes responsible for the synthesis of the polysaccharide are not yet identified.

1.4.2.1 Chemical structures and composition of O-antigen in *V. anguillarum* O3

Although analysis of *V. anguillarum* O-antigens is limited for the O1 serotype and only recently characterised in the O2 serotype, a number of studies have been conducted with the strain V-123 which belongs to the O3 serotype. The structure and composition of the O-antigen repeating unit of this serotype was determined by acid hydrolysis and NMR. The O-polysaccharide repeating unit consists of →3)-α-GalNAcA(amido)-(1→4)-α-GalNFoA-(1→3)-α-QuiNAc-(1→3)-β-VioNAcyl-(1→ where the N-Acyl group is 2,4,-

dihydroxy-3,3,4-trimethylproglutamic acid, Qui is D-quinovosamine, Vio is D-viosamine, and Fo is formyl (Eguchi *et al.*, 1992). The presence of the N-formyl group is rare in O-polysaccharides although 2-deoxy-2-formamido-D-galacturonic acid, 4,6-dideoxy-4-formamido-L-mannose and 1,3 linked 4 formamido-4,6-dideoxy-D-mannose have been found in O-antigens of *Pseudomonas aeruginosa* (Yokota *et al.*, 1986; Knirel *et al.*, 1985), *Yersinia enterocolitica* O9 (Caroff *et al.*, 1984) and *Brucella abortus* (Aragon *et al.*, 1996), respectively.

1.5 Lipopolysaccharide

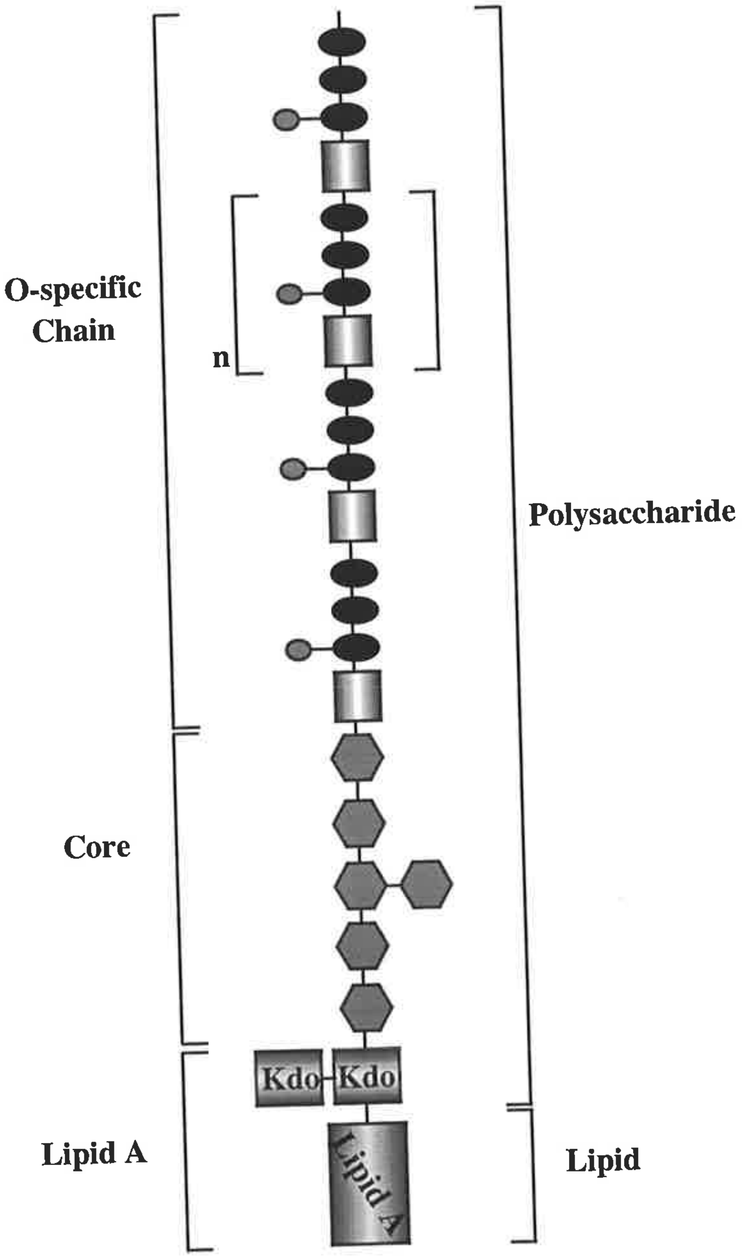
1.5.1 General Introduction

Lipopolysaccharide (LPS) is the most abundant molecule on the surface of Gram-negative bacteria and acts as a protective barrier to hydrophobic agents such as dyes, detergents and other lethal agents (Osborn., 1979).

LPS consists of three components: lipid A, the core oligosaccharide and the O-antigen (Fig. 1.4). Lipid A is situated in the lipid bilayer of the outer membrane with the core oligosaccharide attached to the lipid A via keto-3-deoxy-D-mannose-octulosonic acid (KDO) (Schmidt *et al.*, 1969). The O-antigen which consists of either polysaccharide or monosaccharide polymer of variable length is linked to the core oligosaccharide and constitutes the outer most region of the LPS molecule. O-antigen is the O-serotype determinant in Gram-negative bacteria (Lüderitz *et al.*, 1971; Jann and Westphal, 1975) and in some bacteria is an important virulence determinant that prevents complement-mediated serum killing (Whitfield, 1995). The diversity of O-antigen structures present in Gram-negative bacteria is due to variations in sugar composition, the linkages between

Figure 1.4: Schematic structure of lipopolysaccharide (LPS).

The lipid A, core oligosaccharide and the O-specific chain are indicated on the left side of the figure. Individual O-repeats are indicated, n represents the number of repeating units. The polysaccharide and lipid component of the molecule are shown in brackets on the right side of the figure. Abbreviations: Kdo, 3-deoxy-D-manno-octulonic acid. Diagram adapted from Lindberg *et al.* (1991).


O-specific
Chain
n
Polysaccharide
Core
Kdo
Kdo
Lipid A
Lipid A
Lipid

sugar molecules and the numerous sugar and non-sugar substitutions found on the O-antigen (Whitfield, 1995).

The best characterised "LPS" systems are those that have been described for *Escherichia coli, Salmonella enterica* and *Shigella flexneri.* The structures of these LPS molecules has aided in the understanding of LPS biosynthesis. With this knowledge numerous similarities have emerged between these genera which is perhaps not that surprising given their close phylogenetic relationships. The biosynthesis pathways involved in the production of the LPS molecules have been comprehensively studied although there are still many questions that remain unanswered.

Biosynthesis and assembly of LPS involves a set of complex reactions. In general biosynthesis of LPS occurs via a number of defined steps:

1) lipid A is synthesised at the surface of the cytoplasmic membrane,

2) the core oligosaccharide is synthesised on lipid A,

3) the O-antigen units are synthesised on a lipid carrier (undecaprenol phosphate or bactoprenol),

4) the O-antigen units are polymerised into appropriate length chains, and

5) the O-antigen chain is ligated to the lipid A/core subunit (Fig. 1.5).

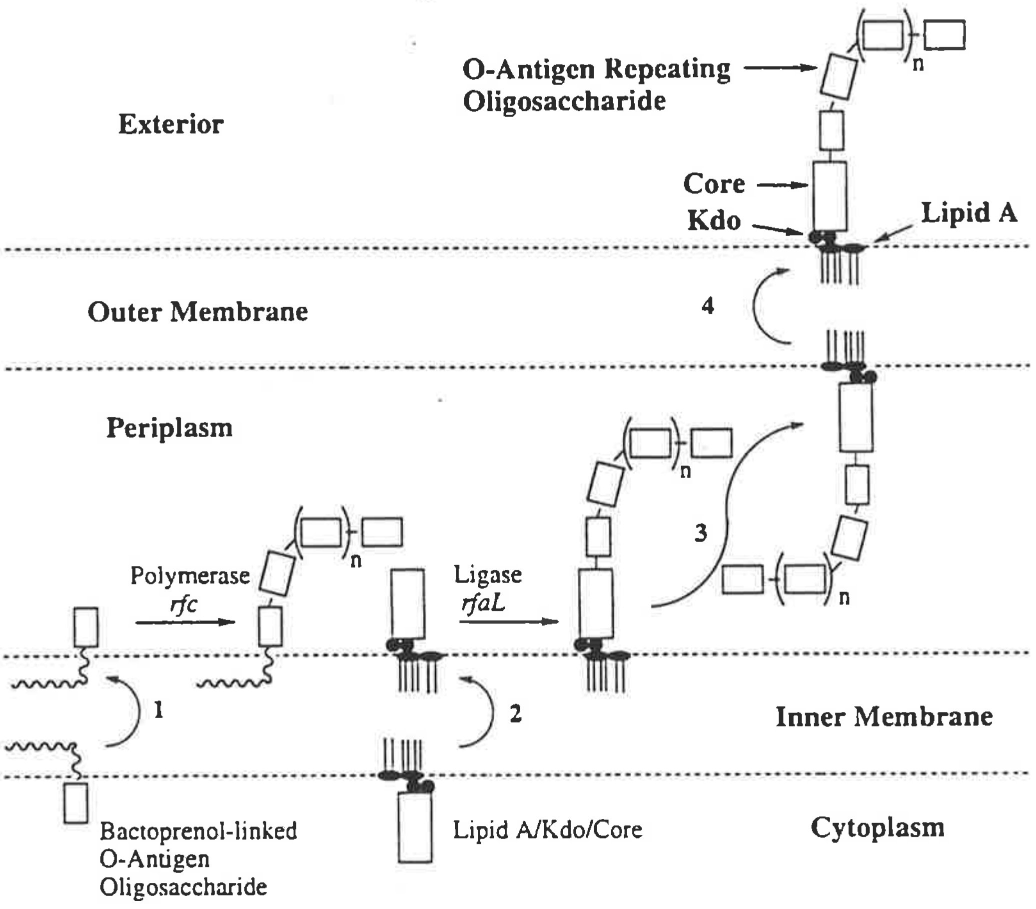
The location of the synthesis of the LPS molecule has not been totally elucidated. It is thought that the core is assembled directly onto the lipid A and that the O-antigen units are synthesised on the lipid carrier at the cytoplasmic face of the inner membrane. This is probably due to the energy requirements of the reactions during O-antigen biosynthesis. It has been shown that *S. enterica* sv Typhimurium requires energy for O-antigen assembly on the lipid carrier (Marino *et al.*, 1991).

Once the individual components are synthesised, the core and the O-antigen units are translocated across the cytoplasmic membrane where the lipid-linked O-units are

Figure 1.5: Model of LPS biosynthesis and assembly.

The biosynthesis and assembly of lipopolysaccharide occurs via a number of steps:

Lipid A/Kdo/Core are synthesised at the surface of the cytoplasmic membrane and translocated to the periplasmic side (2). O-antigen units are synthesised on the lipid carrier and transported to the periplasm (1), and are then polymerised by *rfc*. The polymerised polymer is ligated to lipid A-core by *rfaL*. Subsequently, the finished LPS molecule is exported to the cell surface (3 and 4). Diagram from Raetz (1996).


O-Antigen Repeating Oligosaccharide
n
Exterior
Core
Kdo
Lipid A
Outer Membrane
4
Periplasm
n
3
n
Polymerase
rfc
n
Ligase
rfaL
1
2
Inner Membrane
Cytoplasm
Bactoprenol-linked O-Antigen Oligosaccharide
Lipid A/Kdo/Core

polymerised into correct length chains (McGrath and Osborn, 1991). Two potential pathways have been proposed for the synthesis of LPS O-polysaccharides. Both pathways are initiated by similar reactions and they involve common precursors and intermediates (Whitfield, 1995). The differences lie in the cellular location of the polymerisation reactions and the direction of polymer growth (Whitfield, 1995). Both of these pathways are discussed to some depth later in this chapter. The pre-formed lipid A/core and the newly polymerised chains are then ligated together at the periplasmic face of the plasma membrane (Måkelå and Stocker, 1984). Little is known about the ligation reaction although the ligase, encoded by *rfaL/waaL* has been identified in *S. enterica* sv Typhimurium (MacLachlan *et al.*, 1991), *E. coli* (Roncero and Casadaban, 1992) and *S. flexneri* (Hong and Payne, 1997). After ligation the completed LPS molecule is translocated to the cell surface by unknown mechanisms (Fig. 1.5).

The following sections describe aspects of the different components of LPS molecules. The examples given relate to the most extensively studied organisms although reference is also made to the human pathogen *Vibrio cholerae* as it is related to *V. anguillarum*, the focus of the research in this thesis.

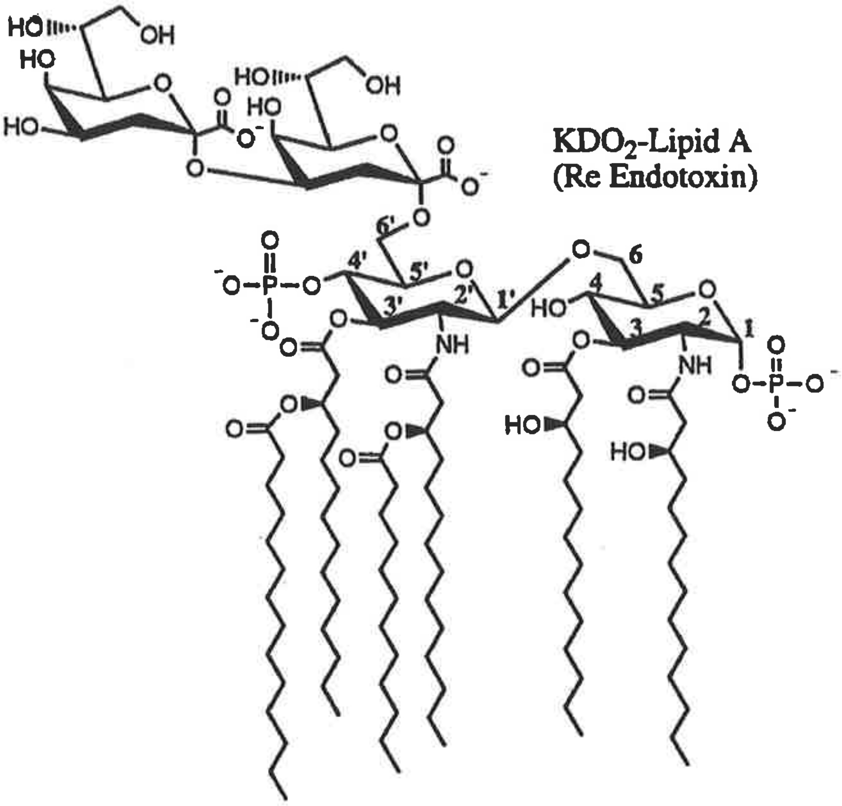
1.5.2 Lipid A

The lipid A component of LPS has been previously described as endotoxin and has been associated with septic shock (Morrison and Ryan, 1987). The biosynthesis of lipid A involves synthesis of sugars and fatty acids. The lipid A of *E. coli* and *Salmonella* consists of a β-1-6 linked disaccharide of glucosamine that is phosphorylated and acylated with R-3-hydroxymyristate which is further substituted with fatty acids (Raetz, 1993; Takayama, 1983) (Fig. 1.6). The genes required for lipid A synthesis in *E. coli* have been cloned and sequenced (Raetz, 1996). The isolation of lipid A mutants in *E. coli* has allowed the

Figure 1.6: Lipid A structure.

The chemical structure for the lipid A component of lipopolysaccharide from *E. coli* K-12.

Abbreviation: Kdo, 3-deoxy-D-manno-octulonic acid. Diagram from Raetz (1996).



KDO$_2$-Lipid A
(Re Endotoxin)

complex biosynthetic pathway to be constructed (Nishijima and Raetz, 1979; Takayama *et al.*, 1983; Anderson *et al.*, 1985). The lipid A subunit and core oligosaccharide of the LPS molecule are linked via 3-deoxy-D-manno-octulosonic acid (KDO) (Strain *et al.*, 1983a, b). In *E. coli* the minimal LPS molecule for cell viability consists of lipid A and KDO disaccharide (Boman and Monner, 1975; Rietschel, 1983). The importance of lipid A in *S. typhimurium* has been further elucidated recently with the discovery that substitutions on the lipid A moiety are vital for virulence (Guo *et al.*, 1997; Khan *et al.*, 1998). A mutation in the *rfaN/waaN* gene which is responsible for the production of an enzyme that catalyses one of two secondary acylation reactions to complete lipid A biosynthesis results in a strain that is essentially identical to wildtype except that it is less lethal (Khan *et al.*, 1998). The mutant produces O-antigen that has altered O-chain length suggesting a relationship between the different components of LPS biosynthesis. O-antigen chain length is regulated by the Rol/Wzz protein which is discussed later in this chapter (see Section 1.5.4.1.4). It has also been shown that differences in lipid A substitutions have an affect on O-antigen chain length in *S. typhimurium* (Baker *et al.*, 1999). The RfaN/WaaN mutant is not as virulent as the wild type, providing direct evidence for the role of lipid A in mouse typhoid infections (Khan *et al.*, 1998). Guo *et* al. (1997) have reported that responses to environmental changes via the *phoP/phoQ* regulatory system in *Salmonella typhimurium* result in altered fatty acid substitutions of lipid A which correlates to decreased expression of cytokines within host tissues. These environmentally induced differences may be a mechanism for bacteria to gain an advantage within the host.

The lipid A of *V. cholerae* is very similar to other Gram-negative lipid A moieties. The lipid A consists of β-1,6-linked D-glucosamine oligosaccharide which is ester-linked and substituted with pyrophosphorylethanolamine. The oligosaccharide is further substituted with a number of fatty acids including tetradecanoic, hexadecanoic and 3-D hydroxydecanoic acid (Raziuddin, 1977; Broady *et al.*, 1981). The lipid A of *V. cholerae* is also linked to the core via KDO but there has been some controversy as to whether one or two KDO residues are present (Brade, 1985).

1.5.3 Core oligosaccharide

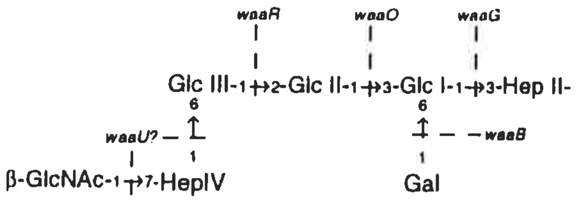
The composition of the core oligosaccharide is highly variable and is specific to the genus/species although similar sugars are found in different core regions. In *E. coli* and *Salmonella* the core region can be divided into the inner core which consists of three heptose units (Coleman and Lieve, 1979; Coleman, 1983) and the outer core containing varying amounts of *N*-acetylglucosamine, glucose and galactose sugars. At present five core types have been identified for *E. coli* (K-12, R1-R4) (Fig. 1.7). The core types are classified according to the substitutions found on the core backbone which may include phosphates, phosphorylethanolamine and other sugars (Heinrichs *et al.*, 1998). In *V. cholerae* the core oligosaccharide is composed of glucose, heptose, fructose, ethanolamine and N-acetyl glucosamine (Kondo *et al.*, 1988; Hisatsune *et al.*, 1989). The structure of the core oligosaccharide is represented in Figure 1.8 (Vinogradov *et al.*, 1995).

The genes responsible for the biosynthesis of the core oligosaccharide are located in the *rfa/waa* locus. The *rfa/waa* genes have been cloned from *Salmonella* and *E. coli* (Austin *et al.*, 1990) and the loci are highly conserved which reflects the high homology between the core structures of these organisms. The differences observed between the *rfa/waa* loci of these organisms is reflected in the chemical composition of their respective core oligosaccharides (Schnaitman and Klena, 1993). The genes involved in core biosynthesis in *E. coli* can be divided into three groups and are found clustered together in distinct operons around the chromosome (Roncero and Casadaban, 1992; Clementz, 1992). The groups are:

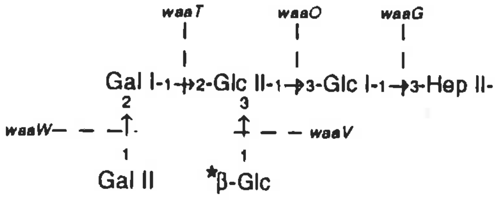
Figure 1.7: The core-oligosaccharide types of *E. coli.*

The structural formulae of the core oligosaccharide types, K-12, R1-R4 are represented. The genes responsible for each linkage are indicated. Abbreviations: Glc, glucose; Hep, L-glyero-D-*manno*-heptose; Gal, galactose; GlcNAc, N-acetyl glucosamine. Diagram adapted from Heinrichs *et al.* (1998).

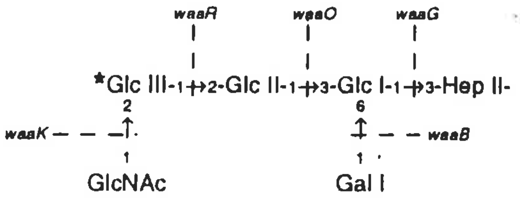
E. coli K-12



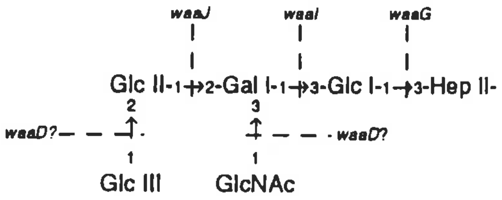
E. coli R1



E. coli R2



E. coli R3



E. coli R4

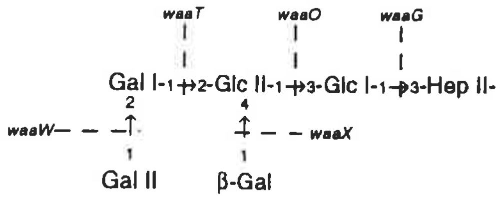
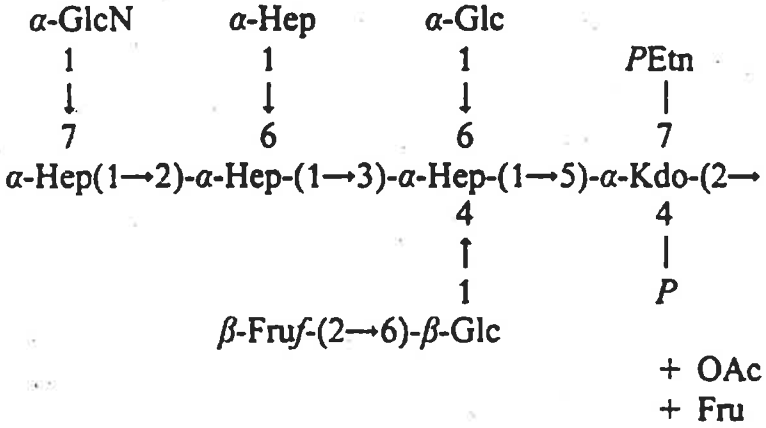


Figure 1.8: The core-oligosaccharide types of *V. cholerae* O1.

The chemical formula of the core-oligosaccharide from *V. cholerae* O1. Abbreviations: Glc, glucose; Hep, L-glyero-D-*manno*-heptose; Fru, fructose; Kdo, 3-deoxy-D-manno-octulonic acid; PEtn, phosphoethanolamine. Diagram from Knirel *et al.* (1997)



$$\overset{\begin{array}{c}\alpha\text{-GlcN}\\1\\\downarrow\\7\end{array}}{\alpha\text{-Hep}}(1\rightarrow2)\text{-}\overset{\begin{array}{c}\alpha\text{-Hep}\\1\\\downarrow\\6\end{array}}{\alpha\text{-Hep}}\text{-}(1\rightarrow3)\text{-}\underset{\begin{array}{c}4\\\uparrow\\1\\\beta\text{-Fru}f\text{-}(2\rightarrow6)\text{-}\beta\text{-Glc}\end{array}}{\overset{\begin{array}{c}\alpha\text{-Glc}\\1\\\downarrow\\6\end{array}}{\alpha\text{-Hep}}}\text{-}(1\rightarrow5)\text{-}\underset{\begin{array}{c}4\\|\\P\end{array}}{\overset{\begin{array}{c}P\text{Etn}\\|\\7\end{array}}{\alpha\text{-Kdo}}}\text{-}(2\rightarrow$$

+ OAc
+ Fru

1. genes involved in sugar biosynthesis,
2. genes encoding transferases,
3. regulatory genes.

Mutations within the *rfa/waa* locus render the bacteria rough (ie no core or O-antigen, R-LPS) and have allowed for extensive study. R-LPS mutants may have varying lengths of core which has led to numerous chemotypes of R-LPS being described (Fig. 1.9) (Lindberg, 1977).

The most studied gene involved in core biosynthesis is *rfaD/gmhD*. *RfaD/gmhD* encodes a ADP-1-glycero-D-mannoheptose-6-epimerase, which is involved in heptose biosynthesis and converts ADP-D-glycero-D-mannoheptose to ADP-L-glycero-D-mannoheptose (Coleman, 1983; Peques *et al.*, 1990). Analysis of this gene in *E. coli* has shown it to be a 37 kDa protein which has a βαβ fold at the N-terminus (Peques *et al.*, 1990) consistent with an ADP-binding domain (Rossman *et al.*, 1974; Wierenga *et al.*, 1986). In *E. coli rfaD/gmhD* is found in an operon along with *rfaF/waaF* and *rfaC/waaC*, which are also involved in heptose biosynthesis, and *rfaL/waaL*, the putative ligase thought to ligate the O-antigen chains to the core oligosaccharide (Peques *et al.*, 1990; Schnaitman *et al.*, 1991; Roncero and Casadaban, 1992; Sirisena *et al.*, 1992, 1994).

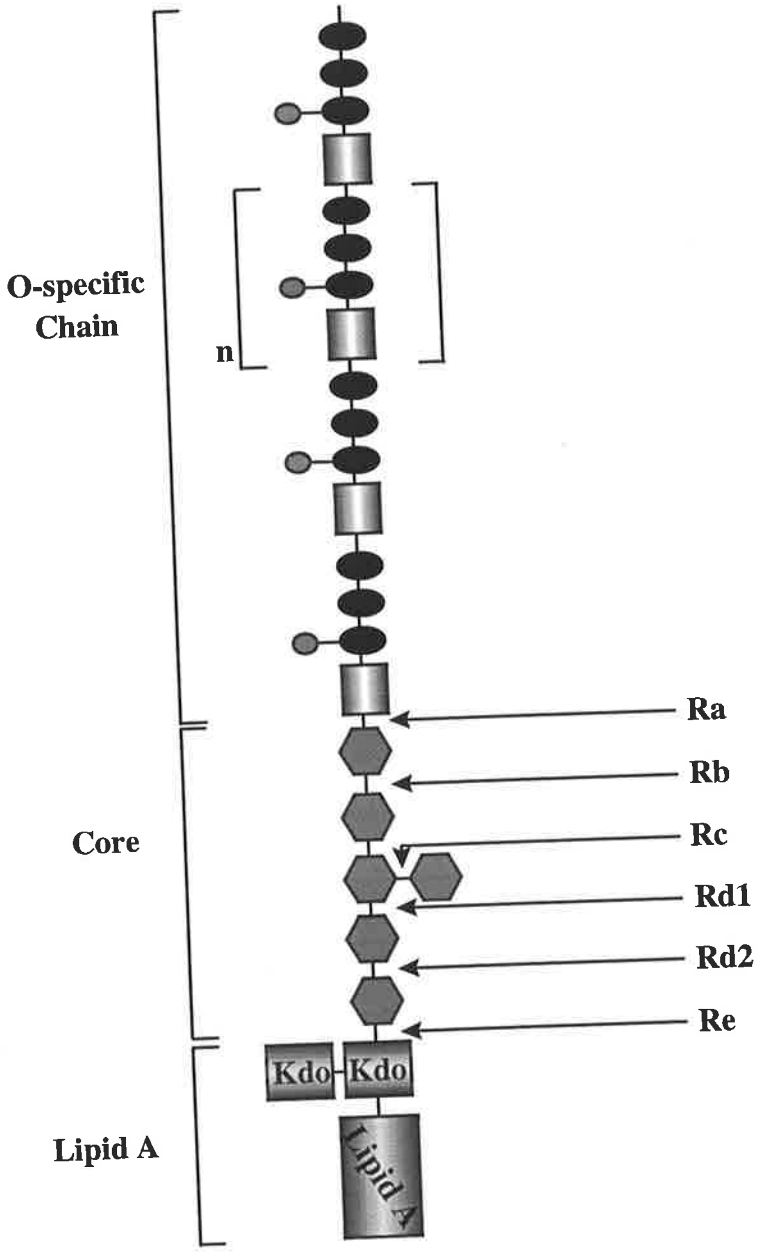
A *rfaD/gmhD* homologue was identified in *V. cholerae* and shows a high homology to the *E. coli rfaD/gmhD* gene (Stroeher *et al.*,1995b). In *V. cholerae rfaD/gmhD* is closely linked to the O-antigen biosynthesis locus and is not a part of an *rfa/waa* locus.

1.5.4 O-antigen

The O-antigen component of the LPS exposed on the surface of the cell and is often the most abundant component. O-antigen is highly variable both within and between bacterial species. The sugars present and their linkage form the basis for variation. The

Figure 1.9: Core oligosaccharide chemotypes of *E. coli*

A schematic representation of the LPS molecule from *E. coli* K-12. The proposed chemotypes of the core oligosaccharide are indicated on the left hand side of the figure to allow comparison. n corresponds to the number of O-repeats. Abbreviations: GlcN, D-glucosamine; Kdo, 3-deoxy-D-manno-octulonic acid; Hep, L-glyero-D-*manno*-heptose; Glc, glucose; Gal, galactose; GlcNAc, N-acetyl glucosamine; Rha, L-rhamnose; Gal*f*, D-galactofuranose; P, phosphate; Ac, acetate. Diagram adapted from Lindberg *et al.* (1991) and Raetz (1996).


O-specific
Chain
n
Ra
Rb
Rc
Rd1
Rd2
Re
Core
Kdo
Kdo
Lipid A
Lipid A

antigenic properties of O-antigen has been long known, with the variation being the basis of serotyping. In *Salmonella enterica* there are 60 known forms of O-antigen and 160 in *E. coli* (Reeves, 1993).

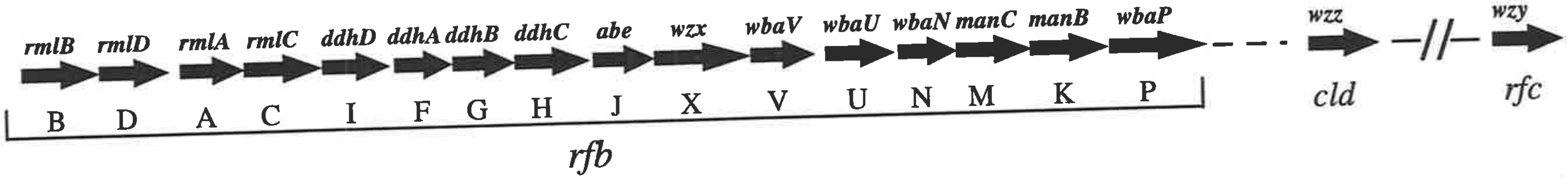
The genetics of O-antigen biosynthesis is best characterised in *Salmonella enterica* sv Typhimurium although with recent advances in molecular biology there is a growing understanding of O-antigen genetics and biosynthesis in a wider variety of species. The genes involved in O-antigen biosynthesis are usually clustered together in a single locus. Recently, an updated nomenclature system for genes involved in polysaccharide biosynthesis was devised by Reeves *et al.*, 1996. In this thesis, the new nomenclature system will be employed, however for clarity the old gene names will also be indicated. The new system allows: 1) genes with the same function to have the same name, 2) genes involved in O-antigen or capsular biosynthesis to be identified by the form used, and 3) genes involved in synthesis of sugars will be identified by the biosynthetic pathway. The new nomenclature can be identified by the form *w**** with O-antigen genes being given the symbols of *wb**, with the * being a specific letter that represents a particular species.

1.5.4.1 *Salmonella enterica* sv Typhimurium (group B)

The O-polysaccharide of group B *Salmonella* consists of a repeat unit of four hexoses (abequose, mannose, rhamnose and galactose). The O-unit is assembled from four nucleotide sugars (CDP-abequose, GDP-mannose, dDTP-rhamnose and UDP-galactose) on the lipid carrier, undecaprenol phosphate. The genes required to synthesise these sugars are located in the O-antigen biosynthesis loci at 42 min on the genetic map in strain LT2 (Sanderson and Roth, 1988). The operon consists of 16 genes (Fig. 1.10) including those involved directly with nucleotide sugar biosynthesis. Numerous other genes are also

Figure 1.10: The *rfb/wba* region of *S. enterica*, group B.

The *rfb/wba* region of *S. enterica* (group B) consists of 16 genes and is located at 42 min on the chromosome. The letters below the operon are the original names, those above the operon represent the present nomenclature as designated by Reeves *et al.* (1996). The diagram was adapted from Reeves *et al*, (1996).


rmlB
rmlD
rmlA
rmlC
ddhD
ddhA
ddhB
ddhC
abe
wzx
wbaV
wbaU
wbaN
manC
manB
wbaP
wzz
wzy
B
D
A
C
I
F
G
H
J
X
V
U
N
M
K
P
cld
rfc
rfb

required for correct assembly of the O-antigen which includes the sugar transferases (*rfb*N/*wbaN*, *rfb*U/*wbaU*, *rfb*V/*wbaV*, *rfb*P/*wbaP*).

Within *Salmonella,* the O-antigen is highly polymorphic. This variation is thought to be important in the pathogenesis of these bacteria (Reeves, 1993). Not all forms of O-antigen are associated with pathogenicity, with *Salmonella* groups A, B and D being the most pathogenic. These groups possess similar O-antigen structures (Roantree, 1967) and the O-antigen biosynthesis regions are genetically related (Jiang *et al.*, 1991).

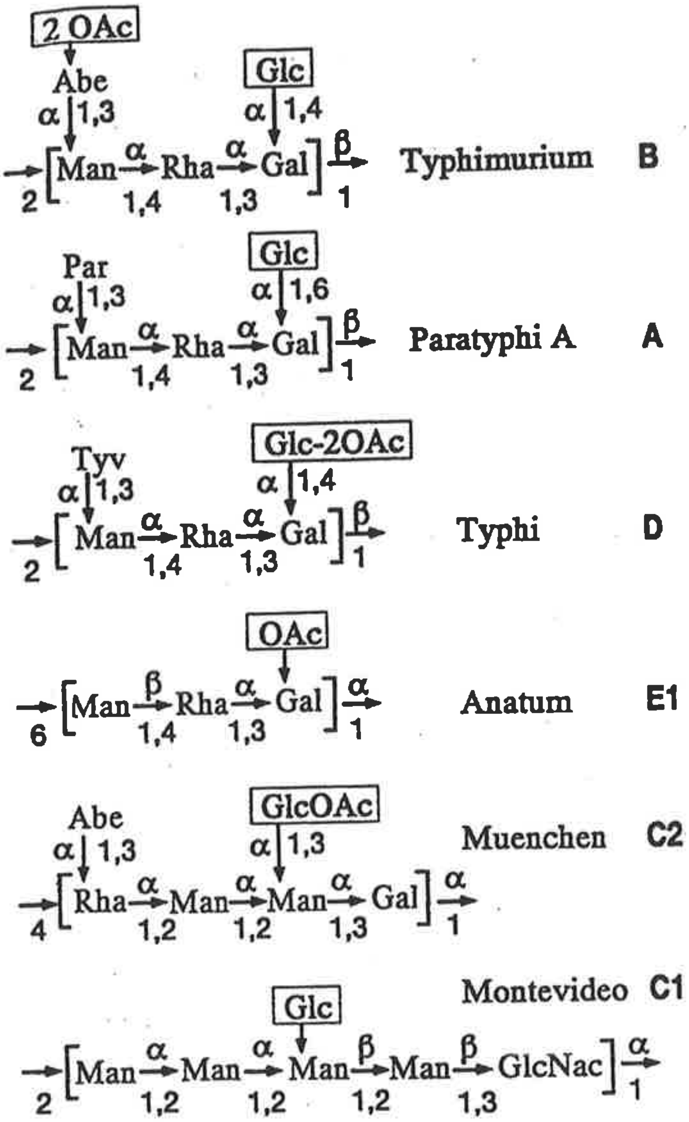
Salmonella groups A (sv Paratyphi), B (sv Typhimurium) and D (Typhi) have paratose, abequose and tyvelose respectively, as the side chain sugar in their O-antigen which are otherwise identical (Fig. 1.11). To these differences there are unique genes involved in the biosynthesis of these nucleotide sugars. The gene *rfb*J/*abe* from group B encodes an abequose synthetase (Wyk and Reeves, 1989), *rfb*S/*prt* is a paratose synthetase found in the group A biosynthesis region, and *rfb*E/*tyv* from group D encodes a CDP-tyvelose epimerase which converts CDP-paratose to CDP-tyvelose (Verma and Reeves, 1989; Liu *et al.*, 1991).

1.5.4.1.1 Biosynthesis of nucleotide sugars

The genes involved in the biosynthesis of the nucleotide sugars, with the exception of UDP-galactose, are located within the O-antigen biosynthesis region The dGDP-mannose pathway requires a gene that is also located outside of the O-antigen biosynthesis operon. This gene (*pmi)* is responsible for the initiation of the mannose biosynthetic pathway and converts fructose-1-phosphate to mannose-6-phosphate (Sanderson and Roth, 1988). The remainder of the genes required for mannose biosynthesis, *rfb*K/*manB* and *rfb*M/*manC* are located in the biosynthesis region (Reeves, 1993).

Figure 1.11: Structures of the repeat units of *S. enterica* groups B, A, D, E1, C2, C1.

The chemical formulae for the repeat units of the individual groups of *S. enterica*. Abequose (Abe) is present in group B and C2, paratose (Par) in group A and tyvelose (Tyv) in group D. Other sugars are represented by the following: Gal, galactose; GlcNAc, N-acetyl glucosamine; Man, mannose; Rha, rhamnose. Diagram from Reeves, (1993).


2 OAc
Abe
α 1,3
Glc
α 1,4
2 [Man α 1,4 Rha α 1,3 Gal] β 1
Typhimurium
B
Par
α 1,3
Glc
α 1,6
2 [Man α 1,4 Rha α 1,3 Gal] β 1
Paratyphi A
A
Tyv
α 1,3
Glc-2OAc
α 1,4
2 [Man α 1,4 Rha α 1,3 Gal] β 1
Typhi
D
OAc
6 [Man β 1,4 Rha α 1,3 Gal] α 1
Anatum
E1
Abe
α 1,3
GlcOAc
α 1,3
4 [Rha α 1,2 Man α 1,2 Man α 1,3 Gal] α 1
Muenchen
C2
Montevideo
C1
Glc
2 [Man α 1,2 Man α 1,2 Man β 1,2 Man β 1,3 GlcNac] α 1

dTDP-rhamnose is synthesised from glucose-1-phosphate. The first two enzymes involved in the dTDP-rhamnose biosynthetic pathway are encoded by *rfbA/rmlA* and *rfbB/rmlB*. RfbA/RmlA is a glucose-1-phosphate thymidyltransferase and RfbB/RmlB is the enzyme dTDP-glucose-4,6-dehydratase. It is possible to distinguish between these enzymes using established enzyme assays (Jiang *et al.*, 1991).

The third and fourth enzymes involved in the biosynthesis of rhamnose are dTDP-4-keto-L-rhamnose-3,5-epimerase (*rfbC/rmlC*) and dTDP-L-rhamnose synthetase (*rfbD/rmlD*) (Jiang *et al.*, 1991). Figure 1.12 shows the rhamnose biosynthetic pathway. The four genes involved in dTDP-rhamnose biosynthesis are highly conserved between *S. enterica*, *Shigella dysenteriae* 1 and *E. coli* K-12 (Jiang *et al.*, 1991).

The major side chain sugars of groups A, B and D share a common biosynthetic pathway (Fig. 1.13). The five genes involved in dCTP-abequose biosynthesis, *rfbF/ddhA*, *rfbG/ddhB*, *rfbH/ddhC*, *rfbI/ddhD* and *rfbJ/abe*, are located in sv Typhimurium. In sv Paratyphi and sv Typhi, the *rfbJ* gene is replaced with *rfbS/prt*. *rfbJ* and *rfbS* have similar functions and are homologous, but have diverged to be only 26% identical at the amino acid level (Verma and Reeves., 1989). In addition both have *rfbE/tyv* which is involved in the conversion of CDP-paratose to CDP-tyvelose. However, the *rfbE* gene in sv Paratyphi has a frameshift mutation and is therefore non-functional preventing the conversion of paratose to tyvelose (Verma and Reeves, 1989).

1.5.4.1.2 O-antigen assembly and biosynthesis

O-antigen oligosaccharide is synthesised on the lipid carrier (undecaprenol phosphate or bactoprenol) before the oligosaccharide is polymerised into O-antigen chains. The polymerised chains are then translocated to lipid A/core.

Figure 1.12: The biosynthetic pathway for the production of rhamnose.

The biosynthetic pathway of rhamnose is represented and the genes responsible for each step are indicated. The new nomenclature for the genes are listed below as described by Reeves *et al*, (1996). *rfbA=rmlA*; *rfbB=rmlB*; *rfbC=rmlC*; *rfbD=rmlD*. Diagram from Raetz, 1996.

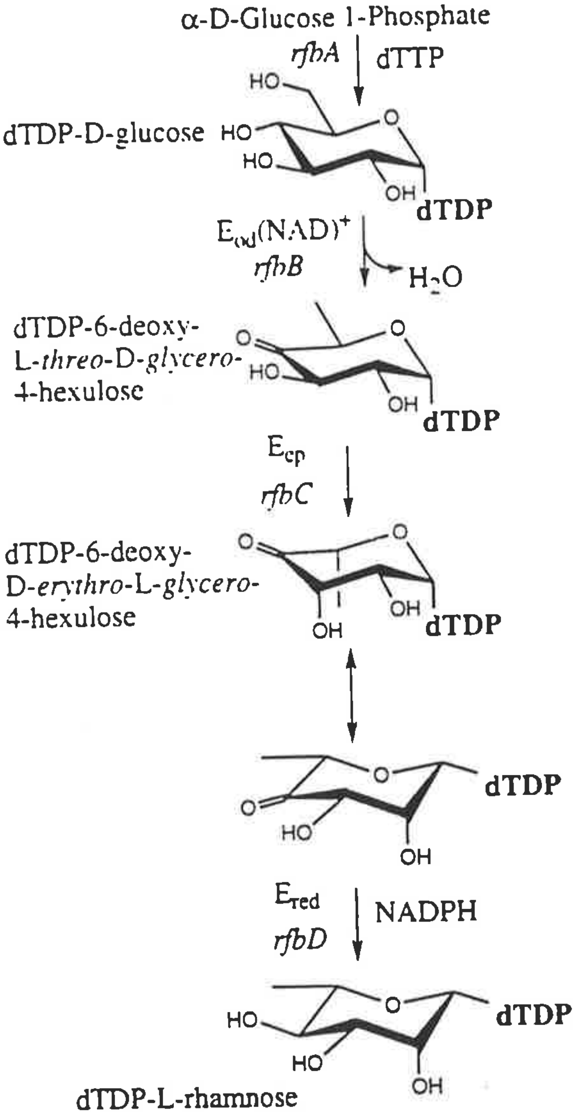
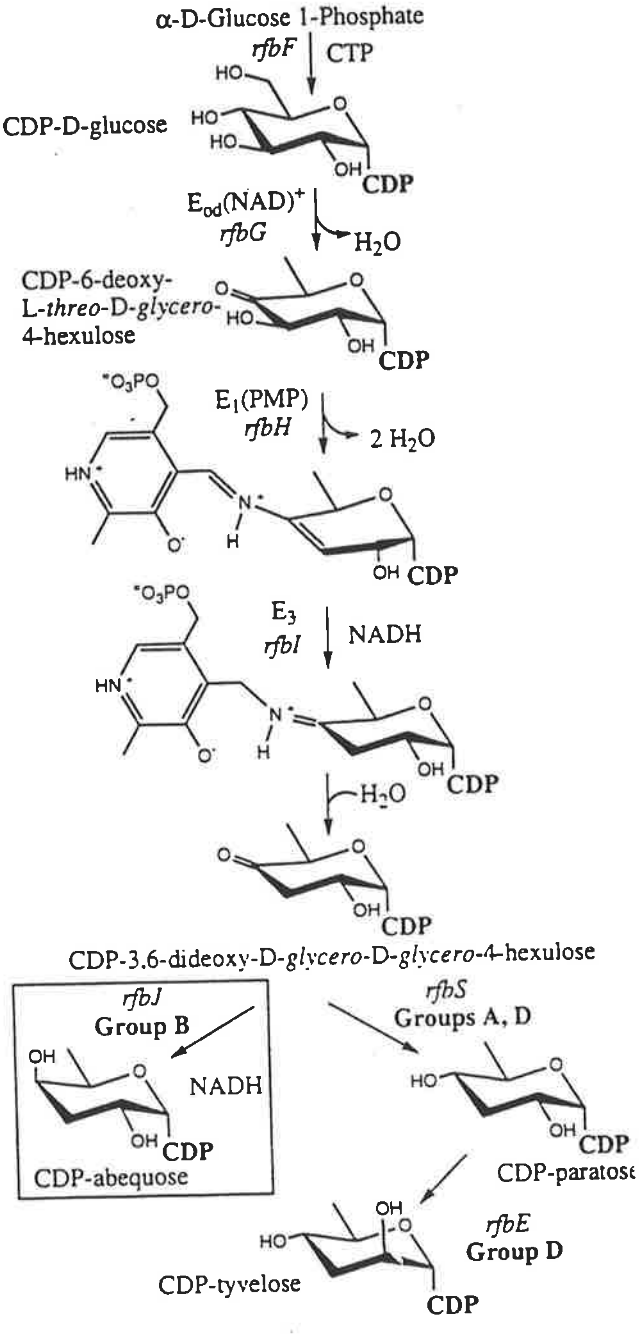

α-D-Glucose 1-Phosphate
rfbA
dTTP
dTDP-D-glucose
HO
HO
HO
O
OH
dTDP
E(ox)(NAD)+
rfbB
H2O
dTDP-6-deoxy-L-threo-D-glycero-4-hexulose
O
HO
O
OH
dTDP
Eep
rfbC
dTDP-6-deoxy-D-erythro-L-glycero-4-hexulose
O
O
OH
OH
dTDP
O
O
dTDP
HO
OH
Ered
rfbD
NADPH
HO
O
dTDP
HO
OH
dTDP-L-rhamnose

Figure 1.13: The biosynthetic pathway for the production of the CDP-3,6 dideoxyhexose sugars.

The biosynthetic pathway for the synthesis of the sugars CDP-abequose, CDP-paratose, CDP-tyvelose is represented and the genes involved at each step are indicated. The new nomenclature for the genes are listed below as described by Reeves *et al.* (1996). *rfbF=ddhA*, *rfbG=ddhB*, *rfbH=ddhC*, *rfbI=ddhD*, *rfbS=prt*, *rfbE=tyv*, *rfbJ=abe*. Diagram from Raetz, (1996).


α-D-Glucose 1-Phosphate
rfbF
CTP
CDP-D-glucose
CDP
E$_{od}$(NAD)$^+$
rfbG
H_2O
CDP-6-deoxy-L-threo-D-glycero-4-hexulose
CDP
E$_1$(PMP)
rfbH
2 H_2O
CDP
E$_3$
rfbI
NADH
CDP
H_2O
CDP
CDP-3.6-dideoxy-D-glycero-D-glycero-4-hexulose
rfbJ
Group B
NADH
CDP
CDP-abequose
rfbS
Groups A, D
CDP
CDP-paratos
rfbE
Group D
CDP-tyvelose
CDP

The initial reactions involved in the biosynthesis of heteropolysaccharide O-antigens are essentially identical within the *Salmonella enterica* serogroups. The initial reaction which is reversible and involves *rfbP/wbaP* (galactosyltransferase) catalyses the transfer of galactose to undecaprenal phosphate. The remaining sugars, TDP-rhamnose and GDP-mannose are then sequentially transferred to the undecaprenol diphosphate-galactose intermediate. Each reaction is catalysed by a specific transferase encoded in the O-antigen biosynthesis operon.

It is important to note that not all O-antigens contain galactose and therefore another initiating enzyme is required for O-antigen biosynthesis. The other known initiating enzyme is *rfe/wecA*, an N-acetyl-glucosamine-1-phosphatetransferase (Meier-Dieter *et al.*, 1992; Whitfield, 1995). Rfe/WecA transfers N-acetyl-glucosamine (Glc*p*NAc) found in heteropolysaccharide O-units of numerous *E. coli* serotypes (Alexander and Valvano, 1994), *Shigella dysenteriae* type 1 (Klena and Schnaitman, 1993) and *Shigella flexneri* (Yao and Valvano, 1994; Whitfield, 1995). Interestingly *rfe/wecA* has also been found associated with the synthesis of homopolysaccharide O-antigens in *E. coli* (Liu and Reeves, 1994; Jann *et al.*, 1982; Rick *et al.*, 1994), *Klebsiella pneumoniae* (Clarke and Whitfield, 1992) even though they do not contain Glc*p*NAc (Whitfield, 1995), the TI antigen of *Salmonella* (Måkelå and Stocker, 1984), and the enterobacterial common antigen (ECA) (Meier-Dieter *et al.*, 1990).

Rfe/WecA and RfbP/WbaP are thought to be structural and functional homologues in O-antigen initiation based on similar hydropathy profiles (Klena and Schanitman, 1993) even though Rfe/WecA is associated with synthesis of other surface polysaccharides, while RfbP is only associated with O-antigen biosynthesis (Whitfield, 1995).

1.5.4.1.3 Polymerisation of O-antigen repeat units: at the reducing terminus

O-antigens from *S. enterica* serogroups A,B,D and E are elongated by the addition of O-repeat units to the reducing end of the growing polymer. For polymerisation in a blockwise fashion to occur complete O-units are required. The repeat units are polymerised into long chains while attached to the lipid carrier (Weiner *et al.*, 1965; Robbins *et al.*, 1966) (Fig. 1.14) (Whitfield., 1995). This reaction requires *rfc/wzy*, which maps outside the O-antigen biosynthesis region in *S. enterica* sv Typhimurium (Naide *et al.*, 1965; Stocker and Måkelå, 1971; Collins and Hackett, 1991). This is unusual as other *rfc*-like genes in numerous other bacteria have been reported to be inside or adjacent the O-antigen biosynthesis locus (Brown *et al.*, 1992; Reeves, 1993; Xiang *et al.*, 1994; Morona *et al.*, 1994). Mutations that result in a *rfc/wzy* defect produce the semi-rough LPS phenotype (ie core oligosaccharide with a single O-antigen unit, SR-LPS) (Morona *et al.*, 1994).

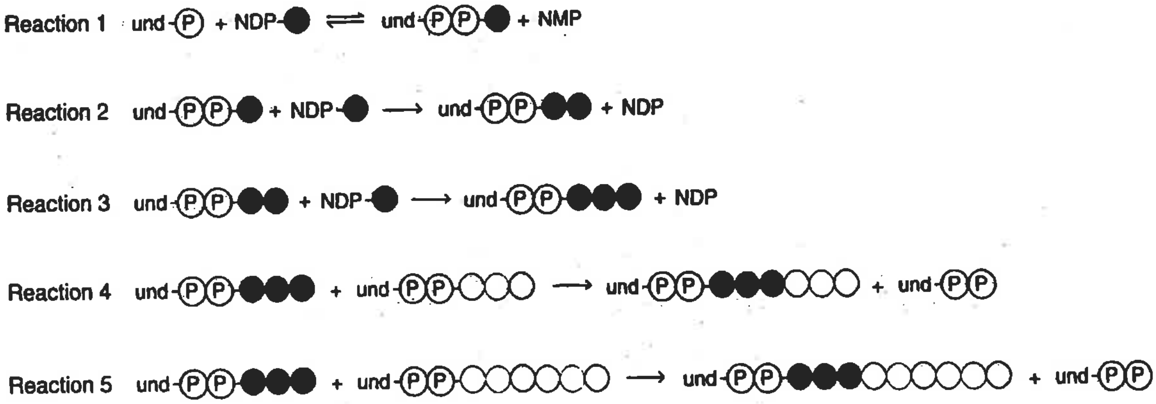
The *rfc/wzy* of *S. typhimurium* and *S. flexneri* have been cloned and sequenced (Collins and Hackett, 1991; Morona *et al.*, 1994). Rfc/Wzy homologues are predicted to be hydrophobic integral membrane proteins with 11-13 transmembrane segments and show little primary sequence homology (Morona *et al.*, 1994; Whitfield, 1995). Recently it was shown that the *rfc/wzy* of *S. flexneri* contains 12 transmembrane segments with two large periplasmic domains and the carboxy-terminal of the protein is located at the cytoplasmic face of the inner membrane (Daniels *et al.*, 1998).

1.5.4.1.4. *rol/cld/wzz*

It has been observed that *rfc*-mediated polymerisation involves the control of the distribution of O-antigen chain length to give a modal chain length distribution. Chain

Figure 1.14: Polymerisation of O-antigen from the reducing terminus.

The reversible initiating step (reaction 1) involves the addition of a sugar-1-phosphate residue from the nucleotide diphosphate (NDP) molecule to undecaprenol phosphate (und-P) the lipid carrier. Reaction 2 and 3 demonstrate the formation of an O-antigen with trisaccharide repeat unit by sequential transfer of each sugar to form und-P-P-linked O-unit. und-P-P linked O-units are the substrates for polymerisation. Polymerisation involves the addition of nascent O-polymer of one und-P-P carrier to the non-reducing terminus of a newly synthesised single O-unit attached to the second und-P-P carrier (reaction 4 and 5). The solid shaded sugars of the O-units indicate those units added most recently. Diagram adapted from Whitfield, (1995).


Reaction 1 und-P + NDP- ⇌ und-P-P- + NMP
Reaction 2 und-P-P- + NDP- ⟶ und-P-P- + NDP
Reaction 3 und-P-P- + NDP- ⟶ und-P-P- + NDP
Reaction 4 und-P-P- + und-P-P- ⟶ und-P-P- + und-P-P
Reaction 5 und-P-P- + und-P-P- ⟶ und-P-P- + und-P-P

length is regulated by proteins known as Rol (regulator of chain length) or Cld (chain length determinant). These proteins are also referred to as Wzz. The mechanism of action of Rol/Wzz is not understood.

The Rol/Wzz protein has been hypothesised to function as a molecular clock that interacts with the Wzy polymerase, so that modal distribution of chain length is determined by a fixed time during which the polymerase can extend the O-side chain, before transfer to the ligase, so that it can covalently link the O-antigen to lipid A-core (Bastin *et al.*, 1993). An alternative mechanism for how Rol/Wzz may function was described by Morona *et al* (1995) in which Rol/Wzz acts as a molecular chaperone which interacts with either the O-antigen ligase (RfaL/WaaL) or Rfc/Wzy allowing a fixed ration of RfaL/WaaL to Rfc/Wzy proteins resulting in a non-random O-antigen chain length. Rol mutants in *S. flexneri* result in random distribution of O-antigen chain length and the bacteria are avirulent (Van Den Bosch *et al.*, 1997).

Recently, numerous Rol/Wzz functional homologues have been described which all show a conservation in primary sequence as well as structure (Schnaitman and Klena, 1993; Morona *et al.*, 1995a). The proteins are characterised by two highly conserved potential transmembrane domains in the N- and C-terminal regions (Whitfield *et al.*, 1997). This leads one to suspect that *rol/wzz* proteins share a common mechanism.

It has been shown that heterologous complementation can occur with Rol/Wzz proteins. However the LPS pattern observed is Rol/Wzz specific (Bastin *et al.*, 1993; Batchelor *et al.*, 1991; Burrows *et al.*, 1997). This suggests that Rol doesn't recognise specific O-antigen sugars but rather interacts with other components of the O-antigen or LPS biosynthesis machinery.

1.5.4.1.5 *rfaL/waaL*

The putative O-antigen ligase encoded by *rfaL/waaL* is responsible for ligating the completed O-antigen chain to the lipid A-core. RfaL/WaaL is an adaptable enzyme that is capable of ligating high or low molecular weight polymers to the lipid A-core (Whitfield *et al.*, 1997). It is known that RfaL/WaaL requires completed lipid A-core molecules to function (Schnaitman and Klena, 1993). If incomplete lipid A-core molecules are produced, they are unable to accept the O-antigen chains which remain attached to the lipid carrier as O-antigen haptens are located in the perisplasm (Beckman *et al.*, 1964; Lindberg *et al.*, 1972). It has been postulated that for ligation to occur, RfaL/WaaL must recognise specific sites on both the core and lipid-carrier linked O-antigen. This is supported by experiments that show complementation of heterologous ligase activities is not possible, and that *rfaL/waaL* appears to require an additional gene product RfaK (WaaK) to function correctly (Whitfield *et al.*, 1997). This suggests that RfaL/WaaL either recognises species-specific LPS-core molecules modified by RfaK/WaaK or a protein:protein interaction is required between RfaK/WaaK and RfaL/WaaL. Despite what is known about the RfaL/WaaL protein, little is understood about the actual mechanism of ligation.

1.5.5 Polymerisation of O-antigen repeat units: at the non-reducing terminus

Another pathway exists for the synthesis of some O-antigens. This pathway does not involve either the Rol/Wzz or Rfc/Wzy proteins. This alternative pathway involves the synthesis of simple O-antigens, usually those that consist of linear homopolymers. Little is known about the details of this pathway which has been studied in *E. coli* O8 and O9, and *K. pneumoniae* O1 (Whitfield, 1995). To date, no heteropolysaccharide O-antigens have been identified using this pathway.

Initiation of O-antigen synthesis involves the transfer of a non-O-antigen residue to the lipid carrier that can then act as the acceptor for the monomers which will become the O-antigen chain. Rfe/WecA is the only known initiating enzyme that has thus far been identified for this pathway. Monomers are added to the acceptor intermediate at the non-reducing terminus, one residue at a time, in a progressive manner (Fig. 1.15) (Weisgerber and Jann, 1982).

O-antigen assembly in this Rol/Rfc independent pathway is different to the Rol/Rfc dependent pathway in the following three areas:

1) the cellular location of the polymerisation reactions,

2) the direction of chain growth during polymerisation, and

3) the requirement for Rfc/Wzy, Rol/Wzz and RfbX/Wzx (the flippase) in one pathway and ABC-2 transporter in the other.

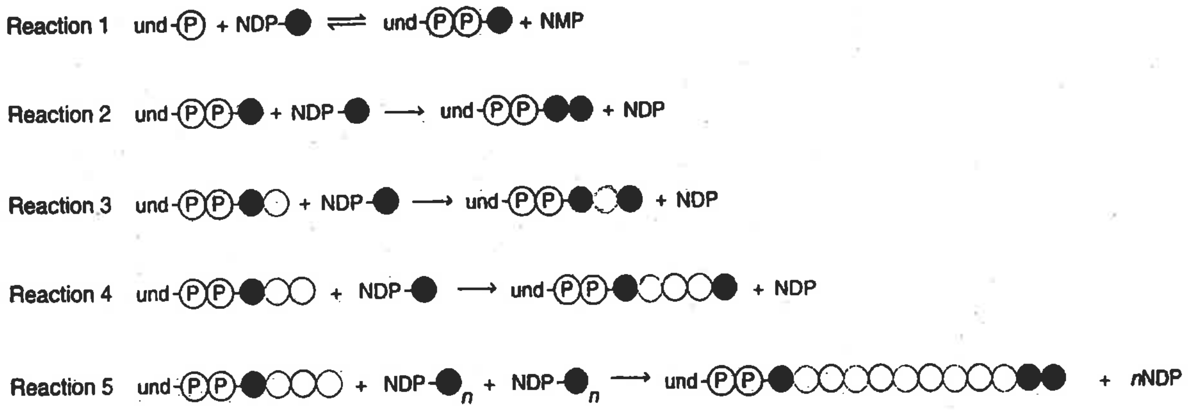
In the Rol/Rfc-dependent pathway, polymerisation of the O-antigen occurs at the reducing terminus at the periplasmic face of the cytoplasmic membrane. In contrast, the Rol/Rfc-independent pathway initiates polymerisation at the non-reducing terminus in the cytoplasm. The polymerised O-antigen chains are then transported across the membrane by an ATP-binding cassette "ABC-2" transporter. In the Rol/Rfc-dependent pathway, the O-antigen units are transported across the plasma membrane by RfbX/Wzx (also known as the flippase) where the O-units are polymerised into O-antigen chains on the periplasmic face of the membrane (Whitfield, 1995).

1.5.5.1 ABC-transporter family

The ABC integral membrane proteins fall into two distinct classes depending on the system in which they are found. Hence, the proteins are either involved in the export of LPS or capsule. The ABC transport systems of capsules require auxiliary proteins to aid in

Figure 1.15: Polymerisation of O-antigen from the non-reducing terminus.

Reaction 1 shows the initiation of polymerisation by the addition of a sugar-1-phosphate to the und-P carrier, in the synthesis of O-antigen with a disaccharide repeating unit. Polymerisation involves the sequential processive transfer of sugars to the non-reducing terminus of the nascent polymer (reaction 2 and 3). There is no block-wise polymerisation as seen in Fig. 1.14, and the polymer is elongated by the processive transfer enzymes (reactions 4 and 5). Sugars are transferred to the growing polymer directly from NDP-sugar precursors. Growth of the polymer occurs at the end of the polymer furthest from the und-P-P carrier (represented by solid circles). Diagram adapted from Whitfield (1995).


Reaction 1
und-P + NDP- ⇌ und-PP- + NMP
Reaction 2
und-PP- + NDP- ⟶ und-PP- + NDP
Reaction 3
und-PP- + NDP- ⟶ und-PP- + NDP
Reaction 4
und-PP- + NDP- ⟶ und-PP- + NDP
Reaction 5
und-PP- + NDP-n + NDP-n ⟶ und-PP- + nNDP

the transport of these high molecular weight molecules while the equivalent proteins in the O-antigen transport do not require these extra proteins as the molecule is smaller in size (Paulsen *et al.*, 1997). Phylogenic relationships of the integral membrane protein component of ABC transporters from both capsule and O-antigen systems have been examined and not surprisingly the proteins from each were found to be clustered together according to similar function, regardless of the organism of origin. Hence capsule proteins were clustered, as were the O-antigen representatives (Paulsen *et al.*, 1997). The LPS export proteins are from species such as: *Yersinia enterocolitica* O:3 (Zhang et *al.*, 1993), *Klebsiella pneumoniae* O1 (Bronner *et al.*, 1994; Whitfield, 1995) and *V. cholerae* O1 (Manning *et al.*, 1995), whereas capsule exporters are found in the group II-type capsule polysaccharides of *E. coli, Haemophilus influenzae* and *Neisseria meningitidis* (Pavelka *et al.*, 1989; Smith *et al.*, 1990; Kroll *et al.*, 1990; Frosch *et al.*, 1991).

In *K. pneumoniae* O1 two genes *rfbA*/*wzm* and *rfbB*/*wzt* are involved in the transport of the O-chain polymer of this organism which consists of the sugar D-Galactan. Figure 1.16 demonstrates how the O-antigen may be polymerised and transported across the membrane for ligation. The transporter is a member of the ABC transporter family and specifically is of the ABC-2 subfamily (Whitfield, 1995). ABC-2 transporters contain an integral membrane protein with multiple spanning domains containing six transmembrane helices and a hydrophilic protein containing the consensus ATP-binding motif (Walker box) (Ames *et al.*, 1992; Higgins., 1992).

1.5.6 A third O-antigen biosynthetic pathway

Recently, a third pathway of O-antigen biosynthesis has been discovered and partially characterised. This pathway is involved in the assembly of the poly-N-acetylmannosamine O-antigen (factor 54) of *S. enterica* sv Borreze (Keenleyside and

Figure 1.16: Model for the transport of O-antigens elongated from the non reducing terminus.

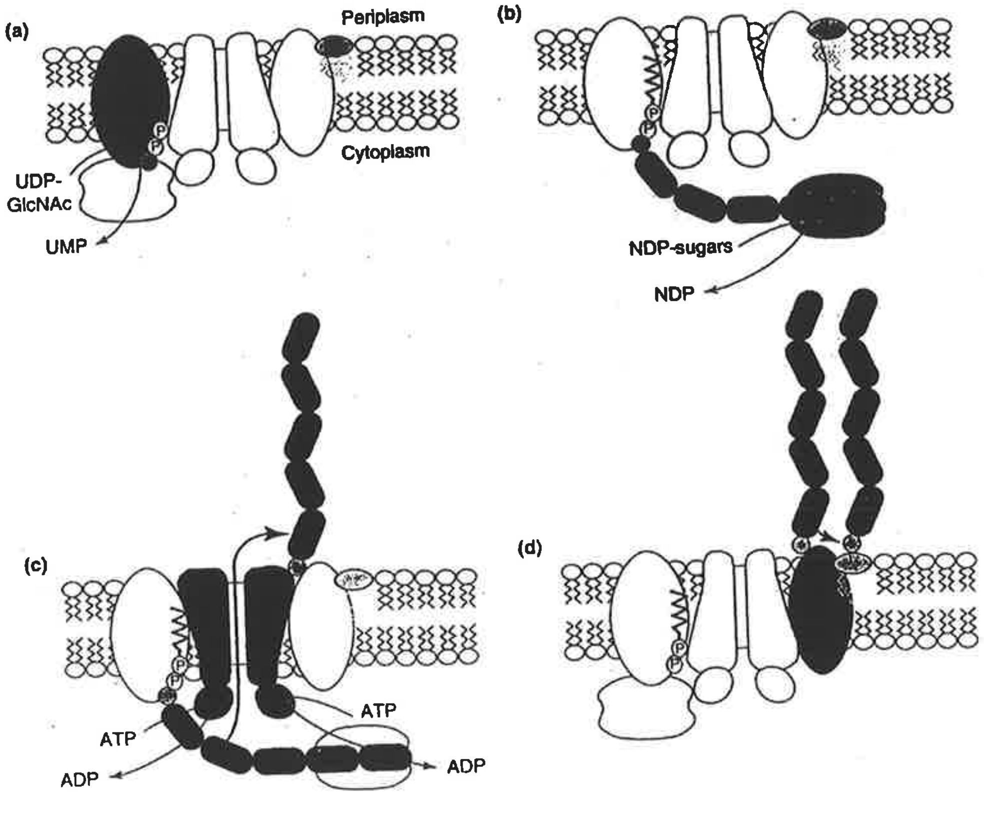
a) Rfe/WecA (black integral membrane protein) initiates polymerisation of the O-antigen polymer,

b) The polymerisation of the polymer involves numerous enzymes. The und-P-P linked polymer is completed in the cytoplasm,

c) Translocation of the polymer across the inner membrane requires the presence of a ATP-binding transporter (ABC-2; Wzm/Wzt).

d) Ligation of the O-antigen to the lipid A-core on the periplasmic face of the inner membrane.

Diagram from Whitfield (1995).


(a)
Periplasm
Cytoplasm
P
P
UDP-
GlcNAc
UMP
(b)
P
NDP-sugars
NDP
(c)
P
P
ATP
ADP
ATP
ADP
(d)
P
P

Whitfield, 1996; Whitfield *et al.*, 1997). This O-antigen is the first homopolymer described in *Salmonella* (Knirel and Kochetkov, 1994). This pathway is similar to other homopolymer systems and requires Rfe/WecA to initiate O-unit synthesis, is Rfc/Wzy-independent, and hence similar to the Rfc/Wzy-independent pathway described above. The O-antigen in this organism is not transported with either an ABC-2 transporter or RfbX/Wzx. Therefore, O:54 synthesis represents a new pathway for O-antigen assembly involving a different mechanism for delivering O-polysaccharide to the LPS O-antigen ligase (Keenleyside and Whitfield, 1996). To date, no export mechanism has been elucidated for this O-antigen pathway. However, Keenleyside and Whitfield (1996) suggested that the O-antigen synthase (WbbF/RfbB) protein of O54 may have a dual function, and therefore may also act as the O-antigen transporter for this system.

1.5.7 LPS export

Although many mechanisms involved in LPS biosynthesis are slowly being determined with better techniques and technology an exception is LPS export or the surface expression of complete LPS molecule, as not much is known about this process. It is known that once the complete LPS molecule is surface expressed it can not be removed from the outer membrane back into the cell proper (Osborn *et al.*, 1972a, b; Jones and Osborn, 1977). There have been a number of models proposed for the export of the LPS molecule across the outer membrane including:

1) the flip-flop of O-antigen, core-lipid A or complete LPS across the membrane

2) budding of vesicles containing LPS and

3) the continuity between the inner and outer membranes involving a filter of LPS (Bayer's Junction, zones of adhesion). To date there is little evidence to support any of these systems.

The only proteins that have been extensively studied which may play a role a in the export of LPS are *htrB* and *msbA* (Polissi and Georgopoulos, 1996; Zhau *et al.*, 1998). It is thought that these genes are at the very least involved in the transport of nascent lipid A-core molecules to the outer membrane. Mutations in both of these genes results in the accumulation of N-acetyl [3H]-glucosamine (a LPS precursor in *E. coli*) in the inner membrane. It has been proposed that *msbA* and/or *htrB* are involved in the export of completed LPS molecules (lipid A- core- O-antigen) to the surface of the cell even though there is no direct evidence for this function (Polissi and Georgopoulos, 1996).

Recently, further evidence for export of capsules in *E. coli* has shown that translocation of group I and II capsules occurs at specific sites where the plasma and outer membrane appear to come in close when examined by electron microscopy (Whitfield and Roberts, 1999). These sites have been previously described as "Bayer's Junctions" and has been subject to some controversy (Whitfield and Valvano, 1993). It now seems from most recent evidence from mutations in the different components of the transport system that translocation involves periplasmic scaffolds comprising transient translocation machinery that exports molecules to the outer membrane (Whitfield and Roberts, 1999).

1.5.8 *V. cholerae* O-antigen

1.5.8.1 Genetics of *V. cholerae* O1 O-antigen biosynthesis

V. cholerae O1 can be divided into two biotypes, classical and El Tor. These biotypes can be further divided into three serotypes, (Inaba, Ogawa and Hikojima) which are based on determinants found on the O-antigen of the LPS. The three serotypes share a common antigenic determinant referred to as the A-antigen. Two specific antigens B and C are found in varying amounts on the different serotypes; Inaba expresses only C, while

Ogawa strains express B and reduced levels of C (Burrows *et al.*, 1946a, b; Sakazaki and Tamura, 1971; Redmond *et al.*, 1973; Redmond, 1979). The third serotype, Hikojima is very unstable and is now thought to belong to the Ogawa serotype (Stroeher *et al.*, 1998).

The genes responsible for O-antigen biosynthesis are encoded at the *rfb/wbe* locus which is located on a 20 kilobase (kb) *Sac*I fragment (Fig. 1.17, Table 1.1). This region has been cloned and expressed in *E. coli* K-12 (Manning *et al.*, 1986; Ward *et al.*, 1987). Recently, three additional genes designated *wbeU, V,* and *W* have been described and found to be important in O-antigen biosynthesis in *V. cholerae* O1 (Fig. 1.17) (Fallarino *et al.*, 1997).

The genetic organisation of the operon is essentially identical between the Inaba and Ogawa serotypes with the sequence diverging less than 0.1% (Stroeher *et al.*, 1992). The *rfb/wbe* operon can be divided into five functional regions:

1). perosamine biosynthesis,

2). O-antigen transport,

3). tetronate biosynthesis,

4). O-antigen modification, and

5). additional genes required for O-antigen expression in *V. cholerae*.

The O-antigen of *V. cholerae* O1 consists of a homopolymer of 4-amino-4,6-dideoxy-mannose (perosamine) which is substituted with 3-deoxy-L-glcero-tetronic acid (tetronate) (Kenne *et al.*, 1979, 1982; Redmond, 1979) (Fig. 1.18). The first four genes *rfbA* (*manC*), *rfbB* (*manB*), *rfbD* (*gmd*) and *rfbE* (*wbeE*) are involved in perosamine biosynthesis (Stroeher *et al.*, 1995c). Some of the genes responsible for perosamine synthesis are homologous to those required for alginate synthesis *in Pseudomonas aeruginosa* (Stroeher *et al.*, 1998). Figure 1.19 shows the proposed pathway for perosamine biosynthesis.

Figure 1.17: The *rfb*/*wbe* region of *V. cholerae* O1.

A schematic representation of the genetic organisation of the O-antigen biosynthesis region of *V. cholerae* O1, contained almost entirely on a 20 kb *Sac*I fragment. The gene designations are based on the scheme proposed by Reeves *et al.* (1996) for the nomenclature of Gram-negative lipopolysaccharide and capsule genes. The genes are represented by boxes. Groups of boxes shaded the same indicate those genes proposed to be involved in specific pathways. The direction of transcription is indicated by arrows. Diagram adapted from Stroeher *et al.* (1998).

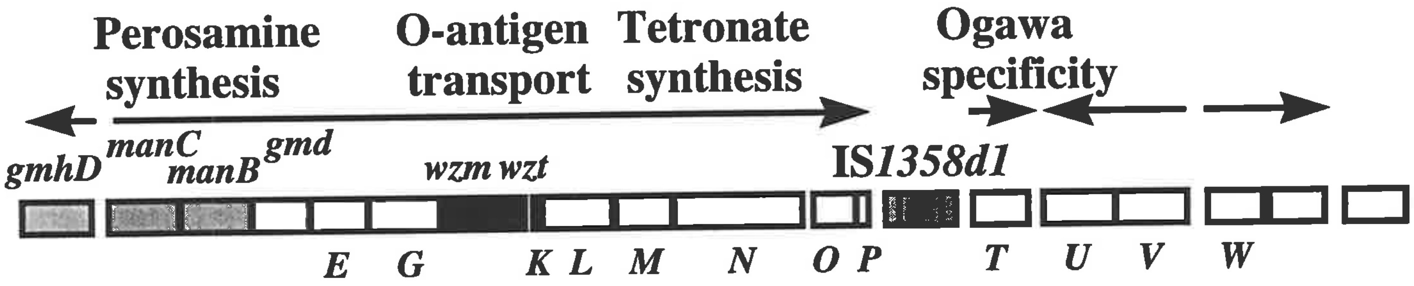

Perosamine synthesis
O-antigen transport
Tetronate synthesis
Ogawa specificity
gmhD
manC
manB
gmd
wzm wzt
IS1358d1
E
G
K L
M
N
O P
T
U
V
W

Table 1.1: Possible functions of *V. cholerae* O1 *rfb* genes

Old ORF/element designation	**New ORF/element designation**	**Putative function**
RfaD	GmhD	core biosynthesis
RfbA	ManC	PMI/GMP activity
RfbB	ManB	PMM activity
RfbD	Gmd/WbeD	oxidoreductase
RfbE	WbeE	pyridoxal binding protein
RfbG	WbeG	unknown
RfbH	Wzm	O-antigen transport (channel?)
RfbI	Wzt	O-antigen transport (energizer?)
RfbK	WbeK	acyl carrier protein
RfbL	WbeL	fatty acid ligase
RfbM	WbeM	alcohol dehydrogenase
RfbN	WbeN	fatty acid reductase
RfbO	WbeO	acetyl Co-A transferase
RfbP	WbeP	unknown
RfbQ,R,S	IS*1358*d1	Insertion sequence
RfbT	WbeT	Ogawa determination
RfbU	WbeU	mannosyl transferase
RfbV	WbeV	LPS (core) biosynthesis
ORF35.7	ORF35.7	dTDP-glucose dehydratase
RfbW	WbeW	galactosyl transferase

Figure 1.18: The O-antigen structure of *V. cholerae* O1.

A) The chemical structure of perosamine, the main constituent of the O-antigen *of V. cholerae* O1.

B) A schematic representation of the LPS molecule of *V. cholerae* O1 showing that the O-antigen consists of a homopolymer of perosamine which is substituted by tetronic acid.

C) The chemical structure of the repeating unit of O-antigen.

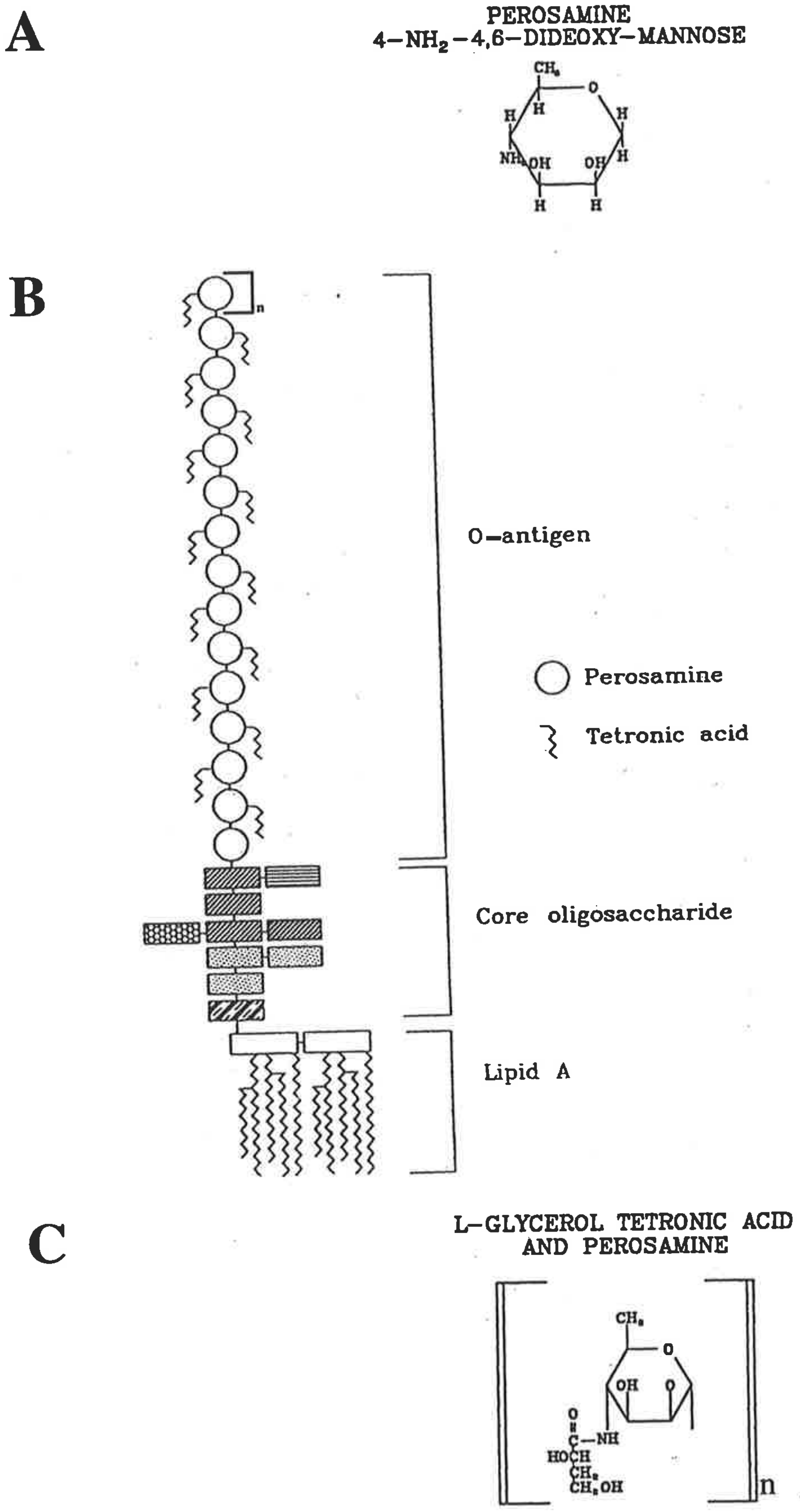
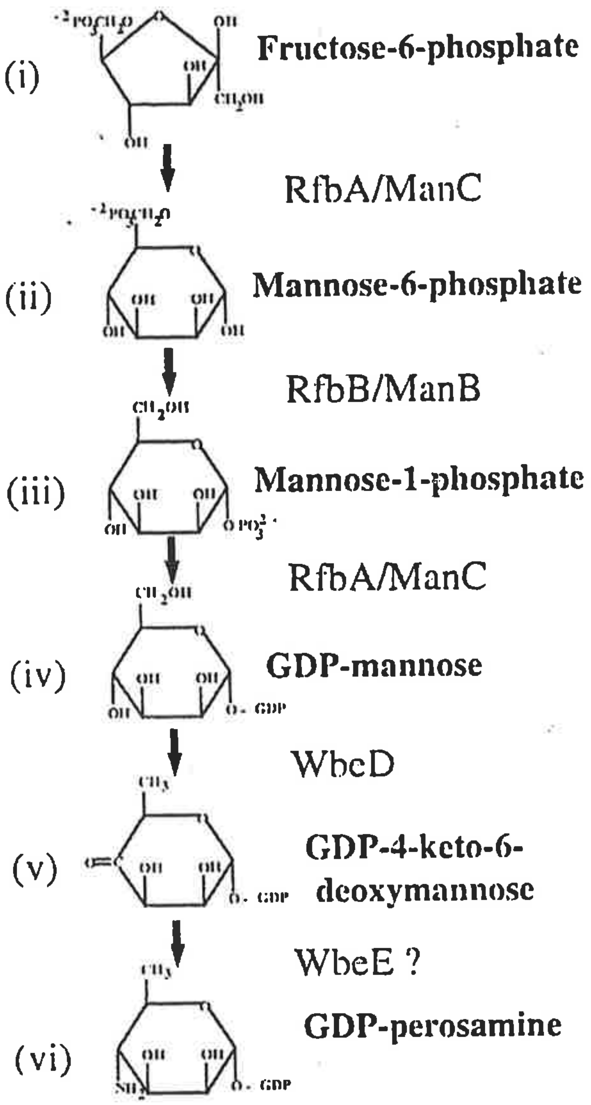

A
PEROSAMINE
4-NH2-4,6-DIDEOXY-MANNOSE
CH3
O
H
NH2
OH
B
n
O-antigen
Perosamine
Tetronic acid
Core oligosaccharide
Lipid A
C
L-GLYCEROL TETRONIC ACID
AND PEROSAMINE
CH3
O
OH
C-NH
HOCH
CH2
CH2OH
n

Figure 1.19: The biosynthesis pathway for the production of perosamine.

The proposed pathway for the biosynthesis of perosamine in *V. cholerae* O1. The pathway is based solely on the homology for each gene (Table 1.1) indicated on the figure between the substrate and product. Figure adapted from Stroeher *et al.* (1998).


(i)
Fructose-6-phosphate
RfbA/ManC
(ii)
Mannose-6-phosphate
RfbB/ManB
(iii)
Mannose-1-phosphate
RfbA/ManC
(iv)
GDP-mannose
WbeD
(v)
GDP-4-keto-6-deoxymannose
WbeE ?
(vi)
GDP-perosamine

The genes required for tetronate biosynthesis are encoded by *wbe/rfb* K,L,M,N and O. Figure 1.20 shows the proposed tetronate biosynthetic pathway (Morona *et al.*, 1995b).

The LPS also contains another sugar, quivosamine, that is thought to cap either the O-antigen or core oligosaccharide. Evidence to date suggests that the genes for this sugar may be encoded in the O1 *rfb/wbe* region because when the region is expressed in *E. coli* the quivosamine is present in the LPS (Hisatsune *et al.*, 1996). Although it is possible that the quivosamine genes were present in the *E. coli* strain used and that the transferase gene is present in the *V. cholerae* O1 region.

V. cholerae O1 O-antigen consists of a homopolymer and as such has the appropriate transport mechanism (Wzm/Wzt). The genes involved in transporting the homopolymer of *V. cholerae* belong to the ABC-2 family of transporters. *rfbH/wzm* is likely to encode the integral membrane protein that forms the channel for O-antigen export. The other protein involved is encoded by *rfbI/wzt* which is likely to be the energiser as it contains an ATP-binding motif (Manning *et al.*, 1994, 1995).

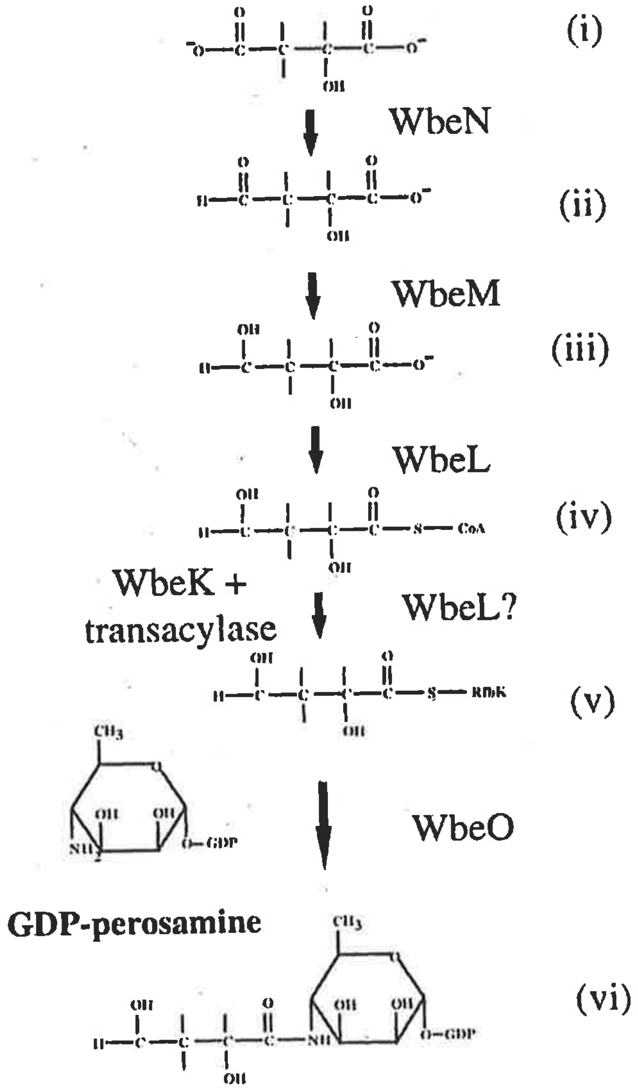
The O-antigen modification gene, *rfbT/wbeT* responsible for expression of the B antigen in the Ogawa serotype is also located within the *rfb/wbe* cluster although it is not apart of the transcriptional unit preceding (*manC* to *wbeO*) (Stroeher *et al.*, 1992; Hisatsune *et al.*, 1993).

1.5.8.2 Genetics of surface polysaccharide synthesis in *Vibrio cholerae* O139

V. cholerae O139 was the first reported non-O1 epidemic strain to cause the cholera. Studies in a number of laboratories were undertaken to discover the major differences between this new disease causing strain, designated O139 and the traditional epidemic strains of *V. cholerae* O1 (Ramamurthy *et al.*, 1993; Johnson *et al.*, 1994;

Figure 1.20: The biosynthetic pathway for the production of tetronate.

The proposed pathway for the biosynthesis of tetronate is shown, although it was devised solely from homologies described for the genes involved (Table 1.1). The starting substrate is malate (i) which is converted to an aldehyde by WbeN (ii) and then di-hydroxy carboxylic acid by WbeM (iii). This intermediate is activated to a Co-A form by the action of WbeL (iv). The substrate 3-deoxy-*l*-glycero-tetronyl Co-A is activated to the acyl carrier protein (ACP) by an unknown enzyme. This ACP activated precursor can be condensed with a molecule of GDP-perosamine to give the complete O-antigen subunit (vi). Diagram from Stroeher *et al.* (1998).


(i)
WbeN
(ii)
WbeM
(iii)
WbeL
CoA
(iv)
WbeK +
transacylase
WbeL?
RfbK
(v)
WbeO
GDP
GDP-perosamine
GDP
(vi)

Karaolis *et al.*, 1994; Manning *et al.*, 1994). The major difference to be reported was the O-antigens of these two serotypes were immunologically different and that *V. cholerae* O139 also expressed a capsule (Manning *et al.*, 1994; Waldor *et al.*, 1994). Extensive research has revealed that in *V. cholerae* O139 there has been precise excision of the original *V. cholerae* O1 *rfb* genes and replacement with new *rfb* genes (Manning *et al.*, 1994; Comstock *et al.*, 1995; Bik *et al.*, 1995; Stroeher *et al.*, 1997; Stroeher and Manning, 1997). The new *rfb* genes of O139 have been cloned and sequenced (Bik *et al.*, 1995, 1996; Comstock *et al.*, 1995, 1996; Stroeher *et al.*, 1995a, 1997; Stroeher and Manning, 1997).

The sugars found in the LPS of O139 are: colitose, glucose, heptose, fructose, glucosamine and quinovosamine (Preston *et al.*, 1995; Knirel *et al.*, 1995). The capsule that has been reported for this serotype is also made up of these sugars. It has been shown that the O-antigen and capsule material are immunologically identical (Waldor *et al.*, 1994)

The region involved in the biosynthesis of the SR-LPS (which constitutes the O-antigen) and capsule is complex and involves numerous sets of genes in independent transcriptional units. Mutational analysis has clearly shown that some genes are involved in both capsule and O-antigen synthesis, and other genes are specific for a particular pathway (Bik *et al.*, 1996; Stroeher *et al.*, 1995a, 1997). Figure 1.21 is a representation of the *V. cholerae* O139 O-antigen/capsule biosynthesis operon. There are numerous similarities between the polysaccharide regions of *V. cholerae* O1 and O139. In O139, like O1, the first gene is *rfaD/gmhD* which is involved in core biosynthesis. However, unlike *V. cholerae* O1, the *rfb* region of O139 is not defined in discreet biosynthetic regions and there are clearly genes that are redundant (Stroeher *et al.*,1998). Several genes in the locus have yet to be designated functions including the *wbfD-H*, *wbfK-N* and *wbfX-Y* (Table 1.2). The genes responsible for the biosynthesis of colitose are encoded by *wbfH*, *wbfI* and *wbfJ*

Figure 1.21: The *rfb/wbf* region of *V. cholerae* O139.

A schematic representation of the genetic organisation of the O-antigen biosynthesis region of *V. cholerae* O139. The gene designations are based on the scheme proposed by Reeves *et al.* (1996) for the nomenclature of Gram-negative lipopolysaccharide and capsule genes. The genes are represented by boxes. Groups of boxes shaded the same indicate those genes proposed to be involved in specific pathways. The direction of transcription is indicated by arrows. Diagram adapted from Stroeher *et al.* (1998).

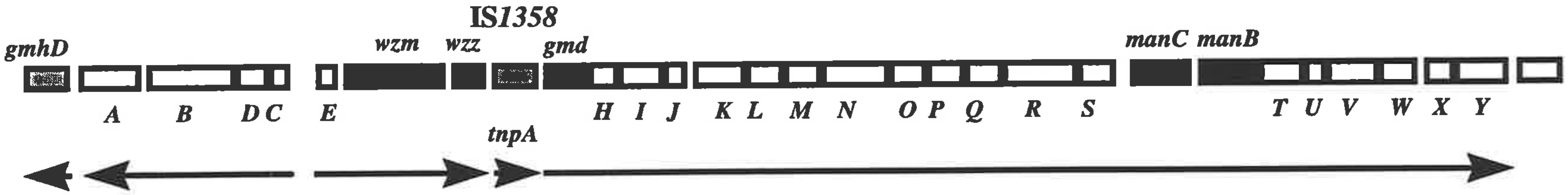

IS1358
gmhD
wzm
wzz
gmd
manC
manB
A
B
D C
E
tnpA
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y

Table 1.2: Possible functions of *V. cholerae* O139 *rfb*/capsule genes

OldORF/element designation	New ORF/element designation	Putative function
RfaD	GmhD	core biosynthesis
OtnH	WbfA	unknown
OtnG	WbfB	unknown
OtnF	WbfC	unknown
OtnE	WbfD	unknown
OtnD	WbfE	unknown
OtnA	Wzm	O-antigen export
OtnB	Wzz	O-antigen chain length regulation
IS*1358*	IS*1358*	Insertion sequence
ORF41.9/RfbD	Gmd	oxidoreductase
ORF34.6	WbfH	colitose biosynthesis
ORF43.9	WbfI	colitose biosynthesis
ORF18.8	WbfJ	colitose biosynthesis
ORF54.4	WbfK	unknown
ORF40.0	WbfL	unknown
ORF40.1	WbfM	unknown
ORF56.2	WbfN	unknown
ORF39.2	WbfO	galactosyl transferase
ORF35.9	WbfP	galactosyl transferase
ORF41.8/ORF1	WbQ (Wzy)	Rfc like (O-antigen polynerase)
ORF71.9/ORF2	WbfR	asparagine bisynthesis
ORF41.3/ORF3	WbfS	galactosyl transferase
ORF50.8/ORF4	ManC	PMI/GMP activity
ORF5	ManB	PMM activity
ORF6	WbfT	UDP-galactose 4-epimerase
ORF7	WbfU	galactosyl transferase
ORF8	WbfV	nucleotide sugar dehydrogenase
ORF9	WbfW	nucleotide sugar epimerase
ORF10	WbfX	unknown
ORF11	WbfY	unknown

based on homology to the colitose biosynthesis genes of *E. coli* O111 (Stroeher *et al.*, 1997). A surprising feature of the O139 region is that there are genes homologous to *wzz* (*rol*) and *wzy* (*rfc*). Mutations in *wzz* appear to effect capsule biosynthesis and not SR-LPS biosynthesis (Bik *et al.*, 1996). The presence of the *wzy* gene is unnecessary in SR-LPS biosynthesis, and therefore Wzy probably acts to polymerise the capsule. Recent studies have shown that the *wzy* gene is involved in capsule biosynthesis and not O-antigen/SR-LPS (U.H. Stroeher, personal communication). Transport of the polysaccharide (O-antigen/SR-LPS and capsule) appears to also utilise the ABC-2 family of transporters although only one (*wzm*) of the two genes required in this system has so far been identified (Bik *et al.*, 1995).

1.6 Insertion Sequences

Analysis of the O-antigen biosynthesis loci of *V. cholerae* O1 and O139 identified the presence of an IS element designated IS*1358*. In *V. cholerae* O1, the element is degenerate (IS*1358*d1) while in the O139 serotype the element and corresponding putative transposase are intact. *V. cholerae* O139 was found to be very closely related to the El Tor strains of *V. cholerae* O1. The main difference between these serotypes are the genes located in their respective O-antigen biosynthesis loci. It was postulated that *V. cholerae* O139 may have arisen from rearrangements in *V. cholerae* O1 and acquisition of new polysaccharide biosynthesis genes from other sources possibly from either non-O1 *V. cholerae* or other *Vibrio* species.

IS elements have been shown to be involved in the acquisition of different types of genes, in particular antibiotic resistance genes. ISs are also used in bacterial systems to acquire other genes that may be advantageous to the host, and indeed have been associated with virulence genes in various animal pathogens such as *Clostridium* (Brynestad *et al.*, 1994; Cornillot *et al.*, 1995), *Escherichia* (So *et al.*, 1979; Hu and Lee, 1988; Collins and Gutman, 1992; Garcia *et al.*, 1994), *Neisseria* (Hammerschmidt *et al.*, 1996), *Yersinia* (Fetherston and Perry, 1994) and *Vibrio* (Stroeher *et al.*, 1995a). Another function of ISs are that they are involved in genetic rearrangements leading to differential gene expression (Mahillon and Chandler, 1998).

1.6.1 IS*As*1 family of Insertion Sequences

In general IS elements have common structural features which allows them to be classified into different families. The ISAs*1* family is a relatively small cluster of 12 members of which 6 are isoforms restricted at present to Gram-negative bacteria (Mahillon and Chandler, 1998). One of the interesting observations of this family of insertion sequences is that a number of the elements have been associated with cell surface biosynthetic genes. The ISs in this family are between 1207 bp (ISPg1) to 1326 bp (IS*1358*) long and have terminal inverted repeats of 14 to 22 bp although not all inverted repeats have been defined (Fig. 1.22). The transposase is encoded by a single ORF (of between 294 and 376 amino acids) and constitutes most of the length of the element. Identity between the elements of this family is between 26% and 50%. Figure 1.22 shows a dendrogram based on the transposase alignments of some of the elements in this family. The transposases of IS*1358* and ISAs*1* have been visualised using protein overexpression systems (Stroeher *et al.*, 1995; Gustafson *et al.*, 1994).

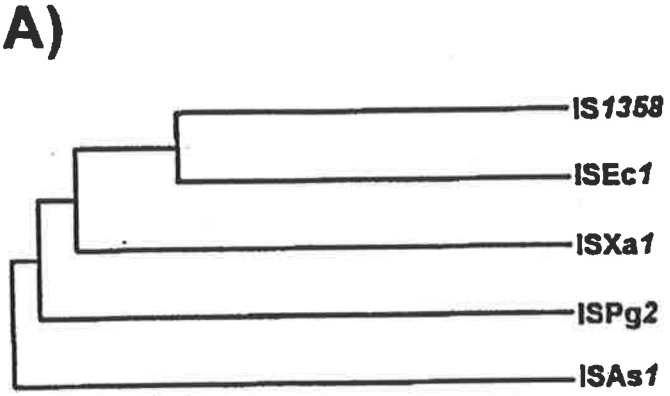
This family of ISs also includes the elements known as “H-repeats”. The H-repeat is not a separate entity but rather is a part of a larger element called the RHS (Rearrangement Hot Spot) element (Zhao *et al.*, 1993; Hill *et al.*, 1994). These elements are found in *E. coli* K-12 with five different elements (RhsA to RhsE) being identified throughout the chromosome. These elements are quite large with sizes between 3.7 and 9.6

Figure 1.22: Relationships between the ISAs1 family of insertion sequences.

A) Dendogram of some of the members of the ISAs1 family of insertion sequences. The dedogram was based on Tpase alignments.

B) The terminal inverted repeats (IRs) of some of the members of the ISAs1 family. Not all IRs have been determined.

Diagram from Mahillon and Chandler (1998).



B)

ISAs*1*	CTGCAGGGCGGGATCGCACTTTG CAGTAGGCCAGGGCGGGAGCACA
ISSe*1*	CAGGGAAAAACACATAAGATTGT AAGGCATGATTACGCCCCCCCGT
IS*1358*	CAGGGAAACCGCATGAAGGTCGT CACGGAAACCGCATGAACATTAC
ISEc*1*	CAGGGAAAGATCATCAAAAGTTC CAGGGCAAGATTACGAAAGCCCG
ISPg2	CAGGGCTGACGCATTAAAAACAG CAGGGGTGACGCATTAAAATTGT

kb. The Rhs element is made up of a number of different components with the biggest segment (core-open reading frame) encoding a possible cell wall, surface ligand binding protein. Every element also contains a H-repeat which displays normal insertion sequence features including inverted repeats. To date however no transposition activity has been detected for this component of the Rhs element.

1.6.2 ISAs*1*

The tetragonal paracrystalline surface protein array (A layer) of the fish pathogen *Aeromonas salmonicida* is a virulence factor which is essential for the bacterium's ability to kill fish (Ishiguro *et al.*, 1981). It has been demonstrated that the production of the A-layer in *A. salmonicida* involves endogenous IS elements, ISAs*1* and ISAs2. These elements are capable of transposing/inserting into either the promoter region or structural gene of the A-layer subunit protein resulting in either reduced levels, truncated forms or complete loss of the protein. IS*As2* does not belong to the ISAs*1* family of insertion sequences (Mahillon and Chandler, 1998). ISAs*1* appears to be a typical insertion sequence element with inverted repeats at either end of the transposase and the production of direct repeats (8 bp) of the target sequence after transposition.

ISAs*1* does not specifically insert into the gene for the A-layer protein and is capable of inserting into other sites on the chromosome although insertion frequency and selection for A-layer mutants showed that this element preferentially inserted into the structural gene (*vapA*) (Gustafson *et al.*, 1994). The mobile nature of ISAs*1* was until recently the only known mobile element in the ISAs1 family of insertion sequences (Mahillon and Chandler, 1998).

1.6.3 IS*1358*

As previously mentioned IS*1358* was first discovered in the *V. cholerae* O1 O-antigen biosynthesis locus it was found to be defective due to point mutations effecting the ORF of the putative transposase. This element consisted of three ORFs which were originally designated *rfbQ*, *rfbR* and *rfbS* before subsequently being renamed IS*1358d1* (Stroeher *et al.*, 1995).

IS*1358* is the second element in the ISAs*1* family of insertion sequences that has been shown to function and hence produce an active transposase. Recently, Dumontier *et al* (1998) demonstrated transposition of one IS*1358* element from *V. cholerae* O22. This suggests that IS*1358* elements with a single ORF are functional and are able to undergo transposition, at least under laboratory conditions.

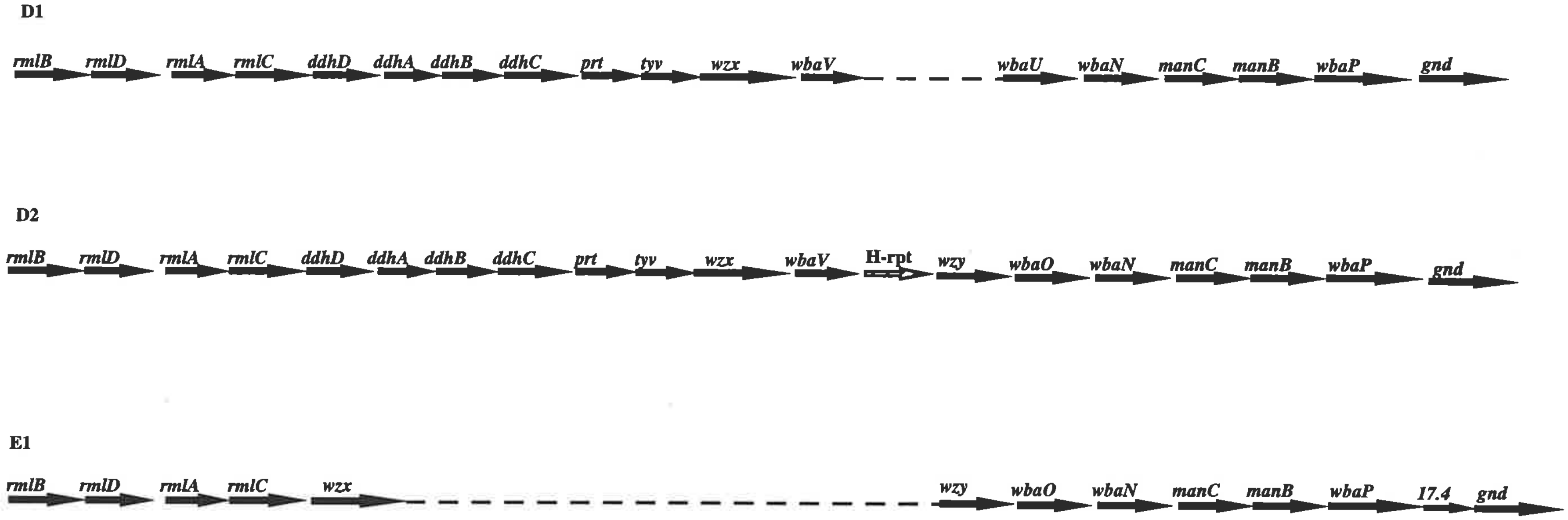
IS*1358* due to its apparent linkage to polysaccharide synthesis genes has been implicated in genetic rearrangements of these regions. The fact that IS*1358* is capable of transposition supports a possible role for this element in rearrangements of polysaccharide biosynthesis genes.

1.6.4 H-repeats of *Salmonella enterica* and their role in genetic rearrangements

The structures of the O-units of the *S. enterica* serogroups A, B, D1, D2 and E are very similar as they all have a mannosyl-rhamnosyl-galactose backbone. The O-antigen biosynthesis regions that encode the O-antigens of these serogroups are similar and have similar genes and genetic arrangements. Xiang *et al* (1994) examined a number of *S. enterica* serogroups for the presence of particular biosynthesis genes and showed that the organisation of the D2 region was very similar to the D1 O-antigen biosynthesis region (Fig. 1.23).

Figure 1.23: The *rfb/wba* regions of *S. enterica* groups D1, D2 and E1.

A comparison of the O-antigen biosynthesis loci of *S. enterica* groups D1, D2 and E1. The new nomenclature for these regions as proposed by Reeves *et al.* (1996) are indicated above the genes. The old ORF designation of these genes can be found in the text. The ORF representing the H-repeat element located in D2 is an arrow with stripes. Diagram adapted from Xiang *et al.* (1994).


D1
rmlB
rmlD
rmlA
rmlC
ddhD
ddhA
ddhB
ddhC
prt
tyv
wzx
wbaV
wbaU
wbaN
manC
manB
wbaP
gnd
D2
rmlB
rmlD
rmlA
rmlC
ddhD
ddhA
ddhB
ddhC
prt
tyv
wzx
wbaV
H-rpt
wzy
wbaO
wbaN
manC
manB
wbaP
gnd
E1
rmlB
rmlD
rmlA
rmlC
wzx
wzy
wbaO
wbaN
manC
manB
wbaP
17.4
gnd

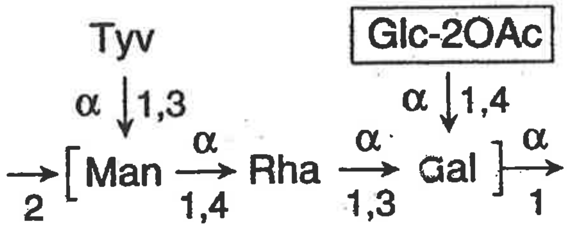
The major differences between the D1 and D2 serogroups are the configuration of the mannose used in the sugar backbone (Fig. 1.24) (Xiang *et al.*, 1994). In D1 α-mannose is used while β-mannose is utilised in the D2 and E serogroups. The E serogroup is also similar to the D serogroup except it does not have the tyvelose side chain sugar (Fig. 1.24). In addition the linkages used to form the O-chains (ie linkage between galactosyl and mannose) during polymerisation are different with D1 having α(1,2) and D2 and E1 using α(1,6) linkages. It was from this information that Xiang *et al* (1994) proposed the suggestion that the O-antigen biosynthesis region of the D2 serogroup was a hybrid of both the D1 and E groups. This suggestion was further supported by immunological studies which showed cross-reactivity of specific groups antigens within these serogroups (Lüderitz *et al.*, 1966).

The proposed mechanism for the generation of the D2 *rfb/wba* region involves recombination of the D1 *rfb/wba* region with the *rfb/wba* region from the E serogroup by intraspecific recombination. Xiang *et al* (1994) propose that a segment of DNA containing the E1 genes (*rfc/wzy* and *rfbO/wbaO*) possibly located on a plasmid replaced *rfbU/wbaU* in the chromosome of the D1 strain (Fig 1.25). This process would facilitate the changes seen in the configuration of the mannosyl residue and the formation of the α(1,6) linkages. Figure 1.25 demonstrates the possible role of the H-repeat in the exchange of DNA to form the new *rfb/wba* region. The H-repeat lies in the junction of the E1 and D1 homologous genes of the D2 *rfb/wba* locus which suggests it did play a role in this recombination event. Xiang *et al* (1994) suggest that the recombination event mediated by the H-repeat can be explained by an attempted H-repeat transposition in which the cointergate is not resolved by site-specific recombination but rather by homologous recombination of the adjacent DNA (Fig. 1.25).

Figure 1.24: Structures of the O-units from *S. enterica* groups D1, D2 and E1.

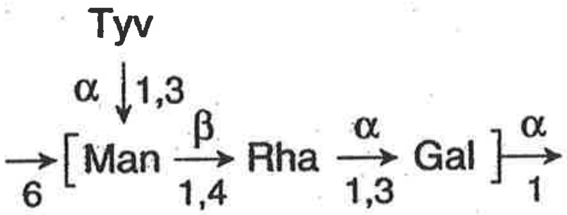
The chemical structures of the O-units from *S. enterica* groups D1 and D2 and E1 are presented for comparison. Abbreviations: Gal, galactose; Man, mannose; Rha, rhamnose; Tyv, tyvelose; Glc, glucose, Oac, acetate. Diagram from Xiang *et al.* (1994).

D1



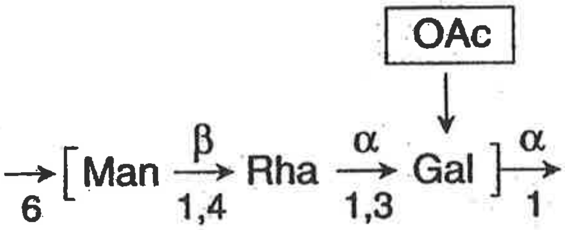
$$\xrightarrow[2]{}\left[\overset{\overset{\text{Tyv}}{\alpha\downarrow 1,3}}{\text{Man}}\xrightarrow[1,4]{\alpha}\text{Rha}\xrightarrow[1,3]{\alpha}\overset{\overset{\boxed{\text{Glc-2OAc}}}{\alpha\downarrow 1,4}}{\text{Gal}}\right]\xrightarrow[1]{\alpha}$$

D2



$$\xrightarrow[6]{}\left[\overset{\overset{\text{Tyv}}{\alpha\downarrow 1,3}}{\text{Man}}\xrightarrow[1,4]{\beta}\text{Rha}\xrightarrow[1,3]{\alpha}\text{Gal}\right]\xrightarrow[1]{\alpha}$$

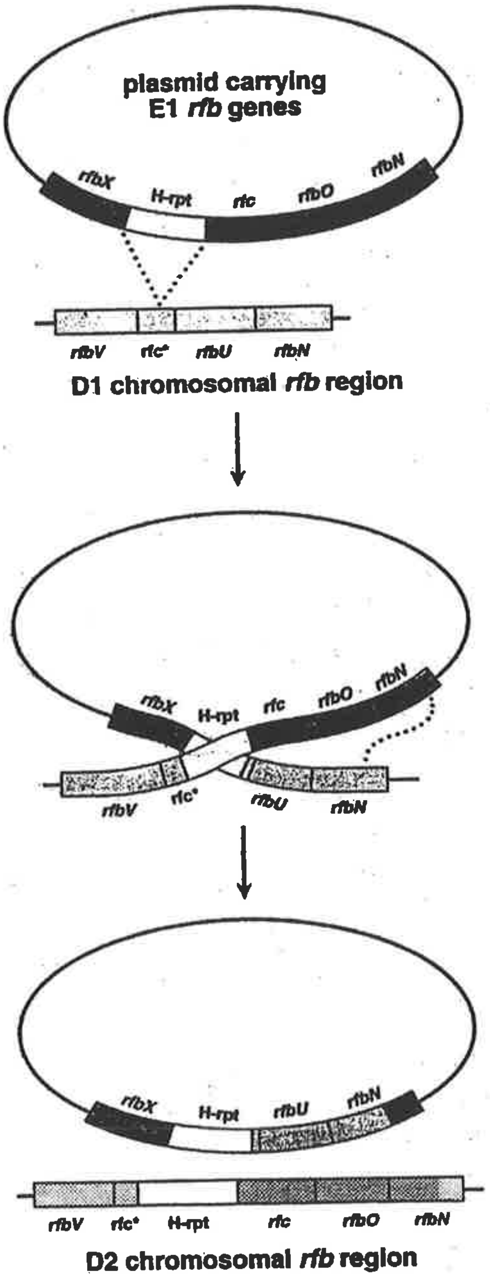
E1



$$\xrightarrow[6]{}\left[\text{Man}\xrightarrow[1,4]{\beta}\text{Rha}\xrightarrow[1,3]{\alpha}\overset{\overset{\boxed{\text{OAc}}}{\downarrow}}{\text{Gal}}\right]\xrightarrow[1]{\alpha}$$

Figure 1.25: Proposed mechanism for the involvement of H-repeats in genetic rearrangements.

The formation of a cointegrate between a chromosome carrying the D1 *rfb* region and a plasmid carrying an E1 *rfb* region form as a result of H-repeat transposition. The dotted lines indicate the sites of two recombination events; the transposition of the H-repeat into *rfc** and a second mediated by homologous recombination which resolves the cointegrate. The new D2 chromosome proposed to have been generated by this mechanism is shown. Diagram from Xiang *et al.* (1994).


plasmid carrying
E1 rfb genes
rfbX
H-rpt
rfc
rfbO
rfbN
rfbV
rfc*
rfbU
rfbN
D1 chromosomal rfb region
rfbX
H-rpt
rfc
rfbO
rfbN
rfbV
rfc*
rfbU
rfbN
rfbX
H-rpt
rfbU
rfbN
rfbV
rfc*
H-rpt
rfc
rfbO
rfbN
D2 chromosomal rfb region

1.7 Aims of this thesis

Vibrio anguillarum is an important fish pathogen that causes mortality in commercial aquaculture. In recent years research has provided some insight into the mechanisms that may be involved in the pathogenesis of this organism. However, little is known about the virulence determinants of this pathogen or how they contribute to the disease process. LPS has been shown to be essential for virulence. To date little is known about the genetics of the biosynthesis region.

Therefore the aims of this thesis are :

1) to identify virulence determinants important in the pathogenesis of this bacteria

2) to examine the relationship between *V. cholerae* and *V. anguillarum* in relation to the previously described insertion sequence, IS*1358* and

3) to isolate, clone, sequence and characterise the O-antigen biosynthesis regions of *V. anguillarum* O1 and O2.

Chapter Two

Materials and Methods

2.1 Bacterial strains and plasmids

The *Vibrio anguillarum* and *Vibrio cholerae* strains used in this study are listed in Table 2.1. Table 2.2 lists the *Escherichia coli* strains used in this study. The plasmids constructed and the cloning vectors used during this study are described in Table 2.3.

2.2 Maintenance of bacterial strains

Strains in routine use were stored as a suspension of freshly grown bacteria in glycerol (32% v/v) and peptone (0.6% w/v) at -70°C. Fresh cultures were prepared from glycerols by streaking the glycerol suspension onto a nutrient agar plate (with or without antibiotic, as appropriate) followed by incubation overnight just prior to use.

For long term storage, all strains were maintained as lyophilised cultures stored *in vacuo* in sealed glass ampoules at 4°C. When strains were required the glass ampoule was opened and its contents resuspended in 100 µl of appropriate sterile broth. Half of the contents were then used to inoculate a 10 ml bottle of nutrient broth and incubated with aeration overnight at the appropriate temperature. The remaining half was streaked for single colonies onto a nutrient plate. If the colony morphology was uniform, single colonies were then selected and used in subsequent experiments. Antibiotics were used when appropriate.

Table 2.1: List of *Vibrio* strains used in this study

Strain:	Serotype/Characterisctics	Source
Vibrio anguillarum		
ATCC43305	O1	P. Hanna [a]
ATCC43306	O2	P. Hanna
ATCC43307	O3	P. Hanna
ATCC43308	O4	P. Hanna
ATCC43309	O5	P. Hanna
ATCC43310	O6	P. Hanna
ATCC43311	O7	P. Hanna
ATCC43312	O8	P. Hanna
ATCC43313	O9	P. Hanna
ATCC43314	O10	P. Hanna
86/3674	O1	J. Carson[b]
85/3954-1	O1	J. Carson
85/3954-2	O1	J. Carson
89/3748	not typed	J. Carson
NB10	O1	D. Milton[c]
KM2	NB10, Tn5*phoA*	D. Milton
KM16	NB10, Tn5*phoA*	D. Milton
KM97	NB10, Tn5*phoA*	D. Milton
KM110	NB10, Tn5*phoA*	D. Milton
Vibrio cholerae		
Type strain collection	O1 to O154	M. J. Albert
O17	O1	M. J Albert
H1	O1	M. J Albert
CA401	O1	M. J Albert
Z17561	O1	M. J Albert
Vibrio parahaemolyticus		
NCTC 10884		IMVS, Australia
AA-3853		M. J. Albert
X-4844		M. J. Albert
Y-17233		M. J. Albert
Vibrio mimicus		
V800		Lab strain
M-33		M. J. Albert
M-35		M. J. Albert
N-4459		M. J. Albert
Vibrio fluvialis		
AL-1577		M. J. Albert
AL-14413		M. J. Albert

[a] Deakin University, Australia
[b] Fish Health Unit, Launceston, Tasmania, Australia
c Umeå University, Sweden

Table 2.2: List of *Escherichia coli* strains used in this study

Strain	Characteristics	Source/Reference
Escherichia coli		
E. coli K-12		
DH5	F,*recA1*,*endA1*,*hsdR17* (r_k^-, m_k^+), *supE*, 44, λ^- ,*thi*1,*gyrA*,*relA1*	Bethesda Research Laboratories
DH5α	F,φ80d,*lacZ*ΔM15Δ (*lacZYA*/*argF*),U169 *recA1*,*endA1*,*hsdR17* (r_k^-, m_k^+), *supE*, 44,λ^-, *thi1*, *gyrA*,*relA1*	Bethesda Research Laboratories
E2096	DH5 + pGP1-2	Laboratory Collection
S17-1	RP4 2-Tc::Mu-Km::Tn7, Pro, res^-. Mod^+, $recA^-$, Tp^R, Sm^R	U.Priefer
SM10	λ*pir*, *thi*, *thr*, *leu*, *tonA*, *lacY*, *supE*, *recA*::RP4-2Tc::Mu	Taylor *et al.*, 1989
E1196	SM10λ*pir*, pRT733, Amp^R, Km^R	Taylor *et al.*, 1989
Shigella flexneri		
2457T	*S. flexneri* 2a	Formal *et al.*, 1958
RMA696	2457T *rol*/*wzz*::Km^R	Van Den Bosch *et al.*, 1997
PE638	*S. flexneri* Y *rpoB* rif^R	Morona *et al.*, 1994
RMA3361	*S. flexneri* *rfbD*::Km^R	C. Daniels
RMA903	*S. flexneri* 2a *bgt*::Km^R	Mavris *et al.*, 1997

Table 2.3: List of plasmids or cloning vectors

Plasmid	Characteristics	Source/Reference
pRT733	Tn*phoA*, Km^R, Amp^R	Taylor *et al.*, 1989
pGP1-2	T7 RNA polymerase Kan^R	Tabor and Richardson, 1985
pGEM5Z+	Amp^R	Promega
pGEM-T	Amp^R	Promega
pBluescript KS$^+$	Amp^R	Stratagene
pBC-KS$^+$	Cml^R	Stratagene
pSUP202	Tet^R, Amp^R, Cml^R	Simon *et al.*, 1983
pTB023	Tet^R, Cml^R, cosmid cloning vector	T. Barnett
pPM2101	Tet^R, Amp^R, cosmid cloning vector	Manning *et al.*, 1985
pPM4901	*virB*/*wbhL* in pGEM-T	This study (Section 4.2.12)
pPM4905	cosmid clone, IS*1358* to *wbhN* in pPM2101	This study (Section 4.2.12)
pPM4910	cosmid clone in pPM2101	This study (Section 5.2.1)
pPM4917	cosmid clone, IS*1358* to *wbhN* in pPM2101	This study (Section 4.2.12)
pPM4918	cosmid clone, IS*1358* to *wbhN* in pPM2101	This study (Section 4.2.12)
pPM4931	0.36 kb *Hin*dIII fragment from pPM4905 in pBC-KS$^+$	This study (Section 5.2.2)
pPM4934	0.40 kb *Hin*dIII fragment from pPM2101 in pBC-KS$^+$	This study (Section 5.2.2)
pPM4940	1.2 kb *Hin*dIII fragment from pPM4905 in pBC-KS$^+$	This study (Section 5.2.2)
pPM4941	*Pst*I-*Eco*RI fragment from pPM4905 in pBC-KS$^+$	This study (Section 5.2.2)
pPM4943	1.2 kb *Hin*dIII fragment from pPM4905 in pBC-KS$^+$	This study (Section 5.2.2)
pPM4944	1.5 kb *Hin*dIII fragment from pPM4905 in pBC-KS$^+$	This study (Section 5.2.2)
pPM4947	IS*1358* from ATCC43305 in pGEM-T	This study (Section 4.2.3)
pPM4948	IS*1358* from 86/3748 in pGEM-T	This study (Section 4.2.3)
pPM4949	IS*1358* from ATCC43307 in pGEM-T	This study (Section 4.2.3)
pPM4950	IS*1358* from ATCC43311 in pGEM-T	This study (Section 4.2.3)
pPM4951	IS*1358* from ATCC43313 in pGEM-T	This study (Section 4.2.3)
pPM4952	IS*1358* from ATCC43306 in pGEM-T	This study (Section 4.2.3)

Table 2.3 continued..

pPM4953	IS*1358* from 85-3954-1in pGEM-T	This study (Section 4.2.3)
pPM4954	IS*1358* from 85-3954-2 in pGEM-T	This study (Section 4.2.3)
pPM4955	1.4 kb *Xba*I fragment from pPM4910 in pBC-KS^{+}	This study (Section 5.2.2)
pPM4958	IS*1358* from ATCC43313 in pGEM-T	This study (Section 4.2.3)
pPM4966	1.8 kb fragment from ATCC43306 in pGEM-T	This study (Section 4.2.6)
pPM4967	2.2 kb fragment from ATCC43306 in pGEM-T	This study (Section 4.2.6)
pPM4968	0.72 kb fragment from ATCC43306 in pGEM-T	This study (Section 4.2.6)
pPM4983	*Pst*I-*Cla*I fragment from pPM4905 in pBC-KS^{+}	This study (Section 5.2.2)
pPM4990	*Eco*RV-*Pst*I fragment from pPM4910 in pBC-KS^{+}	This study (Section 5.2.4)
pPM4991	*Pst*I-*Eco*RV fragment from pPM4910 in pBC-KS^{+}	This study (Section 5.2.2)
pPM4993	*Xba*I fragment from pPM4905 in pBC-KS^{+}	This study (Section 5.2.2)
pPM4995	*Eco*RV-*Pst*I fragment from pPM4905 in pBC-KS^{+}	This study (Section 5.2.2)
pPM4998	4.5 kb *Bgl*II fragment from pPM4905 in pBC-KS^{+}	This study (Section 5.2.3)
pPM5501	*Hinc*II-*Pst*I fragment from pPM5502 inpBluescript-KS^{+}	This study (Section 5.2.4)
pPM5502	PCR product amplified from 85-3954-2, cloned in pGEM-T	This study (Section 5.2.4)
pPM5503	*Hinc*II fragment from pPM5502 in pBluescript-KS^{+}	This stud (Section 5.2.4)
pPM5504	*Hinc*II fragment from pPM5502 in pBluescript-KS^{+}	This study (Section 5.2.4)
pPM5510	*Nde*I Inverse PCR product from ATCC43306 in pGEM-T	This study (Section 4.2.7)
pPM5512	Cosmid clone from 85-3954-2 in pTB023	This study (Section 5.2.1)
pPM5513	Cosmid clone from ATCC43306 in pPM2101	This study (Section 4.2.8)
pPM5514	Cosmid clone from ATCC43306 in pPM2101	This study (Section 4.2.8)
pPM5515	4.8 kb *Hind*III fragment from pPM5512 in pBluescript-KS^{+}	This study (Section 5.2.5)

Strains for long term storage were prepared by suspending several loopfuls of bacteria in a small volume of sterile skim milk. Small aliquots of this suspension are aliquoted into sterile 0.25 x 4 inch freeze drying ampoules and the end of the ampoule plugged with sterile cotton wool. The samples were then lyophilised in a freeze drier. After the sample was completely dried and the vacuum released the cotton wool plug was pushed into the ampoule and the neck of the container is constricted just above the plug. The ampoules are evacuated to a partial pressure of 30 microns and then sealed at the constriction without releasing the vacuum. The ampoules are then labelled and stored at 4°C.

2.3 Growth media

For routine bacterial cultivation of *V. cholerae* and *E. coli* strains the following nutrient media were used. Nutrient broth (NB) which consisted of Lab Lemco (oxoid) (10 g/l), Bactopeptone (oxoid) (10 g/l) and sodium chloride (NaCl) (5 g/l) or Luria-Bertani broth (LB) which is composed of Bacto-tryptone (10 g/l) (Difco), Bacto-yeast (5 g/l) (Difco) and NaCl (5 g/l as described by Miller (1972). For cultivation of *V. anguillarum* strains Tryptic Soy Broth (TSB) (Difco) supplemented with 1.5% (w/v) NaCl was used. Solid media used in this study consisted of nutrient agar (NA) which is composed of the same ingredients as NB except 1.5% (w/v) Oxoid-Bacto Agar has been added. TSA was TSB (Difco) 1.5% (w/v) NaCl and 1.5% (w/v) Oxoid-Bacto Agar. Methionine assay medium (Difco) was reconstituted according to manufacturers instructions, unless otherwise specified.

Antibiotics were added when required to broth and solid media at the following final concentrations: ampicillin (Ap), 50 µg/ml; kanamycin (Km), 50 µg/ml; rifampicin (Rif), 200 µg/ml ; chloramphenicol (Cm), 25 µg/ml; tetracycline (Tet), 4-8 µg/ml depending on the strain used. Colour indicator plates contained X-gal (5-Bromo-4-chloro-3-indolyl-β-D-Galactopyranoside) (Boehringer Mannheim) dissolved in dimethyl formamide at a final concentration of 40 µg/ml and IPTG (Isoropyl-β-D-thiogalactopyranoside) (Boehringer Mannheim) at 24 µg/ml.

Cultures of *V. cholerae* and *E. coli* were incubated at 37°C unless otherwise specified. *V. anguillarum* was grown at 25°C to 30°C. Liquid cultures were normally grown in 20 ml narrow neck McCartney bottles. Optical densities (OD) were measured at 600 nm using a spectrophotometer (Pharmacia LKB, Sweden).

2.4 Chemicals and Reagents

Chemicals were Analar grade. Phenol, sodium dodecyl sulphate (SDS), ethanol, methanol, propan-2-ol, hydrochloric acid, glycerol, ammonium acetate and sucrose were from BDH Chemicals. Tris (Trisma base), herring sperm DNA, Tween 20 and ethidium bromide were from Sigma (St. Louis, MO). Ethylene-diamine-tetra-acetic-acid, disodium salt (EDTA), caesium chloride, citric acid, calcium chloride, magnesium chloride and sodium hydroxide were obtained from Ajax Chemicals, NSW, Australia.

Antibiotics were purchased from Sigma (ampicillin, kanamycin sulphate, rifampicin). and Calbiochem (tetracycline, chloramphenicol). All other anti-microbial agents (dyes, detergents) were purchased from Sigma Chemical Co., BDH Chemicals Ltd., Glaxo, or Calbiochem.

The following electrophoresis grade reagents were obtained from the sources indicated: acrylamide and ammonium persulphate (Bio-Rad), ultra pure N,N'-methylene bis-acrylamide and urea (BRL). N,N,N',N'-Tetramethyl-ethylenediamine (TEMED) was purchased from (Sigma).

The four deoxyribonucleotide triphosphates (dATP, dCTP, dGTP and dTTP), X-gal (5-Bromo-4-chloro-3-indolyl-b-D-galacto-pyranoside) and IPTG (isopropyl-β-D-thiogalacto-pyranoside) were obtained from Boehringer-Mannheim.

Sequencing kits using either dye-labelled primer or dye-labelled terminators were purchased from Perkin Elmer Applied Biosystems (Foster City, California). [35S]-methionine (1,270 Ci/mmole) was purchased from Amersham. Digoxigenin (DIG) labelling and detection kits were purchased from Boehringer-Mannheim.

2.5 Enzymes

Lysozyme was obtained from Sigma, while pronase and proteinase K were obtained from Boehringer-Mannheim. All restriction endonucleases were purchased from Boehringer-Mannheim, New England Biolabs or Progen, and used according to the suppliers instructions. Other DNA modifying enzymes were purchased from the following suppliers: Progen (T4 DNA ligase, calf intestinal phosphatase) and Boehringer-Mannheim (molecular grade alkaline phosphatase). Taq polymerase (Ampli Taq) was purchased from Perkin Elmer Cetus Corp. Long range PCR kits were purchased from Boehringer-Mannheim and were used according to manufacturers instructions. The pGEM-T cloning system was supplied by Promega.

Horseradish peroxidase-conjugated goat anti-rabbit IgG was obtained from Kirkegaard and Perry Laboratories Inc. Anti-digoxigenin-AP (Fab fragments) were from Boehringer-Mannheim.

2.6 Synthesis of oligonucleotides

Oligonucleotides were synthesised on an Applied Biosystems 381A DNA synthesizer in the trityl-off mode and butanol extracted prior to use. Reagents were

Table 2.4: Oligonucleotide primers used in this study

Oligonucleotide#	Sequence: 5'-3'	Location/Description
773	CA(G/C)GGAAACCGCATGAA	IS*1358*, inverted repeats
1065	TTCATGCGGTTTCC(C/G)TG	IS*1358*, inverted repeats
1088	TTTACTCCTATGCACTTG	*virB*, (nt 25691-25709)
1089	GAGATAGTACTGCTAGAA	*virB*, (nt 26850-26832)
2091	GATGAATCCGGCAGCGTT	IS*1358*, (nt 10214-10231)
2092	GCTACAGCCGATTCTTGG	IS*1358*, (nt 10792-10775)
2177	TCTTCCCCATCGGTGATG	pPM2101, Tet^R gene
2178	TTTAGGTACGCCTTCTCG	*wbhK*, (nt 24916-24898)
2181	TCAGCACATCCTTTACTG	*ddhB*, (nt 15176-15194)
2182	GTGAGTTAATGGCCTATG	*ddhA*, (nt 14581-14599)
2193	GCTGGGATGGCATTATCG	IS*1358*, (nt 10102-10120)
2194	TCTCTGTTGCTACAGCCG	IS*1358*, (nt 10800-10782)
2218	CCACAGTATCAGTTTTGG	*otnF*
2219	AAACGCAATGCAGCACGC	*otnG*
2220	TTCACGGGTTGTACGTCG	*otnG*
2221	GAACAACGCTTTCTCCTC	*ddhC*, (nt 15888-15870)
2222	GCTAACCTATTAGCACTG	*ddhC*, (nt 16049-16067)
2223	TTCAGGTTCTGTTGTACC	3' end (nt 30057-30039)
2225	TGTCTCTTGGCATAAAGC	*ddhB*, (nt15685-15703)
2226	GAGTGGCATACATGCACC	1.8 kb fragment between IS*1358* in *V. anguillarum* O2
2227	CTCTGGGCTATAACCAGC	1.8 kb fragment between IS*1358* in *V. anguillarum* O2
2278	GGGCGAGTAAGGATTTAC	1.8 kb fragment between IS*1358* in *V. anguillarum* O2
2280	GGGGTATGATTATGGATC	*wbhG*, (nt 21612-21630)
2281	GGTATGAGCTCTGGTGGA	*wbhG*, (nt 21390-21372)
2282	TGGGCAACCAATGGATAC	*ddhA*, (nt 14613-14631)
2283	ACCCATAAATCTCTGGCG	*wbhE*, (nt 20429-20411)
2284	TGCTCCAAGAACTTTACC	*ddhC*, (nt 15913-15896)
2285	AACCGAGAGTGGGTTTGG	*ddhB*, (nt 15110-15128)
2286	AGATGCATTCACACCCGC	*ddhD*, (nt 13877-13895)
2287	CTCACGCGACAATACTAA	*ddhD*, (nt 13706-13691)

Table 2.4 continued..

2288	TCGAAAGCCTTATCGGGCGGT	reading out of IS*1358*, (nt 11083-11099)
2298	AGGTGGCATTAAACGCAAACGGATGAACTG	IS*1358*, (nt 11021-11051)
2301	GTGAATGCATCTGAGAAAAATGCATCGTTCGGG	*ddhC*, (nt 13890-13857)
2303	TTCTTACATGCGAGATAG	*wbhH*, (nt 22054-22072)
2304	AATGTTATGGGTACGGTC	*ddhB*, (nt 15018-15036)
2305	CATATTGGCTGGATGGCC	*wbhE*, (nt 19937-19919)
2332	GGGGTAGAGAATACTTTC	*wbhB*, (nt 12632-12649)
2333	CAACTATCAATGATACCG	*wbhB*, (nt 12086-12068)
2334	TCTCGAGAGATATTCTCG	*wbhB*, (nt 11494-11512)
2335	TTTATAATCGAGTTTCCAAGTTCGAGTATCCAG	*wbhN*, (nt 28254-28287)
2339	TGGTTGGTACCATTGGTG	*ddhC*, (nt 16387-16405)
2340	CACAGAGCCAGCACCAAC	*wbhJ*, (nt 24427-24409)
2356	GTCAACCATATTTTGAGC	*ddhC*, (nt 16932-16950)
2357	AATACCAGCAGTACACAG	*wbhI*, (nt 23946-23928)
2360	TTCACTCAGATGTTCGAC	*wbhK*, (nt 25412-25394)
2361	TTACTATAAGATTGCGCC	*wbhL*, (nt 25811-25793)
2362	TGGAGAGTTCCTGTCCAC	*wbhN*, (nt 28544-28562)
2363	AAGTTATGGGTAGCAATG	*wbhM*, (nt 27651-27669)
2365	CACGCAATACATCGGCAA	1.8 kb fragment between IS*1358* in *V. anguillarum* O2
2366	CACTGTCCGACCGCTTTG	*Bam*HI site in pPM2101
2372	TTGCAGACTGTTGGCTGG	*wbhC*, (nt 17368-17386)
2373	GTAAAGGTCACTGAAGCG	intergenic region between *wbhH/wbhI*, (nt 23523-23505)
2393	GCTTTTCCGTCACTGCAC	1.8 kb fragment between IS*1358* in *V. anguillarum* O2
2398	CTGACACTCCAATAACAAC	*wbhH*, (nt 23165-23147)
2399	CCAAAATTGTCGCGGCTC	3‘ end, (nt 30413-30395)
2400	TGATCGTGCAGCCCACTC	3‘ end, (nt 29024-29042)
2405	ATCAACTTACCGAGTGGG	*wbhE*, (nt 20561-20543)
2405	GAGAAACTAGATGCTTATTT	*wbhG*, (nt 21178-21160)
2408	TAATGATTTAAAGCAACCAG	*wbhH*, (nt 22492-22510)
2438	AAGCAAATCAGTGGCCTG	*rfe*
2440	CGCTAGGTATTCCCACATG	*wbhC*, (nt 17815-17833)
2473	GCATATAGATTTAGCCCG	*wbhC*, (nt 18204-18222)
2474	GATAAACGCCAGAGATTT	*wbhE*, (nt 20423-20405)
2486	GAGAAGCATATGGTATTG	*wbhC*, (nt 18623-18641)

3034	GTAC**GaATtc**AAAAATGCTTCGTACG	*wbhB*, (nt 12902-12877)
3035	AAAA**gAatTC**GGTAAAGACTATATC	*wbhE*,(nt 19438-19463)
3036	ATT**GaAtTc**TAGATAGAATGTATG	*wbhE*, (nt 20688-20665)
3038	TAGACCTAAGGGCGGTAG	*rmlB*, (nt 2865-2883)
3039	ACCCGCTAAGATAATGCC	*rmlB*, (nt 4035-4018)
3040	AACCTATGCTTAAGAATGAC	*rmlD*, (nt 4827-4847)
3041	ATTTTGACATCAGGAATGTC	*rmlD*, (nt 5815-5795)

nt denotes nucleotide position and corresponds to the sequence of the *wbh* region (Appendix 1). Lower case letters indicate base substitution to insert restriction sites. Cleavage sites for the restriction endonucleases *Eco*RI (GAATTC) and *Pst*I (CTGCAG) are bolded.

purchased from Applied Biosystems or Ajax Chemicals. Some oligonucleotides were purchased from Bresatec or Sequencing Laboratory, IMVS, Adelaide, which were lyophilised and then resuspended in sterile milli-Q water to optimal OD. The oligonucleotides used in this study are listed in Table 2.4.

2.7 Transformation Procedure

Transformation was performed essentially according to the method described by Brown *et al.* (1979). *E. coli* DH5α strains were made competent for transformation with plasmid DNA as follows: an overnight shaken culture (in NB) was diluted 1:20 into NB and incubated with aeration until the culture reached an A_{600} OD of 0.6 (4×10^8 cells/ml). The cells were chilled on ice for 20 min, and pelleted in a bench centrifuge (5, 000 rpm, IEC centrifuge) before resuspending in half volume of cold 100 mM $MgCl_2$, centrifuged again and resuspended in a tenth volume of cold 100 mM $CaCl_2$. The cells were allowed to stand for 60 min on ice before addition of DNA. Competent cells (0.2 ml) were then mixed with 1 µl (100 ng) DNA and left on ice for a further 30 min. The cell/DNA mixture was heated at 37°C for 3 min and left on ice for at least 15 mins. Depending on selection cells were either plated directly or left shaking at 37°C for 45 to 90 min before plating onto selective media. Cells without DNA were included as controls.

2.7.1 Super competent cells

A 10 ml overnight culture in LB was diluted 1:20 into LB and incubated with aeration until an A_{600} OD of 4×10^8 cells/ml was reached. The cells were chilled on ice for 5 min, pelleted at 4°C in a bench centrifuge (5, 000 rpm, IEC centrifuge) and resuspended in 10 ml of solution A (30 mM KAc, 100 mM KCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% glycerol). The suspension was pelleted for 10 minutes at 4°C as above and resuspended in 1 ml of solution B (10 mM MOPS, 75 mM $CaCl_2$, 10 mM KCl, 15% glycerol) and left on ice for 1-2 hours before use. Alternatively the bacterial suspension in solution B was aliquoted into 100 µl lots and stored at -70°C.

2.7.2 Electroporation of *Shigella*

Electrocompetent *S. flexneri* cells were freshly prepared according to the Bio-Rad protocol. An overnight broth of *S. flexneri* (500 µl) was used to inoculate 10 ml Luria broth and incubated at 37°C with aeration until the cells reached an OD_{600} 0.5-0.8. The cells were centrifuged (5, 000 rpm, IEC centrifuge) for 10 min at 5000 rpm and the supernatant discarded. The cells were subsequently resuspended in 5 ml of ice-cold sterile 10% (v/v) glycerol and then centrifuged again for 5 min at 5000 rpm. Finally, the cell pellet was resuspended in 1 ml of ice-cold 10% (v/v) glycerol and kept on ice. The cells were either stored at -70°C or used immediately.

In a sterile microfuge tube on ice, plasmid DNA (in TE or sterile Milli-Q water) as mixed with 100 µl of electrocompetent *S. flexneri* and then transferred to an ice-cold sterile *E. coli* Pulser ™ cuvette (0.2 cm electrode gap, Bio-Rad). The gene pulser (Bio-Rad) was set at 25 µF with the pulse controller at 200 Ω. The *S. flexneri* cells were pulsed at 2.5 kV with time constants of 4.6 - 4.8 msec. Immediately after electroporation, 1 ml of LB was added to the cuvette, the contents mixed and transferred to a sterile microfuge tube, and then incubated at 37°C for 60-90 min. After centrifugation at 15,000 rpm (Hereus Biofuge 15) for 1 min, the supernatant was discarded and the cells gently resuspended in 200 µl of sterile LB and plated onto NA containing appropriate antibiotic.

2.8 Bacterial Conjugation

2.8.1 Method 1

Broth cultures grown in either NB or LB for 18 h were diluted 1:20 and grown to early exponential phase with slow agitation for 4 h. Donor and recipient bacteria were mixed at a ratio of 1:10 and the cells pelleted by centrifugation (5000 rpm, 5 min, IEC centrifuge). The pellet was gently resuspended in 200 µl of broth and spread onto a cellulose acetate membrane filter (0.45 mm, type HA, Millipore Corp.) on a NA plate. Plates were then incubated for 3 h at 37°C. The cells were washed off the filter in 10 ml NB and harvested by centrifugation as above. The cells were subsequently plated onto selective agar and incubated overnight at 37°C.

2.8.2 Method 2

Both the donor and recipient bacterial cells are grown in 10 ml broths for 18 h. The bacteria are then sub-cultured (1/20) and allowed to grow until late log phase ($O.D._{600}$ = 1.0). In an 1.5 ml microfuge tube, 0.5 ml of each culture was centrifuged (1 min, 15k, Heraeus Biofuge 15) and the supernatant removed. To the bacterial pellet 50 µl of fresh broth (no antibiotics) was added and the pellet was resuspended. The resuspended pellets were mixed gently and the mixed bacterial suspensions were then spotted directly onto a TSA plate containing no antibiotics and incubated at 30°C for 16 h. After incubation, the spotted suspension was scrapped off the plate with a sterile loop and resuspended in 2 ml of fresh broth with no antibiotic. The mixture was vortexed to achieve a bacterial suspension and then 25-100 µl of diluted suspension is plated (10^6-10^8) onto the appropriate selective media. The selection used for *V. anguillarum* was either TCBS

Vibrio media (Oxoid) or VAM (sorbitol, 15g (Sigma); yeast extract, 4g; bile salts (Oxoid), 5g; NaCl, 35g; ampicillin, 10mg; cresol red (Sigma), 40mg; bromothymol blue (Sigma), 40mg; agar, 15g dissolved in distilled water, 1000ml) as described by Alsina *et al.*, 1994.

2.9 DNA extraction procedures

2.9.1 Plasmid DNA extraction procedures

2.9.1.1 Method 1

Plasmid DNA was isolated by small scale plasmid purification. The three step alkali lysis method used was a modification of Garger *et al.* (1983). Overnight bacterial cultures (1.5 ml) were transferred to a microfuge tube, harvested by centrifugation (1 min, 15k rpm Heraeus Biofuge 15), and resuspended in 100 µl of solution I (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). The subsequent addition of 200 µl of solution II (0.2 M NaOH, 1% (w/v) SDS) followed by a 5 min incubation on ice resulted in cell lysis. After the addition of 150 µl of solution III (60 ml of 5 M KAc, pH 4.8, 11.5 ml of glacial acetic acid and 28.5 ml of H_2O) and a further 5 min incubation on ice, protein, chromosomal DNA and high molecular weight RNA were collected by centrifugation (8 min, 15k rpm Heraeus Biofuge 15). The supernatant was transferred to a fresh tube and extracted once with TE-equilibrated phenol. Plasmid DNA was precipitated by the addition of 2 volumes of 100% ethanol and a 5 min incubation on ice. The DNA was collected by centrifugation (15 min, 15k rpm Heraeus Biofuge 15), washed with 70% (v/v) ethanol and dried *in vacuo*. The pellet was resuspended in 40 µl of 1x TE.

2.9.1.2 Method 2

Alternatively small scale plasmid preparations were also done essentially as above except after the removal of protein, chromosomal DNA and high molecular weight RNA, the supernatent was transferred to fresh tube and 250 µl of 7.5 M Ammonium acetate pH 7.8 added. This was incubated on ice for a further 15 min and centrifuged for 15 min at 15k rpm (Heraeus Biofuge 15). The resultant supernatant was transferred to a fresh tube and the plasmid DNA precipitated by the addition of 0.8 ml propan-2-ol and incubation on ice for 15 min. The DNA was collected by centrifugation (15 min, 15k rpm Heraeus Biofuge 15), washed with 70% (v/v) ethanol and dried *in vacuo*. The pellet was then resuspended in 50 µl 1 x TE or MQ.

2.9.2 Cosmid preparation

Cosmid DNA was collected by using Triton X-100 cleared lysates from 10 ml overnight cultures by a modification of the procedure described by Kahn *et al.* (1979). Cells were resuspended in 0.4 ml 25% (w/v) sucrose in 50 mM Tris-HCl, pH 8.0. Lysozyme (50 µl, 10 mg/ml freshly prepared in H_2O) and 50 µl of 0.25 M EDTA, pH 8.0 were then added to cells in microfuge tubes and left to stand on ice for 15 min. 0.5 ml TET buffer (50 mM Tris-HCl, 66 mM EDTA, pH 8.0, 0.4% (v/v) Triton X-100) was added and the suspension mixed by inversion of the tubes. The chromosomal DNA was pelleted by centrifugation (20 min, 4°C, Eppendorf) and the supernatant was extracted twice with TE saturated phenol (pH 7.5) and twice with diethyl-ether. Plasmid DNA was precipitated by the addition of 0.6 volumes of propan-2-ol and allowed to stand at -70°C for 30 min. The precipitate was collected (15 min, 15k rpm Heraeus Biofuge 15), washed once with 1 ml 70% (v/v) ethanol, dried *in vacuo* and resuspended in 50 µl 1 X TE buffer.

2.9.3 Preparation of genomic DNA

Genomic DNA from either *V. cholerae* or *V. anguillarum* was prepared according to Manning *et al.* (1986). Cells from a 10 ml shaken overnight culture were pelleted in a bench centrifuge (IEC centrifuge) for 10 min at 4k rpm and resuspended in 2 ml of 25% (w/v) sucrose. Subsequently 1 ml 50 mM Tris-HCl, pH 8.0. with lysozyme (10 mg/ml in 0.25 mM EDTA, pH 8.0) was added and the mixture incubated on ice for 20 min. TE buffer (0.75 ml) and 0.25 ml of lysis solution (5% (w/v) Sarkosyl, 50 mM Tris-HCl, pH 8.0, 0.25 mM EDTA, pH 8.0) were added, together with 2 mg solid pronase. The mixture was gently vortexed, and incubated at 60°C for 60 min. This was followed by three extractions with TE-saturated phenol and one extraction with diethyl-ether. The genomic DNA was precipitated with 9 mls of ice cold 100% ethanol, spooled onto a Pasteur pipette, washed in 70% (v/v) ethanol, and resuspended in 1 ml of MQ water.

2.10 Analysis and manipulation of DNA

2.10.1 DNA quantitation

The concentration of DNA in solutions was determined by measurement of absorption at 260 nm and assuming an A_{260} of 1.0 is equal to 50 mg dsDNA/ml (Miller, 1972).

2.10.2 Restriction endonucleases digestion of DNA

Most cleavage reactions were performed using the restriction enzyme buffer recommended by the manufacturers. 0.1-0.5 µg of DNA or purified restriction fragments were incubated at 37°C, for 1-2 h with 2 units of each restriction enzyme in a final volume of 20 μl. Chromosomal digests (1-5 μg) were incubated overnight at 37°C with 4 units of restriction enzyme. The reactions were terminated by heating at 65°C for 10 min (or for restriction enzymes which can not be heat inactivated, the reaction mix was extracted with 1 volume of Tris equilibrated phenol followed by ethanol precipitation). Prior to loading onto a gel, a one tenth volume of tracking dye (15% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 0.1 mg/ml RNase A) was added.

2.10.3 Calculation of restriction fragment size

The sizes of restriction enzyme fragments were calculated by comparing their relative mobility with that of *Eco*RI digested *Bacillus subtilis* bacteriophage SPP1 DNA. The sizes were: 8.50; 7.35; 6.10; 4.84; 3.59; 2.81; 1.95; 1.86; 1.51; 1.39; .1.16; 0.98; 0.72; 0.48; 0.36; (Franzon and Manning, 1986).

2.10.4 Analytical and preparative separation of restriction fragments

Electrophoresis of digested DNA was carried out at room temperature on horizontal, 0.8% or 1% (w/v) agarose gels (Seakem HGT). Gels were electrophorised at 100V for 1-3 h in either 1x TBE buffer (67 mM Tris base, 22 mM boric acid and 2 mM EDTA, final pH 8.8), or 1x TAE buffer (40 mM Tris acetate and 2 mM EDTA). After electrophoresis the gels were stained in distilled water containing 2 mg/ml ethidium bromide. DNA bands were visualized by trans-illumination with UV light and photographed using either Polaroid 667 positive film, 665 negative film or thermal paper (K65HM) for Mitsubishi Video Copy Processor.

For preparative gels Sea Plaque (Seakem) low-gelling-temperature agarose at a concentration of 0.6% (w/v) was used for separation of restriction fragments, which were recovered by the following methods:

Method 1: DNA bands were excised and the agarose melted at 65°C. Five volumes of 20 mM Tris-HCl, 1 mM EDTA, pH 8.0 buffer were added and the agarose extracted with phenol:water (1:1) and then phenol:chloroform (1:1). Residual phenol was removed with chloroform and the DNA precipitated with two volumes of ethanol and one tenth volume of 3 M sodium acetate, pH 5.0. DNA was collected by centrifugation (4°C, 15 min, 15k rpm, Heraeus) washed once with 70% (v/v) ethanol and dried *in vacuo* before being resuspended in 1 x TE buffer.

Method 2: After electrophoresis, the required DNA bands were excised and then placed inside dialysis tubing. This was then positioned in an electrophoretic tank filled with 0.5 x TAE buffer. A current (100V, 1 h) was applied to electrophorese the DNA out of the gel and into the buffer. The DNA was then extracted with an equal volume of TE saturated phenol and precipitated with two volumes of ethanol and one tenth volume of 3 M sodium acetate, pH 5.0.

Method 3: The Qiagen gel extraction kit was used according to the instructions provided by the manufacturer (Qiagen, Germany).

2.10.5 Dephoshorylation of DNA

Restriction enzyme digested DNA was treated with alkaline phosphatase by the following method. 0.1-0.5 µg of digested plasmid DNA was incubated with 1 unit of alkaline phosphatase (Calf intestinal: CIP) for 30 min at 37°C. The reaction was terminated by the addition of EDTA, pH 8.0 to a final concentration of 3 mM followed by heating at 65°C for 10 min. The reaction mix was then extracted twice with TE saturated phenol and twice with diethyl ether. DNA was precipitated overnight at -20°C with two volumes of ethanol and 1/10 volume of 3 M sodium acetate pH 8.0. The precipitate was collected by centrifugation (4°C, 15 min, 15k rpm Eppendorf), washed once with 1 ml 70% (v/v) ethanol, dried *in vacuo* and dissolved in 1x TE buffer.

2.10.6 Ligation of DNA/in vitro cloning

DNA to be subcloned (200 ng) was cleaved by either single or double restriction enzyme digests. The restriction digest of the insert was combined with 20 ng of similarly cleaved vector DNA. Ligation reactions with T4 DNA ligase were performed in 1 x ligase buffer (20 mM Tris-HCl, 10 mM $MgCl_2$, 0.6 mM ATP, 10 mM DTT and BSA (100 mg/ml)) for cohesive ends, and incubated at 10°C for 16 h. Restriction enzymes were heat inactivated at 65°C prior to ligation.

The ligated DNA was then used directly for transformation of *E. coli* strains. Transformants were screened for insertional inactivation of the appropriate drug resistance gene (eg. Ap or Tc), wherever possible, prior to plasmid DNA isolation.

2.10.6.1 Cosmid Cloning

High molecular weight genomic DNA prepared from strain 85-2954-2 and ATCC 44306 were partially digested with *Sau*3A. Samples were taken at specific time intervals and aliquots were electrophoresed on a 1.0% (w/v) agarose gel in TAE buffer. DNA fragments of approximately 35-45 kb were excised from the gel, electroeluted and extracted using TE-saturated phenol/chloroform. The purified DNA was ligated to pPM2101 which had previously been completely digested with *Bam*HI and dephosphorylated with CIAP. Ligation was allowed to proceed overnight at 4°C. The ligation mix was subsequently packaged onto bacteriophage λ (Collins and Hohn, 1978) using an *in vitro* packaging system (Packagene System, Promega) and then transduced into *the E. coli K-12 strain* S17-1.

2.10.6.2 pGEM-T cloning

For efficient ligation a molar ratio of 1:3 of pGEM-T vector to PCR product was required. Unpurified PCR product was added to a mix containing T4 DNA ligase (1 µl), pGEM-T vector (1 µl), T4 DNA ligase 10x buffer(1 µl) and dH_20 to 10 µl. This mixture was left overnight at room temperature.

The mixture was then transformed into DH5α and plated onto NA/ampicillin/ X-gal (5-Bromo-4-chloro-3-indolyl-b-D-galacto-pyranoside)/ IPTG (isopropyl-β-D-thiogalacto-pyranoside) plates to select for vector with insert.

2.10.7 Random primer labelling

Purified DNA fragments were labelled with DIG-11-UTP (Boehringer Mannheim) or with the described protocol. 5 µl of plasmid or PCR product was heated for 10 min at 100°C before snap freezing on dry ice/ethanol. To the DNA, 2 µl of hexanucleotide mix (Boehringer Mannheim), 2 µl of dNTP labelling mix, 10 µl MQ water and 1 µl Klenow Enzyme were added and incubated overnight at 37°C. The reaction was stopped by the addition of 0.2 M EDTA pH 8.0 (2 µl). The DNA was precipitated with 2 µl 4 M LiCl and 60 µl of cold ethanol and either left at -70 °C for 30 min or 2h at -20°C. Labelled DNA probes were collected by centrifugation and washed with 70% (v/v) ethanol, dried *in vacuo* and resuspended in 50 µl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0)

2.10.8 Southern transfer and hybridisation

Unidirectional transfer of DNA from agarose gels to Hybond-N+ nylon transfer membrane (Amersham) were performed as described by Southern (1975) and modified by Maniatis *et al.* (1982).

Digested chromosomal DNA samples and markers were electrophoresed overnight on 0.8% (w/v) agarose gel, stained with EtBr and photographed. The gel was then denatured by soaking in several volumes of 1.5 M NaCl and 0.5 M NaOH solution for 1 h with slow agitation. The gel was subsequently neutralized with several volumes of 1 M Tris pH 8.0, 1.5 M NaCl for one hour with slow agitation. Transfer of the DNA to nitrocellulose (Schleicher and Schuell) or nylon (Amersham) was performed overnight at room temperature using 10 X SSC as the transfer buffer as described by Maniatis *et al* (1982). After transfer the filter was either fixed for 20 min with 0.4 M NaOH or baked for 1 h at 80°C.

Prior to hybridization filters were incubated for 2 h in pre-hybridization solution (50% (v/v) formamide, 1% (w/v) skim milk, 7% (w/v) SDS, 250 µg/ml single stranded herring sperm DNA (Sigma), 5 X SSPE (Maniatis *et al.*, 1982)). The pre-hybridization solution was removed and denatured DIG-labelled probe in pre-hybridisation was added after pre-heating at 65°C for 10 min (freshly prepared probes were boiled for 10 min and then added to pre-hybridization solution). The probe was left on the filter for 18 h at 42°C. After hybridization, the filters were washed at room temperature twice for 5 min with 2 X SSC containing 0.1% (w/v) SDS followed by 2 x 15 min washes with 0.2xSSC containing (w/v) 0.1% SDS at 65°C.

To detect the bound probe, filters were washed briefly in Buffer 1 (0.1 M Tris-HCl, 0.15 M NaCl pH 7.5) and incubated for at least 1 h in 5% (w/v) skim milk (in Buffer 1) at room temperature (RT) with gentle agitation. After a brief rinse in Buffer 1, antibody conjugate (anti-DIG AP; alkaline phosphatase coupled to anti-DIG Fab fragments) diluted 1/5000 in Buffer 1 was added to the filter and left to incubate for 30 min at RT with gentle agitation. Unbound antibody conjugate was removed after incubation by washing twice for 15 min in Buffer 1. This was followed by a 2 min wash with Buffer 3 (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$). The bound alkaline phosphatase conjugate was detected by the addition of 45 µl of nitroblue-toluidine (NBT) (75 mg/ml) (Boehringer Mannheim) and 35 µl of the substrate bromochloroindolyl phosphate (BCIP) (50 mg/ml) (Boehringer Mannheim) in Buffer 3. The filter was allow to develop in the dark without agitation. To stop the reaction the above solution was removed and 100 mls of 1 x TE added.

2.10.9 Colony blots using labelled DNA probes

Individual colonies are grown in 100 µl of appropriate broth with antibiotics in 96 well microtitre trays (Falcon) at the appropriate temperature without agitation for 16 h. After incubation, 5 µl of the individual cultures were spotted onto H^+ bond nylon (Amersham) and allowed to air dry. The filter is then treated with 0.5 M NaOH for 5 min followed by soaking of the filter in 0.1 M NaOH, 1.5 M NaCl for a further 5 min. The nylon membrane is then treated twice with 1 M Tris-HCl pH 7.5 for 5 min. The filter are subsequently soaked for 5 min in 0.5 M Tris-HCl pH 7.5, 1.5 M NaCl. Each of the above steps was performed by laying the filter membrane onto Whatmann 3MM soaked with the appropriate solution. The filter was then air dried for 35-40 min before the treated colonies were fixed to the membrane by treating the filter with 0.4 M NaOH for 20 min, followed by a brief wash in 5 X SSC. The filter was then incubated at 42°C for 1-2 h with agitiation in prewashing solution (50 mM Tris-HCl pH 8.0, 1 M NaCl, 1 mM EDTA pH 8.0, 0.1% (w/v) SDS). The filter was subsequently treated as for Southern hybridisation beginning with prehybridisation solution (section 2.10.9).

2.11 Polymerase Chain Reaction (PCR) Protocol

2.11.1 Standard PCR Reaction

The procedure used was a modification of Saiki and Gelfand (1984). PCR amplifications were performed in reaction tubes (0.5 ml, Perkin Elmer) in a volume of 50 µl containing PCR buffer (1.5 mM $MgCl_2$, 10 mM Tris pH8.4, 50 mM KCl), 2U of Taq polymerase (Cetus), 100 pmol of each oligonuncleotide primer and 100 ng of chromosomal or plasmid DNA. The dNTPs (deoxynucleoside triphosphate, Boehringer-Mannheim) were used at a final concentration of 2 mM. The reaction was overlaid with a drop of light mineral oil (Nujol, Perkin Elmer). The thermocycler (Perkin Elmer Cetus) was programmed with the following cycle; initial incubation at 95°C for 5 min followed by 25 amplification cycles consisting of 95°C for 30 sec, 50°C for 30 sec, 72°C for 3 min and a final extension at 72°C for 5 min. 10 µl of this reaction product was analysed on a 0.8% agarose gel, the remaining product was purified using a PCR purification kit (Qiagen) and used for cloning.

2.11.2 Inverse PCR

The method used was described by Ochman *et al.* (1988) with modifications outlined below. Chromosomal DNA was digested with appropriate restriction enzymes as described above (section 2.10.2). Digested DNA was purified as described by ethanol precipitation with 0.5 volumes of 3 M sodium acetate plus 2.5 volumes of 95% ethanol and incubated at -20°C for 10 min. The DNA was collected by centrifugation at 15,000 rpm for 15 min (Heraeus Biofuge 15), washed with 70% ethanol, dried *in vacuo* and resuspended in 20 µl of sterile water. The DNA was recircularised by the addition of T4 DNA ligase and 1 X ligase buffer (as described in 2.10.6) followed by a 18 h incubation at 4°C. The ligated DNA was precipitated and resuspended in 15 µl of sterile MQ. PCR amplification was performed as described for standard conditions, with an elongation time ranging from 2 to 5 minutes.

2.11.3 Long Range PCR

The Expand^{TM} Long Template PCR System (Boehringer Mannheim, Germany) was used for long range PCR (according to the manufacturer's instructions). The PCR program used consisted of 25 cycles, after an initial incubation at 92°C for 2 min to allow complete denaturation of the DNA template. The first 10 cycles consisted of denaturation at 94°C for 10 sec, annealing at 65°C for 30 sec, and elongation at 68°C for 5-10 min. Cycles 11-25 consisted of denaturation at 94°C for 10 sec, annealing at 65°C for 30 sec, and elongation at 68°C for 5-10, increasing by 1 min per cycle to cycle 25.

2.11.4 PCR DIG-labelled DNA probes

The protocol used for DIG-labelled PCR is essentially that described in Section 2.10.7. The PCR reaction was performed in 0.5 ml reaction tubes in a 50 µl volume containing supplied Taq buffer, 200 µM each of deoxynucleoside triphosphate dATP, dCTP and dGTP,_190 µM of dTTP and 10 µM of DIG-11-dUTP, 200 ng of plasmid template or genomic DNA and 2.5 U of Taq polymerase (Perkin Elmer). The reaction was overlaid with a drop of light mineral oil and subjected to the cycle detailed in Section 2.11.1.

2.12 DNA sequencing procedures

2.12.1 Sequencing using dye labelled primers

Sequencing reactions were carried out on 1 µg of double stranded plasmid DNA using the protocol provided by Applied Biosystems. In dye-labelled primer sequencing the DNA was divided into four tubes, tubes A and C containing 160 ng and tubes G and T containing 320 ng of DNA respectively. To each tube Ready reaction mix and DNA template were added as follows:

Reagent	**A**	**C**	**G**	**T**
Ready Reaction Mix	4µl	4µl	8µl	8µl
DNA Template	1µl	1µl	2µl	2µl
Total Vol.	5µl	5µl	10µl	10µl

Each reaction was overlayed with 20 µl of light mineral oil and centrifuged briefly.

The samples were then cycle sequenced using the following conditions: 15 cycles (96°C 10 sec; 55°C 5 sec; 70°C for 60 sec), followed by 15 cycles (96°C 10 sec; 70°C 60 sec; 15 cycles total), and were then held at 4°C. Reactions were subsequently combined in 80 µl of 95% (v/v) ethanol with 3 µl of 3 M sodium acetate and precipitated on ice. DNA was harvested at 13,000 rpm for 15 min (Hereaus Biofuge15). Samples were dried *in vacuo* and stored at -20°C.

2.12.2 Sequencing using dye labelled terminators

Plasmid DNA was purified by ethanol precipitation prior to dye terminator sequencing with kits supplied by Applied Biosystems. 0.5 ml thin walled tubes (Gene Amp, Perkin Elmer) containing 1-2 μg of template DNA and 3.2 pmol primer, made up to a final volume of 20 μl with 9.5 μl of pre-mix (Applied Biosystems) and sterile water. The reaction were overlaid with mineral oil (Nujol, Perkin Elmer) and subjected to 25 cycles (96°C 30 sec; 50°C 15 sec; 60°C 4 min) before adding 2 μl of 3 M sodium acetate and 50 μl of ice cold 100% ethanol and precipitating for 2 h at -20°C. The DNA was washed with 70% (v/v) ethanol and drying *in vacuo*.

2.12.3 Electrophoresis of sequence reaction and analysis of DNA sequences

The dried DNA pellets for sequencing were stored at -20°C until required. Prior to loading for sequencing, the reactions were resuspended in 4.5 μl loading buffer (83% deionised formamide, 8.3 mM EDTA pH 8.0) and heated to 95°C for 2 min. Samples were subsequently electrophoresed on a 6% polyacrylamide-8M urea gel in an Applied Biosystems 373A or 377 DNA sequencer. Raw sequencing data from the 373A or 377 automated sequencer were analysed using the Applied Biosystems Seq Ed program version 6.0. The elecrophoresis was carried out by Sequencing Laboratory, IMVS, Adelaide. Sequencing data were analysed using the LKB DNA and protein analysis programs, DNASIS and PROSIS (Hitachi Software).

2.13 Protein Analysis

2.13.1 SDS-Polyacrylamide Gel Electrophoresis

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on either 15-20% polyacrylamide gels for proteins or lipopolysaccharides using a modification of the procedure of Lugtenberg *et al.* (1975) as described previously by Achtman *et al.* (1978). Samples were heated at 100°C for 3 min in 1 X SDS sample buffer (25 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, 15% (w/v) bromophenol blue) prior to loading. Gels were generally electrophoresed at 180 V for 2.5 h (15-20% gels) or 13 mA constant current for 16 h (15% PAGE gels). Proteins were stained with gentle agitation overnight at RT in 0.06% (w/v) Coomassie Brilliant Blue G250 (dissolved in 5% (v/v) perchloric acid). Destaining was accomplished with several changes of 5% (v/v) acetic acid, with gentle agitation for 6-24 h. Size markers (Pharmacia) were phosphorylase B (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20.1 kDa) and a-Lactalbumin (14.4 kDa). Gels for LPS were not stained by the above method (see 2.15.2).

2.13.2 Protein overexpression using the T7 RNA Polymerase Expression System

The method employed is essential that described by Tabor and Richardson (1985). The plasmid pGP1-2 carries the T7 RNA polymerase under the control of the lambda P_L promoter (Tabor and Richardson 1985). The plasmid pGP1-2 was transformed into *E. coli* strains containing a plasmid with the specific gene of interest under control of the T7 RNA polymerase promoter. 10 ml of LB broth with Amp (50 μg/ml) and Km (50 μg/ml) was inoculated with a single colony and grown with aeration at 30°C for 16 h. The culture was subcultured 1:10 and incubated with constant aeration at 30°C. When an A_{600} O.D. of 0.6 was reached, the cells were pelleted (15 k rpm, 10 min, IEC centrifuge) and washed in M9 media supplemented with 0.2 mg/ml $MgSO_4$, 0.5% (w/v) glucose and 50 mg/ml thiamine HCl before incubation at 30°C for 1 h. The cells were pelleted again, resuspended in Methionine Assay Media (Difco) and the temperature was shifted to 42°C to induce the pGP1-2 P_L promoter by the inactivation of the repressor cI_{ts}. This allows the expression of the T7 RNA polymerase from the l_{pL} promoter. Subsequently, rifampicin was added to a final concentration of 200 µg/ml to inactivate the *E. coli* host RNA polymerase, and incubation was continued at 42°C for a further 10 minutes. The culture was then left for at least 2 h shaking at 37°C to allow expression of the gene under the control of the T7 promoter. 1 ml of culture was transferred to microfuge tubes, centrifuged (15k rpm, 2 min Heraeus Biofuge15) to pellet the cells and resuspended in 100 µl of 1 X SDS sample buffer. 10 µl of sample was boiled for 2 min and loaded onto SDS-PAGE gels for analysis. Gels were subsequently stained with Coomassie G250. For radioactive labelling the procedure is the same except the samples were pulsed for 5 min at 30°C with 5 uCi L-[^{35}S]-methionine after expression at 37°C.

2.13.3 Autoradiography

SDS-PAGE gels were dried on Whatman 3MM chromatography paper at 60°C for 2 h on a Bio-Rad gel drier. L-[^{35}S]-methionine labeled gels were subjected to autoradiography which was performed at room temperature for 1-7 days without intensifying screens using Kodak XR-100 film.

2.13.4 Colony Blotting and Western Transfer

The procedure used was a modification of that described by Towbin *et al.* (1979). Samples were subjected to SDS-PAGE and transferred to nitrocellulose (Schleicher and Schuell) at a transfer current of 200 mA for 2 h in a Trans-Blot Cell (Biorad). The transfer buffer used was 25 mM Tris-HCl pH 8.3, 192 mM glycine and 5% (v/v) methanol. For LPS transfer the buffer contained 20% (v/v) methanol and the transfer time was 1 h at 500 mA at 4°C. After transfer, the nitrocellulose sheet was incubated for 30 min in 5% (w/v) skim milk powder in TTBS (0.05% (v/v) Tween 20, 20 mM Tris-HCl pH 8.0, 0.9% (w/v) NaCl) to block non-specific protein binding sites. The primary antiserum was diluted 1/1000 to 1/5000 in TTBS with 0.02%(w/v) skim milk powder and incubated with gentle agitation at RT overnight. The unbound antibody was removed by washing the filter three times for 10 min in TTBS with agitation. Bound antibody were detected by incubating the filter for 2 h (gentle agitation) with goat anti-rabbit IgG coupled with horseradish peroxidase (Nordic Immunology) at a dilution of 1/5,000 in TTBS. The filter was then washed twice for 5 min with TTBS, followed by three 5 min washes in TBS (20 mM Tris-HCl pH 8.0, 0.9% (w/v) NaCl). The antigen-antibody complexes were then visualized using peroxidase substrate (9.9 mg 4-chloro-1-napthol dissolved in 3.3 ml cold methanol (-20°C) added to 16.5 ml TBS containing 15 µl hydrogen peroxide) and allowed to incubate for 10-15 min with shaking, as described by Hawkes *et al.* (1982).

2.13.5 Purification of Proteins by Electroelution

Whole cell samples were electrophoresed on polyacrylamide gels as described in Section 2.13.1. The gel was stained and destained as described above. Proteins can be electroeluted from gel slices by the method described by Leppard *et al.* (1993). After destaining the gel the corresponding protein band cut out of the gel ensuring that no other proteins were also excised. The gel slice was placed into small length dialysis tubing containing 1 ml of 0.2 M Tris acetate (pH 7.4), 1.0 % SDS and 100 mM dithiothreitol (DDT) per 0.1 g of wet polyacrylamide gel. The tubing was then placed in a horizontal electrophoresis tank and submerged in running buffer (50 mM Tris acetate (pH 7.4), 0.1% SDS and 0.5 mM sodium thioglycolate) and allowed to electrophorese for 3 h at 100V. The gel slices were removed from the tubing and stained with Coomassie blue to check that the protein had been electroeluted. The solution containing the protein of interest was dialysed against distilled water. The 100 µl of purified protein was then electrophoresed on a 15 % SDS polyacrylamide gel in preparation for N-terminal sequencing (2.13.5). The purified protein was stored at -20°C.

2.13.5.1 Cell fractionation

The cell fractionation procedure was a modification of that described by Osborn *et al.*, (1972). Cells were grown in LB to mid-exponential phase at 37°C (50 ml, OD_{600} of 0.6). Cells were pelleted in a Beckman SS-34 rotor, (7,000 rpm, 10 min, 4°C) and resuspended in 1 ml of 20% (w/v) sucrose, 30 mM Tris-HCl pH 8.1, transferred to SM-24 tubes and chilled on ice. Cells were converted to sphaeroplasts with 0.1 ml of 1 mg/ml lysozyme in 0.1 M EDTA pH 7.3 for 30 min on ice. Cells were centrifuged as above and the supernatant collected (periplasmic fraction). The cell pellet was frozen in an ethanol dry ice bath for 30 min, thawed and dispersed vigorously in 3 ml 3 mM EDTA, pH 7.3. Cells were lysed with a Branson Ultrsonifier (50% cycle, intermittent), by successive freeze- thawing. Unlysed cells and inclusion bodies were removed by slow speed centrifugation (7,000 rpm, 10 min, 4°C). The supernatant containing the membranes and the cytoplasm was centrifuged at 35, 000 rpm using a 50Ti or 80Ti rotor for 90 min at 20°C in a Beckman L8-80 ultracentrifuge. The supernatant (cytoplasmic fraction) was collected and the whole membrane pellet was resuspended in 1 ml H_2O. Five hundred microlitres of Triton solution (4% Triton X-100, 2 mM $MgCl_2$, 50 mM Tris pH 7.5) was added to an equal volume of the whole membrane sample which was vortexed intermittently for 30 min at RT. The inner (soluble) membrane fraction was separated from the outer (insoluble) membrane fraction by centrifugation at 35,000 rpm for 90 min in a 50Ti or 80Ti rotor (Beckman L8-80) at 20°C. The outer membrane fraction was resuspended in Milli Q H_2O

2.13.6 Transfer of proteins for N-terminal sequencing

Samples were run in SDS on 15 % polyacrylamide gels, stained with Coomassie blue G250 and destained with 5 % (v/v) acetic acid. The samples were subsequently transfered to polyvinylidene difluoride (PVDF) (BioRad) at a transfer current of 200 mA for 2 hr in a Trans-Blot Cell (Biorad). The transfer buffer used was 25 mM Tris-HCl pH 8.3, 192 mM glycine and 10 % (v/v) methanol. Proteins were stained for 10 min in 0.025 % (w/v) Coomassie blue R-250 in 40 % (v/v) methanol and then air dried. The membrane was sent to Dr A. Gooley (Macquarie University, Sydney) to be analysed on a 470A Applied Biosystems Protein Sequencer.

2.14 Preparation of rabbit antisera

Adult rabbits were obtained from the Central Animal House of the University of Adelaide. Heat killed bacteria were used to immunize the rabbit. 50 ml of NB was inoculated with *V.anguillarum* and left shaking at 37°C overnight. The cells were pelleted (10 min, 5,000 rpm, IEC centrifuge), washed in saline, recentifuged as previously described and resuspended in saline. The bacterial saline suspension were then boiled (100°C) for 1-2

h. Bacterial concentration was determined by plating serial dilutions of pre-boiled bacteria onto NA. The following immunization protocol was used;

Day 1- 1 ml of 10^7 bacteria was mixed with 4% Na Alginate with 250 µl injected subcutaneously. This is repeated 3 more times on consecutive days.

Day 3- 500 µl of 1×10^7 bacteria injected intravenously (i/v)

Day 4- 500 µl of 5×10^7 bacteria injected i/v

Day 9- 500 µl of 1×10^8 bacteria injected i/v

Day 14- 500 µl of 5×10^8 bactria injected i/v

Day 18- 500 µl of 1×10^9 bacteia injected i/v

Two weeks after the last immunization the rabbit is bled out (aortic puncture). The titre of the antiserum is determined by testing different dilution's of antiserum by Western blotting.

2.14.1 Affinity Purification of antisera

Crude antiserum against the O1 serotype of *V. anguillarum* was purified by immunoaffinity chromatography as described by Salamitou *et al.* (1994). Whole cell samples of the strain 85-3954-2 (*V. anguillarum* O1) were electrophoresed on a 15% polyacrylamide SDS gel and transfered to nitrocellulose filter (2.13.4). The filter was incubated with 2% Ponceau S (Sigma) for 5 min and the protein bands stained. The filter was cut such that only LPS bound to the nitrocellulose was further treated. The membrane strip was incubated at RT for 1 h in Tris-buffered saline (TBS) containing 0.1% Tween 20 (Sigma) (TTBS) and 5 % skim milk powder to block non-specific sites. The filter was then incubated at RT for 16 h with 4 mls of unpurified antiserum. The unbound portion of antiserum was then removed and the membrane washed 3 times with TTBS followed by a single wash in TBS. The membrane was then cut into small pieces and the specific LPS antibodies were eluted by treatment with 1050 µl 0.2 M HCl-glycine buffer pH 2.2 at RT for 15 min. The eluate was neutralised by the addition of 450 µl 1 M K_2HPO_4 and dialysed against PBS at 4°C for 16 h. After dialysis, the purified antibodies were stored at -20°C. The antibody solution was subsequently used at dilutions of 1/500 to 1/1000.

2.14.2 Absorption

Antiserum was absorbed with either live or heat killed bacteria. Overnight culture (100 ml) was centrifuged at 7, 000 rpm for 10 min, (IEC Bench centrifuge) and resuspended in saline, and boiled for 1-2 h. The suspension was centrifuged at 4,500 rpm for 10 min and resuspended in 1 ml of saline (heat killed bacteria).

2-Loopfuls of live bacteria were isolated from freshly streaked TSA plates or 100 µl aliquots of heat killed bacteria were resuspended in 1 ml aliquots of antisera. Suspensions of antisera were left at 4°C for 16 h. The suspensions were centrifuged to pellet the bacteria (Hereaus Biofuge 15, 13, 000 rpm, 2 min) and the serum was collected. The procedure was repeated each day for 7 to 14 days before testing absobed serum by Western blotting.

2.15 Lipopolysaccharide procedures

2.15.1 Isolation of LPS: Method 1

Whole cell lysates (WCL) were prepared by the method of Hitchcock and Brown (1983). Cells were grown overnight in NB and 1.5 ml of bacterial suspension were collected by centrifugation in a microfuge (5 min, 15k rpm Heraeus Biofuge 15). The pellets were solubilized in 50 µl of lysing buffer containing 2% (w/v) SDS, 4% (v/v) β-mercaptoethanol, 10% (v/v) glycerol, 1 M Tris-HCl pH 6.8, and 0.1% (w/v) bromophenol blue. Lysates were heated at 100°C for 10 min. 2.5 µg of Proteinase K in 10 µl of lysing buffer was added to each sample and incubated at 60°C for 4 h. Samples were stored at -20°C.

2.15.2 LPS isolation: Method 2

Ultra pure LPS was isolated from the method described by Darveau and Hancock, (1983). *V. anguillarum* cells were grown in 1 litre of TSA supplemented with 1.5% NaCl to an O.D$_{600nm}$ of 0.6-0.8. The bacterial cells were harvested at 7, 000 rpm (Beckmann J2-M1, JA-10 rotor) for 15 min. The pellets were resuspended in 15 ml of 10 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$, 100 µg of DNase and 25 µg RNase per ml. The cell suspension was then passed twice through a French pressure cell at 15, 000 psi to ensure complete cell breakage. The cell lysate was sonicated (Branson Ultrasonifer) for two 30 sec bursts at a probe intensity of 75. To ensure efficient nucleic acid digestion, DNase and RNase were added again to final concentrations of 200 and 50 µg/ml, respectively. The suspension was incubated at 37°C for 2 h. After the incubation, 5 ml of 0.5 M EDTA/10 mM Tris-HCl, pH 8.0; 2.5 ml of 20% (w/v) SDS/10 mM Tris, pH 8.0 and 2.5 ml of 10 mM Tris-HCl, pH 8.0 was added to give a final volume of sample of 25 ml. The sample was vortexed to ensure solubilisation of the components and was then subjected to centrifugation at 20,000 rpm (Beckmann, J2-M1, JA-20 rotor) for 30 min at 20°C to remove peptidoglycan. The supernatent was decanted, and Pronase was added to give a final concentration of 200 µg/ml. The sample was then incubated at 37°C with constant shaking for 16 h. After this incubation, two volumes of 0.375 M $MgCl_2$ in 95% ethanol were added, mixed and cooled to 0°C by placing the samples in a dry ice ethanol-water bath or -20°C freezer. After the sample had cooled to 0°C, it was centrifuged (Beckmann J2-M1, JA-20 rotor) at 10, 000 rpm for 15 min at 0-4°C. The pellet obtained was resuspended in 25 ml of 2% (w/v) SDS, 0.1 M EDTA/10 mM Tris-HCl, pH 8.0 and sonicated as described above. The solution was then incubated at 85°C for 10 to 30 min to ensure removal of SDS-resistant proteins and then cooled to RT. Once at RT the pH was increased to 9.5 by the addition of 4 M NaOH. Pronase was then added to 25 μg/ml and the sample incubated at 37°C with constant shaking for 16 h. After the incubation, the LPS was precipitated with 2 volumes of 0.375 M $MgCl_2$ in 95% ethanol at 0°C as described above. The sample was then centrifuged at 10, 000 rpm for 15 min at 0°C (Beckmann J2-M1, JA-20 rotor). The pellet was resuspended in 15 ml of 10 mM Tris-HCl, pH 8.0, sonicated as previously described and centrifuged at 1, 000 rpm (IEC Centra-4X Bench centrifuge) for 5 min to remove insoluble Mg/EDTA complexes. The supernatant was then centrifuged at 45,000 rpm for 2 h (Beckmann L8-88, 50Ti/80Ti rotor). The pellet which contained the LPS was resuspended in distilled water.

2.15.3 LPS silver staining

Silver staining of LPS in polyacrylamide gels was performed using a modified method of Tsai and Frasch (1982). Distilled, deionized water which had been passed through a series of Millipore filters and had a conductivity of not more than 18 Mega ohms/cm was used to rinse all glassware and in the preparation of solutions. All reagents were analytical grade.

The following procedure was used:

i) fixation for 2 h in 40% (v/v) ethanol, 10% (v/v) acetic acid;

ii) oxidation for 5 min with 0.7% (w/v) periodic acid in 40% (v/v) ethanol, 10% (v/v) acetic acid;

iii) 3 washes with (MQ) water at 15 min each;

iv) staining for 10 min, in a solution containing 28 ml 0.1 N NaOH, 2 ml concentrated NH_4OH and 5 ml 20% (w/v) $AgNO_3$ in a total volume of 150 ml;

v) 3 washes with MQ water for 10 min each;

vi) developing in a solution of 50 mg citric acid and 0.5 ml formaldehyde in 1 litre. The citric acid was dissolved in MQ and heated to 37°C prior to the addition of formaldehyde which was added just before use. The reaction was stopped by the addition of 980 ml of MQ and 20 ml of acetic acid for 20 min. The developed gel is stored in the dark, submerged in MQ.

2.16 Microscopy studies

2.16.1 Immunogold electronmicroscopy

Immuno-gold electron microscopy was performed using an adaption of the method from Levine *et al.* (1984). The protein-A-gold particles were purchased from Amersham and were 10-20 µm in size. Colloidin-coated copper grids (300 mesh, TAAB) were placed coat down in 20 µl of washed bacterial suspension (PBS with 1 mM $MgCl_2$) for 5 min. Excess liquid was removed and then the grid was placed face down in 20 µl of antiserum (diluted 1:50) for 15 min. After two washes in PBS (pH 7.2) with 1 % (w/v) BSA, the grids were placed on 10 µl of Protein-A-gold (diluted 1:20) for a further 15 min. The grids were washed 2 X in PBS/BSA and once in distilled water and left at RT to dry. The grids were examined with a Phillips TM-100 electron microscope at an acceleration voltage of 80 kV and a magnification rnage of x27, 500 to 29, 000 was used.

2.16.2 Indirect Immunofluorescence

The technique is a minor modification of a published method by Klauser *et al.*, 1990. Round glass cover slips were boiled for 1 min in O.1 M HCl before storage in 95% ethanol prior to air-drying for use. Poly-L-lysine (100 µl, 0.1 mg/ml) was pipetted onto cover slips placed in a 24 well flat bottomed tissue culture tray (Costar) and incubated at RT for 5 min before being washed with PBS. Cells were formalin killed by treating 10 ml of pelleted culture (IEC bench centrifuge, 5000 rpm, 10 min) with 1 ml 2% paraformaldehyde, 0.1% gluteraldehyde (Sigma) for 20 min. The cells were washed 2 x PBS and resuspended in a final volume of 1 ml PBS. 100 µl of of the cell suspension was pipetted gently over the cover slip and 400 µl of sterile PBS added before centrifugation at 800 rpm (Heraeus Labofuge 400R) for 10 mins. The wells were aspirated dry and washed 3 times with PBS. The primary antibody (1:100) in 10% foetal calf serum (FCS) in PBS was added and incubated at 37°C for 90 min. After 3 washes with PBS the secondary antibody (1:80) (FITC conjugated goat anti-rabbit) (Sigma) in 10% FCS in PBS was added and left at 37 C for 30 mins followed by a further two PBS washes. The cover slips were gently aspirated dry and mounted upside down on a clean glass microscope slide with 3 µl of Moviol 4-88 and sealed with clear acrylic nail polish. Bacteria were photographed with Kodak TMAX400 film using an Olympus fluorescence microscope.

Chapter Three

General characterisation of *Vibrio anguillarum* serotypes O1-O10

3.1 Introduction

Vibrio anguillarum is a pathogen of feral and cultured fish. The disease process is not very well understood although some virulence factors have been partially characterised. The best characterised virulence factor in *V. anguillarum* is the iron-sequestering system that has been shown to be essential for virulence (Crosa, 1980; Crosa *et al.*, 1980). *V. anguillarum* is closely related to the human pathogen *V. cholerae,* which is the aetiological agent for cholera, as shown by the similarities of their *recA* genes (Stroeher *et al.*, 1994). This close relationship has lead some researchers to suggest that *V. anguillarum* may follow a similar infectious process as *V. cholerae* (Pierce *et al.*, 1985).

Since the first description of *V. anguillarum* infections, there have been numerous studies that have examined the composition and function of its outer membrane. These studies have focussed mainly on the use of lipopolysaccharides (LPSs) in serotyping (Sorenson and Larson, 1986) and on their role in virulence (Norqvist and Wolf-Watz, 1993). To date little is understood about the genetics of LPS (O-antigen) biosynthesis in *V. anguillarum* despite the role LPS plays in the disease process. Recently, there has been renewed interest in the role outer membrane proteins play in the pathogenicity of *V. anguillarum,* in particular that of a major outer membrane protein (MOMP) whose apparent molecular weight varies between 30 to 40 kDa in different strains (Suzuki *et al.*, 1994; Simon *et al.*, 1996).

In this chapter the focus is on characterising the *V. anguillarum* strains used in this study. Antisera were generated against *V. anguillarum* O1 and O4 and were found to react with LPS-related material and cross-reactive proteins. One protein of approximately 36 kDa was further investigated by N-terminal sequencing and localised via immunogold electron microscopy. The LPS profiles of the different *V. anguillarum* serotypes were examined by SDS-PAGE and silver staining. Finally, the distribution of the insertion sequence IS*1358* in *V. anguillarum* was investigated as this element is clearly associated with polysaccharide biosynthesis genes (ie. LPS) .

3.2 Results

3.2.1 Generation of an antiserum to *V. anguillarum* O1 strain 85-3954-2

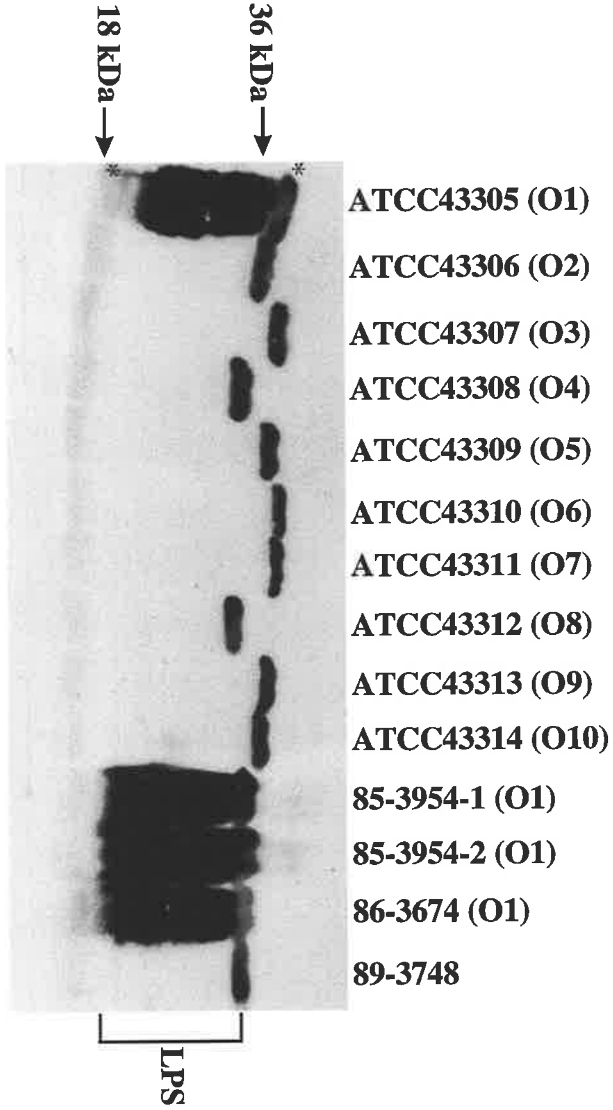
As the most predominant strains in disease outbreaks in cultured fish belong to *V. anguillarum* serotype O1, an antiserum was produced to enable further analysis, and screening of both a transposon mutant library and a cosmid bank. The antiserum was generated against heat killed bacteria of *V. anguillarum* O1 strain 85-3954-2 using the protocol described in the Materials and Methods (Section 2.14).

A Western immunoblot was performed to characterise the antiserum. The antiserum was found to react only with two proteins (one strong and one weak band) common to all the serotypes tested, and with material that could be the LPS of the O1 serotype due to the banding pattern observed (Fig. 3.1). Little cross-reaction was observed with *E. coli* K-12 or *V. cholerae* O139. Cross-reacting antibodies could be removed by absorption with the latter strains (data not shown).

The cross reacting antibodies to the smaller molecular weight protein (~18 kDa) were easily removed by absorbing the antiserum with whole cells of the O2 serotype (strain

Figure 3.1 Western immunoblot analysis of *V. anguillarum* serotypes O1-O10 using anti-O1 serum.

Whole cell lysates (equivalent to 1 x 10^8 cells) of *V.anguillarum* serotypes O1-O10 were electrophoresed on a SDS 15% polyacrylamide gel and transferred to nitrocellulose. Proteins were detected using polyclonal anti-*V. anguillarum* O1 serum. The strains and corresponding serotype are indicated at the top of the figure. The cross-reactive proteins of 36 kDa and 18 kDa are indicated by a * at the left side of the figure with relative size markers. The material with a LPS-like banding pattern is indicated by the abbreviation LPS on the right side of the figure.


18 kDa
36 kDa
ATCC43305 (O1)
ATCC43306 (O2)
ATCC43307 (O3)
ATCC43308 (O4)
ATCC43309 (O5)
ATCC43310 (O6)
ATCC43311 (O7)
ATCC43312 (O8)
ATCC43313 (O9)
ATCC43314 (O10)
85-3954-1 (O1)
85-3954-2 (O1)
86-3674 (O1)
89-3748
LPS

ATCC 43306)(data not shown). Antibodies against the larger protein (~36 kDa), which varies in size between the different serotypes, were not removed by absorption with either whole cells or outer membrane preparations of the O2 serotype (data not shown).

As the antiserum was generated with heat-killed bacteria, the reactions observed against the two proteins indicate either that these proteins are heat stable or that particularly immunogenic epitopes are recognised despite the denaturing of the protein during immunogen preparation.

The reactive material with LPS-like banding pattern was confirmed to be LPS by treating whole cells of the O1 serotype with proteinase K, and subsequent Western immunoblotting showed that the material was still detected. The absence of the reactive material in samples from other O-serotypes, and its proteinase K resistance strongly suggested that the reactive material is most likely to be LPS of *V. anguillarum* O1 (Fig. 3.2).

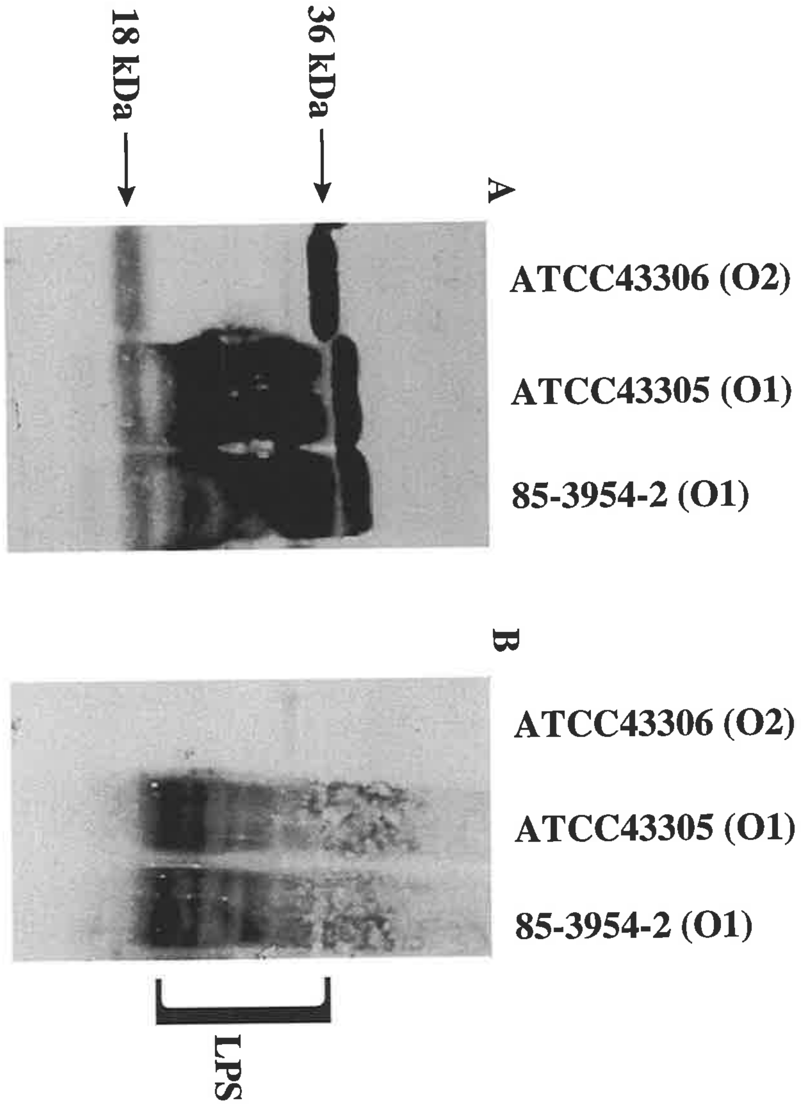
3.2.2 Characterisation of the 36 kDa protein

The outer membrane profile of *V. anguillarum* has been examined in numerous studies (Buckley *et al.*, 1981; Pazos *et al.*, 1993). A major outer membrane protein (MOMP) of approximately 36 kDa has previously been reported, however it was described as a weak immunogen (Chart and Trust, 1984). In addition, a 40 kDa protein was described by Simon *et al* (1996). This protein was found to be common to all 10 serotypes of *V. anguillarum* and varied in size. Antibodies generated against one protein were found to be cross-reactive to the equivalent protein in other serotypes. This protein was found to be a general diffusion porin.

In this study, the antiserum generated against heat killed bacteria of the O1 serotype reacted strongly to a predominant protein of ~ 36 kDa (and above) in whole cell (Fig. 3.1)

Figure 3.2 Western immunoblot analysis of proteinase K treated whole cell lysates of *V. anguillarum*.

Whole cell lysates of *V.anguillarum* strains were either treated with proteinase K (B) or not treated (A), electrophoresed on a SDS 15% polyacrylamide gel and transferred to nitrocellulose. The blot was incubated with polyclonal rabbit anti-*V. anguillarum* O1 serum used at a 1:1000 dilution. The strains and corresponding serotype are indicated at the top of the figure. The cross-reactive proteins of 36 kDa and 18 kDa are indicated by an arrow at the left side of the figure with relative size markers. The LPS-like material is indicated by the abbreviation LPS on the right side of the figure. Samples equivalent to 1 x 10^8 cells


18 kDa
36 kDa
A
ATCC43306 (O2)
ATCC43305 (O1)
85-3954-2 (O1)
B
ATCC43306 (O2)
ATCC43305 (O1)
85-3954-2 (O1)
LPS

and outer membrane samples (Section 2.13.5.1), of all serotypes tested (data not shown). Coomassie Brilliant Blue staining and SDS-PAGE showed that the protein was produced at high levels (in comparison to other proteins) in all the serotypes and that the protein varied in size between the serotypes with the smallest being ~36 kDa in ATCC 43312 (O8) and the largest at ~40 kDa in ATCC 43310 (O6) (Fig. 3.3). The size of the protein observed on the Coomassie Brilliant Blue stained gel matched the size of the protein bands previously observed by Western blotting (Fig. 3.1). The variation in the size of this protein appeared to be the main difference in the protein profile of the different serotypes (Fig. 3.3). The only strain that did not appear to produce this predominant protein was G20. G20 was isolated from goldfish and classified as *V. anguillarum* based on phenotypes (J. Carson, personal communication). In this thesis I have consistently found G20 to display different characteristics to the *V. anguillarum* type strains (data not shown) and therefore suggest that this strain does not belong with this species.

3.2.3 Lipopolysaccharide (LPS) profile of *V. anguillarum* serotypes O1-O10

The LPS profiles of the O1 and O2 serotypes have previously been reported (Chart and Trust, 1984; Amor and Mutharia, 1995; Tiainen *et al.*, 1997; Santos *et al.*, 1995) however little data has been reported for the remaining serotypes. Figure 3.4 shows the LPS profiles of the type *V. anguillarum* strains as determined by silver staining of LPS separated on a SDS 15% polyacrylamide gel. Each serotype has a unique LPS profile which would be reflected in the genes located within each O-antigen biosynthesis locus. Serotype O1 and O2 demonstrate profiles similar to those that have previously been reported (Fig. 3.4) (Chart and Trust, 1984; Santos *et al.*, 1995). Serotype O1 has a S-LPS (smooth LPS) with a very tight O-antigen modal length distribution which is in contrast to

Figure 3.3 Analysis of *V. anguillarum* O1-O10 proteins by SDS-PAGE.

Whole cell lysates (equivalent to 1 x 10^8 cells) of *V.anguillarum* serotypes O1-O10 were electrophoresed on a SDS 15% polyacrylamide gel and stained with Coomassie Brilliant Blue. The strains and corresponding serotype are indicated at the top of the figure. The MOMP is indicated with a * in each lane. Migration positions of the molecular mass standards (Pharmacia) are indicated on the left side of the figure (in kDa): soybean trypsin inhibitor (20.1), carbonic anhydrase (30), ovalbumin (43), bovine serum albumin (67) and phosphorylase b (94).

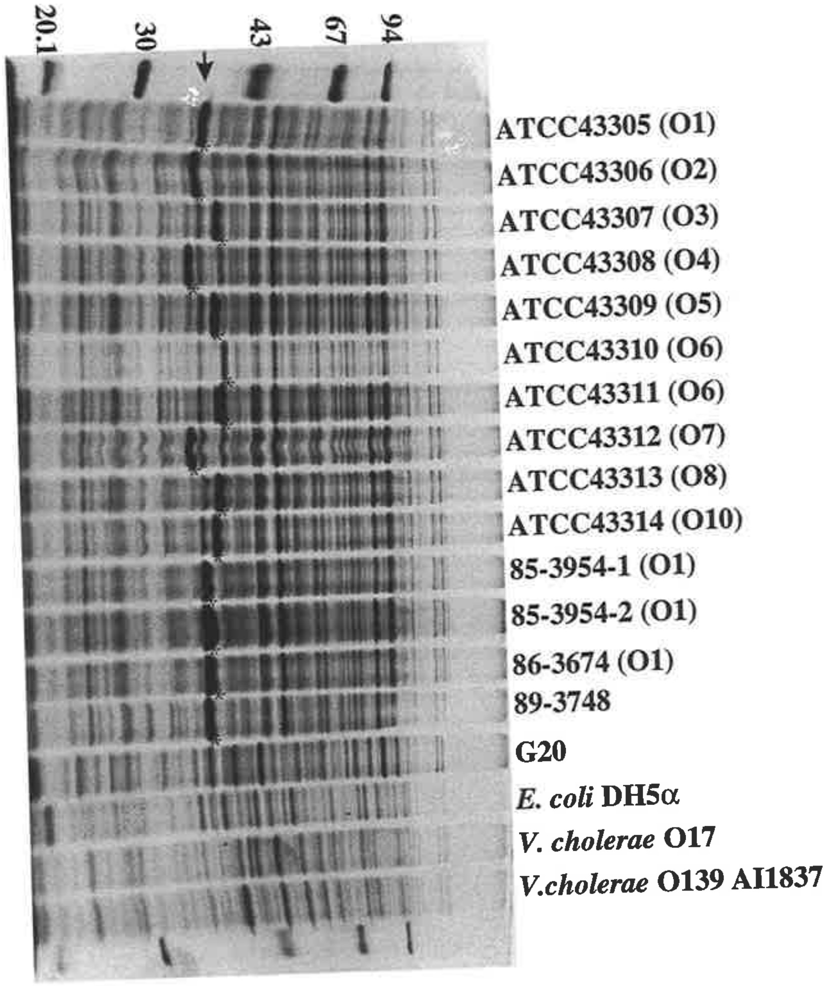
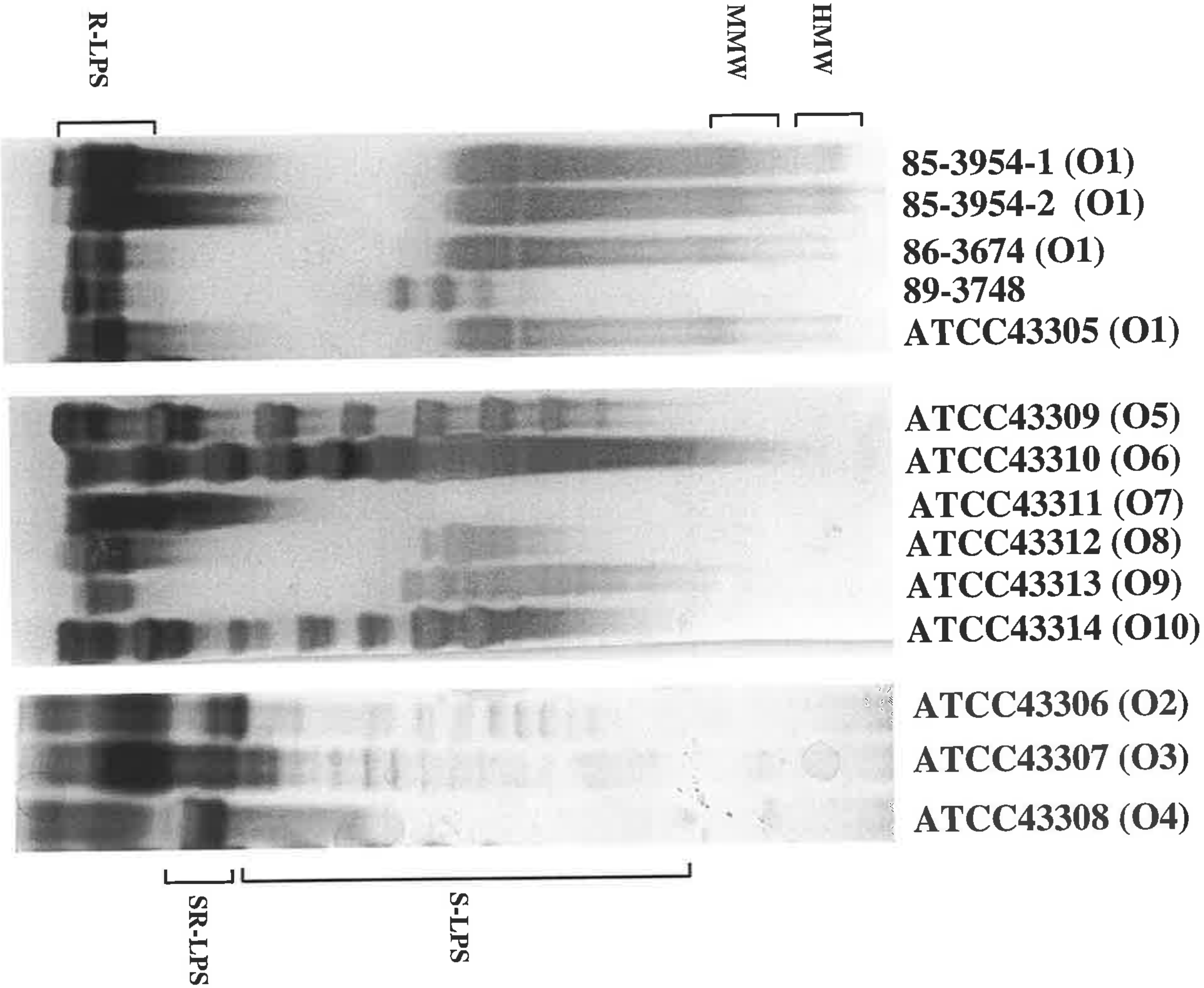

20.1
30
43
67
94
ATCC43305 (O1)
ATCC43306 (O2)
ATCC43307 (O3)
ATCC43308 (O4)
ATCC43309 (O5)
ATCC43310 (O6)
ATCC43311 (O6)
ATCC43312 (O7)
ATCC43313 (O8)
ATCC43314 (O10)
85-3954-1 (O1)
85-3954-2 (O1)
86-3674 (O1)
89-3748
G20
E. coli DH5α
V. cholerae O17
V.cholerae O139 AI1837

Figure 3.4 Analysis of LPS from *V. anguillarum* O1-O10.

This shows a silver-stained 15% SDS polyarylamide gel with lipopolysaccharides prepared from the indicated strains as described in Materials and Methods (Section 2.15.1).

The smooth LPS (S-LPS) consisting of lipid A-core and O-antigen, and the semi-rough LPS (SR-LPS) are indicated on the right side of the figure. Medium and high molecular weight material (MMW, HMW, respectively) and lipid A-core oligosaccharide (R-LPS) are indicated on the left side of the figure. Samples represent ~1 x 10^8.


R-LPS
MMW
HMW
85-3954-1 (O1)
85-3954-2 (O1)
86-3674 (O1)
89-3748
ATCC43305 (O1)
ATCC43309 (O5)
ATCC43310 (O6)
ATCC43311 (O7)
ATCC43312 (O8)
ATCC43313 (O9)
ATCC43314 (O10)
ATCC43306 (O2)
ATCC43307 (O3)
ATCC43308 (O4)
SR-LPS
S-LPS

O2 which has a ladder of S-LPS, and also a predominant semi-rough LPS species (SR-LPS) (Fig. 3.4). The O1 serotype also possesses high and medium molecular weight material which is thought to be S-LPS with longer O-antigen chains (Boesen *et al.*, 1999). It has been reported that strains within serotype O2 display heterogeneity of their LPS molecules and hence have been subdivided into two groups (Rassmussen, 1987). The type strain (ATCC 43306) used in this study to represent the O2 serotype belongs to the subgroup O2A.

Serotypes O3 to O10 display different LPS profiles and with the exception of serotypes O8 and O9 do not show a typical *Vibrio* LPS profile (ie S-LPS with a tight distribution of O-antigen chains) (Manning *et al.*, 1986; Alekseeva *et al.*, 1998). Serotype O4 has a single O-antigen unit on the lipid A/core and its LPS can therefore be described as being semi-rough (SR-LPS). Serotype O7 also appears to have only one or two O-antigen units attached to the lipid A/core and hence its LPS could also be described as SR-LPS. Serotypes O3, O5, O6 and O10 have a S-LPS phenotype in which a ladder pattern is observed.

3.2.4 Generation of an antiserum against the 36 kDa protein

A specific antiserum was produced against the 36 kDa protein so that it could be further investigated and characterised. Analysis of the LPS of *V. anguillarum* strain ATCC43308 showed that its LPS lacked O-antigen repeat units (Fig. 3.4; Section 3.2.3). It was therefore used to produce a comparable antiserum to that produced for the O1 serotype strain (Section 3.2.1). ATCC 43308 was heat killed and used to immunise a rabbit to raise an antiserum specific to the 36 kDa protein as described in Materials and Methods (Section 2.14). Western immunoblotting with this antiserum showed that it recognised a single protein in all the serotypes tested (Fig. 3.5). The antiserum was used at a dilution of 1:1000, which was equivalent to the antiserum generated against serotype O1, however, in this case no reaction was observed either to any other proteins or to LPS (Fig. 3.5).

3.2.5 Immunoelectron microscopy using polyclonal antiserum

To demonstrate that the 36 kDa protein was expressed on the cell surface, immunogold electron-microscopy was performed using the antisera generated against *V. anguillarum* O4 described above. When samples were labelled with the antiserum generated against O4, gold particles were only observed on the surface of *V. anguillarum* O4 (Fig. 3.6). Most of the binding was concentrated on the flagella, although some binding was observed on the body of the bacteria. When this antiserum was used in immunogold electron microscopy against the *V. anguillarum* O1 and O2 serotypes no gold particles were found to bind to the bacteria (Fig. 3.6). Binding of the gold particles observed on the body (and flagella) of the serotype O4 bacteria is likely to be due to antibodies recognising the 36 kDa protein. The absence of gold particle binding on bacteria of the O1 and O2 serotypes suggests that the 36 kDa protein, although present in these strains, is masked by the LPS O-antigen chains produced by these bacteria (Fig. 3.4).

3.2.6 N-terminal sequencing of the major outer membrane protein from *V. anguillarum* serotype O1.

To further analyse the 36 kDa protein of *V. anguillarum*, the protein from an outer membrane fraction of an O1 strain (85-3954-2) was partially purified by electroelution

Figure 3.5: Western immunoblot analysis of *V. anguillarum* serotypes O1-O10 using strain ATCC43308 (serotype O4) serum.

Whole cell lysates (equivalent to 1 x 10^8 cells) of *V.anguillarum* serotypes O1-O10 were electrophoresed on a SDS 15% polyacrylamide gel and transferred to nitrocellulose. Proteins were detected using polyclonal anti-*V. anguillarum* O4 serum (1:1000 dilution). The strains and corresponding serotype are indicated at the top of the figure. The protein band in each lane has a molecular weight of approximately 36-40 kDa.

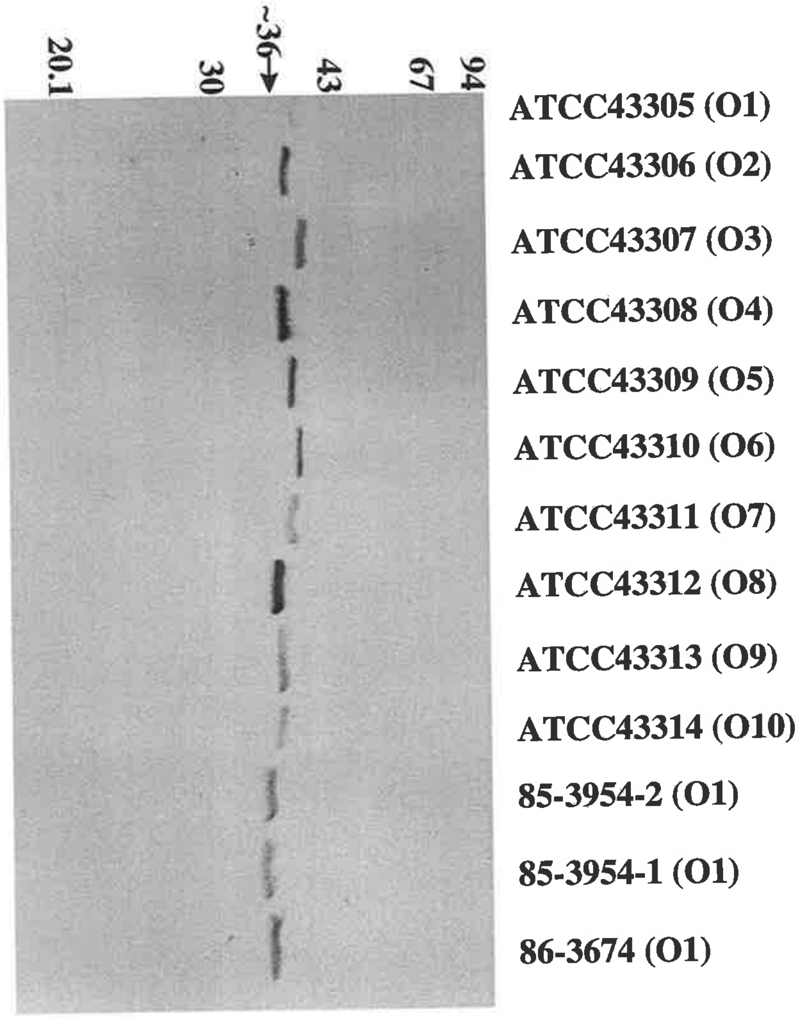

20.1
30
~36→
43
67
94
ATCC43305 (O1)
ATCC43306 (O2)
ATCC43307 (O3)
ATCC43308 (O4)
ATCC43309 (O5)
ATCC43310 (O6)
ATCC43311 (O7)
ATCC43312 (O8)
ATCC43313 (O9)
ATCC43314 (O10)
85-3954-2 (O1)
85-3954-1 (O1)
86-3674 (O1)

Figure 3.6: Immunogold Electronmicroscopy using anti-*V. anguillarum* O4 serum.

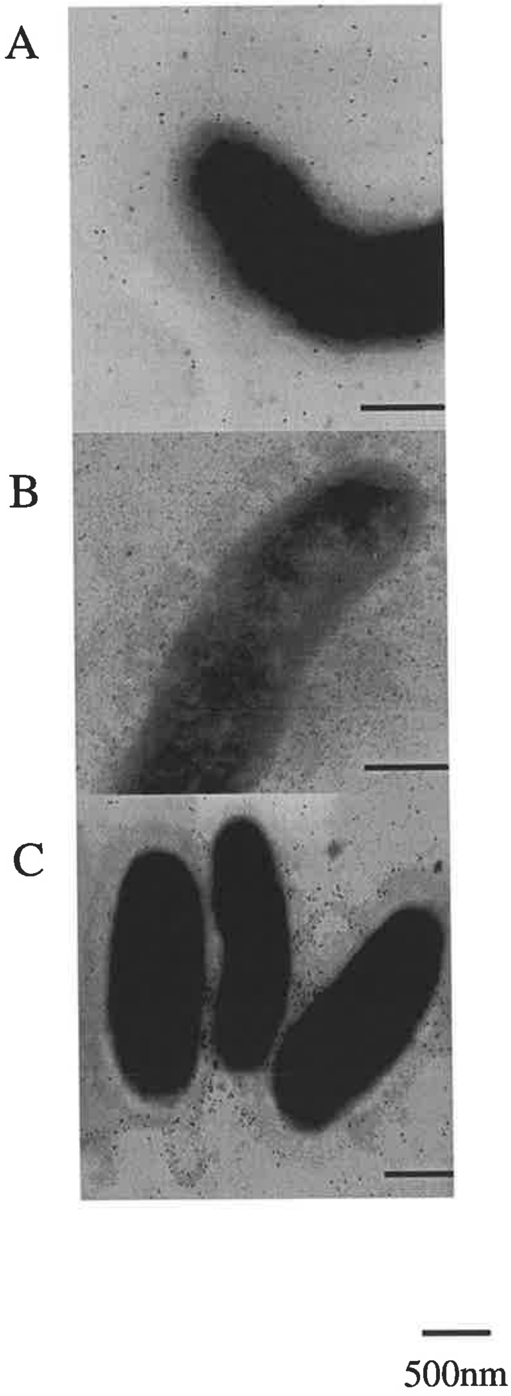
Electronmicrographs showing binding of polyclonal antibodies generated against *V. anguillarum* O1 (85-3954-2) to *V. anguillarum* O1 (85-3954-2), O2 (ATCC43306) and O4 (ATCC43308).

The antiserum was used at 1:50 dilution, the gold particles were 10 nm in diameter and the magnification used ranged from 28.5K-29.5K.

A: 85-3954-2 (O1)

B: ATCC43306 (O2)

C: ATCC43308 (O4)


A
B
C
500nm

from SDS-15% polyacrylamide gel (Materials and Methods, Section 2.13.6) (Fig. 3.7). The purified protein was transferred to PVDF (Polyvinylidene Difluoride) membrane and its N-terminal sequence was determined (Materials and Methods, Section 2.13.6). The N-terminal sequence obtained (Fig. 3.8) was analysed using the BlastP and the Genbank, SWISS PROT, PIR databases. The 36 kDa protein was not homologous to any previously reported *V. anguillarum* protein in the databases. However, the sequence shown in Figure 3.9 was 94% identical (as determined by PROSIS) to the published N-terminal sequence of the MOMP reported by Suzuki *et al.* (1994, 1996) and Simon *et al.* (1996), which has not been lodged in the databases. The sequence of the 36 kDa protein and the MOMP differed by only one amino acid (Valine/Tyrosine at position 4, Fig. 3.8) suggesting that they are the same protein. The MOMP was also 38% identical to an outer membrane protein (P2) from *Haemophilias influenzae* (Munson *et al.*, 1989a, b; Forbes *et al.*, 1992; Chong *et al.*, 1993; Bell *et al.*, 1994) (Fig. 3.8). The P2 protein has been shown to be immunologically important in *H. Influenzae,* and also shows size variation between serotypes as observed in the serotypes of *V. anguillarum* described herein.

3.2.7 Cloning the outer membrane protein of *V. anguillarum* O1

Using a cosmid bank constructed from the *V. anguillarum* O1 strain 85-3954-2, an attempt was made to isolate the gene responsible for the production of the major outer membrane protein (MOMP). Approximately 1000 cosmid clones were constructed (Materials and Methods, Section 2.9.2) and screened for the production of the serotype O1 MOMP by colony immunoblotting of lysed and unlysed cells with antiserum generated against serotype O4 (ATCC 43308). Although numerous positive clones were detected,

Figure 3.7:Analysis of a ~36 kDa protein (MOMP) purified from *V. anguillarum* O1

A) The MOMP of *V. anguillarum* O1 from strain 85-3954-2 was purified by electroelution (Section 2.13.5) and 10 µg was electrophoresed on a SDS 15% polyacrylamide gel which was then stained with Coomassie Brilliant Blue.

B) Purified MOMP (10 µg) and a whole cell sample (representing 1 x 10^8 cells) of *V. anguillarum* O1 were analysed by Western immunoblotting using the serum generated against *V. anguillarum* O4 (1:1000 dilution). Migration positions of the molecular mass standards (Pharmacia) are indicated on the left side of the figure (in kDa): soybean trypsin inhibitor (20.1), carbonic anhydrase (30), ovalbumin (43), bovine serum albumin (67) and phosphorylase b (94).

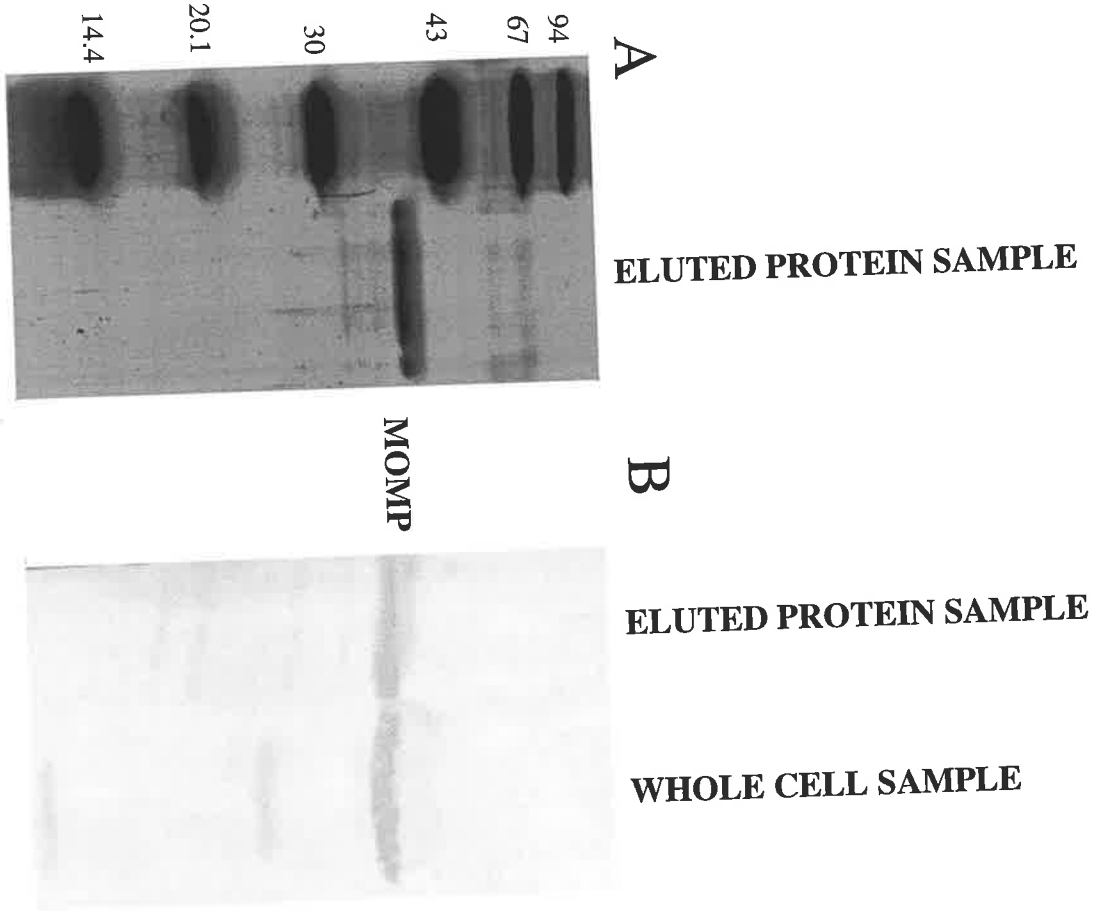

14.4
20.1
30
43
67
94
A
ELUTED PROTEIN SAMPLE
MOMP
B
ELUTED PROTEIN SAMPLE
WHOLE CELL SAMPLE

Figure 3.8: N-terminal alignment of major outer membrane protein (MOMP)

V. ang^a	G	E	L	V	N	Q	D	G	T	S	L	E	M	G	G	R
V. ang^b	G	E	L	Y	N	Q	D	G	T	S	L	E	M	G	G	R
H. inf	A	V	V	Y	N	N	E	G	T	N	V	E	L	G	G	R
H. som	T	T	V	Y	N	Q	N	G	T	K	V	D	V	G	G	R
			.		*	.	.	*	*		.	.	.	*	*	*

V. ang^a, to *V. anguillarum* O1 major outer membrane protein (this study)
V. ang^b, to *V. anguillarum* O1 major outer membrane protein (Suzuki *et al.*, 1996)
H. inf, to *Haemophilus influenzae* outer membrane protein, P2 (GB, U08205) (Bell *et al.*, 1994)
H. som, to *H. sommus* outer membrane protein (PIR, A49226)

(*) corresponds to identical amino acid residues
(.) corresponds to similar amino acid residues
The shaded residues are those that are identical in all of the aligned proteins. The accession number/database or reference for each protein is provided. GB, Genbank.

further analysis by Western immunoblotting showed that the clones did not express the protein (data not shown). Hence no gene for the MOMP was obtained.

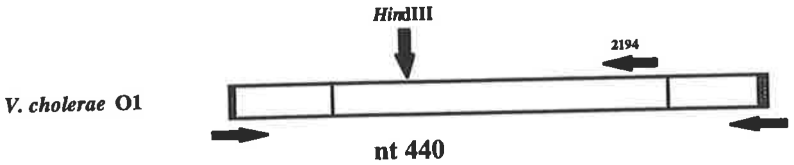
3.2.8 IS*1358* in *V. anguillarum*

V. cholerae and *V. anguillarum* are the most closely related species within the genus Vibrionaceae (Dorsch *et al.*, 1992). Several elements are common between these species including the novel genetic element called VCR (C.A Clark, personal communication). Therefore it was possible that the recently described insertion sequence IS*1358* (Fig. 3.9) would be found in *V. anguillarum*. I was interested in defining the distribution of IS*1358* in *V. anguillarum* with the view to obtaining data on the O-antigen biosynthesis genes given the previous association with polysaccharide loci in *V. cholerae* O1, O69 and O139 (Stroeher *et al.*, 1995; Bik *et al.*, 1996).

Using a DIG-PCR labelled probe of the entire IS*1358* element (1326 bp) from *V. cholerae* O139 (generated using oligonucleotide #773, Table 2.4) (Section 2.11.4), the strains which represented the ten serotypes of *V. anguillarum* were initially screened by colony hybridisation for the presence of the element (data not shown). Based upon the colony hybridisation data which showed that some of the serotypes contained the element, Southern hybridisation of *Hin*dIII digested chromosomal DNA was performed to confirm the above data and to determine the copy number of IS*1358*-like elements in each of the positive serotypes (Fig. 3.10a). *Hin*dIII is known to cut IS*1358* elements at nt 440 (Stroeher *et al.*, 1995) (Fig. 3.9). Hence two fragments constitute one IS*1358* element assuming this restriction site is present in all IS*1358* elements. The Southern hybridisation data confirmed the colony blot data which had indicated that the IS*1358* elements are not present in the type strains representing serotypes O5, O6, O8 and O10 (Fig. 3.10a, Table 3.1). Of the other strains tested, serotypes O1, O3, O4 and the non-typed strain (89-3748)

Figure 3.9: Schematice diagram of the IS*1358* elements from *V. cholerae* O1 and O139.

Representation of the genetic organisation of the IS*1358* elements found in *V. cholerae* O1 and O139. The position of oligonucleotide #2194 is indicated, as is the conserved *Hin*dIII restriction site. Vertical lines represent the stop codons of the truncated ORFs of *V. cholerae* O1. The inverted repeats are indicated by arrows.


HindIII
2194
V. cholerae O1
nt 440

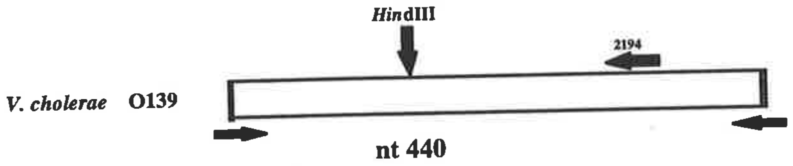
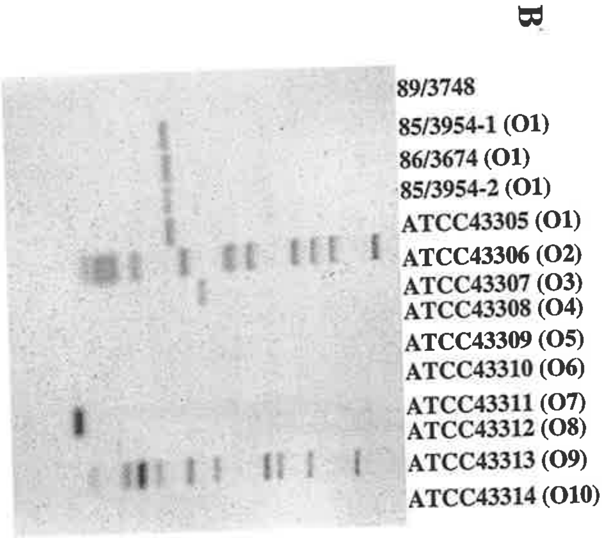
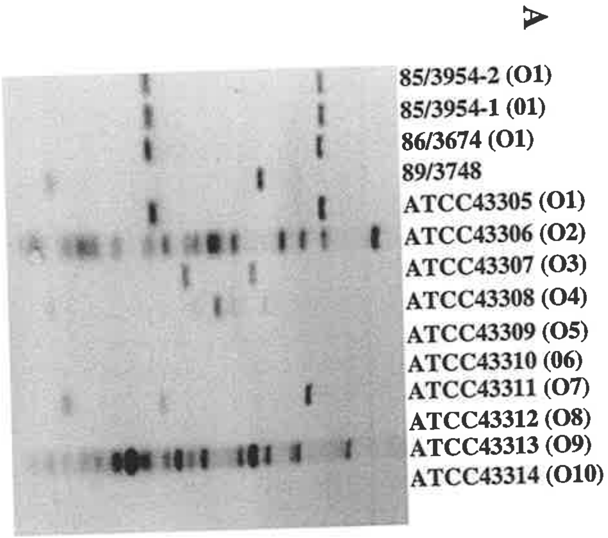

HindIII
2194
V. cholerae O139
nt 440

Figure 3.10: Detection of IS*1358* in *V. anguillarum*.

Chromosomal DNA from the various *V. anguillarum* strains were digested with the restriction enzyme *Hin*dIII. Strain used are indicated on the figure.

Initially, the number of copies were determined by Southern hybridisation with a PCR-DIG labelled probe of IS*1358* (generated with oligonucleotide #773) (Table 2.4). The assumption was that every IS*1358* element contains a conserved *Hin*dIII restriction site, and therefore two bands correspond to a single IS*1358* element (A). The use of a DIG-labelled oligonucleotide probe (#2194) (Table 2.4) in Southern hybridisation confirmed the copy number determined by the initial southern (B). Oligonucleotide #2194 binds to the second *Hin*dIII fragment of IS*1358* as shown in Fig. 3.10.


B
89/3748
85/3954-1 (O1)
86/3674 (O1)
85/3954-2 (O1)
ATCC43305 (O1)
ATCC43306 (O2)
ATCC43307 (O3)
ATCC43308 (O4)
ATCC43309 (O5)
ATCC43310 (O6)
ATCC43311 (O7)
ATCC43312 (O8)
ATCC43313 (O9)
ATCC43314 (O10)


A
85/3954-2 (O1)
85/3954-1 (01)
86/3674 (O1)
89/3748
ATCC43305 (O1)
ATCC43306 (O2)
ATCC43307 (O3)
ATCC43308 (O4)
ATCC43309 (O5)
ATCC43310 (06)
ATCC43311 (O7)
ATCC43312 (O8)
ATCC43313 (O9)
ATCC43314 (O10)

all have a single copy of IS*1358*. Three serotypes were found to contain more than one copy of IS*1358*. Serotype O7 (strain ATCC 43311) contained 2 copies while the remaining two serotypes, O2 (strain ATCC 43306) and O9 (strain ATCC 43313), were found to have multiple copies of (greater than 6) IS*1358* elements (Fig. 3.10a). The serotypes that contain an IS*1358* element show restriction fragment length polymorphism (RFLP) of the *Hin*dIII fragments which is not surprising as the DNA flanking the element is likely to be unique to each serotype. Importantly however, there is homogeneity between strains in the O1 serotype which were tested in this study even though they are from different geographical locations. In the O1 serotype, IS*1358* hybridises to conserved *Hin*dIII fragments of approximately 4.8 kb and 1.8 kb. The other fragment sizes from the various strains and serotypes are listed in Table 3.1.

Southern hybridisation using an internal oligonucleotide probe (#2194) (Fig. 3.9; Table 2.4) to IS*1358* showed that the O2 serotype contained 10-11 copies and the O9 serotype possessed 10 copies of IS*1358* (Fig. 3.10b). Serotype 07 was found to contain one complete IS*1358* element and a partial element. Further Southern hybridisations using different restriction enzymes with both the complete *V. cholerae* O139 IS*1358* and oligonucleotide probe (#2194) confirmed the number of IS*1358* elements determined here to be accurate (data not shown).

3.3 Summary and Discussion

This chapter described the production and characterisation of two anti- *V. anguillarum* sera. One antiserum contained antibodies to both the LPS and the MOMP of *V. anguillarum* O1. The second antiserum was generated specifically for studying the MOMP and was used in Western immunoblot analysis and immunogold electron microscopy which showed that the MOMP was most likely masked by the O-antigen

Table 3.1: Distribution of IS*1358* in *Vibrio anguillarum*

Strain	Serotype	Presence by colony blot/PCR	number of copies	*Hind*III fragments (kb)
ATCC43305	O1	+/+	1	4.8/1.8
ATCC43306	O2	+/+	9	#
ATCC43307	O3	+/+	1	3.4/2.6
ATCC43308	O4	+/-	1	2.8/1.0
ATCC43309	O5	-/-	0	n/a
ATCC43310	O6	-/-	0	n/a
ATCC43311	O7	+/+	2	4.7/1.9/1.0
ATCC43312	O8	-/-	0	n/a
ATCC43313	O9	+/+	10	#
ATCC43314	O10	-/-	0	n/a
86/3674	O1	+/+	1	4.8/1.8
85/3954-1	O1	+/+	1	4.8/1.8
85/3954-2	O1	+/+	1	4.8/1.8
89/3748	non-typed	+/-	1	3.6/1.2

+ : IS*1358* could be detected,- : no copy of IS*1358* could be detected
# : numerous bands of different sizes

component of LPS. Masking of outer membrane proteins by S-LPS has been reported previously for *Escherichia coli* by van der Ley *et al.* (1986).

The LPS profiles of the strains used in this study were examined by SDS-PAGE and silver staining. The LPS pattern of each serotype was different. Most of the serotypes possessed smooth LPS (S-LPS) with the exception of O4 and O7 which appeared to have semi-rough LPS (SR-LPS). The LPS of *V. anguillarum* O1 possesses high molecular weight O-antigen which is thought to play a role in serum resistance and hence survival in the fish host (Boesen *et al.*, 1999).

During the course of this study the outer membrane protein described in this chapter was isolated and its two dimensional structure was determined (Simon *et al.*, 1996). These data confirmed the protein to be an outer membrane porin with characteristics of a general diffusion porin which is similar to OmpF from *E. coli* (Benz and Bauer, 1988). The gene for the MOMP, despite the attempt in this study, has yet to be cloned. It is not clear why the cloning attempt failed, however, it is possible that either expression of the MOMP in *E. coli* caused outer membrane instability or the protein was unable to insert into the outer membrane, therefore not allowing the gene to be isolated using the method described in this chapter. Further investigation is required to determine a potential role in pathogenicity given the homology to the immunogenic P2 protein of *H. influenzae* as reported in this chapter.

Finally, this chapter also determined that the recently characterised insertion sequence IS*1358* was widely distributed in *V. anguillarum* with serotypes O1, O2, O3, O4, O7, O9, and the non-typed strain (89/3748), all containing at least one copy of the element. Further characterisation of the DNA sequences adjacent to IS*1358 in V. anguillarum* is described in the next chapter.

Chapter Four

Identification and characterisation of IS*1358* in *V. anguillarum* and its linkage to polysaccharide biosynthesis related genes

4.1 Introduction

IS*1358* contains a putative transposase (*tnpA*) which shows homology to an open reading frame (ORF) found in RHS (Rearrangement Hot Spot) elements of *E. coli* called H-repeats (Stroeher *et al.*, 1995). These RHS elements are involved in chromosomal rearrangements (Lin *et al.*, 1984: Zhao *et al.*, 1993). IS*1358* has 17 bp inverted repeats at either end, with another smaller 12 bp inverted repeat covering the ribosome binding site of *tnpA* (Stroeher *et al.*, 1995). This genetic arrangement is observed in other IS elements such as IS*10* and IS*50*, and is thought to be involved in the regulation of expression of the transposase (Davis *et al.*, 1985; Schulz & Reznikoff, 1991). Most recently, a H-repeat homologue has been found located in the type 3 capsule locus of *Streptococcus pneumoniae* and is the first example of this type of element in a Gram-positive pathogen (Yother *et al.*, 1997). IS*1358* also shows strong homology to ISAs*1*, a known mobile genetic element which is associated with the expression of a paracrystalline surface protein array (A-layer) in *Aeromonas salmonicida* (Gustafson *et al.*, 1994).

This chapter describes the distribution, sequence and location of IS*1358* elements in different *Vibrio spp* and shows that these elements are associated with polysaccharide biosynthesis-related genes in *V. anguillarum*.

4.2 Results

4.2.1 Distribution of IS*1358* in *Vibrio* species

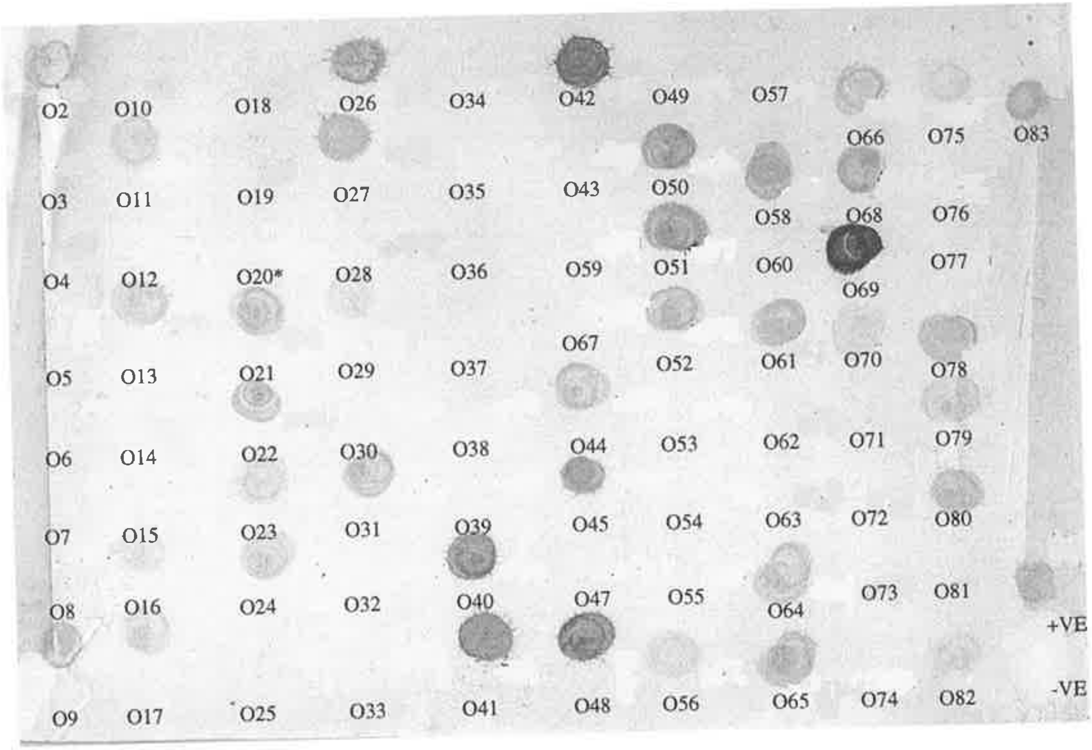
The presence of IS*1358* has already been reported in *V. cholerae* O1 and O139 (Stroeher *et al.*, 1992; Stroeher *et al.*, 1995). The presence of IS*1358* in a number of *V. cholerae* serotypes has been reported for O2, O37, O45, O56, O60, O69 (Bik *et al* 1995) and O22, O155 (Dumontier *et al.*, 1998). In this study it was decided to further investigate the distribution of IS*1358* in all of the 155 *V. cholerae* serotypes. Table 4.1 summarises the findings of dot-blot DNA-hybridisations (Fig. 4.1) using a PCR DIG-labelled probe of the whole IS*1358* element generated from *V. cholerae* O139 (AI-1837) against all the Vibrionaceae tested. The only non-O1 *V. cholerae* serotype not tested in this study was O20 which was not available. Of the 154 *V. cholerae* strains tested in this study (including those serotypes previously reported to possess a IS*1358* element), 65 proved to be positive for IS*1358*. Interestingly the DNA hybridisation (Fig. 4.1) performed in this study did not indicate the presence of an IS*1358* element in *V. cholerae* serotypes O37 and O60, which may be explained either by the use of different strains or hybridisation methods. These serotypes were previously reported by Bik *et al* (1995) to possess IS*1358*. Recent experiments have shown that some of the non-O1 *V. cholerae* serotypes found to contain the IS*1358* element in this study also have it in multiple copies (U.H. Stroeher, personal communication; Dumontier *et al.*, 1998).

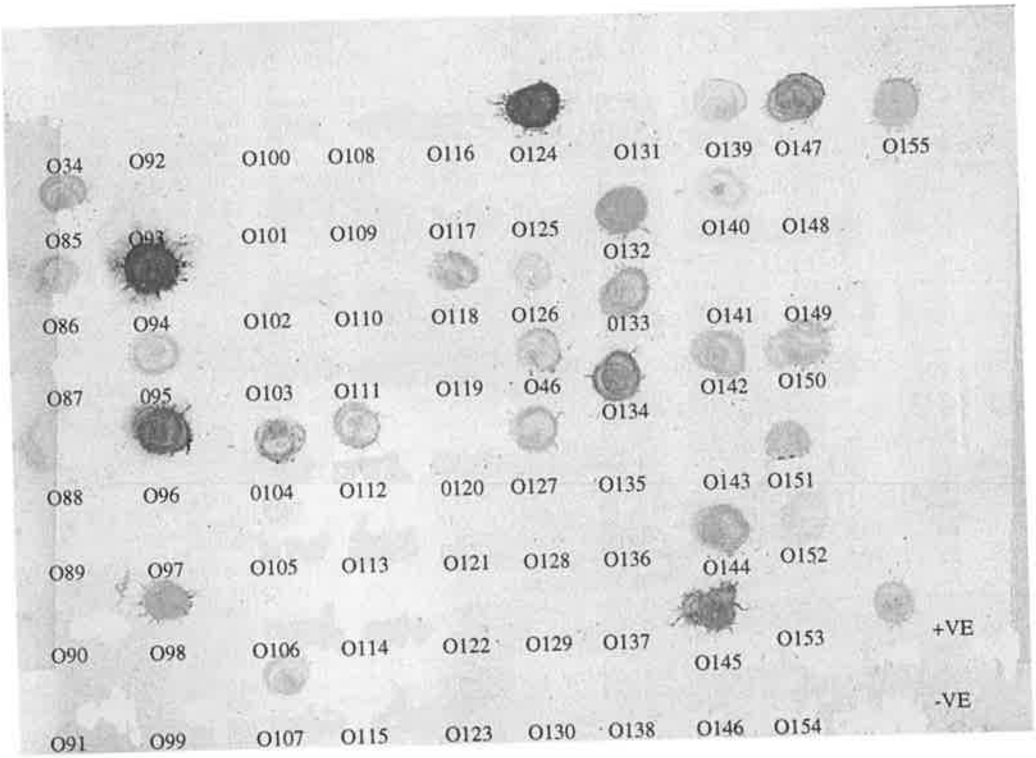
To determine if IS*1358* was more widespread, a number of non-*V. cholerae* Vibrionaceae were tested for the presence of the element using either dot-blot DNA hybridisation or Southern hybridisation (probed with the complete *V. cholerae* O139 IS*1358* element, Fig. 4.1). The strains that were tested belonged to species of *V. anguillarum*, *V. fluvialis*, *V. mimicus* and *V. parahaemolyticus*. The only strains that were

Figure 4.1: DNA hybridisation of *Vibrio spp.* using DIG-labelled IS*1358*.

Cultures were grown in microtitre trays to an OD_{600} equivalent to 0.6. 10 µl samples were spotted onto nylon membrane and treated as described in Materials and Methods (Section 2.10.9). The probe used to determine the presence of IS*1358* in the *V. cholerae* serotypes was a PCR-DIG-labelled IS*1358* element generated from chromosomal DNA from *V. cholerae* O139 using oligonucleotide #773 (Table 2.4). DNA hybridisation was performed as described in Materials and Methods (Section 2.10.10)

The serotype/strain probed are indicated on the figure and are listed in Table 2.1. +ve, *V. cholerae* O1; -ve, *E. coli* DH5α.


O2
O10
O18
O26
O34
O42
O49
O57
O66
O75
O83
O3
O11
O19
O27
O35
O43
O50
O58
O68
O76
O4
O12
O20*
O28
O36
O59
O51
O60
O77
O69
O67
O5
O13
O21
O29
O37
O52
O61
O70
O78
O6
O14
O22
O30
O38
O44
O53
O62
O71
O79
O7
O15
O23
O31
O39
O45
O54
O63
O72
O80
O8
O16
O24
O32
O40
O47
O55
O64
O73
O81
+VE
-VE
O9
O17
O25
O33
O41
O48
O56
O65
O74
O82


O34
O92
O100
O108
O116
O124
O131
O139
O147
O155
O85
O93
O101
O109
O117
O125
O132
O140
O148
O86
O94
O102
O110
O118
O126
O133
O141
O149
O87
O95
O103
O111
O119
O46
O134
O142
O150
O88
O96
O104
O112
O120
O127
O135
O143
O151
O89
O97
O105
O113
O121
O128
O136
O144
O152
O90
O98
O106
O114
O122
O129
O137
O145
O153
+VE
-VE
O91
O99
O107
O115
O123
O130
O138
O146
O154

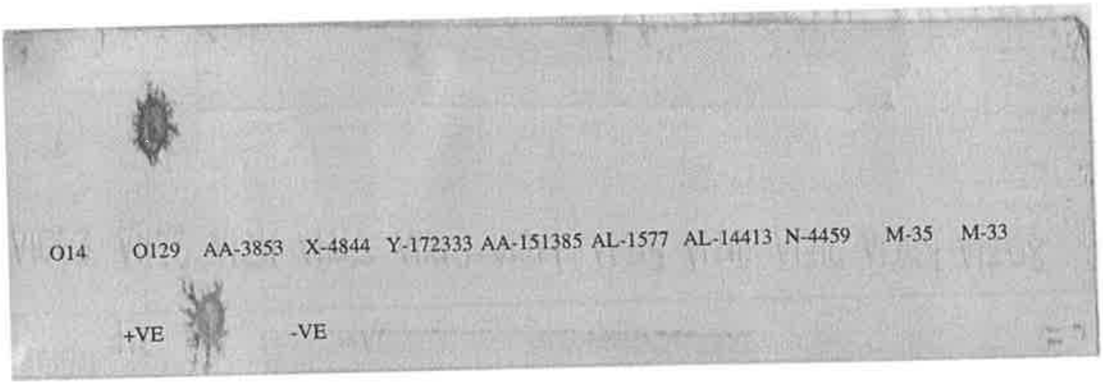

O14
O129
AA-3853
X-4844
Y-172333
AA-151385
AL-1577
AL-14413
N-4459
M-35
M-33
+VE
-VE

Table 4.1: Presence of IS*1358* in Vibrionaceae

Serogroup	IS*1358*	Serogroup	IS*1358*	Serogroup	IS*1358*	Serogroup	IS*1358*	Serogroup	IS*1358*
O1	+	O41	+	O81	-	O121	-	ATCC43306	+
O2	+	O42	+	O82	+	O122	-	ATCC433O7	+
O3	-	O43	-	O83	+	O123	-	ATCC433O8	+
O4	-	O44	+	O84	-	O124	+	ATCC433O9	-
O5	-	O45	+	O85	+	O125	-	ATCC433O10	-
O6	-	O46	+	O86	+	O126	+	ATCC43311	+
O7	-	O47	-	O87	-	O127	+	ATCC43312	-
O8	-	O48	+	O88	-	O128	-	ATCC43313	+
O9	+	O49	-	O89	-	O129	+	ATCC43314	-
O10	-	O50	+	O90	-	O130	-	89/3748	+
O11	+	O51	+	O91	-	O131	-	*V. para*	
O12	-	O52	+	O92	-	O132	+	10884	-
O13	+	O53	-	O93	-	O133	+	AA-3853	-
O14	-	O54	-	O94	+	O134	+	X-4844	-
O15	-	O55	-	O95	+	O135	-	Y-17233	-
O16	+	O56	+	O96	+	O136	-	*V.fluv*	
O17	+	O57	-	O97	-	O137	-	AA-15385	-
O18	-	O58	-	O98	+	O138	-	AL-1577	-
O19	-	O59	-	O99	-	O139	+	AL-14413	-
O20	Nd	O60	-	O100	-	O140	+	*V.mim*	
O21	+	O61	+	O101	-	O141	-	V800	-
O22	+	O62	-	O102	-	O142	+	M-33	-
O23	+	O63	-	O103	-	O143	-	M-35	-
O24	+	O64	+	O104	+	O144	+	N-4459	-
O25	-	O65	+	O105	-	O145	+		
O26	+	O66	+	O106	-	O146	-		
O27	+	O67	-	O107	+	O147	+		
O28	-	O68	+	O108	-	O148	-		
O29	+	O69	+	O109	-	O149	-		
O30	-	O70	+	O110	-	O150	+		
O31	+	O71	-	O111	-	O151	+		
O32	-	O72	-	O112	+	O152	+		
O33	-	O73	-	O113	-	O153	-		
O34	-	O74	-	O114	-	O154	-		
O35	-	O75	+	O115	-	O155	+		
O36	-	O76	-	O116	-	*V. ang.*			
O37	-	O77	-	O117	-	85-3954-2	+		
O38	-	O78	+	O118	+	85-3954-1	+		
O39	-	O79	-	O119	-	86/3674	+		
O40	+	O80	+	O120	-	ATCC43305	+		

All strains belong to the species *V. cholerae* unless otherwise stated.
V.ang: *V. anguillarum*
V. para: *V. parahaemolyticus*
V. fluv:*V.fluvialis*
V.mim:*V.mimicus*
+: present; -: absent
Nd: not done

found to be positive were those that belonged to some serotypes of *V. anguillarum* (Chapter 3, Section 3.2.8). This may indicate that IS*1358* is either specific to *V. cholerae* and *V. anguillarum* or that more strains in these other species need to be tested to determine the true distribution of this element. However, it should be noted that amongst the Vibrionaceae, *V. cholerae* and *V. anguillarum* are the most closely related at the DNA level based on analysis of the *recA* gene (Stroeher *et al.*, 1994).

4.2.2 PCR Amplification of IS*1358*

The Southern and dot-blot hybridisation probing of various serotypes of *V. anguillarum* with IS*1358* (Fig. 3.10a, b) indicated that 10 out of the 14 strains tested contained an IS*1358* element. Using an oligonucleotide (#773) that recognises the 17 bp inverted repeats at the 5' and 3' ends of previously sequenced IS*1358* elements (Stroeher *et al.*, 1995), I attempted to PCR-amplify the equivalent elements from the *V. anguillarum* strains. Using chromosomal DNA from each strain as a template, the elements were successfully amplified from 8 out of the 10 strains that were positive by Southern hybridisation. The elements could not be amplified from strains ATCC43308 (O4) and 89/3748 (non-typed). This suggests that the inverted repeat sequence which is used in the oligonucleotides to amplify the element was either missing or divergent in these particular IS*1358*-related sequences. Alternatively, the IS*1358* elements in these strains may only be partial copies which has been observed previously with one of the IS*1358* copies in *V. cholerae* O139 (Stroeher *et al.*, 1995), and with H-repeats in *E. coli* (Hill *et al.*, 1994: Lin *et al.*, 1984). Using an internal IS*1358* oligonucleotide probe (#2194) (Table 2.4), serotype O4 (ATCC43308) and strain 89/3748 were shown to contain partial copies of IS*1358* (Fig. 3.10b, Chapter 3), explaining why these elements could not be PCR amplified.

4.2.3 Cloning, sequencing and analysis of different IS*1358* elements

The PCR-amplified DNA corresponding to IS*1358* elements from the different *V. anguillarum* serotypes were purified and cloned into pGEM-T. The cloned IS*1358* elements (pPM4947 to pPM4954, pPM4958) were sequenced using -21 M13 (forward) and M13 (reverse) primers, and the sequence of each element was then completed using internal oligonucleotide primers (#2091, #2092, #2193, #2194) (Table 2.4; Fig. 4.4). Multiple clones of each element were sequenced to reduce the likelihood of sequencing PCR errors and regions of ambiguity were clarified by re-sequencing. Figure 4.2 and Table 4.2 summarise the similarity between the cloned IS*1358* elements of *V. anguillarum* from this study and *V. cholerae* (Stroeher *et al.*, 1995; U.H. Stroeher, personal communication). The IS*1358* sequences can be accessed from Genbank using the accession numbers U93587 to U93597. The sequence of the IS*1358* elements from the different *V. anguillarum* serotypes were virtually identical with the only differences being single base-pair substitutions and deletions (Fig. 4.2). The elements from the different strains within the *V. anguillarum* O1 serotype were identical with the exception of the element from the strain 86/3674 which was slightly divergent (99.8%) (Table 4.2).

Analysis of the completed nucleotide sequences of the IS*1358* elements from *V. anguillarum* revealed that some serotypes contained elements with one uninterrupted ORF whilst others encoded elements with three small ORFs. Serotypes O1 and O3 contained single elements that consist of three small ORFs, a similar organisation to that observed in *V. cholerae* O1 (Stroeher *et al.*, 1992). Elements cloned from serotypes O2, O7 and O9, which all contain multiple copies of IS*1358*, were shown to encode single uninterrupted ORFs of 1133 bp, 1109 bp and 1133 bp respectively (Fig. 4.4). The size of these ORFs correspond to a predicted protein of approximately 42 kDa. *V. cholerae* O139 also encodes

Figure 4.2: Alignment of IS*1358* elements sequences from *V. anguillarum*

DNA sequence alignment of the of the sequenced IS*1358* elements from *V. anguillarum*. The letters represent A, ATTC 43305 (O1); B, 86/3674 (O1); C, ATCC 43307 (O3); D, ATCC 43311 (O7); E, ATCC 43313 (O9); F, ATCC 43306 (O2); G, 85-3954-1 (O1); H, 85-3954-2 (O1). * represents identical nucleotide bases.

```
IS1358A    CAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGAATCCGCT    60
IS1358B    CAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGAATCCGCT    60
IS1358C    CAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGAATCCGCT    60
IS1358D    CAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGAATCCGCT    60
IS1358E    CAGGGAAACCGCATGAAAGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGATTCCGCT    60
IS1358F    CAGGGAAACCGCATGAAAGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGATTCCGCT    60
IS1358G    CAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGAATCCGCT    60
IS1358H    CAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAAAGATGCTAAGAATCCGCT    60
           ***************** *********************************** ******

IS1358A    TTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358B    TTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358C    TTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358D    TTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358E    TTAAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358F    TTAAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358G    TTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
IS1358H    TTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTG    120
           ** *********************************************************

IS1358A    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358B    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358C    TAGATCATTTTCGTTGTTTTCTAATGTCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358D    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358E    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358F    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358G    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
IS1358H    TAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATAGGAGCCTAAATATGAGCG    180
           *************************  *********************************

IS1358A    AGTTAACCAACCCATTTATGCATTTCCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
IS1358B    AGTTAACCAACCCATTTATGCATTTCCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
IS1358C    AGTTAACCAACCCATTTATGCATTTTCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
IS1358D    AGCTAACCAACCCATTTATGCATTTCCAAATCATTAAAGACTACCGACAAGAAAGCAAAG    240
IS1358E    AGTTAACCAACCCATTTATGCACTTTCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
IS1358F    AGTTAACCAACCCATTTATGCACTTTCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
IS1358G    AGTTAACCAACCCATTTATGCATTTCCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
IS1358H    AGTTAACCAACCCATTTATGCATTTCCAAATCATTAAAGATTATCGACAGGAAAATAAAA    240
           ** ******************* ** ************** ** ***** ****  ***

IS1358A    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGCGGCGTCCTGTCGGGTCATG    300
IS1358B    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGCGGCGTCCTGTCGGGTCATG    300
IS1358C    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGTGGCGTCTTGTCGGGTCATG    300
IS1358D    TAGATCACAAATTATCAGATATTATTTTGCTGACAATTTGCGGCGTCCTGTCGGGTCATG    300
IS1358E    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGTGGCGTCTTGTCGGGTCATG    300
IS1358F    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGTGGCGTCTTGTCGGGTCATG    300
IS1358G    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGCGGCGTCCTGTCGGGTCATG    300
IS1358H    TAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGCGGCGTCCTGTCGGGTCATG    300
           ***  ************** ******************** ****** ************

IS1358A    -AAGGCTGGGATGGCATTATCGATTTTTGGGCAATGCTCGCTTAGATTTCCTCAAACGAT    359
IS1358B    -ATGGCTGGGATGGCATTATCGATTTT--GGCAATGCTCGCTTAGATTTCCTCAAACGAT    357
IS1358C    GACGGTTGGGATGGCATTATTGACTTT--GGCAATGCTCGCTTAGATTTCCTAAAACGAT    358
IS1358D    -ATGGCTGGGATGGCATTATCGATTTT--GGCAATGCTCGCTTAGATTTCCTCAAACGAT    357
IS1358E    -ACGGT-GGGATGGCATTAT-GACTTT--GGCAATGCTCGCTTAGATTTCCTAAAACGAT    357
IS1358F    -ACGGTTGGGATGGCATTATTGACTTT--GGCAATGCTCGCTTAGATTTCCTAAAACGAT    357
IS1358G    -ATGGCTGGGATGGCATTATCGATTTT--GGCAATGCTCGCTTAGATTTCCTCAAACGAT    357
IS1358H    -ATGGCTGGGATGGCATTATCGATTTT--GGCAATGCTCGCTTAGATTTCCTCAAACGAT    357
            * **  ************* ** ***  *********************** *******

IS1358A    ATGGGTCACTTTGACGCTGGTATTCCTTCTGCAGATACGCTGTCTCGTGTGATGGGCATG    419
IS1358B    ATGG-TCACTTTGACGCTGGTATTCCTTCTGCAGATACGCTGTCTCGTGTGATGGGCATG    416
IS1358C    ATGG-TCATTTTGAGGCAGGCATTCCATCGGCAGATACTTTGTCTCGTGTGATGGGCATG    417
IS1358D    ATGG-TCACTTTGACGCTGATATTCCTTCTGCAGATACGCTGTCTCGTGTGATGGGCATG    416
IS1358E    ATGG-TCATTTTGAGGCAGGCATTCCATCGGCAGATACTTTGTCTCGTGTGATGGGCCTG    416
IS1358F    ATGG-TCATTTTGAGGCAGGCATTCCATCGGCAGATACTTTGTCTCGTGTGATGGGCATG    416
IS1358G    ATGG-TCACTTTGACGCTGGTATTCCTTCTGCAGATACGCTGTCTCGTGTGATGGGCATG    416
IS1358H    ATGG-TCACTTTGACGCTGGTATTCCTTCTGCAGATACGCTGTCTCGTGTGATGGGCATG    416
           **** *** ***** ** *  ***** ** ********  ***************** **
```

```
IS1358A    ATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTG    479
IS1358B    ATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTG    476
IS1358C    ATTAATCCGGCAGGGTTGCAAAAAAGCTTTATATCTTGGATGAAGGACTGCCATACACTG    477
IS1358D    ATTAATCCGGCAGCGTTGCAAAGAAGCTTTATGGCTTGGATGAAGGACTGCCATACACTG    476
IS1358E    ATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTG    476
IS1358F    ATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTG    476
IS1358G    ATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTG    476
IS1358H    ATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTG    476
           ** ********** ******** *********  **************************

IS1358A    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCTCTTATGACCGCTCG    539
IS1358B    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCTCTTATGACCGCTCG    536
IS1358C    ACGGGTGGAGAATTCATCGCTATCGACGGTAAAACATTACGTGGTTCTTATGATCGCTCT    537
IS1358D    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCTCTTATGACCGCTCG    536
IS1358E    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCTCTTATGACCGCTCG    536
IS1358F    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCCCTTATGACCGCTCG    536
IS1358G    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCTCTTATGACCGCTCG    536
IS1358H    ACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCTCTTATGACCGCTCG    536
           **** ******* *******************************  ******* *****

IS1358A    AAGGGTAAAGGAACAATTCACATGGTG-AATGCTCTTGCTACTGCAAACGGAATGAGTAT    598
IS1358B    AAGGGTAAAGGAACAATTCACATGGTG-AATGCTCTTGCTACTGCAAACGGAATGAGTAT    595
IS1358C    AAGGGTAAAGGAACGATTCACATGGTG-AATGCTTTTGCTACTGCAAACAGAATGATTAT    596
IS1358D    AAGGGTAAAGGAACGATTCACATGGTG-AATGCTTTTGCTACTGCAAACGGAATGAGTAT    595
IS1358E    AAGGGTAAAGGAACGATTCACATGGTG-AATGCTTTTGCTACTGCAAACGGAATGAGTAT    595
IS1358F    AAGGGTAAAGGAACGATTCACATGGTGCAATGCTTTTGCTACTGCAAACGGAATGAGTAT    595
IS1358G    AAGGGTAAAGGAACAATTCACATGGTG-AATGCTCTTGCTACTGCAAACGGAATGAGTAT    595
IS1358H    AAGGGTAAAGGAACAATTCACATGGTG-AATGCTCTTGCTACTGCAAACGGAATGAGTAT    595
           ************** ************ ****** ************** ****** ***

IS1358A    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    658
IS1358B    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    655
IS1358C    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCCATCCCCAAGCTACTTGA    656
IS1358D    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    655
IS1358E    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    655
IS1358F    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    655
IS1358G    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    655
IS1358H    CGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCCCCAAGCTACTTGA    655
           ****************************************** *****************

IS1358A    CTTGCTAGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    718
IS1358B    CTTGCTAGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATCCGAAAAGAAAAT    715
IS1358C    CTTGCTATAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    716
IS1358D    CTTGCTAGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    715
IS1358E    CTTGCTCGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    715
IS1358F    CTTGATCGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    715
IS1358G    CTTGCTAGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    715
IS1358H    CTTGCTAGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGGATGCCAAAAGAAAAT    715
           **** *  *************************************** * **********

IS1358A    AGCGCAAGAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGGAATTAGGG    778
IS1358B    AGCGCAAGAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGGAATTAGGG    775
IS1358C    AGCGCAAAAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGTAATCAGGG    776
IS1358D    AGCGCAAAAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGTAATCAGGG    775
IS1358E    AGCGCAAAAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGTAATCAGGG    775
IS1358F    AGCGCAAAAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGGAATTAGGG    775
IS1358G    AGCGCAAGAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGGAATTAGGG    775
IS1358H    AGCGCAAGAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGAAAGGGAATTAGGG    775
           ******* ******************************************* *** ****

IS1358A    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGG    838
IS1358B    AATGCTTGACCAACCTTTTGATGATTATTTTGGAATGAACATGCTTCAAGACTTTGACGG    835
IS1358C    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGG    836
IS1358D    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTCCAAGACTTTGACGG    835
IS1358E    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGG    835
IS1358F    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGG    835
IS1358G    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGG    835
IS1358H    AATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGG    835
           ********* *** * *************** ***** ******* **************
```

```
IS1358A    CAGTTCTTACAGTACTCAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   898
IS1358B    CAGTTCTAACAGAATTCAAGAAAAAATTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   895
IS1358C    CAGTTCTTACAGTACTCAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   896
IS1358D    CAGTTCTTACAGTACTCAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   895
IS1358E    CAGTTCTTACAGTATACAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   895
IS1358F    CAGTTCTTACAGTACACAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   895
IS1358G    CAGTTCTTACAGTACTCAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   895
IS1358H    CAGTTCTTACAGTACTCAAGAAAAAAGTCACGGAAGAACGGAAACGAGAGTGGCTTTAGT   895
           ******* **** *  ********** *********************************

IS1358A    GAGCCGTGATTTGTCGGTTTTAGGTTGATATTGAGCATGAATGGCCCGAGCTAAAATCAA   958
IS1358B    GACCCGTGATTTGTCGGTTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAATCAA   954
IS1358C    GAACCGTGATTTGTCGGTTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAATCAA   955
IS1358D    GAACCGTGATTTGTCGGTTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAACCAT   954
IS1358E    GAACCGTGATTTGTCGGTTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAATCAA   954
IS1358F    GAACCGTGATTTGTCG-TTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAATCAA   953
IS1358G    GAGCCGTGATTTGTCGGTTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAATCAA   954
IS1358H    GAGCCGTGATTTGTCGGTTTTAGGT-GATATTGAGCATGAATGGCCCGAGCTAAAATCAA   954
           ** ************* ******** ****************************** **

IS1358A    TGGGCATCGTC-GCTTCGATTCCGCCAAGAAATCCGGCTGTTGCCAACAGAGCAAGATGT  1017
IS1358B    TGGGCATCGTC-GCTTCGATTCGCC-AAGAA--TCGGCTGTAGC-AACAGAGCAAGATGT  1009
IS1358C    TGG-CATCGTCCGCTTCGATTCGCC-AAGAA--TCGGCTGTAGC-AACAGAGCA-GATGT  1009
IS1358D    GGGCCATCGTC-GCTTCGATACGCC-CAGAA--TCGGCTGTAGC-AACAGAGCAAGATGT  1009
IS1358E    TGGGCATCGTC-GCTTCGATTCGCC-AAGAA--TCGGCTGTAGC-AACAGAGCAAGATGT  1009
IS1358F    TGGGCATCGTC-GCTTCGATTCGCC-AAGAA--TCGGCTGTAGC-AACAGAGCAAGATGT  1008
IS1358G    TGG-CATCGTC-GCTTCGATTCGCCCAAGAA--TCGGCTGTAGC-AACAGAGCAAGATGT  1009
IS1358H    TGGGCATCGTC-GCTTCGATTCGCC-AAGAA--TCGGCTGTAGC-AACAGAGCAAGATGT  1009
            ** ******* ******** *  *  ****   ******* ** ********* *****

IS1358A    GAGTATTCGTTACTATATTTGTTCTAAAGAATAAGAAGCCCAAACTCTGCTTGAAGCGAC  1077
IS1358B    GAGTATTCGTTACTATATTTGTTCTAAAGAATAAGAAGCCCAAACTCTGCTTGAAGCGAC  1069
IS1358C    GAGTATTCGTTATTATATATGTTCTAAAGAATTAGAAGCCCAAACGCTGCTTGAAGCGAC  1069
IS1358D    GAGTATTCGTTACTATATATGTTCTAAAGAATTAGAAGCCCAAACGCTGCTTGAAGCGAC  1069
IS1358E    GAGTATTCGTTACTATATCTGTTCTAAAGAATTAGAAGCCCAAACGCTGCTTGAAGCGAC  1069
IS1358F    GAGTATTCGTTACTATATCTGTTCTAAAGAATTAGAAGCCCAAACGCTGCTTGAAGCGAC  1068
IS1358G    GAGTATTCGTTACTATATTTGTTCTAAAGAATAAGAAGCCCAAACTCTGCTTGAAGCGAC  1069
IS1358H    GAGTATTCGTTACTATATTTGTTCTAAAGAATAAGAAGCCCAAACTCTGCTTGAAGCGAC  1069
           ************ ***** ************* ************ **************

IS1358A    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTAGTGAGGA  1137
IS1358B    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTAGTGAGGA  1129
IS1358C    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCTCTTGATACCGCATTTAGTGAGGA  1129
IS1358D    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTGGTGAGGA  1129
IS1358E    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTAGTGAGGA  1129
IS1358F    GCGTTCTCATTGGGCCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTAGTGAGGA  1128
IS1358G    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTAGTGAGGA  1129
IS1358H    GCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTAGTGAGGA  1129
           ************** ********************* *************** *******

IS1358A    CAACTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATCAGGCGGATGTG  1197
IS1358B    CAACTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATCAGGCAGATGTG  1189
IS1358C    CAATTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATCAGGCAGATGTG  1189
IS1358D    CAATTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATTAGGCAGATGTG  1189
IS1358E    CAACTCTCGTATTAGAGCCGATGATCGTGCAGAGGCTTTAGCAAGGATCAGGCAGATGTG  1189
IS1358F    CAACTCTCGTATTAGAGCCGATGATCGTGCAGAGGCTTTAGCAAGGATCAGGCAGATGTG  1188
IS1358G    CAACTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATCAGGCAGATGTG  1189
IS1358H    CAACTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATCAGGCAGATGTG  1189
           *** *********************** ******************** **** ******

IS1358A    TTTGAACCTATTAAAGAGTGAAAC-ACCTTTAAAGGTGGCATTAAACGCAAACGGATGAA  1256
IS1358B    TTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGCAAACGGATGAA  1249
IS1358C    TTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGTAAACGGATGAA  1249
IS1358D    TTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGTAAACGGATGAA  1249
IS1358E    TTTGAACCTATTAAAGAGTGAAAC-ACCTTTAAAGGTGGCATTAAACGCAAACGGATGAA  1248
IS1358F    TTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGCAAACGGATGAA  1248
IS1358G    TTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGCAAACGGATGAA  1249
IS1358H    TTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGCAAACGGATGAA  1249
           ************************ *********************** ***********
```

```
IS1358A   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTATCGGGCGGTGATG  1316
IS1358B   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTATCGGGCGGTGATG  1309
IS1358C   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTACCATGCGGTGATG  1309
IS1358D   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCTTTACAGGGCGGTGATG  1309
IS1358E   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTACCGGGCGGTGATG  1308
IS1358F   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTACCGGGCGGTGATG  1308
IS1358G   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTATCGGGCGGTGATG  1309
IS1358H   CTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTATCGGGCGGTGATG  1309
          ******************************************* ***    *********

IS1358A   TTCATGCGGTTTCCCTG                                             1333
IS1358B   TTCATGCGGTTTCCGTG                                             1326
IS1358C   TTCATGCGGTTTCCCTG                                             1326
IS1358D   TTCATGCGGTTTCCCTG                                             1326
IS1358E   TTCATGCGGTTTCCGTG                                             1325
IS1358F   TTCATGCGGTTTCCGTG                                             1325
IS1358G   TTCATGCGGTTTCCGTG                                             1326
IS1358H   TTCATGCGGTTTCCGTG                                             1326
          ************** **
```

Table 4.2: Sequence similarity of IS*1358* in different strains from *V. anguillarum* and *V. cholerae*

	O139	O17	H1	CA401	Z17561	A (O1)	B (01)	C (O3)	D (O7)	E (O9)	F (O2)	G (O1)	H (O1)
O139	100 (0)	96 (52)	96 (53)	95.4 (61)	95 (65)	88.2 (155)	88.2 (157)	87.7 (162)	88.8 (147)	87.7 (164)	87.9 (161)	88.3 (155)	88.3 (155)
O17		100 (0)	99.8 (2)	99.2 (10)	98.9 (15)	87.6 (163)	87.6 (164)	86.9 (173)	88 (158)	87 (162)	87.2 (159)	87.7 (163)	87.7 (163)
H1			100 (0)	99.1 (12)	98.9 (15)	87.6 (163)	87.6 (165)	86.9 (173)	88 (156)	87 (172)	87.2 (170)	87.7 (163)	87.7 (163)
CA401				100 (0)	99.5 (7)	87 (171)	87 (173)	86.3 (181)	87.4 (166)	86.4 (182)	86.6 (178)	87.1 (171)	87.1 (171)
Z17561					100 (0)	86.7 (175)	86.7 (177)	86.1 (184)	87.2 (169)	86.2 (183)	86.4 (181)	86.8 (175)	86.8 (175)
A (O1)						100 (0)	99.8 (2)	96.8 (42)	97.6 (32)	97.3 (35)	97.4 (33)	100 (0)	100 (0)
B (O1)							100 (0)	96.6 (44)	97.4 (34)	97.4 (35)	97.4 (35)	99.8 (2)	99.8 (2)
C (O3)								100 (0)	96.4 (48)	98 (26)	98.1 (24)	96.8 (42)	96.8 (42)
D (O7)									100 (0)	96.3 (48)	96.5 (47)	97.5 (32)	97.5 (32)
E (O9)										100 (0)	99.8 (2)	97.4 (35)	97.4 (35)
F (O2)											100 (0)	97.5 (33)	97.5 (33)
G (O1)												100 (0)	100 (0)
H (O1)													100 (0)

% similarity between different IS*1358* sequenced from various *V. cholerae* (U.H. Stroeher, personal communication) and *V. anguillarum* serogroups. The table shows the % similarity and the number of base pair differences (presented in brackets ( ) between the IS*1358* elements for comparison. The letters in the table represent the following strains: A, ATCC 43305 (O1); B, 86/3674 (O1); C, ATCC 43307 (O3); D, ATCC 43311 (O7); E, ATCC 43313 (O9); F, ATCC 43306 (O2); G, 85-3954-1 (O1); H, 85-3954-2 (O1).

an IS*1358* element with a single ORF (1127 bp), and it has been proposed that the single ORF encodes a functional transposase (~42 kDa) and hence may be a mobile genetic element (Stroeher *et al.*, 1995).

4.2.4 Protein expression from the IS*1358* elements of *V. anguillarum*

To confirm the sequence data obtained for the different IS*1358* elements, the T7 RNA polymerase expression system (Tabor and Richardson, 1985) in *E. coli* DH5 was employed using the plasmids pPM4947 to pPM4954, and pPM4958, containing the cloned IS*1358* related sequences from the various *V. anguillarum* serotypes. A protein corresponding to the predicted molecular weight of ~42 kDa could be detected in tracks C, D, E and F which represent IS*1358* from *V. anguillarum* O3, O7, O9, and O2 respectively (Fig. 4.3). The sequenced IS*1358* gene from serotype O3 did not have an ORF, but closer inspection of the sequence identified an amber stop codon (UAG) at base number 663, which was different to the other interrupted IS*1358* elements (Fig. 4.4). The significance of this stop codon is that the *E. coli* strain (DH5) used in the over-expression has a *supE* mutation which results in the stop codon being misread and an amino acid (glutamine) being added, thereby resulting in a product from the IS*1358* element cloned from *V. anguillarum* O3. Proteins produced by IS*1358* elements with the smaller ORFs were not detected in this system indicating that these small proteins are either rapidly degraded or are not translated. A similar situation has been seen in *V. cholerae* O1 (Stroeher *et al.*, 1995). The expression data showed that proteins corresponding to the predicted molecular weight (42 kDa) can be synthesised which suggests that the uninterrupted ORFs of IS*1358* in *V. anguillarum* O2, O7 and O9 may potentially be functional.

Figure 4.3: Detection of IS*1358* encoded proteins by T7 RNA polymerase overexpression.

The proteins were overexpressed by using the T7 promoter/RNA polymerase system (Tabor & Richardson, 1985). Proteins were labelled with L-[^{35}S] methionine, electrophoresed on a SDS-15% polyacrylamide gel. The gel was dried and proteins detected by autoradiography. Each lane contains E2096 with a pGEM-T clone of the IS*1358* element from the indicated *V. anguillarum* strain. Molecular size markers (Pharmacia-LKB) are indicated as follows: 94kDa (phosphorylase b), 67kDa (bovine serum albumin), 43kDa (ovalbumin), 30kDa (carbonic anhydrase).

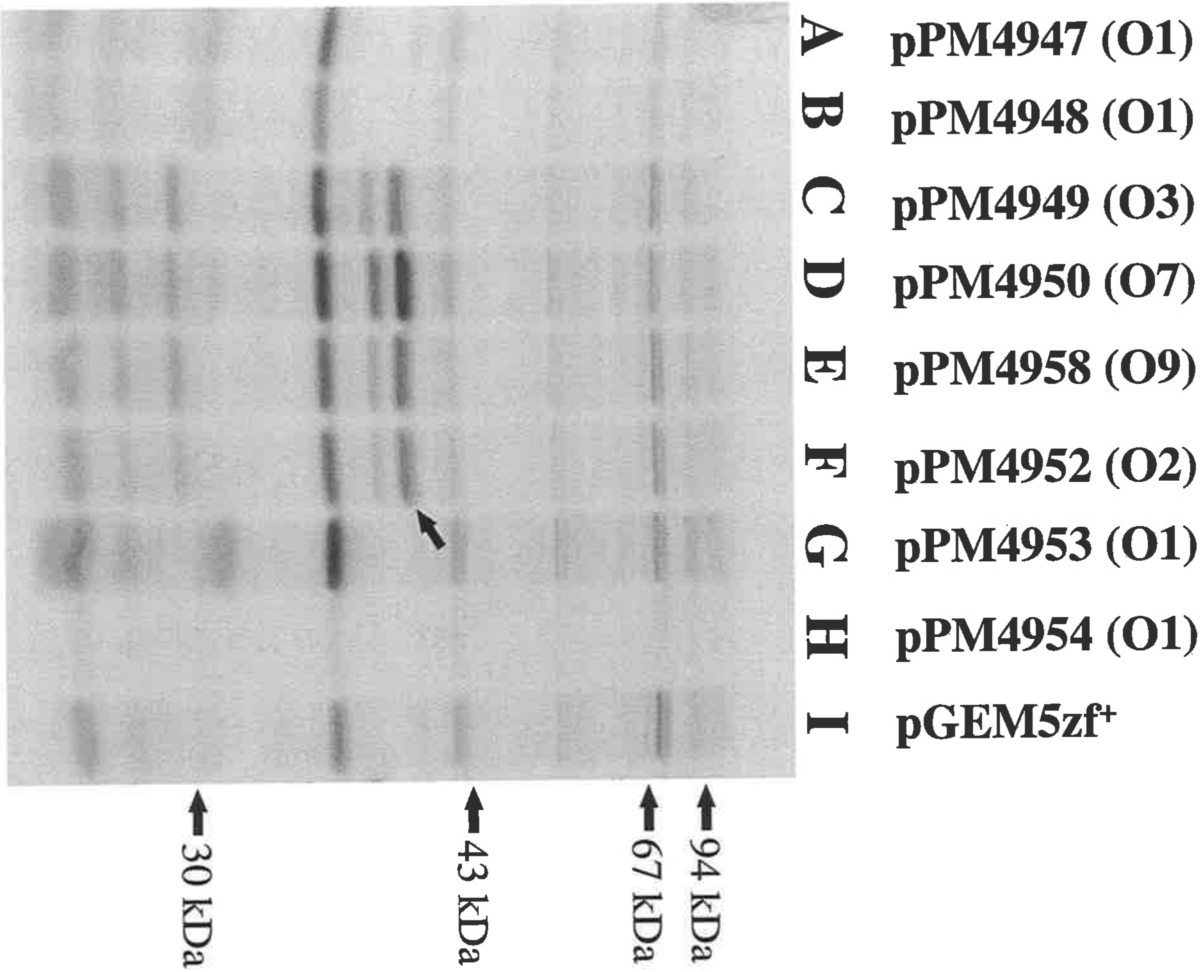
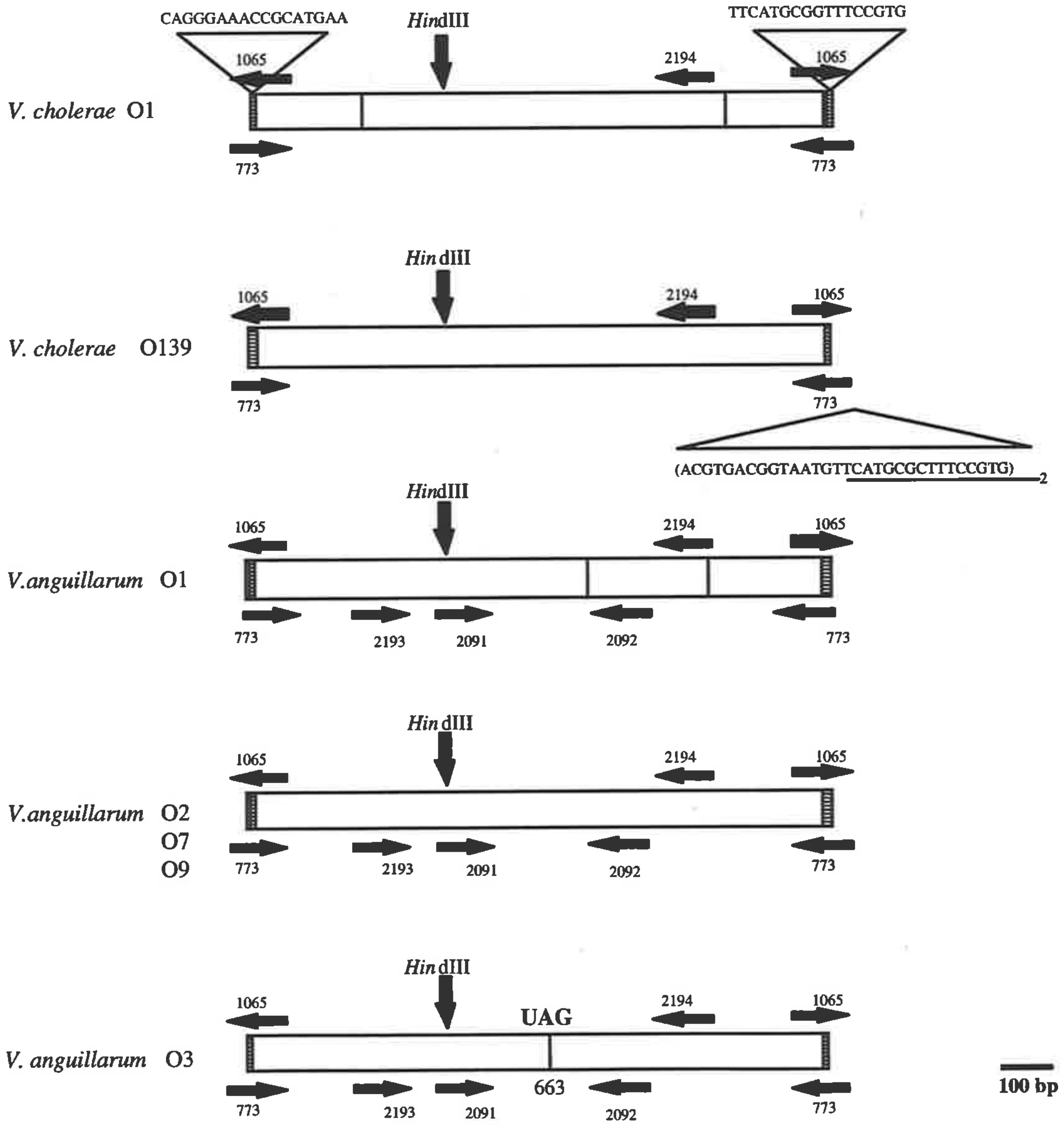

A pPM4947 (O1)
B pPM4948 (O1)
C pPM4949 (O3)
D pPM4950 (O7)
E pPM4958 (O9)
F pPM4952 (O2)
G pPM4953 (O1)
H pPM4954 (O1)
I pGEM5zf⁺
30 kDa
43 kDa
67 kDa
94 kDa

Figure 4.4: Genetic organisation of the IS*1358* elements found in *V. cholerae* O1, O139 and *V. anguillarum* O1, O2 and O3.

The sequence above the triangle represents the inverted repeat flanking IS*1358*. The large triangle below the *V. cholerae* O139 IS*1385* is the sequence of the 31bp repeat which occurs twice, and the underlined portion represents the same sequence as the inverted flanking repeats. The position of oligonucleotides #773, #1065 and #2194 are indicated as is the conserved *Hin*dIII site. The amber stop codon is shown in the *V. anguillarum* O3 IS*1358* element. Vertical lines represent the stop codons of the various truncated ORFs.


CAGGGAAACCGCATGAA
HindIII
TTCATGCGGTTTCCGTG
1065
2194
1065
V. cholerae O1
773
773
HindIII
1065
2194
1065
V. cholerae O139
773
773
(ACGTGACGGTAATGTTCATGCGCTTTCCGTG)2
HindIII
1065
2194
1065
V.anguillarum O1
773
2193
2091
2092
773
HindIII
1065
2194
1065
V.anguillarum O2
O7
O9
773
2193
2091
2092
773
HindIII
1065
UAG
2194
1065
V. anguillarum O3
773
2193
2091
663
2092
773
100 bp

4.2.5 PCR amplification of DNA between the multiple copies of IS*1358* in *V. anguillarum* serotypes O2 and O9

The presence of multiple copies of IS*1358* in serotypes O2 and O9, and the single uninterrupted ORFs noted for the elements sequenced from these strains, suggest that IS*1358* may be mobile. During the course of this study, IS*1358* from *V. cholerae* was shown to possess transposition activity (Dumontier *et al.*, 1998) which further suggests that the elements found in *V. anguillarum* may also be able to transpose.

To determine if there was any relationship in the O2 and O9 serotypes of *V. anguillarum* between IS*1358* and *rfb*/polysaccharide genes, the ORFs located between some of the elements were investigated. PCR was used with an oligonucleotide complementary (#1065) to the inverted repeats of IS*1358*, allowing DNA between the elements to be amplified. PCR reactions were performed using standard conditions but with extension times ranging from 2 to 8 min. Chromosomal DNA from strains ATCC 43306 (O2) and ATCC 43313 (O9) which contained multiple copies of the complete IS*1358* element were used in the PCR reactions. While it was possible to amplify three bands of 2.2 kb, 1.8 kb and 0.72 kb from the O2 serotype, no products were generated using O9 DNA (data not shown).

4.2.6 *rfb*-like genes are located between IS*1358* elements in *V. anguillarum* O2

The amplified DNA products from strain ATCC43306 were cloned into pGEM-T yielding pPM4966, pPM4967, and pPM4968 which were then sequenced using -21 M13 forward and M13 reverse primers. The initial sequence indicated that only the 1.8 kb PCR product (pPM4966) had any significant homology to any genes currently in the database. Oligonucleotides (#2226, #2227, #2278, #2365, #2393) were designed from the sequence data and this enabled the sequence of the 1.8 kb fragment to be completed (Fig. 4.5) The completed sequence was analysed using BlastX and similarities to capsule and O-antigen biosynthesis proteins were detected. Two ORFs were identified: one (ORF1) showed homology to *bplA/wlpA* from the LPS biosynthesis operon of *Bordetella pertussis* (64% identical at DNA and amino acid level; Fig. 4.6) (Allen & Maskell, 1996) and the other (ORF2) was homologous to a Vi polysaccharide biosynthesis protein, VipA from *Salmonella typhi* (66% identical at the DNA level and 62.5% identical at the amino acid level; Fig. 4.7) (Hashimoto *et al.*, 1993: Waxin *et al.*, 1993). Table 4.3 shows the proteins to which ORF2 is most homologous. The homologues of ORF 1 and 2 are both thought to be dehydrogenase enzymes involved in polysaccharide biosynthesis. ORF 1 which shows significant homology to *bplA* is a complete ORF (954 bp) predicted to encode a protein of 35 kDa, whereas ORF 2 (710 bp) is interrupted by the IS*1358* element (Fig. 4.8).

Southern hybridisation was performed to determine which IS*1358* elements correlated with the 1.8 kb fragment. *Hin*dIII digested DNA of ATCC43306 (O2) and 85-3954-2 (O1) was used with either PCR DIG-labelled IS*1358* (generated from chromosomal DNA of 85-3954-2 and oligonucleotide # 773 (Table 2.4)), or pPM4966 (the 1.8 kb fragment cloned into pGEM-T) was DIG-labelled using the primers -21M13 (Forward) and M13 (Reverse)) as probes. Figure 4.9 clearly shows that the 1.8 kb fragment is located on two *Hin*dIII fragments of 4.8 and 2.6 kb. Two fragments are detected as there is a *Hin*dIII restriction site at nt 1415 (Fig. 4.5) which lies within ORF 2. This data also indicates that the IS*1358* element adjacent ORF 2 does not contain a *Hin*dIII site which was found to be conserved between all IS*1358* elements sequenced (Section 4.2.3, U.H. Stroeher, personal communication). The Southern blot hybridised with the IS*1358* probe showed the multiple

Figure 4.5: The nucleotide and deduced amino acid sequence of the DNA between IS*1358* elements in *V. anguillarum* O2 (ATTC 43306)

The nucleotide sequence is numbered in accordance with the Genbank accession number AF156256 and is shown from nt 1 tont 1830. The amino acid translation of each ORF is represented by single letter code below the first nt of each codon. Possible ribosome binding sites are bolded. The oligonucleotides used to sequence the region are underlined and indicated by corresponding number as in Table 2.4.

```
          10        20        30        40        50        60
5'TTCATGCGGTTTCCGTGGACTATCACTGAGATAAAAAGTAGCAAATTAGAGTGATTAATA
                                        ←IS1358      orf1   M

        70        80        90       100       110       120
TGATAAAAAATAGAAAAATTCGAATAGCAGTCGTTGGTTGTGGGCGTATTTCTAAAAATC
  I  K  N  R  K  I  R  I  A  V  V  G  C  G  R  I  S  K  N  H

       130       140       150       160       170       180
ATTTTGGCTCAATTGAGCAGCTTGATTCAGAATATGAGCTGGTCGCGGTGTGCGATAACA
  F  G  S  I  E  Q  L  D  S  E  Y  E  L  V  A  V  C  D  N  N

       190       200       210       220       230       240
ACCCAGAAGTGTTAGAGCTGCATTCAAATAAATACGGCGTTCCGGGTTATCACTCGATAG
  P  E  V  L  E  L  H  S  N  K  Y  G  V  P  G  Y  H  S  I  D

       250       260       270       280       290       300
                                                    # 2226
ATGATCTGCTCGCTTCACAAACACTTGATATCGTCACCTTGTGTACCCCGAGTGGCATAC
  D  L  L  A  S  Q  T  L  D  I  V  T  L  C  T  P  S  G  I  H

       310       320       330       340       350       360
ATGCACCACAAACTATTAAAGCTGCGAAAGCCGGTGTACATGTGATTACCGAAAAACCGA
  A  P  Q  T  I  K  A  A  K  A  G  V  H  V  I  T  E  K  P  M

       370       380       390       400       410       420
TGGCGACCAAGTGGGAAGATGGACTAGCAATGGTTAAAGCGTGTGATGACGCTGGGGTGC
  A  T  K  W  E  D  G  L  A  M  V  K  A  C  D  D  A  G  V  R

       430       440       450       460       470       480
                                                    # 2393
GTTTATTTGTGGTTAAACAAAACCGCCGTAATTCAACATTGCAATTACTTAAGCGTGCAG
  L  F  V  V  K  Q  N  R  R  N  S  T  L  Q  L  L  K  R  A  V

       490       500       510       520       530       540
TGACGGAAAAGCGTTTTGGTAAAATCCATATGGTGCATTTAAATGTGTTTTGGACCAGAC
  T  E  K  R  F  G  K  I  H  M  V  H  L  N  V  F  W  T  R  P

       550       560       570       580       590       600
CTCAAGAGTATTACGACCGCGCTGCTTGGTCTGGTACTTGGCACATGGATGGTGGTGCTT
  Q  E  Y  Y  D  R  A  A  W  S  G  T  W  H  M  D  G  G  A  F

       610       620       630       640       650       660
TTATGAATCAGGCGACTCATTATGTCGATCTTCTCCATTGGCTAGTGGGGCCTATTGAAA
  M  N  Q  A  T  H  Y  V  D  L  L  H  W  L  V  G  P  I  E  T

       670       680       690       700       710       720
CCATCCATGCTATGACCTCCACTCATCGTGATATTGAAGTTGAAGATACTGGTGTGGTGA
  I  H  A  M  T  S  T  H  R  D  I  E  V  E  D  T  G  V  V  N

       730       740       750       760       770       780
ACATCAAATGGCGTAATGGCGCGTTAGGCTCAATGGCGGTCACTATGTGTACTTATCCCA
  I  K  W  R  N  G  A  L  G  S  M  A  V  T  M  C  T  Y  P  N

       790       800       810       820       830       840
ATAATTTAGAGGGTTCAATCACTATTTTTGGGGAAAAAGGGACAGTGCGTGTTGGTGGCG
  N  L  E  G  S  I  T  I  F  G  E  K  G  T  V  R  V  G  G  V

       850       860       870       880       890       900
TGGCGGTTAATGAAATCCAAGAATGGAATTTTGCAGAATCAAAAGATTACGATCAGCAAA
  A  V  N  E  I  Q  E  W  N  F  A  E  S  K  D  Y  D  Q  Q  I

       910       920       930       940       950       960
TTGAACAAGCCAACTACCAAACCACTTCAGTCTATGGATTTGGTCATCCACCTTACTTTA
  E  Q  A  N  Y  Q  T  T  S  V  Y  G  F  G  H  P  P  Y  F  K
```

```
       970       980       990      1000      1010      1020
       # 2365
AAAACGTTGCCGATGTATTCCGTGGCGAGGCAGAGCCAGAAACCGACGGCCTGTAGGGTT
   N  V  A  D  V  F  R  G  E  A  E  P  E  T  D  G  L  *

      1030      1040      1050      1060      1070      1080
TAAAATCACTTGAGCTATTGATCAGCATTTATCGCTCAGCTCGCGATCATAAAGAAATTG

      1090      1100      1110      1120      1130      1140
GTTTGCCATTGAATCTGTAATTACGAGTAAAGAAAAATGAAGTTAGAAAATGTAAAGATT
                               orf2  M  K  L  E  N  V  K  I

1150    1160      1170      1180      1190      1200
GGCATTATTGGTTTAGGCTATGTTGGCTTACCGCTTGCGGTTGAGTTTGGTAAGAAGTTT
 G  I  I  G  L  G  Y  V  G  L  P  L  A  V  E  F  G  K  K  F

      1210      1220      1230      1240      1250      1260
                                                    # 2278
AGCACGGTAGGTTTTGATATTAACCAGAAAAGGGTCGCTGAATTACACTCTGGCATTGAT
 S  T  V  G  F  D  I  N  Q  K  R  V  A  E  L  H  S  G  I  D

      1270      1280      1290      1300      1310      1320
TTAACCTTGGAATGTTCCGAAGAGGAACTGCGTAGTGCTCCATTGCTGAGTTACACTCAC
 L  T  L  E  C  S  E  E  E  L  R  S  A  P  L  L  S  Y  T  H

      1330      1340      1350      1360      1370      1380
CAATTGGAAGAGATTAAAGACTGTAATTTTTATATCGTTACCGTGCCTACACCAATCACA
 Q  L  E  E  I  K  D  C  N  F  Y  I  V  T  V  P  T  P  I  T

      1390      1400      1410      1420      1430      1440
GATGAAAAAACACCGGATCTTAACCCACTAAAAAAAGCTTCTGAAGCTCTTGCGAAAGTG
 D  E  K  T  P  D  L  N  P  L  K  K  A  S  E  A  L  A  K  V

      1450      1460      1470      1480      1490      1500
GTCTCCCACGGTGATATTGTCGTTTTCGAATCAACAGTTTATCCCGGTGCAACGGAAGAA
 V  S  H  G  D  I  V  V  F  E  S  T  V  Y  P  G  A  T  E  E

      1510      1520      1530      1540      1550      1560
GTGTGCCTTCCTATCATCGAGAAAAATTCAGGTTTAGTGTTTAATCGCGATTTCTTTGCT
 V  C  L  P  I  I  E  K  N  S  G  L  V  F  N  R  D  F  F  A

      1570      1580      1590      1600      1610      1620
  # 2227
GGTTATAGCCCAGAGCGTATCAATCCAGGAGATAAAGTGAATCGTTTAACTACGATCATG
 G  Y  S  P  E  R  I  N  P  G  D  K  V  N  R  L  T  T  I  M

      1630      1640      1650      1660      1670      1680
AAAATAACTTCTGGCTCTACGCCTGAAGTGGCTGATTTTGTTGATCAAGTATACTCTTCG
 K  I  T  S  G  S  T  P  E  V  A  D  F  V  D  Q  V  Y  S  S

      1690      1700      1710      1720      1730      1740
ATAGTCACTGCTGGTACCCATAAAGCGCCATCAATCAAGGTGGCAGAAGCCGCTAAAGTT
 I  V  T  A  G  T  H  K  A  P  S  I  K  V  A  E  A  A  K  V

      1750      1760      1770      1780      1790      1800
ATTGAAAATACCCAGCGTGATCTAAATATTGCTGTAATCAATGAGTTTGCCAAAATTTTT
 I  E  N  T  Q  R  D  L  N  I  A  V  I  N  E  F  A  K  I  F

      1810      1820      1830
AATCGACTAGGTATCAGGGAAACCGCATGAA 3'
 N  R  L  G  I  R  E  T  A  IS1358→
```

Figure 4.6: Alignment of ORF 1 with homologous proteins

Alignment of the amino acid sequence of ORF 1 from *V. anguillarum* O2 with WlpA/BplA from *Bordetella pertussis* (gbX90711), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: gb, Genbank.

```
Orf1        -VINMIKNRKIRIAVVGCGRISKNHFGSIEQLDSEYELVAVCDNNPEVLELHSNKYGVPG   59
WlpA        MSSLPITDRKIRFGLVGCGRISKNHIGAIAQHGDRAELVEICDTNPEALQAAEAATGARP   60
                 *  ****  .********** *.* *     *** .**.*** *.      *

Orf1        YHSIDDLLASQTLDIVTLCTPSGIHAPQTIKAAKAGVHVITEKPMATKWEDGLAMVKACD  119
WlpA        FSSLSDMLAQGNADALVLATPSGLHPWQAIEVAQAGRHVVSEKPMATRWEDGKRMVKACD  120
            . *. *.**  . * . *.****.*  *.*  *.** **..******.****  ******

Orf1        DAGVRLFVVKQNRRNSTLQLLKRAVTEKRFGKIHMVHLNVFWTRPQEYYDRAAWSGTWHM  179
WlpA        EAGVRLFVVKQNRRNATLQLVKKAIEQGRFGRIYMVTVNVFWTRPQEYYDAARWRGKWEW  180
            .**************.****.*.*.*. . ***.* ** .************ * * * *

Orf1        DGGAFMNQATHYVDLLHWLVGPIETIHAMTST-HRDIEVEDTGVVNIKWRNGALGSMAVT  238
WlpA        DGGAFMNQASHYVDLLDWLVGPVESVYAYTATLARRIEAEDTGVAALRWRHGAMGSINVT  240
            *********.****** *****.*.. * *.*  * ** *****  ..**.**.**. **

Orf1        MCTYPNNLEGSITIFGEKGTVRVGGVAVNEIQEWNFAESKDYDQQIEQANYQTTSVYGFG  298
WlpA        MLTYPQNLEGSITILGEKGTVRVGGVAVNRIDEWKFAEPHPDDDKIREANYETTSVYGFG  300
            * ***.******** ************** * ** *** .  * .* .***.********

Orf1        HPPYFKNVADVFRGEAEPETDGL---------------------------           321
WlpA        HPLYYDNVINCLRGDCEPETDGREGLQSLALLTAIYRSARDGVRIPLPLD           350
            ** *. **    **..******
```

Figure 4.7: Alignment of ORF2 with homologous proteins

Alignment of the amino acid sequence of ORF2 (VaOrf2), *Salmonella typhi* VipA (StVipA) (spQ04972), *Shigella sonnei* ORF1 (SsOrf1) (gbU34305), *E. coli* ORF7 (EcOrf7) (Z21706) and *Staphylococcus aureus* CapL (SaCapL) (spP39861), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: sp, SWISS PROT; gb, Genbank.

```
VaOrf2     -MKLE--NVKIGIIGLGYVGLPLAVEFGKKFSTVGFDINQKRVAELHSGIDLTLECSEEE   57
StVipA     MFGID--EVKIAIIGLGYVGLPLAVEFGKSRQVVGFDVNKKRILELKNGVDVNLETTEEE   58
SsOrf1     -MKFDTLNAKIGIIGLGYVGLPLAVEFGKKVTTIGFDINKSRIDELRNGHDSTLECSNLE   59
EcOrf7     MISKD--DVKIAIIGLGYVGLPLAAEFGKIRQVVGFDVNHKRILELKEGIDVNLETTEEE   58
SaCapL     ------MNRNIAVVGLGYVGLPVAVTFGNKHKVIGFDINESRIKELKNNYDRTNEVTEN-   53
                     * ..********.*  **     .***.*  *. **. . * . * .

VaOrf2     LRSAPLLSYTHQLEEIKDCNFYIVTVPTPITDEKTPDLNPLKKASEALAKVVSHGDIVVF  117
StVipA     LREARYLKFTSEIEKIKECNFYIITVPTPINTYKQPDLTPLIKASETVGTVLNRGDIVVY  118
SsOrf1     LLEATKLTYACSLDALKECNVFIVTVPAPIDKHKQPDLTPLIKASETLGKIIKKGDVIIY  119
EcOrf7     LRDARHLTFTSSIDEIRECNFYIITVPTPINDYKQPDLTPLIMASETVGKVLRPGDIVVY  118
SaCapL     KLKNTNIEYTSNAEDLKKADFIIIAVPTPIDKHNKPDLLPLLKASETVGKVITPDTIVVY  113
                 . ..   . .. .   *..**.**     *** **  ***..  ..    ....

VaOrf2     ESTVYPGATEEVCLPIIEKNSGLVFNRDFFAGYSPERINPGDKVNRLTTIMKITSGSTPE  177
StVipA     ESTVYPGCTEEECVPILARMSGMTFNQDFYVGYSPERINPGDKKHRLTNIKKITSGSTAQ  178
SsOrf1     ESTVYPGATEEDCIPVVEKVSGLKFNIDFFAGYSPERINPGDKEHRVTNILKVASGSTPD  179
EcOrf7     ESTVYPGCTEEECVPILERMSAMTFNKDFFVGYSPERINPGDKKYRLTTIKKITSGSTAE  178
SaCapL     ESTVYPGATEEECVPVLEKYSGLVCGKDFFVGYSPERINPGDKVHTFETITKVVSGQTLE  173
           *******.*** *.*.. . * .  . **.************     .* *. ** *

VaOrf2     VADFVDQVYSSIVTAGTHKAPSIKVAEAAKVIENTQRDLNIAVINEFAKIFNRLGIRETA  237
StVipA     IAELIDEVYQQIISAGTYKAESIKVAEAAKVIENTQRDLNIALVNELAIIFNRLNIDTEA  238
SsOrf1     VAEYVDQLYKLIITVGTHKASSIKVERLQSN-VNTQRDVNIALINELSIIFNKLGIDTLE  238
EcOrf7     TACLVDEIYRQIITAGTYKTESIKIAERKGD-----------------------------  209
SaCapL     VLEIVADVYSSVVTAGVHKASSIKVAEAAKVIENTQRDVNIALMNELAIIFDKLDIDTNE  233
               .  .*  ... *  *. ***.           ..

VaOrf2     ------------------------------------------------------------  237
StVipA     VLRAAGSKWNFLPFRPGLVGGHCIGVDPYYLTHKSQGIGYYPEIILAGRRLNDNMGNYVS  298
SsOrf1     VLEAAGTKWNLLPFRPGLVGGHCIGVDPYYLTHKAQSVGYHPEMILAGRRLNDSMGQYVV  298
EcOrf7     ------------------------------------------------------------  209
SaCapL     VLKASGTKWNFLNFKPGLVGGHCIGVDPYYLTHKAQEVGHHPEVILAGRRINDNMAKYIA  293


VaOrf2     ------------------------------------------------------------  237
StVipA     EQLIKAMIKKGINVEGSSVLILGFTFKENCPDIRNTRIIDVVKELGKYSCKVDIFDPWVD  358
SsOrf1     SQLVKKMLKQRIQVEGANVLVMGLTFKENCPDLRNTKVIDIISELKEYNINIDIIDPWCS  358
EcOrf7     ------------------------------------------------------------  209
SaCapL     SNVIKELLKQGLEVQGATVNVLGLTFKENCPDLRNTKVIHIIEELKEYGLNVTVNDVEAD  353


VaOrf2     ------------------------------------------------------------  237
StVipA     AEEVRREYGIIPVSEVKSSHYDAIIVAVGHQQFKQMGSEDIRGFGKDKHVLYDLKYVLPA  418
SsOrf1     TDEAQHEYGLTLCEDPKVNHYDAIIIAVAHNEFREMGESAIRALGKDEARFVRFKICA--  416
EcOrf7     ------------------------------------------------------------  209
SaCapL     KNEAKKFFGLDLIDTKELKMVDVVLFAVPHKDYMEN-KKDYINLVKDCGIVFDIKGIINS  412


VaOrf2     ------------                                                  237
StVipA     EQSDVRL-----                                                  425
SsOrf1     ------------                                                  416
EcOrf7     ------------                                                  209
SaCapL     DELNVSQRLWRL                                                  424
```

Table 4.3: Homology table for ORF 2

	%identity[a]				
	VaOrf2	**StVipA**	**SsOrf1**	**EcOrf7**	**SaCapL**
VaOrf2	100	62.5 (235)	64.2 (229)	65.2 (201)	59.0 (227)
StVipA		100	59.7 (412)	82.4 (205)	52.6 (416)
SsOrf1			100	58.6 (198)	58.1 (389)
EcOrf7				100	54.5 (198)
SaCapL					100

[a] Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaOrf2:	*V. anguillarum* ORF2
StVipA:	*Salmonella typhi* VipA (spQ04972),
SsOrf1:	*Shigella sonnei* ORF1 (gbU34305),
EcOrf7:	*E. coli* ORF7 (gbZ21706)
SaCapL:	*Staphylococcus aureus* CapL (spP39861)

Figure 4.8: The genetic organisation of the 1.8 kb fragment located between IS*1358* elements in *V. anguillarum* O2.

A schematic representation of pPM4966 which contains the 1.8 kb PCR product generated using the oligonucleotide #1065. Restriction sites are indicated on the figure for reference. The arrows indicate the location and direction of the oligonucleotides used.

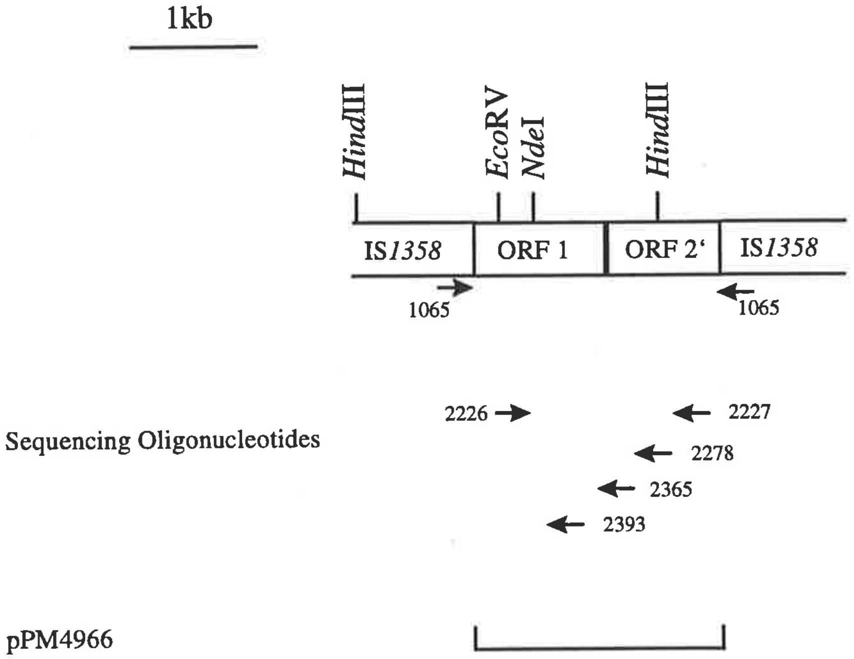
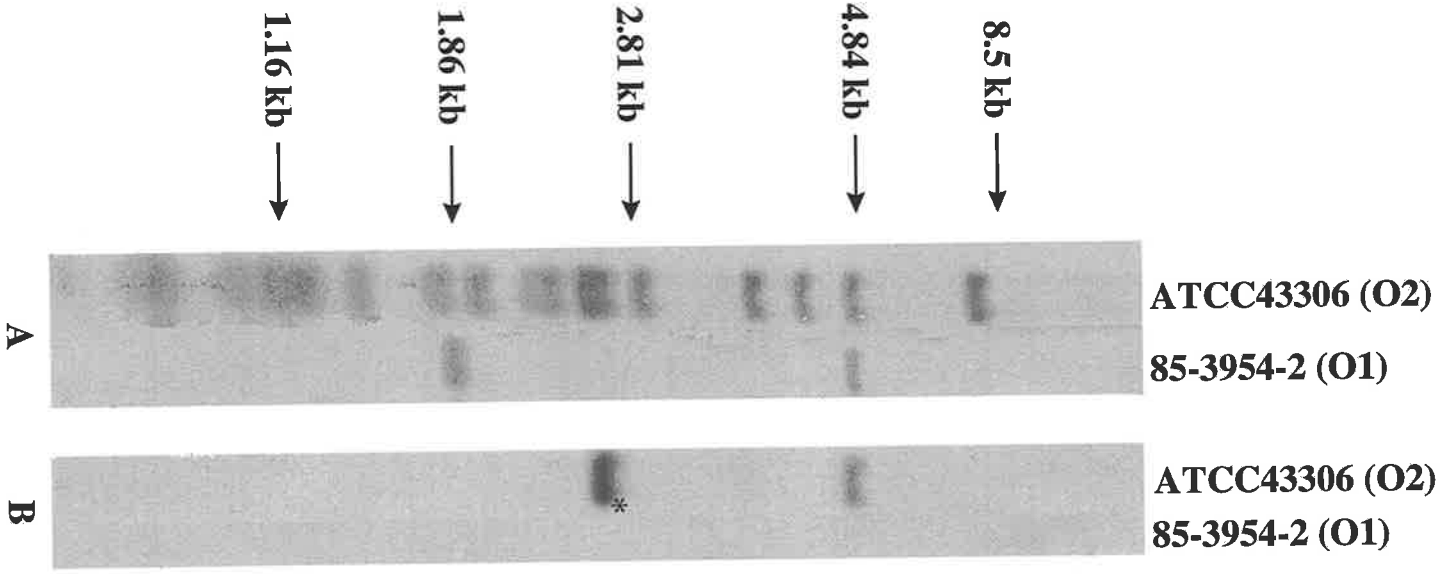

1kb
HindIII
EcoRV
NdeI
HindIII
IS1358
ORF 1
ORF 2'
IS1358
1065
1065
2226
2227
Sequencing Oligonucleotides
2278
2365
2393
pPM4966

Figure 4.9: Southern hybridisation analysis of the 1.8 kb PCR product

Southern blots of *Hin*dIII restricted DNA of ATCC 43306 (O2) and 85-3954-2 (O1) were probed with PCR DIG-labelled IS*1358* (labelled using oligonucleotide #773, Table 2.4)(A) and pPM4966 (labelled using -21M13 and M13 primers) (B) probes. The appropriate size markers are indicated on the left side of the figure. The smaller molecular weight band in panel B (labelled *) appears to be darker than the higher molecular weight band in the same lane. This is because the *Hin*dIII restriction site is located near the end of the 1.8 kb fragment, and therefore a greater proportion of the probe hybridises to the smaller band (labelled *).


1.16 kb
1.86 kb
2.81 kb
4.84 kb
8.5 kb
A
ATCC43306 (O2)
85-3954-2 (O1)
B
*
ATCC43306 (O2)
85-3954-2 (O1)

bands characteristic of the O2 serotype. However there are two IS*1358* bands (Fig. 4.9 (A)) that correspond to the size of the fragments (4.8 and 2.6 kb) observed with the pPM4966 probe (Fig. 4.9 (B)) suggesting that these elements flank the region. In addition, the 1.8 kb fragment does not hybridise to any fragments in the O1 serotype strain 85-3954-2 (see also Section 4.2.12), and hence ORF 1 and ORF 2 are not present in this strain.

4.2.7 Inverse PCR

Inverse PCR was used to further characterise the DNA flanking these closely linked IS*1358* elements. PCR using oligonucleotide #1065 (complementary to the inverted repeats of IS*1358*) (Fig. 4.8) was performed on ligated chromosomal DNA that had previously been digested with *Nde*I. *Nde*I was used as it cuts inside the 1.8 kb fragment but not within IS*1358* (Fig. 4.8). Southern hybridisation using ORF 1 as a probe indicated a fragment of approximately 3.0 kb should be generated by PCR if *Nde*I was used to digest the chromosome (data not shown). PCR amplified a single product corresponding to the expected size (3.0 kb). The product was cloned into pGEM-T and the resulting plasmid, pPM5510, was partially sequenced using forward, reverse and specific primers (Fig. 4.10) (Table 2.4). Subsequent BlastX searches revealed the presence of ORF1 (*bplA*/*wlpA*) at one end of the cloned sequence, and at the other end another ORF with homology to *wzz* (*cld* or *rol)* genes from *E. coli* (Bastin *et al.*, 1993; Morona *et al.*, 1995) and other pathogenic bacteria (Fig. 4.11). Wzz has been shown to regulate the chain length of O-antigen in these bacteria (Chapter 1, Section 1.5.4.1.4).

Another partial ORF (ORF 4) was detected adjacent to the *wzz* homologue (ORF 3); this gene was transcribed in the opposite direction to ORF 3 and has homology to *rfaD/gmhD* from *V. cholerae* O1 (60% identical at amino acid level) (Fig. 4.11). The RfaD/GmhD protein of *V. cholerae* is involved in LPS core sugar biosynthesis.

Figure 4.10: The nucleotide and deduced amino acid sequence of a Wzz homologue from *V. anguillarum* O2 (ATCC 43306).

The nucleotide sequence is numbered in accordance with the Genbank accession number AF156257 and is shown from nt 1 to nt 1415. The amino acid translation of Wzz is represented by single letter code below the first nt of each codon. The potential initiation codons are bolded. The beginning of *rfaD*/*gmhD* is indicated at nt 283.

10 20 30 40 50 60
5'TAGCGGCTTAGATAGTGGCTGTAAACACAAAAGCTCGCTTTTCTCATTGAACGTCTAGCC

70 80 90 100 110 120
TTGAAAAGGGTTTTGCTTTAAACTTCGGCACGCAAGGTGGCAAGGTGGCAAGGTGGCAAG

130 140 150 160 170 180
GTGGCAAGGTGGCAAGGTGGCTCGAAATTGACAAATTCGATAAAGCTTGTACAACAATTT

190 200 210 220 230 240
AAACAGTCAGCCTTTAGGAGCGGATACCTAAGGGCGGTAGCGTGCCTAAAAACCAATGAT

← *rfaD/gmhD*
250 260 270 280 290 300
AGGATCAACAGTGTGAATCAATCCCCTTTTTCGCACTCTTCATACCCACCA**GTG**CCTTCT
wzz → V P S

310 320 330 340 350 360
TACTATCAGCAAAGCAATAATGATGAAATTGACCTGCGTGAGTTGTTTAGT**GTG**CTATGG
Y Y Q Q S N N D E I D L R E L F S V L W

370 380 390 400 410 420
CAAGGGAAATTGACCATA**GTG**CTGTGCACTGTTTTGTTTGCT**GTGATG**GCGGTGGTATAT
Q G K L T I V L C T V L F A V M A V V Y

430 440 450 460 470 480
GCATTGACGGCGCAGCAGTGGTGGAGCGCTAAAGCAACCGTTACCCAGCCTGAATTAAGC
A L T A Q Q W W S A K A T V T Q P E L S

490 500 510 520 530 540
CAAGTTGTAACCTTCCTGCAACAGGTGAAGCAGTATCAGCCATTGTTTGATCTCTACCAA
Q V V T F L Q Q V K Q Y Q P L F D L Y Q

550 560 570 580 590 600
GAAGATGGCACCATTATTGTGAGTAAAGAGCTTGATGGCTTAATTAACCCAGCCACTATT
E D G T I I V S K E L D G L I N P A T I

610 620 630 640 650 660
TTCCAGCGTTTTGTGCGTGCATTTAATACTAATGACAATAAAAAGCGCTTTATGCAAAGT
F Q R F V R A F N T N D N K K R F M Q S

670 680 690 700 710 720
AACCCTACTTTTTTAGCCATTGAGCAGCAACAGCTTAAAGTGAATGAAAATAACGAAACT
N P T F L A I E Q Q Q L K V N E N N E T

730 740 750 760 770 780
GATGATGTGGCGCGCTTTTATCAAAGTTGGTATGGCAAAATTACGGCTCAAGAAGTGGGT
D D V A R F Y Q S W Y G K I T A Q E V G

790 800 810 820 830 840
AAGACACAAGCGGGTGATTTTTATCTCAGCTTTCAGTCAATTGATAAAGCCAGCAGTTTA
K T Q A G D F Y L S F Q S I D K A S S L

850 860 870 880 890 900
GTGTTGCTTAACGCCTATATCGACTTTATTAACCAACAACTGAACCAGCAGCTAACCAAT
V L L N A Y I D F I N Q Q L N Q Q L T N

910 920 930 940 950 960
GACCTTACCAGTAAATTGACGGTAAAACATAACCAACTTAGCCAGCAATACAGTAGTTTG
D L T S K L T V K H N Q L S Q Q Y S S L

970 980 990 1000 1010 1020
CAGCAGCAAACCCAATTGCGCCTGCAGGTGGAATTGGCGCGAGTGCAGAATGCATTGGCA
Q Q Q T Q L R L Q V E L A R V Q N A L A

1030 1040 1050 1060 1070 1080
ATTGCAAAAGCGGCCAATATTAATGAGCCTGTACAAAATCTGAATGAAGAGAAGCTGTTT
I A K A A N I N E P V Q N L N E E K L F

1090 1100 1110 1120 1130 1140
GCCATTAGTATTGGTAGTAAAGCATTGCAAGCCAAAGTCGATGCATTAAAAAGTATAACG
A I S I G S K A L Q A K V D A L K S I T

1150 1160 1170 1180 1190 1200
AATTTAAGTGTGTTTGAACCACGCTTAGCGCTGCTGCAAGCCCAAGTTCAGCAGGTTGAG
N L S V F E P R L A L L Q A Q V Q Q V E

1210 1220 1230 1240 1250 1260
CTATTAGGTAAAGTAAAACCAGCGCAAGTGCAAGGTTACGCTTACTTAGAGCAACCAGAA
L L G K V K **P** A Q V Q G Y A Y L E Q **P** E

1270 1280 1290 1300 1310 1320
GCGCCGATTAGCCGTGATGAGCCCAAACGAGCATTGATAGCGGTGCTTGGAACCTTACTG
A **P** I S R D E **P** K R A L I A V L **G** T L L

1330 1340 1350 1360 1370 1380
GGCGGCATGCTTGGCGTTGCGATTGTGCTGGTGAGGTTTGCGTTTAGAAAAGAAGAAGAA
G **G** M L **G** V A I V L V R F A F R K E E E

1390 1400 1410
AAAGCTTAGACCAGGCAGCAGGGAAACCGCATGAA 3′
K A * IS*1358* →

Additionally, a putative JUMPstart sequence (5' TAGGAGCGGATACCTAAGGGCG GTAGCGTGC 3') (Fig. 4.11), often found at the beginning of *rfb* operons (Hobbs and Reeves, 1994; Marolda and Valvano, 1998), was identified between the divergently transcribed ORFs 3 and 4.

4.2.8 Further characterisation

In an attempt to clone the DNA further downstream of ORF 2-IS*1358* (Fig. 4.8; Fig. 4.11), a *Sau*3A cosmid bank library of ATCC43306 was constructed using the cosmid vector pPM2101 (Table 2.3) as described in Materials and Methods (Section 2.10.6.1). Cosmids clones were initially screened using a DIG-labelled IS*1358* probe in colony dot blot hybridisations. Out of 500 clones, 12 hybridising to the IS*1358* probe were found. These clones were screened with a DIG-labelled 800 bp *Dra*I fragment of ORF2 probe (Fig 4.8). Two clones, pPM5513 and pPM5514 were found to hybridise to this fragment, however pPM5514 was later found to contain only 200 bp of ORF2 (data not shown) and was no longer used to characterise this region. To further map the region around ORF2-IS*1358*, pPM5513 and chromosomal DNA (ATTC43306) were digested with *Hin*dIII, *Nde*I and *Eco*RV, and used in Southern hybridisation with the ORF2 probe described above. The bands that hybridised were the same size in both pPM5513 and the chromosome (Fig. 4.12). The size of the bands were: *Hin*dIII, 4.8 and 2.6 kb; *Nde*I, 7.3 kb and *Eco*RV, ~8.5 kb (Fig. 4.12). The *Nde*I digest appears slightly different between pPM5513 and chromosome as the cosmid DNA had not cut to completion (Fig. 4.12). The restriction sites were mapped onto the physical map of the region (Fig. 4.11).

Figure 4.11: The genetic organsiation of the putative polysaccharide synthesis region of *V. anguillarum* O2 (ATCC 43306).

A map representing the DNA located around two IS*1358* elements of *V. anguillarum* O2. The map was derived using Southern hybridisation, inverse PCR and sequencing data obtained from various subclones. The genes within this region are indicated on the map. ORF1 corresponds to the gene homologue, *bplA,* and ORF2' corresponds to the gene homologue, *vipA*. ORF2' is not a complete open reading frame and is truncated due to the adjacent IS*1358* element. The *wzz* (ORF 3) and *rfaD/gmhD* (ORF 4) homologues are also represented. The arrows indicate the direction of transcription. The putative JUMPstart sequence is represented by the sequence above the large triangle above the figure. The bolded nucleotides indicate an identical match to the JUMPstart consensus sequence. The underlined nucleotides show homology to some JUMPstart sequences and that these bases show variation. The location of the 800 bp *Dra*I fragment described in the text as the ORF 2 probe is indicated for reference.

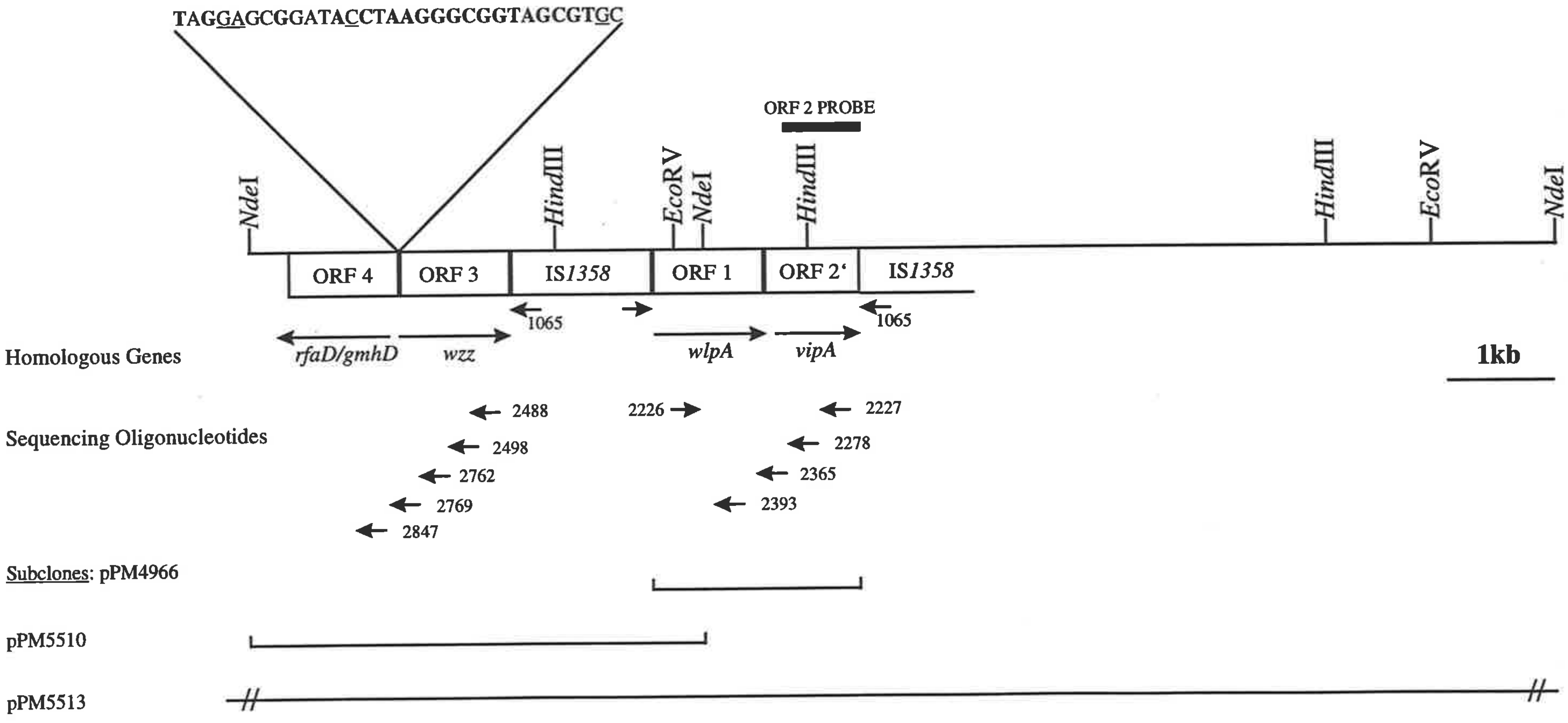
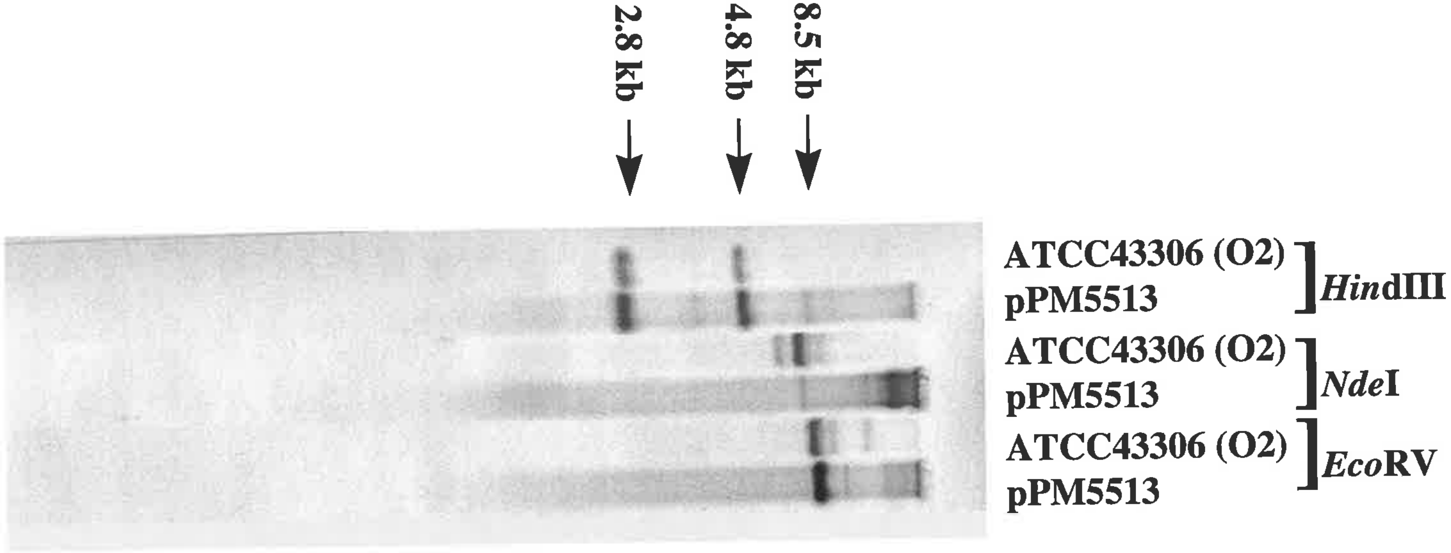

TAGGAGCGGATACCTAAGGGCGGTAGCGTGC
ORF 2 PROBE
NdeI
HindIII
EcoRV
NdeI
HindIII
HindIII
EcoRV
NdeI
ORF 4
ORF 3
IS*1358*
ORF 1
ORF 2'
IS*1358*
1065
1065
Homologous Genes
rfaD/gmhD
wzz
wlpA
vipA
1kb
Sequencing Oligonucleotides
2488
2498
2762
2769
2847
2226
2227
2278
2365
2393
Subclones: pPM4966
pPM5510
pPM5513

Figure 4.12: Southern hybridisation analysis of *V. anguillarum* O2 (ATCC 43306) chromosomal DNA and cosmid DNA from pPM5513.

Southern blots of *Hind*III, *Nde*I, *Eco*RV restricted DNA of ATCC 43306 (O2) and pPM5513 were probed with DIG-labelled probe of ORF2 (*Dra*I fragment, nt 1020 - nt 1830, Fig. 4.5; Fig. 11) (Section 4.2.8). The approximate size of bands are shown in kilobases (kb). The analysis was performed to further map the region in *V. anguillarum* O2 downstream of the second putative IS*1358* element. The order of loading and the restriction enzyme used are indicated at the top of the figure.


2.8 kb
4.8 kb
8.5 kb
ATCC43306 (O2)
pPM5513
*Hin*dIII
ATCC43306 (O2)
pPM5513
*Nde*I
ATCC43306 (O2)
pPM5513
*Eco*RV

4.2.9 Comparison of *V. anguillarum* O2 O-antigen biosynthesis regions

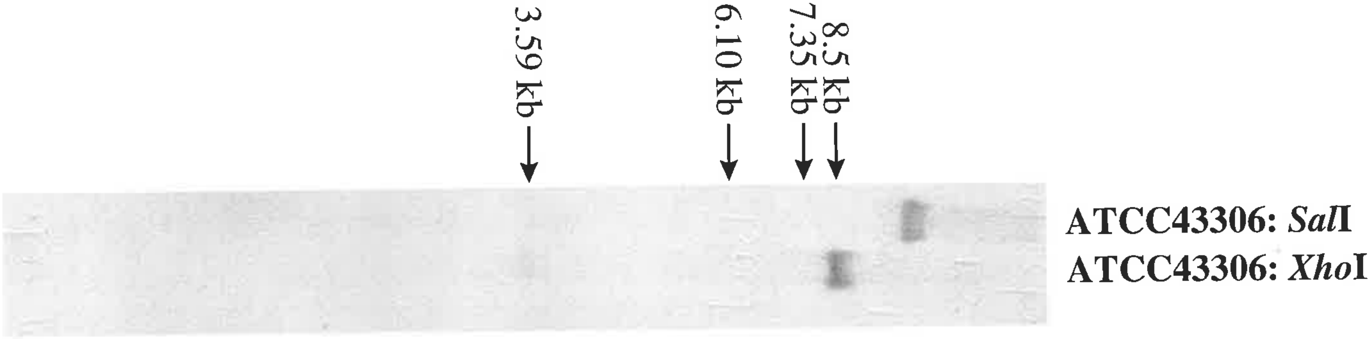
During the course of this study the *rfb* locus responsible for O-antigen biosynthesis in *V. anguillarum* O2 from the strain ATCC19264 were cloned and expressed in *E. coli* (Amor and Mutharia, 1995). A physical map of the region indicating *Sal*I, *Acc*65I and *Nae*I and *Xho*I restriction sites was reported. Using the restriction enzymes *Sal*I and *Xho*I, Southern hybridisation was performed using the 1.8 kb fragment (pPM4966, Section 4.2.6) as a probe to determine if the region sequenced in this study (using the strain ATCC43306) was the same as the region previously published. Figure 4.13 shows that the sizes of the bands for the *Sal*I (<8.5 kb) and *Xho*I (8.5 kb and 3.6 kb) digests do not correlate precisely with the restriction fragments reported by Amor and Mutharia (1995). The reported map indicates the presence of five *Xho*I fragments (5.3, 4.5, 2.5, 2.0, and 1.3 kb) and four *Sal*I fragments (10.2, 2 x 4.0 and 2.5 kb). It is possible that the different strains used in the independent studies accounts for the restriction fragment length polymorphism observed. It is also possible that the region obtained in this study is adjacent the previously mapped locus and therefore the probe used in this study would not hybridise to common bands. Further characterisation of cosmid pPM5513 will allow the region(s) to be more extensively compared.

4.2.10 Analysis of the Rol/Wzz protein from *V. anguillarum* O2

The *wzz* homologue (ORF 3) adjacent to IS*1358* contains the highly conserved proline/glycine rich domain at the carboxy terminal end of the protein which is thought to be important for Wzz function (Bastin *et al.*, 1993; Becker *et al.*, 1995) (Fig. 4.10). The

Figure 4.13: Southern hybridisation analysis of *V. anguillarum* O2 (ATCC 43306)

Southern blots of *Sal*I and *Xho*I restricted DNA of ATCC 43306 (O2) was probed with DIG-labelled pPM4966. The approximate size of bands are shown in kilobases (kb). The order of loading and the restriction enzyme used are indicated at the top of the figure. The analysis was performed to compare the published *V. anguillarum* O2 O-antigen biosynthesis locus restriction map with the region sequenced in this study.


3.59 kb
6.10 kb
7.35 kb
8.5 kb
ATCC43306: *Sal*I
ATCC43306: *Xho*I

predicted *wzz* gene product of 365 amino acids also shows homology to *otnB* from *V. cholerae* O139. OtnB has been shown to be involved with capsule biosynthesis and has some similarity to other known Wzz proteins (Bik *et al.*, 1996). Figure 4.14 shows a CLUSTAL alignment of the Wzz protein from *V. anguillarum* O2 in comparison to other Wzz proteins from *E. coli* and *S. enterica sv* Typhimurium. Table 4.4 summarises the homology between Wzz$_{\mathrm{Va}}$ and related Wzz proteins.

Wzz proteins have a highly conserved structure with the N and C terminal ends being transmembrane domains, with the remaining portion of the protein being located in the periplasm (Morona *et al.*, 1995). Figure 4.15 demonstrates that the Wzz protein from *V. anguillarum* O2 has an identical hydropathy plot to other Wzz proteins. The sequence of the *wzz* gene in *V. anguillarum* indicates the presence of five potential initiation codons at nucleotide #292 (GTG), # 352 (GTG), # 379 (GTG), #403 (GTG) and # 406 (ATG) which are indicated on the sequence in bold letters (Fig. 4.10). Surprisingly none of the potential starts have a ribosome binding site which resembles the consensus sequence (AGGAGGT). The computer program TMpredict (Hofmann and Stoffel, 1993) indicates that the initiation codon at nucleotide #292 (GTG) as the likely initiation codon as this provides the Wzz protein with an appropriate length N terminal transmembrane domain of ~20 amino acids (Morona *et al.*, 1995) in the correct location. This start also provides the best homology at the N-terminal end in comparison to other Wzz proteins (Fig. 4.15).

To determine if Wzz$_{\mathrm{Va}}$ protein had activity, the plasmid containing ORF3, pPM5510, was electroporated into a *Shigella flexneri wzz* mutant (RMA696) (Van den Bosch *et al.*, 1997) to determine if the *V. anguillarum wzz* homologue was capable of complementing this defect. Analysis of LPS by SDS-PAGE and silver staining showed that the *V. anguillarum wzz* homologue was capable of low level complementation (10-15%) compared to the *S. flexneri* wildtype strain (Fig. 4.16). It is possible that the reason

Figure 4.14: Clustal alignment of Wzz_{Va} with other Wzz related proteins

Alignment of the amino acid sequence of Wzz_{Va} (VaWzz), *E. coli* Wzzb2 (EcWzb2) (spQ05032), *E. coli* Wzzb8 (EcWzb8) (pirS33670), *E. coli* Wzz_{Ec} (EcWzz) (gbD90840), *S. typhimurium* Wzzb (StWzzb) (spQ04866), *E. coli* Cld (EcCld) (spQ05033) and *V. cholerae* OtnB (VcOtnB) (pirS54287), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: gb, Genbank; sp, SWISS PROT.

```
VaWzz           ------VPS----YYQQ---SNNDE-IDLRELFSVLWQGKLTIVLCTVLFAVMAVVYALT    46
EcWzb2          ----MSISDN--NMSGR---SHDPEQIDLIDLLMQLWRGKVTIIICIVVAIALAVGYLAV    51
EcWzb8          ----MRVENN--NVSGQ---NLDPEQIDLIDLLVQLWRGKMTIIISVIVAIVLAIGYLVV    51
EcWzz           ---MMRVENN--NVSGQ---NLDPEQIDLIDLLVQLWRGKMTIIISVIVAIVLAIGYLVV    52
StWzzb          ----MTVDSN--TSSGR---GNDPEQIDLIELLLQLWRGKMTIIVAVIIAILLAVGYLMI    51
EcCld           ----MSISDN--NMSGR---SHDPEQIDLIDLLMQLWRGKVTIIICIVVAIALAVGYLAV    51
VcOtnB          MQEQKLTPPNYLTYPAQPTAIGDDE-IDLRELFRALWKGKWTIIAITLVFAIGSVIFAIM    59
                                .       * *** .*   **.** **.   ..    .. .

VaWzz           AQQWWSAKATVTQPELSQVVTFLQQVKQYQPLFDLYQEDGTIIVSKELDGLINPATIFQR   106
EcWzb2          AKEKWTSIAIVTQPDAGQLAS-------YNNAMDVLFGANAPEMTDVQAGFIG------R    98
EcWzb8          AKEKWTSTAIVTQPDVGQIAG-------YNNAINVIYGSAAPKVSEIQSILIG------R    98
EcWzz           AKEKWTSTAIVTQPDVGQIAG-------YNNAINVIYGSAAPKVSEIQSILIG------R    99
StWzzb          AKEKWTSTAIITQPDAAQVAT-------YTNALNVLYGGNAPKISEVQANFIS------R    98
EcCld           AKEKWTSIAIVTQPDAGQLAS-------YNNAMDVLFGANAPEMTDVQAGFIG------R    98
VcOtnB          QPNIYKAEALIAPASEQQGGG--------LSALASQFGGLASLAGVNLGGKGG----VDK   107
                      . * ..     *                      .                  .

VaWzz           FVRAFNTNDNKKRFMQSNPTFLAIEQQQLKVNENNETDDVARFYQSWYGKITAQEVGKTQ   166
EcWzb2          FSSAFSALA--------------------ETLDN----------QQEPEKLSIDAAVKGQ   128
EcWzb8          FSTTFSALA--------------------ETLDN----------QEEPEKLTIEPTVKNQ   128
EcWzz           FSTTFSALA--------------------ETLDN----------QEEPEKLTIEPTVKNQ   129
StWzzb          FSSAFSALS--------------------EVLDN----------QKEREKLTIEQSVKGQ   128
EcCld           FSSAFSALA--------------------ETLDN----------QQEPEKLSIDAAVKGQ   128
VcOtnB          TQMAIEVLKS-------------------RQFTS----------EFIQQHNILADIMAAK   137
                   .                                        .    .         .

VaWzz           AGDFYLSFQSIDKASSLVLLNAYIDFINQQLNQQLTNDLTSKLTVKHNQLSQQYSSLQQQ   226
EcWzb2          TLPLKVSYTGKTSEEAQKTLAQYIQQVDEGVAKELNADLSMSMETRLADLQKSLAAQEAV   188
EcWzb8          SLPLAVSYVGQSPEAAQKQLAQYIQQVDDQVNDELEKDLKDNIALRMKNLQDSLKTQEVV   188
EcWzz           SLPLAVSYVGQSPEAAQKQLAQYIQQVDDQVNDELEKDLKDNIALRMKNLQDSLKTQEVV   189
StWzzb          ALPLSVSYVSTTAEGAQRRLAEYIQQVDEEVAKELEVDLKDNITLQTKTLQESLETQEVV   188
EcCld           TLPLKVSYTGKTSEEAQKTLAQYIQQVDEGVAKELNADLSMSMETRLADLQKSLAAQEAV   188
VcOtnB          KWDNSADQLIYDPALYNAQTHTWVREAKAPFKPEPSMQEAYKVFSKLVTVNKAKDSGMVT   197
                                       .         .        .  .   .     .

VaWzz           TQLRLQVELARVQNALAIAKAANINEP-VQNLNEE-KLFAISIGSKALQAKVDALKSITN   284
EcWzb2          AKEQKTLRIAQMTQALKVAQQSNIKLPQVQQVDQVSQDSMFMLGSEALSSMVENEATRP-   247
EcWzb8          AQEQKELRIRQIQEALQYANQAQVTKPQIQQTQDVTQDTMFLLGSDALESMVKHEASRP-   247
EcWzz           AQEQKELRIRQIQEALQYANQAQVTKPQIQQTQDVTQDTMFLLGSDALESMVKHEASRP-   248
StWzzb          AQEQKDLRIKQIEEALRYADEAKITQPQIQQTQDVTQDTMFLLGSDALKSMIQNEATRP-   247
EcCld           AKEQKTLRIAQMTQALKVAQQSNIKLPQVQQVDQVSQDSMFMLGSEALSSMVENEATRP-   247
VcOtnB          IAVEHLSPTVAQQWVTWLIEDINKVMK-ERDVAEAHRSTAFLNEQIALTNVADIRTVLY-   255
                                             .      .         **

VaWzz           LSVFEPRLALLQAQVQQVELLGKVKPAQVQGYAYLEQPEAPISRDEPKRALIAVLGTLLG   344
EcWzb2          -LTFSDQYYQTRQNLLEVQAL-EVAPDSVHAYRYVMKPTLPIRRDSPKKAITLVLAVLIG   305
EcWzb8          -LVFSSTYYQTRQNLLDIESL-KVDDLDIHAYRYVMKPTLPIRRDSPKKAITLILAVLLG   305
EcWzz           -LVFSSTYYQTRQNLLDIESL-KVDDLDIHAYRYVMKPTLPIRRDSPKKAITLILAVLLG   306
StWzzb          -LVFSPAYYQTKQTLLDIKNL-KVTADTVHVYRYVMKPTLPVRRDSPKTAITLVLAVLLG   305
EcCld           -LTFSDQYYQTRQNLLEVQAL-EVAPDSVHAYRYVMKPTLPIRRDIPKKAITLVLAVLIG   305
VcOtnB          ------KLIEEQAKTIMFAEV---RDEYVFKT--IDPALAPEEKAKPKRALICVLGTLLG   304
                           .        .       .     .     *  .  ** *.  .*  *.*

VaWzz           GMLGVAIVLVRFAFRKEEEKA---------                                 365
EcWzb2          GMIGAGVVLGRNALRGYKAKAE--------                                 327
EcWzb8          GMVGAGIVLGRNALRNYNAK----------                                 325
EcWzz           GMVGAGIVLGRNALRNYNAK----------                                 326
StWzzb          GMIGAGIVLGRNALRSYKPKAL--------                                 327
EcCld           GMIGAGVVLGRNALRGYKAKAE--------                                 327
VcOtnB          GMLGVAIVLVRFAFRKAGKNHEHADVRETA                                 334
                **.*  .** * * *
```

Table 4.4: Homology Table for Wzz$_{Va}$

	%identity[a]						
	VaWzz	**EcWzzb2**	**EcWzzb8**	**EcWzz**	**StWzzb**	**EcCld**	**VcOtnB**
VaWzz	100	26.2	27.7	27.7	28.5	26.2	32.1
EcWzzb2		100	64.9	64.9	64.7	99.7	19.9
EcWzzb8			100	100	70.5	64.6	18.6
EcWzz				100	70.5	64.6	18.6
StWzzb					100	64.4	20.3
EcCld						100	19.9
VcOtnB							100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Identity is over the entire length of the protein. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaWzz: *V. anguillarum* Wzz$_{Va}$

EcWzzb2: *E. coli* Wzzb2 (spQ05032)

EcWzzb8: *E. coli* Wzzb8 (pirS33670),

EcWzz: *E. coli* Wzz$_{Ec}$ (gbD90840),

StWzzb: *S. typhimurium* Wzzb (spQ04866),

EcCld: *E. coli* Cld (spQ05033)

VcOtnB: *V. cholerae* OtnB (pirS54287)

Figure 4.15: Hydropathy plot of Wzz proteins

The hydropathy plots were generated by the method of Kyte and Doolittle (1982) and aligned using PROFILEGRAPH (Hofmann and Stoffel, 1989). Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 10th amino acid is marked on each X-axis. VAWZZ, *V. anguillarum* Wzz_{292}; WZB2, *E. coli* Wzzb2 (spQ05032); WZB8, *E. coli* Wzzb8 (pirS33670); ECWZZ, *E. coli* Wzz (gbD90840), and OTNB, *V. cholerae* OtnB (pirS54287). Abbreviations: gb, Genbank; sp, SWISS PROT.

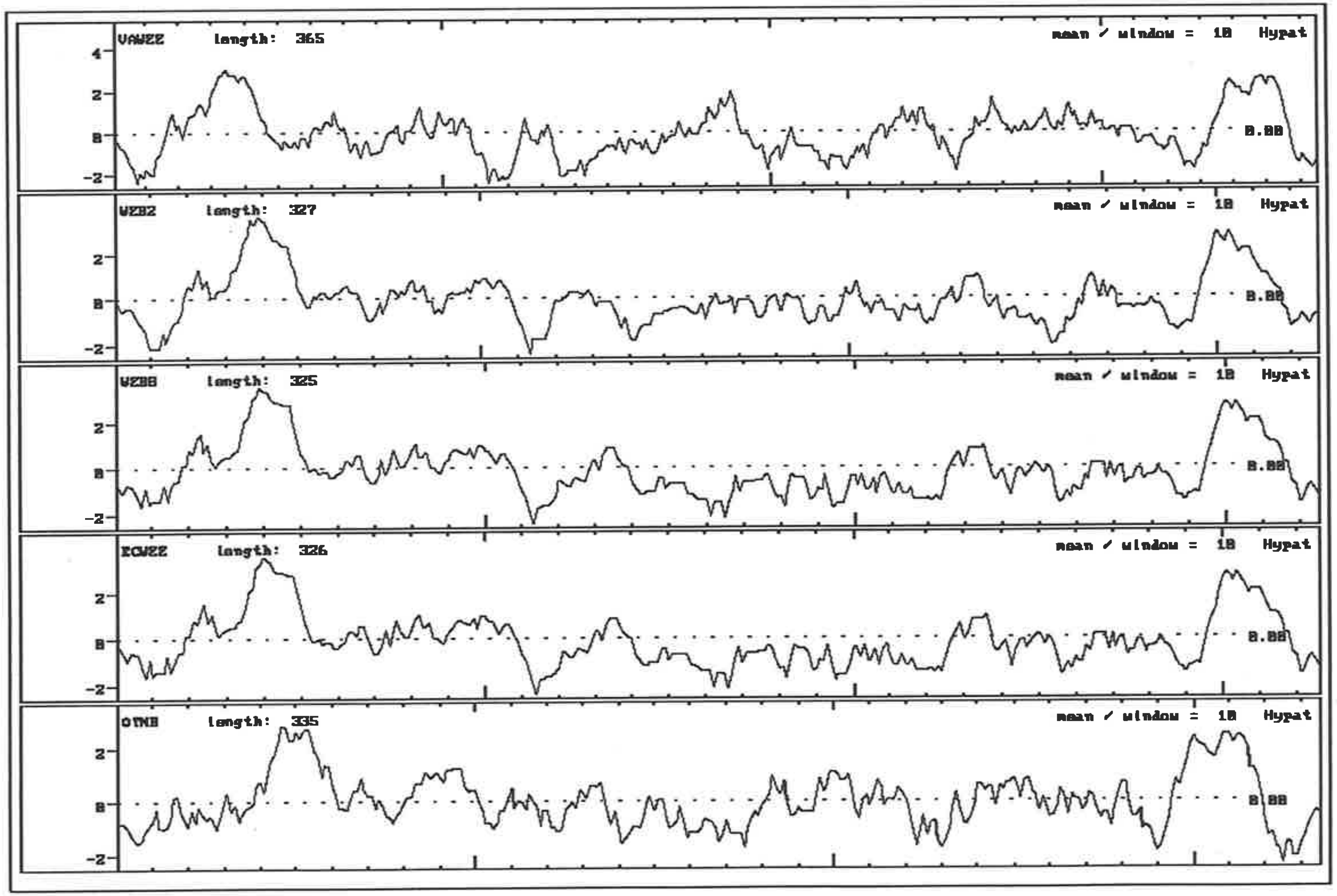
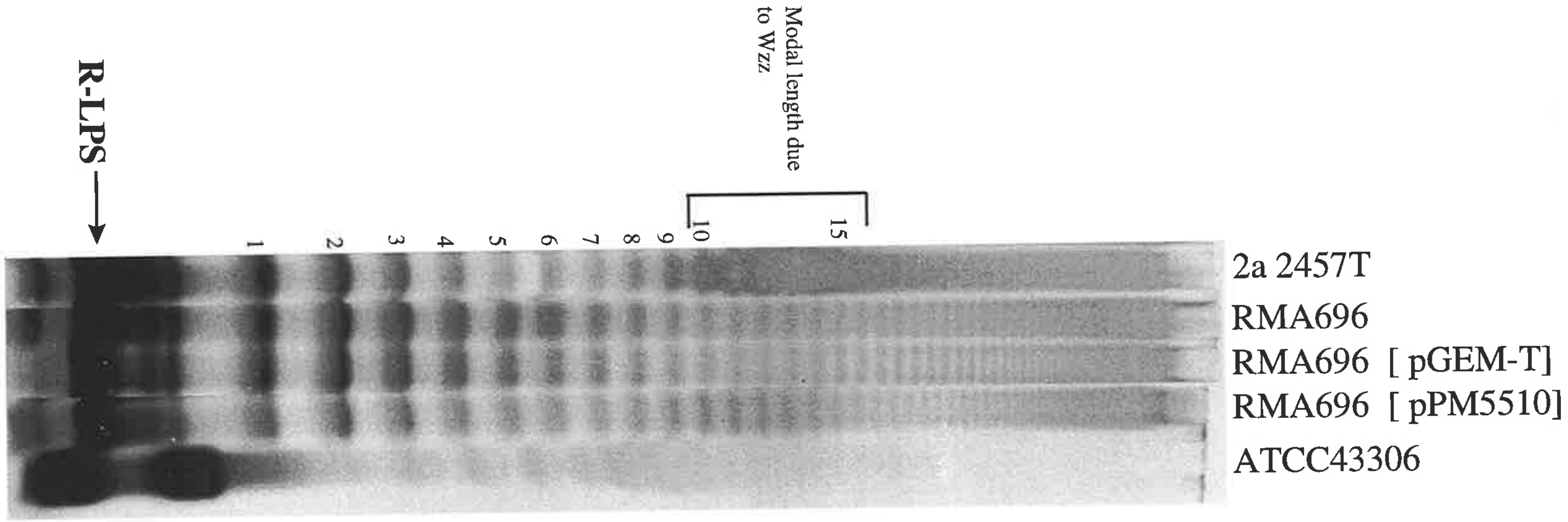

VAWZZ
length: 365
mean / window = 10
Hypat
4
2
0
-2
0.00
WZB2
length: 327
mean / window = 10
Hypat
2
0
-2
0.00
WZBB
length: 325
mean / window = 10
Hypat
2
0
-2
0.00
ECWZZ
length: 326
mean / window = 10
Hypat
2
0
-2
0.00
OTNB
length: 335
mean / window = 10
Hypat
2
0
-2
0.00

Figure 4.16: Complementation analysis of Wzz$_{Va}$

This shows a silver-stained 15% SDS-polyacrylamide gel with lipopolysaccharide prepared from the indicated strains by proteinase K treatment (Materials and Methods, 2.15.1). The first lane contains the *S. flexneri* 2a wildtype strain (2457T). The remaining lanes contain RMA696 (*S. flexneri wzz* mutant), RMA696 [pGEM5zf$^+$], RMA696 [pPM5510] and the wildtype strain of *V. anguillarum* O2 (ATCC 43306). The lipid A-core oligosaccharide (R-LPS) and the number of O-antigen repreat units are indicated on the left side of the figure. The LPS O-antigen modal length due to Wzz is shown on the right side of the figure. Samples represent ~1 X 10^8 cells.


R-LPS
Modal length due to Wzz
1
2
3
4
5
6
7
8
9
10
15
2a 2457T
RMA696
RMA696 [pGEM-T]
RMA696 [pPM5510]
ATCC43306

for the low level of complementation is due to poor expression as the Wzz_{Va} gene lacks a RBS.

4.2.11 Presence of ORF 1 and ORF 2 in other *V. anguillarum* and *V. cholerae* serotypes

Due to the presence of IS*1358* in many different strains of both *V. cholerae* and *V. anguillarum,* it was of interest to determine if either ORF 1 or ORF2 were present in any of these strains representing the different serotypes of these species. This was investigated in an attempt to provide indirect evidence for the possible transfer of these genes to *V. anguillarum* O2. Initial studies using dot-blot DNA hybridisation indicated the presence of both these ORFs in a number of non-O1 *V. cholerae* serotypes. Table 4.5 shows the non-O1 *V. cholerae* serotypes that contain ORF 1 and ORF 2 by dot blot using PCR DIG-labelled probes for either ORF 1 or ORF 2. The presence of ORF 1 and/or ORF 2 could not be detected by Southern hybridisation in any of the other *V. anguillarum* serotypes. The presence of IS*1358* is also indicated in Table 4.5. Interestingly, some of the strains that contain these ORFs also contain at least one copy of IS*1358,* including one *V. cholerae* serotype (O94) which contained all three: ORF1, ORF2 and IS*1358*. ORF 2 is also present in strains lacking IS*1358* (O8, O10).

4.2.12 IS*1358* and *virBA*

Due to the prevalence of IS*1358* in the *V. anguillarum* strains tested in this study, and the location of these elements in the chromosome of *V. anguillarum* O2 in relation to *rfb*/ polysaccharide biosynthesis genes, it was decided to further investigate the association of IS*1358* with polysaccharide biosynthesis-related genes in another *V. anguillarum* serotype. Initially, a *Sau*3A cosmid gene library was constructed using the cosmid vector

Table 4.5: Presence of ORF 1 and ORF 2 in *V. cholerae*

Serotype	ORF 1	ORF 2	Presence of IS*1358*
V. anguillarum			
O2	Y	Y	Y
V.cholerae			
O8	N	Y	N
O10	N	Y	N
O27	Y	N	Y
O30	Y	N	N
O38	Y	N	N
O72	Y	N	N
O88	Y	N	N
O94	Y	Y	Y
O98	Y	N	Y
O110	Y	N	N
O111	Y	N	N
O128	Y	N	N
O132	Y	N	Y

The table summarises data obtained from Southern hybridisation and dot blots of *V. cholerae* and *V. anguillarum* (Section 4.2.12). The DIG-labelled probes used: ORF1 corresponds to nucleotide #290 to #985 (Fig. 4.5), ORF2 corresponds to nucleotide #1020 to #1830 (Fig. 4.5; Fig. 4.11), IS*1358* generated from *V. anguillarum* O1 using oligonucleotide #773 (Table 2.4).
Y: the probe hybridised
N: not detected

pPM2101 (Table 2.3) and chromosomal DNA of *V. anguillarum* O1 serotype (strain 85-3954-2) which was then transfected into an *E. coli* (S17-1) background

Colony dot-blot hybridisation was performed using a DNA probe of IS*1358* in an attempt to isolate cosmid clones that contained IS*1358* and surrounding DNA. Approximately 1000 cosmid clones were screened, with 5 hybridising to the IS*1358* probe (data not shown). To determine if the complete IS*1358* element was contained on the cosmid clones (pPM4905, pPM4906, pPM4914, pPM4917, pPM4918), PCR amplification was performed using oligonucleotide #773 derived from the inverted repeat sequences. Three of the clones (pPM4905, pPM4917, pPM4918) contained the complete element (data not shown). The cosmids were then investigated for linkage of IS*1358* to two previously described genes, *virA* and *virB*, that have been shown to be involved in LPS biosynthesis in *V. anguillarum* O1 (Norqvist & Wolf-Watz, 1993). Oligonucleotides #1088 and #1089 (Table 2.4) were designed based on the published sequence (accession number: L08012) to allow amplification of *virB*. PCR reactions with all 5 cosmid clones resulted in amplification of a 1.1 kb fragment consistent with the size of *virB* from pPM4905, pPM4917 and pPM4918. The PCR products were cloned into pGEM-T yielding pPM4901 (from pPM4905) and pPM4902 (from pPM4917), and confirmed to be *virB* by sequencing (data not shown). This result suggests that IS*1358* and *virB* are located in the same region of the chromosome. A putative function for the *virB* gene product was not reported (Norqvist and Wolf-Watz, 1993). However, when a BlastX search was performed on the sequence, an ORF with significant homology (43% identity, 70 % similarity at the amino acid level) to glycosyl transferase genes from *Haemophilus influenzae* and *Neisseria gonorrhoea* was identified. This suggests *virB* (and *virA*) may function as transferases in the LPS O-antigen biosynthetic pathway. This is consistent with the observation that mutations in *virBA* cause a loss of LPS O-antigen in *V. anguillarum* O1 (Norqvist & Wolf-Watz, 1993).

Southern hybridisation was performed to determine the proximity of IS*1358* to *virB*. Southern blot analysis using IS*1358* and *virB* probes on *Eco*RI digested chromosomal of 85-3954-2 and cosmid DNA of pPM4905, respectively, showed that the genes were located on the same large (~20 kb) *Eco*RI fragment (Fig. 4.17). Thus, the genes are in close proximity, and as in *V. cholerae* O1 and O139 (Stroeher *et al.*, 1995, 1998), IS*1358* is again associated with a potential *rfb*/O-antigen biosynthesis gene locus.

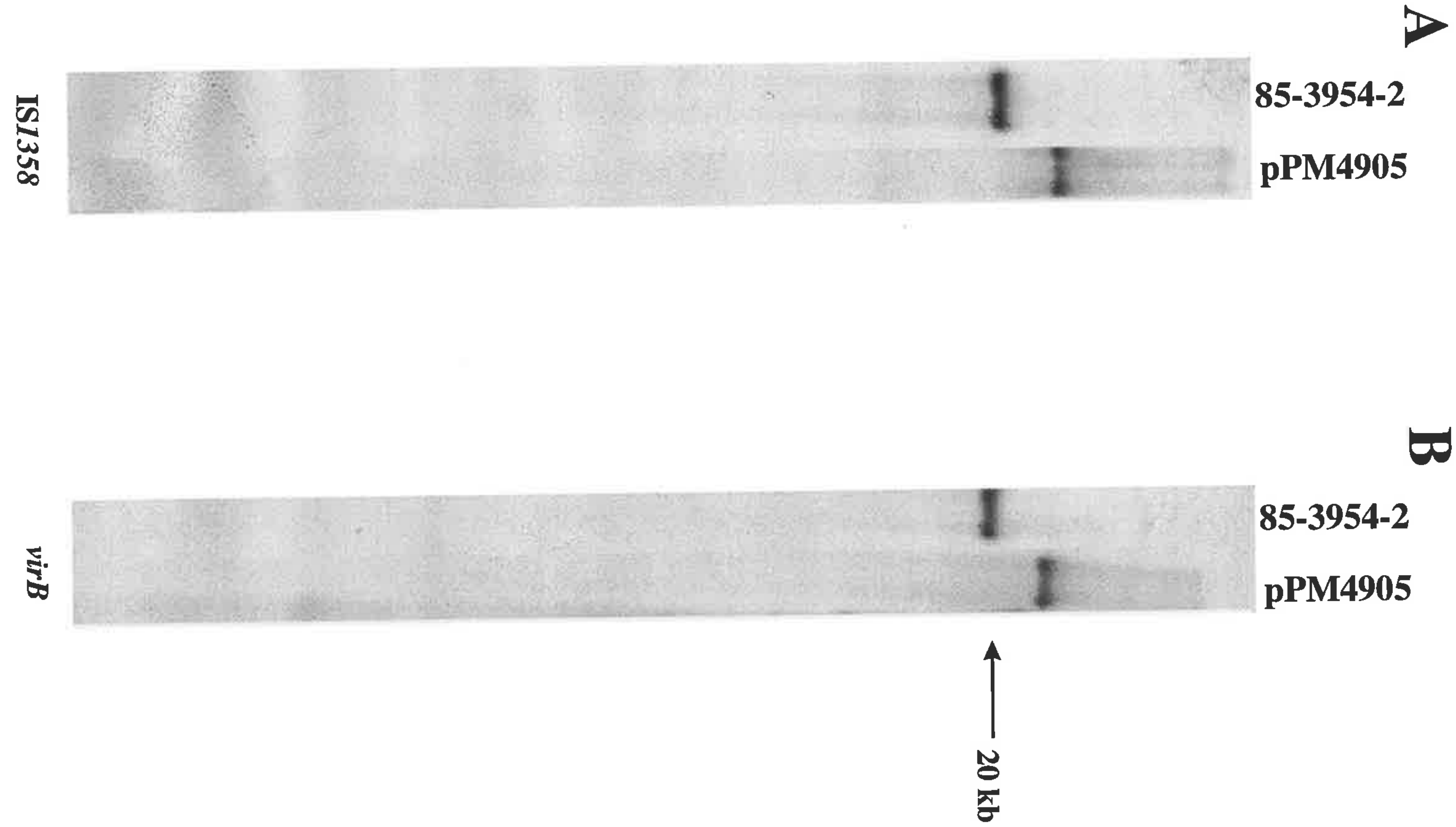
4.3 Summary and Discussion

In this chapter, the distribution of the insertion sequence, IS*1358* was described for *V. anguillarum* and *V. cholerae*. The IS*1358* element was found to be widely distributed in these species as it was found in 7 out of 11 serotypes in *V. anguillarum*, and 65 out of 154 serotypes in *V. cholerae*. The elements in *V. anguillarum* were further characterised by cloning, sequencing, and T7 overexpression. Interestingly, IS*1358* has again been found to be located either near or within gene clusters involved in polysaccharide biosynthesis. This was the case for both *V. anguillarum* serotypes O1 and O2. The *rfb* gene cluster of *V. anguillarum* O1 is further characterised and discussed in the following chapter.

During the course of this study the *rfb* locus from the O2 serotype of *V. anguillarum* was cloned and expressed in *E. coli* by Amor and Mutharia (1996). The published chromosomal restriction enzyme map of this region (Chapter 1) was used to determine if the *V. anguillarum* O2 genes sequenced in this study were located in the described *rfb* locus. However, a number of discrepancies with our Southern hybridisation data in comparison to the published restriction enzyme map made it difficult to determine if the two sets of genes were located in this region. This may explained by the use of

Figure 4.17: Southern hybridisation analysis of *V. anguillarum* O1 (85-3954-2) chromosomal and pPM4905 DNA

Southern blots of *Eco*RI restricted DNA of 85-3954-2 (O1) and pPM4905 were probed with DIG-labelled probes of IS*1358* (generated using oligonucleotide #773) (A) and *virB* (nt 25700 - nt 26788, Appendix 1) (B). The approximate size of the bands are shown in kilobases (kb). This analysis was performed to demonstrate that IS*1358* and *virB* were located on the same fragment on the chromosome and cosmid.


A
IS1358
85-3954-2
pPM4905
B
virB
85-3954-2
pPM4905
20 kb

different strains in the two independent studies, use of an inappropriate probe or it may also suggest the presence of other *rfb*-like genes in the O2 chromosome that may contribute to LPS (O-antigen) expression. Alternatively, the genes identified may either be involved in capsule/core oligosaccharide synthesis, or perhaps they are redundant and not expressed *in vivo*.

Recently Dumontier *et al.* (1998) demonstrated that the IS*1358* element from *V. cholerae* O22, which has an intact ORF and multiple copies of the element in the chromosome, was capable of transposition. This suggests that IS*1358* elements that contain a single ORF are capable of transposition. Indirect evidence that IS*1358* may be mobile in *V. anguillarum* stems from the region characterised in the O2 serotype. The presence of two ORFs, one intact (ORF 1) and the other interrupted (ORF 2), suggests that the IS*1358* element adjacent ORF 2 may have either transposed into this gene, or complex recombination events occurred resulting in the interruption to ORF 2 and the possible introduction of ORF 1. It is conceivable the introduction of a new gene of similar function (and disruption of the old gene) could alter the antigenic properties of the O-antigen.

Chapter Five

Cloning, sequencing and analysis of the *wbh* operon of *V. anguillarum* O1

5.1 Introduction

Gene clusters or operons determining expression of the O-antigen component of LPS have been well characterised in a number of Gram negative bacteria. Analysis of the genes located within these clusters have played a vital role in understanding the molecular basis of O-antigen expression and composition.

The gene cluster responsible for O-antigen biosynthesis in *V. anguillarum* O1 has yet to be identified. However two genes, *virBA* have been isolated and shown to be associated with virulence. Mutations in these genes result in the loss of O-antigen production (Norqvist and Wolf-Watz, 1993) and therefore it is possible that these genes are part of the gene cluster responsible for the biosynthesis of this component of LPS.

In this chapter the isolation and characterisation of the O-antigen biosynthesis operon of *V. anguillarum* O1 is described. Using the new nomenclature for polysaccharide synthesis genes, the operon has been designated the name *wbh*.

5.2 Results

5.2.1 Isolation and expression of the O-antigen biosynthesis gene cluster using cosmid cloning

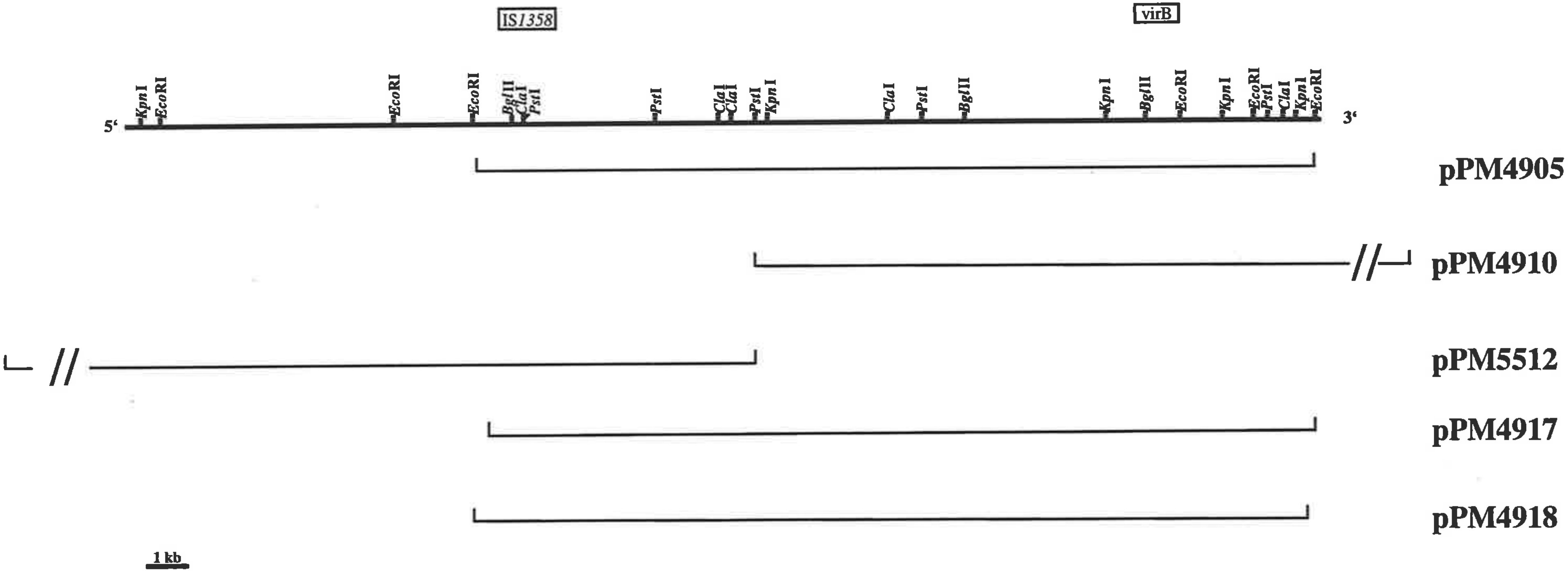
The cosmid clones isolated in the section 4.2.12 (Chapter 4) were used to initiate the cloning and characterisation of the O-antigen biosynthesis locus. *E. coli* K-12 strain S17-1 harbouring the cosmids pPM4905, pPM4917, and pPM4918 were tested by colony and Western immunoblotting for expression of *V. anguillarum* O1 O-antigen. The antiserum used was raised against the strain 85-3954-2 (O1 serotype) and affinity purified such that the serum only contained antibodies to the *V. anguillarum* O1 LPS (Section 2.14.1). None of the *E. coli* strains containing the cosmid clones tested were positive with the antiserum. This suggested that either more genes were required for expression or the O-antigen was not able to be expressed in an *E. coli* K-12 background for an unknown reason(s). Subsequent dye-terminator sequencing of the ends of the DNA insert in the cosmid clones using oligonucleotides #2177 and #2366 (Table 2.4) showed that all three cosmids clones were essentially identical (siblings) and that the same *rfb*-like genes were truncated at both the 5' and 3' end (Fig. 5.1). This suggested, in addition to the negative expression of LPS by the cosmid clones, that additional genes may be required for expression.

In order to locate and isolate the additional genes, the original cosmid library was re-screened using DIG-labelled probes generated from fragments of the cosmid pPM4905. This resulted in an additional cosmid, (pPM4910) being isolated which spanned the region at the *virB* end of the DNA insert in pPM4905 (Fig. 5.1). Random subcloning of cosmid pPM4910 and subsequent sequencing showed more regions of homology *rfb*/polysaccharide synthesis-like genes which differ from those detected in sequences derived from pPM4905 (data not shown).

A cosmid clone spanning the region upstream of IS*1358* was required to ensure that the entire O-antigen biosynthesis region of *V. anguillarum* O1 was obtained. Additional cosmid libraries were constructed according to the method described in section 2.10.6.1. These libraries were constructed using cosmid cloning vectors pOU61cos (Amp^R) (Knott *et al.*, 1988) and pTB023 (T. Barnett, personal communication) which confers chlorampenicol resistance but is otherwise identical to pPM2101 (Table 2.3). DIG labelled

Figure 5.1: Map indicating relative positions of cosmids pPM4905, pPM4910, pPM4917, pPM4918 and pPM5512.

Schematic representation of the *Sau*3A-derived cosmid clones (pPM4905, pPM4910, pPM4917, pPM4918, and pPM5512) isolated by colony hybridisation using either DIG-labelled probes of IS*1358* (generated using oligonucleotide #773, Table 2.4) or *virB* (nt 25700-nt 26788, Appendix 1). Cosmids were sequenced with oligonucleotides #2177 and #2366 to determine the points of insertion relative to the *wbh* region. // represents undefined distance. IS*1358* and *virB* are marked on the map for reference. Restriction sites were determined by Souther hybridisation and sequencing (Section 5.2.6).


IS1358
virB
5'
KpnI
EcoRI
EcoRI
EcoRI
BglII
ClaI
PstI
PstI
ClaI
ClaI
PstI
KpnI
ClaI
PstI
BglII
KpnI
BglII
EcoRI
KpnI
EcoRI
PstI
ClaI
KpnI
EcoRI
3'
pPM4905
pPM4910
pPM5512
pPM4917
pPM4918
1 kb

IS*1358* was used as a probe in colony hybridisation to screen 500 cosmid clones. One cosmid clone (pPM5512) was isolated from the bank made with the pTB023. pPM5512 was sequenced using the oligonucleotides #2177 and #2366 (Table 2.4). pPM5512 was found to overlap with pPM4905, pPM4917, and pPM4918 but also contained novel DNA not previously identified on other cosmid clones (Fig. 5.1).

5.2.2 Random subcloning of cosmids pPM4905 and pPM4910

The sequence between and adjacent IS*1358* and *virB* was likely to be the putative *wbh*/O-antigen biosynthesis region of *V. anguillarum* O1 and therefore this region was sequenced using the cosmid clones obtained. Three different cloning techniques were used. Cosmids pPM4905 and pPM4910 were digested in single or double digests using combinations of six different restriction enzymes. The resulting fragments were ligated into like-wise digested pBC-KS$^+$. The clones obtained with this approach are shown in Figure 5.2. The clones were sequenced using -21M13 and M13 primers and oligonucleotide primers designed from initial sequence obtained (Table 2.4). Analysis of the sequence allowed overlapping clones to be joined, and contiguous DNA sequence of this part of the region was created. To ensure no errors occurred during both the generation of the cosmids and subcloning, the subclones were DIG-labelled and used in Southern hybridisation (Fig. 5.3). Figure 5.3 demonstrates, as an example that the DIG labelled probes (pPM4953, pPM4955, *virB*) derived from pPM4905 are located on identical restriction fragments in the chromosome and cosmids. pPM4910 (Fig. 5.3(c)) does not hybridise with the pPM4943 probe as the cosmid does not span this region (Fig. 5.2). The Southern blots also served the purpose of allowing the region surrounding the subclones to be mapped (Fig. 5.2) with respect to the restriction sites in the cosmid cloning vector,

Figure 5.2: Map indicating relative positions of sub-clones of pPM4905 and pPM4910.

Schematic representation of the subclones obtained from random sub-cloning of the cosmids pPM4905 and pPM4910, digested with the restriction enzymes listed below. The shaded box indicates the vector, pPM2101. Gap A, B, and C are regions where no sub-clones were obtained using shot-gun cloning and are marked for reference. The capital letters above the end of the cloned region indicates the restriction enzymes used to generate the subclones: X, *Xba*I; E, *Eco*RV; P, *Pst*I; C, *Cla*I; H, *Hind*III; Ec, *Eco*RI. IS*1358* and *virB* are marked on the map for reference.

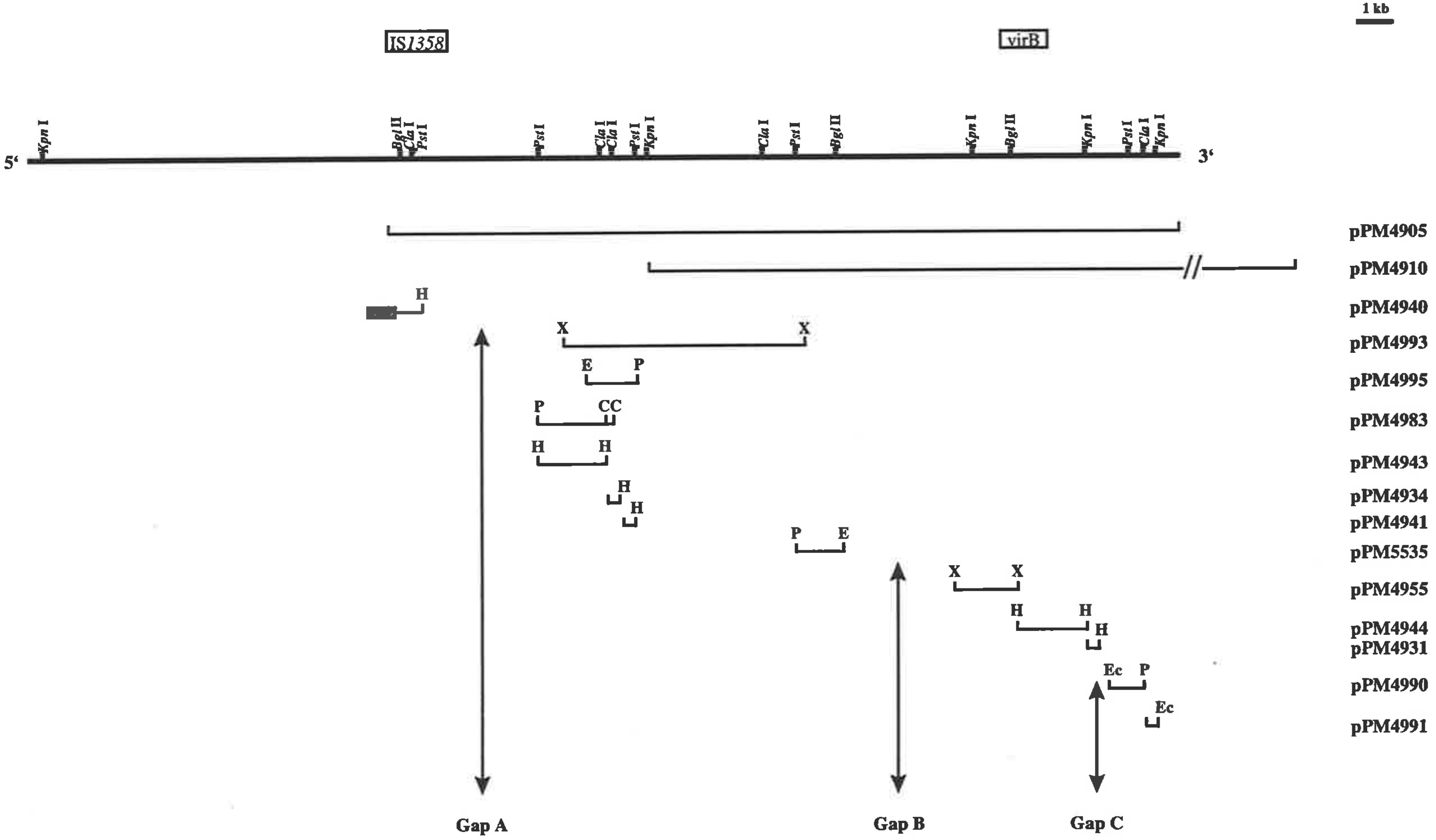

1 kb
IS1358
virB
5'
3'
Kpn I
Bgl II
Cla I
Pst I
Pst I
Cla I
Cla I
Pst I
Kpn I
Cla I
Pst I
Bgl II
Kpn I
Bgl II
Kpn I
Pst I
Cla I
Kpn I
pPM4905
pPM4910
H
pPM4940
X
X
pPM4993
E
P
pPM4995
P
CC
pPM4983
H
H
pPM4943
H
pPM4934
H
pPM4941
P
E
pPM5535
X
X
pPM4955
H
H
pPM4944
H
pPM4931
Ec
P
pPM4990
Ec
pPM4991
Gap A
Gap B
Gap C

Figure 5.3: Southern hybridisation analysis of *V. anguillarum* O1 (85-3954-2) chromosomal DNA and cosmid DNA from pPM4905 and pPM4910.

Southern blots of *Hin*dIII/*Cla*I restricted DNA of 85-3954-2, pPM4905 and pPM4910 were probed with PCR DIG-labelled probes of pPM4943 (A), pPM4955 (B) and *virB* (nt 25700- nt 27688, Appendix 1) (C). This analysis was performed to demonstrate that arrangement of chromosomal DNA and the cloned genes on the cosmids were identical.

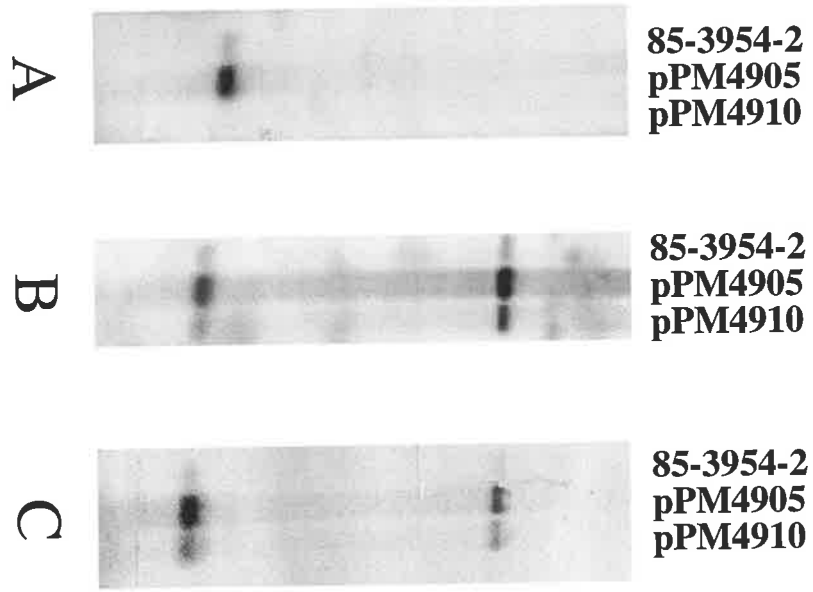

A
85-3954-2
pPM4905
pPM4910
B
85-3954-2
pPM4905
pPM4910
C
85-3954-2
pPM4905
pPM4910

Table 5.1:Sizes of restricted fragments which hybridise to various sublcones of the wbh region of *V. anguillarum* O1.

	Sizes[a] of restricted fragments which hybridise to :									
Restriction enzyme	**IS*1358***	**pPM4940**	**pPM4990**	**pPM4991**	**pPM4993**	**pPM5535**	**pPM4943**	**pPM4995**	**pPM4983**	**pPM4955**	***vir*B**
***Eco*RI**	18	18	1.3	1.3	NA	NA	NA	NA	NA	18	18
***Hin*dIII**	4.8 1.8	4.8	NA	NA	NA	NA	1.1	NA	NA	3.7 0.9	4.0 0.9
***Cla*I**	>10 5.2	NA	2.5	NA	5.2 4.0 0.2	NA	1.6 0.2	5.2 4.0 0.2	5.2 0.2	10	NA
***Pst*I**	>10 3.5	NA	1.3	1.3	2.5 4.2	8.5 2.5	2.5	2.5	2.4	NA	9.0
***Bgl*II**	>10 0.4	NA	1.6	NA	NA	>10 4.5	>10	>10	NA	5.0 4.5	NA
***Eco*RV**	NA	NA	NA	NA	NA	3.8	NA	NA	NA	NA	NA
***Xba*I**	NA	NA	NA	NA	6.2	NA	>15 5.2	NA	NA	1.4	1.5 1.4
***Hinc*II**	NA	NA	NA	NA	NA	NA	4.8	NA	NA	3.8	NA
***Eco*RI/PstI**	NA	NA	1.0	0.3	NA	NA	NA	NA	NA	NA	NA
***Cla*I/PstI**	>10 3.5	NA	1.0		3.3 1.7 0.8 0.6 0.2	NA	1.6 0.2	1.7 0.6 0.2	1.6 0.2	NA	NA
***Bgl*II/*Cla*I**	5.2 0.4 0.2	NA	NA	NA	NA	NA	NA	5.2 4.0 0.2	NA	NA	NA
***Eco*RV/*Pst*I**	NA	NA	NA	NA	NA	1.1	NA	NA	NA	NA	NA
***Hin*dIII/*Cla*I**	NA	NA	NA	NA	NA	NA	1.1	NA	NA	3.7 0.9	NA
***Hin*dIII/*Xba*I**	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.0 0.3	1.5 1.4
***Bgl*II/HincII**	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.5 1.3	NA
***Bgl*II/*Xba*I**	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.0 0.4	NA

[a] size is approximate and is in kilobases (kb)

PCR DIG labelled probes specific to subclones or DNA fragments: IS*1358* (nt. 9801-11139), *vir*B (nt. 25700-26788), pPM4940 (nt. 9506-10239), pPM4943 (nt. 14289-15442), pPM4955 (nt. 24832-26226), pPM4983 (nt 13668-15531), , pPM4990 (nt. 29305-30324), pPM4991 (nt. 30324-30623)pPM4993 (nt. 14416-20628), pPM4995 (nt. 14895-16139), pPM5535 (nt. 20329-21426).

nt. corresponds to the nucleotide numbers in Appendix 1.

NA: not analysed by Southern hybridisation pPM2101. The data obtained from the numerous Southern hybridisation blots performed are summarised in Table 5.1.

5.2.3 Generation of specific subclones

Random cloning in conjunction with Southern hybridisation allowed the generation of a physical chromosomal map that highlighted gaps in the contiguous sequence. Clones were required to obtain sequence between pPM4940 and pPM4983 near IS*1358* (Gap A) and between pPM5535 and pPM4955 (Gap B) towards the 3' end of the region (Fig. 5.2). Hence the second strategy used to complete the sequencing of this region involved the generation of defined clones.

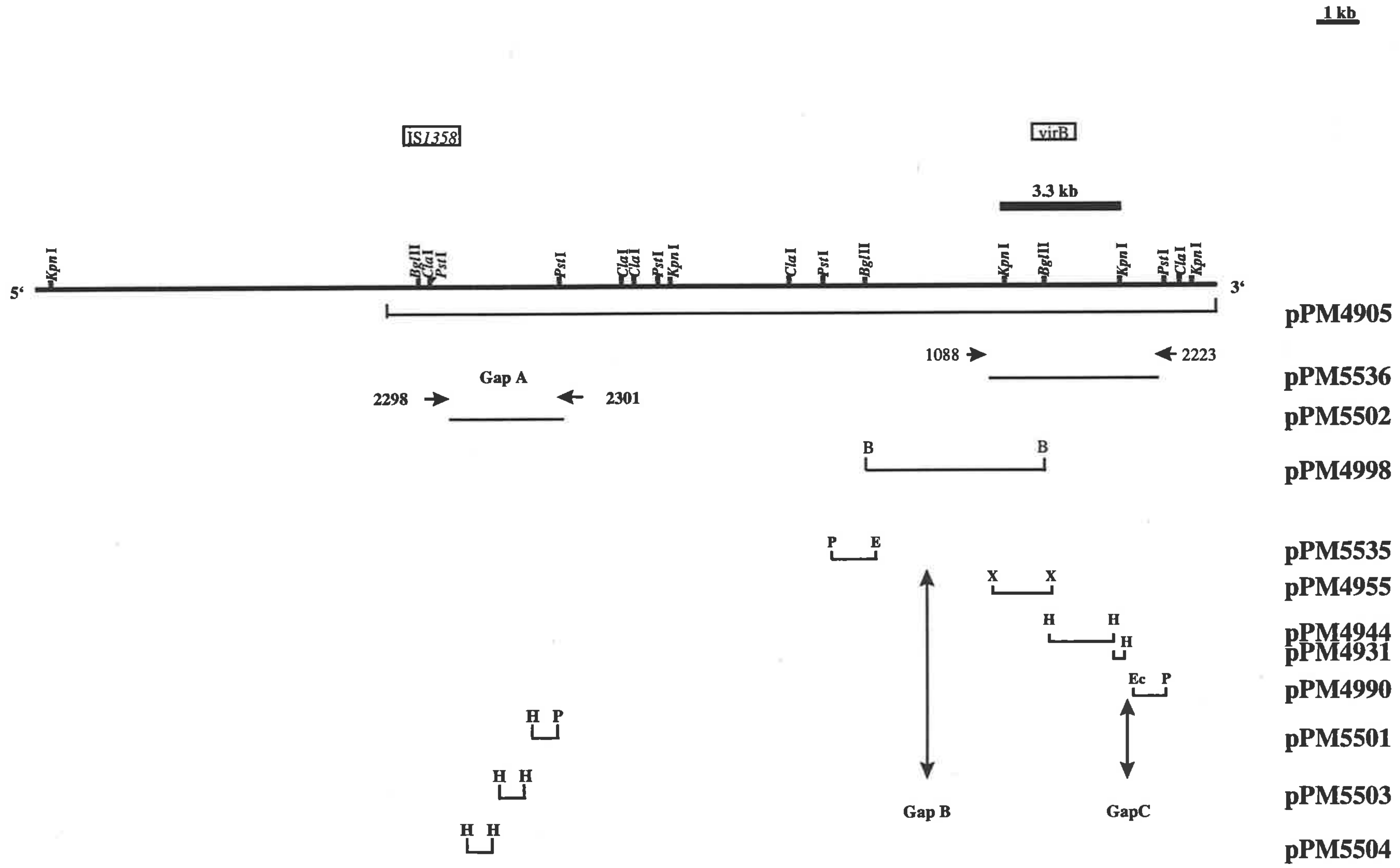
Southern hybridisation was used to determine the fragments that were needed to be cloned to sequence the gap (B) between pPM4955 and pPM5535. DNA from chromosomal and cosmid (pPM4905) preparations were digested with *Bgl*II and probed with DIG-labelled pPM4955 and pPM5535. The Southern hybridisation data indicated that both pPM4955 and pPM5535 hybridised to a DNA fragment of 4.5 kb (Table 5.1). The cosmid pPM4905 was digested with *Bgl*II, electrophoresed on a 0.8% agarose gel and the DNA band of 4.5 kb was excised and ligated into *Bam*HI digested pBC-KS$^+$. The clone obtained designated pPM4998 (Fig. 5.4) was subsequently sequenced. The data obtained allowed generation of a larger contiguous sequence of the locus (Gap B; Fig. 5.4).

5.2.4 PCR generated subclones

PCR was also employed to obtain specific clones which were required to close gaps in the sequence. It was evident from the data generated that there was a small fragment (<100 bp) missing near the 3' end between pPM4931 and pPM4990 of the initial map (Gap

Figure 5.4: Schematic representation of specific subclones generated to completely sequence Gap A, Gap B and Gap C.

The cloned PCR products inserts of pPM5502 (#2298/#2301) and pPM5536 (#1088/#2223) were generated by using standard PCR conditions. Oligonucleotides used in the reactions are indicated by arrows. pPM5502 subclones are indicated on the map with the capital letters representing the restriction enzyme used to clone the fragment into pBC KS$^+$. pPM4998 contains the 4.5 kb *Bgl*II fragment from pPM4905. The bold line marked 3.3 kb represents the previously published *virBA* sequence (Norqvist and Wolf-Watz, 1993) (Accession number, L08012. Restriction sites are as follows; B, *Bgl*II; E, *Eco*RV; H, *Hinc*II; P, *Pst*I; X, *Xba*I. IS*1358* and *virB* are indicated for reference.


1 kb
IS1358
virB
3.3 kb
KpnI
BglII
ClaI
PstI
PstI
ClaI
ClaI
PstI
KpnI
ClaI
PstI
BglII
KpnI
BglII
KpnI
PstI
ClaI
KpnI
5'
3'
pPM4905
1088
2223
pPM5536
Gap A
2298
2301
pPM5502
B
B
pPM4998
P
E
pPM5535
X
X
pPM4955
H
H
pPM4944
H
pPM4931
Ec
P
pPM4990
H
P
pPM5501
H
H
pPM5503
H
H
pPM5504
Gap B
GapC

C; Fig. 5.2). Utilising the previously published sequence of *virB* (on a 3.3 kb *Kpn*I fragment) (Norqvist and Wolf-Watz, 1993) and sequence obtained from subclone pPM4990 (Fig. 5.2), PCR was performed using oligonucleotide #1088 (*virB*) and #2223 (pPM4990) (Fig. 5.4; Table 2.4). This resulted in a PCR product of 3.6 kb which was cloned into pGEM-T (pPM5536) and sequenced. The sequence data obtained from pPM5536 was used to complete the sequence between *virB* and the end of the DNA insert in pPM4905 (Fig.5.2). The sequence was also used to check the junctions between the subclones, pPM4955, pPM4944, pPM4931, and pPM4990 (Fig. 5.2; Fig. 5.4) ensuring the region was correctly sequenced and no small restriction enzyme fragment had been omitted.

PCR was also used to generate a clone that connected pPM4940 to pPM4993 (Gap A; Fig. 5.2). Southern hybridisation using DIG-labelled IS*1358* and numerous DIG-labelled subclones (Table 5.1) indicated that approximately 2.8 kb of DNA from this region had to be cloned and sequenced. Using oligonucleotide primers #2298 (IS*1358*), and #2301 (pPM4943) (Fig. 5.4; Table 2.4), a PCR was performed on DNA isolated from both the chromosome (85-3954-2) and pPM4905 which resulted in a product of approximately 2.8 kb (data not shown). The PCR product was cloned into pGEM-T (pPM5502) subcloned and sequenced (Fig. 5.4). The generation of this PCR product allowed closure of the gap (A) between IS*1358* and the rest of the sequenced region.

5.2.5 Cloning of the 5' end of the *wbh* region

To obtain sequence preceding IS*1358*, which was thought to contain the beginning of the *wbh* operon, DNA from the cosmid pPM5512 (Fig. 5.1) was digested with the restriction enzyme *Hind*III. *Hind*III was used as Southern hybridisations using IS*1358* as the DIG-labelled probe on *Hind*III-digested pPM5512 and 85-3954-2 chromosomal DNA had indicated the presence of two fragments one of 4.8kb and another of 1.8kb (data not shown). The 1.8 kb fragment was previously sequenced from pPM4905 and included in the completed region (Section 5.2.4), and hence it could be concluded that the 4.8 kb fragment was located in the unsequenced region towards the 5' end of the operon. The *Hin*dIII digested pPM5512 was electrophoresed on a 0.6% agarose gel and the DNA bands of around 4.8 kb were isolated and cloned into pBluescript KS$^+$. During this band isolation an additional fragment of 3.4 kb *Hin*dIII fragment (pPM5516) was also cloned (Fig. 5.5). The clones, pPM5515 (4.8 kb fragment) and pPM5516 were sequenced using -21M13 and M13 primers and with specific oligonucleotides (Table 2.4). Analysis of the sequence from pPM5516 revealed the beginning of the operon by the presence of sequences homologous to non-polysaccharide biosynthesis genes. However the two clones, pPM5515 and pPM5516 did not have overlapping sequences, indicating the presence of additional DNA between (Fig. 5.5). Analysis of the genes present on the plasmids pPM5515 and pPM5516 indicated that approximately 2 kb of sequence was missing, based on the size of homologous genes present. Oligonucleotide primers were designed to the 3' end of pPM5516 (# 2751) and the 5' end of pPM5515 in the complementary strand (#2752) to obtain the connecting DNA. A PCR product of 1.8 kb was generated (Fig. 5.5) which was subsequently cloned into pGEM-T (pPM5517) and sequenced.

5.2.6 Construction of a physical map of the *wbh* cluster of *V. anguillarum* O1

Figure 5.6 shows the physical restriction map of the putative *wbh* cluster of *V. anguillarum* O1 generated from the data obtained from subcloning, PCR, Southern hybridisation and DNA sequencing. Restriction sites were mapped with respect to known sites in the vector pPM2101 and from the sequence data obtained. The *wbh* region

Figure 5.5: Cloning of the 5‘ end of the *wbh* locus.

Schematic representation of the subclones generated from pPM5512. A *Hin*dIII digest was performed on DNA from cosmid pPM5512. pPM5515 and pPM5516 were obtained as *Hin*dIII-*Hin*dIII (H) subclones in pBluescript SK$^+$of pPM5517. To complete the sequence, PCR was performed to amplify the 1.8 kb fragment indicated using oligonucleotides #2751 and #2752 (marked with arrows). //, indicates an undefined distance. IS*1358* and *virB* are indicated for reference. Some *Hin*dIII restriction sites are marked for reference.

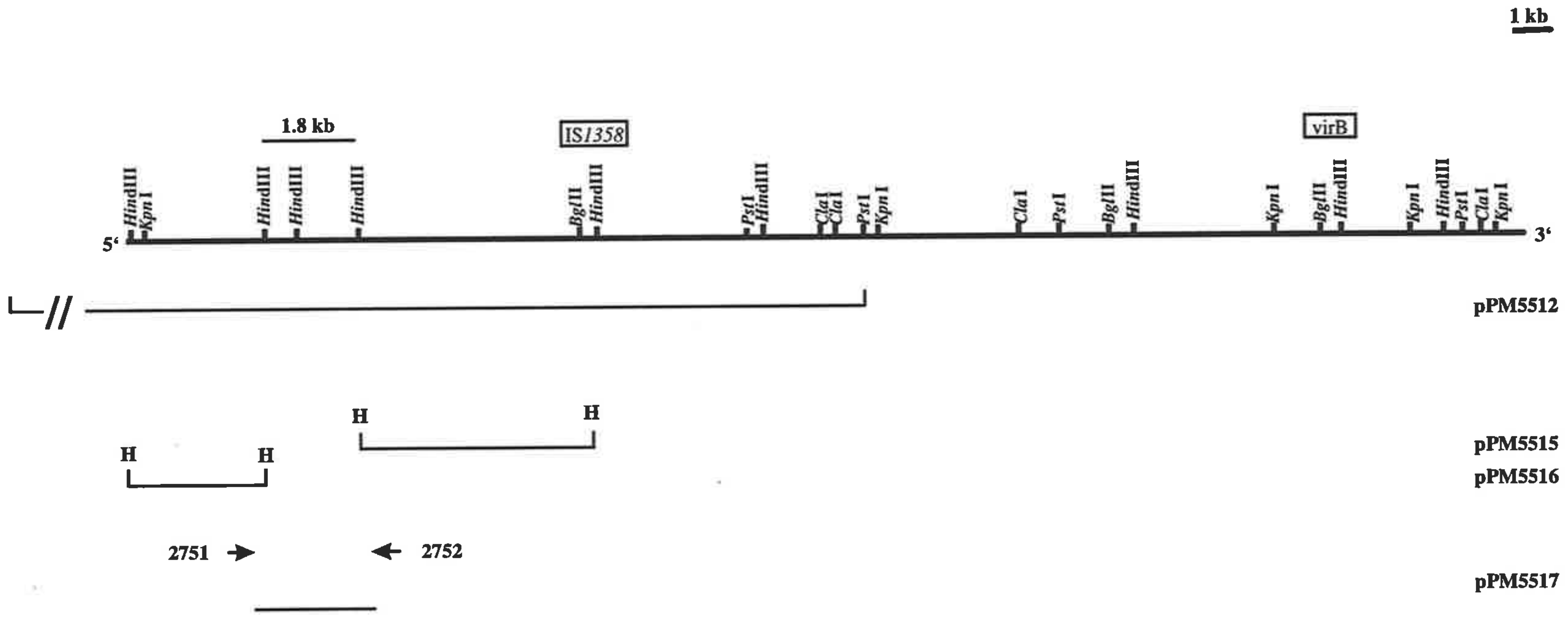
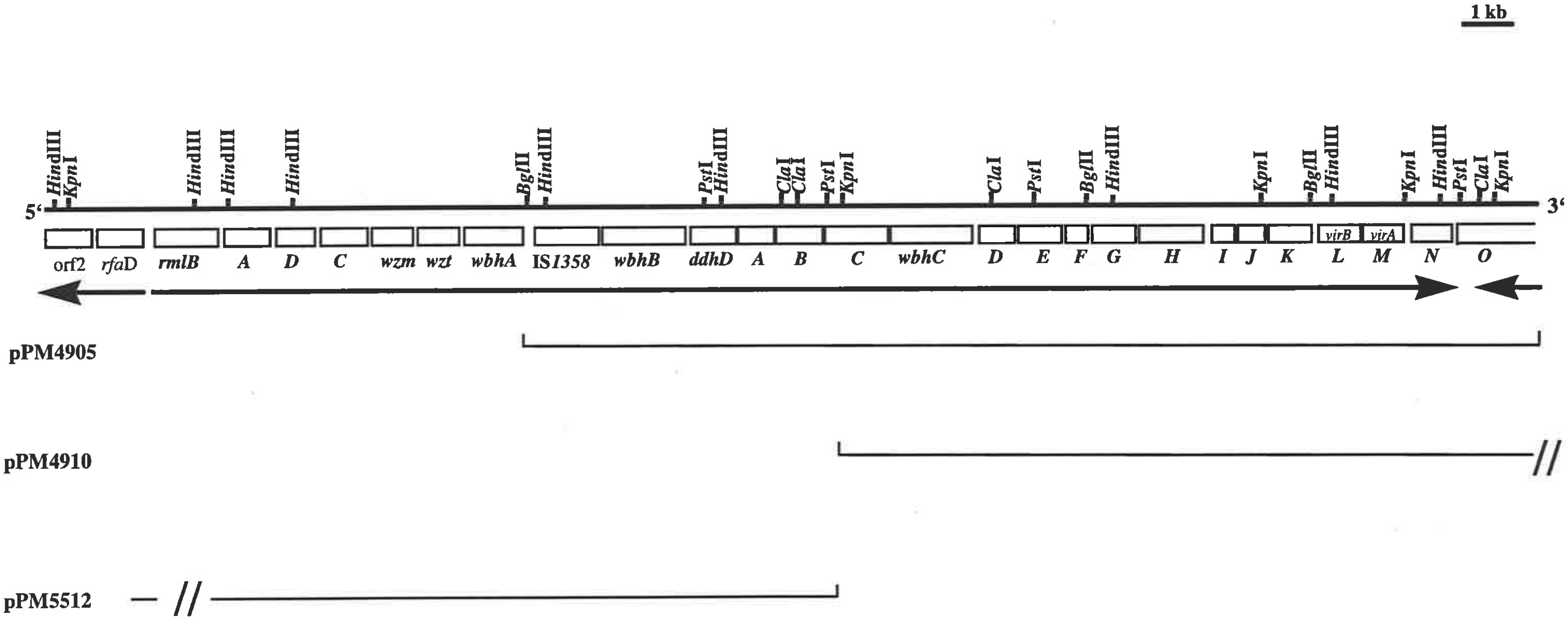

1 kb
1.8 kb
IS1358
virB
HindIII
Kpn I
HindIII
HindIII
HindIII
BglII
HindIII
PstI
HindIII
ClaI
ClaI
PstI
Kpn I
ClaI
PstI
BglII
HindIII
Kpn I
BglII
HindIII
Kpn I
HindIII
PstI
ClaI
Kpn I
5'
3'
pPM5512
H
H
H
H
pPM5515
pPM5516
2751
2752
pPM5517

Figure 5.6: Physical chromosomal map of the *wbh* locus of *V. anguillarum* O1- putative O-antigen biosynthesis operon.

Schematic representation of the entire 30,627 bp sequenced from cosmids pPM4905, pPM4910, and pPM5512. The ORFs have been designated names based on the nomenclature scheme for polysaccharide genes proposed by Reeves *et al.*, (1996). Analysis of the sequence indicates that there are 24 ORFs, with three potential ORFs transcribed in the opposite direction. Arrows on the diagram indicate the direction of transcription. Restriction sites were mapped by Southern hybridisation and from analysis of the sequence.


1 kb
HindIII
KpnI
HindIII
HindIII
HindIII
BglII
HindIII
PstI
HindIII
ClaI
ClaI
PstI
KpnI
ClaI
PstI
BglII
HindIII
KpnI
BglII
HindIII
KpnI
HindIII
PstI
ClaI
KpnI
5'
3'
orf2
rfaD
rmlB
A
D
C
wzm
wzt
wbhA
IS1358
wbhB
ddhD
A
B
C
wbhC
D
E
F
G
H
I
J
K
L
M
N
O
virB
virA
pPM4905
pPM4910
pPM5512

presented in Figure 5.6 is contained on the cosmid clones pPM4905, pPM4910 and pPM5512.

5.2.7 Identification of Open Reading Frames

Sequencing of the putative *wbh* region of *V. anguillarum* O1 shows that it consists of at least 30627 bp (Appendix 1). Evidence from this study and others (D. Milton, affiliation, personal communication, Umea University, Sweden) suggests that additional genes further downstream (RHS of Fig. 5.6) may also contribute to O-antigen biosynthesis (Section 6.2.10) and hence the contiguous DNA reported here may not represent the entire *wbh* region. The sequence of the ORFs are accessible through Genbank using the accession number AF025396.

Analysis of the nucleotide sequence of the 30627 bp revealed 24 ORFs, some of which demonstrate strong homology to polysaccharide biosynthesis genes from other bacterial species including: *Salmonella enterica*, *Yersinia enterocolitica* O3 and *Bordetella pertussis*. The location and several properties including %G+C content of each of the ORFs designated *rmlBADC*, *wzm*, *wzt*, *ddhDABC* and *wbhA-O* are summarised in Table 5.2. Significant similarities with other known proteins, revealed by comparison with sequence databases (eg BLASTX), are described below. Identity to DNA sequences (as determined by FASTA and BLASTN) are also described for those *wbh* genes which demonstrated significant homology to other genes with known function.

5.2.7.1 *rmlBADC*

The genes *rmlBADC* encode for the enzymes involved in dTDP-rhamnose biosynthesis (Jiang *et al.*, 1991; Reeves, 1993; Zhang *et al.*, 1993) and are present between nt 2920 and nt 6323. The *rml* genes are most homologous to the dTDP-rhamnose

Table 5.2 Summary of the ORFs in the *wbh* region.

ORF	G+C content[a]	Location in Sequence	Predicted Molecular Weight (kDa)[b]	#amino acids	Hydrophobicity index[b]
rfaD/gmhD	43.7	2606-1667	35	313	-0.32
rmlB	43.3	2928-3993	40	354	-0.16
rmlA	45.8	4016-4894	32	292	-0.37
rmlD	48.2	4891-5778	32	294	-0.16
rmlC	40.5	5781-6326	21	181	-0.53
wzm	35.7	6483-7287	31	268	+0.88
wzt	33.9	7304-8642	50	446	-0.19
wbhA	31.7	8670-9785	44	371	-0.70
wbhB	30.2	11141-12790	64	549	+0.46
ddhD	33.9	12930-13901	37	324	-0.15
ddhA	38.3	13927-14700	32	264	-0.31
ddhB	41.1	14703-15785	41	360	-0.50
ddhC	38.4	15785-17098	49	438	-0.11
wbhC	36.4	17115-18812	60	567	-0.08
wbhD	32.5	18903-19532	24	204	-0.27
wbhE	39	19522-20631	40	365	-0.05
wbhF	19.4	20689-21075	15	130	+0.97
wbhG	32.3	21072-22008	36	312	+0.07
wbhH	32.4	22062-23408	51	447	-0.02
wbhI	36.3	23622-24020	15	133	-0.38
wbhJ	36.6	24038-24493	16	153	+0.16
wbhK	37.8	24490-25605	41	369	-0.06
wbhL	31.2	25700-26788	42	363	-0.38
wbhM	30.9	26798-27712	36	303	-0.33
wbhN	37.5	27880-28763	34	294	-0.01

a. Percent guanine plus cytosine (G+C) of coding region.

b. According to Kyte and Doolittle (1982), as implemented in PROSIS.

biosynthesis genes from *Actinobacillus actinomycemcomitans* (Nakano *et al.*, 1998) although there is also similarity to the dTDP-rhamnose biosynthesis genes of *Shigella flexneri* (Macpherson *et al.*, 1994). These genes are discussed further in section 6.2. Immediately upstream of the start of *rmlD* (ie the beginning of the *wbh* region in *V. anguillarum* O1 is a putative JUMPstart sequence with the sequence 5'TAGCCGCTTGTAGGAGTTAGACCTAAGGGCGTAGCATA 3'(Hobbs and Reeves, 1994; Bailey *et al.*, 1997; Marold~~o~~a and Valvano, 1998) (Appendix 1).

5.2.7.2 *wzm* and *wzt*

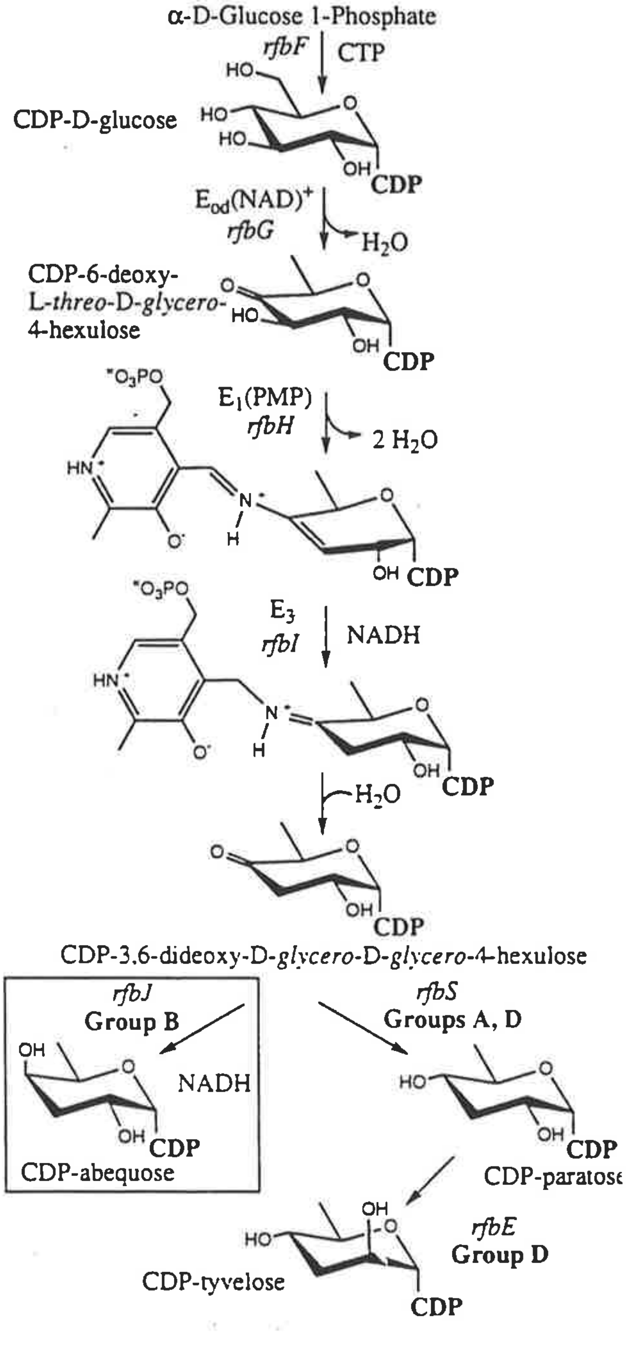
Following the dTDP-rhamnose biosynthesis genes from nt 6483- nt 7287 and nt 7304- nt 8642 are two genes designated *wzm* and *wzt* which are thought to be involved in homopolymer O-antigen export. Wzm and Wzt show significant homology to the ABC-transport family. Wzm is homologous to the *rfbA/wzm* gene product from *Klebsiella pneumoniae* O1 which is the integral membrane protein component. Wzt is the second component of the system and demonstrates homology to the hydrophilic protein that acts as an energiser for the transport system as it has an ATP binding domain. Wzm and Wzt are discussed further in Chapter 6 (see section 6.2).

5.2.7.3 The CDP-3,6-dideoxyhexose pathway

After IS*1358*, between nt 12930 to nt 17098, are the genes which demonstrate homology to the first four genes involved in the CDP-dideoxyhexose biosynthetic pathway, *ddhDAB,C* which synthesise the sugars abequose, paratose and tyvelose. Figure 5.7 shows the biosynthetic pathway in which these genes are involved in *S. enterica* (Jiang *et al.*, 1991). The first gene, $ddhD_{Va}$ was found to be homologous to AscD/$DdhD_{Ye}$, a CDP-6-

Figure 5.7 : The CDP-3,6-dideoxyhexoase biosynthesis pathway.

Biosynthetic pathway for abequose, paratos and tyvelose which indicates at which points the various *Salmonella enterica* gene products act and the intermediates produced by these enzymes. The old nomenclature is inidated however the genes correspond to the following: *rfbF* = *ddhA*, *rfbG* = *ddhB*, *rfbH* = *ddhC*, *rfbI* = *ddhD*, *rfbJ* = *abe*, *rfbS* = *prt*, *rfbE* = *tyv*. (The figure was reproduced from Reeves et al., 1993).


α-D-Glucose 1-Phosphate
rfbF
CTP
CDP-D-glucose
HO
HO
HO
O
OH
CDP
E$_{od}$(NAD)$^+$
rfbG
H$_2$O
CDP-6-deoxy-L-threo-D-glycero-4-hexulose
O
HO
OH
CDP
$^-$O$_3$PO
HN$^+$
O$^-$
N$^+$
H
E$_1$(PMP)
rfbH
2 H$_2$O
OH
CDP
E$_3$
rfbI
NADH
H$_2$O
CDP-3.6-dideoxy-D-glycero-D-glycero-4-hexulose
rfbJ
Group B
NADH
CDP-abequose
rfbS
Groups A, D
CDP-paratose
rfbE
Group D
CDP-tyvelose

deoxy-delta-3,4-glucoseen reductase from *Yersinia pseudotuberculosis* (Thorson *et al.*, 1994). DdhD also showed significant homology to RfbI/DdhD$_{St}$, which has the same function as AscD/DdhD$_{Yp}$, from *S. enterica* sv Typhimurium (Fig. 5.8; Table 5.3). Identity at the DNA level was also significant with *ddhD*$_{Va}$ demonstrating 60% identity to *ascD/ddhD*$_{Yp}$. The identity between *ddhD*$_{Va}$ and *rfbI/ddhD*$_{St}$ was considerably less (<40%). This suggests that *ddhD*$_{Va}$ is more closely related to *ascD/ddhD*$_{Yp}$ than to *rfbI/ddhD*$_{St}$.

DdhA demonstrates most similarity to the equivalent protein RfbF/DdhA in *S. enterica sv* Typhimurium (79.4% identity at the amino acid level, 71% identity at the DNA level) and to AscA/DdhA$_{Yp}$ from *Y. pseudotuberculosis* (76.3% identity at the amino acid level, 69% identity at the DNA level) (Thornson *et al.*, 1994) (Fig. 5.9; Table 5.4) and is the first enzyme in the biosynthetic pathway. DdhA converts glucose-1-phosphate to CDP-D-glucose (Fig. 5.7).

The second enzyme in the pathway is encoded by *ddhB*. DdhB converts CDP-D-glucose to CDP-4-keto-6-deoxy-D-glucose. The DdhB protein from *V. anguillarum* O1 shows most similarity to an equivalent protein in *Y. pseudotuberculosis* AscB/DdhB$_{Yp}$ (71.1% identity at the amino acid level, 66% identity at the DNA level) (Thornson *et al.*, 1994). Homology at a significant level (70.1% identity at the amino acid level, 67% identity at the DNA level) was also observed to the RfbG/DdhB$_{St}$ protein of *S. typhimurium* (Fig. 5.10; Table 5.5). DdhB is an enzyme that requires NAD$^+$ to function. The NAD$^+$ binding domain (GXXGXX) was found in the amino terminal end of the protein sequence which is consistent with its proposed function (Thornson *et al.*, 1994).

The last gene, *ddhC* is most homologous to the equivalent gene product in *Y. pseudotuberculosis* AscC/DdhC$_{Yp}$ (82.2% identity at the amino acid level, 73% identity at the DNA level). DdhC$_{Va}$ also showed similarity to RfbH/DdhC from *S. typhimurium*

Figure 5.8: Alignment of DdhD with homologous proteins

```
VaDdhD    MEFKVIVKPSGVEYQSG--RNILDDAFASSISLEHSCKTGDCGVCCAEVISGLVENEN--   56
YpAscD    MSLNVKLHPSGIIFTSDGTSTILDAALDSNIHIEYSCKDGTCGSCKAILISGEVDSAENT   60
StRfbI    MSHIIKIFPSNIEFSGREDESILDAALSAGIHLEHSCKAGDCGICESDLLAGEVVDSKG-   59
          *   . . **.. .       *** *  . * .* *** * ** * . ...* *

VaDdhD    ---GELVTQGHILTCQSKAKSDVVLKANYYPELVDIKQQTISCKVASFEFVTKDIVSIRF  113
YpAscD    FLTEEDVAKGAILTCCSKAKSDIELDVNYYPELSHIQKKTYPCKLDSIEFIGEDIAILSL  120
StRfbI    ---NIFGQGDKILTCCCKPKTALELNAHFFPELAGQTKKIVPCKVNSAVLVSGDVMTLKL  116
                     **** .* *. . *  ...***    ..   **. *   .  *.  .

VaDdhD    RFSPKTIFNYLPGQYVDLSFRGVKRSYSIANAKSKSNELELHIRKVPNGFMSELLFEHLK  173
YpAscD    RLPPTAKIQYLAGQYIDLIINGQRRSYSIANAPGGNGNIELHVRKVVNGVFSNIIFNELK  180
StRfbI    RTPPTAKIGFLPGQYINLHYKGVTRSYSIANSDESNG-IELHVRNVPNGQMSSLIFGELQ  175
          *  * .   .* ***. *   *  *******.    . .***.* * **  * ..*  *.

VaDdhD    ENQLMRIEGPKGTFFVRDNIKPLIFIATGTGIAPIKAIVEELIAKEDKRNVYIYWGMRYK  233
YpAscD    LQQLLRIEGPQGTFFVREDNLPIVFLAGGTGFAPVKSMVEALINKNDQRQVHIYWGMPAG  240
StRfbI    ENTLMRIEGPCGTFFIRESDRPIIFLAGGTGFAPVKSMVEHLIQGKCRREIYIYWGMQYS  235
           . *.***** ****.*.   *..*.* *** **.*..** **    .* . *****

VaDdhD    NEIYCDELSLLAAENQNIFFNLVLSREFEVSPDYKKGYVQDAVIRDFNSLKDIEVYACGS  293
YpAscD    HNFYSDIANEWAIKHPNIHYVPVVSGD-DSTWTGATGFVHQAVLEDIPDLSLFNVYACGS  299
StRfbI    KDFYSALPQQWSEQHDNVHYIPVVSGD-DAEWGGRKGFVHHAVMDDFDSLEFFDIYACGS  294
             *.      .  . *. .  *.* . .       *.*. **. *   *    .*****

VaDdhD    SKMIECAKALLLQHQLPNDAFFSDAFTPAK------                         323
YpAscD    LAMITAARNDFINHGLAENKFFSDAFVPSK------                         329
StRfbI    PVMIDASKKDFMMKNLSVEHFYSDAFTASNNIEDNL                         330
            ** ...   . . *    *.****  .
```

Alignment of the amino acid sequences of DdhD (VaDdhD), *Y. pseudotuberculosis* AscD (YpAscD) (spP37911), and *S. typhimurium* RfbI (StRfbI) (spP26395), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: sp, SWISS PROT.

Table 5.3 Homology Table for DdhD

	%identity[a]		
	VaDdhD	**StRfbI**	**YpAscD**
VaDdhD	100	45.2 (330)	42.3 (326)
StRfbI		100 (257)	50.8 (329)
YpAscD			100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: sp, SWISS PROT.

VaDdhD	*V. anguillarum* DdhD
YpAscD	*Y. pseudotuberculosis* AscD (spP37911)
StRfbI	*S. typhimurium* RfbI (spP26395)

Figure 5.9: Alignment of DdhA with homologous proteins

```
VaDdhA     ----MKVVILAGGLGTRLSEETSVKPKPMVEIGG-PILWHIMKQYSAHGINDFIICCGYK   56
StRfbF     ----MKAVILAGGLGTRLSEETIVKPKPMVEIGGKPILWHIMKMYSVHGIKDFIICCGYK   56
YeDdhA     MECKVKAVILAGGLGTRLSEETVIKPKPMVEIGGMPILWHIMKLYSSYGINDFIICCGYK   60
YpDdhA     MEIQVKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKLYSSYGINDFVICCGYK   60
YpAscA     ----MKAVILAGGLGTRLSEETVVKPKPMVEIGGKPILWHIMKLYSSYGINDFVICCGYK   56
               .* *************** .********** ******** **  ** **.******

VaDdhA     GYIIKEYFANYFLHMSDVTFDMKENKMEVHHKRAEPWTVTLVDTGDNSMTGGRLARVADY  116
StRfbF     GYVIKEYFANYFLHMSDVTFHMAENRMEVHHKRVEPWNVTLVDTGDSSMTGGRLKRVAEY  116
YeDdhA     GYVIKEYFANYFMHMSDITFSMRDNEMKVHQKRVEPWNVTLVDTGEHSMTGGRLKRVKDY  120
YpDdhA     GYVIKEYFANYFMHMSDITFCMRDNEMIVHQKRVEPWNVTLVDTGEDSMTGGRLRRVKDY  120
YpAscA     GYVIKEYFANYFMHMSDITFCMRDNEMVVHQKRVEPWNVTLVDTGEDSMTGGRLRRVKDY  116
           **.*********.****.** * .* * **.** ***.*******. ******* ** .*

VaDdhA     VKDEEAFCFTYGDGVSDIDITKSIEFHQAHGKQATLTATFPPGRFGALDITSGKVDNFKE  176
StRfbF     VKDDEAFLFTYGDGVADLDIKATIDFHKAHGKKATLTATFPPGRFGALDIRAGQVRSFQE  176
YeDdhA     VKDDDAFCFTYGDGVSDINISELINFHKSHGKKATLTATYPPGRFGALDIENKQVRSFKE  180
YpDdhA     VKDDEAFCFTYGDGVSDVNIAELIEFHKSHGKQATLTATYPPGRFGALDIKDKQVRSFKE  180
YpAscA     VKDDEAFCFTYGDGVSDVNIAELIAFHKSHGKQATLTATYPPGRFGALDIKDKQVRSFKE  176
           ***..** *******.*. *   * **..***.******.**********   .*  *.*

VaDdhA     KPRGDGAMINGGFFVLSPRVLQLIDSDSCIWEQYPLNRLADDGELMAYEHNGFWQPMDTL  236
StRfbF     KPKGDGAMINGGFFVLNPSVIDLIDNDATTWEQEPLMTLAQQGELMAFEHPGFWQPMDTL  236
YeDdhA     KPKGDGALINGGYFVLSPKVIDLIDGDKSIWEQEPLMILADNGELMAYEHTGFWQPMDTL  240
YpDdhA     KPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMTLAAQGELMAFEHAGFWQPMDTL  240
YpAscA     KPKGDGALINGGYFVLSPKVIDLIDGDKSTWEQEPLMTLAAQGELMAFEHAGFWQPMDTL  236
           **.****.****.*** * *. *** * . *** **  **  *****.** *********

VaDdhA     RDKLYLDELWQAGKAPWKIWE                                         257
StRfbF     RDKVYLEGLWEKGKAPWKTWE                                         257
YeDdhA     RDKIYLQQLWEEGQAPWKVWE                                         261
YpDdhA     RDKIYLHELWEEGRAPWKVWE                                         261
YpAscA     RDKIYLHELWEEGRAPWKVWE                                         257
           ***.**  **. *.**** **
```

Alignment of the amino acid sequences of DdhA (VaDdhA), *S. typhimurium* RfbF (StRfbF) (spP26395), *Y. enterocolitica* DdhA (YeDdhA) (gbU46859), *Y. pseudotuberculosis* DdhA (YpDdhA) (pirC47070), *Y. pseudotuberculosis* AscA (YpAscA) (gbL27130) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.4 Homology Table for DdhA

	%identity[a]				
	VaDdhA	**StRfbF**	**YeDdhA**	**YpDdhA**	**YpAscA**
VaDdhA	100	79.4 (257)	76.3 (257)	76.3 (257)	76.3 (257)
StRfbF		100 (257)	79 (257)	80.5 (257)	80.9 (257)
YeDdhA			100	90.4 (261)	90.4 (257)
YpDdhA				100	98.8 (257)
YpAscA					100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaDdhA	*V. anguillarum* DdhA
StRfbF	*S. typhimurium* RfbF (spP26395)
DdhAYe	*Y. enterocolitica* DdhA (gbU46859)
DdhAYp	*Y. pseudotuberculosis* DdhA (pirC47070)
YpAscA	*Y. pseudotuberculosis* (gbL27130)

Figure 5.10: Alignment of DdhB with homologous proteins

```
VaDdhB      -MNPTFWHGKKVFITGHTGFKGGWLSLWLQEMGAIVKGYSLPAPTTPSLFEQGKVWAGMR 59
YpYepA      MINNSFWQGKRVFVTGHTGFKGGWLSLWLQTMGATVKGYSLTAPTVPSLFETARVADGMQ 60
YpAscB      MINNSFWQGKRVFVTGHTGFKGGWLSLWLQTMGATVKGYSLTAPTVPSLFETARVADGMQ 60
YpDdhB      MINNSFWQGKRVFVTGHTGFKGGWLSLWLQTMGATVKGYSLAPPTVPSLFETARVADGMQ 60
StrfbG      MIDKNFWQGKRVFVTGHTGFKGSWLSLWLTEMGAIVKGYALDAPTVPSLFEIVRLNDLME 60
YeDdhB      MINEKFWQDKRVFITGHTGFKGSWLSLWLQNMGAKVKGYALDAPTKPSLFEAANVADGMI 60
             .   **. *.**.******** ******  *** ****.*  ** *****   .   *

vaDdhB      TEEGDIRDFTHMRQSMYEFKPEIVFHMAAQPLVRLSYHEPIETYSTNVMGTVYLLEAVKQ 119
YpYepA      SEIGDIRDQNKLLESIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRH 120
YpAscB      SEIGDIRDQNKLLESIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRH 120
YpDdhB      SEIGDIRDQNKLLEAIREFQPEIVFHMAAQPLVRLSYSEPVETYSTNVMGTVYLLEAIRH 120
StRfbG      SHIGDIRDFEKLRNSIAEFKPEIVFHMAAQPLVRLSYEQPIETYSTNVMGTVHLLETVKQ 120
YeDdhB      SEVGDIRNYAQLLKSINDFKPEIVFHMAAQSLVRLSYLEPVETYSTNVMGTVYLLEALRH 120
            .  ****   ..  .. .*.********** ****** .*.*********** ***....

VaDdhB      VGGVKAVVNITSDKCYENREWVWGYREDEPMGGYDPYSNSKGCAELVASSYRQSFFNKEN 179
YpYepA      VGGVKAVVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPAN 180
YpAscB      VGGVKAVVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPAN 180
YpDdhB      VGGVKAVVNITSDKCYDNKEWIWGYRENEAMGGYDPYSNSKGCAELVTSSYRNSFFNPAN 180
StRfbG      VGNIKAVVNITSDKCYDNREWVWGYRENEPMGGYDPYSNSKGCAELVASAFRNSFFNPAN 180
YeDdhB      VGGAKAIVNVTSDKCYENKEWVWGYREDESMGGYDPYSNSKGCAELVTSSYRNSFFNPVK 180
            **. **.**.******.*.**.***** * *****************.*..*.****

VaDdhB      YHQHRCALASVRAGNVIGGGDWAEDRLIPDMLKAFSNNQTVEIRS-PHAIRPWQHVLEPL 238
YpYepA      YGQHGTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRN-PHAIRPWQHVLEPL 239
YpAscB      YGQHGTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRN-PHAIRPWQHVLEPL 239
YpDdhB      YGQHGTAVATVRAGNVIGGGDWALDRIVPDILRAFEQSQPVIIRN-PHAIRPWQHVLEPL 239
StRfbG      YEQHGVGLASVRAGNVIGGGDWAKDRLIPDILRSFENNQQVIIRN-PYSIRPWQHVLEPL 239
YeDdhB      YEQHGTALATQRAGNVIGGGDWALDRVVPDMLRAFENSEPVIIRNNPYAIRPWQHVLEPL 240
            * **   .*. ************ **..**.*..* . . * **  * .***********

VaDdhB      SGYLTIAEKLYTVGPGFAEGWNFGPRDEDAKPVDWIVNRLTELWGGGARWSLCDGEHPHE 298
YpYepA      SGYLLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHE 299
YpAscB      SGYLLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQNVKYWGEGASWQLDGNAHPHE 299
YpDdhB      SGYLLLAQKLYTDGAEYAEGWNFGPNDADATPVKNIVEQMVKYWGEGASWQLDGNAHPHE 299
StRfbG      SGYIVVAQRLYTEGAKFSEGWNFGPRDEDAKTVEFIVDKMVTLWGDDASWLLDGENHPHE 299
YeDdhB      SGYLLLAQRLYIDGIKYAEGWNFGPDDTDSSQVQSIVDKMVKYWGGDACWQLDDHEHPHE 300
            ***. .*..**  *  ..******* * *.  *  ** .    **  * * *    ****

VaDdhB      AHYLKLDCSKAKMRLDWQPVWDVDTTLEKIVSWHKAWLAKQDMHQYTINEIKQYMTARAG 358
YpYepA      AHYLKLDCSKAKMQLGWHPRWNLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK-- 357
YpAscB      AHYLKLDCSKAKMQLGWHPRWNLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK-- 357
YpDdhB      AHYLKLDCSKAKMQLGWHPRWTLNTTLEYIVGWHKNWLSGTDMHEYSITEINNYMNTK-- 357
StRfbG      AHYLKLDCSKANMQLGWHPRWGLTETLGRIVKWHKAWIRGEDMLICSKREISDYMSATTR 359
YeDdhB      ANYLKLDCSKAKVLLGWQSRWNLDCTLEYIVAWHKAWLRGVNMHEYSIVEINKYMASK-- 358
            *.********* . * *.  * .  **  ** *** *.    *   .  **  ** .

VaDdhB      KE                                                           360
YpYepA      --                                                           357
YpAscB      --                                                           357
YpDdhB      --                                                           357
StRfbG      --                                                           359
YeDdhB      --                                                           358
```

Alignment of the amino acid sequences of DdhB (VaDdhB), *Y. pseudotuberculosis* YepA (YpYepA) (gbL33181), *Y. pseudotuberculosis* AscB (YpAscB) (spS72887), *Y. pseudotuberculosis* DdhB (YpDdhB) (gbD47070), *S. typhimurium* RfbG (StRfbG) (spP26395), *Y. enterocolitica* DdhB (YeDdhB) (gbU46859), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. The shaded residues represent the NAD+ binding site required for enzymatic function (Thornson *et al.*, 1994). Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.5: Homology table for DdhB

	%identity[a]					
	VaDdhB	**YpYepA**	**YpAscB**	**YpDdhB**	**StRfbG**	**YeDdhB**
VaDdhB	100	71.1 (356)	71.1 (356)	70.5 (356)	70.1 (355)	68.3 (357)
YpYepA		100	98.6 (357)	98.9 (357)	73.3 (356)	78.2 (358)
YpAscB			100	98.6 (357)	73 (356)	77.9 (356)
YpDdhB				100	72.8 (356)	77.4 (358)
StRfbG					100	69.5 (357)
YeDdhB						100

[a] Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaDdhB	*V. anguillarum* DdhB
YpYepA	*Y. pseudotuberculosis* YepA (gbL33181)
YpAscB	*Y. pseudotuberculosis* AscB (pirS72887)
DdhBYp	*Y. pseudotuberculosis* DdhB (gbD47070)
StRfbG	*S. typhimurium* RfbG (spP26395)
DdhBYe	*Y. enterocolitica* DdhB (gbU46859)

(80.8% identical at the amino acid level and 72% at the DNA level) (Fig. 5.11; Table 5.6). The final enzymatic step in the production of abequose, paratose and tyvelose is encoded by the genes *rfbJ* (*abe*), *rfbS* (*prt*) and *rfbE* (*tyv*) (Jiang *et al.*, 1991), however no DNA or amino acid sequence homologous to these genes could be detected in *V. anguillarum* O1. Interestingly, some of the ORFs following *ddhC,* including *wbhC* which directly follows, have homology to synthetase enzymes from *E. coli.* One of these may function as the enzyme in last step of the pathway producing an alternative (3,6 dideoxy) sugar.

5.2.7.4 *wbhC, wbhD* and *wbhE*

wbhC and *wbhE* show homology to synthetase genes from *E. coli* and other bacteria. *wbhC* (nt 17115- nt 18812) has similarity to *ilvG* from *E. coli* (33.1% identity at the amino acid level) (Fig. 5.12). *ilvG* encodes the enzyme acetolactate synthetase which is involved in the biosynthesis of the amino acids valine, isoleucine and leucine (Haughn *et al.*, 1985; Coppala *et al.*, 1991). Based on the homologues found, the role of *wbhC* in polysaccharide biosynthesis in the *V. anguillarum* O1 *wbh* operon is unclear.

The next gene in the operon is *wbhD*. The gene product of *wbhD* does not show any significant similarity to any sequences currently in the databases. The initiation codon of the next ORF *wbhE* (nt 19522- nt 20631) overlaps with the termination codon of the preceding ORF (*wbhD)* which indicates that *wbhD* and *wbhE* are translationally coupled. The *wbhE* gene product is most homologous to the perosamine synthetase of *Methanbacterium thermoautotrophicum.* WbhE also demonstrates homology to probable perosamine synthetase genes *E. coli* and *V. cholerae* O1 (*wbeE*/*rfbE*), LpsC, which is involved in the production of S-LPS in *Caulobacter crescentus,* and SpsC involved in spore coat polysaccharide synthesis in *Synechocystis sp* (Fig. 5.13; Table 5.7). At the DNA level *wbhE* shows 55% identity to a *rfbE* gene (perosamine synthetase) from *E. coli* O157.

Figure 5.11: Alignment of DdhC with homologous proteins

```
VaDdhC      MSKEQLRSQIAELVAQYAVIEYAPKMFIGGESVVPPSGKVLGAKELQLMVDASLDGWLTT   60
YpDdhC      MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTKELQLMVEASLDGWLTT   60
YpAscC      MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTKELQLMVEASLDGWLTT   60
StRfbH      MTANNLREQISQLVAQYANEALSPKPFVAGTSVVPPSGKVIGAKELQLMVEASLDGWLTT   60
Amorf6      MVDSARSDLLTQVRDYHDAQSPGPAPFIPGQTPILSSGAVLDAADREALVAAALDLRIAS   60
            *        ....          *  *  * . .  ** *. . . .* *.**  ...

VaDdhC      GRFNDAFEKRLGEYLGVPFVLTTTSGSSANLLALTTLTSPKLGDRQLKPGDEVITVAAGF  120
YpDdhC      GRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKLGVRALKPGDEVITVAAGF  120
YpAscC      GRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKLGVRALKPGDEVITVAAGF  120
StRfbH      GRFNDAFEKKLGEFIGVPHVLTTTSGSSANLLALTALTSPKLGERALKPGDEVITVAAGF  120
Amorf6      GTIATKFESSFARKLKRRKAVLTNSGSSANLLALSTLTSPTLGDEQLKPGDEVVTVAAGF  120
            *     **      .     . *.**********..**** **   *******.******

VaDdhC      PTTVNPTIQNGLIPVFVDVDIPTYQIKPEMIEAAVSEKTKAIMVAHTLGNTFDLTEARRV  180
YpDdhC      PTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTKAIMIAHTLGNLFDLAEVRRV  180
YpAscC      PTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTKAIMIAHTLGNLFDLAEVRRV  180
StRfbH      PTTVNPAIQNGLIPVFVDVDIPTYNIDASLIEAAVTEKSKAIMIAHTLGNAFNLSEVRRI  180
Amorf6      PTTVNPILQNGMIPVFVDIELGTYNTTAERVEAAISPRTRAIMIAHTLGNPFPVAEIAEI  180
            ****** .***.******... **.     .***.. ...***.****** * ..*   .

VaDdhC      ADKYNLWLIEDCCDALGSTYNGKMVGTIGDIATVSFYPAHHITMGEGGAVFTKDKELRKL  240
YpDdhC      ADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYPAHHITMGEGGAVFTQSAELKSI  240
YpAscC      ADKYNLWLIEDCCDALGSTYDGKMAGTFGDIGTVSFYPAHHITMGEGGAVFTQSAELKSI  240
StRfbH      ADKYNLWLIEDCCDALGTTYEGQMVGTFGDIGTVSFYPAHHITMGEGGAVFTKSGELKKI  240
Amorf6      AQRHGCHLIEDNCDAVGSTYQGQLTGTFGRFSTVSFYPAHHLTMGEGGMLLSSDIKLAKL  240
            * . .  **** ***.*.** *.. ** *   *********.****** . .    *  .

VaDdhC      IESFRDWGRDCYCAPGCDNTCGKRFDQQLGSLPQGYDHKYTYSHLGYNLKITDMQAACGL  300
YpDdhC      IESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGYDHKYTYSHLGYNLKITDMQAACGL  300
YpAscC      IESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGYDHKYTYSHLGYNLKITDMQAACGL  300
StRfbH      IESFRDWGRDCYCAPGCDNTCGKRFGQQLGSLPQGYDHKYTYSHLGYNLKITDMQAACGL  300
Amorf6      AKSFRDWGRDCWCEPGEDDRCLKRFDLHLGTLPEAYDHKYVFSEVGYNLKSTDIHAALGL  300
              ********* * ** *  * ***  .**.**  ***** .* .***** **..** **

VaDdhC      AQMDRVEEFVQARKENFAYLKNGLASCEEFIILPEATENSEPSWFGFPITIKDDSGISRV  360
YpDdhC      AQLERIEEFVEKRKANFKYLKDALQSCADFLELPEATENSDPSWFGFPITLKEDSGVSRI  360
YpAscC      AQLEPIEEFVEKRKANFKYLKDALQSCADFIELPEATENSDPSWFGFPITLKEDSGVSRI  360
StRfbH      AQLERVEEFVEQRKANFSYLKQGLQSCTEFLELPEATEKSDPSWFGFPITLKETSGVNRV  360
Amorf6      SQLAKLPDFVDARKRNWRQLREGLDG-LPGLILPEATPGSDPSWFGFIITVKPDAGFDVP  360
            .*.  . .**  ** *   *.  *      . *****  *.****** **.*  .*

VaDdhC      DLLKFMDQHKIGTRLLFAGNLTRQPYFEHVKYRVVGELTNTDLIMNNTFWIGVYPGLTLA  420
YpDdhC      DLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKYRVVGELTNTDRIMNQTFWIGIYPGLTHD  420
YpAscC      DLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKYRVVGELTNTDRIMNQTFWIGIYPGLTHD  420
StRfbH      ELVKFLDEAKIGTRLLFAGNLIRQPYFANVKYRVVGELTNTDRIMNQTFWIGIYPGLTTE  420
Amorf6      EVVEYLEARNIRTRRLFAGNLTRHPAYIGKPQRIAGPLTNSDIVTERTFWIGVYPGITTE  420
            ... ...   . ** ****** *.* .     *. * ***.* .   *****.***.*

VaDdhC      HLDFVLEKFEEFFGVNF                                            437
YpDdhC      HLDYVVSKFEEFFGLNF                                            437
YpAscC      HLDYVVSKFEEFFGLNF                                            437
StRfbH      HLDYVVSKFEEFFGLNF                                            437
Amorf6      MIEYVIGTLKEFVATH-                                            436
             ...*.    **   .
```

Alignment of the amino acid sequences of DdhC (VaDdhC), *Y. pseudotuberculosis* DdhC (YpDdhC) (gbE47070), *Y. pseudotuberculosis* AscC (YpAscC) (gbL33181), *S. typhimurium* RfbH (StRfbH) (spP26395), *and Amycolatopsis mediterranei* orf6 (Amorf6) (gbAF040570) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.6: Homology table for DdhC

	%identity[a]				
	VaDdhC	**YpDdhC**	**YpAscC**	**StRfbH**	**Am orf6**
VaDdhC	100	82.2 (437)	82.2 (437)	80.8 (437)	51.3 (431)
YpDdhC		100	99.5 (437)	87.4 (437)	50.2 (430)
YpAscC			100	87 (437)	50.2 (430)
StRfbH				100	50.2 (432)
Am orf6					100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaDdhB	*V. anguillarum* DdhA
YpDdhC	*Y. pseudotuberculosis* DdhC (gbE47070)
YpAscC	*Y. pseudotuberculosis* AscC (gbL33181)
StRfbH	*S. typhimurium* RfbH (spP26395)
Amorf6	*Amycolatopsis mediterranei* orf6 (gbAF040570)

Figure 5.12: Alignment of WbhC with IlvG

```
vawbhC      MKSSDAIAHLLSLNNVTVGFELIGGMITHLVDSINELGKTKLISLHHEQAAAFAAGGVAR   60
ecilvG      MNGAQWVVHALRAQGVNTVFGYPGGAIMPVYDALYDGGVEHLLCR-HEQGAAMAAIGYAR   60
            *  .  . * *  ..*.  *   ** *  . *.. . *  .*..  *** ** ** * **

vawbhC      ATNNEQVGLALGTSGPGATNLITGIADCWLDSYPCIFITGQVNTYELKDKRPIRQQGFQE  120
ecilvG      ATG--KTGVCIATSGPGATNLITGLADALLDSIPVVAITGQVSAPFIG------TDAFQE  112
            **.  . *... ************.**. *** * . ***** .  .          ***

vawbhC      LDIVSLVDSITKYSIQVKTVEQLLIEIQKAISIARSGRPGPVLIDIPMDLQRK--ELDIT  178
ecilvG      VDVLGLSLACTKHSFLVQSLEELPRIMAEAFDVACSGRPGPVLVDIPKDIQLASGDLEPW  172
            .*.. *  . ** *  *...*.*   .  *  .* ********.*** *.*    .*.

vawbhC      FDDIA-RLVVPSAEEEMNSGFFSVDNALKEAQKPLFIIGGGACAEVQFSAWQKKISSLGI  237
ecilvG      FTTVENEVTFPHAEVEQAR------QMLAKAQKPMLYVGGGVGMAQAVPALREFLAATKM  226
            *  .   .  * ** *         . *  ****.  .***        * .  ...  .

vawbhC      PHVSSLKGS-ERTSNYPEYLGMIGAYGTRAANYAVQNADIIIVLGSRLDIRQTGANVADF  296
ecilvG      PATCTLKGLGAVEADYPYYLGMLGMHGTKAANFAVQECDLLIAVGARFDDRVTG-KLNTS  285
            *  ..***     . ** ****.*  **.***.*** .*..* .*.* * * **  .

vawbhC      ARNAKKIIQIDVDKGQIDNRITTHLNIVSKCNSYFEHFLSEDYIINCSLWREKLKETFRK  356
ecilvG      APHAS-VIHMDIDPAEMNKLRQAHVALQGDLNALLPALQQPLNQYDWQQHCAQLRDEHSW  344
            * .*  .*..*.*  ..     .*. .    *.                   .*..

vawbhC      KFIDEYEAYRFSPFKIMQTLSEKFSGKIVHYIPDVGNHQMWLAHSLFIEPQQKIHHSGGL  416
ecilvG      RYDHPGDAI-YAPLLLKQLSDRKPADCVVT--TDVGQHQMWAAQHIAHTRPENFITSSGL  401
            ..    .*  ..*  . *    * .  .*    ***.**** *. .     .    * **

vawbhC      GAMGFSLPTAIGVRVVTGN-YVVSISGDGGFQLNIQELDVINRDKIPILIIILNNKSLGM  475
ecilvG      GTMGFGLPAAVGAQVARPNDTVVCISGDGSFMMNVQELGTVKRKQLPLKIVLLDNQRLGM  461
            *.*** **.*.* .*   *  **.***** * .*.***  . * ..*. *..* *. ***

vawbhC      VKNFQDMYFNGRNKPTYWGGYSCSFSQVGEAYGIESHLIKNDLEFSALVESYVKNARPLL  535
ecilvG      VRQWQQLFFQERYSETTLTD-NPDFLMLASAFGIHGQHITRKDQVEAALDTMLNSDGPYL  520
            *.. * ..*. *   *        *  .  *.**  . *    .  * ... .    * *

vawbhC      IEVSLEDVTVCKPRLVYGKSIDEQYPFDEQ                                565
ecilvG      LHVSIDELENVWPLVPPGASNSEMLEKLS-                                549
            . **....    * .  * *  *
```

Alignment of the amino acid sequences of WbhC (VaWbhC), and *E.coli* IlvG (EcIlvG) (spP00892) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: sp, SWISS PROT.

Figure 5.13: Alignment of WbhE with homologous proteins

```
VaWbhE      -MINEMILTAGPSITEKEISYVTDAVKNGWNNNWNNYLLKFEKSLADYVGVKHSLSTSSA   59
Mt          -----MIPILEPSIGNEELENVLEALKSGWISSRGEFITKFEEKFARYHGMGFGISTSNG   55
SspSpsC     ----------------MKLSVLTVAFSVSALFET--------TIFAEFCGTKYAVAVSSG   36
EcRfbE      ---MKYIPVYQPSLTGKEKEYVNECLDSTWISSKGNYIQKFENKFAEQNHVQYATTVSNG   57
VcRfbE      -----MIPVYEPSLDGNERKYLNDCIDSGWVSSRGKYIDRFETEFAEFLKVKHATTVSNG   55
CcLpsC      MSDLPRISVAAPRLDGNERDYVLECMDTTWISSVGRFIVEFEKAFADYCGVKHAIACNNG   60
                                                   .  .  *   .

VaWbhE      TGALHLSMLACGIGPGDEVIVPEISWVASASAVAYVGATPVFCDIDPVSWCLDIESAARL  119
Mt          TTALHLALKSVGIKKGDEVIVPSLTFAATANAVIYCNAKPVLVDSHPDYWCIDPSRIEEK  115
SspSpsC     TAALHLALLALNIGAGDEVIVPTLSFIATANAVTYTGAKPIFVDSEWETWNINPDLIEAA   96
EcRfbE      TVALHLALLALGISEGDEVIVPTLTYIASVNAIKYTGATPIFVDSDNETWQMSVSDIEQK  117
VcRfbE      TVALHLAMSALGITQGDEVIVPTFTYVASVNTIVQCGALPVFAEIEGESLQVSVEDVKRK  115
            * ****.. .*  *******  .  *.  ..   .* *.  .       .

VaWbhE      LTPKTKAILPVHIYGHPANMPAIMEFARANNILIIEDAAPSIGAEVDGKKTGSFGDAAAF  179
Mt          INKNTRAIIPVHLYGHPCDMGWIGDIADEHGLIIIEDAAEAHGAEYRGKKIGTFGDVSCF  175
SspSpsC     ITPRTKAIMPVHLYGHPAKMDKILDIAQRYHLAVIEDAAEAHGATYQGKTVGSLGDLGIF  156
EcRfbE      ITNKTKAIMCVHLYGHPCDMEQIVELAKSRNLFVIEDCAEAFGSKYKGKYVGTFGDISTF  177
VcRfbE      INKKTKAVMAVHIYGQACDIQSLRDLCDEHGLYLIEDCAEAIGTAVNGKKVGTFGDVSTF  175
CcLpsC      ITPRTKAIMPVHLYGQICDMDPILEVARRHNLLVIEDAAEAVGATYRGKKSGSLGDCATF  180
            ..  *.*.. **.**. . .  . . .    . .***.* .* .   **  *. **   *

VaWbhE      SFQGAKILSTGEGGMFVSNNDEIFNRVKSLNDHGRDPSQPFASVEVGYKYKMSNLQAAMG  239
Mt          SFYGNKIITTGEGGMCLTNDEELAERMAVLRDHGMRPEKRYWHEEVGFNYRMTNLQAAVG  235
SspSpsC     SFYGNKIVTTGEGGMIVTDDEELAQKIRILKDHGMSKKQRYWHPILGYNYRITNIQAALG  216
EcRfbE      SFFGNKTITTGEGGMVVTNDKTLYDRCLHFKGQGLAVHRQYWHDVIGYNYRMTNICAAIG  237
VcRfbE      SFFGNKTITSGEGGMVVSNSDIIIDKCLRLKNQGVVAGKRYWHDLVAYNYRMTNLCAAIG  235
CcLpsC      SFFGNKIITTGEGGMITTNDDDLAAKMRLLRGQGMDPNRRYWFPIVGFNYRMTNIQAAIG  240
            ** * * ...*****  .    .  .      .*    . .    . . *...*. **.*

VaWbhE      LAQIERVEELVNKKREINSIYQELLKDCT---AVKVTTELPDCKSIHWMTSVELLGFDYD  296
Mt          VAQLEKIDGFINRKREIAEIYNRGLAELHENEELMLHPEMLWARSVYWVYSIVLKKR---  292
SspSpsC     VAQMERINKIPEAKRRIAQLYEQELLQIQG---LTLPPRQPWAESVFWLYTILINQDKLE  273
EcRfbE      LAQLEQADDFISRKREIADIYKKNINSL-----VQVHKESKDVFHTYWMVSILTRTAEE-  291
VcRfbE      VAQLERVDKIIKAKRDIAEIYRSELAGLP----MQVHKESNGTFHSYWLTSIILDQEFEV  291
CcLpsC      LAQLERVDEHLAARERVVGWYEQKLARLGN--RVTKPHVALTGRHVFWMYTVRLGEG-LS  297
            .**.*.       .  .   *   .        .             *. ..

VaWbhE      -KRQRFMGKLRENLVDSRPVFSPLSSLPMFEPRVKN---PVALRIGQSAINLPSGHNLIL  352
Mt          -KNMDLIRYLNSCGVDTRPFFYPLNLMPPYKDDSS---YPVAENLSYGGLSLPSSVEIGD  348
SspSpsC     LNRDQLMSRLQEKGIETRPLFIPIHRQPIYNTHQS---LPVAESLSKNGLSLPSFVTLSN  330
EcRfbE      --REELRNHLADKLIETRPVFYPVHTMPMYSEKYQK--HPIAEDLGWRGINLPSFPSLSN  347
VcRfbE      -HRDGLMTFLENNDIESRPFFYPAHTLPMYEHLAEKTAFPLSNSYSHRGINLPSWPGLCD  348
CcLpsC      TTRDQVIKDLDALGIESRPVFHPMHIMPPYAHLATDD-LKIAEACGVDGLNLPTHAGLTE  356
                     *    ...** * *    * .          ..         **.   .

VaWbhE      EQLEHVATTIKKTRIAI                                             369
Mt          EAIEYVVDRVREFFE--                                             363
SspSpsC     ENLYQIIDSIKKVIS--                                             345
EcRfbE      EQVIYICESINEFYSDK                                             364
VcRfbE      DQVKEICNCIKNYFNCI                                             365
CcLpsC      ADIDRVIAALDQVLV--                                             371
              .  .   .
```

Alignment of the amino acid sequences of WbhE (VaWbhE), *Methanobacterium thermoautotrophicum* perosamine synthetase (Mt) (gbAE000818), *Synechocystis* sp SpsC (SspSpsC) (gbD90911), *E. coli* O157 RfbE (EcRfbE) (pirS83460), *V. cholerae* RfbE (VcRfbE) (pirS28471) and, *Caulobacter crescentus* (CcLpsC) (gbAF062345) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.7: Homology table for WbhE

	%identity					
	WbhEVa	**Mt**	**SpsCSsp**	**RfbEEc**	**RfbEVc**	**CcLpsC**
WbhEVa	100	39.5 (362)	38.2 (327)	38.4 (354)	36.8 (364)	38.1 (365)
Mt		100	50.2 (329)	46.3 (363)	41.5 (369)	45.6 (366)
SpsCSsp			100	48.0 (327)	41.2 (330)	45.6 (366)
RfbEEc				100	54.7 (364)	46.4 (362)
RfbEVc					100	43.1 (357)
CcLpsC						100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank.

VaWbhE	*V. anguillarum* WbhE
Mt	*Methanobacterium thermoautotrophicum,* (gbAE000818),
SspSpsC	*Synechocystis* sp SpsC (gbD90911),
EcRfbE	*E. coli* O157 RfbE (pirS83460)
VcRfbE	*V. cholerae* O1 RfbE (pirS28471)
CcLpsC	*Caulobacter crescentus* (gbAF062345)

The putative perosamine synthetase in V. *cholerae* O1 (*wbeE/rfbE*) is the last enzyme in the biosynthetic pathway of perosamine and is thought to convert GDP-4-keto-6-deoxymannose to GDP-4-amino-4,6 dideoxymannose (GDP-perosamine) (Fig. 5.14) (Stroeher *et al.*, 1995). The role of *wbhE* in *V. anguillarum* O1 is not known.

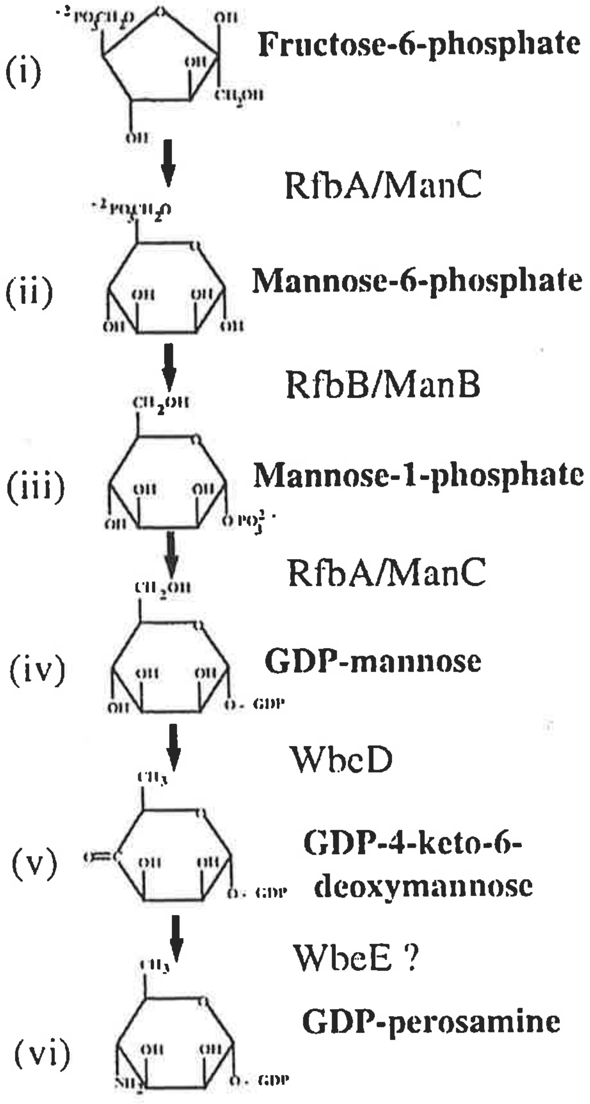
5.2.7.5 *wbhG*

wbhG lies between nt 21072 to nt 22008 (Appendix 1) and WbhG showed the best similarity to proteins in *Bacillus subtilis* (CshB), *E. coli* (o306), and *S. flexneri* (Bgt) (Fig. 5.15; Table 5.8). The function of these proteins is to transfer UDP-glucose to the bactoprenol carrier during polysaccharide synthesis and glucosylation of O-antigen (Mavris *et al.*, 1997). These related proteins also demonstrate structural and sequence similarity to a number of dolichol mannosyltransferase enzymes including Dpm1 of *Saccharomyces cereviceae*, a well characterised dolichol phosphate mannose synthetase which catalyses the formation of dolichol-phosphate mannose from dolichol-phosphate and GDP-mannose (Orlean *et al.*, 1988).

For correct functioning, proteins of this family (glycosyl transferases) require an amino terminal motif with the sequence DXSXD and DXD (Saxena *et al.*, 1995; Keenleyside and Whitfield, 1996). The motif sequence is bolded in Figure 5.15. Interestingly WbhG does not have the exact consensus sequence, suggesting it may not be functional. In addition, analysis of the sequence preceding *wbhG* indicates that it does not possess a potential RBS (Appendix 1). An attempt to show functionality of WbhG in a *S. flexneri bgt* Tn*phoA* mutant (RMA903) (Mavris *et al.*, 1997) proved to be unsuccessful (data not shown) which further indicates that this protein is either not functional or possibly that its substrate specificity is different to that described for *bgt* (Mavris *et al.*, 1997).

Figure 5.14: The pathway for the biosynthesis of perosamine in *V. cholerae*.

This figure describes the putative perosamine biosynthesis pathway in *V. cholerae* O1 which is based entirely on the homology to other enzymes that have been biochemically assayed. The pathway commences with fructose-1-phosphate which is converted to mannose-6-phosphate by ManC. This intermediate is converted to mannose-1-phosphate by ManB which is then made into GDP-mannose by ManC. GDP-mannose is converted to GDP-4-keto-6-deoxymannose by WbeD/Gmb, before the final step mediated by WbeE which produces GDP-perosamine. (Figure was reproduced from Stroeher *et al.*, 1998).



(i)
Fructose-6-phosphate
RfbA/ManC
(ii)
Mannose-6-phosphate
RfbB/ManB
(iii)
Mannose-1-phosphate
RfbA/ManC
(iv)
GDP-mannose
WbcD
(v)
GDP-4-keto-6-deoxymannose
WbeE ?
(vi)
GDP-perosamine

Figure 5.15: Alignment of WbhG to homologous proteins

```
VaWbhG      -----------------MKKKKISIVTPCFNEEDNVELLYEKVKEEFEKLDAYLYEHIFI   43
SspDpml     MVTNSPSRSDISVYNGDRQGPMYSLIVPIYNEEDNIPVLYERLKAVMD--QLASTELVLI   58
BsCsbB      -----------------MKQGLISIIIPSYNEGYNVKLIHESLKKEFKN-IHYDYEIFFI   42
Eco306      --------------------MKISLVVPVFNEEEAIPIFYKTVREFEEL-KSYEVEIVFI   39
SfBgt       --------------------MKISLVVPVFNEEEAIPVFYKTVREFQEL-KPYEVEIVFI   39
                                   *.. * .**   . .    ..         *   *

VaWbhG      DNNSVDKTVEKLRSIAKRDCNVKVILNSRNFGPVRSPHYGLLQGSGDATMLVVADLQDPP   103
SspDpml     NDGSGDRSLEMIRALHDQDKRVCYLSFARNFGHQVAVTAGLNFARGQAVIILDADLQDPP   118
BsCsbB      NDGSVDDTLQQIKDLAATCSRVKYISFSRNFGKEAAILAGFEHVQGEAVIVMDADLQHPT   102
Eco306      NDGSKDATESIINALAVSDPLVVPLSFTRNFGKEPALFAGLDHATGDAIIPIDVDLQDPI   99
SfBgt       NDGSKDATESIINALAVSDPLVVPLSFTRNFGKEPALFAGLDHASGDAVIPIDVDLQDPI   99
              .* * .   .  .      *  .  .****   .   *     * * . .  *** *

VaWbhG      ELIPEFIAKWEDGNDIVIGVK-SESDESPAMYMIRKVYYNFSNKLSESKLVKNYYGFGLY   162
SspDpml     ELVPQLVERWQAGYSVVYAQRVKRRQESWFKRLTAYGFYRLLQRLADVRIPADTGDFCLM   178
BsCsbB      YLLKEFIKGYEEGYDQVIAQR-NRKGDSFVRSLLSSMYYKFINKAVEVDLRDGVGDFRLL   161
Eco306      EVIPHLIEKWQAGADMVLAKRSDRSTDGRLKRKTAEWFYKLHNKISNPKIEENVGDFRLM   159
SfBgt       EVIPHLIEKWQAGADMVLAKRSDRSTDGRLKRKTAEWFYKLHNKISTPKIEENVGDFRLM   159
             ..   .   . *   *   .     .          .*   ..     .      * *

VaWbhG      DKKIIEILDSIDDPFPYTRGMIMDLGFKLEKIYYRQPIRRRGVTSTNFLSLYDIAMLGIC   222
SspDpml     DRQVVDLLNTMPERNRYIRGLRAWVGFPQTGVKFERDPRHAGEVKYTFRKSLRLAINSLV   238
BsCsbB      SRQAVNALLKLSEGNRFSKGLFCWIGFDQKIVFYENVERKNGTSKWSFSSLFNYGMDGVV   221
Eco306      SRDVVENIKLMPERNLFMKGILSWVGGKTDIVEYVRAERIAGDTKFNGWKLWNLALEGIT   219
SfBgt       SREVVENIKLLPERNLFMKGILSWVGGQTDVVEYVRAERVAGISKFNGWKLWNLALEGIT   219
             .  .  .  . .  . .*.   .*     . .    *  *           .    .

VaWbhG      SHSKIPLRIATISGFILSLISFTLAIIFLALKIIYWDDFPMGTAPIIIGMFFLGSVQIFF   282
SspDpml     SFSIVPLRLATYLGLLAALLAIAMMILVLYWRLSETNSPLDGFATVVIANLFFGAVQLIC   298
BsCsbB      SFNHKPLRLCFYTGIFILLLSIIYIIATFVKILTNGI-SVPGYFTIISAVLFLGGVQLLS   280
Eco306      SFSTFPLRIWTYIGLVVASVAFIYGAWMILDTIIFGN-AVRGYPSLLVSILFLGGIQMIG   278
SfBgt       SFSTFPLRVWTYIGLFVASISFLYGAWMIIDTLVFGN-PVRGYPSLLVSILFLGGVQLIG   278
            *    ***.    *     ..           .        *   ..    * * .*.

VaWbhG      IGLVGEYVGHLVAKLSKFPLVIEKERINFD-------------------              312
SspDpml     IGILGEYIGRIYDEVKGRPLYTLAEMAGFEQLL----------------              331
BsCsbB      LGIIGEYIGRIYYETKKRPHYLIKEANIPNKDLPETNELKSMRRLTKMH              329
Eco306      IGVLGEYIGRTYIETKKRPKYIIKRVKK---------------------              307
SfBgt       IGVLGEYIGRIYIEVKNRPKYIIKKSHRGNP------------------              308
            .*..***.*.        *
```

Alignment of the amino acid sequences of WbhG (VaWbhG), *Synechocystis* sp Dpml (SspDpml) (spP14020), *B. subtilis* (BsCsbB) (spQ45539), *E. coli* o306 (Eco306) (spP77293) and, *S. flexneri* bacteriophage SfII (SfBgt) (gbAF021347) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. The shaded amino acids represent the functional motif DXDXD and DXD described for this family of proteins involved in the transfer of glucose to bactoprenol (Saxena *et al.*, 1995; Keenleyside and Whitfield, 1995; Mavris *et al.*, 1997). Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.8: Homology table for WbhG

	%identity					
	VaWbhG	**BsCsbB**	**Eco306**	**SfBgt**	**SspDpm1**	**ScDpm1**
VaWbhG	100	32.2 (307)	32.0 (303	31.4 (303)	28.0 (307)	27.1 (258)
BsCsbB		100	38.5 (301)	39.8 (304)	38.0 (303)	26.5 (147)
Eco306			100	89.9 (306)	41.1 (299)	16.0 (75)
SfBgt				100	43.7 (299)	16.0 (75)
SSpDpm1					100	30.2 (258)
ScDpm1						100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaWbhG	*V. anguillarum* WbhG
SspDpm1	*Synechocystis* sp Dpm1 (spP14020),
BsCsbB	B. subtilis (spQ45539),
Eco306	*E. coli* o306 (spP77293)
SfBgt	*S. flexneri* bacteriophage SfII (gbAF021347)
ScDpm1	Saccharomyces cerevivae Dpm1 (spJ04184)

The protein WbhG from *V. anguillarum* was found to be hydrophilic at its amino end with two potential transmembrane (hydrophobic) domains being located near its carboxy terminal (Fig. 5.16). Proteins of this type all contain this structural organisation, a comparison of which is shown in Fig. 5.16.

5.2.7.6 Modification and chain elongation enzymes

Towards the end of the sequenced region lie numerous ORFs that show homology to transferase genes from various different bacterial species. The ORFs are designated *wbhJ*, *wbhK*, *wbhL*, *wbhM* and *wbhN*. Enzymes involved in the transfer of sugars or other residues associated with O-antigens are often found located at the end of the O-antigen biosynthesis operon. This genetic arrangement is observed in *S. enterica* (Jiang *et al.*, 1991), *S. flexneri* (Morona *et al.*, 1994, 1995) and *B. pertussis* (Allen and Maskell, 1996).

5.2.7.6.1 *wbhJ*

WbhJ is similar to numerous proteins in the databases. The ORF that encodes WbhJ lies between nt 24038-nt 24493, and is most homologous to BlpB (41.1% identity in 151 amino acid overlap) from the *Bordetella pertussis rfb* region (Fig. 5.17). BlpB was assigned the role of an acetyl transferase, responsible for the transfer of an acetyl group to sugars groups, in the synthesis of *B. pertussis* O-antigen. BlpB and WbhJ are among a set of proteins that can be divided into two families. The first, acyltransferases encoded by *lxpA* and *lxpD* are enzymes involved in the transfer of fatty acyl chains from acyl carrier protein to glucosamine in the biosynthesis of lipid A (Raetz, 1993; Allen and Maskell, 1996). The second family are enzymes involved with the transfer of acetyl groups (ie acetyl transferases) encoded by *cysE*. This gene transfers acetyl groups from acetyl Co-A

Figure 5.16: Hydropathy plot of WbhG and related proteins.

The hydropathy plots were generated by the method of Kyte and Doolittle (1982) and aligned using PROFILEGRAPH (Hofmann and Stoffel, 1989). Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 10th amino acid is marked on each X-axis. Abbreviations: gb, Genbank; sp, SWISS PROT. VaWbhG, *V. anguillarum* WbhG; SspDpm1, *Synechocystis* sp Dpm1 (spP14020); BsCsbB, *B. subtilis* CsbB (spQ45539); Eco306, *E. coli* o306 (spP77293); SfBgt, *S. flexneri* SfII Bgt (gbAF021347); ScDpm1, *Saccharomyces cerevicae* Dpm1 (spJ04184)

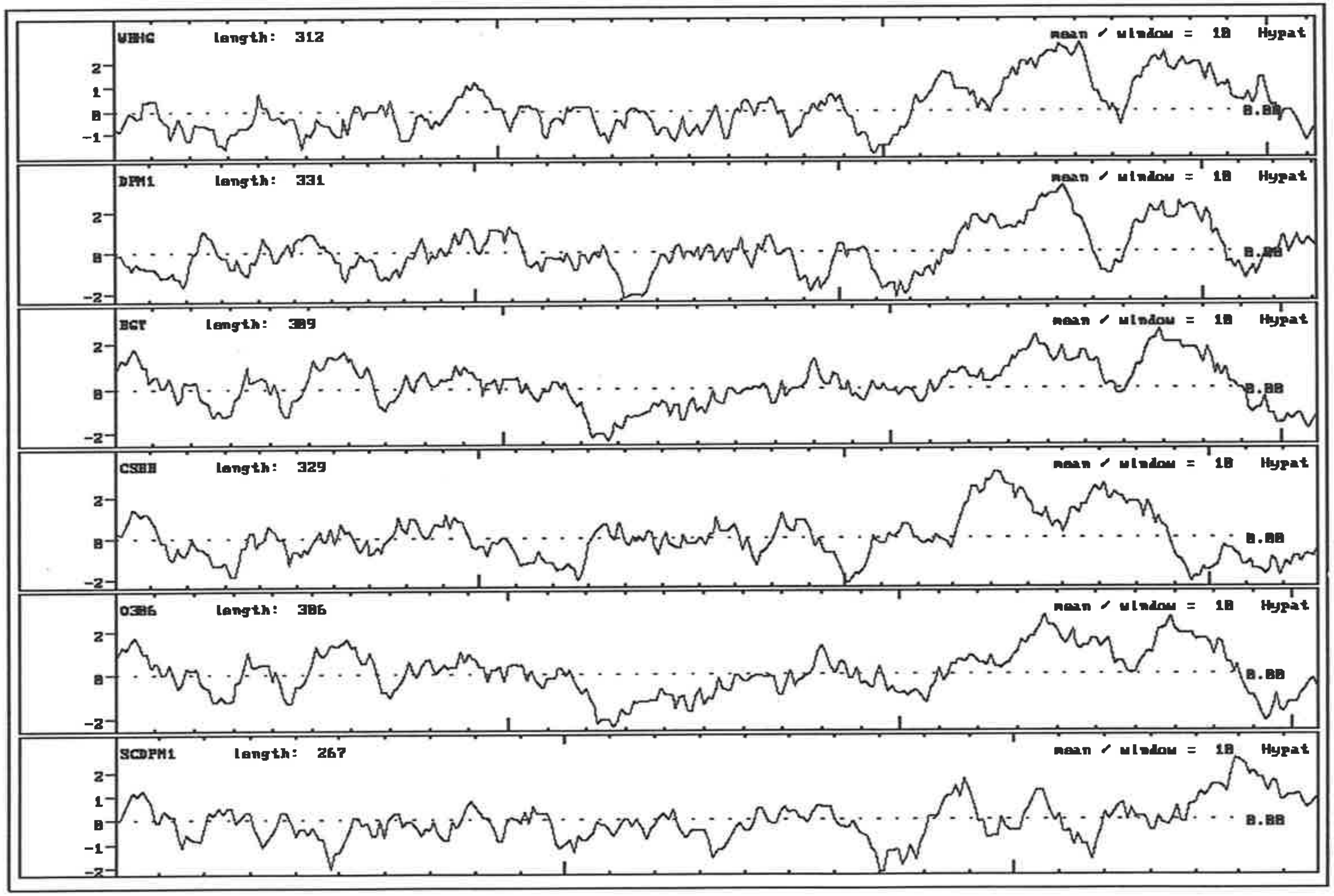

WBHG length: 312
mean / window = 10 Hypat
2
1
0
-1
0.00
DPM1 length: 331
mean / window = 10 Hypat
2
0
-2
0.00
BGT length: 309
mean / window = 10 Hypat
2
0
-2
0.00
CSBB length: 329
mean / window = 10 Hypat
2
0
-2
0.00
O306 length: 306
mean / window = 10 Hypat
2
0
-2
0.00
SCDPM1 length: 267
mean / window = 10 Hypat
2
1
0
-1
-2
0.00

Figure 5.17: Alignment of WbhJ with BplB

```
VaWbhJ     MTFIHELADVQS-SNIGINTKIWQFSVVLPNAIIGKNCNICSHTFIENDVTIGNNVTIKC   59
BpBplB     MTTIHPTAIVDEGARIGANSRIWHWVHICGGAEIGAGCSLGQNVFVGNRVRIGDRVKIQN   60
           ** **  * *   . ** *..**.   .  .* ** .* .  . *. * * **  * *.

VaWbhJ     GVQIWDGILIGNNVFIGPNATFTNDMYPRSKQ-YPDEFMKTVVCDNASIGANTTILPGVT  118
BpBplB     NVSVYDNVFLEDDVFCGPSMVFTNVYNPRAAIERKNEYRDTLVRQGATLGANCTIVCGAT  120
           .* . *.. .   ** **   ***   **.      *.  *.*  .*..***.**. * *

VaWbhJ     IGEGALVGAGSVVTKDVKPFTIVAGNPAREIEK---------------------------  151
BpBplB     VGRYAFVGAGAVVNKDVPDFALVVGVPARQIGWMSRHGEQLDLPLAGNGQARCPHTGDLY  180
           .*  * ****.**.***  *..* * ***.*

VaWbhJ     -----------                                                  151
BpBplB     ILENGVCRLGE                                                  191
```

Alignment of the amino acid sequences of WbhJ (VaWbhJ) and B. pertussis BplB (BsBplB) (pirS70673) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues.

to serine in the biosynthesis of cysteine (Wigley *et al.*, 1990; Gagnon *et al.*, 1994; Lai and Baumann, 1992). WbhJ (and BplB) display similarity to proteins from both these families which ranges from 20-25 % identity.

It is possible that WbhJ is involved in acetyl group transfer for the *V. anguillarum* O1 O-antigen similar to that of BlpB for *B. pertussis*. It is unlikely that WbhJ is involved in lipid A synthesis as lipid A genes are usually found in complex operons with other lipid A biosynthesis genes (Raetz, 1993).

5.2.7.6.2 *wbhK*

Immediately following *wbhJ* is the gene designated *wbhK*. The initiation codon of *wbhK* overlaps with the termination codon of *wbhJ* at nt 24490. WbhK shows significant homology to numerous proteins in the database involved in the synthesis of sugar containing antibiotics. It is most homologous to a transaminase enzyme (DesV) in the biosynthetic pathway responsible for production of macrolide antibiotics from *Streptomyces venezuelae* (Xue *et al.*, 1998), EryC from *Saccharopolyspora erythraea*, an enzyme involved in the synthesis of dTDP-D-desosamine in the erythromycin biosynthesis pathway (Dhillon *et al.*, 1989; Salah-Bey *et al.*, 1998), and DegT from *Bacillus stearothermophilus* (Takagi *et al.*, 1990) (Fig. 5.18; Table 5.9). Further similarities are to polysaccharide biosynthesis-related proteins involved in the transfer of amino groups to form amino sugars. These include BlpC from *B. pertussis* (Allen and Maskell, 1996), WbpE from *Pseudomans aeruginosa* (Burrows *et al.*, 1996) and the putative perosamine synthetase enzyme, RfbE/WbeE from *V. cholerae* (Manning *et al.*, 1994; Stroeher *et al.*, 1995c) (Fig. 5.18). WbhK was also similar to WbhE, a perosamine synthetase homologue from *V. anguillarum* O1 (27.4% over 354 amino acid overlap) (this study). Interestingly WbhK also has similarity to DdhC from *Salmonella typhimurium*, *Yersinia*

Figure 5.18: Alignment of WbhK with homologous proteins

```
VaWbhK      --------MISFLDLKKLNAQYEQELKDACTRVIDS-------------------GWYIL    33
SeEryC      -------MDVPFLDLQAAYLELRSDIDQACRRVLGS-------------------GWYLH    34
BsDegT      -------MNVPMLDLSEQYEQLKPEIMRVLDEVMRS-------------------SRFIL    34
BpBplC      ---------MQFIDLKTQYQALRDTINPRIQAVLDH-------------------GQFIM    32
PaWbpE      --------MIEFIDLKNQQARIKDKIDAGIQRVLRH-------------------GQYIL    33
SvDesV      MSSRAETPRVPFLDLKAAYEELRAETDAAIARVLDS-------------------GRYLL    41
VcRfbE      --------MIPVYEPSLDGNERKYLNDCIDSGWVSS--------------------RGKY    32
StRfbH      MTANNLREQISQLVAQYANEALSPKPFVAGTSVVPPSGKVIGAKELQLMVEASLDGWLTT    60
VaDdhC      MSKEQLRSQIAELVAQYAVIEYAPKMFIGGESVVPPSGKVLGAKELQLMVDASLDGWLTT    60
YpDdhC      MSQEELRQQIAELVAQYAETAMAPKPFEAGKSVVPPSGKVIGTKELQLMVEASLDGWLTT    60

VaWbhK      GNEVSEFEKEFAAYCNVEHCLGVANGLDALILILRAYIELGVMS----KGDEVIVPSNTY    89
SeEryC      GPENEAFEAEFAAYCENAHCVTVGSGCDALELSLVA---LGVG-----QGDEVIVPSHTF    86
BsDegT      GDYVKKLEADIAAYSRAKHGIGCGNGSDAIHIALQA---AGVG-----PGDEVITTAFTF    86
BpBplC      GPEVKELEAALCAYTGAKHCITVASGTEALLISLMA---LGVK-----AGDEVITTSFTF    84
PaWbpE      GPEVTELEDRLADFVGAKYCISCANGTDALQIVQMA---LGVG-----PGDEVITPGFTY    89
SvDesV      GPELEGFEAEFAAYCETDHAVGVNSGMDALQLALRG---LGIG-----PGDEVIVPSHTY    93
VcRfbE      IDR---FETEFAEFLKVKHATTVSNGTVALHLAMSA---LGIT-----QGDEVIVPTFTY    84
StRfbH      GRFNDAFEKKLGEFIGVPHVLTTTSGSSANLLALTALTSPKLGERALKPGDEVITVAAGF   120
VaDdhC      GRFNDAFEKRLGEYLGVPFVLTTTSGSSANLLALTTLTSPKLGDRQLKPGDEVITVAAGF   120
YpDdhC      GRFNDAFEKKLGEYLGVPYVLTTTSGSSANLLALTALTSPKLGVRALKPGDEVITVAAGF   120
                   *                 *  *  .         .       *****     .

VaWbhK      IASILAISEAGLVPVLVEPCEVTFNLDPNLIELAVTTRTKAILTVHLYGQVSGMDQIQKI   149
SeEryC      IATWLGVP-VGAVPVPVEPEGVSHTLDPALVEQAITPRTAAILPVHLYGHPADLDALRAI   145
BsDegT      FATAGSIARAGAKPVFVDIDPVTFNIDPAQVEAAVTEKTKAIIPVHLYGQMADMEAIAAI   146
BpBplC      VATAEVIALLGAKPVFVDVEPDTCNIKVSEIEAKITPRTKAIIPVSLYGQCGDMDEVNAV   144
PaWbpE      VATAETVALLGAKPVYVDIDPRTYNLDPQLLEAAITPRTKAIIPVSLYGQCADFDAINAI   149
SvDesV      IASWLAVSATGATPVPVEPHEDHPTLDPLLVEKAITPRTRALLPVHLYGHPADMDALREL   153
VcRfbE      VASVNTIVQCGALPVFAEIEGESLQVSVEDVKRKINKKTKAVMAVHIYGQACDIQSLRDL   144
StRfbH      PTTVNPAIQNGLIPVFVDVDIPTYNIDASLIEAAVTEKSKAIMIAHTLGNAFNLSEVRRI   180
VaDdhC      PTTVNPTIQNGLIPVFVDVDIPTYQIKPEMIEAAVSEKTKAIMVAHTLGNTFDLTEARRV   180
YpDdhC      PTTVNPTIQNGLIPVFVDVDIPTYNVNASLIEAAVSDKTKAIMIAHTLGNLFDLAEVRRV   180
             ..       *  **  .       .    .   .  .. *..     *          .

VaWbhK      ARKHNLKVIEDCAQAHGALYDSKEGIKKVGSIGDAAGFSFYPGKNLGALGDAGAVTTNDP   209
SeEryC      ADRHGLALVEDVAQAVGARHRGH----RVGAGSNAAAFSFYPGKNLGALGDGGAVVTTDP   201
BsDegT      AKRHGLVVIEDAAQAIGAKYNGK----CVGELGTAATYSFFPTKNLGAYGDGGMIITNDD   202
BpBplC      AARHGLPVIEDAAQSFGATYKGR----KSCNLSTIGCTSFFPSKPLGCYGDGGALFTNDD   200
PaWbpE      ASKYGIPVIEDAAQSFGASYKGK----RSCNLSTVACTSFFPSKPLGCYGDGGAIFTNDD   205
SvDesV      ADRHGLHIVEDAAQAHGARYRGR----RIGAGSSVAAFSFYPGKNLGCFGDGGAVVTGDP   209
VcRfbE      CDEHGLYLIEDCAEAIGTAVNGK----KVGTFGDVSTFSFFG-NKTITSGEGGMVVSNSD   200
StRfbH      ADKYNLWLIEDCCDALGTTYEGQ----MVGTFGDIGTVSFYP-AHHITMGEGGAVFTKSG   236
VaDdhC      ADKYNLWLIEDCCDALGSTYNGK----MVGTIGDIATVSFYP-AHHITMGEGGAVFTKDK   236
YpDdhC      ADKYNLWLIEDCCDALGSTYDGK----MAGTFGDIGTVSFYP-AHHITMGEGGAVFTQSA   236
            .   .. ..** . . *.    .               **.      . *. * . .

VaWbhK      ELASTISALRNYGS---------------------------HEKYRNIFKGLNSRLDEIQ   242
SeEryC      ALAERIRLLRNYGS---------------------------KQKYVHEVRGTNARLDELQ   234
BsDegT      ELAEKCRVIRVHGS---------------------------KPKYYHHVLGYNSRLDEMQ   235
BpBplC      ELAQAMREIRVHGQ---------------------------SGRYYHARIGVGGRMDTLQ   233
PaWbpE      ELATAIRQIARHGQ---------------------------DRRYHHIRVGVNSRLDTLQ   238
SvDesV      ELAERLRMLRNYGS---------------------------RQKYSHETKGTNSRLDEMQ   242
VcRfbE      IIIDKCLRLKNQGVVA-------------------------GKRYWHDLVAYNYRMTNLC   233
StRfbH      ELKKIIESFRDWGRDCYCAPGCDNTCGKRFGQQLGSLPQGYDHKYTYSHLGYNLKITDMQ   296
VaDdhC      ELRKLIESFRDWGRDCYCAPGCDNTCGKRFDQQLGSLPQGYDHKYTYSHLGYNLKITDMQ   296
YpDdhC      ELKSIIESFRDWGRDCYCAPGCDNTCKKRFGQQLGSLPFGYDHKYTYSHLGYNLKITDMQ   296
                        *                              .*       . ..  .

VaWbhK      AAMLRVKLRYLDDEISLRRKVASRYLNEVR-NPLIKVPLVE--DDNAHVWHLFVVVVE--   297
SeEryC      AAVLRVKLRHLDDWNARRTTLAQHYQTELKDVPGITLPETH--PWADSAWHLFVLRCE--   290
BsDegT      AAILSVKFPHLDRWTEQRRKHAATYTRLLEEAVGDLVVTPKEVDGRYHVFHQYTIRAP--   293
BpBplC      CAVVLGKLERFDWEIAQRIKIGARYQQLLADLPGGACTVTVRPD-RDSVWAQFTVMVP--   290
PaWbpE      AAILLPKLEIFEEEIALRQKVAAEY-----DLS---------------------------   266
SvDesV      AAVLRIRLAHLDSWNGRRSALAAEYLSGLAGLPGIGLPVTA--PDTDPVWHLFTVRTE--   298
VcRfbE      AAIGVAQLERVDKIIKAKRDIAEIYRSELAGLPMQVHKESN---GTFHSYWLTSIILDQE   290
StRfbH      AACGLAQLERVEEFVEQR-KANFSYLKQGLQSCTEFLELPEATEKSDPSWFGFPITLKET   355
VaDdhC      AACGLAQMDRVEEFVQAR-KENFAYLKNGLASCEEFIILPEATENSEPSWFGFPITIKDD   355
YpDdhC      AACGLAQLERIEEFVEKR-KANFKYLKDALQSCADFLELPEATENSDPSWFGFPITLKED   355
            .*    .    .     .      *
```

```
VaWbhK      ---DRDKLVEHLSENQIQSLIHYPIPPHKQEAYMEWN---NSAFPLSEKMHKQVLSLPLS  351
SeEryC      ---NRDHLQRHLTDAGVQTLIHYPTPVHLSPAYADLGL-PPGSFPVAESLAGEVLSLPIG  346
BsDegT      ---KRDELQAFLKEQGIATMVYYPLPLHLQPVFASLGY-KEGQLPEAEKAAKEALSLPMF  349
BpBplC      ---NREAVIAQLKEAGIPTAVHYPRPIHAQPAYEQYAE-GAGATPVSDDLAARVMSLPMH  346
PaWbpE      -----------LKQVGIGTP--------------FIGS-G--------------------  280
SvDesV      ---RRDELRSHLDARGIDTLTHYPVPVHLSPAYAGEAP-PEGSLPRAESFARQVLSLPIG  354
VcRfbE      FEVHRDGLMTFLENNDIESR-PFFYPAHTLPMYEHLAE--KTAFPLSNSYSHRGINLPSW  347
StRfbH      SGVNRVELVKFLDEAKIGTRLLFAGNLIRQPYFANVKYRVVGELTNTDRIMNQTFWIGIY  415
VaDdhC      SGISRVDLLKFMDQHKIGTRLLFAGNLTRQPYFEHVKYRVVGELTNTDLIMNNTFWIGVY  415
YpDdhC      SGVSRIDLVKFLDEAKVGTRLLFAGNLTRQPYFHDVKYRVVGELTNTDRIMNQTFWIGIY  415
                       •    • •

VaWbhK      SVLDKSEIDKVIDVLNSFLG-----                                      371
SeEryC      PHLSREAADHVIATLKAGA------                                      365
BsDegT      PELKEEQQQYVVEKIAEFYRHFA--                                      372
BpBplC      PDLDEATQDKIVAALRQALN-----                                      366
PaWbpE      -------------------------                                      280
SvDesV    PHLERPQALRVIDAVREWAERVDQA                                        379
VcRfbE      PGLCDDQVKEICNCIKNYFNCI---                                      369
StRfbH      PGLTTEHLDYVVSKFEEFFGLNF--                                      438
VaDdhC      PGLTLAHLDFVLEKFEEFFGVNF--                                      438
YpDdhC      PGLTHDHLDYVVSKFEEFFGLNF--                                      438
```

Alignment of the amino acid sequences of WbhK (VaWbhK), *Saccharopolyspora erythreae,* EryC (SeEryC*)* (spP14290*), B. subtilis* DegT (BsDegT) (spP15263), *B. pertussis* BplC (BsBplC) (pirS70674), *P. aeuringosa* WbpE (PaWbpE) (gbU50396), *Streptomyces venezuelae* transaminase, DesV (SvDesV) (gbAF079762), *V. cholerae* RfbE (VcRfbE) (pirS28471*), S. typhimurium* RfbH (StRfbH) (spP26398), *V. anguillarum* DdhC (VaDdhC) and, *Y. psuedotuberculosis* DdhC (YpDdhC) (gbE47070) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.9: Homology table for WbhK

					%identity[a]					
	VaWbhK	**SeEryC**	**BsDegT**	**BpBplC**	**PaWbpE**	**SvDesV**	**StRfbH**	**YpDdhC**	**VaDdhC**	**VcRfbE**
VaWbhK	100	47.8 (370)	40.2 (373)	36.6 (374)	42.3 (284)	48.1 (368)	36.0 (200)	35.5 (200)	35.0 (200)	29.7 (354)
SeEryC		100	44.7 (365)	40.5 (365)	43.5 (271)	66.8 (361)	33.0 (194)	32.0 (194)	33.0 (194)	31.4 (344)
BsDegT			100	45.9 (362)	51.5 (264)	44.1 (365)	34.2 (193)	33.3 (195)	43.8 (137)	29.2 (363)
BpBplC				100	63.2 (269)	40.4 (364)	35.1 (202)	36.6 (202)	34.7 (202)	29.4 (337)
PaWbpE					100	43.5 (276)	32.4 (185)	33.0 (185)	31.9 (185)	31.2 (234)
SvDesV						100	27.1 (210)	29.6 (162)	27.1 (210)	29.6 (365)
StRfbH							100	33.5 (203)	34.7 (202)	33.3 (204)
YpDdhC								100	80.8 (437)	87.4 (437)
VaDdhC									100	82.2 (437)
VcRfbE										100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. VaWbhK. *V. anguillarum* WbhK, SsEryC. *Saccharopolyspora erythreae,* EryC (spP14290), BsDegT *B. subtilis* DegT (spP15263), BsBplC *B. pertussis* BplC (pirS70674), PaWbpE *P. aeuringosa* WbpE (gbU50396), SvDesV *Streptomyces venezuelae* transaminase, DesV (gbAF079762), VcRfbE *V. cholerae* RfbE (pirS28471), StRfbH *S. typhimurium* RfbH (spP26398), VaDdhC *V. anguillarum* DdhC, YpDdhC. *Y. psuedotuberculosis* DdhC (gbE47070). Abbreviations: gb, Genbank; sp, SWISS PROT.

pseudotuberculosis and *V. anguillarum* (Reeves, 1994; Kessler *et al.*, 1993; this thesis) which are enzymes involved in the biosynthesis of CDP-4-keto-6-deoxy-D-glucose in the production of CDP-dideoxyhexose sugars (Table 5.9).

DegT and related proteins were first thought to function as protein kinase sensor elements in two component regulatory systems (Takagi *et al.*, 1990), however it is now thought that these proteins are required for the synthesis of 2,6-, 3,6-, 4,6- dideoxyhexoses, and for transamination of sugars (Thorson *et al.*, 1993; Allen and Maskell, 1996) which is interesting given the other homologues described above are proteins involved in the production of 3,6 dideoxyhexose and amino sugars.

5.2.7.6.3 *wbhL/virB* and *wbhM/virA*

The next putative transferase is encoded by the gene *wbhL* which lies between nt 25700 and nt 26788. *wbhL* and the following ORF, *wbhM,* are identical to the previously described genes *virBA*, respectively from *V. anguillarum* O1 and have been shown to be important in the virulence of this organism (Norqvist and Wolf-Watz, 1993). Mutations in *virBA* result in the loss of LPS/O-antigen production (Norqvist and Wolf-Watz, 1993). The arrangement of the genes, *whhL(virB)* and *wbhM (virA)* are identical to that described previously with the genes having a 9 bp gap between the termination codon of *wbhL* and the initiation codon of *wbhM*. WbhM does not show any significant similarity to other sequences and therefore its function is unknown.

WbhL demonstrates homology to a number of proteins predicted to be glycosyl transferases. WbhL is most similar to the putative glycosyl transferase genes of *Haemophilus influenzae* (LgtD) (Fleischmann *et al.*, 1995) and *Yersinia enterocolitica* type O:8 (WbcG) (Zhang *et al.*, 1996) (Fig. 5.19; Table 5.10). Figure 5.19 shows an alignment of WbhL with homologous proteins, the residues that are shaded indicate a motif for

Figure 5.19: Alignment of WbhL with homologous proteins

```
VaWbhL     MHLSDSKPPLVSLCILTFNHEKYITKAINSCLAQSYSNIEIIIVDNNSSDGTVNKIRSDF   60
VaVirB     MHLSDSKPPLVSLCILTFNHEKYITKAINSCLAQSYSNIEIIIVDNNSSDGTVNKIRSDF   60
HiLgtD     MNM-----PLISIIMPVYNAECYLNQGILSCLNQSYQNIELILIDDGSTDKSI-EIINNI   54
NgLgtD     -MQ-----PLVSVLICAYNAEKYFAQSLAAVVGQTWRNLDILIVDDGSTDGTP-AIARHF   53
YeWbcG     MIDD----IDVSVVIPVYNAERFIRTAISSVLSQEYVNIEVIIIDDGSTDSCG-KIIQSI   55
                     .*. .  .* * .    . . . *   *......* .*.* .   *

VaWbhL     KNELEVGEIKLFDLEHNTYPSHGFNYALKKSQ----GEYVSLFSGDDTLCLNKVERQIGI  116
VaVirB     KNELEVGEIKLFDLEHNTYPSHGFNYALKKSQ----GEYVSLFSGDDTLCLNKVERQIGI  116
HiLgtD     IDKDKR--VKLFFTPTNQGPAAARNIGLEKAQ----GDYITFLDSDDFIANDKLEKQLNF  108
NgLgtD     QEQDGR--IRIISNPRNLGFIASLNIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGE  111
YeWbcG     ND--DR--IKYFKKENG-GIVSALNFAIPKVH----ANIIARMDADDIMEPLRLRKQLDY  106
                   ..      .       *  .           ..    **      . . .

VaWbhL     MVKEG-LSNLFTWVNIINDKDE-----IIKCD----YLESIFNRNYNSQQIKEHFIHSGN  166
VaVirB     MVKEG-LSNLFTWVNIINDKDE-----IIKCD----YLESIFNRNYNSQQIKEHFIHSGN  166
HiLgtD     MLQN----------HLVMTHGN-----YAFCD----LEGNQIKLVTTSKKIDYLTLLQGN  149
NgLgtD     MEKDRSIIAMGAWLEVLSEENNKSVLAAIARNGAIWDKPTRHEDIVAVFPFGNP-IHNNT  170
YeWbcG     MARYN-LDVVGGNIKLIDENDR-----IIGRK----KFPTNHFDIISSLPFINPLCHPAT  156
           * .            ..                                           .

VaWbhL     MLSALSVMLSRDVFDRYGHFDERLVQLQDFDFWLRMASNDDLNLLTEKLSNYRLRDDGGN  226
VaVirB     MLSALSVMLSRDVFDRYGHFDERLVQLQDFDFWLRMASNDDLNLLTEKLSNYRLRDDGGN  226
HiLgtD     QFKIMTVLVER-----------------------------------ESIKLLRFPN----  170
NgLgtD     MIMRRSVIDGGLR------FDPAYIHAEDYKFWYEAGKLGRLAYYPEALVKYRFHQDQTS  224
YeWbcG     MIRTAVLKKANG-------YSLGTDGAEDFDLWCRLSRVCTFGSVPEDLLSYRLTSN--S  207
                 .                                       * .   *

VaWbhL     LSLANHKSRQLRTDFEEVYVYRHLLNFDLKTIQSVVGMLNKDQSIAMALHGYYHNENKMK  286
VaVirB     LSLANHKSRQLRTDFEEVYVYRHLLNFDLKTIQSVVGMLNKDQSIAMALHGYYHNENKMK  286
HiLgtD     ---IKHED------------YAFFLDCLKEVKQSIL---------------YSHQ-----  195
NgLgtD     SKYNLQQRR-------TAWKIKEEIRAGYWKAAGIAVG--------ADCLNYGLLKSTAY  269
YeWbcG     ISQTLHLHR-----------ISLCTNSIRKNIPNAP------------------------  232
                .

VaWbhL     LAKGFLLSIYEELGTNIVFPSSHYSYFFDIYSKCEFFNSDENDEVSKLKEKIHVYENSRA  346
VaVirB     LAKGFLLSIYEELGTNIVFPSSHYSYFFDIYSKCEFFNSDENDEVSKLKEKIHVYENSRA  346
HiLgtD     -ASSFVRIGKVSVSSNK-FKSAIWT--FNIY-----FK----------REKLGVVKS--I  235
NgLgtD     ALYEKALSGQDIGCLRLFLYEYFLSLEKYSLTDLLDFLTD--------RVMRKLFAAP-Q  320
YeWbcG     -IVFYSFSLYLSKSLAVLIYAAKYA--YITK-----FK----------SFYYMISGFIRA  274
                                   .           *                .

VaWbhL     IRFTSLIVSWLGRLKK-                                            362
VaVirB     IRFTSLIVSWLGRLKK-                                            362
HiLgtD     YYFILYAYNGFIKYKK-                                            251
NgLgtD     YRKILKKMLRPWKYRSY                                            337
YeWbcG     LNFSFKIHKGKNEC---                                            288
```

Alignment of the amino acid sequences of WbhL (VaWbhL), *V. anguillarum* VirB (VaVirB) (gbL08012), *H. influenzae* LgtD (HiLgtD) (spQ57022), *N. gonnorrhoea* LgtD (NgLgtD) (gbU14554) and, *Y. enterocolitica* WbcG (YeWbcG) (gbU46859) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Shaded amino acids represent the motif for glycosyltransferases as described by Morona et al. (1995). The shaded residues are identical to the proposed consensus sequence. Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.10: Homology table for WbhL

	%identity[a]				
	VaWbhL	**VaVirB**	**HiLgtD**	**YeWbcG**	**NmLgtD**
VaWbhL	100	100 (362)	34.2 (244)	26.0 (223)	24.1 (241)
VaVirB		100	34.2 (244)	26.0 (223)	24.1 (241)
HiLgtD			100	38.1 (126)	32.5 (120)
YeWbcG				100	27.1 (144)
NmLgtD					100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaWbhL	*V. anguillarum* WbhL
VaVirB	*V. anguillarum* VirB (gbL08012),
HiLgtD	*H. influenzae* LgtD (spQ57022),
YeWbcG	*Y. enterocolitica* WbcG (gbU46859)
NgLgtD	*N. gonnorrhoea* LgtD (gbU14554)

glycosyltransferases which was described by Morona *et al.* (1995a) (see also Section 5.2.7.6.4). Although not as significant as the homology seen to the glycosyl transferases, *wbhL* is also similar to a N-acetylglucosaminyl transferase (21.8 % identity at the amino acid level over 188 amino acid overlap) from *Streptococcus pneumoniae* (Kolkman *et al.*, 1996).

5.2.7.6.4 *wbhN*

The final ORF designated *wbhN* lies between nt 27828 and nt 28763 and has three potential initiation codons: valine (GTG) at position #27828, methionine (ATG) at position #27841, and leucine (TTG) at position # 27879. The potential translation starts each have a putative ribosome binding site (RBS), however the best RBS was found with the leucine (TTG) start (Appendix 1). When the valine initiation codon at nt 27828 was used to translate the sequence, a stop codon (TAG) at position nt 27867 was found. A similar situation was also found for the methionine start at nt 27841 with a stop codon (TAA) at nt 27895 (Fig. 5.20). Translation of the sequence encoding *wbhN* using DNASIS indicated that the leucine start at position nt 27879 is the most probable site of initiation. Although the valine and methionine starts would produce truncated proteins, they can not be discounted as potential initiation sites. *wbhN* was found to have an interesting feature of a homopolymeric tract of 11 cytidine residues (poly (dC)) at nt 27850 to nt 27861. Depending on the initiation codon used to translate *wbhN,* the poly (dC) tract either lies in the 5' end of the gene or in the intergenic region preceding the beginning of *wbhN* (Appendix 1) (Fig. 5.20).

WbhN shows similarity rhamnosyl transferase enzymes, RfbF from *S. flexneri* (Morona *et al.*, 1994; 1995a), RfbQ from *S. dysenteriae* (Klena and Schnaitman, 1993) and RfbC from *Y. enterocolitica* (Zhang *et al.*, 1993) (Fig. 5.21; Table 5.11). At the DNA

Figure 5.20: Potential Initiation codons of *wbhN*.

The sequence used in this figure was derived from Appendix 1. Analysis of the sequence encoding *wbhN* suggests that the ORF may have three potential initiation codons. The initiation codons are indicated in bold type and are valine (GTG) at position nt 27828, methionine (ATG) at nt 27841 and leucine at nt 27879. The cytidine tract is bolded and underlined with double lines. The termination codons are bolded.

A: INITIATION CODON: Valine (GTG) at position # 27828

A

27730 27740 27750 27760 27770 27780 27790 27800 27810 27820 27830 27840

CTAGAGTAATTTGATTTGGTATTCTATTTTCATATTTCACGTAGTTCGACTATTATATTTTTAGTGCTCTGCCCACAGATTTTTTATATAAATATTATAGTGACATT**GTG**AGCATTCGAT

wbhN V S I R Y

27850 27860 27870 27880 27890 27900 27910 27920 27930 27940 27950 27960

ATGTATCTT**CCCCCCCCCCC**AAAGTA**TAG**GAGTATTTTTTGTCGATTTCATCCGTAATTATCACTTTTCAACCCGATCTATCGAATGTCTCCCAGTTACTATTCGCTTGTATATCCTTTG

V S S P P P K V * E Y F L S I S S V I I T F Q P D L S N V S E L L F A C I S F G

B: INITIATION CODON: Methionine (ATG) at position #27841

B

27730 27740 27750 27760 27770 27780 27790 27800 27810 27820 27830 27840

CTAGAGTAATTTGATTTGGTATTCTATTTTCATATTTCACGTAGTTCGACTATTATATTTTTAGTGCTCTGCCCACAGATTTTTTATATAAATATTATAGTGACATTGTGAGCATTCGAT

27850 27860 27870 27880 27890 27900 27910 27920 27930 27940 27950 27960

ATGTATCTT**CCCCCCCCCCC**AAAGTATAGGAGTATTTTTTGTCGATTTCATCCG**TAA**TTATCACTTTTCAACCCGATCTATCGAATGTCTCCCAGTTACTATTCGCTTGTATATCCTTTG

wbhN M Y L P P P P K Y R S I F C R F H P *

C: INITIATION CODON: Leucine (TTG) at position #27879

27730 27740 27750 27760 27770 27780 27790 27800 27810 27820 27830 27840

CTAGAGTAATTTGATTTGGTATTCTATTTTCATATTTCACGTAGTTCGACTATTATATTTTTAGTGCTCTGCCCACAGATTTTTTATATAAATATTATAGTGACATTGTGAGCATTCGAT

27850 27860 27870 27880 27890 27900 27910 27920 27930 27940 27950 27960

ATGTATCTT**CCCCCCCCCCC**AAAGTA**TAG**GAGTATTTT**TTG**TCGATTTCATCCGTAATTATCACTTTTCAACCCGATCTATCGAATGTCTCCCAGTTACTATTCGCTTGTATATCCTTTG

* E Y F L S I S S V I I T F Q P D L S N V S E L L F A C I S F G

wbhN →

Figure 5.21: Alignment of WbhN with homologous proteins

```
VaWbhN          -----LSISSVIITFQPDLSNVSQLLFACISFGNKAVVIDNGSNN--AEELQDICRLFEH   53
SfRfbF          ---MNSNIYAVIVTYNPELKNLNALITELKEQNCYVVVVDNRTN----FTLKDKLADIEK   53
SdRfbQ          --MIKKKVAAIIITYNPDLTILRESYTSLYKQVDKIILIDNNSTN--YQELKKLFEKKEK   56
YeRfbC          MNASQYSISAVVVTENPEVEVFKQLIKSIIPQVDQIIIVDNGSLCEIRSGLEEIVTNYQN   60
                             *  *                      **         *
                       . .... .. ..                 ...   .               .

VaWbhN          VKLIRLDENVGIASAQNIAISNLNGNEDDIIVFFDQDSSID----NGYLSKVELAYNRLE  109
SfRfbF          VHLICLGRNEGIAKAQNIGIRYSLEKGAEKIIFFDQDSRIR----NEFIKKLSCYMDNEN  109
SdRfbQ          IKIVPLSDNIGLAAAQNLGLNLAIKNNYTYAILFDQDSVLQDNGINSFFFEFEKLVSEEK  116
YeRfbC          LYLISLSDNMGIASAQNYGIKKSFSMESSHVLLLD[illegible]SIPAEDMVEKLLSLELKLLSQGR  120
                     *  * * * ***                 * **
                . ..       .       .           .   .

VaWbhN          SDFGRGIVLGPRFYNRVS-KFEYPVIKFNIFGLRSRIYPSESRYPIEASCIISSGMAVRK  168
SfRfbF          -----AKIAGPVFIDRDK-SHYYPICNIKKNGLREKIHVTEGQTPFKSSVTISSGTMVSK  163
SdRfbQ          ---LNIVAIGPSFFDEKTGRRFRPTKFIGPFLYPFRKITTKN-PLTEVDFLIASGCFIKL  172
YeRfbC          ----QVGAVGPTSVDRRTSTRSGFVRKSG--ILIKRIYPDDSKGFVKTDFLIASGTLIRT  176
                         **                        .               *.**  .

VaWbhN          NILDSVGVMDDSLFIDYVDTEWSLRARYLGNLILVDPQLVMGHEIGTDNLKLFK---WRV  225
SfRfbF          EVFEIVGMMDEELFIDYVDTEWCLRCLNYGILVHIIPDIEMVHAIGDKSVKICG---INI  220
SdRfbQ          ECIKSAGMMTESLFIDYIDVEWSYRMRSYGYKLYIHNDIHMSHLVGESRVNLGL---KTI  229
YeRfbC          EVLHNIGLMKENYFIDHVDTEWCFRAVNCGIELFGCGDAFLNHTLGDSVIRIWMGRWREI  236
                      *.* .  *** .* **. *    *  .       . * .*   . .       .

VaWbhN          PVHSASRRYYRIRNSFFLFRYPHIPRLVCTREVTFSILHQLFLVLLTNEKKAHWKS----  281
SfRfbF          PIHSPVRRYYRVRNAFLLLRKNHVPLLLSIREVVFSLIHTTLIIATQKNKIEYMKKHILA  280
SdRfbQ          SLHGPLRRYYLFRNYISILKVRYIPLGYKIREGFFNIGRFLVSMIITKNRKTLILY----  285
YeRfbC          PKHSPLRNYYIFRNTINMVTSTPMSWSWKLAHIYRLMIFFVFFMIAAKPRYRRFIM----  292
                  *   * **  **
                                 .     .            .      .     .

VaWbhN          ---LWRGIKDGV-----FYKS                                        294
SfRfbF          TLDGIRGITGGG-----RYNA                                        296
SdRfbQ          ---TIKAIKDGINNEMGKYKG                                        303
YeRfbC          ---MIKGGFDGIRGISGKLN-                                        309
                          *
                     .
```

Alignment of the amino acid sequences of WbhN (VaWbhN), *S. flexneri* RfbF (SfRfbF) (spP37782), *S. dysenteriae* RfbQ (SdRfbQ) (pirS34966) and *Y. enterocolitica* RfbC (YeRfbC) (pirS28579) as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. Shaded amino acids represent the motif for glycosyltransferases as described by Morona *et al.* (1995). The lighter shading are residues that are 100% identical to proposed consensus sequence. Abbreviations: gb, Genbank; sp, SWISS PROT.

Table 5.11: Homology table for WbhN

	%identity[a]			
	VaWbhN	**SfRfbF**	**SdRfbQ**	**YeRfbC**
VaWbhN	100	37.2 (274)	29.0 (290)	27.9 (294)
SfRfbF		100	30.2 (298)	29.5 (298)
SdRfbQ			100	30.0 (300)
YeRfbC				100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: sp, SWISS PROT.

VaWbhN	*V. anguillarum* WbhN
SfRfbF	*S. flexneri* RfbF (spP37782),
SdRfbQ	*S. dysenteriae* RfbQ (pirS34966)
YeRfbC	*Y. enterocolitica* RfbC (pirS28579)

level, *wbhN* showed some homology to *rfbC* from *Y. enterocolitica* (55%), although this similarity was concentrated at the first 170 bp of the respective genes. Morona *et al.* (1995a) suggested that a specific motif in the amino terminal existed for sugar transferases that involved 6-deoxy hexoses. This motif highlighted in Figure 5.21 was found in the protein sequence of WbhN suggesting that it belongs to this family of related proteins. Examination of the other putative transferases in the locus showed that *wbhL* also contained this motif suggesting that it also belongs to this family of transferases (Fig. 5.22). Figure 5.22 also shows a CLUSTAL alignment of WbhN and WbhL and highlights the shared motif. It is evident however, that the similarity between these proteins is quite low (22.1% over 77 amino acid overlap), although they may have a similar function.

5.2.7.7 ORFs of unknown function within the *wbh* region

The *wbh* region sequenced in this study has numerous ORFs that do not show any significant homology to any genes in the Genbank database. These genes have been designated *wbhA* (nt 8670- nt 9785), *wbhB* (nt 11141- nt 12790), *wbhD* (nt 18903- nt 19532), *wbhF* (nt 20689- nt 21075), *wbhH* (nt 22062- nt 23408) and *wbhI* (nt 23622- nt 24020). An analysis using TMpredict (Hofmann and Stoffel, 1993) and the hydrophobicity analysis program of PROSIS (Kyte and Doolittle, 1982) indicated that WbhA and WbhD were hydrophilic proteins while WbhB, WbhF and WbhH had transmembrane domains. WbhB was predicted to have 11 transmembrane domains (Fig. 5.23), WbhF to have 4 potential transmembrane domains (Fig. 5.24) and WbhH was suggested to be mainly hydrophilic with two transmembrane domains in the carboxy terminal end of the protein (Fig. 5.25).

Figure 5.22: Alignment of WbhL and WbhN

```
WbhL      MHLSDSKPPLVSLCILTFN-HEKYITKAINSCLAQSYSNIEIIIVDNNSSDGTVNKIRSD   59
WbhN      --LS------ISSVIITFQPDLSNVSQLLFACISFGN---KAVVIDNGSNN---AEELQD   46
            **      .*  *.**.     ... . .*..        ...**.*          *

WbhL      FKNELEVGEIKLFDLEHNTYPSHGFNYALKKSQGEYVSLFSGDDTLCLNKVERQIGIMVK  119
WbhN      ICRLFEH--VKLIRLDENVGIASAQNIAISNLNG--------------NEDD--------   82
               *   .**  *. *   .   * *.   .*              *  .

WbhL      EGLSNLFTWVNIINDKDEIIKCYLESIFNRNYNSQQIKEHFIHSGNMLSALSVMLSRDVF  179
WbhN      ---------IIVFFDQDSSIDNGYLSKVELAYN--RLESDFGRG--------IVLGPRFY  123
                   . .  *.*  *     *     **  ..   * .         ..*    .

WbhL      DRYGHFDERLVQLQDFDFWLRMASNDD---LNLLTEKLSNYRLRDDGGN-LSLANHKSRQ  235
WbhN      NRVSKFEYPVIKFNIFGLRSRIYPSESRYPIEASCIISSGMAVRKNILDSVGVMDDSLFI  183
           *  .*.  ... . *    *.   .    .   .   *.  .*      . .

WbhL      LRTDFEEVYVYRHLLNFDLKTIQSVVGMLN-KDQSIAMALHGYYHNENK--MKLAKGFLL  292
WbhN      DYVDTEWSLRARYLGNLILVDPQLVMGHEIGTDNLKLFKWRVPVHSASRRYYRIRNSFFL  243
             * *     * * *  *   * *.*     *.      .   *   .   ..   * *

WbhL      SIYEELGTNIVFPSSHYSYFFDIYSKCEFFNSDENDEVSKLKEKIHVYENSRAIRFTSLI  352
WbhN      FRYPHIPRLVCTREVTFSILHQLFLVLLTN-----------EKKAHWKSLWRGIKDGVFY  292
            *  .   .      .*    ..                   * *     * *.

WbhL      VSWLGRLKK                                                    362
WbhN      KS-------                                                    294
           *
```

Alignment of the amino acid sequences of WbhL and WbhN as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues. The shaded regions show identical amino acids with respect to the conserved motif for this class of glycosyl transferases as described by Morona *et al.* (1995).

Figure 5.23: Predicted transmembrane topology and hydropathy plot for *wbhB*.

Using the computer program TMpredict (Hofmann and Stoffel, 1993), a model for the topology of *wbhB* was proposed. The suggested transmembrane domains are shown in the table. The corresponding hydropathy plot shows that *wbhB* has 11-12 potential transmembrane domains. Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 100th amino acid is marked on the X-axis.

Suggested model for transmembrane topology by TMpredict

TM #	Start	End	Length	Orientation
1	11	30	20	i-o
2	69	90	22	o-i
3	99	118	20	i-o
4	126	145	20	o-i
5	151	169	19	i-o
6	177	199	23	o-i
7	225	246	22	i-o
8	278	297	20	o-i
9	303	320	18	i-o
10	324	347	24	o-i
11	363	381	19	i-o

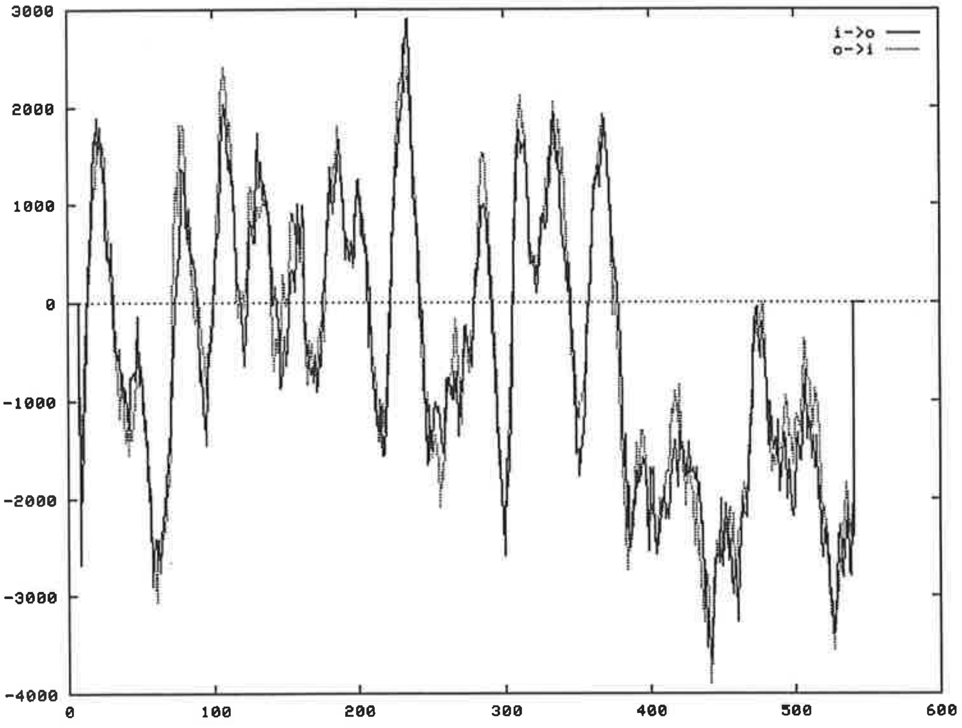


Figure 5.24: Predicted transmembrane topology and hydropathy plot for *wbhF*.

Using the computer program TMpredict (Hofmann and Stoffel, 1993), a model for the topology of *wbhF* was proposed. The suggested transmembrane domains are shown in the table, with two potential models described. The corresponding hydropathy plot shows that *wbhF* has 4 potential transmembrane domains. Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 20th amino acid is marked on the X-axis.

Suggested models for transmembrane topology

A

TM #	Start	End	Length	Orientation
1	15	38	24	o-i
2	42	60	19	i-o
3	76	95	20	o-i
4	104	121	18	i-o

B

TM #	Start	End	Length	Orientation
1	20	38	19	i-o
2	42	60	19	o-i
3	76	92	17	i-o
4	106	124	19	o-i

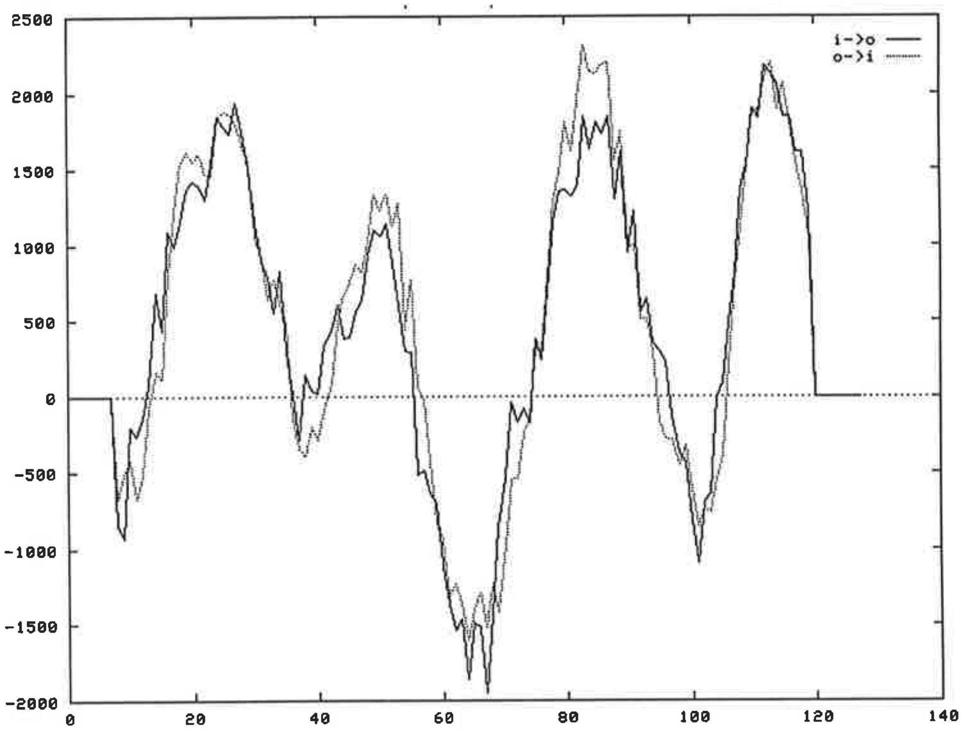


Figure 5.25: Predicted transmembrane topology and hydropathy plot for *wbhH*.

Using the computer program TMpredict (Hofmann and Stoffel, 1993), a model for the topology of *wbhH* was proposed. The suggested transmembrane domains are shown in the table, with two potential models described. The corresponding hydropathy plot shows that *wbhH* has 2 potential transmembrane domains. Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 50th amino acid is marked on the X-axis.

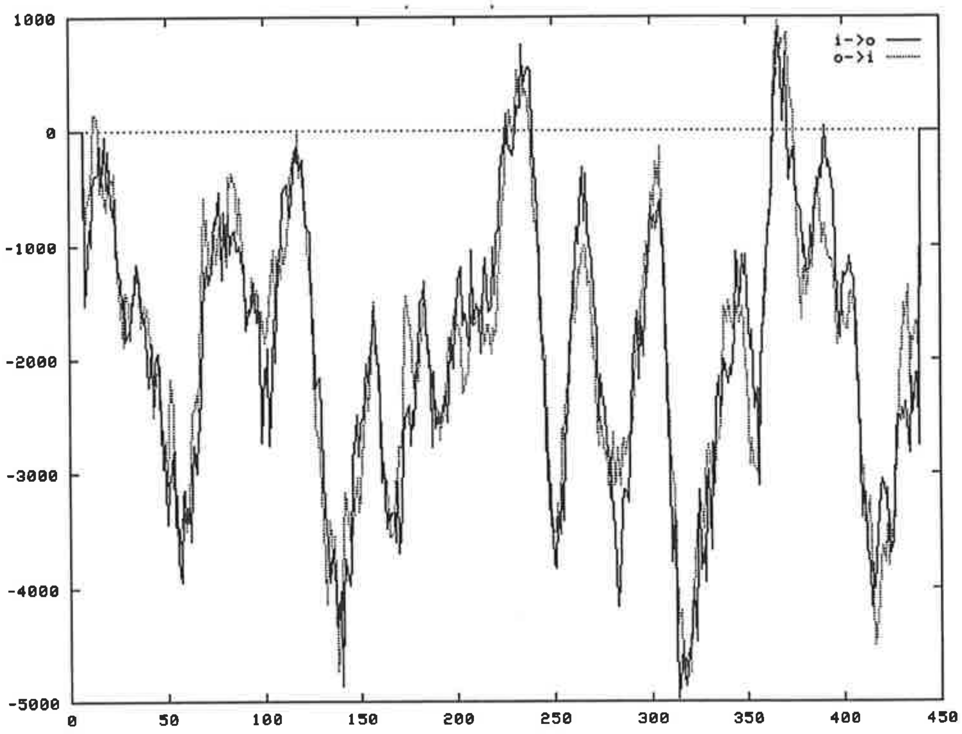
Suggested models for transmembrane topology

A

TM #	Start	End	Length	Orientation
1	227	245	19	i-o
2	357	378	22	o-i

B

TM #	Start	End	Length	Orientation
1	225	244	20	o-i
2	357	379	23	i-o



5.2.7.8 Analysis of Complementary Strand

Analysis of the complementary strand revealed 3 potential ORFs which showed homology to genes located in the O-antigen biosynthesis operon of *V. cholerae* O1 (Manning et al., 1986). *orf2* and *gmhD* are located at the 5' end of the operon. ORF2 has not been assigned a function and GmhD is a RfaD homologue involved in core heptose biosynthesis. WbhO, which is a truncated ORF, shows similarity to TrsG of *Yersinia entercolitica* O3 (Skurnik *et al.*, 1995) and RfbV/WbeH *of V. cholerae* O1 (Fallarino *et al.*, 1997). TrsG is proposed to be involved in outer core oligosaccharide biosynthesis in *Yersinia enterocolitica* (Skurnik *et al.*, 1995). In *V. cholerae* O1, the role of the TrsG homologue WbeH is not known although it is essential for O-antigen biosynthesis and therefore unlikely to be involved in core biosynthesis as mutations in *wbeH* do not effect lipid A- core motility by SDS- PAGE and silver staining (Fallarino *et al.*, 1997).

5.2.8 Promoters and Translational Coupling

Transcriptional coupling within the *wbh* region of *V. anguillarum* O1 is evident (Appendix 1). However, there are numerous regions of intergenic sequence of greater than 90 bp corresponding to nts 9785-11140, 12790-12930, 18810-18902, 23410-23620, 25606-25699 and 27713-27879 (Appendix 1). This is reminiscent of the *V. cholerae* O139 O-antigen/capsule biosynthesis locus which also shows pattern of integenic regions (Stroeher *et al.*, 1998). Hence, it is likely that the *wbh* region is transcribed in discreet blocks.

Preceding the JUMPstart sequence there appears to be a good promoter (P_1) corresponding with nucleotides 2665 to 2689 with the -10 box being identical to the consensus sequence, TATAAT (Table 5.12; Fig. 5.26; Appendix 1). Other potential promoters were located in this region however they did not show as good

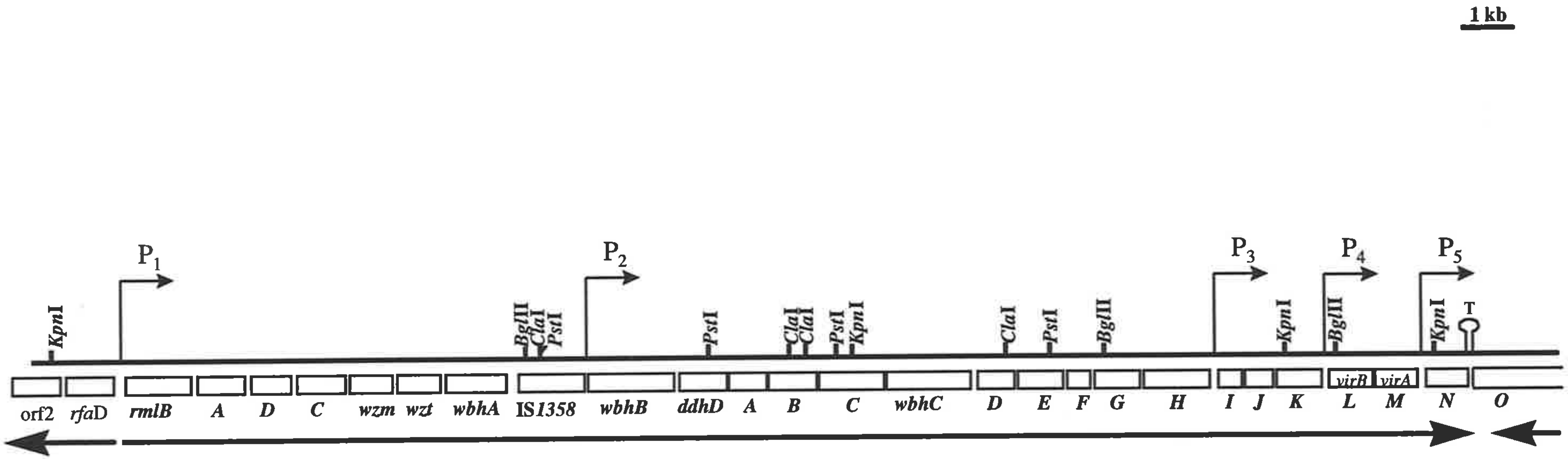
Table 5.12: Putative promoter regions in the *wbh* locus

Promoter #[a]	Location[b]	Sequence
P_1	nt 2665-2689	CATACAaaaattcctatggtTATAAT
P_2	nt 10787-10821	TGTAGCaacagagcaagatgtgagtattcgttacTATATT
P_3	nt 23456-23478	TTAGGAtgaaaatataatataaTATAAT
P_4	nt 25644-25660	TTGGAAaaactgtttacTATTAA
P_5	nt 27731-27771	TTGATTtggtattctattttcatatttcacgtagttcgacTATTAT

Legend:
The putative promoter regions are listed above and refer to those in Figure 5.28[a] and Appendix 1[b]. The capital letters refer to those bases that are identical to the consensus sequence for the -35 (TTGACA) and -10 (TATAAT) regions. Those letters underlined do not match the consensus. The lower case letters indicate the sequence between the -35 and -10 boxes.

Figure 5.26: The promoter and terminator regions of the *wbh* locus.

A schematic representation showing the location and direction of the promoters P_1-P_5. The potential terminator structure (T) is also indicated with a stem loop diagram after *wbhN*.


1 kb
P1
P2
P3
P4
P5
KpnI
BglII
ClaI
PstI
PstI
ClaI
ClaI
PstI
KpnI
ClaI
PstI
BglII
KpnI
BglII
KpnI
T
virB
virA
orf2
rfaD
rmlB
A
D
C
wzm
wzt
wbhA
IS1358
wbhB
ddhD
A
B
C
wbhC
D
E
F
G
H
I
J
K
L
M
N
O

homology/similarity to the consensus sequence for either the -35 (TTGACA) or -10 (TATAAT).

As it appeared that the *wbh* region was not completely transcriptionally coupled, the intergenic regions were searched for other potential promoters. The putative promoters located in these intergenic regions have strong consensus to the -10 box but the homology to the -35 box was variable (Table 5.12). Between *wbhA* and *wbhB* lies the interrupted ORFs of IS*1358*. Within this region (nt 9785-11140 bp) is a putative promoter (P_2) at nt 10787 to 10821, however, the spacing between the -35 and -10 boxes is not ideal. The -35 box is TGTAGC and the -10 box is TATATT (Table 5.12; Fig. 5.26).

The next potential promoter (P_3) lies between *wbhH* and *wbhI*. There are numerous potential -10 boxes, all being identical to the consensus sequence. However no strong -35 box was located, with the closest being TTAGGA at nt 23456. The corresponding -10 box is located at nt 23478 (Table 5.12).

In the intergenic region preceding *wbhL* there appears to be a promoter (P_4) with a -35 box (TTGGAA) and a -10 box (TAAAAT) with reasonable homology to the consensus sequence (Table 5.12; Fig. 5.26). Norqvist and Wolf-Watz (1993) reported this as the potential promoter for *virB*. The final putative promoter (P_5) located in the sequenced locus was found between *wbhM* and *wbhN*. The promoter was found to have a -35 box at nt. 27731 (TTGATT) and a -10 box at nt 27771 (TATTAT) which closely resembles the consensus sequences, however the spacing is not ideal (Table 5.12; Fig. 5.26).

Although numerous potential promoter regions were located, only one stem loop structure resembling a terminator was observed. This terminator was found at the 3' end of *wbhN* at nt 28886 to nt 28916 with the inverted repeated sequence 5' CTTtAATATCTggaaagtagtAGATATTgAAG 3' (Fig. 5.26; Appendix 1).

Most of the genes described in the region display ribosome binding sites that are similar to the consensus sequence (TAGGAGGTGATC) (Table 5.13). The ribosome binding sites are bolded on the sequence in Appendix 1 and listed in Table 5.13. Most of the genes/ORFs in the *wbh* region initiate and terminate within 1-20 bp. There is some indication that some of the genes/ORFs are translationally coupled. The rhamnose biosynthetic genes (*rmlBADC*) are closely linked with the initiation codon (ATG) of *rmlD* overlapping with the termination codon (TGA) of *rmlA*. Other genes that show this organisation are *ddhA*/*ddhB*, *ddhB*/*ddhC*, *wbhD*/*wbhE*, *wbhF*/*wbhG* and *wbhJ*/*wbhK*. Termination codon usage in the locus corresponds to the usual *E. coli* preferences with TAA being used 12 times as a single stop codon (Appendix 1). TGA and TAG are used five and seven times, respectively as single termination codons.

5.2.9 Protein expression using T7 RNA polymerase expression system

A set of plasmids containing most of the 24 ORFs identified in the *V. anguillarum* O1 *wbh* region were used in the T7 expression system (section 2.13.2) to identify corresponding gene products (Fig. 5.27; Fig. 5.28). Initially, the overexpression assay was performed without the addition of L-[^{35}S]-methionine to detect gene products that were highly expressed. Gene products could be detected for: *rfaD* (35 kDa), orf2 (20 kDa) (pPM5516); *rmlB* (40 kDa) (pPM5518); *rmlA* (30 kDa) (pPM5517); *rmlC* (22 kDa), *wzm* (30 kDa), *wbhA* (40 kDa) (pPM5515); *ddhA* (29 kDa) (pPM5526); *ddhB* (40 kDa) (pPM5527); *wbhD* (25 kDa) (pPM5528), *wbhE* (40 kDa) (pPM5524); *wbhJ* (17 kDa), *wbhK* (41 kDa) (pPM5533) and *wbhN* (32 kDa) (pPM5534) (Fig. 5.27). The sizes of the proteins detected correlate to the predicted molecular weights of each gene product (Table 5.14). The smaller molecular weight gene product of 21 kDa observed in pPM5526 may

Table 5.13: Features of the initiation regions of the *wbh* genes

Gene/ORF	Sequence[a]	Space between SD and first codon
rmlB	tatTAttcgGGTaAcaatATGaagatt	4
rmlA	aatTAAGttGGcGATttcATGaaaggc	3
rmlD	taaagAGGAatccCattcatgATGcgtgtt	7
rmlC	ctaTAAGaGtgaATagcaATGaaagtg	4
wzm	tccTAAGGtaGcgAaaaATGtataaa	3
wzt	actatAGGgaagaAgtatttgATGaatgat	8
wbhA	tacaAAGGttcTtATatATGtttaat	2
wbhB	aacTAAcGgaGTtgaaATGgtaaat	4
ddhD	tatTcAGaaatatagtttATGgaattt	11
ddhA	ttagttGGAGtcatataATGaaagta	7
ddhB	aaaTttGGgaGTaataATGaatcca	4
ddhC	cgcacgaGcagGtaaagagtaATGtcaaaa	9
wbhC	taaaTttaGgatGtaatatATGaaatcc	6
wbhD	catTAAGataagGccagataATGagtgat	7
wbhE	aatTAtGGAGcTaaaaaATGataaat	5
wbhF	atcTAAtagtGaatATGaatatt	4
wbhG	aatttttgggtttttaaaaaATGaaaaaa	-
wbhH	caacAAtGAtGTaaTattcttacATGcgagat	8
wbhI	tatTAAttAGGtaaGgcaATGgacatt	3
wbhJ	taaagaAAGacaGGTtttaaATGacattt	5
wbhK	tgcTAgaGAaattgaaaaATGatatcc	9
wbhL	ccATtAccAGGTttaCtcctATGcacttg	4
wbhM	aaaTAAGtgaGaaaaATGaagtat	4
wbhN	agtatAGGAGtattttTTGtcgatt	6

Capital letters indicate one of the following sites: potential Shine-Dalgarno (SD) sites (bases that are identical to TAAGGAGGTGATC) or the presumed initiation codon (also underlined), either ATG or TTG. The termination codon of the preceding gene is indicated by a bar above, if it is in the region shown. The space between the last consensus base of the Shine-Dalgarno sequence and the initiation codon is also shown for reference.
a: sequence corresponds to Appendix 1.

Figure 5.27: Protein over-expression of ORFs within the *wbh* region.

The strain E2096 (pGP1-2) harbouring the plasmids pPM4901, pPM5515-pPM5520, pPM5522, pPM5524, pPM5525-pPM5531, pPM5533, pPM5534, pGEM5z$^+$ and pBluescript SK$^+$ were induced at 42°C to express protein from the T7 promoter as described in Materials and Methods (section 2.13.2). Whole cell lysates were electrophoresed on a SDS-15% polyacrylamide gel which was then stained with Coomassie Brilliant Blue stain and destained in 5% acetic acid. * indicates the over-expressed protein in each lane. Migration positions of the molecular mass standards (Pharmacia) are indicated on the left side (in kilodaltons): soybean trypsin inhibitor (20.1), carbonic anhydrase (30), ovalbumin (43), bovine serum albumin (67) and phosphorylase b (94).

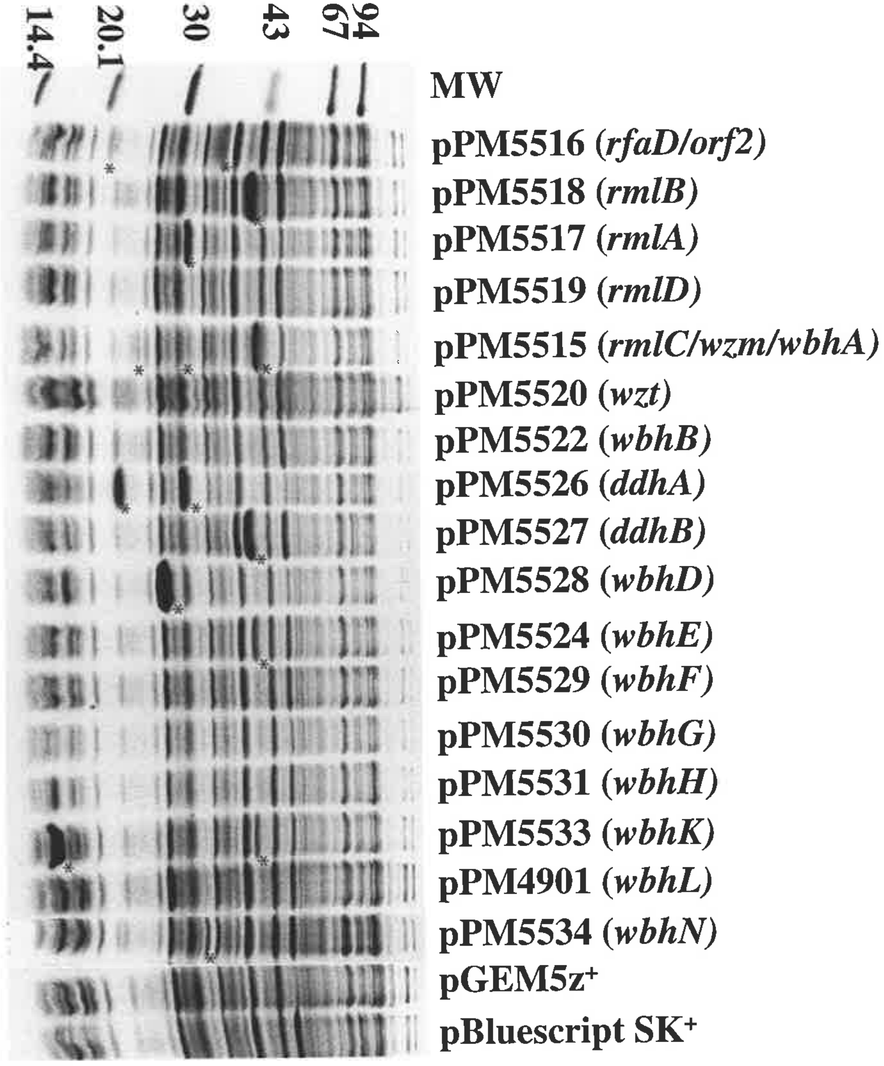

14.4
20.1
30
43
67
94
MW
pPM5516 (*rfaD/orf2*)
pPM5518 (*rmlB*)
pPM5517 (*rmlA*)
pPM5519 (*rmlD*)
pPM5515 (*rmlC/wzm/wbhA*)
pPM5520 (*wzt*)
pPM5522 (*wbhB*)
pPM5526 (*ddhA*)
pPM5527 (*ddhB*)
pPM5528 (*wbhD*)
pPM5524 (*wbhE*)
pPM5529 (*wbhF*)
pPM5530 (*wbhG*)
pPM5531 (*wbhH*)
pPM5533 (*wbhK*)
pPM4901 (*wbhL*)
pPM5534 (*wbhN*)
pGEM5z+
pBluescript SK+

Figure 5.28: Detection of *wbh* gene product using L-[^{35}S]-methionine labelling.

The strain E2096 (pGP1-2) harbouring the plasmids pPM5519, pPM5520, pPM522, pPM5530, pPM5531, pPM5529, pGEM5z^{+} and pBluescript SK^{+} were induced at 42°C to express protein from the T7 promoter. The proteins were radioactively labelled with L-[^{35}S]-methionine and whole cell lysates were electrophoresed on a SDS-15% polyacrylamide. The gel was stained with Coomassie Brilliant Blue stain and destained in 5% acetic acid before drying down onto Whatmann 3MM paper. X-ray film was exposed to the gel and developed after 72 h. * indicates the over-expressed protein in each lane. Migration positions of the molecular mass standards (Pharmacia) are indicated on the left side (in kilodaltons): soybean trypsin inhibitor (20.1), carbonic anhydrase (30), ovalbumin (43), bovine serum albumin (67) and phosphorylase b (94).

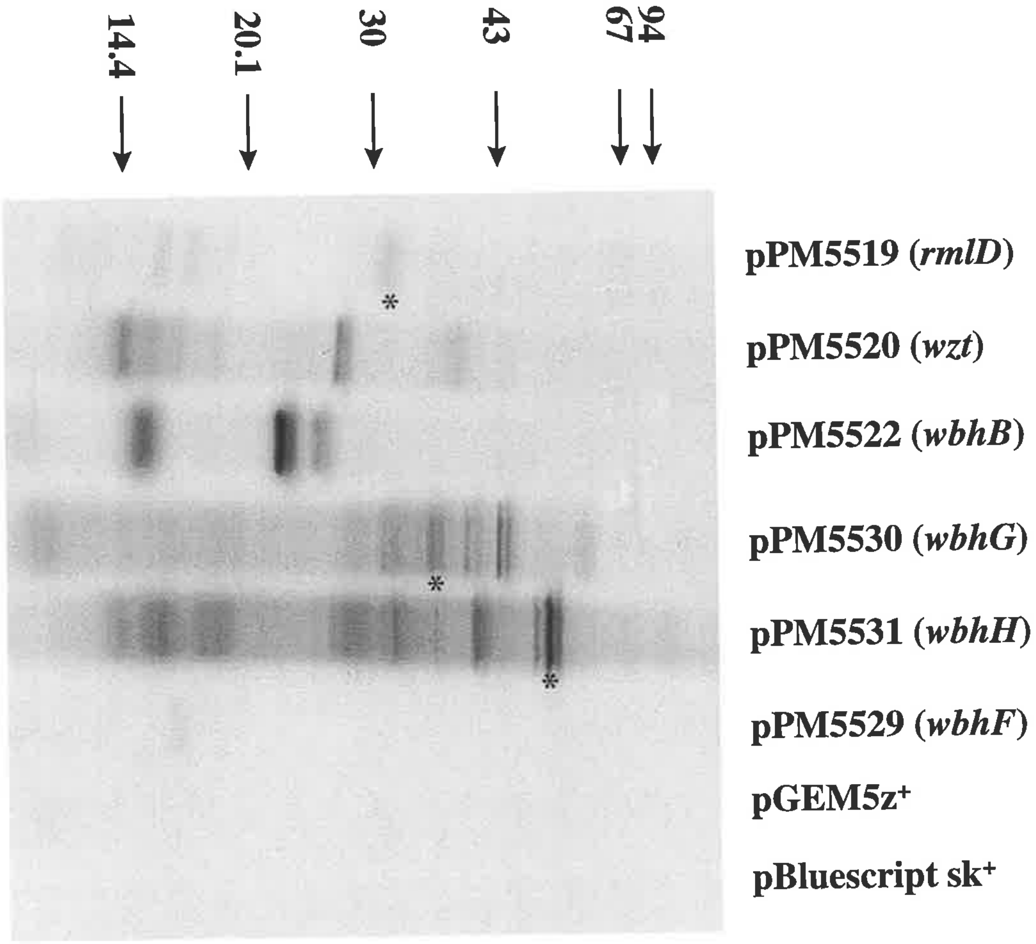

14.4
20.1
30
43
67
94
*
pPM5519 (*rmlD*)
pPM5520 (*wzt*)
pPM5522 (*wbhB*)
pPM5530 (*wbhG*)
pPM5531 (*wbhH*)
pPM5529 (*wbhF*)
pGEM5z+
pBluescript sk+

Table 5.14: Predicted Molecular Weight of ORFs in the *wbh* region of *V. anguillarum* O1

ORF	Location in Sequence	Predicted Molecular Weight (kDa)[a]	#amino acids	Detected by T7 overexpression	Size of Protein detected (kDa)
rfaD/gmhD	complementary strand, 2606-1667	35	313	+	35
rmlB	2928-3993	40	354	+	40
rmlA	4016-4894	32	292	+	30
rmlD	4891-5778	32	294	+	32
rmlC	5781-6326	21	181	+	22
wzm	6483-7287	31	268	+	30
wzt	7304-8642	50	446	-	-
wbhA	8670-9785	44	371	+	40
wbhB	11141-12790	64	549	-	-
ddhD	12930-13901	37	324	+	21
ddhA	13927-14700	32	264	+	29
ddhB	14703-15785	41	360	+	40
ddhC	15785-17098	49	438	-	-
wbhC	17115-18812	60	567	-	-
wbhD	18903-19532	24	204	+	25
wbhE	19522-20631	40	365	+	40
wbhF	20689-21075	15	130	-	-
wbhG	21072-22008	36	312	+	36
wbhH	22062-23408	51	447	+	49
wbhI	23622-24020	15	133	-	-
wbhJ	24038-24493	16	153	+	17
wbhK	24490-25605	41	369	+	41
wbhL	25700-26788	42	363	-	-
wbhM	26798-27712	36	303	-	-
wbhN	27880-28763	34	294	+	32

[a]: According to Kyte and Doolittle (1982), as implemented in PROSIS
+: protein was expressed, -: no protein was detected
nt: corresponds to sequence in Appendix 1.

correlate to a truncated protein for *ddhD* which has a predicted molecular weight of 36 kDa (Fig. 5.27). Analysis of the clone pPM5526 showed that the insert due to PCR error does not contain the entire *ddhD* gene. Those gene products that could not be detected by this approach were detected by labelling with L-[^{35}S]-methionine in the same T7 expression system (Fig. 5.28). Using this method, gene products could be detected for *rmlD* (32 kDa) (pPM5519); *wbhG* (36 kDa) (pPM5530); and *wbhH* (49 kDa) (pPM5531). Gene products could not be detected for *wzt*, *wbhB*, *wbhF* or *wbhL*. Plasmids containing the genes *ddhC*, *wbhC*, *wbhI*, and *wbhM* were not tested in these assays.

5.3 Summary and Discussion

In this chapter, the *wbh* operon, encoding the genes required for O-antigen biosynthesis in *V. anguillarum* O1 was cloned using cosmid libraries. The *wbh* region contained on three cosmid clones was sequenced and found to consist of 30627 bp. Analysis of the sequenced revealed the presence of 24 complete ORFs. The ORFs were analysed using BLAST searches and numerous sequences homologous to polysaccharide biosynthesis genes were found. Genes responsible for the biosynthesis of rhamnose and dideoxyhexose sugars, O-antigen transport and sugar transferases were found within the region. This has allowed putative functions according to homology to be assigned for some of the ORFs (Fig. 5.29; Table 5.15).

Numerous potential promoters were found throughout the region however not all of these were ideal with respect to the consensus sequence and optimal spacing for the (-35/-10) boxes. Analysis of the sequence revealed the presence of only one potential terminator after the ORF designated *wbhN*. Further investigation is required to determine which promoters are functional in this region.

Figure 5.29: The O-antigen biosynthesis operon of *V. anguillarum* O1.

The putative *V. anguillarum* O1 *wbh* region encodes for 24 ORFs. The ORFs have been designated names based on the nomenclature scheme for polysaccharide genes proposed by Reeves *et al.*, (1996). The arrows indicate the direction of transcription. Putative functions have been assigned to some ORFs and are labelled accordingly.

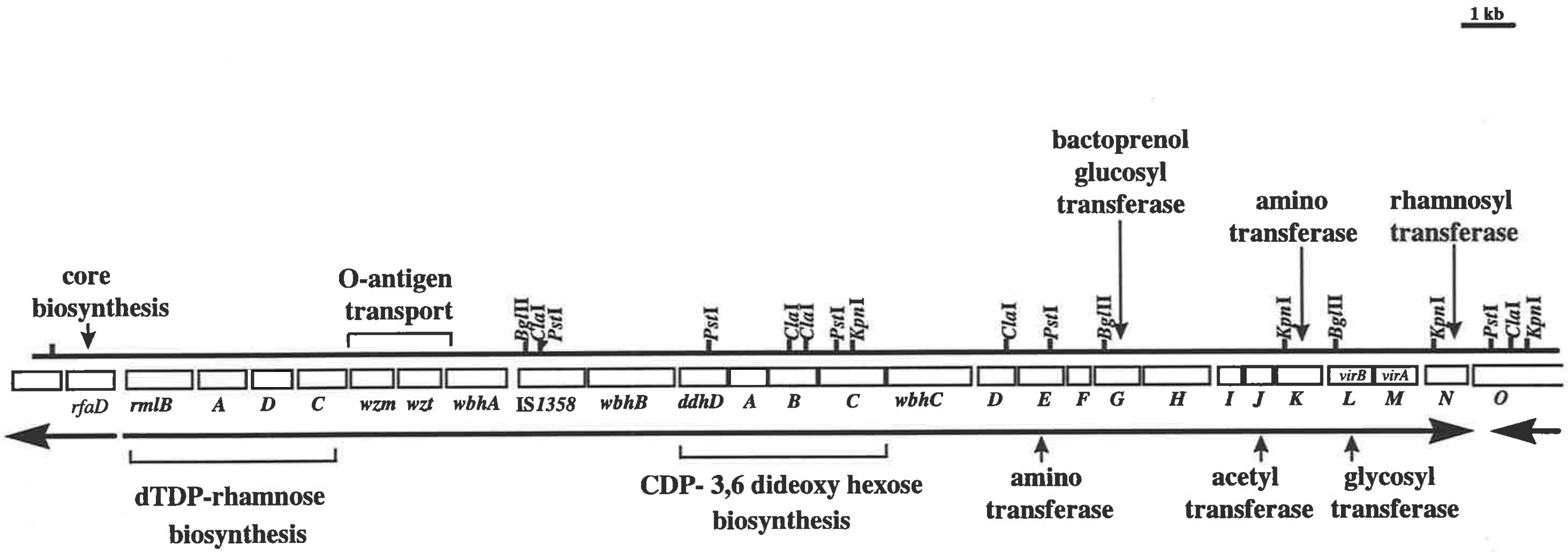

1 kb
core biosynthesis
O-antigen transport
bactoprenol glucosyl transferase
amino transferase
rhamnosyl transferase
BglII
ClaI
PstI
PstI
ClaI
ClaI
PstI
KpnI
ClaI
PstI
BglII
KpnI
BglII
KpnI
PstI
ClaI
KpnI
virB
virA
rfaD
rmlB
A
D
C
wzm
wzt
wbhA
IS1358
wbhB
ddhD
A
B
C
wbhC
D
E
F
G
H
I
J
K
L
M
N
O
dTDP-rhamnose biosynthesis
CDP- 3,6 dideoxy hexose biosynthesis
amino transferase
acetyl transferase
glycosyl transferase

Table 5.15: Summary of the *wbh* genes found in *V. anguillarum* O1

ORF	Putative Function	Similarity to	Accession Number
rfaD/gmhD	core biosynthesis	*V. cholerae* O1 *rfaD*	X59554[a]
rmlB	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB010415[a]
rmlA	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB010415[a]
rmlD	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB002668[a]
rmlC	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB010415[a]
wzm	O-antigen export (channel)	*B. subtilis tagG*	P42953[b]
wzt	O-antigen export (energiser)	*En. faecalis* ATP binding protein	AF071085[a]
wbhA	unknown		
IS*1358*	insertion sequence		
wbhB	unknown		
ddhD	CDP-dideoxyhexose pathway	*Y. pseudotuberculosis ascD*	P37911[b]
ddhA	CDP-dideoxyhexose pathway	*S. typhimurium rfbF*	P26396[b]
ddhB	CDP-dideoxyhexose pathway	*Y. pseudotuberculosis*	L33181[a]
ddhC	CDP-dideoxyhexose pathway	*Y. pseudotuberculosis*	E47070[a]
wbhC	acetolactate synthetase	*E. coli ilvG*	P00892[b]
wbhD	unknown		
wbhE	perosamine synthetase	*M. thermoautotrophicum*	AE000818[a]
wbhF	unknown		
wbhG	bactoprenol glucosyl transferase	*B. subtilis csbB*	Q45539[b]
wbhH	unknown		
wbhI	unknown		
wbhJ	acetyl transferase	*B. pertussis blpB*	X90711[a]
wbhK	transaminase	*S. venezuelae*	AF079762[a]
wbhL	glycosyl transferase	*V. anguillarum virB* *H. influenazae lgtD*	L08012[a] Q57022[b]
wbhM	unknown	*V. anguillarum virA*	L08012[a]
wbhN	rhamnosyl transferase	*S. flexneri rfbF*	S51266[c]

a. Genbank; b. SWISS PROT; c. PIR.

Interestingly, there are 3 ORFs in the complementary strand which are homologous to *orf2*, *rfaD* and *trsG*. The organisation of the *wbh* region at the 5' and 3' ends is identical to the O-antigen biosynthesis operons of *V. cholerae* O1, O139 (Stroeher *et al.*, 1998) and V. *anguillarum* O2 (this study, Chapter 4).

Protein expression assays were performed using plasmids containing most of the *wbh* genes and gene products were produced from the different ORFs which correlated to the predicted molecular weight (Table 5.14).

Chapter Six

Characterisation of the *wbh* operon by mutational analysis

6.1 Introduction

Analysis of O-antigen biosynthesis loci has relied on the isolation and characterisation of mutants defective in the production of this component of LPS. Determination of enzyme function has allowed the biosynthetic pathways for many of the different sugars found within various O-antigens to be elucidated. For example the enzymatic mechanisms of the RfbH/DdhD and RfbI/DdhD proteins involved in the synthesis of 3,6 dideoxyhexose in *Y. pseudotuberculosis* have been recently determined with the development of better assays (Thornson *et al.*, 1994).

Mutations in O-antigen biosynthesis genes have varying effects on the structure of LPS, depending on the gene which is defective. LPS phenotypes associated with mutation in various genes are described as smooth (S-LPS), semi rough (SR-LPS), rough (R-LPS) or unregulated. S-LPS constitutes lipid A, core sugars and O-antigen; SR-LPS consists of lipid A, core sugars and a single O-antigen repeat unit; R-LPS contains lipid A and core sugars; and unregulated LPS is S-LPS which does not display a modal chain length. These differences are observable by silver staining LPS separated by SDS-PAGE.

In this chapter, *wbh*::Tn5*phoA* (Cm^R) transposon insertion mutants are characterised by silver staining, Western immunoblotting, serum cross-absorption studies using LPS specific sera and complementation.

6.2 Results

6.2.1 Characteristics of *wbh*::Tn*phoA* (Cm^R) mutants

The mutants described in this study were constructed by transposon mutagenesis using a chloramphenicol derivative of Tn*5phoA* (D. Milton, personal communication). The mutants were isolated as their LPS no longer reacted by Western immunoblots to a polyclonal antiserum raised against formalin-killed *V. anguillarum* O1. The insertion site was sequenced to reveal the location of the mutation within the *wbh* region (D. Milton, affiliation, personal communication). The mutants were isolated from the *V. anguillarum* O1 strain NB10, and were designated KM2 (*otnE*), KM16 (*wbhE*), KM97 (*wzm*) and KM110 (*wbhB*) (D. Milton, personal communication). Table 6.1 describes the location of the transposon insertion in the mutants.

6.2.2 LPS profile of *wbh* mutants

On SDS-PAGE gels LPS migrates as a ladder of bands, which represent individual S-LPS molecules of different chain lengths which are ligated to lipid A-core oligosaccharide (Palva and Måkelå, 1980). The wildtype O1 strains, 85-3954-2 and NB10 have an identical S-LPS pattern (data not shown). Analysis of the LPS from *wbh* mutants by SDS-PAGE and silver staining showed that in comparison to the wildtype strain (NB10) the mutants have either an altered migration profile, which was unexpected given the method of isolation, or LPS with no O-antigen (ie R-LPS) (Fig. 6.1).

KM2 has a mutation in a gene which is homologous to *otnE* from *V. cholerae* O139. This gene maps outside of the sequenced region from this study and will be discussed later in this chapter (section 6.2.10). The LPS profile of KM2 shows that the

Table 6.1: Tn*phoA* (Cm^R)insertion mutants of *V. anguillarum* O1

Mutant	Gene (*wbh*)	Location of the gene[a]	Transposon insertion point[b]
KM16	*wbhE*	19522-20631	20357
KM97	*wzm*	6483-7287	6924
KM110	*wbhB*	11141-12790	12426
KM2	*otnE*	NA	NA

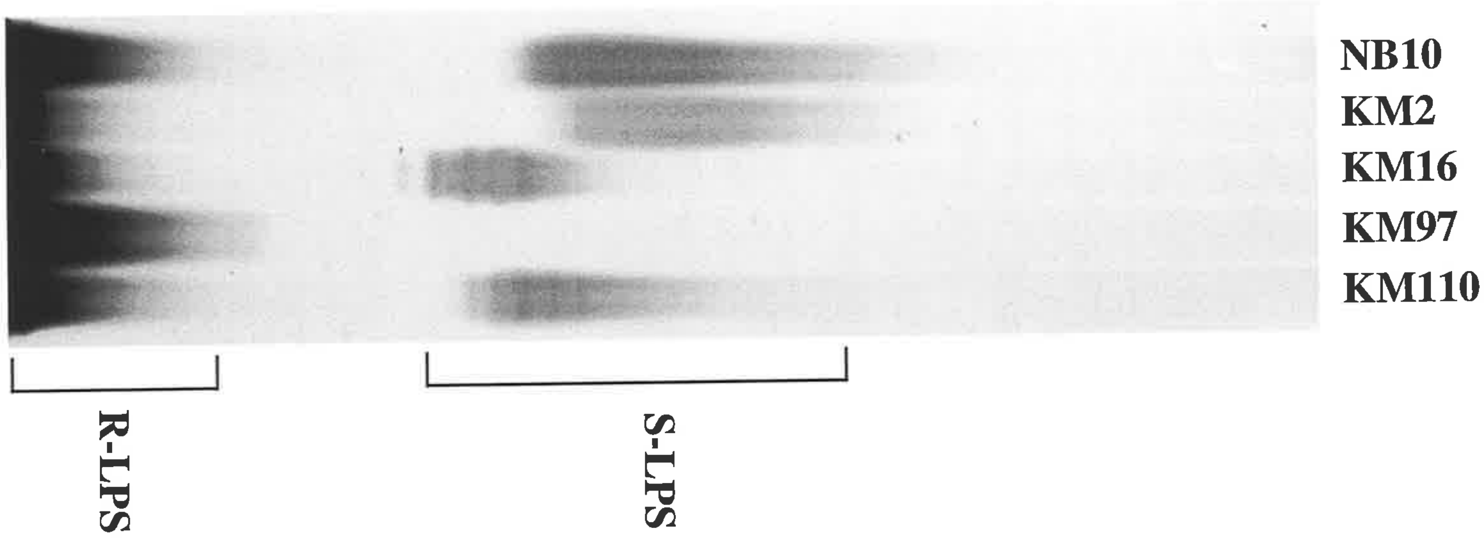
a. corresponds to nucleotide number in Appendix 1.

b. determined by sequencing of the Tn*phoA* insertion point (D. Milton, affiliation, personal communication) and comparison to the *wbh* sequence in Appendix 1.

NA: data not available.

Figure 6.1: Analysis of *V. anguillarum* O1 LPS/O-antigen mutant lipopolysaccharides.

This shows a silver stained 15% SDS polyacrylamide gel with lipopolysaccharide prepared from the indicated strains by proteinase K treatment (Section 2.15.1). The first lane contains the NB10 wildtype strain. The remaining lanes contain the Tn5*phoA* (Cm^R) mutants derived from NB10 indicated at the top of the figure. The S-LPS and R-LPS are indicated on the right side of the figure. Samples represent ~1 x 10^8 cells.


NB10
KM2
KM16
KM97
KM110
R-LPS
S-LPS

lipid A-core linked O-chains has a slight increase in apparent molecular weight/modal length compared to the wildtype. No further analysis of this mutant was performed.

The second mutant, KM16 has an insertion in the gene *wbhE*, which shows homology to *rfbE/wbeE*, the perosamine synthetase from *E. coli* O157 and *V. cholerae* O1 (Section 5.2.7.4). The LPS profile is different to the wildtype as the S-LPS molecules have a decreased molecular weight (Fig. 6.1). A similar LPS pattern was observed for KM110, however the relative decrease in S-LPS size was not as great (Fig. 6.1). The insertion in KM110 is located in the gene designated, *wbhB*. *wbhB* has not been assigned a function as no DNA or protein homologues were found in database searches (Section 5.2.7.7). Another interesting observation when comparing KM16 and KM110 to the wildtype is that the mutants have a slightly more defined banding pattern than the wildtype, which appears fuzzy on the silver stained SDS-polyacrylamide gel (Fig. 6.1). This fuzzy appearance may indicate either acetylation or that some other variable modification is absent in these mutants.

The LPS of mutant KM97 which has an insertion in the *wzm* gene displays a different phenotype to the other three LPS mutants studied. Silver staining showed that KM97 had R-LPS, ie. lacking O-antigen (Fig. 6.1). Wzm has a proposed transport function as it demonstrates homology to ABC transporters. A mutation in this gene would presumably prevent O-antigen being expressed on the surface as a lipid A, core sugar linked molecule.

6.2.3 Western blot analysis of the *wbh* mutants, KM16, KM97 and KM110

Analysis of LPS samples from the mutants KM16, KM97 and KM110 by Western immunoblotting showed that polyclonal antiserum specific to LPS generated against the *V.*

anguillarum O1 strain, 85-3954-2 (anti-VagO1) (Chapter 3, Section 3.2.1), when used at a dilution of 1:1000 was able to recognise the altered LPS of the mutants KM16 and KM110, and the reaction was equivalent to that for the wildtype LPS. This suggests that the O-antigen expressed by the wildtype strains (85-3954-2 and NB10) and mutants (KM16 and KM110) have similar antigenic epitopes. KM97 did not react to the antiserum, as expected since it is has R-LPS (Fig. 6.2). The lack of reactivity with R-LPS suggests most antibodies in the antiserum react with the O-antigen component of the LPS.

6.2.3.1 Serum Absorption studies

To determine if the O-antigen expressed by the wildtype and mutants were similar, absorption of the anti-VagO1 serum was performed using heat-killed bacteria of strains KM16, KM97, and KM110. The antiserum absorbed with heat killed KM97 was used in Western immunoblotting analysis against the wildtype strains and *wbh* mutants described above. The reaction previously observed against the mutants KM16 and KM110 using the unabsorbed antibodies had been abolished as no reaction to the LPS of these mutants could be detected with the absorbed serum (Fig. 6.3). Reaction to the LPS from wildtype strains remained unchanged (Fig. 6.3). Therefore KM97 which produces R-LPS has epitopes that remove antibodies cross-reactive with the altered LPS produced by KM16 and KM110. This data in conjunction with the silver staining (Section 6.2.2) suggested that KM97 was in fact producing a form of O-antigen (O-antigen hapten) which is apparently not linked to lipid A. It also suggests that the O-antigen hapten which is produced by KM97 while sufficiently different to the wildtype LPS is similar to KM16 and KM110 S-LPS O-antigen as the absorption removed only antibodies against the altered O-antigens produced by the latter mutants.

Figure 6.2: Western immunoblot of LPS from *V. anguillarum* O1 wildtype and Tn5*phoA* (Cm^R) mutant strains.

Proteinase K whole cell lysates (Section 2.15.1) of the indicated strains were electrophoresed on a SDS 15% polyacrylamide gel and transferred to nitrocellulose (Section 2.13.4). LPS was detected using antiserum generated against 85-3954-2 (*V. anguillarum* O1) as described in Chapter 3. Lanes contain the samples (equivalent to 1 x 10^8 cells) indicated at the top of the figure. The O-antigen and lipid A-core oligosaccharide are indicated on the right side of the figure.

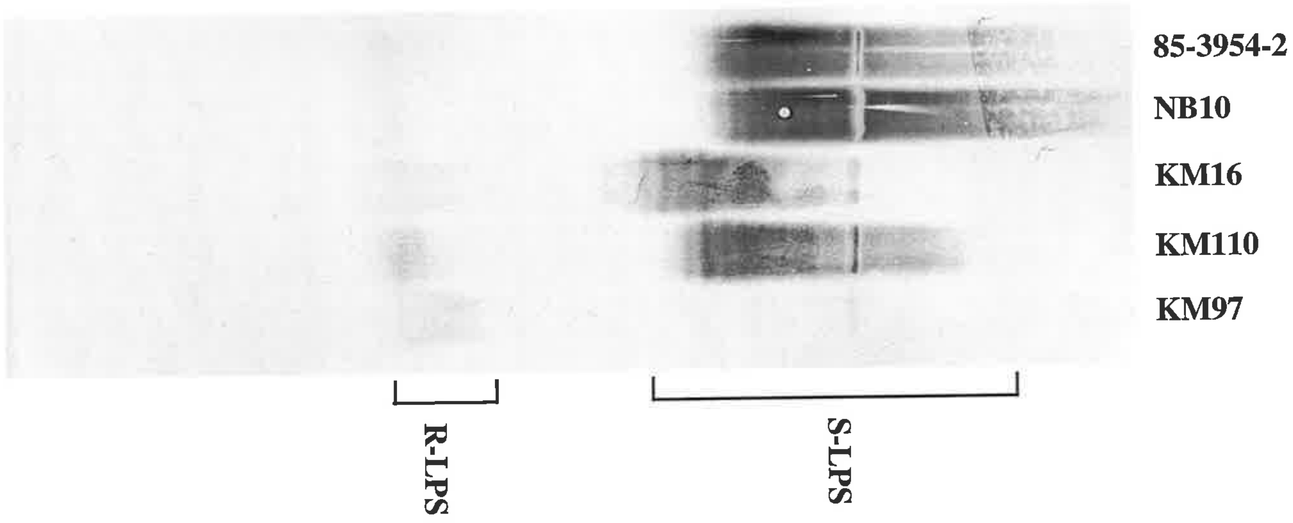
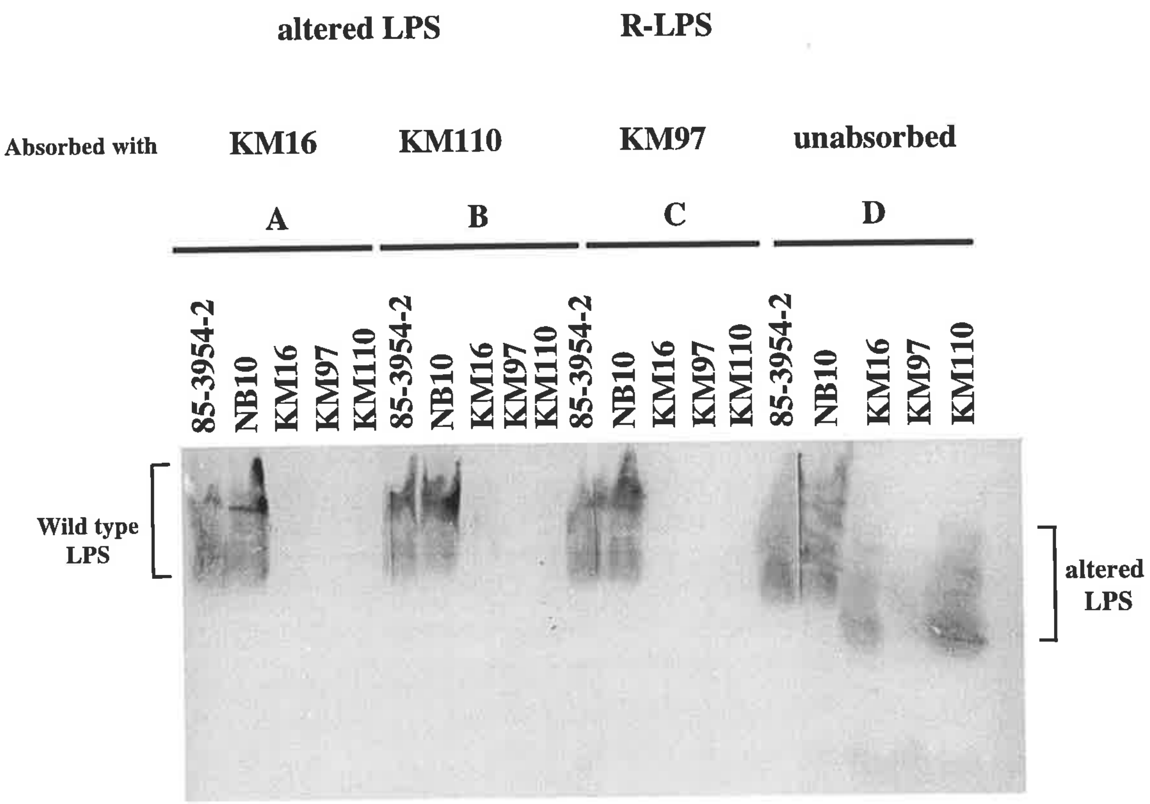

85-3954-2
NB10
KM16
KM110
KM97
R-LPS
S-LPS

Figure 6.3: Western immunoblot of LPS from various *V. anguillarum* strains using *V. anguillarum* O1 serum absorbed against KM16, KM97 and KM110.

Proteinase K whole cell lysates (Section 2.15.1) were electrophoresed on a SDS 15% polyacrylamide gel and transferred to nitrocellulose (Section 2.13.4). LPS was detected using an antiserum generated against 85-3954-2 (*V. anguillarum* O1) which was either unabsorbed (D), or absorbed with heat killed KM16 (A), KM97 (C) or KM110 (B). Lanes contain the samples (equivalent to 1 x 10^8 cells) indicated at the top of the figure. The position of wild type and altered LPS are indicated on the figure.


altered LPS
R-LPS
Absorbed with
KM16
KM110
KM97
unabsorbed
A
B
C
D
85-3954-2
NB10
KM16
KM97
KM110
85-3954-2
NB10
KM16
KM97
KM110
85-3954-2
NB10
KM16
KM97
KM110
85-3954-2
NB10
KM16
KM97
KM110
Wild type LPS
altered LPS

The anti-VagO1 serum was also absorbed with KM16 and KM110 and used in Western immunoblot analysis (dilution 1:1000) against the same strains. Reaction against LPS from the mutants was not observed, however the antisera did recognise the wildtype O-antigen as well as the unabsorbed serum, suggesting the absorption was not removing antibodies against these epitopes (Fig. 6.3).

These data indicate that the O-antigen of the S-LPS produced by the mutants KM16, KM110 and the O-antigen hapten of KM97 are different to the majority of S-LPS O-antigens expressed by the wildtype strains 85-3954-2 and NB10. However, the anti-VagO1 antiserum generated does recognise the altered O-antigen produced by the mutants suggesting that the wildtype strain 85-3954-2 expresses these altered O-antigens at either low levels or that the O-antigens share common epitopes.

6.2.4 Immunogold Electron microscopy of *V. anguillarum* and *wbh* mutants

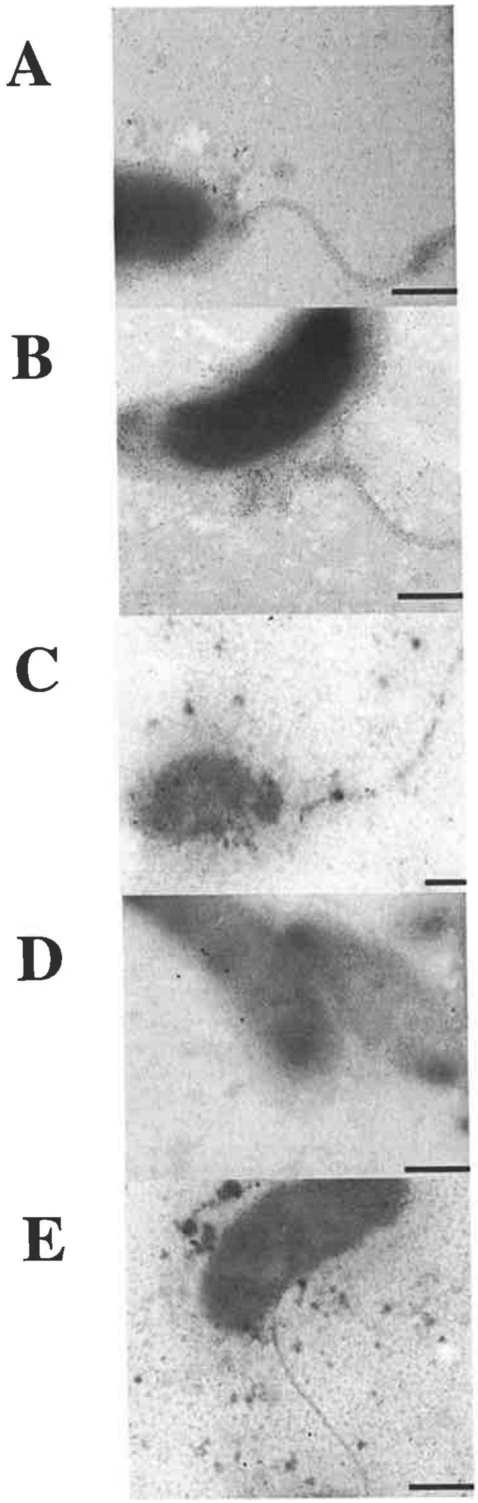
To further examine the reaction of LPS observed by Western immunoblotting with the various unabsorbed and sera absorbed with, KM16, KM97 and KM110, immunogold electron microscopy was undertaken to detect O-antigens on the surface of the *V. anguillarum* wildtype and mutant strains.

The two wildtype strains NB10 and 85-3954-2 appeared to express the same level of O-antigen according to protein-A gold binding to the cell surface (Fig. 6.4 (A) (B)). As expected the R-LPS mutant, KM97 did not bind protein A-gold particles (Fig. 6.4 (C)). The altered LPS mutants, KM16 and KM110 also did not exhibit binding of protein A-gold particles (Fig. 6.4 (D) (E)), a result which was unexpected as the antiserum reacted to their respective LPS in Western immunoblotting (Section 6.2.3). IEM was also performed using

Figure 6.4: Immunogold electronmicrographs of 85-3954-2, NB10, KM2, KM16 and KM110.

Electron micrographs showing binding of polyclonal antibodies generated against *V. anguillarum* O1 (85-3954-2) to LPS of *V. anguillarum* O1 wildtype strains and Tn5*phoA* (Cm^R) mutants: KM16, KM97 and KM110, detected with protein A-gold (10 nm). Antibodies were used at 1:50 dilution, and the magnification microscope (TM100, Philips) ranged from 27.5 to 29K. The bar represents 500 nM.

A: NB10 (28.5 K)
B: 85-3954-2 (29 K)
C: KM16 (27.5 K)
D: KM97 (29 K)
E: KM110 (29 K)


A
B
C
D
E

the antisera absorbed with KM16, KM97 or KM110. The same result was obtained with these sera as the unabsorbed anti-VagO1 serum (data not shown).

The inability of the anti-VagO1 sera to detect the LPS of the mutants, despite their reaction with the antiserum by Western immunoblotting, indicated that immunogold electron microscopy may have been too insensitive and therefore labelling could not be detected.

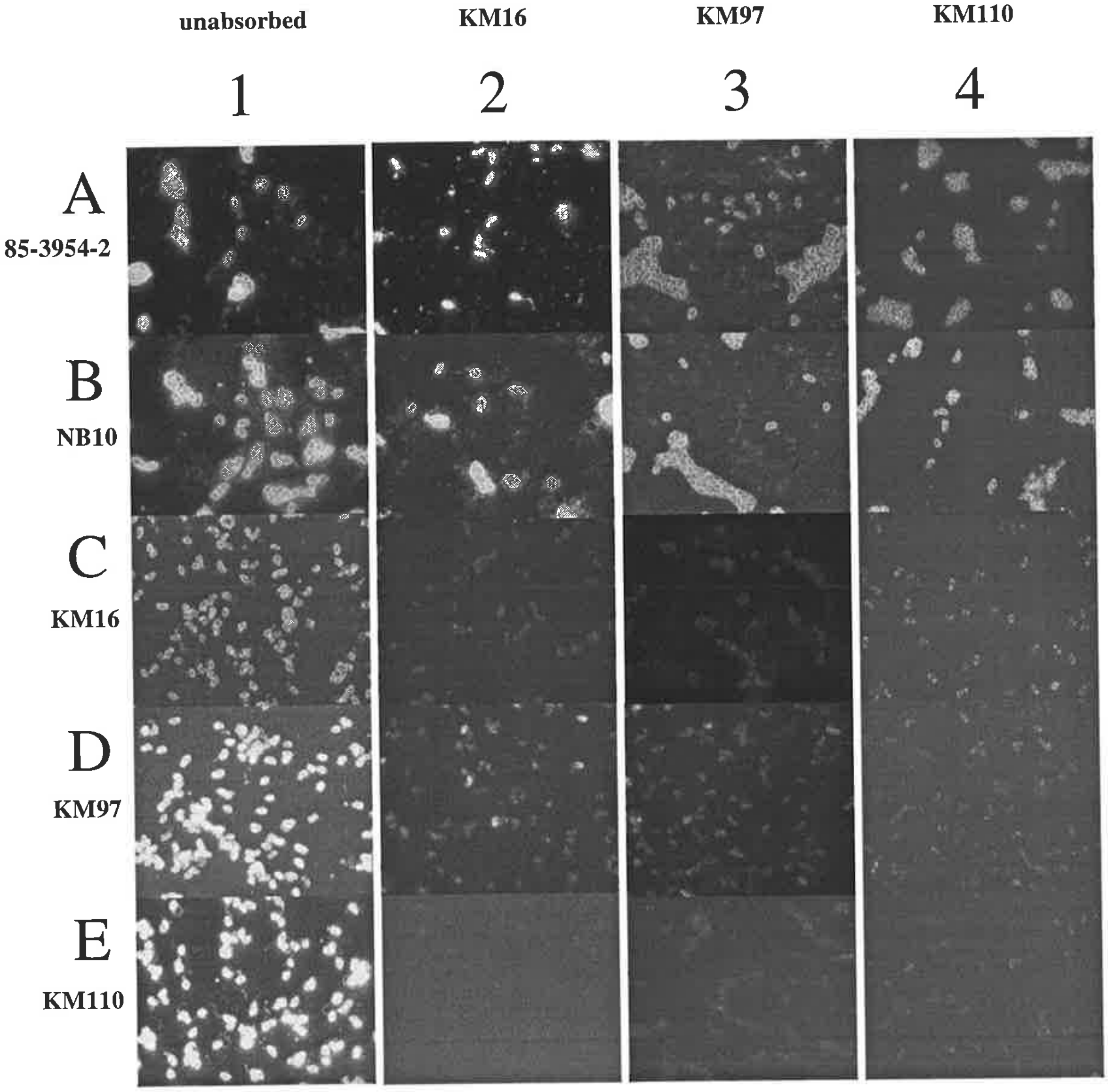
6.2.5 Immunofluorescence Microscopy

Due to the apparent insensitivity of immunogold electron microscopy, immunofluorescence was undertaken to visualise the O-antigen on the surface of the various *V. anguillarum* strains and to confirm the Western immunoblotting data which suggested that the LPS molecules of the wild type and mutants had both cross-reactive and specific epitopes.

Using the unabsorbed antiserum (anti-VagO1), the wildtype and the mutant strains were labelled over the entire bacterial cell surface, as expected (Fig. 6.5 (A-E, column 1)). Unexpectedly, KM97 could also be labelled. This indicated that the haptenic form of O-antigen expressed by KM97, although not lipid A-core linked, was expressed on the cell surface (Fig. 6.5 (D1)). When the anti-VagO1 serum absorbed with the mutant KM97 was used, wildtype strains were labelled but little or no labelling of the mutants KM16 and KM110 was observed (Fig. 6.5 (A3, B3, C3, E3)). A low level of labelling was observed on a proportion of KM97 cells (~<10%) (Fig. 6.5 (D3)). The labelling appeared to be concentrated to one pole (Fig. 6.5 (D2)). When the assay was performed using anti-VagO1 sera absorbed with KM16 and KM110 a similar result was obtained; the wildtype strains were labelled and the mutants showed little or no fluorescence (Fig. 6.5 (A-E, column 2, column 4)). KM97 showed a low level of polar fluorescence on some bacteria with the

Figure 6.5: Detection of LPS/O-antigen on *V. anguillarum* cells by indirect immunofluorescence.

Cells of *V. anguillarum* strains 85-3954-2 (A), NB10 (B), KM16 (C), KM97 (D), KM110 (E) were stained with anti-VagO1 serum (1:200 dilution) that was either unabsorbed (1), or absorbed with heat killed KM16 (2), KM97 (3) or KM110 (4). Bound antibodies were detected with an FITC-conjugated goat anti-rabbit immunoglobulin secondary antibody. Images were photographed with a 100 x objective.


unabsorbed
KM16
KM97
KM110
1
2
3
4
A
85-3954-2
B
NB10
C
KM16
D
KM97
E
KM110

anti-VagO1 serum absorbed against KM16 (~<10%) (Fig. 6.5 (D3)). Little or no reaction against KM97 was observed using the anti-VagO1-KM110 absorbed serum (Fig. 6.5 (D4)).

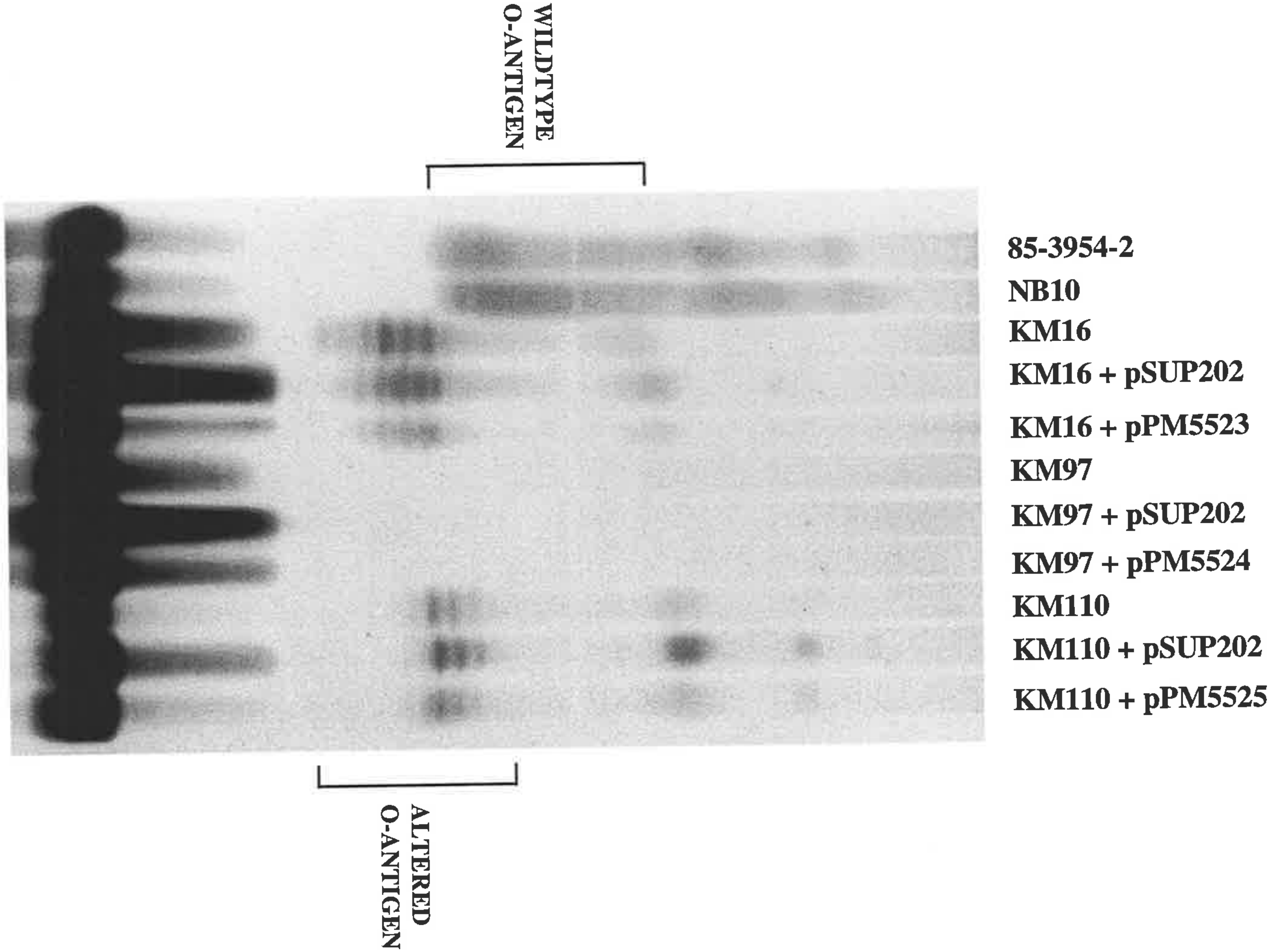
The IF data obtained correlates with the Western immunoblot analysis (Fig. 6.2, Fig. 6.3) and indicates that the O-antigen expressed by the mutants possesses antigenically different epitopes to the predominant VagO1 LPS O-antigen, although the latter O-antigens may have common epitopes. However, it is possible that the wildtype strains expresses both the wildtype and mutant (altered) O-antigens.

6.2.6 Complementation studies of KM16, KM97 and KM110.

To determine if complementation of the Tn*phoA* (Cm^R) mutation in the O-antigen mutants KM16, KM97 and KM110 would restore wildtype O-antigen production, plasmids containing the appropriate gene were constructed in the vector pSUP202. *V. anguillarum* strains are naturally resistant to ampicillin and the mutants have a chloramphenicol cartridge within the transposon, hence pSUP202 was used to allow selection of tetracycline resistant ex-conjugants. The plasmids pPM5523 (*wbhE*), pPM5524 (*wzm*) and pPM5525 (*wbhB*) (Table 2.3) were constructed by PCR using oligonucleotide primers (#3031-#3036, Table 2.4) containing *Eco*RI and *Pst*I sites, as these were unique to the vector and not present in the genes. The plasmids pPM5523, pPM5524 and pPM5525 were introduced into S17-1 and then conjugated into KM16, KM97 and KM110, respectively. The resultant ex-conjugants were tested for the presence of the complementing plasmid, and proteinase K treated whole cell lysates were made to assess the LPS phenotype. The LPS preparations were electrophoresed on a SDS-15% polyacrylamide gel and LPS was detected by silver staining (Fig. 6.6). Complementation was not achieved, suggesting the Tn*phoA* mutations are polar and a single gene is not able to restore the correct function. Attempts were also

Figure 6.6: Complementation of KM16, KM97 and KM110.

This shows a silver stained 15% SDS polyacrylamide gel with lipopolysaccharide prepared from the indicated strains (Section 2.15.1). The first two lanes contain LPS from 85-3954-2 and NB10 wildtype strains. The remaining lanes contain the Tn5*phoA* (Cm^R) mutants derived from NB10 and constructed strains harbouring the appropriate plasmid as indicated at the top of the figure. The wildtype and altered O-antigens are indicated on the figure. Samples represent ~1 x 10^8 cells.


WILDTYPE
O-ANTIGEN
85-3954-2
NB10
KM16
KM16 + pSUP202
KM16 + pPM5523
KM97
KM97 + pSUP202
KM97 + pPM5524
KM110
KM110 + pSUP202
KM110 + pPM5525
ALTERED
O-ANTIGEN

made to conjugate the cosmids pPM4905 and pPM5512 (Section 5.2.1) into the mutants however this was not successful probably due to the size of each cosmid and the presence of pJM1 in *V. anguillarum* which has been shown to prevent conjugal entry of large plasmids (Singer *et al.*, 1992).

6.2.7 Composition of the LPS from *V. anguillarum* O1

The sugar analysis presented in this section was performed by M. Jansson and A. Weintraub (Stockholm University, Sweden).

Crude sugar analysis of ultra-pure LPS (Materials and Methods, section 2.15.2) from NB10 and 85-3954-2 indicated that both of these wild type *V. anguillarum* O1 strains contain glucose and rhamnose. Glucose is in greater proportion to rhamnose with the ratio being 10:1. The LPS preparation and analysis used does not separate the LPS core sugar oligosaccharide from the O-antigen component and hence it is unclear what proportion of the sugars are present in each component. The analysis also indicated the presence of glucosamine in the LPS (glucose:rhamnose:glucosamine (10:1:1)) (Table 6.2). No other sugars were found in this initial investigation of the LPS of *V. anguillarum* O1.

Table 6.2: Summary of the sugar analysis on *V. anguillarum* O1

	Ratio of sugars present:		
	glucose	**rhamnose**	**glucosamine**
strain			
wild type	10	1	1
KM16 (*wbhE*)	10	1	-
KM97 (*wzm*)	5	1	-
KM110 (*wbhB*)	2.5	1	-

-: sugar was not detected

Analysis of the sugar composition of the LPS from the transposon mutants KM16, KM97 and KM110 was also performed. The LPS in KM16 showed a glucose to rhamnose ratio of 10:1 with no glucosamine detected (Table 6.2). KM16 has an insertion in *wbhE*, encoding a putative perosamine (or amino sugar) synthetase (Chapter 5, Section 5.2.7.4).

Analysis of KM97 LPS sugars detected the presence of glucose and rhamnose in a ratio of 5:1. KM97 was shown by sequencing to be a mutant in *wzm*, whose gene product is likely to be involved in O-antigen transport. KM110, a mutant in *wbhB* which has no known homologues, had a glucose to rhamnose in a ratio of 2.5:1. Neither the LPS prepared from KM97 or KM110 had detectable glucosamine (Table 6.2).

6.2.8 The rhamnose biosynthesis genes

The sugar analysis revealed the presence of rhamnose in the LPS of *V. anguillarum* O1. The *wbh* region described in this study supports this finding as genes are present that synthesise dTDP-rhamnose, based on homology to previously described dTDP-rhamnose biosynthesis genes from *Actinobacillus actinomycetemcomitans, N. meningitidis, S. enterica* and *S. flexneri* (Section 5.2.7.1).

dTDP-rhamnose biosynthesis is known to proceed through the pathway described in Chapter 1 (Fig. 1.12). The first two steps are encoded by *rfbA*/*rmlA* and *rfbB*/*rmlB* with the last step involving two enzymes, *rfbC*/*rmlC* and *rfbD*/*rmlD*. The gene order does not follow the enzyme order in the biosynthetic pathway which is not uncommon for these genes with the order being *rmlBDAC* in *S. enterica, E. coli* K-12 (Reeves *et al.*, 1996) and *S. flexneri* (Macpherson *et al.*, 1995). The order of the dTDP-rhamnose genes in *V. anguillarum* O1 are *rmlBADC*.

The DNA sequences of *rmlBADC* were used in a search employing the BlastX and BlastN databases to find similar sequences. Significant homologies were found to numerous functionally identical and related proteins and are presented as CLUSTAL alignments and homology tables. The genes for these proteins also showed significant identity at the DNA level, which is also described below.

RmlA$_{Va}$ is the first enzyme in the pathway and encodes a glucose-1-phosphate thymidyl transferase. The RmlA$_{Va}$ protein shows a high level of similarity to identical proteins in *A. actinomycetemcomitans* (81.7% identity at the amino acid level, 72% identity at the DNA level), and *N. meningitidis* (81.5% identity at the amino acid level, 68% identity at the DNA level) (Fig. 6.7; Table 6.3). RmlA$_{Va}$ also shows similarity to the related proteins in *S. flexneri* (67.8% identity at the amino acid level, 63% identity at the DNA lecel) and *S. enterica* sv Typhiumurium (65.7% identity at the amino acid level, 62% identity at the DNA level).

The RmlB$_{Va}$ protein belongs to a family of dehydratases and is the second enzyme in the dTDP-rhamnose biosynthetic pathway. RmlB$_{Va}$ is significantly homologous to the same enzyme in *A. actinomycetemcomitans* (82.6% identity at the amino acid level, 74% identity at the DNA level), *Neisseria meningitidis* (80.1% identity at the amino acid level, 69% identity at the DNA level), *S. flexneri* (73.4% identity at the amino acid level, 67% identity at the DNA level) and *S. enterica* sv Typhimurium (73.1% identity at the amino acid level, 66% identity at the DNA level) (Fig. 6.8; Table 6.4).

It has been previously suggested that an important feature of RmlB and related proteins is the presence of an NAD binding domain (GXXGXXG) at the amino terminal end of the protein (Macpherson *et al.*, 1995). This site is important for bacterial enzymes which use NAD/NADP as a co-factor. RmlB$_{Va}$ was found to contain this essential motif (Fig. 6.8).

The next enzyme in the pathway is encoded by *rmlC$_{Va}$*. RmlC is a dTDP-4-keto-6-deoxy-D-glucose-3,5-eprimerase/synthetase. RmlC$_{Va}$ shows highest homology to RmlC from *S. flexneri* (69.7% identity at the amino acid level, 67% identity at the DNA level), *A.*

Figure 6.7: Alignment of RmlA$_{Va}$ with homologous proteins

```
VaRmlA      ---MKGIILAGGSGTRLYPITRGVSKQLLPIYDKPMIYYPLSTLMLAGIRDILIITTPED   57
AaRmlA      ---MKGIILAGGSGTRLYPITRGVSKQLLPVYDKPMIYYPLSVLMLAGVRDILIITTPED   57
NmRmlA      ---MKGIILAGGSGTRLYPITRGVSKQLLPVYDKPMIYYPLSVLMLAGIRDILVITAPED   57
SfRmlA      MKTRKGIILAGGSGTRLYPVTMAVSKQLLPIYDKPMIYYPLSTLMLAGIRDILIISTPQD   60
StRmlA      MKTRKGIILAGGSGTRLYPVTMAVSKQLLPIYDKPMIYYPLSTLMLAGIRDILIISTPQD   60
                ***************.*  *******.*********** *****.****.*..*.*

VaRmlA      NESFKRLLGDGSDFGIHLQYAIQPSPDGLAQAFLIGEEFIGNDSVCLVLGDNIFYGQSFS  117
AaRmlA      NESFKRLLGDGSEFGVNLQYAIQPSPDGLAQAFLIGEGFINGDSCCLVLGDNIFYGQNFT  117
NmRmlA      NASFKRLLGDGSDFGISISYAVQPSPDGLAQAFIIGEEFIGNDNVCLVLGDNIFYGQSFT  117
SfRmlA      TPRFQQLLGDGSQWGLNLQYKVQPSPDGLAQAFIIGEEFIGGDDCALVLGDNIFYGHDLP  120
StRmlA      TPRFQQLLGDGSQWGLNLQYKVQPSPDGLAQAFIIGEEFIGHDDCALVLGDNIFYGHDLP  120
            .  *..******  *. . * .***********.*** **. *  .**********.

VaRmlA      KTLKSAASREHGATVFGYQVKDPERFGVVEFDEQMRAISIEEKPLKPKSNYAVTGLYFYD  177
AaRmlA      QMLQQAVARPYGATVFGYLVKDPGRFGVVEFDENFKAVSIEEKPAQPKSNYAVTGLYFYD  177
NmRmlA      QTLKQAARQTHGATVFAYQVKNPERFGVVEFNENFRAVSIEEKPQRPKSDWAVTGLYFYD  177
SfRmlA      KLMDTAVNRESGATVFAYHVNDPERYGVVEFDDNGTPISLEEKPQQPKSNYAVTGLYFYD  180
StRmlA      KLMEAAVNKESGATVFAYHVNDPERYGVVEFDQKGTAVSLEEKPLQPKSNYAVTGLYFYD  180
            . .  *  .  ***** * *  * *.*****      .*.**** .***  *********

VaRmlA      NRVVELAKQVKPSARGELEITTLNEMYLNDGSLNVELLGRGFAWLDTGTHESLHEASSFV  237
AaRmlA      NRVVDFAKQVKPSARGELEITTLNEMYLKDGSLNVQLLGRGFAWLDTGTHESLHEGASFV  237
NmRmlA      NRAVEFAKQLKPSARGELEISDLNRMYLEDGSLSVQILGRGFAWLDTGTHESLHEAASFV  237
SfRmlA      NDVVEMAKNLKPSARGELEITDINRIYMDQGRLS-AMMGRGYACLDTDTSK-PIEASNFI  238
StRmlA      NSVVEMAKNLKPSARGELEITDINRIYMEQGRLSVAMMGRGYAWLDTGTHQSLIEASNFI  240
            *  *. **..**********. .* .*.  * *   ..***.* *** *     * . *.

VaRmlA      ETVQHIQGLKIACLEEIAWRNGWLSSGQLLECAKPMLKNDYGQYLAGLAKEESHS        292
AaRmlA      RTVESVQGLQVACLEEIAWRNGWLTSEQVETLARPMVKNEYGQYLLRLINEEK--        290
NmRmlA      QTVQNIQNLHIACLEEIAWRNGWLSDEKLEELARPMAKNQYGQYLLRLLKK----        286
SfRmlA      ATNEDRQGLKVSCPEEIAHRKGFIDAEQVKVLAEPLKKNAYGQYLLKMIKGY---        290
StRmlA      ATIEERQGLKVSCPEEIAFRKNFINAQQVIELAGPLSKNDYGKYLLKMVKGL---        291
             * .  *.*...* **** * . .   ..   * *. ** **.**  .
```

Alignment of the amino acid sequences of RmlA (VaRmlB), *A. actinomycetemcomitans* RmlA (AaRmlA) (AB010415), *N. meningitidis* RmlA (NmRmlA) (L09189), *S. flexneri* RmlA (SfRmlA) (P37779) and *S. typhimurium,* RmlA (StRmlA) (P26393), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues.

Table 6.3: Homology table for RmlA$_{Va}$

	%identity[a]				
	VaRmlB	**AaRmlA**	**NmRmlA**	**StRmlA**	**SfRmlA**
VaRmlA	100	81.7 (289)	81.5 (287)	67.8 (286)	65.7 (286)
AaRmlA		100	79.1 (287)	65.7 (286)	65.4 (286)
NmRmlA			100	64.7 (286)	62.9 (286)
StRmlA				100	89.3 (291)
SfRmlA					100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaRmlA	*V. anguillarum* RmlA
AaRmlA	*A. actinomycetemcomitans* RmlA (gbAB010415)
NmRmlA	*N. meningitidis* RmlA (gbL09189)
StRmlA	*S. typhimurium* RmlA (spP26394)
SfRmlA	*S.flexneri* RmlA (spP37779)

Figure 6.8: Alignment of RmlB$_{Va}$ with homologous proteins

```
VaRmlB     ------MKILVTGGAGFIGSAVVRHIIQNTQDSVINIDKLTYAGNLESLQGIDSSDRYAF   54
AaRmlB     ----MLKTILVTGGAGFIGSAVVRYIIENTQDSVVNVDKLTYAGNLESLEAVKNNPRYIF   56
NmRmlB     MQTANKKTILVTGGAGFIGSAVVRHIIRNTQDSVVNLDKLTYAGNLESLTDIADNPRYAF   60
SfRmlB     ------MKILVTGGAGFIGSAVVRHIINNTQDSVVNVDKLTYAGNLESLADVSDSERYAF   54
StRmlB     ------MKILITGGAGFIGSAVVRHIIKNTQDTVVNIDKLTYAGNLESLSDISESNRYNF   54
                   **.************* ** ****.*.*.************  .    ** *

VaRmlB     EPVDICNRSELDRVFYVHQPDAIMHLAAESHVDRSIDGPAAFIETNIVGTYTLLEAARSY  114
AaRmlB     EQVDICDAKALARIFEQHQPDAVMHLAAESHVDRSIDGPATFIETNIVGTYILLEAARAY  116
NmRmlB     EQVDICDRAELDRVFAQHRPDAVMHLAAESHVDRSIGSAGEFIQTNIVGTFNLLEAARAY  120
SfRmlB     EHADICDAVAMSRIFAQHQPDAVMHLAAESHVDRSITGPAAFIETNIVGTYVLLEAARNY  114
StRmlB     EHADICDSAEITRIFEQYQPDAVMHLAAESHVDRSITGPAAFIETNIVGTYALLEVARKY  114
           *  ***    . *.*   .***.*************     **.******. *** ** *

VaRmlB     WNTLESNKKSAFRFHHISTDEVYGDLEGTDDL--------FTEETPYSPSSPYSASKASS  166
AaRmlB     WSSLIDEKKAGFRFHHISTDEVYGDLDGTNNL--------FTETTPYSPSSPYSASKASS  168
NmRmlB     WQQMPSEQHEAFRFHHISTDEVYGDLHGTDDL--------FTETTPYAPSSPYSASKASS  172
SfRmlB     WSALNDEKKKSFRFHHISTDEVYGDLPHPDEANNNEALPLFTETTAYAPSSPYSASKASS  174
StRmlB     WSALGEDKKNNFRFHHISTDEVYGDLPHPDEVENSVTLPLFTETTAYAPSSPYSASKASS  174
           *  .   ..  ***************              *** * *.************

VaRmlB     DHLLRAWLRTYGFPTIVTNCSNNYGPYHFPEKLIPLMILNALDGKPLPVYGDGMQIRDWL  226
AaRmlB     DHLVRAWLRTYSLPTIVTNCSNNYGPKVFPEKLIPLIILNALDGKPLPVYGNGQQIRDWL  228
NmRmlB     DHLVRAWLRTYGLPTIVTNCSNNYGPYHFPEKLIPLMILNALDGKPLPVYGDGMQIRDWL  232
SfRmlB     DHLVRAWKRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEGKALPIYGKGDQIRDWL  234
StRmlB     DHLVRAWRRTYGLPTIVTNCSNNYGPYHFPEKLIPLVILNALEGKPLPIYGKGDQIRDWL  234
           ***.*** ***  *************  ********.*****.** **.** * ******

VaRmlB     FVEDHARALYKVVTEGQVGETYNIGGHNEKANIEVVNTLCALLEDLVPNKPANITHYADL  286
AaRmlB     FVEDHARALYKVVTEGKIGETYNIGGHNEKANIDVVRTICALLEELVPDKPAGVTKYEDL  288
NmRmlB     FVEDHARALYQVVTEGVVGETYNIGGHNEKANIEVVKTICTLLEELVPEKPAGVARYEDL  292
SfRmlB     YVEDHARALYTVVTEGKAGETYNIGGHNEKKNIDVVLTICDLLDEIVPKE----KSYREQ  290
StRmlB     YVEDHARALHMVVTEGKAGETYNIGGHNEKKNLDVVFTICDLLDEIVPKA----TSYREQ  290
           .********  *****  ************ *..** *.* **...**        * .

VaRmlB     ITYVKDRPGHDVRYAIDASKIERELGWKPEETFESGIRKTVEWYLNNREWWSRVLDGSYS  346
AaRmlB     ITYVKDRPGHDVRYAIDATKISRELGWKPQETFESGIRKTVEWYLNNRKWWSRVLDGSYN  348
NmRmlB     ITFVQDRPGHDVRYAVDAAKIRRDLGWLPLETFESGLRKTVQWYLDNKTWWQNVLNGSYR  352
SfRmlB     ITYVADRPGHDRRYAIDADKISRELGWKPQETFESGIRKTVEWYLANTNWVENVKSGTYQ  350
StRmlB     ITYVADRPGHDRRYAIDAGKISRELGWKPLETFESGIRKTVEWYLANTQWVNNVKSGAYQ  350
           **.* ****** ***.** ** *.*** * ******.****.*** *  *   *  *.*

VaRmlB     --CERLGSGQ-                                                  354
AaRmlB     --RERLGSQ--                                                  355
NmRmlB     --LERLGTGK-                                                  360
SfRmlB     SWIEQNYEGRQ                                                  361
StRmlB     SWIEQNYEGRQ                                                  361
              *.
```

Alignment of the amino acid sequences of RmlB (VaRmlB), *A. actinomycetemcomitans* RmlB (AaRmlB) (AB010415), *N. meningitidis* RmlB (NmRmlB) (P55294), *S. flexneri* RmlB (SfRmlB) (P37777) and *S. typhimurium* RmlB (StRmlB) (P26391),as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). The NAD binding motif (GXXGXXG) essential for these homologous proteins as described by Macpherson *et al.*, 1995) is highlighted. *, identical residues; •, similar residues.

Table 6.4: Homology table for RmlB$_{Va}$

	%identity[a]				
	VaRmlB	**AaRmlB**	**NmRmlB**	**SfRmlB**	**StRmlB**
VaRmlB	100	82.6 (350)	80.1 (352)	73.4 (353)	73.1 (353)
AaRmlB		100	79.8 (352)	77.0 (352)	74.4 (352)
NmRmlB			100	72.4 (352)	71.0 (352)
SfRmlB				100	88.4 (361)
StRmlB					100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaRmlB	*V. anguillarum* RmlB
AaRmlB	*A. actinomycetemcomitans* RmlB (gbAB010415)
NmRmlB	*N. meningitidis* RmlB (spP55294)
SfRmlB	*S. flexneri* RmlB (spP37777)
StRmlB	*S. typhimurium* RmlB (spP26391)

actinomycetemcomitans (68.7% identity at the amino acid level, 68% identity at the DNA level), *N. meningitidis* (67.1% identity at the amino acid level, 71% identity at the DNA level), and *S. enterica* sv Typhimurium (67.2% identity at the amino acid level, 67% idnetity at the DNA level) (Fig. 6.9; Table 6.5).

The final enzyme in the biosynthetic pathway is RmlD which encodes a dTDP-4-rhamnose reductase. RmlD$_{\mathrm{Va}}$ shows similarity to the same enzyme in *A. actinomycetemcomitans*, *S. flexneri* and *S. enterica* sv Typhimurium. RmlD shows 61% identity at the amino acid and DNA level to the equivalent gene and protein from *A. actinomycetemcomitans* (Fig. 6.10; Table 6.6). Homology to the related proteins from *S. flexneri* and *S. enterica* sv Typhimurium was considerably less significant at the amino acid level (38.7% and 41.2% respectively) and less than 30% at the DNA level.

6.2.8.1 Complementation of *S. flexneri rfbD* mutant with *V. anguillarum* rhamnose genes

To show functionality of the rhamnose genes from *V. anguillarum* O1, clones of the individual genes (pPM5515 (*rmlC*), pPM5517 (*rmlA*), pPM5518 (*rmlB*), pPM5519 (*rmlD*)) were electroporated into a *rfbD*/*rmlD* mutant of *S. flexneri* (RMA3361) (C. Daniels, personal communication) which no longer expressed O-antigen. The plasmid (pPM5519) containing *rmlD* from *V. anguillarum* was shown by dot blot with anti-*S. flexneri* O-antiserum to partially restore O-antigen biosynthesis in the *S. flexneri rfbD* (*rmlD*) mutant (Fig. 6.11). This indicates that *rmlD*$_{Va}$ is a functional homologue of *rfbD*/*rmlD* from *S. flexneri* which has been assigned the function of being a dTDP-rhamnose synthetase enzyme and confirms the function of *rmlD*$_{\mathrm{Va}}$.

Figure 6.9: Alignment of RmlC$_{\mathrm{Va}}$ with homologous proteins

```
VaRmlC          -----MKVIETDIPDVKIIEPTVFGDERGFFMETWNQQRFEELVTGKPTQFVQDNHSKSK   55
AaRmlC          -----MKVIDTKIPDVKLLEPQVFGDERGFFMETFRDEWFRKNVA--DRIFVQENHSKSI   53
NmRmlC          MKDKKMNIIDTAIPDVKLLEPQVFGDARGFFMETFRDEWFKTQVC--ERTFVQENHSKSG   58
SfRmlC          -----MNVIKTEIPDVLIFEPKVFGDERGFFMESFNQKVFEEAVG-RKVEFVQDNHSKST   54
StRmlC          ----MMIVIKTAIPDVLILEPKVFGDERGFFFESYNQQTFEELIG-RKVTFVQDNHSKSK   55
                     * .* * **** . ** **** **** *.     *   .      ***.*****

VaRmlC          KGILRGLHYQT-ENTQGKLVRVVSGEVFDVAVDIRKDSPTFGKWVGVHLSAENKRQLWVP  114
AaRmlC          KGVLRGLHYQT-ENTQGKLVRVISGAVFDVAVDMRGNSLTFGQWVGEILSAENKHQLWVP  112
NmRmlC          KGVLRGLHYQT-ENTQGKLVRVVVGEVFDVAVDMREGSPTFGKWVGEILSAQNQCQLWIP  117
SfRmlC          KGVLRGLHYQLEPYAQGKLVRCVVGEVFDVAVDIRKSSPTFGKWVGVNLSAENKRQLWIP  114
StRmlC          KNVLRGLHFQRGENAQGKLVRCAVGEVFDVAVDIRKESPTFGQWVGVNLSAENKRQLWIP  115
                *..*****.*    .******   * *******.*  * ***.***  ***.*. ***.*

VaRmlC          EGFAHGFYVTSEEAEFVYKCTNYYNPEAEHTLLWNDESINIDWPNKTCPLLSEKDVHGKH  174
AaRmlC          EGFAHGFYVLTDEAEFTYKCTDYYNPKAEHSLIWNDETIGIEWPLYGEPSLSAKDLAGKV  172
NmRmlC          EGFAHGFCVLGDAAEVVYKCTDYYKPEAEQVLIWNDPTVGIGWPLQTAPLLSPKDLAGKA  177
SfRmlC          EGFAHGFCVLSDEAEFVYKTNNFYSKMQERGILWSDKSINIEWPV-QNPLLSDKDINGQK  173
StRmlC          EGFAHGFVTLSEYAEFLYKATNYYSPSSEGSILWNDEAIGIEWPFSQLPELSAKDAAAPL  175
                *******    . **  **.. .*    *  ..* *  ...* **    * ** **

VaRmlC          LSDIIWL-----------------------------------------------------  183
AaRmlC          LAETVKF-----------------------------------------------------  180
NmRmlC          WAQAKSSALRFPDKKCRLNVSDGIFSDRLPTRLQYALCKEKHPGNEERQTQPRSGGIPVL  237
SfRmlC          FVDADYF-----------------------------------------------------  179
StRmlC          LDQALLTE----------------------------------------------------  183

VaRmlC          ------------------------------------------------------------  183
AaRmlC          ------------------------------------------------------------  180
NmRmlC          QIQRKRRIRAAVGHGEGNNKRHNHTDQAFTDNQAGGEQGSDALGVFQTAFAVFRFFTDDV  297
SfRmlC          ------------------------------------------------------------  179
StRmlC          ------------------------------------------------------------  183

VaRmlC          ------------------------------------                        183
AaRmlC          ------------------------------------                        180
NmRmlC          FQHRRQHRACHDGHIDGRRQINTHTDRKDGQGKFAA                        333
SfRmlC          ------------------------------------                        179
StRmlC          ------------------------------------                        183
```

Alignment of the amino acid sequences of RmlC (VaRmlC), *A. actinomycetemcomitans* RmlC (AaRmlC) (AB010415), *N. meningitidis* RmlC (NmlC) (L09189), *S. flexneri* RmlC (SfRmlC) (P37780) and *S. typhimurium* RmlC (StRmlC) (P26394), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues.

Table 6.5: Homology table for RmlC$_{Va}$

			%identity[a]		
	VaRmlC	**SfRmlC**	**AaRmlC**	**StRmlC**	**NmRmlC**
VaRmlC	100	69.7 (178)	68.7 (179)	67.2 (180)	67.1 (173)
SfRmlC		100	59.1 (181)	71.9 (178)	75.6 (176)
AaRmlC			100	61.5 (179)	61.8 (178)
StRmlC				100	59.6 (178)
NmRmlC					100

[a] Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaRmlC	*V. anguillarum* RmlC
AaRmlC	*A. actinomycetemcomitans* RmlC (gbAB010415)
SfRmlC	*S. flexneri* RmlC (spP37780)
StRmlC	*S. enterica* RmlC (spP37777)
NmRmlC	*N. meningitidis* RmlC (gbL09189)

Figure 6.10: Alignment of RmlD$_{Va}$ with homologous proteins

```
VaRmlD          -MRVLVTGSHGQVGYCLAHQLNSMVDVEFLAVDREQ--LDITNPCHVNSVVNEFKPNIII    57
AaRmlD          MAKFLITGANGQVGYCLTQQLQG--KHEILAVDHDE--LDITNQNAVKKTVENFRPDVVI    56
SfRmlD          -MNILLFGKTGQVGWELQRALAPLGNLIALDVHSTDYCGDFSNPEGVAETVKKIRPDVIV    59
StRmlD          -MNILLFGKTGQVGWELQRSLAPVGNLIALDVHSKEFCGDFSNPKGVAETVRKLRPDVIV    59
                    *. *  ****  * . *        * *       * .*   *   *   .* ...

VaRmlD          NAAAHTAVDRAEQEAELSYAINRDGPKHLAQAAHKVGAAILHISTDYVFSGDKPSAYIEA   117
AaRmlD          NAAAHTAVDRAETEIELSEAINVKGPQYLAETAKSVGAAILHISTDYVFDGRRAGKYKET   116
SfRmlD          NAAAHTAVDKAESEPNFAQLLNATCVEAIIKAANEVGAWVIHYSTDYVFPGNGDTAWLET   119
StRmlD          NAAAHTAVDKAESEPELAQLLNATSVEAIAKAANETGAWVVHYSTDYVFPGTGDIPWQET   119
                *********.** *   .  .*      .  .*   **  ..* ****** *       *.

VaRmlD          DSTSPQGVYGKSKLAGEIAVAQACPRHIILRTAWVFGEHGNNFVKTMLRLAQTRDALGVV   177
AaRmlD          DAVDPQGIYGKTKLAGEQAVVEANDKFIVLRTAWVFCEHGSNFVKTMLRLAKTRDTLGVV   176
SfRmlD          DATAPLNVYGGTKLAWGKALQDHCGKHLIFRTSWVYAGKGNNFAKTMLRLAKDRETLAVI   179
StRmlD          DATSPLNVYGKTKLAGEKALQDNCPKHLIFRTSWVYAGKGNNFAKTMLRLAKERQTLSVI   179
                *.  * ..** .***   *.     . .. **.**.  .* ** *******. * .* *.

VaRmlD          SDQYGGPTYAGDIANALLIMAKAIVEGKSTAFGVYHFSGTPHVSWYDFACAIFDKAREQG   236
AaRmlD          ADQIGGPTYAGDIAAALIQIAEKIIAGESVEYGIYHFTGEPYVSWCDFARAIFDEAVSQN   235
SfRmlD          TDQFGAPTGALDLLADCTRHAIWLAANKPEVAGLYHLVAGGTTTWHDYPALVFKEARRAG   239
StRmlD          NDQYGAPTGA-ELLADCTAHAIRVALNKPEVAGLYHLVAGGTTTWHDYAALVFDEARKAG   239
                 ** * ** * ..       *  .  .     *.**       .* *.   .*  *   .

VaRmlD          LLNKPLQVNAITTKDYPTPAKRPANSKLETQKIHQHFGVVASDWRAALNHLIDYKSE---   293
AaRmlD          MLEKAPLVNAIITADYPTPAKRPANSCLDLTKIQQAFGIQPSDWQRALKNIKAYAE----   291
SfRmlD          FNLPLNKLNAVPTTAYPTPARGPHNFRLNTEKFQQNFALVLPDWQVGVKRMLNELFTTTA   299
StRmlD          ITLALTELNAVPTSAYPTPASRPGNSRLNTEKFQRNFDLILPQWELGVKRMLTEMFTTTT   299
                       .**. *  *****  * *  *   * .. * .    *   .  .

VaRmlD          -                                                              293
AaRmlD          -                                                              291
SfRmlD          I                                                              300
StRmlD          I                                                              300
```

Alignment of the amino acid sequences of RmlD (VaRmlD), *A. actinomycetemcomitans* RmlD (AaRmlD) (AB002668), *S. flexneri* RmlD (SfRmlD) (P37778) and *S. typhimurium* RmlD (StRmlD) (P26392), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues.

Table 6.6: Homology table for RmlD$_{Va}$

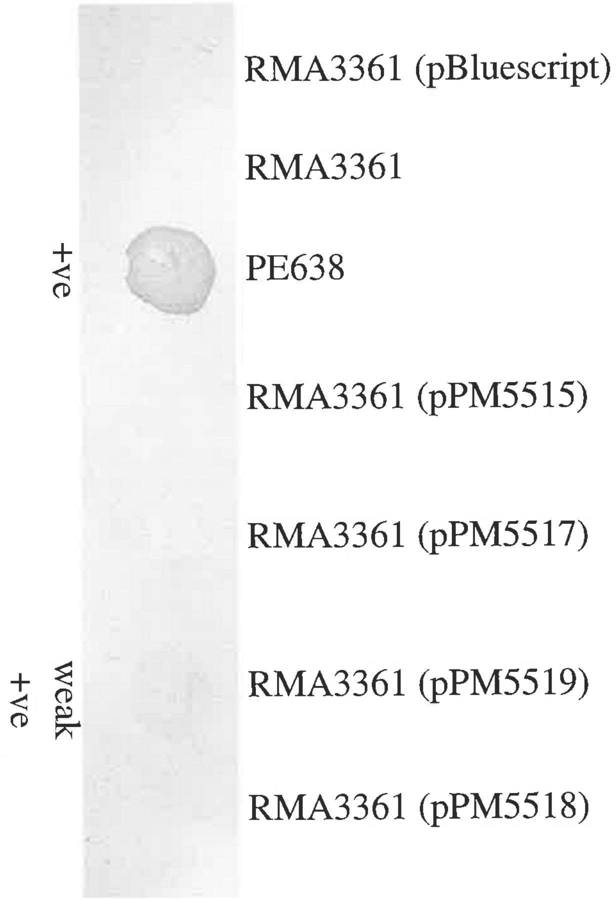
	%identity[a]			
	VaRmlD	**AaRmlD**	**StRmlD**	**SfRmlD**
VaRmlD	100	61.0 (290)	41.2 (291)	38.7 (292)
AaRmlD		100	40.8 (289)	39.1 (289)
StRmlD			100	81.0 (300)
SfRmlD				100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs.Abbreviations: gb, Genbank; sp, SWISS PROT.

VaRmlD	*V. anguillarum* RmlD
AaRmlD	*A. actinomycetemcomitans* RmlD (gbAB002668)
SfRmlD	*S. flexneri* RmlD (spP37778)
StRmlD	*S. typhimurium* RmlD (spP26392)

Figure 6.11: Complementation of *S. flexneri rfbD* mutant (RMA3361).

Dot immunoblot detected with anti-LPS (*S. flexneri* ZF01, Murex) at a 1:1000 dilution. The strains used are indicated above the samples (Table 2.2). PE638 (*S. flexneri* wild type) was used as the positive control. A weak positive reaction was observed with RMA3361 (pPM5519) which contained the cloned *rmlD* gene from *V. anguillarum* O1. No reaction was detected with RMA3361 (pPM5515), RMA3361 (pPM5517) or RMA3361 (pPM5518) which contained the other *rml*$_{\mathrm{Va}}$ genes.


RMA3361 (pBluescript)
RMA3361
+ve
PE638
RMA3361 (pPM5515)
RMA3361 (pPM5517)
weak +ve
RMA3361 (pPM5519)
RMA3361 (pPM5518)

6.2.9 *wzm* and *wzt*

The presence of *wzm* and *wzt* in the *wbh* region strongly indicates that the O-chain of *V. anguillarum* is a homopolymer. These genes belong to the ABC-2 transport family of LPS exporters, and have previously been described as exporters of homopolymer O-antigens and capsules (Whitfield *et al.*, 1997). Transporters belonging to this super family are able to facilitate the export of either small molecules or macromolecules via ATP-hydrolysis energised processes (Paulsen *et al.*, 1997).

Wzm shows homology to integral membrane proteins including those of other ABC-2 transport systems. TagG from *Bacillus subtilis* to which Wzm is most homologous (34.8% identity at the amino acid level) is described as a highly hydrophobic integral membrane protein involved in teichoic acid translocation (Lazarevic and Karamata, 1995) (Fig. 6.12; Table 6.7). Wzm also shows similarity to RfbA/Wzm from *Klebsiella pnuemoniae* (25.9% identity at the amino acid level) (Bronner *et al.*, 1994) (Fig. 6.12; Table 6.6). RfbA/Wzm is the integral membrane translocator protein involved in the export of the O-antigen homopolymer in this organism. At the DNA level, no significant homologies were obtained across the entire gene, however *wzm* did show 54% identity to the *tagG* gene in the first 327 nucleotide bases.

The integral membrane protein of a typical ABC transporter spans the cytoplasmic membrane six times. Hydropathy analysis and the use of the program TMpredict showed that Wzm has six transmembrane domains. Table 6.8 shows the predicted transmembrane regions as indicated by TMpredict. Hydropathy plots of functionally identical proteins are shown in comparison to Wzm (Fig. 6.13).

The second protein in the ABC-2 transport system is the energiser which has an ATP-binding domain (Higgins, 1992). This protein is highly hydrophilic and is located in the cytoplasm. Wzt is highly homologous to other ATP-binding proteins from

Figure 6.12: Alignment of Wzm with homologous proteins

```
VaWzm           MYKLPSKFVHDVFSSYELLFPMAKRDIRSRYLGSLFGGVWAFIQPLVTILVIWFVFQVGF   60
BsTagG          MNDL-LRILREQITSFPLILRLAAYETKSKYQMNYLGVLWQFLNPLIQMLAYWFVFGMGI   59
KpWzm           ---MSIKMKYNLGYLFDLLVVITNKDLKVRYKSSMLGYLWSVANPLLFAMIYYFIFKLVM   57
                   .  .        . *.  ..  . . .*     * .*   .**.  .   *.* .

VaWzm           K--AQATDDG--VPFTLWLVSAMIPWFFVAEMLSSGTNSIIEQANIVKKIVFKVSLLPIV  116
BsTagG          RKGGPVTTGAGEVPFIIWMLAGLIPWFFISPTILDGSNSVFKRINMVAKMNFPISSLPSV  119
KpWzm           R----VQIPN----YTVFLITGLFPWQWFASSATNSLFSFIANAQIIKKTVFPRSVIPLS  109
                .             . . ... . **   .        *     ... *  *  * .*

VaWzm           KIISALMIHIFFIVILFTVSIAYGYYPKVGWLQ-VPYYILCSLVLMLGISWITSSIIVFF  175
BsTagG          AIASNLFSYMIMMVIYIIVLLVNGVFPSVHWLQ-YIYYFICMIAFMFSFSLFNSTISVLI  178
KpWzm           NVMMEGLHFLCTIPVIVVFLFVYGMTPSLSWVWGIPLIAIGQVIFTFGVSIIFSTLNLFF  169
                 .       .  . .        *  * . *.       .  .      *   *.. .

VaWzm           RDVGQIIAVLIQLGFWATPIFWNITMVP----IEYQWVLKLNPVFYITEGYRNTITTDLW  231
BsTagG          RDYQFLLQAVTRLLFFLLPIFWDVNAKLGQSHPELVPVLKLNPLFYIIEGFRNSFLDGAW  238
KpWzm           RDLERFVSLGIMLMFYCTPILYASDMIP----EKFSWIITYNPLASMILSWRDLFMNGTL  225
                **    .     * *   **                 ..  **.  .    *
VaWzm           FWESFLWTAYYWIFTSFTLLVGIVCFKKLRPHFADVL                        267
BsTagG          FFHDMKYTLYFWLFTFLLLLVGSILHMKFRDKFVDFL                        274
KpWzm           NYE---YISILYFTGIILTVVGLSIFNKLKYRFAEIL                        258
                       .           .**     * . .* . *
```

Alignment of the amino acid sequences of Wzm (VaWzm), *B. subtilis* TagG (BsTagG) (P42953) and *K. pneumoniae* Wzm (KpWzm) (Q48475), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). *, identical residues; •, similar residues.

Table 6.7: Homology table for Wzm_{Va}

	%identity^{a}		
	VaWzm	**BsTagG**	**KpWzm**
VaWzm	100	34.8 (267)	25.9 (255)
BsTagG		100	22.9 (262)
KpWzm			100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: sp, SWISS PROT.

VaWzm	*V. anguillarum* Wzm
BsTagG	*B. subtilis* TagG (spP42953)
KpRfbA	*K. pneumoniae* Wzm (spQ48375)

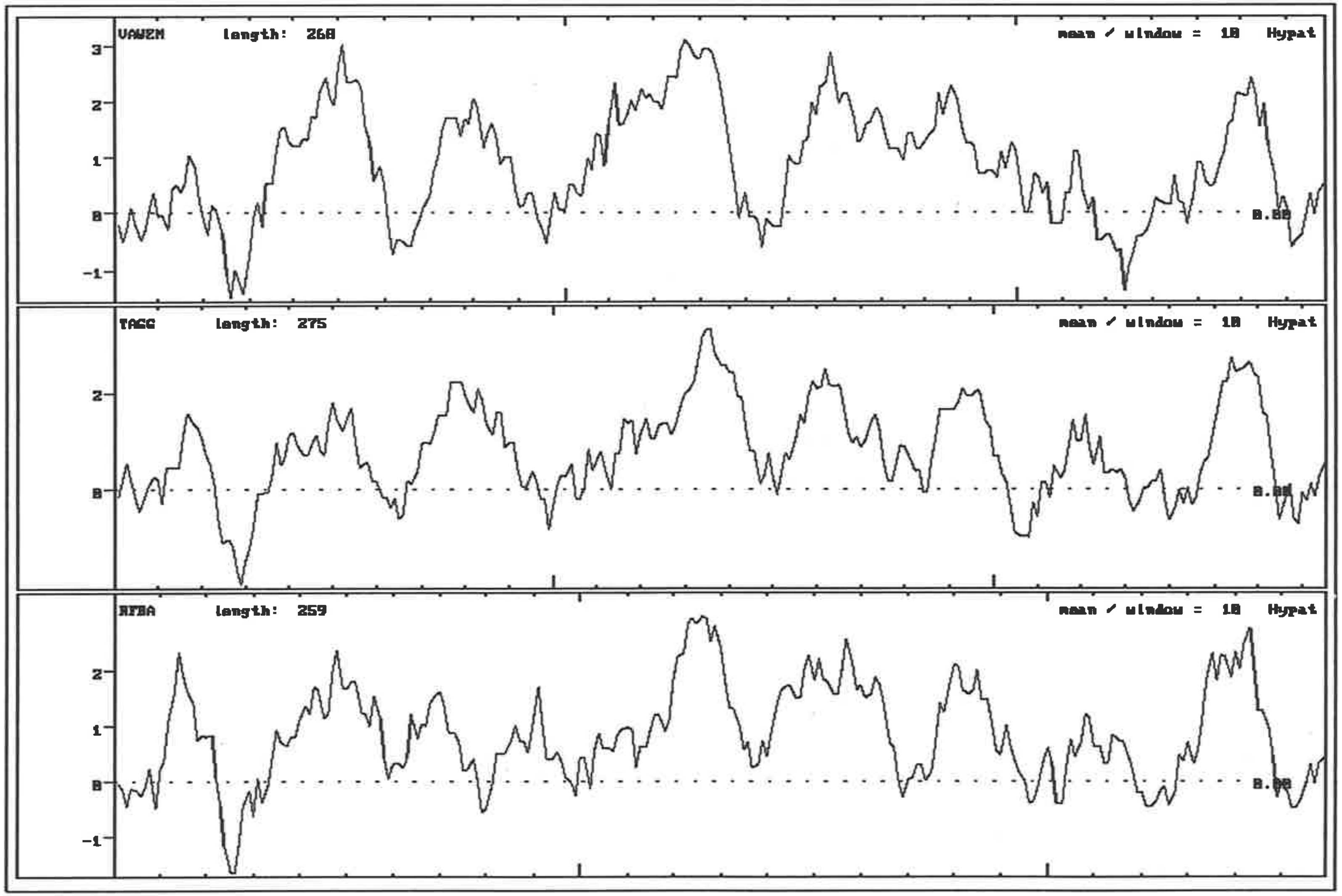
Table 6.8: Predicted transmembrane regions of *wzm*$_{\mathrm{Va}}$

Tm #	Start	End	Length	Orientation
1	31	56	26	i-o
2	68	90	23	o-i
3	118	142	25	i-o
4	145	175	31	o-i
5	178	204	27	i-o
6	236	258	22	o-i

Transmembrane spanning domains as predicted by the computer program TMpredict (Hofmann and Stoffel, 1993). Only the most likely model is presented. The numbers in the columns represent amino acid residues in the *wzm* translated sequence.

Figure 6.13: Hydropathy plot of Wzm and related proteins.

The hydropathy plots were generated by the method of Kyte and Doolittle (1982) and aligned using PROFILEGRAPH (Hofmann and Stoffel, 1989). Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 10th amino acid is marked on each X-axis. VaWzm, *V. anguillarum* Wzm; TagG, *B. subtilis* TagG (spP42953); RfbA, *K. pneumoniae* Wzm (spQ48375). Abbreviations: sp, SWISS PROT.


VAWZM
length: 268
mean / window = 10
Hypat
3
2
1
0
-1
0.00
TACC
length: 275
mean / window = 10
Hypat
2
0
0.00
RFBA
length: 259
mean / window = 10
Hypat
2
1
0
-1
0.00

Enterococcus faecalis (56% identity at the amino acid level), *E. coli* (34% identity at the amino acid level) and *Aeromonas salmonicida* (42% identity at the amino acid level) (Fig. 6.14; Table 6.9). Wzt is also homologous to RfbB (45.5% identity at the amino acid level) and TagH (42.3% identity at the amino acid level) from *K. pnuemoniae* and *B. subtilis*, respectively, however the degree of homology is not as high as that seen for the integral membrane protein component (Table 6.9). Homology at the DNA level was observed for *wzt* and the gene for the ATP binding protein from *E. faecalis* (63%). For Wzt to bind and utilise ATP it requires an ATP-binding motif or Walker box (Higgins *et al.*, 1992; Walker *et al.*, 1982). The Walker box consists of two short motifs separated by approximately 100 amino acids which is characteristic of many nucleotide binding proteins (Walker *et al.*, 1982; Higgins *et al.*, 1992). The sequence of the Walker box is generally well conserved with different energisers showing between 30 and 50% identity, depending on the proteins being compared. An amino acid sequence alignment of Wzt in comparison to other ABC transport energisers is presented in Figure 6.14. The conserved Walker box domains (W_A and W_B) are underlined.

The ABC transport system, Wzm and Wzt of *V. anguillarum* O1 also shows homology to RfbH and RfbI, respectively which are the equivalent proteins in *V. cholerae* O1 (Manning *et al.*, 1995). *V. cholerae* O1 O-antigen consists of a homopolymer of perosamine and therefore would be expected to have an ABC transport system to translocate its polymerised O-antigen chain to the periplasm for ligation to the lipid A core oligosaccharide component of LPS. The homology observed is 24.1% identity (over 220 amino acid overlap) for Wzm/RfbH and 34.6% identity (over 208 amino acid overlap) for Wzt/RfbI at the amino acid level.

Figure 6.14: Alignment of Wzt with homologous proteins

```
VaWzt           MNDVVISCENITKIYPMYDDHRDRFKEVFHPFRKVYHKKFHALDDVSFEVKRGETVGIVG   60
EfATP           MSEYAIDIQHITKTYNMYKKPSDRFKEALSPTKKSYHDLFYALDDVTMQIKKGEMIGFVG   60
AsABCA          MSEPVLAVSGVNKSFPIYRSPWQALWHALNP--KADVKVFQALRDIELTVYRGETIGIVG   58
EcATP           --MSYIRVNNVGKAYRQYHSKTGRLIEWLSPLNTKRHNLKWILSDINFEVAPGEAVGIIG   58
                     .    . * .  *            *           * *.   .  ** .* .*

VaWzt           KNGAGKSTLLKIITGVLTPTSGSVKLNGVVSSLLELGTGFNPDLTGIENIYLNSSLMGIQ  120
EfATP           ENGSGKSTMLKIITGVLTPTSGSMVINGKISALLELGSGFNPEYSGYENIFLNGMVLGFS  120
AsABCA          HNGAGKSTLLQLITGVMQPDCGQITRTGRVVGLLELGSGFNPEFTGRENIFFNGAILGMS  118
EcATP           INGAGKSTLLKLITGTSRPTTGEIEISGRVAALLELGMGFHSDFTGRQNVYMSGQLLGLS  118
                        WA
                 **.****.*..***   * .* .   * .  ***** **. . .* .*..    ..*

VaWzt           KEDIDKKLAQVIAFADIGEHIHQPVRGYSSGMFARLAFSVAISVEPDILIVDEALAVGDA  180
EfATP           REEMEERVQDVIDFADIGDHLYQPVKTYSSGMFVRLAFAVAINVDPDILIVDEALAVGDL  180
AsABCA          QREMDDRLERILSFAAIGDFIDQPVKNYSSGMMVRLAFSVIINTDPDVLIIDEALAVGDD  178
EcATP           SEKITELMPQIEEFAEIGDYIDQPVRVYSSGMQVRLAFSVATAIRPDVLIIDEALSVGDA  178
                                                                     WB
                   .   .  .  ** **. . ***. *****  ****.*     **.**.****.***

VaWzt           AFVNKCYGKINELKSKGMTLLFVSHSLGAVSELCTKAILIDNGKCLLVSDVETVVNQYNR  240
EfATP           EFQLKCMEKFTEIKNSGKTILFVSHDVNSIRRFCDRTFWVATRESP--------------  226
AsABCA          AFQRKCYARLKQLQSQGVTILLVSHAAGSVIELCDRAVLLDRGEVLLQGEPKAVVHNYHK  238
EcATP           YFQHKSFERIRKFRQEGTTLLLVSHDKQAIQSICDRAILLNKGQIEMEGEPEAVMDYYNA  238
                 *  *.  .    .  * *.* ***   ..   * ..  .

VaWzt           MIRSSFDKNLQENENIEETTFDSSEKDTSEISLISNVKSKDIFNINNEFMNKHSKTRYGK  300
EfATP           ------------------------------------------------------------  226
AsABCA          LLHMEGDERARFRYHLRQTGRGDSYISDESTSEPKIKSAPGILSVDLQPQ---STVWYES  295
EcATP           LLAD------KQNQSIKQVEHNG---KTQTVSGTGEVTISEVHLLDEQGN-VTEFVSVGH  288

VaWzt           GGARIVNVEIIEGDDNKNRLFSYNTKIKLRVYIKSDQNLNKLNCGYFIRTDKGLSIVGNN  360
EfATP           ------------------------------------------------------------  226
AsABCA          KGAVLSDVHIESF-----------------------------------------------  308
EcATP           RVSLQVNVEVKD-------------------------DIPELVVGYMIKDRLGQPIFGTN  323

VaWzt           LESSKFELRDLKANRKVIIDFDIPLTIKSGIYSLTVVLGASDSENNEFIVDWVDLADSFE  420
EfATP           ------------------------------------------------------------  226
AsABCA          ------------------------------------------------------------  308
EcATP           TYHLNQTLTSLKKGEKRSFLFSFDARLGVGSYSVAVALHTSSTHLG-KNYEWRDLAVVFN  382

VaWzt           VFHSKDQYSFEYMVHIQNKISMNYID                                  446
EfATP           --------------------------                                  226
AsABCA          --------------------------                                  308
EcATP           VVNTEQQE-FVGVSWLPPELEIS---                                  404
```

Alignment of the amino acid sequences of Wzt (VaWzt), *E. faecalis* ATP (EfATP) (AF071085), *E. coli* ATP (EcATP) (AB010150) and *Aeromonas salmonicida* ABCA (AsABCA) (Q07698), as determined using the default settings of the program CLUSTAL (Higgins and Sharp, 1988). The Walker box motifs are underlined and marked according to text. *, identical residues; •, similar residues.

Table 6.9: Homology table for Wzt$_{Va}$

	%identity[a]					
	VaWzt	**EfATP**	**EcATP**	**AsABCA**	**KpRfbB**	**BsTagH**
VaWzt	100	55.5 (220)	33.9 (381)	41.6 (310)	45.5 (244)	42.3 (215)
EfATP		100	44.6 (213)	49.8 (217)	41.0 (222)	42.3 (215)
EcATP			100	46.3 (244)	43.0 (242)	38.3 (256)
AsABCA				100	42.6 (244)	40.3 (236)
KpRfbB					100	40.5 (242)
BsTagH						100

a Percentage of identical amino acids determined with FASTA as implemented in PROSIS. Numbers in the brackets indicate the number of amino acids over which the % identity occurs. Abbreviations: gb, Genbank; sp, SWISS PROT.

VaWzt	*V. anguillarum* Wzt
EfATP	*E. faecalis* ATP binding protein (gbAF071085)
EcATP	*E. coli* ATP binding protein (gbAB010150)
AsABCA	*A. salmonicida*, ATP binding protein (spQ07698)
KpRfbB	*K. pneumoniae*, RfbB (spQ48479)
BsTagH	*B. subtilis*, TagH (spP42954)

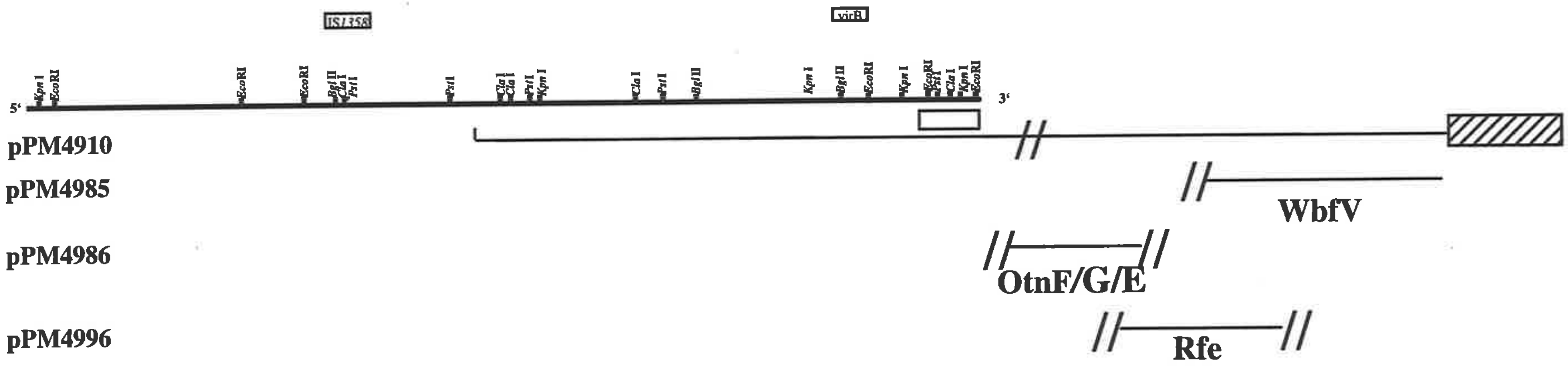
6.2.10 Additional genes involved in the synthesis of O-antigen in *V. anguillarum* O1

It was evident from further cloning (data not shown) and the initial analysis of the Tn*phoA* (Cm^R) insertion mutant KM2 (Section 6.2.2) that additional genes are required for the synthesis of LPS in *V. anguillarum* O1. KM2 was found have a Tn*phoA* (Cm^R) insertion in an ORF homologous to *otnE* (Bik *et al.*, 1996) from *V. cholerae* O139 (D. Milton, personal communication). *otnE* has yet to be assigned a specific function, however it is thought to play a role in capsule biosynthesis (Stroeher *et al.*, 1998; Bik *et al.*, 1996). KM2 was shown to have LPS with slightly higher molecular weight O-antigen (increased modal chain length) (Section 6.2.2). This suggests a role for this *otnE* homologue in O-antigen biosynthesis in *V. anguillarum* O1. *otnE* could not be linked to the *wbh* region sequenced in this study (data not shown). Further cloning and sequencing of the region downstream of *wbhO* on pPM4910 (Chapter 5, Section 5.2.1) has revealed the presence of other ORFs with homology to *rfb*/polysaccharide genes (Fig. 6.15). Restriction mapping and Southern hybridisation analysis using fragments of these additional genes (cloned in pBC-KS$^+$ and DIG-labelled), and long range PCR between the end of *wbhN* and the additional cloned regions, failed to directly link these new genes to the *wbh* region described in this study (data not shown) suggesting that the distance between the regions was too great for amplification.

Some of the subclones obtained (pPM4985, pPM4986, pPM4996) showed homology to *rfb*-like genes including other *otn* genes (nucleotide sugar dehydrogenase (*wbfV*), *otnF* and *otnG*) from *V. cholerae* O139; *otnF* (*wbfC*) and *otnG* (*wbfB*) do not have defined functions. The nucleotide sugar dehydrogenase (*wbfV*) is homologous to an ORF adjacent the *E. coli* O111 *rfb* region (Bastin and Reeves, 1995). It is therefore possible that the region involved in O-antigen synthesis in *V. anguillarum* O1 will need to be extended

Figure 6.15: Presence of additional polysaccharide gene homologues in *V. anguillarum* O1.

A schematic representation of the fragments obtained from random subcloning of the cosmid clone pPM4910. // : undefined location with respect to the other clones and the sequenced *wbh* region. Box with stripes represents the cosmid vector pPM2101. Blank box represents the gene, *wbhN* which is marked for reference. Protein homologues obtained from sequencing the fragments and performing BLASTX searches are shown beneath the respective subclone.


IS1358
virB
5'
Kpn I
EcoRI
EcoRI
EcoRI
Bgl II
Cla I
Pst I
Pst I
Cla I
Cla I
Pst I
Kpn I
Cla I
Pst I
Bgl II
Kpn I
Bgl II
EcoRI
Kpn I
EcoRI
Pst I
Cla I
Kpn I
EcoRI
3'
pPM4910
pPM4985
WbfV
pPM4986
OtnF/G/E
pPM4996
Rfe

to include these additional genes, although it is also possible that these genes are redundant and not functional.

One subclone of interest (pPM4996) had sequence homologous to *rfe* (Meier-Dieter, 1992) (data not shown). Rfe/WecA is one of only two known initiating enzymes and catalyses the addition of GlcNAc to bactoprenol in the first step of O-antigen biosynthesis (Whitfield, 1995). Rfe/WecA has been associated with initiating polymerisation in the homopolymers O-antigens of *E. coli* O8, O9 (Jann *et al.*, 1982; Rick *et al.*, 1994) and *Klebsiella pneumoniae* O1 (Clarke and Whitfield, 1992). The presence of this homologous sequence (47% identical at the amino acid level) suggests that the O-antigen polymerisation in *V. anguillarum* O1 is initiated by the transferase activity of the Rfe/WecA homologue.

6.2.11 Comparison of the *V. anguillarum* O1 wildtype strains 85-3954-2 and NB10 by Southern Hybridisation and PCR

The two *V. anguillarum* O1 wildtype strains, 85-3954-2 and NB10 used in this study were isolated from different geographical locations. NB10 was the wildtype strain used to construct the *wbh* mutants. However, the *wbh* region was cloned and sequenced from the O1 strain 85-3954-2. It was therefore decided to confirm the presence of the other *wbh* genes in NB10. Southern hybridisation was used to compare the strains to ensure the strains were identical and contained the same *wbh* genes. The DNA sequences obtained (D. Milton, personal communication) from sequencing the insertion point of the Tn*phoA* (Cm^R) transposon in the mutants were compared to the sequence obtained in this study (Appendix 1). The sequences were found to be 99.9% identical to those for 85-3954-2, with only a few nt differences. Previously Norqvist and Wolf-Watz (1993) sequenced a 3.3

kb *Kpn* I fragment of the region from strain 775 (Accession Number L08012). Comparison of this sequence also indicated only a few nt differences (data not shown).

Chromosomal DNA from 85-3954-2 and NB10 was digested with various restriction enzymes found in the *wbh* region, and DIG-labelled subclones (pPM4901, pPM4943) and IS*1358* were used as probes to determine similarity. The Southern blot using pPM4943 as a probe showed that the two O1 strains had identical restriction fragments for all the restriction enzymes used (Fig. 6.16 (A); Table 6.10). This was also the case for the blot with the probe pPM4901 (Fig. 6.16 (B); Table 6.10). Unexpectedly, a difference was observed when the IS*1358* probe was used. Southern hybridisation preformed previously in this study using IS*1358* as a probe on *Hin*dIII digested chromosomal DNA of *V. anguillarum* O1 strains showed the presence of conserved 4.8 and 1.8 kb fragments (Chapter 3, Section 3.2.8). NB10 when probed with IS*1358*, showed the conserved 1.8 kb fragment but the second band was smaller in size (approximately 3.6 kb). However, when chromosomal DNA was digested with *Cla*I the same expected DNA fragments were detected (Fig. 6.16 (C); Table 6.10). This is the first *V. anguillarum* O1 strain analysed that has shown a difference in the *Hin*dIII restriction pattern.

PCR using oligonucleotides (Table 2.4, Fig. 6.17) combinations to amplify different *wbh* genes were also performed. The products produced were identical between the two strains tested suggesting the sequenced *wbh* operon from 85-3954-2 was present in NB10 (Fig. 6.17).

6.3 Summary and Discussion

In this chapter, the *wbh* locus was further characterised by the analysis of mutants with transposon insertions in three different *wbh* genes. The LPS from these mutants were characterised by Western immunoblotting, SDS-PAGE, and silver staining.

Figure 6.16: Southern hybridisation: A comparison of the *V. anguillarum* O1 strains, 85-3954-2 and NB10.

Southern blots of DNA from 85-3954-2, pPM4905 and pPM4910 digested with various restriction enzymes, were probed with DIG-labelled probes of pPM4943 (A), pPM4901 (B) and IS*1358* (C). The approximate size of the hybridised bands are shown in kilobases (kb) in accordance with Spp1 (*Eco*RI) markers.

A: Southern blot probed with DIG-labelled pPM4943 (*ddhD*, *ddhA*)

a, 85-3954-2 digested with *Hin*dIII; b, NB10 digested with *Hin*dIII; c, 85-3954-2 digested with *Cla*I; d, NB10 digested with *Cla*I; e, 85-3954-2 digested with *Sac*I; f, NB10 digested with *Sac*I; g, 85-3954-2 digested with *Bgl*II; h, NB10 digested with *Bgl*II; i, 85-3954-2 digested with *Pst*I; j, NB10 digested with *Pst*I. The probe was labelled randomly (Section 2.10.7) and corresponds to nt 14289- nt 15442 (Appendix 1)

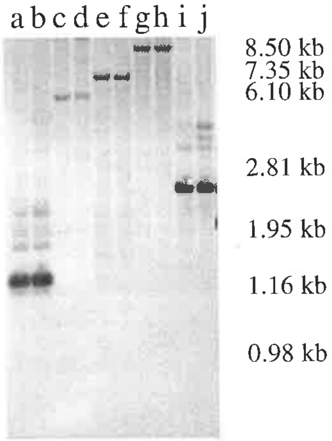
B: Southern blot probed with PCR DIG-labelled pPM4901 (*virB/wbhL*)

a, 85-3954-2 digested with *Eco*RI; b, NB10 digested with *Eco*RI; c, 85-3954-2 digested with *Pst*I; d, NB10 digested with *Pst*I; e, 85-3954-2 digested with *Hin*dIII; f, NB10 digested with *Hin*dIII. The probe was PCR DIG-labelled (Section 2.11.4) using oligonucleotides #1088 and #1089 (Table 2.4).

C: Southern blot probed with PCR DIG-labelled IS*1358*

a, 85-3954-2 digested with *Hin*dIII; b, NB10 digested with *Hin*dIII; c, 85-3954-2 digested with *Cla*I; d, NB10 digested with *Cla*I. The probe was PCR-DIG labelled (Section 2.11.4) using oligonucleotide #773 (Table 2.4).

A


a b c d e f g h i j
8.50 kb
7.35 kb
6.10 kb
2.81 kb
1.95 kb
1.16 kb
0.98 kb

B


a b c d e f
8.5 kb
4.84 kb
3.59 kb.
1.16 kb
0.98 kb

C

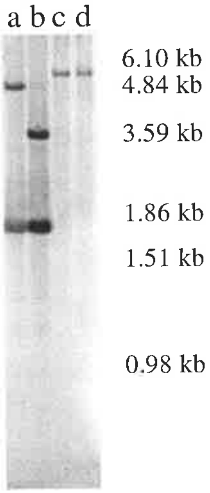

a b c d
6.10 kb
4.84 kb
3.59 kb
1.86 kb
1.51 kb
0.98 kb

Table 6.10: Sizes of restriction enzyme fragments which hybridise to pPM4943, pPM4901 and IS*1358*

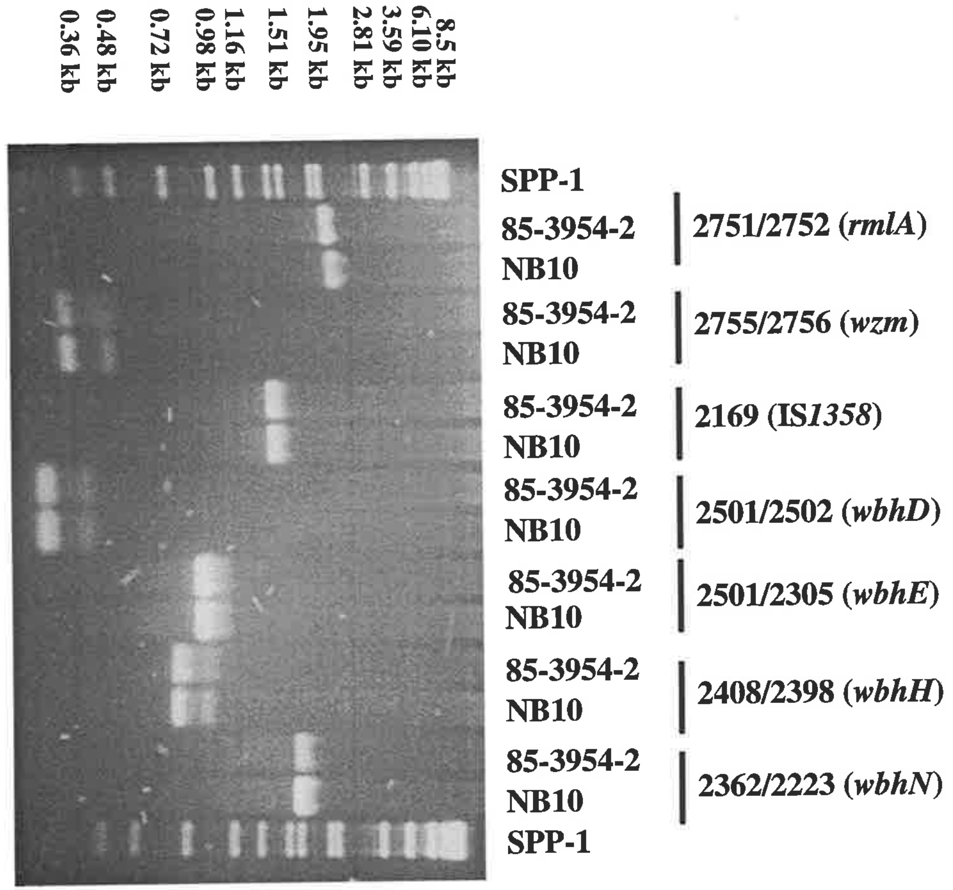
	DIG-labelled probe used[a]					
	pPM4943		pPM4901		IS*1358*	
Restriction enzyme	DNA used in Southern analysis					
	85-3954-2	**NB10**	**85-3954-2**	**NB10**	**85-3954-2**	**NB10**
***Bgl*II**	>8.5 kb	>8.5 kb	NA	NA	NA	NA
***Cla*I**	5.0 kb	5.0 kb	NA	NA	5.0 kb	5.0 kb
***Eco*RI**	NA	NA	18 kb	18 kb	NA	NA
***Hind*III**	1.2 kb	1.2 kb	4.0 kb	4.0 kb	4.8 kb	3.6 kb
			1.2 kb	1.2 kb	1.8 kb	1.8 kb
***Pst*I**	2.8kb	2.8 kb	>8.5 kb	>8.5 kb	NA	NA
***Sac*I**	7.3 kb	7.3 kb	NA	NA	NA	NA

Sizes of the restricted fragments are approximate and were measured in kilobases (kb). a, the DIG-labelled probes correspond to nucleotide numbers in Appendix 1: pPM4943, nt 14289-15442; pPM4901, nt 25700-26788; IS*1358*, nt 9801-11139. NA: not available

Figure 6.17: PCR analysis of the *wbh* region in 85-3954-2 and NB10.

PCR products were amplified using standard conditions as described in Materials and Methods (Section 2.11.1) using an extension time of 2 min 30 sec. The products were electrophoresed on a 1% agarose gel followed by staining with ethidium bromide. SPP-1 digested with *Eco*RI was also electrophoresed as size markers and the fragments are indicated on the left side of the figure.

The bolded lines with small letters correspond to the following oligonucleotide combinations (Table 2.4). The *wbh* genes amplified from these reactions are indicated below.


0.36 kb
0.48 kb
0.72 kb
0.98 kb
1.16 kb
1.51 kb
1.95 kb
2.81 kb
3.59 kb
6.10 kb
8.5 kb
SPP-1
85-3954-2
NB10
2751/2752 (*rmlA*)
85-3954-2
NB10
2755/2756 (*wzm*)
85-3954-2
NB10
2169 (IS*1358*)
85-3954-2
NB10
2501/2502 (*wbhD*)
85-3954-2
NB10
2501/2305 (*wbhE*)
85-3954-2
NB10
2408/2398 (*wbhH*)
85-3954-2
NB10
2362/2223 (*wbhN*)
SPP-1

The transposon mutants were initially isolated by screening using Western immunoblotting analysis with antisera raised against formalin-killed bacteria, and were initially thought be R-LPS (ie no O-antigen) mutants (D. Milton, unpublished results). Silver staining of the LPS from these mutants showed that KM2, KM16 and KM110 possess LPS with altered O-antigens with an increased or decreased molecular weight, in comparison to the wildtype. The fourth mutant, KM97 appeared to have a true R-LPS phenotype.

Western immunoblotting analysis of KM16, KM97 and KM110 using the rabbit antiserum raised against heat killed 85-3954-2 (*V. anguillarum* O1) indicated that the altered O-antigens produced by KM16 and KM110 contained similar epitopes to the wildtype O-antigen. LPS from KM97 showed no reaction with this antiserum by Western immunoblotting. Serum absorption studies using heat killed suspensions of the mutants indicated that the O-antigen produced by the mutants had sufficiently different epitopes to the wildtype (85-3954-2, NB10) as only antibodies against the altered O-antigen produced by the mutants were removed by absorption (Fig. 6.3). This suggests that either the wildtype strain is capable of producing both wildtype O-antigen and altered O-antigen or that cross-reacting antibodies recognise similar/common epitopes on both O-antigens. However, if this was the case, a decrease in detection of the wild type LPS would be expected in Western immunoblotting. Absorption probably removed any common epitope antibodies leaving only antibodies specific to epitopes on the wildtype O-antigen. Indirect immunofluorescence further supported the serum absorption studies. Electron microscopy did not support the data using the KM97 absorbed serum. The most likely reason is that this method has less sensitivity. The finding in this study of altered O-antigens is reminiscent of what has been shown previously in *S. flexneri* (Morona *et al.*, 1995a) and *E. coli* (Saeki *et al.*, 1993), and is therefore a form of antigenic variation.

Initial analysis of the sugar composition of LPS from the wildtype and mutant strains used in this study indicated the presence of glucose, rhamnose and glucosamine in the LPS of *V. anguillarum* O1. The wildtype strains, 85-3954-2 and NB10 used in this study were found to contain glucose and rhamnose in a ratio of 10:1. In addition glucosamine was also found at the same ratio as rhamnose (ie 10:1:1). The rhamnose biosynthesis genes have been located based on gene homology and the synthetase gene (*rmlD*) was found to complement a *S. flexneri rfbD*/*rmlD* mutant. Indirect evidence that the O-antigen of *V. anguillarum* O1 is a homopolymer comes from the O-antigen transport system (*wzm*/*wzt*) available.

Analysis of the mutant LPS from KM16, KM97 and KM110 indicated that no glucosamine was present and that the ratio of glucose to rhamnose was different for each mutant. KM16 was shown to have a wildtype ratio of glucose to rhamnose (ie. 10:1). It is unclear from the preliminary sugar analysis of KM97 and KM110 if the change in sugar ratio observed in these strains is due to either an increase in rhamnose or decrease in glucose. Silver staining the LPS of KM16 and KM110 showed that these strains had O-antigens with decreased molecular weights with no apparent effect on the core oligosaccharide component (Section 6.2.2). The Tn*phoA* (Cm^R) insertions in the mutants resulted in polar mutations as indicated by the inability of a single gene to complement the mutation and therefore it is likely that downstream genes are effected. Hence caution is required in the interpretation of the data obtained in the analysis of these strains, which will be discussed in detail in Chapter 7.

A comparison of the two wildtype strains used in this study suggests that the *wbh* locus in each strain is identical, however restriction fragment length polymorphism (RFLP) was noted for *Hin*dIII when NB10 was probed with IS*1358*. Additionally, other genes to those characterised in detail (Chapter 5) were detected in the region. One of these (*otnE*) is required for O-antigen biosynthesis in *V. anguillarum* O1, although the precise function of this, and the other genes is unknown.

Chapter Seven

Discussion

7.1 Introduction

V. anguillarum O1 is a significant pathogen that causes vibriosis in both feral and farmed fish. Little is known either about the precise mechanisms involved in the pathogenesis or the virulence determinants that facilitate this process.

A commercial vaccine based on lipopolysaccharide is available for *V. anguillarum,* however the immunity conferred is serotype specific and ineffective unless injected intraperitoneally. Studies investigating the LPS of *V. anguillarum* have focussed on the chemical composition of these molecules, while little is known about the genetics of LPS biosynthesis in *V. anguillarum*. In one study, the O-antigen of the O3 serotype (Eguchi *et al.*, 1992 a, b) was examined however the genes required for the biosynthesis have not yet been isolated. The O-antigen biosynthesis locus of *V. anguillarum* O2 has been cloned and expressed in *E. coli* (Amor and Mutharia, 1995) however the region has not been sequenced and no information on the biosynthetic genes was presented. The structure of the O-antigen of this serotype has been determined (Sadovskaya *et al.*, 1996). The strains which have been associated with the most severe disease outbreaks belong to the O1 serotype of *V. anguillarum*, and it has been shown that the LPS of the O1 serotype is an essential virulence factor (Norqvist and Wolf-Watz, 1993). Sugar composition data is available for the *V. anguillarum* O1 LPS core oligosaccharide (Banoub and Shaw, 1981) but information on the O-antigen component of the LPS of this serotype is unreliable (Banoub *et al.*, 1987) as the strain used was poorly defined.

In this thesis the LPS/O-antigen biosynthesis locus (*wbh*) of *V. anguillarum* O1 was isolated, cloned and sequenced. The functions of a majority of the *wbh* gene products have been assigned, based on either homology to other proteins of known function, and mutational analysis in conjunction with either sugar analysis or gene complementation. In addition, the association of IS*1358* with O-antigen biosynthesis loci in *V. cholerae* and *V. anguillarum* was investigated, and its relevance to genetic rearrangements involving O-antigen biosynthesis loci will be discussed in detail.

7.2 Major Outer Membrane Protein (MOMP)

In addition to the LPS/O-antigen, this study also attempted to identify other potential virulence factors in this pathogen. The generation of two polyclonal antisera against *V. anguillarum* O1 and O4 allowed detection of a common immunogenic protein (designated MOMP) in all *V. anguillarum* serotypes (O1-O10) tested in this study. The protein varied in size between the serotypes with the smallest being ~36 kDa and the largest ~40 kDa. The protein was purified from the O1 serotype and the N-terminal sequence was determined. The sequence obtained was found to be 38% identical and 75% similar to an immunogenic protein, P2, in *Haemophilus influenzae*. No other matches were found using the databases searched with the program BLASTP. During this study, Suzuki *et al.* (1996) determined the N-terminal sequence of the MOMP from three *V. anguillarum* serotypes (O1, O2, O3) and found that the amino acids were identical within the strains used. The N-terminal sequence determined from *V. anguillarum* O1 in this study was identical to that reported by Suzuki *et al.* (1996), except for a change at amino acid #4 were

Valine (V) is substituted for Tyrosine (Y) (Chapter 3). Suzuki and colleagues did not report any similarity to the P2 protein. However, they did indicate that their sequence had 30-45% identity to other bacterial porins (OmpF and OmpC). In this study no similarity to OmpF and OmpC was found, using the BLASTP associated databases.

It has been demonstrated that the P2 protein from *H. influenzae* type b has porin activity (Vachon *et al.*, 1986) and that antibody directed against P2 is protective in an infant rat model (Munson *et al.*, 1983). The size of the P2 protein varies between different *H. influenzae* isolates, with the size variation being found to occur at the DNA level which results in proteins of corresponding lengths. Not surprisingly, the P2 proteins from different isolates share common antigenic epitopes (Munson *et al.*, 1983). Given the similarity between P2 and MOMP proteins allows one to speculate on a potential protective role for the MOMP in *V. anguillarum*.

During the course of this study Simon *et al* (1996, 1998) described the MOMP from *V. anguillarum* O1 as a general diffusing porin which had similar properties to OmpF from *E. coli* (Mizano and Mizushima, 1990) as determined by crystallisation and image processing of electron micrographs. In addition, Simon *et al*, (1998) have examined the immunospecificity of antiserum generated against one serotype of *V. anguillarum* against other serotypes, by immunoblotting, ELISA, and dot-blotting and found that the MOMPs from the different *V. anguillarum* serotypes contained cross-reactive epitopes. This correlates to the Western immunoblot data from this study (Chapter 3).

Further work is required to determine the antigenic properties of MOMP and its potential for a cross-protective vaccine for all *V. anguillarum* serotypes. The data presented in this study (Chapter 3, Section 3.2.5) which showed that antibodies against the MOMP were unable to bind, to *V. anguillarum* O1 and O2, as determined by immunoelectron microscopy, but were able to bind to the SR-LPS strain representing serotype O4, suggests that the MOMP is masked by other components of the outer membrane, in particular LPS (Chapter 3, Section 3.3). Therefore it is most unlikely that this protein would be a good vaccine candidate. However, as the protein shows homogeneous size on SDS 15% polyacrylamide gels within the typed O1 strains indicates that the MOMP may be useful in further characterising and typing *V. anguillarum* strains isolated from diseased fish and the environment. This is becoming increasingly important given the number of new strains and serotypes emerging, some of which can not be typed using current methods (Pedersen *et al.*, 1999).

7.3 The O-antigen/LPS biosynthesis locus of *V. anguillarum* O1

The LPS/O-antigen biosynthesis locus (*wbh*) of *V. anguillarum* was isolated and partially characterised in this study. Three cosmid clones (pPM4905, pPM4910, pPM5512) were found to contain what is likely to be most of the *wbh* locus. Sequencing of the region revealed the presence of 24 ORF's designated: *rmlBADC*, *wzm*, *wzt*, *wbhAB*, *ddhDBAC* and *wbhC-wbhN*. The predicted function of their protein products are listed in Table 7.1. Although specific functions of some of the genes will be discussed later in this chapter, the putative functions of the genes were inferred by their sequence homology of their gene products to proteins of known function, gene complementation and information obtained from the preliminary analysis of LPS sugar composition.

7.4 Transcription and Translation

Numerous promoter regions consisting of -35 and -10 consensus sequences were located throughout the *wbh* locus (Section 5.2.8; Appendix 1). It is not known which

Table 7.1: Functions of the *wbh* genes found in *V. anguillarum* O1

ORF	G+C content[a]	Putative Function	Homology to	Accession Number
rfaD/gmhD	43.7	core biosynthesis	*V. cholerae* O1 *rfaD*	X59554
rmlB	43.3	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB010415
rmlA	45.8	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB010415
rmlD	48.2	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB002668
rmlC	40.5	rhamnose biosynthesis	*A. actinomycetemcomitans*	AB010415
wzm	35.7	O-antigen export (channel)	*B. subtilis tagG*	P42953
wzt	33.9	O-antigen export (energiser)	*En. faecalis* ATP binding protein	AF071085
wbhA	31.7	unknown		
IS*1358*	43.2	insertion sequence		
wbhB	30.2	unknown- transferase		
ddhD	33.9	CDP-dideoxyhexose pathway	*Y. pseudotuberculosis ascD*	P37911
ddhA	38.3	CDP-dideoxyhexose pathway	*S. typhimurium rfbF*	P26396
ddhB	41.1	CDP-dideoxyhexose pathway	*Y. pseudotuberculosis*	L33181
ddhC	38.4	CDP-dideoxyhexose pathway	*Y. pseudotuberculosis*	E47070
wbhC	36.4	acetolactate synthetase	*E. coli ilvG*	P00892
wbhD	32.5	unknown		
wbhE	39.0	glucosamine biosynthesis	*M. thermoautotrophicum*	AE000818
wbhF	19.4	unknown		
wbhG	32.3	bactoprenol glucosyl transferase	*B. subtilis csbB*	Q45539
wbhH	32.4	unknown		
wbhI	36.3	unknown		
wbhJ	36.6	acetyl transferase	*B. pertussis blpB*	X90711
wbhK	37.8	transaminase	*S. venezuelae*	AF079762
wbhL	31.2	glycosyl/rhamnose transferase	*V. anguillarum virB*	L08012
			H. influenazae lgtD	Q57022
wbhM	30.9	unknown	*V. anguillarum virA*	L08012
wbhN	37.5	rhamnosyl transferase	*S. flexneri rfbF*	S51266

a. Percent guanine plus cytosine (G+C) of coding region promoters are functional and further investigation is required. However, the only obvious stemmed-loop structure, likely to be a transcriptional terminator, was located after the gene designated *wbhN*. This may suggest that the *wbh* locus is transcribed as a single mRNA. Although, another possible explanation is that the terminator is used during the transcription of *wbhN*, which potentially has its own promoter (Section 5.2.8).

A sequence resembling the JUMPstart sequence was located in the intergenic region between *gmhD* and *rmlB*. The JUMPstart sequence is involved in transcriptional regulation (Marolda and Valvano, 1998) of polysaccharide biosynthesis loci in *E. coli kps, E. coli* K-12 *waa, S. flexneri rfb*, *S. enterica wba*, *V. cholerae wbe*, and *Y. enterocolitica wbb* (Hobbs and Reeves, 1994). Regulation involving the JUMPstart sequence involves the known transcription elongation factor *rfaH* (*waaH*). RfaH/WaaH permits transcription to proceed over long distances (Bailey *et al*., 1997). Any mutation effecting *rfaH/waaH* or the regulatory sequence leads to premature termination of transcription (Marolda and Valvano, 1998). It is therefore likely that the regulation of transcription of the *wbh* region of *V. anguillarum* involves *rfaH/waaH* and other unknown factors.

All of the genes in the *wbh* region, with the exception of *wbhG*, were found to have putative RBSs which closely resemble the Shine-Dalgarno consensus sequence and were positioned at an optimal distance from the initiation codon (Chapter 5). The initiation codon of *wbhG* was found to overlap with the termination codon of *wbhF* suggesting that these genes may be translationally coupled. It is likely that *wbhG* is not highly expressed which is supported by the overexpression assays performed in this study where *wbhG* was not highly expressed *in E. coli* (Chapter 5, Section 5.2.9). Overlap of initiation and termination codons was found throughout the *wbh* operon, with *rmlA*/*rmlD*, *ddhB*/*ddhC*, *wbhD*/*wbhE* and *wbhJ*/*wbhK* all displaying this feature. It is highly unlikely that the whole *wbh* region is translationally coupled given the large intergenic regions and hence most of the gene products are likely to be translated from their own RBS (Section 5.2.8).

7.5 G+C content

The O-antigen/LPS biosynthesis locus of *V. anguillarum* O1 appears to be comprised of genes that may have been acquired from different sources. The %G+C content of parts of the *V. anguillarum wbh* region do not correlate to the average %G+C content for this organism (44.4%) suggesting that some of the genes in this locus may have been acquired from different sources, which is not uncommon for O-antigen/polysaccharide biosynthesis loci (Xiang *et al.*, 1993). Generally the %G+C content of the genes in the *wbh* locus were below 40% with the exception of *rmlBADC* and *ddhB*; *rmlD* had the highest %G+C content of 48.2%, and *wbhF* possessed the lowest 19.4% (Table 7.1).

Further evidence that some of the *wbh* genes may have been obtained by horizontal gene transfer from different sources, at some point in the evolution of the species, stems from DNA sequence identity. For example, the *rml* genes of *V. anguillarum* O1 show between 61-74% identity to the *rml* genes from *A. actinomycetemcomitans*, 57-68% identity to the equivalent genes from *N. meningitidis*, and 62-67% identity to the *S. enterica* group B and *S. flexneri* rhamnose genes (Chapter 6), indicating these genes originated from a common source, and were acquired by the respective bacteria wherein mutations over time have lead to sequence differences.

7.6 Composition of the LPS from *V. anguillarum* O1.

Preliminary sugar analysis performed on the LPS of the wildtype strains of *V. anguillarum* O1 (NB10 and 85-3954-2) has indicated that the molecule is composed of three sugars: glucose, rhamnose and glucosamine, in a ratio of 10:1:1. Unfortunately, the sugar analysis was performed on the whole LPS molecule which makes it difficult to allocate sugars to the either the core or O-antigen region, although analysis of the LPS from several *wbh* O-antigen mutants allows one to make some reasonable interpretations.

SDS-PAGE and silver staining of LPS indicated high levels of core oligosaccharide material (R-LPS) with respect to S-LPS in both the wildtype (NB10 and 85-3954-2) and mutant (KM2, KM16, KM97, KM110) strains suggesting that the core oligosaccharide is not effected by the insertions in the *wbh* genes of the above mutants. As the sugar analysis indicated that all the strains tested have high levels of glucose it is possible that this sugar constitutes part of the core region, although it may also be a modification on, or component of, the O-antigen, as these possibilities can not be ruled out from the data obtained from this study (see also Section 7.9.1).

The presence of Wzm/Wzt in the *wbh* region of *V. anguillarum* O1 supports the likelihood that the O-antigen of this serotype is a homopolymer, as these proteins show significant identity to proteins of the ABC-2 transporter family (Chapter 6). The ABC-2 transporter family have to date only been associated with the export of homopolymers (Whitfield, 1995).

7.7 Mutagenesis and phenotypic analysis of mutants

Sugar analysis of the *wbhE*::Tn*phoA* (Cm^R) mutant KM16 indicated that this strain had a ratio of glucose to rhamnose identical to the wild type (ie. 10:1), but no glucosamine. Based on the lack of glucosamine, in association with its protein homology to other proteins, *wbhE,* has been assigned a putative role in the biosynthesis of glucosamine. WbhE shows homology to putative perosamine synthetases from *E. coli* and *V. cholerae* which are involved in the production of perosamine, an amino sugar. The observation that KM16 still produces S-LPS with O-antigen (albeit altered) indicates that glucosamine is not the main sugar in the polymer and it is most likely a modification on the O-antigen. The location of the rhamnose is unknown as no mutants were unavailable in the *rml* genes, and hence the sugar could be either in the core or O-antigen (see also Section 7.9.1).

The mutant designated KM97 has a Tn*phoA* (Cm^R) insertion in *wzm*. Wzm shows protein sequence and structural similarity to the integral membrane component of the ABC-2 O-antigen transport system (Chapter 6). Silver staining and Western immunoblotting indicated that no LPS associated O-antigen was produced, and that KM97 had a R-LPS phenotype, as expected for a mutation in O-antigen transport genes. However, subsequent analysis of KM97 by immunofluorescence showed that this strain produced cell surface expressed material that reacted with anti-VagO1. Interpretation of all of the data suggested that KM97 produces O-antigen hapten that is unlinked to lipid A-core which is exported to the surface of the cell by an alternative export pathway. A similar observation has previously been reported in *E. coli* (Morona *et al.*, 1995a). Despite the "hapten" of KM97 not being lipid-A linked, the material was purified using the method described in Section 2.15.2 (Chapter 2) allowing sugar composition analysis to be undertaken. The most likely explanation for this is that the hapten was co-purified with R-LPS (ie. lipid A-core). The sugar analysis of the hapten and R-LPS of this strain indicated the presence of glucose and rhamnose (5:1) but no glucosamine. Assuming glucosamine is a modification (described above) on the O-antigen of *V. anguillarum* O1, its loss from the hapten of KM97 may be explained as modifications are thought to either occur during O-antigen export or be associated with the linkage of the O-antigen to the lipid A-core. However, the absence of glucosamine may also be due to polar effects (Section 6.2.6) on genes downstream of *wzm* required for glucosamine biosynthesis and transfer (Fig. 7.2).

KM110 has a Tn*phoA* insertion in *wbhB* and has S-LPS of decreased molecular weight as observed by SDS-PAGE and silver staining. The sugar analysis of this strain showed an altered glucose:rhamnose ratio (ie. 2.5:1), with no glucosamine detected. The apparent increase in rhamnose may also be interpreted as a decrease in glucose. The role of *wbhB* in the overall biosynthesis of O-antigen/LPS in *V. anguillarum* O1 is harder to elucidate as no homology to other sequences in the database were located. The putative topology of *wbhB* suggests that it is probably an integral membrane protein which spans the membrane 11-12 times (Section 5.2.7.7), a structure suggestive of a polysaccharide-modifying sugar transferase enzyme. This structure has previously been reported for a glucosyl transferase (GtrII) and O-acetylase (Oac) in *S. flexneri* (Mavris *et al.*, 1997; Clark *et al.*, 1991). It is therefore possible that *wbhB* acts as a O-antigen glucosamine transferase. However, as the mutation in this strain was polar (it could not be complemented by *wbhB*, Section 6.2.6), genes downstream are also likely to effect the S-LPS profile.

7.8 The rhamnose biosynthesis and transferase genes

The analysis of the sugars present in *V. anguillarum* O1 showed that rhamnose was present which was expected from the sequence as *rmlBADC* were located at the 5‘ end of the region; these genes demonstrated significant similarity to genes previously shown to determine of synthesis dTDP-L-rhamnose. In addition *rmlD* was confirmed to be the dTDP-rhamnose synthetase gene by complementation of an *S. flexneri rmlD* mutant (Chapter 6, Section 6.2.8.1).

For rhamnose to be constituent of the LPS of *V. anguillarum* O1 rhamnosyl transferases are required. A potential rhamnosyl transferase gene was located near the 3‘

end of the sequenced region. *wbhN* shows significant homology to the rhamnosyl transferase (*rfbF*) of *S. flexneri*. Interestingly, *wbhL* (*virB*) which shows similarity to glycosyl transferases has a functional motif similar to *wbhN*. This motif was previously described by Morona *et al* (1995a), for *rfbF* as being characteristic for 6-deoxyhexose transferases. Therefore, it is postulated that *wbhL* and *wbhN* are both rhamnosyl transferases although it is not known if both are functional. *wbhL* (*virB*) has previously been shown to be important in LPS biosynthesis, as it encoded an essential protein, as determined by mutagenesis (Norqvist and Wolf-Watz, 1993). However, from the data presented by Norqvist and Wolf-Watz (1993) it is unclear as to whether the *virB* mutant prevented O-antigen biosynthesis or if the mutant had an altered LPS profile (see Section 7.11). T7 overexpression allowed expression of *wbhN* but no product was detected for *wbhL*. However, in a previous study (Norqvist and Wolf-Watz, 1993) a gene product for *wbhL*/*virB* was reported. It is possible that the plasmid vector used in this study did not allow the *wbhL* gene product to be expressed.

7.8.1 A role for the homopolymeric tract of cytidine residues?

Homopolymeric tracts of cytidine residues have been reported to be involved in capsule and lipooligosaccharide (LOS) phase variation in *Neisseria meningitidis* by causing slipped-strand mispairing, resulting in the addition of different sugar residues (Hammerschmidt *et al.*, 1996; Kahler and Stephens, 1998). The novel mechanism involved in the slipped-strand mispairing is postulated to be the result of *recA*-independent events during DNA replication and repair that causes an insertion or deletion of nucleotide residues (Levinson and Gutman, 1987). The homopolymeric tract of cytidine residues in

N. meningitdis is present in the 5‘ end of a gene which encodes a glycosyl transferase enzyme (Hammerschmidt *et al.*, 1996; Kahler and Stephens, 1998).

Hammerschmidt *et al.* (1996) have shown a direct correlation between the ability of *N. meningitidis* to invade epithelial cells (and hence cause disease) and the expression of capsule. Only unencapsulated variants are able to enter epithelial cells. Examination of encapsulated and unencapsulated variants showed that the activity of the enzyme, polysialyltransferase required for the production of α-2,8 polysialic acid capsule, was absent in capsule-negative bacteria. Further analysis of the gene (*siaD*) encoding this enzyme indicated the involvement of a homopolymeric tract of seven cytidine residues in the regulation of the essential enzyme. DNA sequence analysis showed that variants (ie unencapsulated) had either a deletion or insertion of a cytidine residues which results in frameshift and hence premature termination. This reversible mechanism allows bacterium to gain an advantage with the host allowing it to cause disease (Hammerschmidt *et al.*, 1996). A similar mechanism has been proposed for variation in LOS of *N. meningitidis* (Kahler and Stephens, 1998).

Analysis of the *wbhN* sequence revealed the presence of 11 cytidine residues adjacent and within the 5‘ end of the ORF (Section 5.2.7.6.4) suggesting that a similar mechanism of variation to that described above may be occurring in *V. anguillarum* O1, although no direct evidence has been obtained to show any role this homopolymeric tract may have in gene expression and hence O-antigen/LPS production. Protein expression data from this study showed that a PCR derived cloned gene of *wbhN* (with all three potential initiation codons) was capable of expressing a 32 kDa protein in *E. coli*, however due to the close proximity of the three start codons (within 50 bp), the size of the protein does not allow the start codon to be determined (Chapter 5). In addition, the overexpression system used in this study allows read through of a TAG stop codons early in the gene due to a *supE* mutation in the *E. coli* strain used and hence the expression seen may be artificial.

The addition or deletion of a cytidine residue in *wbhN* would have no effect on a protein expressed exclusively from the leucine initiation codon which is located after the homopolymer tract. However a frameshift would occur for either the methionine or valine start (Chapter 5, Fig. 5.20). For expression to occur from the valine codon (GTG) it would depend on the suppression or (readthrough) of a TAG stop codon to produce a protein. However, should a cytidine residue be deleted, a frameshift would occur such that no protein would be expressed due to two stop codons (TGA, TAA). Addition of a cytidine residue restores the situation back to the original reading frame. For expression to occur from the methionine residue, a frameshift by the addition of a single cytidine residue is required and would also depend on suppression or readthrough of the TAG stop codon. Deletion of a cytidine residue in this scenario would ensure no protein would be produced or expressed. Naturally occurring suppression or readthrough of an internal amber codon has previously been reported for expression of proteins involved in the biosynthesis and assembly of CS3 pili in *E. coli* (Jalajakumari *et al.*, 1989).

Norqvist and Wolf-Watz (1993) reported the sequence of *virBA* which also included the beginning of the ORF designated *wbhN* in this study. The homopolymeric tract of cytidine residues in the O1 strain (775.17B) sequenced in the previous study was found to contain only 10 cytidine residues in comparison to the 11 residues in the strain used in this study (85-3954-2), suggesting that the number of cytidine residues present in this region are capable of variation and hence it is tempting to speculate that this homopolymeric tract may play in the regulation of this putative glycosyltransferase expression.

If the entire *wbh* region sequenced is transcriptionally coupled it would be expected that the transposon mutants (KM16, KM97, KM110) analysed in this study would effect the expression of the *wbhN* gene product (putative rhamnosyl transferase), as the mutations were found to be polar (Chapter 6). However, the mutants produce S-LPS with altered O-antigens, as determined by SDS-PAGE and silver staining which have been shown by serum absorption studies to be antigenically different to the wild type O-antigen (Chapter 6) but still contain similar sugars (rhamnose, glucose) to that of the wild type. A number of scenarios are possible to explain these observations: 1) the *wbhN* gene is transcribed from the promoter located at its 5' end (Chapter 5) and hence is not effected by the polar mutations, 2) another transferase is located elsewhere in the chromosome which compensates for the polar mutation, 3) the polar mutation effects other *wbh* genes which normally function as the transferase enzymes in a wild type situation allowing *wbhN* to be expressed due to frameshift, 4) *wbhN* is not involved in the transfer of sugars (rhamnose) to the O-antigen polymer, or 5) *wbhN* is a redundant gene and doesn't produce a protein, or the protein is non-functional.

7.9 Putative functions for other *wbh* genes

The number of genes present in the sequenced *wbh* region suggests a more complicated O-antigen structure instead of a simple homopolymer with modifications as implied by the sugar analysis. It is difficult to assign functions to all of the genes and further investigations are required to determine which genes are essential for O-antigen expression.

The gene products of *ddhDABC* demonstrate homology to the enzymes required for synthesis of CDP-3,6 dideoxyhexose sugars in *Y. pseudotuberculosis* and *S. enterica* sv *typhimurium*. Initial sugar analysis of the LPS from *V. anguillarum* O1 did not show the presence of any known 3,6 dideoxyhexose suggesting that these genes may not be expressed. A relevant observation was the notable absence of a homologue to the final enzyme in the CDP 3,6 dideoxyhexose pathway which is present in *Salmonella* and *Yersinia*. Given the high percentage of homology for DdhDABC to their respective homologues this does suggest that the final enzyme is missing in *V. anguillarum*. Another possibility is that one of the genes in the *wbh* locus with unknown function may actually undertake the role of being the last enzyme and hence a novel sugar may be produced that was not detected during the sugar analysis.

Following the *ddh* genes are genes with no homologues (*wbhD*, *wbhF*, *wbhH*, *wbhI*, *wbhM*), a putative amino sugar synthetase (*wbhE*), putative modification enzymes (*wbhG*, *wbhJ*, *wbhK*) and sugar transferases (*wbhL*, *wbhN*).

In addition to the genes in the contiguous *wbh* region there appear to be additional genes required for expression of O-antigen. This was supported by an O-antigen mutant (KM2) and sequencing downstream of *wbhN*. These additional genes were not linked to the *wbh* loci described in this study. However, given the phenotype of KM2 which has LPS with a slightly increased molecular weight O-antigen, it does appear the LPS/O-antigen biosynthesis locus of *V. anguillarum* O1 that has been sequenced is incomplete.

7.9.1 Previous studies of *V. anguillarum* O1 LPS: composition and structure

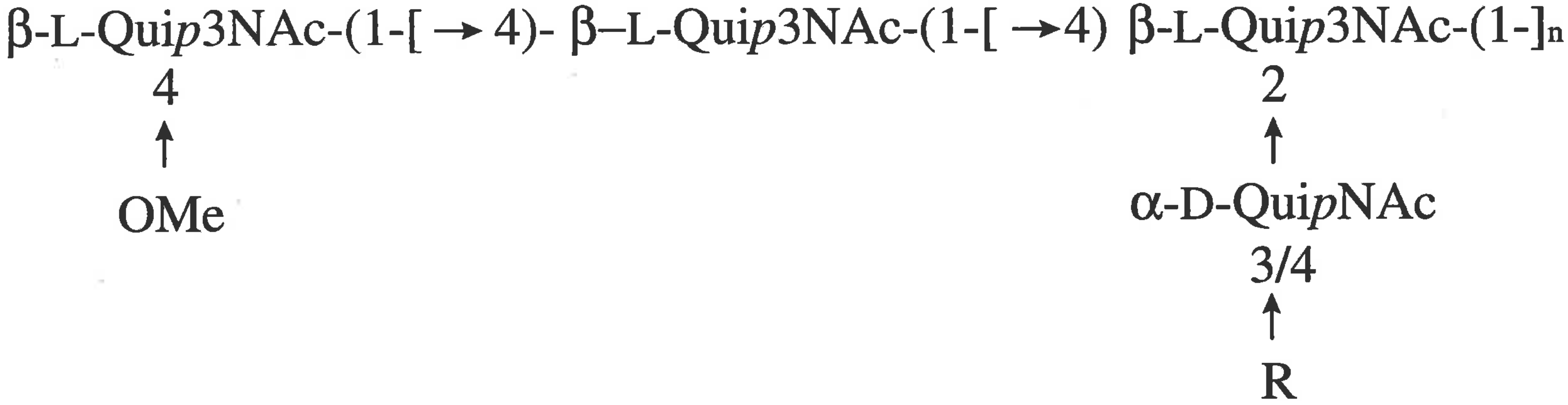
There has been some controversy as to whether *V. anguillarum* produces more than one type LPS. Different researchers (Banoub and Shaw, 1981; Eguchi *et al.*, 1992; Norqvist and Wolf-Watz, 1993) have isolated LPS from both the phenol and water phase which they suggest indicates the presence of two different LPS structures which has also been reported in other bacteria (Helander *et al.*, 1992). The investigation reported in this thesis of the LPS profile of *V. anguillarum* O1 and its mutants suggests that the strains within this serotype are able to produce S-LPS with different O-antigens.

A structural study has been performed on the LPS from a strain (#225/SJ-41) believed to belong to the O1 serotype by Banoub and colleagues (1981, 1987). However the researchers were unclear as to the history and origin of the strain and therefore did not conclusively state the serotype of strain #225. In addition, the researchers indicated that the strain (#225) had short O-antigen chains, although a silver stained gel was not shown for comparison to the strains used in this study: it does indicate that #225 may be a variant and not a wild type O1 strain. The analysis of a number of O1 strains in this study suggests that they have relatively long O-antigen chains (Chapter 3, Chapter 6). The analysis of the LPS from this strain #225 indicated that the core oligosaccharide was composed of L-rhamnose, D-glucose, 3-acetomido-3,6-dideoxy-L-glucose, and L-*glycero*-D-*manno*-heptose in molar ratio of 3:1:0.5:3. They reported the O-antigen to be a regular heteropolysaccharide of (1→4)-linked 3-acetamido-3,6-dideoxy-β-L-glucose residues alternatively substituted through O-2 with a side chain residues of 2-acetamido-2,6-dideoxy-α-D-glucose (N-acetylquinovosamine) which was also substituted with propionyl groups (Fig. 7.1).

As the repeating unit of the O-polysaccaharide of this strain consisted of 3-acetamido-3,6-dideoxy-β-L-glucose residues, I propose that the O-polymer is in fact a homopolymer that is substituted by N-acetylquinovosamine (Fig. 7.1). It is interesting that the initial sugar analysis performed on the typed O1 strains 85-3954-2 and NB10 in this study did not indicate the presence of these amino-sugars but rather glucosamine, although it should be pointed out that not all amino sugars were tested in this initial study. The

Figure 7.1: Chemical composition and structure of *V. anguillarum* strain #225/SJ-41.

A schematic representation of the chemical structure of the *V. anguillarum* strain #225/SJ-41, which is tentatively designated as a O1 serotype strain. Abbreviations: L-Qui3Nac, 3-acetamido-3,6-dideoxy-L-glucose; D-QuiNAc, 2-acetamido-2,6-dideoxy-D-glucose; R, propionyl groups; OMe, O-methyl. Diagram adapted from Banoub *et al.*, 1987.



$$\begin{array}{ccc} \beta\text{-L-Qui}p\text{3NAc-}(1\text{-}[\rightarrow 4)\text{-} & \beta\text{–L-Qui}p\text{3NAc-}(1\text{-}[\rightarrow 4) & \beta\text{-L-Qui}p\text{3NAc-}(1\text{-}]_n \\ 4 & & 2 \\ \uparrow & & \uparrow \\ \text{OMe} & & \alpha\text{-D-Qui}p\text{NAc} \\ & & 3/4 \\ & & \uparrow \\ & & \text{R} \end{array}$$

presence of rhamnose and glucose is consistent between the studies although the molar ratios are quite different. The previous study indicated that there was more rhamnose than glucose (3:1) while the analysis conducted on NB10 and 85-3954-2 showed the opposite (1:10).

Although there are similarities between the analyses, there are also significant differences which may indicate that the strain used by Banaub and colleagues was not a an O1 strain, although it may be a variant/mutant. Assuming that the strain used in the 1987 study was an O1, the genetics of the LPS/O-antigen biosynthesis *wbh* locus sequenced and characterised in this study supports a homopolymer O-antigen, which is present in strain #225 and is consistent with my interpretation of the data presented by Banoub *et al.* (1987).

The rhamnose and glucose residues found to be present in the LPS of the strains used this study and that of the strain #225 were shown to be components of the core oligosaccharide. An important observation is that the molar ratios of these residues are different in the independent studies again suggesting that the strains may not be comparable. However it may also suggest that the *rml* genes present in the *wbh* locus are actually core biosynthesis genes, and in the absence of defined mutations of these genes this possibility can not be ruled out. The finding that glucose was a part of the core was not overly surprising as sugar analysis of the mutants had indicated that this may be the case.

Although the other sugars present in the LPS of strain #225 were not found in either the wildtype or mutant strains in the preliminary analysis performed in this thesis, genes present in the *wbh* region could be involved in the synthesis of these sugar residues and modifications. For example, the 3-acetoamido-3,6-dideoxy-L-glucose may be synthesised from the enzymes encoded by *ddhDBAC,* and the L-glycero-D-manno-heptose from *rfaD* which is known to produce this sugar. Other genes in the *wbh* region could then do the necessary acetylation and amination reactions. Assuming that this is the case, this means that the *rfa/waa* genes (core oligosaccharide) and O-antigen biosynthesis genes (*wbh*) are located in the same locus, which is an unusual arrangement, but may explain the presence of the numerous genes that did not seem to be required for the synthesis of a homopolymer O-antigen. It is evident that defined mutations and extensive biochemical work is required for all the genes within the *wbh* region to determine their function in LPS biosynthesis.

7.10 Antigenic Variation

Random mutagenesis of *V. anguillarum* O1 using the transposon Tn5*phoA* (Cm^R) and subsequent screening resulted in the isolation of four mutants (KM2, KM16, KM97, KM110) that did not express O-antigen as determined using an antiserum generated against formalin-killed bacteria. In this study however, three of these mutants (KM2, KM16, KM110) were found to express S-LPS with altered O-antigens, while the remaining mutant (KM97) expressed an O-antigen "hapten "that was not lipid A-core oligosaccharide linked as determined by SDS-PAGE and silver staining (Chapter 6). In addition, further analysis of three of the mutants (KM16, KM97 and KM110) showed that the altered O-antigens and hapten reacted to the anti-VagO1 serum generated against heat-killed bacteria.

The recognition of epitopes on the O-antigen of the mutants by the antiserum raised against heat killed *V. anguillarum* O1, and the data demonstrating that the mutants produce antigenically different LPS (Chapter 6) suggests that the wildtype strain produces small quantities of the altered LPS, undetectable by Western immunoblotting and by silver staining. The immunofluorescence, serum absorption studies and Western immunoblotting of the wildtype strains and mutants strongly suggests that antigenic variation occurs to some degree in *V. anguillarum*. These different LPS epitopes may be occurring by simple spontaneous mutations within the bacterial population that result in altered LPS production reminiscent of the mutants KM16 and KM110. In this study, three different LPS/O-antigen phenotypes were observed: two visible by SDS-PAGE and silver staining that had S-LPS with decreased molecular weight O-antigens (of different degrees), and a haptenic form that could not be visualised as it was not lipid-A linked.

Antigenic variation by simple mutation has been reported in *E. coli* with *S. flexneri* *rfb* genes. Mutations in the *wzx/rfbX* (putative flippase) and *rfbF* (rhamnosyl transferase), when expressed in an *E. coli* background have been shown to result in the biosynthesis of O-antigen that is smooth but antigenically altered presumably due to less rhamnose residues in the O-repeat unit. The altered O-antigen reacted against antisera generated from wildtype *S. flexneri*. However absorptions of polyclonal sera and the use of monoclonal antibodies indicated that the O-antigens were different and that each had specific epitopes (ie sugar linkages) (Morona *et al.*, 1995a). This allows conclusions about the production of O-antigen to be drawn. Defects in particular O-antigen biosynthesis genes may lead to compensation for the loss of that gene product by another similar or related protein leading to the expression of a different O-antigen, which also has implications for the bacteria in relation to evading host immune responses (Morona *et al.*, 1995a).

In *E. coli* O8 a similar scenario has been noted by Sacki *et al.* (1993). *E. coli* K-12 strains harbouring the *rfb* genes of *E. coli* O8 on a plasmid produce antigenically different LPS with some common epitopes. The antigenic differences are thought to be due to the requirement of additional genes to add modifications onto the O8 O-antigen which are not present in *E. coli* K-12. Although the examples of antigenic variation in *S. flexneri* and *E. coli* O8 were performed in heterologous hosts, it suggests that mutations in a wildtype background of genes involved in modification or elongation of the O-antigen could effect the composition of the O-antigen that is expressed. This would have implications on the virulence and antigenicity of the LPS molecule. The observations made in *V. anguillarum* O1 are consistent with these previous reports.

7.11 Vaccines and Serotyping

The fact that there is possibly heterogeneity of O-antigen within a strain of *V. anguillarum* O1 has implications on the serotyping schemes currently in use which are based on LPS and antiserum developed from this molecule. It has long been established that the commercial vaccines available based on LPS preparations are not extremely effective, and that the strain used for the preparation of the vaccine is important for protection. That is, strains of the same serotype from one geographical location may not confer protection if used in a different region or location. The reason for this has not yet been determined. It is therefore possible that full protection may also require the minor/altered S-LPS types such as those observed in this study. The LPS mutants analysed in this study were originally thought to have R-LPS based on Western immunoblotting using an antiserum generated against formalin-killed bacteria (D. Milton, unpublished data). However, subsequent analysis of these mutants using serum generated against heat-killed bacteria showed that these mutants did produce LPS and that it was antigenically different to the wild type LPS, although the molecules may share common epitopes. This suggests that the method of antiserum preparation is important when investigating LPS phenotypes and for the classification of *V. anguillarum* isolates. At this point in time it is not known why the different methods of sera preparation would produce different antibodies, although, it is postulated that the use of formalin may have destroyed O-antigen epitopes, explaining the results obtained with the antiserum raised with heat killed bacteria.

These observations also have implications on the study performed by Norqvist and Wolf-Watz (1993) who described the loss of LPS material for a *virB* (*wbhL*) mutant (putative glycosyl transferase) using Western immunoblotting analysis with an antiserum generated against formalin-killed bacteria. This mutant may actually have LPS with altered O-antigen but was undetected due to the serum used. The results of this study suggest that caution is required when using sera to serotype and determine phenotypes related to the LPS of *V. anguillarum* isolates/strains.

7.12 Virulence of the *wbh* mutants

Virulence assays were performed on the *wbh* mutants (KM2, KM16, KM97 and KM110) by D. Milton (unpublished data) to determine if these mutants were defective at any stage of the disease process as tested by the experimental fish model. From the virulence assays, it is clear that the putative additional genes in *V. anguillarum* O1, thought to be required for O-antigen biosynthesis (Chapter 6), effect the virulence of this bacteria. This conclusion can be drawn as assays performed by both immersion and intraperitoneal injection of the mutant KM2 (which has S-LPS with a slightly increased molecular weight) showed that it possesses wildtype virulence for immersion but is avirulent when injected into the peritoneum. This implies that *otnE*, in to which the transposon has inserted, has a role in invasion and hence the O-antigen chain length of this bacterium is important. Virulence assays with the other mutants were also performed, however the virulence plasmid pJM1, essential for virulence, was lost during the experiment when KM97 (*wzm*) and KM110 (*wbhB*) were isolated, invalidating the data obtained. Strain KM16 [*wbhE* + pJM1] having S-LPS with altered O-antigen of decreased molecular weight, had a loss of virulence for both immersion and injection. This is not overly surprising as Norqvist and

Wolf-Watz (1993) had previously shown a loss of virulence by these methods for the R-LPS mutants, *wbhL* (*virB*) and *wbhM* (*virA*). In addition a recent study by Boesen *et al.* (1999) demonstrated O-antigen length was critical for *V. anguillarum* to survive in fish in relation to complement and serum resistance.

7.13 Presence of IS*1358* in the *wbh* locus of *V. anguillarum* O1 and the putative O-antigen/capsule biosynthesis region of O2.

An interesting feature of the *wbh* locus of *V. anguillarum* O1 was the discovery that the previously characterised insertion sequence, IS*1358,* was associated with this region and aided in the localisation and isolation of this region.

IS*1358*, a member of the IS*As1* family of insertion sequences (Mahillon and Chandler, 1998) has been associated with O-antigen and capsule biosynthesis loci in *V. cholerae* O1, O139 (Stroeher *et al.*, 1995) and O69 (Bik *et al.*, 1996). Interestingly other insertion elements of this family including IS*As1* of *Aeromonas salmonicida* (Gustafson *et al.*, 1994) and the H-repeats of *E. coli* (Hill *et al.*, 1994) have also been associated with cell surface biosynthesis genes. It is therefore postulated that elements in the family IS*As1* have preference for these genetic sequences although they are not located exclusively within these. The mechanism (or the basis) for this preference is unknown. It is known that transposable elements exhibit a degree of preference for a particular insertion point, despite the fact that their mobile nature is homology independent (Dalrymple *et al.*, 1984; Gustafson *et al.*, 1994).

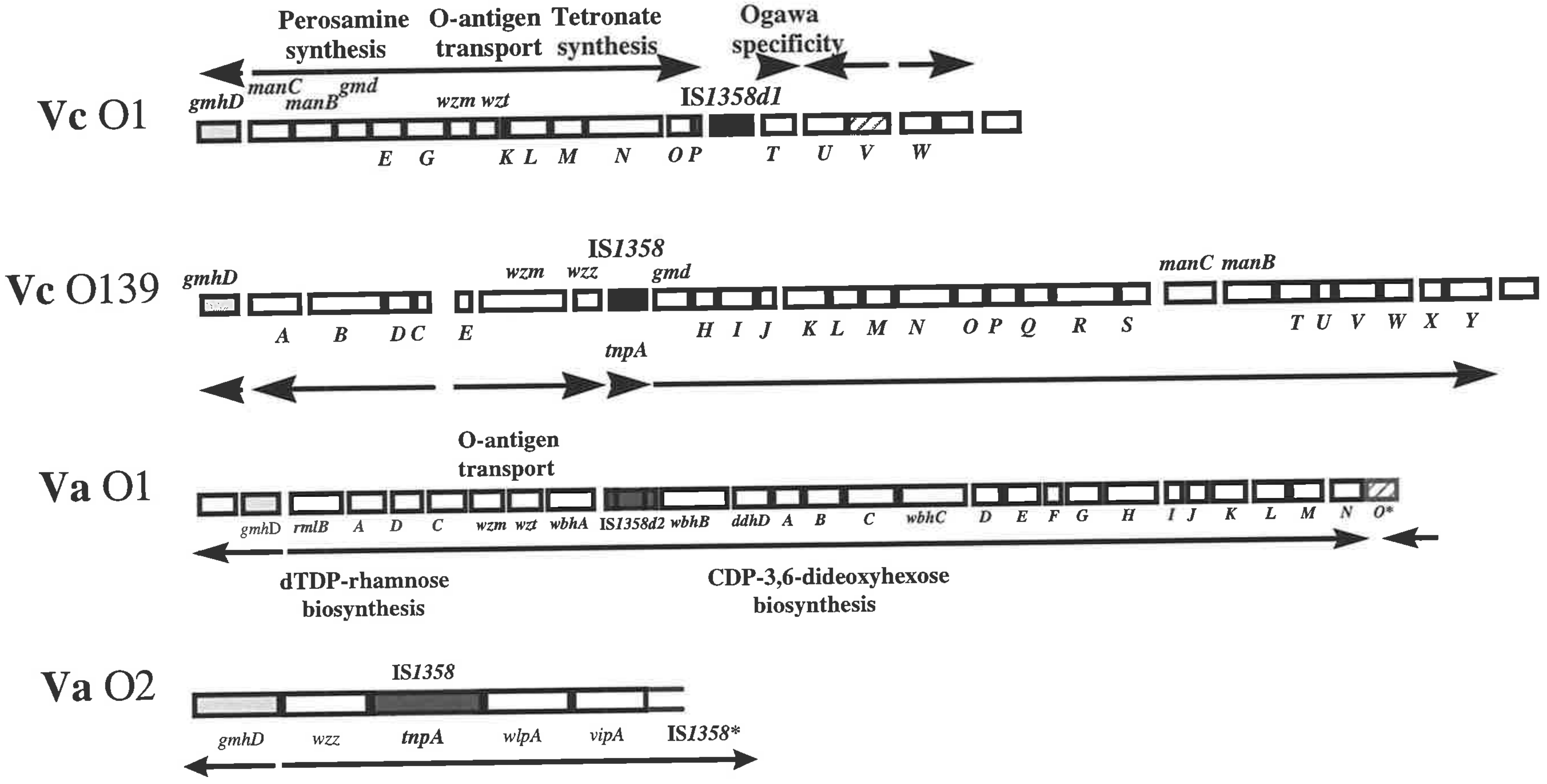
In the *V. anguillarum* O2 serotype two putative *rfb*/polysaccharide genes were located between two IS*1358* elements. These gene products show significant homology to polysaccharide biosynthesis-related genes, including *bpl*A from *Bordetella pertussis* and Vi polysaccharide biosynthesis protein (VipA) from *Salmonella typhi*. The *bplA* homologue was an intact ORF, while the ORF which showed similarity to *vipA* was interrupted by IS*1358* (Chapter 4). Investigations to determine the identity of the sequence beyond the 3‘ end of this IS*1358* element were not successful, although the isolation of cosmid clones will allow future investigations of this region. Two different outcomes may arise from the discovery of the DNA flanking the 3‘ end of this element. If the remainder of the *vipA* gene is present, this suggests that IS*1358* transposed directly into the gene, however should different sequences be revealed this indicates the possibility that the two IS*1358* elements in this region jumped into the chromosome using mechanisms similar to compound transposons.

The organisation of the *wbh* region of *V. anguillarum* O1 (this study) and the putative O-antigen/capsule region in *V. anguillarum* O2 (this study) are similar to that observed in *V. cholerae* O1, O139 and O69. Figure 7.2 compares the O-antigen biosynthesis loci of *V. anguillarum* O1 and O2, *V. cholerae* O1 and O139, and highlights the common features. All of these loci commence with *rfaD/gmhD* transcribed 3' to 5', followed by a JUMPstart sequence and then the polysaccharide biosynthesis genes. The *rfaD/gmhD* gene is involved in core biosynthesis and is involved in biosynthesis of the heptose component of the *V. cholerae* core oligosaccharide. It has been suggested that the conservation of the core structure in *V. cholerae* O1 and O139 allows for compatible linkage of the different O-antigens in these serotypes (Stroeher *et al.*, 1998). The conservation of *rfaD/gmhD* in *V. anguillarum* is not as easily explained, although it does appear to have a similar core oligosaccharide sugars to *V. cholerae,* as determined by chemical analysis (Banoub *et al.*, 1981). It is most likely that *gmhD* is conserved so that it can act as a site of homologous recombination between O-antigen biosynthesis regions

Figure 7.2: Comparison of the O-antigen biosynthesis loci of *V. anguillarum* O1, O2 and *V. cholerae* O1, 0139.

A schematic representation of the O-antigen biosynthesis loci of the *V. cholerae* O1 (Vc O1), O139 (Vc O139) and *V. anguillarum* O1 (Va O1) (Chapter 5), O2 (Va O2) (Chapter 4). Functions that have been assigned to gene blocks are indicated on the figure. The arrows indicate the direction of transcription of the various genes. The lighter shading represents the homology seen between the *rfaD*/*gmhD* genes, the darker shading the IS*1358* element(s) associated with the region. * indicates the presence of a partial ORF.

The O-antigen biosynthesis regions of *V. anguillarum* O1, O2 and *V. cholerae* O1, O139



(Stroeher *et al.*, 1998; Bik *et al.*, 1996) (see below). At the DNA level the *gmhD* genes from *V. cholerae* and *V. anguillarum* O1 are 78 % identical.

In addition to the arrangement observed at the 5' end of the region, *V. anguillarum* O1 and *V. cholerae* O1 have identical arrangements near the ends of their respective O-antigen biosynthesis operons. The *trsG* (*wbhO/wbeV*) homologue is found located at the end of these loci and is divergently transcribed. Interestingly, the *trsG/wbeV* homologue found in *V. cholerae* O1 has been shown to be essential for O-antigen expression in this bacterium (Fallarino *et al.*, 1997). In *V. anguillarum* the role of *trsG* (*wbhO*) in LPS/O-antigen expression has yet to be elucidated.

7.14 A putative role for IS*1358* in shuffling polysaccharide biosynthesis genes?

For IS*1358* to be involved in rearrangements of polysaccharide biosynthesis loci by a process involving transposition it would require a functional transposase. In this study, I showed that only strains with multiple copies of IS*1358* produced a protein of the correct molecular weight (42 kDa) by the T7 RNA polymerase expression system (Chapter 4). It is proposed that the presence of multiple copies indicates the active nature of some of the elements within serotypes O2, O7 and O9. *V. anguillarum* serotypes O2 and O9 have greater than 9 IS*1358* elements as indicated by Southern hybridisation (Chapter 3). The organisation of the IS*1358* elements within the chromosome is reminiscent of complex transposon-like structures (compound) in which two insertion sequences flank the mobile regions. Thus, either excision or transposition of this region would allow the genes trapped between the insertion elements to move.

The single IS*1358* element from *V. cholerae* O139 was shown to produce a protein, and Stroeher *et al.* (1995) suggested that this indicated the element was still potentially active and therefore may be able to transpose. Experiments to show transposition of IS*1358* from *V. cholerae* O139 have been unsuccessful in our laboratory (U. Stroeher, personal communication). Recently, Dumontier *et al.* (1998) reported that IS*1358* from *V. cholerae* O22, which has multiple copies of the element on the chromosome, was able to undergo simple insertion onto a plasmid in *E. coli,* producing 10 bp direct repeats. Given the similarities between the IS*1358* elements of *V. cholerae* O139 and *V. anguillarum* O2, O7 and O9, it is highly probable that these elements would be able to transpose and therefore may be involved in *rfb*/polysaccharide gene rearrangements.

An apparent role for IS*1358* in genetic rearrangements has been suggested (Stroeher *et al.*, 1995, 1998; Bik *et al.*, 1996). Therefore one of the questions asked in this study was whether IS*1358* may have played a role in the acquisition of these genes. Although the question still remains to be answered, data collected in this study provides further indirect evidence from both *V. anguillarum* O1 and O2 that IS*1358* could be involved in gene rearrangements.

In *V. anguillarum* O2, the two ORFs (ORF1, ORF2) located between the IS elements were found to be present in a number of *V. cholerae* serotypes by dot-blot hybridisation (Chapter 4). Some of these also contained the IS*1358* element, and therefore it is conceivable that these genes were acquired from or are being swapped between different species by a mechanism that may involve IS*1358*.

The researchers (Dumontier *et al.*, 1998) who have shown IS*1358* to transpose believe it is unlikely that IS*1358* is involved in the acquisition of novel O-antigen and capsule biosynthesis genes because they had shown that the element mediates direct transposition rather than formation of cointegrates, which are commonly associated with large genome rearrangements. However, this conclusion contradicts their own experimental data. To perform mobility assays they constructed a compound transposon of

IS*1358* flanking a kanamycin cartridge. This construct was capable of transposition (Dumontier *et al.*, 1998). This suggests that any genes trapped between these elements, as observed in *V. anguillarum* O2, would also be transposed providing a mechanism for the acquisition of new genes (or gene shuffling).

The presence of IS*1358* in numerous O-antigen biosynthesis loci is not a coincidence. The development and expression of new O-antigen biosynthesis loci would most likely occur over time with the scenario involving numerous rearrangements to acquire genes that are capable of functioning together to produce new O-antigen material. Even if IS*1358* transposed without additional genes attached, its presence on the chromosome would allow homologous recombination to occur with DNA from an exogenous source that also contained IS*1358* and possibly *rfaD*/*gmhD* or *trsG*. Using the data presented in this study and from other studies, it is unreasonable to dismiss IS*1358* as a player in genetic rearrangements and hence the evolution of new *rfb*/capsule loci until evidence is provided to the contrary.

Therefore I suggest that the most likely explanation for the involvement of IS*1358* in the evolution of new polysaccharide operons is by a mechanism of rearrangement that requires initial transposition of IS*1358* into the region containing polysaccharide genes and then subsequent homologous recombination events between polysaccharide operons containing the element.

7.15 Future studies

There remains unanswered questions regarding LPS/O-antigen biosynthesis *in V. anguillarum*. To further determine the role of LPS in virulence and vaccine development defined mutants of the *wbh* locus of *V. anguillarum* O1 are required. This will enable specific functions and role in LPS/O-antigen biosynthesis to be determined. In addition, detailed biochemical analysis of the *wbh* gene products would also be required to confirm proposed functions. The role of the additional genes located near the *wbh* locus also need to be further investigated to determine their precise role in LPS/O-antigen biosynthesis and their location in the chromosome. A structure analysis of the S-LPS of a defined *V. anguillarum* O1 type strain (eg 85-3954-2/NB10) is essential to elucidate the functions of the *wbh* genes.

There is considerable interest in determining the origins of genes in bacteria and hence the role that IS*1358*, and other related elements, may play in the evolution of polysaccharide biosynthesis loci in Vibrionaceae needs to be further investigated. It is important therefore to demonstrate transposition activity of IS*1358* in an autologous host (eg *V. cholerae* or *V. anguillarum*), as this element has only be shown to be mobile in a heterologous host (*E. coli*).

BIBLIOGRAPHY

Achtman, M., Manning P.A., Edelbluth, C., and Herrlich, P. (1979) Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor *tra* cistrons in *Escherichia coli* minicells. *Proc. Natl. Acad. Sci. USA.* **76:** 4837-4841.

Actis, L.A., Fish, W., Crosa, J.H., Kellerman, K., Ellenberger, S.R., Hauser, F.M., and Sanders Loehr, J. (1986) Characterization of anguibactin, a novel siderophore from *Vibrio anguillarum* 775(pJM1) *J. Bacteriol.* **167:** 57-65.

Actis, L.A., Tolmasky, M.E., Farrell, D., and Crosa, J.H. (1988) Genetic and molecular characterisation of essential components of the *Vibrio anguillarum* plasmid-mediated iron transport system. *J. Biol. Chem.* **263:** 28853-28860.

Actis, L.A., Potter, S.A., and Crosa, J.H. (1985) Iron-regulated outer membrane protein OM2 of *Vibrio anguillarum* is encoded by virulence plasmid pJM1. *J. Bacteriol.* **161:** 736-42.

Actis, L.A., Tolmasky, M.E., Crosa, L.M., and Crosa, J.H. (1995) Characterisation of the expression of the FatB, an iron transport protein encoded by the pJM1 virulence plasmid. *Mol. Microbiol.* **17:** 197-204.

Agre, P., Johnson, P., and Mcknight, S. (1989) Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science* **246:** 922-925.

Alekseeva, L.P., Iia, C., OI, S., and OS, B. (1998) A monoclonal antibody-based study of the antigenic interrelations of typical and R- forms of *Vibrio cholerae. Zh Mikrobiol Epidemiol Immunobiol.* **4:** 9-12. Russian.

Alexander, D. C., and Valvano, M.A. (1994) Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O-specific polysacchardies containing *N*-acetylglucosamine. *J. Bacteriol.* **176:** 7079-7084.

Allen, A., and Maskell, D. (1996) The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in *Bordetella pertussis. Mol Microbiol.* **19**: 37-52.

Alm, R.A. Stroeher, U.H., and Manning, P.A. (1988) Extracellular proteins of *Vibrio cholerae*: nucleotide sequence of the structural gene (*hlyA*) for the haemolysin of the haemolytic El Tor strain O17 and characterisation of the *hlyA* mutation in the non-haemolytic classical strain 569B. *Mol. Microbiol.* **2:** 481-488.

Alm, R.A., Mayhofer, G., Kotlarski, I., and Manning, P.A. (1991) Amino-terminal domain of the El-Tor haemolysin of *Vibrio cholerae* O1 is expressed in classical strains and is cytotoxic. *Vaccine* **9:** 588-594.

Alsina, M., Martinez Picado, J., Jofre, J., and Blanch, A.R. (1994) A medium for presumptive identification of *Vibrio anguillarum. Appl. Environ. Microbiol.* **60:** 1681-3.

Ames, G.F.-L., Mimura, C.S., Holbrook, S.R., and Shyamala, V. (1992) Traffic ATPases: a superfamily of transport proteins operating from *Escherichia coli* to humans. *Adv. Enzymol.* **65:** 1-47.

Amor, P.A., and Mutharia, L.M. (1995) Cloning and expression of rfb genes from *Vibrio anguillarum* serotype O2 in *Escherichia coli*: evidence for cross-reactive epitopes. *Infect-Immun.* **63:** 3537-42.

Anderson, M.S., Bulawa, C.E., and Raetz, C.R.P. (1985) The biosynthesis of gram negative endotoxin: formation of lipid A precursors from UDP-GluNAc in extracts of *Escherichia coli. J. Biol. Chem.* **260:** 15536-15541.

Aoki, T., Kitao, T., and Kawano, T. (1981) Changes in drug resistance of *Vibrio anguillarum* in cultured ayu, *Plecoglossus altivelis* Temmick and Schegel, in Japan. *J. Fish Dis.* **4:** 223-230.

Aoki, T., and Hirono, I. (1991) Cloning and characterisation of the haemolysin determinants from *Aeromonas hydrophila*. *J. Fish Dis.* **14:** 303-312.

Aragon, V., Diaz, R., Moreno, E., and Moriyon, I. (1996) Characterisation of Brucella abortus and *Brucella melitensis* native haptens as outer membrane O-type polysaccharides independent from the smooth lipopolysaccharide. *J. Bacteriol.* **178:** 1070-1079.

Austin, B., Alsina, M., Austin, D.A., Blanch, A.R., Grimont, F., Grimont, P.A.D., Jofre, J., Koblavi, S., Larsen, J.L., Pederson, K., Tianen, T., Verdonck, L., and Swings, J. (1995) Identification and typing of *Vibrio anguillarum*: a comparison of different methods. *Syst. Appl. Microbiol.* **18:** 285-302.

Austin, E.A., Graves, J.F., Hite, L.A., Parker, C.T., and Schnaitman, C.A. (1990) Genetic analysis of lipopolysaccharide core biosynthesis by *Escherichia coli* K-12: insertion mutagenesis of the *rfa* locus. *J. Bacteriol.* **172:** 5312-5325.

Bailey, M.J., Hughes, C., and Koronakis, V. (1997) RfaH and the *ops* element, components of a novel system controlling bacterial transcription elongation. *Mol. Microbiol.* **26:** 845-851.

Baker, S.J., Gunn, J.S., and Morona, R. (1999) The *Salmonella typhi* melittin resistance gene *pqaB* affects intracellular growth in PMA-differentiated U937 cells, polymyxin B resistance and lipopolysaccharide. *Micro.* **145:** 367-378.

Banoub, J.H., Michon, F., and Hodder, H.J. (1987) Structural elucidation of the O-specific polysaccharide of the phenol-phase soluble lipopolysaccharide of *Vibrio anguillarum*. *Biochem-Cell-Biol.* **65:** 19-26.

Banoub, J.H., and Shaw, D.H. (1981) Isolation and characterization of 3-acetamido-3,6-dideoxy-L-glucose from the core oligosaccharides obtained from the aquatic gram-negative bacteria *Aeromonas hydrophila* and *Vibrio anguillarum*. *Can-J-Biochem.* **59:** 877-9.

Bastin, D.A., and Reeves, P.R. (1995) Sequence analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. *Gene* **164:** 17-23.

Bastin, D.A., Stevenson, G., Brown, P.K., Haase, A., and Reeves, P.R. (1993) Repeat unit polysaccharides of bacteria: a model for polymerisation resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* **7:** 725-734.

Batchelor, R.A., Haraguchi, G.E., Hull, R.A., and Hull, S.I. (1991) Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *J. Bacteriol.* **173:** 5699-5704.

Beachey, E.H. (1981) Bacterial adherence: adhesion-receptor interactions mediating the attachment of bacteria to mucosal surface. *J. Infect. Dis.* **143:** 325-345.

Beckman, I., Subbaiah, T.V., and Stocher, B.A.D. (1964) Rough mutants of *Salmonella typhimurium*. 2. Serological and chemical investigations. *Nature* (London) **1:** 1299-1301.

Bell, J., Grass, S., Jeanteur, D., and Munson, R.S. Jr. (1994) Diversity of the P2 protein among nontypeable *Haemophilus influenzae* isolates. *Infect. Immun.* **62:** 2639-2643.

Bell, A., Bain, M., and Hancock, R.E.W. (1991) *Pseudomonas aeruginosa* outer membrane protein OprH: expression from cloned genes and function in EDTA and gentamycin resistance. *J. Bacteriol.* **173:** 6657-6664.

Benz, R., and Bauer, K. (1988) Permeation of hydrophilic molecules through the outer membrane of gram-negative bacteria. *Eur. J. Biochem.* **176:** 1-19.

Bik, E.M., Bunschoten, A.E., Gouw, R.D., and Mooi, F.R. (1995) Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J.* **14:** 209-216.

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

Boesen, H.T., Pedersen, K., Larsen, J.L., Koch, C., and Ellis, A.E. (1999) *Vibrio anguillarum* resistance to rainbow trout (Oncorhynchus mykiss) serum: role of O-antigen structure of lipopolysaccharide. *Infect. Immun.* **67:** 294-301.

Boliches, J., Lemos, M.L., Fouz, B., Cambra, M., Larsen, J.L., and Toranzo, A.E. (1990) Serological relationships among *Vibrio anguillarum* strains. *J. Aquat. Anim. Health.* **2:** 21-29.

Boman, H.G., and Monner, D.A. (1975) Characterisation of lipopolysaccharides from *Escherichia coli* K-12 mutants. *J. Bacteriol.* **121:** 455-464.

Bowser, P.R., Rosemark, R., and Reiner, C.R. (1981) A preliminary report of vibriosis in cultured American lobsters, *Hormarus americanus*. *J. Invertebr. Pathol.* **37:** 80-85.

Brade, H. (1985) Occurrence of 2-keto-deoxyoctonic 5-phosphate in lipopolysaccharides of *V. cholerae* Ogawa and Inaba. *J. Bacteriol.* **161:**795-798.

Broady, K.W., Rietschel, E., and Luderitz, O. (1981) The chemical structure of the lipid A component of lipopolysaccharides from *Vibrio cholerae*. *Eur. J. Biochem.* **115:**463-468.

Bronner, D., Clarke, B.R., and Whitfield, C. (1994) Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccahride O-antigen side chains across the cytoplasmic membrane of *Klebsiella pnuemoniae* serotype O1. *Mol. Microbiol.* **14:** 505-519.

Brown, M.C.M., Weston, A., Saunders, J.R., and Humphreys, G.O. (1979) Transformation of *E. coli* C600 by plasmid DNA at different phases of growth. *FEMS Microbiol. Lett.* **5:** 219-222.

Brown, P.K., Romana, L.K., and Reeves, P.R. (1992) Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar *muenchen* (strain M67), the genetic basis of the polymorphism between groups C2 and B. *Mol. Microbiol.* **6:** 1385-1394.

Brynestad, S.. Synstad, B., and Granum, P.E. (1997) The *Clostridium perifringens* eneterotoxin gene is on a transposable elelment in type A human food poisoning strains. *Microbiol.* **143:** 2109-2115.

Buckley, J.T., Howard, S.P., and Trust, T.J. (1981) Influence of virulence plasmid and geographic origin on outer membrane proteins of *Vibrio anguillarum*. *FEMS Microbiol. Lett.* **11:** 41-46.

Bullen, J.J. (1981) The significance of iron in infection. *Rev Infect Dis.* **3:** 1127-1138.

Burrows, L.L., Chow, D., and Lam, J.S. (1997) *Pseudomonas aeruginosa* B-band O-antigen chain length is modulated by Wzz (Rol). *J. Bacteriol.* **179:** 1482-1489.

Burrows, L.L., Charter, D.F., and Lam, J.S. (1996) Molecualr characterisation of the *Pseudomonas aeruginosa* O5 (PAO1) B-band lipopolysaccharide gene cluster. *Mol. Microbiol.* **22:** 481-495.

Burrows, W., Mather, A.N., McGann, V.G., and Wagner, S.M. (1946a) Studies on Immunity to Asiatic cholera. Part I: Introduction. *J. Infect. Dis.* **79:** 159-167.

Burrows, W., Mather, A.N., McGann, V.G., and Wagner, S.M. (1946b) Studies on Immunity to Asiatic cholera. Part II: The O and H antigenic structure of the cholera and related vibrios. *J. Infect. Dis.* **79:** 168-179.

Caroff, M., Bundle, D.R., and Perry, M.B. (1984) Structure of the O-chain of the phenol-phase soluble cellular lipopolysaccharide of *Yersinia enterocolitica* serotype O9. *Eur. J. Biochem.* **139:** 195-200.

Chart, H., and Trust, T.J. (1984) Characterization of the surface antigens of the marine fish pathogens *Vibrio anguillarum* and *Vibrio ordalii*. *Can. J. Microbiol.* **30:** 703-10.

Chen, Q., and Crosa, J.H. (1996) Antisense RNA, fur, iron, and the regulation of iron transport genes in *Vibrio anguillarum*. *J. Biol. Chem.* **271:** 18885-91.

Chen, D., and Hanna, P.J. (1992) Attachment of *Vibrio* pathogens to cells of rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J. Fish Dis.* **15:** 331-337.

Chen, D., and Hanna, P.J. (1994) Immunodetection of specific *Vibrio* bacteria attaching to tissues of the giant tiger prawn Penaeus monodon. *Dis. Aquatic Organsims*. **20:** 159-162.

Chen, D., Hanna, P.J., Altman, K., Smith, A., Moon, P., and Hammond, L.S. (1992) Development of monoclonal antibodies that identify *Vibrio* species commonly isolated from infections of humans, fish and shellfish. *Appl. Environ. Micro*. **58:** 3694-3700.

Chen, Q., Wertheimer, A.M., Tolmasky, M.E., and Crosa, J.H. (1996) The AngR protein and the siderophore anguibactin positively regulate the expression of iron-transport genes in *Vibrio anguillarum*. *Mol. Microbiol.* **22:** 127-34.

Chong, P., Yang, Y-P., Fahim, R., McVerry, P., Sia, C., and Klein, M. (1993) Immunogenicity of overlapping synthetic peptides covering the entire sequence of *Haemophilus influenzae* type b outer membrane protein P2. *Infect. Immun.* **61:** 2653-2661.

Clarke, C.A., Beltrame, J., and Manning, P.A. (1991) The *oac* gene encoding a lipopolysaccharide O-antigen acetylae maps adjacent to the intergrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. *Gene* **107:**43-52.

Clarke, B.R., and Whitfield, C. (1992) Molecular cloning of the *rfb* region of *Klebsiella pneumoniae* serotype O1:K20: the *rfb* gene cluster is responsible for the synthesis of the D-galactan I O polysaccharide. *J. Bacteriol.* **174:** 4614-4621.

Clementz, T. (1992) The gene coding for 3-deoxy-D-*manno*-octulosonic acid transferase and the *rfaQ* gene are transcriced from divergently arranged promoters in *Escherichia coli*. *J. Bacteriol.* **174:** 7750-7756.

Coleman, W.G. Jr. (1983) The *rfaD* gene codes for ADP-L-glycero-D-mannoheptose-6-epimerase. *J. Biol. Chem.* **258:** 1985-1990.

Coleman, W.G. Jr., and Leive, L. (1979) Two mutations which affect the barrier function of the *Escherichia coli* K-12 outer membrane. *J. Bacteriol.* **139:** 899-910.

Collins, C.M., and Gutman, D.M. (1992) Insertional inactivation of an *Escherichia coli* urease gene by IS*3411*. *J. Bacteriol.* **174:** 883-888.

Collins, L.V., and Hackett, J. (1991) Molecular cloning, characterisation, and nucleotide sequence of the *rfc* gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. *J. Bacteriol.* **173:** 2521-2529.

Comstock, L.E., Johnson, J.A., Michalski, J.M., Morris Jr, J.G., and Kaper, J.B. (1996) Cloning and sequencing of a region encoding a surface polysaccharide of *Vibrio cholerae* O139 and characterisation of the insertion site in the chromosome of *Vibrio cholerae* O1. *Mol. Microbiol.* **19**: 815-826.

Comstock, L.E., Maneval Jr, D., Panigrahi, P., Joseph, A., Levine, M.M., Kaper, J.B., Morris Jr, J.G., and Johnson, J.A. (1995) Capsule and O antigen in *Vibrio cholerae* O139 Bengal associated with a genetic region not present on *Vibrio cholerae* O1. *Infect. Immun.* **63**: 317-323.

Coppola, G., Huang, F., Riley, J., Cox, J.L., Hantzopoulos, P., Zhou, L-B., and Calhoun, D.H. (1991) Sequence and transpciptional activity of the *Escherichia coli* K-12 chromosome region between *rrnC* and *ilvGMEDA*. *Gene* **97:** 21-27.

Cornillot, E., Saint-Joanis, B., Daube, G., Katayama, S., Granum, P.E., Canard, B., and Cole, S.T. (1995) The enterotoxin gene (cpe) of *Clostridium perfringens* can be chromosomal or plasmid-borne. *Mol. Microbiol.* **15:** 639-647.

Crosa, J.H. (1984) The relationship of plasmid mediated iron transport and bacterial virulence. *Annu. Rev. Microbiol.* **38:** 69-89.

Crosa, J.H. (1989) Genetics and molecular biology of siderophore-mediated iron transport in bacteria. *Microbiol. Rev.* **53:** 517-30.

Crosa, J.H. (1980) A plasmid associated with virulence in the marine fish pathogen *Vibrio anguillarum* specifies an iron-sequestering system. *Nature* **284:** 566-8.

Crosa, J.H., Actis, L.A., Mitoma, Y., Perez Casal, J., Tolmasky, M.E., and Valvano, M.A. (1985) Plasmid-mediated iron sequestering systems in pathogenic strains of *Vibrio anguillarum* and *Escherichia coli*. *Basic Life Sci.* **30:** 759-74.

Crosa, J.H., Schiewe, M.H., and Falkow, S. (1977) Evidence for the plsamid contribution to the virulence of fish pathogen *Vibrio anguillarum*. *Infect. Immun.* **18:** 509-513.

Crosa, J.H., Hodges, L., and Schiewe, M. (1980) Curing of a plasmid is correlated with an attenuation of virulence in the marine fish pathogen *Vibrio anguillarum*. *Infect. Immun.* **27:** 897-902.

Dalrymple, B., Caspers, P., and Arber, W. (1984) Nucletoide sequence of the prokaryotic mobile genetic element IS*30*. *EMBO J.* **3:** 2145-2149.

Daniels, C., Vindurampulle, C., and Morona, R. (1998) Overexpression and topology of the *Shigella flexneri* O-antigen polymerase (Rfc/Wzy) *Mol. Microbiol.* **28:** 1211-1222.

David, V.A., Deutch, A.H., Sloma, A., Pawlyk, D., Ally, A., and Durham, D.R. (1992) Cloning, sequencing and expression of the gene encoding the extracellular neutral protease, vibriolysin, of *Vibrio proteolyticus*. *Gene* **112:** 107-112.

Davis, M.A., Simons, R.W., and Kleckner, N.(1985) Tn*10* protects itself at two levels from fortuitous activation by external promoters. *Cell* **43**:379-387.

Darveau, R.P., and Hancock, R.E.W. (1983) Procedure for isolation of bacterial lipopolysaccharides from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhimurium* strains. *J. Bacteriol.* **155:** 831-838.

Davey, M.L., Hancock, R.E., and Mutharia, L.M. (1998) Influence of culture conditions on expression of the 40-kilodalton porin protein of *Vibrio anguillarum* serotype O2. *Appl-Environ-Microbiol.* **64:** 138-46.

Devine, J.H., Shadel, G.S., and Baldwin, T.O. (1989) Identification of the operator of the *lux* regulon from the *Vibrio fisheri* strain ATCC7744. *Proc. Natl. Acad. Sci. USA*. **86:** 5688-5692.

Dhillon, N., Hale, R.S., Cortes, J., and Leadlay, P.F. (1989) Molecular characterisation of a gene from *Saccharopolyspora erythraea* which is involved in erythromycin biosynthesis. *Mol. Microbiol.* **3:** 1405-1414.

DiSalvo, L.H., Blecka, J., and Zebal, R. (1978) *Vibrio anguillarum* and larval mortality in California coastal shellfish hatchery. *Appl. Environ. Microbiol.* **35:** 219-221.

Distler, J., Mansouri, K., Mayer, G., Stockmann, M., and Piepersberg, W. (1992) Streptomycin biosynthesis and its regulation in *Streptomycetes*. *Gene* **115:** 105-111.

Dorsch, M., Lane, D., and Stackebrandt, E. (1992) Towards a phylogeny of the genus *Vibrio* based on 16S rRNA sequences. *Int. J. Syst. Bacteriol.* **42:** 58-63.

Dumontier, S., Trieu-Cuot, P., and Berche, P. (1998) Structural and functional characterisation of IS*1358* from *Vibrio cholerae*. *J. Bacteriol.* **180:** 6101-6106.

Eguchi, H., Kaya, S., and Araki, Y. (1992) Occurrence of 2,4-dihydroxy-3,3,4-trimethylpyroglutamic acid as an N-acyl substituent in the O-polysaccharide chain of the lipopolysaccharide of *Vibrio anguillarum* V-123. *Carbohydr. Res.* **231:** 147-58.

Eguchi, H., Kaya, S., Araki, Y., Kojima, N., and Yokota, S. (1992) Structure of the O-polysaccharide chain of the lipopolysaccharide of *Vibrio anguillarum* V-123. *Carbohydr. Res.* **231:** 159-69.

Evelyn, T.P.T. (1984) Immunisation against pathogenic vibriosis. In: Symposium on Fish Vaccination. DeKinkelin, P. (ed) Office International des epizooties, Paris, pp. 121-150.

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

Fetherston, J.D., and Perry, R.D. (1994) The pigmentation locus of *Yersinia pestis* KI16+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. *Mol. Microbiol.* **13:** 6970708.

Finkelstein, R.A., Boesman-Finkelstein, M., Chang, Y., and Hase, C.C. (1992) *Vibrio cholerae* hemagglutinin/protease, colonial variation, virulence and deteachment. *Infect. Immun.* **60:** 472-478.

Fleischmann, R.D. *et al.* (1995) Whole-genome sequencing and assembly of *Haemophilus influenzae*. *Science* **269:** 496-512.

Forbes, K.J, Bruce, K.D., Ball, A., and Pennington. T.H. (1992) Variation and length and sequence of porin (ompP2) alleles of non-capsualte *Haemophilus influenzae*. *Mol. Microbiol.* **6:** 2107-2112.

Formal, S.B., Dammin, G.J., LaBrec, E.H., and Schneider, H. (1958) Experimental *Shigella* infections: characteristics of a fatal infection produced in guinea pigs. *J. Bacteriol.* **75:** 604-610.

Frosch, M., Edwards, U., Bousset, K., Krauβe, B., and Weisgerber, C. (1991) Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. *Mol. Microbiol.* **5:** 1251-1263.

Gagnon, Y., Breton, R., Putzer, H., Pelchat, M., Grunberg-Manago, M., and Lapointe, J. (1994) Clustering and co-transciption of the *Bacillus subtilis* genes encoding the aminoacyl-tRNA synthetases specific for glutamate and for cysteine and the first enzyme for cysteine biosynthesis. *J. Biol. Chem.* **269:** 7473-7482.

Garcia, M.I., Labigne, A., and Le Bouguenec. (1994) Nucleotide sequence of the afimbrial-adhesin encoding *afa*-3 gene cluster and its translocation via flanking IS*1* insertion sequences. *J. Bacteriol.* **176:** 7601-7613.

Garger, S.J., Griffith, O.M., and Grill, L.K. (1983) Rapid purification of plasmid DNA by a single centrifugation in a two step caesium chloride- ethidium bromide gradient. *Biochem. Biophys. Res. Commun.* **117:** 835-842.

Griffiths, E. (1987a) Iron in biological systems. In: Iron and Infection. Bullen, J.J., and Griffiths, E. (eds), John Wiley and Sons Ltd, Chichester, United Kingdom. pp 1-25.

Griffiths, E. (1987a) Iron in biological systems. In: Iron and Infection. Bullen, J.J., and Griffiths, E. (eds), John Wiley and Sons Ltd, Chichester, United Kingdom. pp 69-137.

Grisez, L., and Ollivier, F. (1995) Comparative serology of the marine fish pathogen *Vibrio anguillarum. Appl. Environ. Microbiol.* **61:** 4367-4373.

Guerry, P., Alm, R.A., Power, M.E., Logan, S.M., and Trust, T.J. (1991) Role of two flagellin genes in *Campylobacter* motility. *J. Bacteriol.* **173:** 4757-4764.

Guerry, P., Alm, R.A., Power, M.E., and Trust, T.J. (1992) Molecular and structural analysis of *Camplylobacter* flagellin. In: *Campylobacter jejuni*: current status and future trends, American Society for Microbiology, Washington DC. Nachamkin, I., Blaser, M.J., and Tompkins, L.S. (eds) pp. 267-281.

Guo, L., Lim, K.B., Gunn, J.S., Bainbridge, B., Darveau, R.P., Hackett, M., and Miller, S.I. (1997) Regulation of lipid A modifications by *Salmonella typhimurium* virulence genes *phoP/phoQ. Science.* **276:** 250-253

Gustafson, C.E., Chun, S., and T.J. Trust.(1994) Mutagenesis of the paracrystalline surface protein array of *Aeromonas salmonicida* by endogenous insertion sequence. *J Mol Biol.* **237**: 452-463.

Hammerschmidt, S., Muller, A., Sillmann, H., Muhlenhoff, M., Borrow, R., Fox, A., van Putten, J., Zollinger, W.D., Gerady-Schahn, R., and Frosch, M. (1996) Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (*siaD*): correlation with bacterial invasion and the outbreak of meningococcal disease. *Mol. Microbiol.* **20:** 1211-1220.

Harrell, L.W. (1978) Vibriosis and current vaccination procedures in Puget Sound, Washington. *Mar. Fish. Rev.* **40:** 24-25.

Häse, C.C., and Finkelstein, R.A. (1991) Cloning and nucleotide sequence of the *Vibrio cholerae* hemagglutinin/protease (HA/protease) gene and construction of an HA/protease-negative strain. *J. Bacteriol.* **173:** 3311-3317.

Hashimoto, Y., Li, N., Yokoyama, H., and Ezaki, T. (1993) Complete nucleotide sequence and molecular characterisation of ViaB region encoding Vi antigen in *Salmonella typhi. J. Bacteriol.* **175**: 4456-4465.

Hastein, T., and Holt, G. (1972) The occurrence of vibrio disease in wild Norwegian fish. *J. Fish Biol.* **4:** 33.

Haughn, G.W., Squires, C.H., DeFelice, M., Largo, C.T., and Calvo, J.M. (1985) Unusual organisation of the *ilvIH* promoter of *Escherichia coli. J. Bacteriol.* **160:** 186-198.

Hawkes, R., Niday, E., and Gordon, J. (1982) A dot-immuno binding assay for monoclonal and other antibodies. *Anal. Biochem.* **119:** 142-147.

Heinrichs, D.E., Yethon, J.A., and Whitfield, C. (1998) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica. Mol. Microbiol.* **30:** 221-232.

Helander, I.M., Hurme, R., Haikara, A., and Moran, A.P. (1992) Seperation and characterisation of two chemically distinct lipopolysaccharides in two *Pectinatus* species. *J. Bacteriol.* **174:** 3348-3354.

Higgins, D.G., and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene.* **73:** 237-244

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

Hill, C.W., Sandt, C.H., and Vlazny, D.A. (1994) *Rhs* elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. *Mol. Microbiol.* **12**: 865-871.

Hirono, I., and Aoki, T. (1991) Nucleotide sequence and expression of an extracellular hemolysin gene of *Aeromonas hydrophilia*. *Microb. Pathog.* **11:** 189-197.

Hirono, I., and Aoki, T. (1993) Cloning and characterisation of three hemolysin genes from *Aeromonas salmonicida*. *Microb. Pathog.* **15:** 269-282.

Hirono, I., Masuda, T., and Aoki, T. (1996) Cloning and detection of the hemolysin gene of *Vibrio anguillarum*. *Microb. Pathog.* **21:** 173-82.

Hisatsune, K., Hayashi, M., Haishima, Y., and Kondo, S. (1989) Relationship between structure and antigenicity of O1 *Vibrio cholerae* lipopolysaccharide. *J. Gen. Microbiol.* **135:**1901-1907.

Hisatsune, K., Kondo, S., Iguchi, T., Isshiki, Y., and Hiramatsu, K. (1996) Lipopolysaccharide of *Escherichia coli* K-12 strains that express cloned genesd for the Ogawa and Inaba antigens of *Vibrio cholerae* O1: identification of O-antigenic factors. In: The thirty second joint conference U.S.-Japan cooperative medical science program cholera and related diarrheal diseases panel. U.S.-Japan Cooperative Medical Science program.

Hisatsune, K., Kondo, S., Iguchi, T., Ito, T., and Hiramatsu, K. (1993) Occurrence of 2-O-methyl-N-(3-deoxy-L-*glycero*-tetronyl)-D-perosamine (4-amino-4,6-dideoxy-D-manno-pyranose) in lipopolysaccharide from the Ogawa but not from the Inaba O forms of O1 *Vibrio cholerae*. *Biochem. Biophys. Res. Commun.* **190:** 302-307.

Hitchcock, P.J., and Brown, T.M. (1983) Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *J. Bacteriol.* **54:** 269-277

Hobbs. M., and Reeves, P.R. (1994) The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Mol. Microbiol.* **12:** 855-856.

Hofmann,K., and Stoffel, W. (1992) PROFILEGRAPH: an interactive graphical tool for protein sequence analysis. *Comput. Appl. Biosci.* **8:** 331-337

Hofmann,K., and Stoffel, W. (1993) TMBASE- A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler.* **374:** 166.

Hong, M., and Payne, S.M. (1997) Effect of mutations in *Shigella flexneri* chromosomal and plasmid-encoded lipopolysaccharide genes on invasion and serum resistance. *Mol. Microbiol.* **24:** 779-791.

Horne, M.T., and Baxendale, A. (1983) The adhesion of *Vibrio anguillarum* to host tissues and its role in pathogenesis. *J. Fish Dis.* **6:** 461-471.

Horne, M.T., Tatner, M., McDerment, S., and Aguis, C. (1982) Vaccination of rainbow trout at low temperatures and the long term persistance of protection. *J. Fish Dis.* **5:** 343-345.

Howe, T.R., and Iglewski, B.H. (1984) Isolation and characterisation of alkaline protease-deficient mutants of *Pseudomonas aeruginosa in vitro* and in mouse eye model. *Infect. Immun.* **43:**1058-1063.

Hu, S.T., and Lee, C.H. (1988) Characterisation of the transposon carrying the STII gene of enterotoxigenic *Escherichia coli. Mol. Gen. Genet.* **214:** 490-495.

Iguchi, T., Kondo, S., and Hisatsune, K. (1989) [Chemical properties of lipopolysaccharide (LPS) isolated from *Vibrio anguillarum* PT514]. *Nippon-Saikingaku-Zasshi.* **44:** 805-12.

Ingole, K.V., Jalgaonkar, S.V., and Fule, R.P. (1998) Study of proteases and other enzymes of *Vibrio cholerae* O1 El Tor and O139 serotypes isolated in Yavatmal (Maharashtra). *Indian. J. Pathol. Microbiol.* **41:** 419-422.

Ishiguro, E.E., Kay, W.W., Ainswoth, T., Chamberlain, J.B., Buckley, J.T., and Trust, T.J. (1981) Loss of virulence during culture of *Aeromonas salmonicida* at high temperature. *J. Bacteriol.* **148:** 333-340.

Itami, T., and Kusuda, R. (1980) Studies on spray vaccination against vibriosis in cultures ayu. II. Duration of vaccination efficacy and effects of different vaccine prparations. *Bull. Jpn. Soc. Sci. Fish.* **46:** 699-703.

Jalajakumari, M.B., Thomas, C.J., Halter, R., and Manning, P.A. (1989) Genes for biosynthesis and assembly of CS3 pili of CFA/II enterotoxigenic *Escherichia coli*: novel regulation of pilus production by bypassing an amber codon. *Mol. Microbiol.* **12:** 1685-1695.

Jalal, M., Hossain, M.B., van der Helm, D., Sanders-Loehr, J., Actis, L.A., and Crosa, J.H. (1989) Structure of anguibactin, a unique plasmid-related bacterial siderophore from the fish pathogen *Vibrio anguillarum*. *J. Am. Chem. Soc.* **111:** 292-296.

Jann, K., Goblemann, G., Weisberger, C., Wolf-Ullisch, C., and Kanegasaki, S. (1982) Biosynthesis of the O9 antigen of *Escherichia coli*. *Eur. J. Biochem.* **127:** 157-164.

Jann, K., and Westphal, O. (1975) Microbial polysaccharides. In: The Antigens, Vol 3. Academic press, New York-San Fransisco-London. pp 1-125.

Jann, B., Reske, K., and Jann, K. (1975) Heterogeneity of lipopolysaccharides. Analysis of polysaccharide chain lengths by sodium dodecylsulfate-polyacrylamide gel electrophorsis. *Eur. J. Biochem.* **60:** 239-246.

Jiang, X.M., Neal, B., Santiago, F., Lee, S.J., Romano, L.K., and Reeves, P.R. (1991) Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2) *Mol. Microbiol.* **5:** 695-713.

Johnson, J.A., Salles, C.A., Panigrahi, P., Albert, M.J., Wright, A.C., Johnson, R.J., and Morris Jr, J.G. (1994) *Vibrio cholerae* O139 synonym Bengal is closely related to *Vibrio cholerae* El Tor but has important differences. *Infect. Immun.* **62**: 2108-2110.

Jones, N. C., and Osborn, M.J. (1977) Translocation of phospholipids between the outer and inner membranes in *Salmonella typhimurium. J. Biol. Chem. 252:*7405-7412.

Kahler, C.M., and Stephens, D.S. (1998) Genetic basis for biosynthesis, structure, and function of meningococcal lipooligosaccharide (endotoxin). *Crit. Rev. Micriobiol.* **24:** 281-334.

Kahn, M., Kotler, T., Thomas, C., Figurski, D., Meyer, R., Renault, E., and Helinski, D.R. (1979) Plasmid cloning vechicles derived from plasmids ColE1, RGK and RK2. *Methods Enzymol.* **68:** 268-280.

Khan, S.A., Everest, P., Servos, S., Foxwell, N., Zahringer, U., Brade, H., Rietschel, E.T., Dougan, G., Charles, I.G., and Maskell, D. J. (1998) A lethal role for lipid A in *Salmonella* infections. *Mol. Microbiol.* **29:** 571-579.

Karaolis, D.K., Lan, R., and Reeves, P.R. (1994) Molecular evolution of the seventh pandemic clone of *Vibrio cholerae* and its relationship to other pandemic and epidemic *V. cholerae* isolates. *J. Bacteriol.* **176**: 6206.

Keen, M.G., and Hoffman, P.S. (1989) Characterisation of a *Legionella pnuemophila* extracellular protease exhibiting hemolytic and cytotoxic activities. *Infect. Immun.* **57:** 732-738.

Keenleyside, W.J., and Whitfield, C. (1996) A novel pathway for O-polysaccharide biosynthesis in *Salmonella enterica* serovar Borreze. *J. Biol. Chem.* **271:** 28581-28592.

Kenne, L., Lindberg, B., Unger, P., Holme, T., and Holmgren, J. (1979) Structural studies of *Vibrio cholerae* O-antigen. *Carbohydr. Res.* **68:** C16-C17.

Kenne, L., Lindberg, B., Unger, P., Holme, T., and Holmgren, J. (1982) Structural studies of *Vibrio cholerae* O-antigen. *Carbohydr. Res.* **100:** 341-349.

Kessler, A.C., Haase, A., and Reeves, P.R. (1993) Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serogroup IIA. *J. Bacteriol.* **175:** 1412-1422.

Kitao, T., Aoki, T., Fukudome, M., Kawano, K., Wada, J., and Mizuno, Y. (1983) Serotyping of *Vibrio anguillarum* isolated from diseased freshwater fish in Japan. *J. Fish Dis.* **6:** 175-181.

Klauser, T., Pohlner, J., and Meyer, T.F. (1990) Extracellular transport of cholera toxin B subunit using *Neisseria gonorrhoeae* IgA protease β-domain: conformation-dependent outer membrane transport. *EMBO J.* **9:** 1991-1999.

Klena, J.D., and Schnaitman, C.A. (1993) Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. *Mol. Microbiol.* **9:** 393-402.

Klena, J.D., Pradel, E., and Schnaitman, C.A. (1992) Comparison of lipopolysaccharide biosynthesis genes *rfaK, rfaL, rfaY and rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium. J. Bacteriol.* **171:** 3629-3633.

Knirel, Y.A., Paredes, L., Jansson, P.E., Weintraub, A., and Albert, M.J. (1995) Structure of the capsular polysaccharide of *Vibrio cholerae* O139 synonym Bengal containing D-galactose 4,6-cyclophosphate. *Eur. J. Biochem.* **232:** 391-396.

Knirel, Y.A., Widmalm, G., Senchenkova, S.N., Jansson, P-E., and Weintraub, A. (1997) Structural studies on the short-chain lipopolysacchardie of *Vibrio cholerae* O139 Bengal. *Eur. J. Biochem.* **247:** 402-410.

Knirel, Y.A., and Kochetkov, N.K. (1994) *Biochem* (Moscow) **59:** 1325-1383.

Knirel, Y.A., Vinogradov, E.V., Shashkov, A.S., Dmitriev, B.A., Kochetkov, N.K., Stanislavsky, E.S., and Mashilova, G.M. (1985) Somatic antigens of *Pseudomonas aeruginosa.* The structure of the O-specific polysaccharide chains of lipopolysaccharides of *P. aeruginosa* serogroup O4 (Lanyi) and related serotype O6 (Habs) and immunotype 1 (Fisher). *Eur. J. Biochem.* **150:** 541-550.

Knott, V., Rees, D.J., Cheng, Z., and Brownlee, G.G. (1988) Randomly picked cosmid clones overlap *pyrB* and *oriC* gap in the physical map of the *E. coli* chromosome. *Nucleic acid Res.* **16:** 2601-2612.

Kolkman, M.A., Morrison, D.A., Van Der Zeijst, B.A., and Nuijten, P.J. (1996) The capsule polysaccharide synthesis locus of *Streptococcus pnuemoniae* serotype 14: Identification of the glycosyl transferase gene *cps14E. J. Bacteriol.* **178:** 3736-3741.

Kondo, S., Iguchi, T., and Hisatsune, K. (1988) A comparitive study of the sugar composition of lipopolysaccharides isolated from *Vibrio cholerae*, *"Vibrio albensis"* and *Vibrio metschnikovii. J. Gen. Microbiol.* **134:** 1699-1705.

Koster, W.L., Actis, L.A., Waldbeser, L., Tolmasky, M.E., and Crosa, J.H. (1991) Molecular characterisation of the iron transport system mediated by the pJM1 plasmid in *Vibrio anguillarum* 775. *J. Biol. Chem.* **266:** 23829-23833.

Kostrzynska, M., Betts, J.D., Austin, J.W., and Trust, T.J. (1991) Identification, characterisation and spatial localisation of two flagellin species in *Helicobacter pylori* flagella. *J. Bacetriol.* **173:** 937-946.

Kothary, M.H., and Kreger, A.S. (1985) Production and partial characterisation of an elastolytic protease of *Vibrio vulnificus. Infect. Immun.* **50:** 534-540.

Kothary, M.H., and Kreger, A.S. (1987) Purification and characterisation of an elastolytic protease of *Vibrio vulnificus. J. Gen. Microbiol.* **133:** 1783-1791.

Kroll, J.S., Loynds, B., Brophy, L.N., and Moxon, E.R. (1990) The bex locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. *Mol. Microbiol.* **4:** 1853-1862.

Krovacek, K., Faris, A., Eriksson, L., Jansson, E., Ljungberg, O., and Mansson, I. (1987) Virulence, cytotoxic and inflammatory activities of *Vibrio anguillarum* and Aeromonas salmonicida isolated from cultivated salmonid fish in Sweden. *Acta Vet Scand.* **28:** 47-54.

Kyte, J., and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105-132.

Lai, C-Y., and Baumann, P. (1992) Sequence analysis of a DNA fragment from *Buchnera aphidicola* (an endosymbiont of aphids) containing genes homologous to *dnaG*, *rpoD*, *cysE*, and *secB*. *Gene* **119:** 113-118.

Lazarevic, V., and Karamata, D. (1995) The *tagGH* operon of *Bacillus subtilis* 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids. *Mol. Microbiol.* **16:** 345-355.

Leppard, K., Totty, N., Waterfield, M., Harlow, E., Jenkins, J., and Crawford, L. (1983) Purification and partial amino acid sequence analysis of the cellular tumour antigen p53, from mouse SV-40-transformed cells. *EMBO J.* **2:** 1993-1999.

Levine, M.M., Ristaina, P., Morley, G., Smyth, C., Knutton, S., Boedecjer, E., Black, R., Young, C., Clements, M.L., Cheney, C. and Patnaik, R. (1984) Coli surface antigens 1 and 3 of coloninsation factor antigen II-positive enterotoxigenic *Escherichia coli*: Morphology, purification and immune responses in humans. *Infect. Immun.* **44:** 409-420.

Levinson, G., and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4:** 203-221.

Lin, R.J., Capage, M., and Hill, C.W. (1984) A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K-12 chromosome. *J. Mol. Biol.* **177**:1-18

Lindberg, A.A. (1977) In: *Surface Carbohydrates of the Prokaryotic cell*, Sutherland, I.W. (ed), Academic Press, N.Y., and London. pp. 289-356.

Lindberg, A.A., Holme, T., Hellerqvist, C.G., and Svensson, S. (1972) Studies of a lipopolysaccharide hapten isolated from cell-walls of the rough mutant *Salmonella typhimurium* 395 MR10. *Acta Pathol. Microbiol. Scand. Sect.* **B80:** 743-750.

Lindberg, A.A., Karnell, A., and Weintraub, A. (1991) The lipopolysaccharide of *Shigella* bacteria as a virulence factor. *Rev. Infect. Diseases.* **13:** S279-284.

Liu, D., Verma, N.K., Romana, L.K., and Reeves, P.R. (1991) Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. *J. Bacteriol.* **173:** 4814-4819.

Liu, D., and Reeves, P.R. (1994) *Escherichia coli* K-12 regains its O antigen. *Microbiol.* **140:** 49-57.

Lüderitz, O., Westphal, O., Staub, A.A., and Nikaido, H. (1971) Isolation and chemical and immunological characterisation of bacterial lipopolysaccharides, p145-233. In *Microbial toxins, vol. 4. Bacterial endotoxins*, G. Weinbaum, S. Kadis, and S. Ajl (ed). Academic Press, Inc., New York.

Lüderitz, O., Westphal, O., and Staub, A.A. (1966) Immunochemistry of O and R antigens of *Salmonella* and related Enterobacteriaceae. *Bacteriol. Rev.* **30:** 192-255.

Lugtenberg, B., Meijers, J., Peters, R., van der Hoek, P., and van Alphen, L. (1975) Electrophoretic resolution of the major outer membrane protein of *Escherichia coli* K-12 into four bands. *FEBS Lett.* **58:** 254-258.

MacLachlan, P.R., Kadam, S.K., and Sanderson, K.E. (1991) Cloning, characterisation and DNA sequence of the *rfaLK* region for lipopolysaccharide synthesis in *Salmonella typhimurium. J. Bacteriol.* **173:** 7151-7163.

Macnab, R.M. (1992) Genetics and biogenesis of bacterial flagella. *Annu. Rev. Genet.* **26:** 131-158.

Macpherson, D.F., Manning, P.A., and Morona, R. (1994) Characterisation of the dTDP-rhamnose biosynthetic genes encoded in the *rfb* locus *of Shigella flexneri. Mol. Microbiol.* **11:** 281-292.

Mahillon, J., and Chandler, M. (1998) Insertion Sequences. *MMBR.* **62:** 725-774.

Måkelå, P.H., and Stocker, B.A.D. (1984) Genetics of lipopolysaccharide. In *Handbook of Endotoxin, Vol. 1. Chemistry of endotoxin*. E.T. Rietschel (ed.), Amsterdam: Elsevier. pp. 59-137.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratories. Cold Spring Harbour, N.Y.

Manning, P.A., Heuzenroeder, M.W., Yeadon, J., Leavesley, D.I., Reeves, P.R., and Rowley, D. (1986) Molecular cloning and expression in *Escherichia coli* K-12 of the O-antigens of the Ogawa and Inaba serotypes of the lipopolysaccharide of *Vibrio cholerae* O1 and their potential for vaccine development. *Infect. Immun.* **53:** 272-277.

Manning, P.A., Stroeher, U.H., and Morona, R. (1994) Molecular basis for O-antigen biosynthesis in *Vibrio cholerae* O1: Ogawa:Inaba switching. In: Wachsmuth, I.K., Blake, P.A., and Olsvik, O. (Eds), *Vibrio cholerae* and Cholera: Molecular to Global Perspectives. American Society for Microbiology, Washington DC, pp77-94.

Manning, P.A., Stroher, U.H., Karageorgos, L.E., and Morona, R. (1995) Putative O-antigen transport genes within the *rfb* region of *Vibrio cholerae* O1 are homologous to those of capsule transport. *Gene* **158**:1-7.

Marino, P.A. McGrath, B.C., and Osborn, M.J. (1991) Energy dependence of O-antigen synthesis in *Salmonella typhimurium. J. Bacteriol.* **173:** 3128-3133.

Marolda, C.L., and Valvano, M.A. (1998) Promoter region of the *Escherichia coli* O7- specific lipopolysaccharide gene cluster: structural and functional characterisation of an upstream untranslated mRNA sequence. *J. Bacteriol.* **180:** 3070-3079.

Mavris, M., Manning, P.A., and Morona, R. (1997) Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri. Mol. Microbiol.* **26:** 939-950.

McGee, K., Horstedt, P., and Milton, D.L. (1996) Identification and characterization of additional flagellin genes from *Vibrio anguillarum*. *J. Bacteriol.* **178:** 5188-98.

McGrath, B.C., and Osborn, M.J. (1991) Localisation of the terminal steps of O-antigen synthesis in *Salmonella typhimurium*. *J. Bacteriol.* **173:** 649-654.

Meier-Dieter, U., Starman, R., Barr, K., Mayer, H., and Rick, P.D. (1990) Biosynthesis of enterobacterial common antigen in *Escherichia coli*. Biochemical characterisation of Tn*10* insertion mutants defective in ECA synthesis. *J. Biol. Chem.* **265:** 13490-13497.

Meier-Dieter, U., Barr, K., Starman, R., Hatch, L., and Rick, P.D. (1992) Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. *J. Biol. Chem.* **267:** 746-753.

Miller, J. (1972) Experiments in Molecular Genetics. Cold Spring Harbour Laboratory, Cold Spring Harbor, N.Y.

Milton, D.L., Hardman, A., Camara, M., Chhabra, S.R., Bycroft, B.W., Stewart, G.S., and Williams, P. (1997) Quorum sensing in *Vibrio anguillarum*: characterization of the vanI/vanR locus and identification of the autoinducer N-(3-oxodecanoyl)-L-homoserine lactone. *J. Bacteriol.* **179:** 3004-12.

Milton, D.L., Norqvist, A., and Wolf Watz, H. (1992) Cloning of a metalloprotease gene involved in the virulence mechanism of *Vibrio anguillarum*. *J. Bacteriol.* **174:** 7235-44.

Milton, D.L., Norqvist, A., and Wolf Watz, H. (1995) Sequence of a novel virulence-mediating gene, *virC*, from *Vibrio anguillarum*. *Gene* **164:** 95-100.

Milton, D.L., O. Toole, R., Horstedt, P., and Wolf Watz, H. (1996) Flagellin A is essential for the virulence of *Vibrio anguillarum*. *J. Bacteriol.* **178:** 1310-9.

Mizuno, T., and Mizushima, S. (1990) Signal transduction and gene regulation through the phosphorylation of two regulatory components: the molecular basis for the osmotic regulation of the porin genes. *Mol. Microbiol.* **4:** 1077-1082.

Morona, R., Mavris, M., Fallarino, A., and Manning, P.A. (1994) Characterization of the *rfc* region of *Shigella flexneri*. *J. Bacteriol.* **176:** 733-747.

Morona, R., Macpherson, D.F., Van Den Bosch, L., Carlin, N.I., and Manning, P.A. (1995a) Lipopolysaccharide with an altered O-antigen produced in *Escherichia coli* K-12 harbouring mutated, cloned *Shigella flexneri rfb* genes. *Mol. Microbiol.* **18:** 209-223.

Morona, R., Van Den Bosch, L., and Manning, P.A. (1995) Molecular, genetic, and topological characterisation of O-antigen chain length regulation in *Shigella flexneri*. *J. Bacteriol.* **177:** 1059-1068.

Morona, R., Brown, M.H., Yeadon, J., Heuzenroeder, M.W., and Manning, P.A. (1991) Effects of lipopolysaccharide core biosynthesis mutations on the production of *Vibrio cholerae* O-antigen in *Escherichia coli* K-12. *FEMS Microbiol. Lett.* **66**: 279-286.

Morona, R., Stroeher, U.H., Karageorgos, L.E., Brown, M.H., and Manning, P.A. (1995b) A putative pathway for biosynthesis of the O-antigen component, 3-deoxy-L-*glycero*-tetronic acid, based on the sequence of the *Vibrio cholerae* O1 *rfb* region. *Gene* **166**: 19-31.

Morrison, D.C., and Ryan, J.L. (1987) Endotoxins and disease mechanisms. *Annu. Rev. Med.* **38:** 417-432.

Munn, C.B. (1978) Haemolysin production by *Vibrio anguillarum*. *FEMS Micro. Lett.* **3:** 265-268.

Munn, C.B. (1980) Production and properties of a haemolysin toxin by *Vibrio anguillarum*. In: Ahne, W. Ed Fish Diseases. Berlin, Springer Verlag, pp. 69-74.

Munson Jr, R., and Tolan Jr, R.W. (1989a) Molecular cloning, expression, and primary sequence of outer membrane protein P2 of *Haemophilus influenzae* type b. *Infect. Immun.* **57:** 88-94.

Munson Jr, R., Bailey, C., and Grass, S. (1989b) Diversity of the outer membrane protein P2 from major clones of *Haemophilus influenzae* typbe b. *Mol. Microbiol.* **3:** 1797-1803.

Munson, R.S., Shenep, J.L., Barenkamp, S.J., and Granoff, D.M. (1983) Purification and comparison of outer membrane protein P2 from *Haemophilus influenzae* type b isolates. *J. Clin. Invest.* **72:** 677-684.

Mutharia, L.M., and Amor, P.A. (1994) Monoclonal antibodies against *Vibrio anguillarum* O2 and *Vibrio ordalii* identify antigenic differences in lipopolysaccharide O-antigens. *FEMS Microbiol. Lett.* **123:** 289-98.

Mutharia, L.W., Raymond, B.T., Dekievit, T.R., and Stevenson, R.M. (1993) Antibody specificities of polyclonal rabbit and rainbow trout antisera against *Vibrio ordalii* and serotype 0:2 strains of *Vibrio anguillarum*. *Can. J. Microbiol.* **39:** 492-9.

Naide, Y., Nikaido, H., Måkelå, P.H., Wilkonson, R.G., and Stocker, B.A.D. (1965) Semirough strains of *Salmonella*. *Proc. Natl. Acad. Sci. USA*. **53:**147-153.

Nei, M. (1987) Molecular evolutionary genetics. Columbia University Press, New York, N.Y.

Neilands, J.B. (1981) Iron absorption and transport in microorgansims. *Annu. Rev. Nutr.* **1:** 27-46.

Nishibuchi, M., and Kaper, J.B. (1985) Nucleotide sequence of the thermostable direct hemolysin gene in *Vibrio parahaemolyticus*. *J. Bacteriol.* **162:** 558-564.

Nishijima, M., and Raetz, C.R.H. (1979) Membrane lipid biogensis in *Escherichia coli*: identification of genetic loci for phosphatidylglycerophosphate synthetase and construction of mutants lacking phosphatidylglycerol. *J. Biol. Chem.* **254:** 7837-7844.

Nakano, Y., Yoshida, Y., Yamashita, Y., and Koga, T. (1998) A gene cluster for 6-deoxy-L-talan synthesis in *Actinobacillus actinomycetemcomitans*. *Biochem. Biophys. Acta*. **1442:** 409-414.

Norqvist, A., Hagstrom, A., and Wolf Watz, H. (1989) Protection of rainbow trout against vibriosis and furunculosis by the use of attenuated strains of *Vibrio anguillarum*. *Appl. Environ. Microbiol.* **55:** 1400-5.

Norqvist, A., Norrman, B., and Wolf Watz, H. (1990) Identification and characterization of a zinc metalloprotease associated with invasion by the fish pathogen *Vibrio anguillarum*. *Infect. Immun.* **58:** 3731-6.

Norqvist, A., and Wolf Watz, H. (1993) Characterization of a novel chromosomal virulence locus involved in expression of a major surface flagellar sheath antigen of the fish pathogen *Vibrio anguillarum*. *Infect. Immun.* **61:** 2434-44.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**: 621-625

Ogierman, M.A., Fallarino, A., Riess, T., Williams, S.G., Attridge, S.R., and Manning, P.A. (1997) Characterisation of the *Vibrio cholerae* El Tor lipase operon *lipAB* and a protease gene downstream of the *hly* region. *J. Bacteriol.* **179:** 7072-7080.

Ohnishi, K., and Muroga, K. (1977) *Vibrio* sp. As a cause of disease in rainbow trout cultured in Japan II. Physiological characteristics and pathogenicity. *Fish Pathol.* **12:** 51-55.

Okada, N., Sasakawa, C., Tobe, T., Talukder, K.A., Komatsu, K., and Yoshikawa, M. (1991a) Construction of a physical map of the chromosome of *Shigella flexneri* 2a and the direct assignment of nine virulence-associated loci identified by Tn*5* insertions. *Mol. Microbiol.* **5:** 2171-2180.

Okada, N., Sasakawa, C., Tobe, T., Yamada, M., Nagai, S., Talukder, K.A., Komatsu, K., Kanegasaki, S., and Yoshikawa, M. (1991b) Virulence- associated chromosomal loci of *Shigella flexneri* by random Tn*5* insertion mutagenesis. *Mol. Microbiol.* **5:** 187-195.

Olsson, J.C., Westerdahl, A., Conway, P.L., and Kjelleberg, S. (1992) Intestinal colonisation potential of turbot (Scophthalmus maximus)- and Dab (Limanda limanda)- associated bacteria with inhibitory effects against *Vibrio anguillarum*. *Appl. Environ. Micro.* **58:** 551-556.

Orlean, P., Albright, C., and Robbins, P.W. (1988) Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein. *J. Biol. Chem.* **263:** 17499-17507.

Osborn, M.J. (1979) Biosynthesis and assembly of the lipopolysaccharide of the outer membrane. Pp. 15-34 In: Bacterial outer membranes. (ed Inouye, M) John Wiley and Sons, Inc., New York.

Osborn, M.J., Gander, J.E., Parisi, E., and Carson, J. (1972) Mechanism of assembly of the outer membrane of *Salmonella typhi. J. Biol. Chem.* **247:** 3962-3971.

Osborn, M.J., Gander, J.E., and Parisi, E. (1972b) Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. Site of synthesis of lipopolysaccharides. *J. Biol. Chem.* **247:** 3973-3985.

O. Toole, R, Milton, D.L., and Wolf Watz, H. (1996) Chemotactic motility is required for invasion of the host by the fish pathogen *Vibrio anguillarum*. *Mol. Microbiol.* **19:** 625-37.

Paulsen, I.T., Beness, A.M., and Saier Jr, M.H. (1997) Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiol.* **143:** 2685-2699.

Pavelka, M.S., Hayes, S.F., and Silver, R.P. (1994) Characterisation of KpsT, the ATP-binding component of the ABC-transporter involved with the export of capsular polysialic acid in *Escherichia coli* K1. *J. Biol. Chem.* **269:** 20149-20158.

Palva, T., and Måkelå, P.H. (1980) Lipopolysaccharide heterogeneity in *Salmonella typhimurium* analysed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis. *Eur. J. Biochem.* **107:** 137-143.

Pavlovskis, O.R., and Wretlind, B. (1979) Assessment of protease (elastase) as a *Psuedomonas aeruginosa* virulence factor in experimental mouse burn infection. *Infect. Immun.* **24:** 181-187.

Pazos, F., Santos, Y., Majarinos, B., Bandin, I., Nunez, S., and Toranzo, A.E. (1993) Phenotypic characterisitics and virulence of *Vibrio anguillarum*-related organsims. *Appl. Environ. Microbiol.* **59:** 2969-2976.

Pedersen, K., Grisez, L., van Houdt, R., Tiainen, T., Ollevier, F., and Larsen, J.L. (1999) Extended serotyping scheme for *Vibrio anguillarum* with the definition and characterization of seven provisional O-serogroups. *Curr. Microbiol.* **38:** 183-9.

Pegues, J.C., Chen, L., Gordon, A.W., Ding, L., and Coleman, W.G. (1990) Cloning, expression and characterisation of the *Escherichia coli* K-12 *rfaD* gene. *J. Bacteriol.* **172:** 4652-4660.

Phalipon, A., Kaufmann, M., Michetti, P., Cavaillon, J.M., Huerre, M., Sansonetti, P., and Kraehenbuhl, J.P.(1995) Monoclonal immunoglobulin A antibody directed against serotype-specific epitope of *Shigella flexneri* lipopolysaccharide protects against murine experimental shigellosis. *J. Exp. Med.* **182**:769-78

Pierce, N.F., Kaper, J.B., Mekalanos, J.J., and Cray Jr, W.C. (1985) Role of cholera toxin in enteric colonisation by *Vibrio cholerae* O1 in rabbits. *Infect. Immun.* **50:** 813-816.

Polissi, A., and Georgopoulos, C. (1996) Mutational analysis and properties of the *msbA* gene of *Escherichia coli*, coding for an essential ABC family transporter. *Mol. Microbiol.* **20:** 1221-1233.

Preston, L.M., Xu, Q., Johnson, J.A., Joseph, A., Maneval Jr, D.R., Husain, K., Reddy, G.P., Bush, C.A., and Morris Jr, J.G. (1995) Preliminary structure determination of the capsular polysaccharide of *Vibrio cholerae* O139 Bengal AI1837. *J. Bacteriol.* **177:** 835-838.

Pu, W., and Struhl, K. (1991) The leucine zipper symmetrically positions the adjacent basic regions for specific DNA binding. *Proc Natl Acad Sci USA* **88:** 6901-6905.

Radar, A.E., and Murphy, J.R. (1988) Nucleotide sequences and comparison of the haemolysin determinants of *Vibrio cholerae* El Tor RV79 (Hly+) and RV79 (Hly-). *Infect. Immun.* **56:** 1414:1419.

Raetz, C.R.H. (1993) Bacterial endotoxins: extraordinary lipids that activate Eukaryotic signal transduction. *J. Bacteriol.* **175:** 5745-5753.

Raetz, C.R.H. (1996) Bacterial lipopolysaccharides: A remarkable family of bioative macroamphiphiles. In: *Escherichia coli* and *Salmonella. Cellular and Microbiology*. Neidhardt, F.C., Curtiss, R. III., Ingraham, J.L., Lin, E.C.C., and Low, K.B. *et al.* (eds). Wsahington, DC: American Society for Microbiology Press, pp. 1035-1063.

Ramamurthy, T., Garg, S., Sharm, R., Bhattachary, S.K., Nair, G.B., Shimada, G.B., Takeda, T., Karasawa, T., Kurazano, H., Pal, A., and Takeda, Y. (1993) Emergence of novel strain of *Vibrio cholerae* with epidemic potenitial in southern and eastern India. *Lancet* **341:** 703-704.

Rasmussen, H.B. (1987a) Subgrouping of lippolysaccharide O-antigens from *Vibrio anguillarum* serogroup O2 by immunoelectrophorestic analyses. *Curr. Microbiol.* **16:** 39-42.

Rasmussen, H.B. (1987b) Immunocelectrophoretic study of lipopolysaccahride antigens from *Vibrio anguillarum* strains, serogroup O2. *Curr. Microbiol.* **15:** 219-222.

Raziuddin, S. (1977) Studies of the polysaccharide fraction from the cell lipopolysaccharides (O-antigen) of *Vibrio cholerae*. *Ind. J. Biochem. Biophys.* **14:** 262-263.

Redmond, J.W. (1979) The structure of the O-antigenic side chain of the lipopolysaccharide of *Vibrio cholerae* 569B (Inaba). *Biochem. Biophys. Acta.* **584:** 346-352.

Redmond, J.W., Korsch, M.J., and Jackson, G.D.F. (1973) Immunochemical studies of the O-antigens of *Vibrio cholerae*. Partial characterisation of an acid-labile antigenic determinant. *Aust. J. Exp. Biol. Med. Sci.* **51:** 229-235.

Reeves, P. (1993) Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *T.I.G.* **9:** 17-22.

Reeves, P. (1994) Biosynthesis and assembly of lipopolysaccharide pp. 281-317.In *Bacterial Cell Wall.* Ghuysen, J.M. and Hakenbeck, R. (*eds*), Elsevier Science B.V.

Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C.R.H., and Rick, P.D. (1996) Bacterial polysaccharide synthesis and gene nomenclature. *T.I.M.* **4:** 495-503.

Rick, P.D., Hubbard, G.L., and Barr, K. (1994) Role of the *rfe* gene in the synthesis of the O8 antigen in *Escherichia coli* K-12. *J. Bacteriol.* **176:** 2877-2884.

Rietschel, E.T., Wollenweber, H.-W., Siorczyk, U., and Lüderitz, O. (1983) Analysis of the primary structure of lipid A. pp.195-219. In *Bacterial lipopolysaccharides. Structure, synthesis and biological activities*. L. Anderson, F.M. Unger (eds) Washington DC: American Chemical Society.

Roantree, R.J. (1967) *Salmonella* O antigens and virulence. *Annu. Rev. Micro.* **21:** 443-466.

Robbins, P. W., Wright, A., and Dankert, M. (1966) Polysaccharide biosynthesis. *J. Gen. Physiol.* **49:** 331-346.

Roncero, C., and Casabadan, M.J. (1992) Genetic analysis of the genes involved in synthesis of the lipopolysaccharide core in *Escherichia coli* K-12: Three operons in the *rfa* locus. *J. Bacteriol.* **174:** 3250-3260.

Rossman, M.G., Moras, D., and Olsen, K.W. (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* (London) **250:** 194-199.

Sadosky, A.B., Davidson, A., Lin, R.J., and Hill, C.W. (1989) Rhs gene family of *Escherichia coli* K-12. *J. Bacteriol.* 171: 636-42.

Sadovskaya, I., Brisson, J.R., Altman, E., and Mutharia, L.M. (1996) Structural studies of the lipopolysaccharide O-antigen and capsular polysaccharide of *Vibrio anguillarum* serotype O:2. *Carbohydr. Res.* **283:** 111-27.

Sadovskaya, I., Brisson, J.R., Khieu, N.H., Mutharia, L.M., and Altman, E. (1998) Structural characterisation of the lipopolysaccharide *Vibrio ordalii* serotype O:2. *Eur. J. Biochem.* **253:** 319-327.

Saeki, A., Kido, N., Sugiyama, T., Ohta, M., Iwashita, T., Uchiya, K., and Kato, N. (1993) Isolation of *rfb* gene clusters directing the synthesis of O polysaccharides consisting of mannose homopolymers and serological analysis of lipopolysaccharides. *Microbiol. Immunol.* **37:** 601-606.

Saiki, R.K., and Gelfand, D.H. (1989) Introducing AmpliTaq DNA polymerase. *Amplification* **1:** 4-6.

Sakazaki, R., and Tamura, K. (1971) Somatic antigen variation in *Vibrio cholerae*. *Jpn. J. Med. Sci. Biol.* **24:** 93-100.

Sakurai, J., Matsuzaki, A., Takeda, Y., and Mitawani, T. (1974) *Infect. Immun.* **9:** 777-780.

Salah-Bey, K., Doumith, M., Michel, J.M., Haydock, S., Cortes, J., Leadley, P.F., and Raynal, M.C. (1998) Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharpolyspora erythraea*. *Mol. Gen. Genet.* **257:** 542-553.

Salamitou, S., Lemaire, M., Fujino, T., Ohayon, H., Gounon, P., Beguin, P., and Aubert, J-P. (1994) Subcellular localisation of *Clostridium thermocellum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. *J. Bacteriol.* **176:** 2828-2834.

Salati, F., and Kusuda, R. (1986) Chemical composition of the lipopolysaccharide O-antigens from *Vibrio anguillarum* and *Vibrio ordalii*. *Fish Pathol.* **21:** 193-199.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: A Laboratory Manual, vol.1, 2 and 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanderson, K.E., and Roth, J.R. (1988) Linkage map of *Salmonella typhimurium*. Edition VII. *Micro. Rev.* **52:** 485-532.

Santos, Y., Pazos, F., Bandin, I., and Toranzo, A.E. (1995) Analysis of antigens present in the extracellular products and cell surface of *Vibrio anguillarum* serotypes O1, O2, and O3. *Appl. Environ. Microbiol.* **61:** 2493-8.

Saxena, I.M., Brown Jr., R.M., Fevre, M., Geremia, R.A., and Henrissat, B. (1995) Multidomain architecture of β-glucosyl transferases: implications for mechanism of action. *J. Bacteriol.* **177:** 1419-1424.

Schmidt, G., Jann, B., and Jann, K. (1969) Immunochemistry of R lipopolysaccharides of *Escherichia coli.* Different core regions in the lipopolysaccharide of the O group 8. *Eur. J. Biochem.* **10:** 501-510.

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

Schniatman, C.A., Parker, C.T., Klena, J.D., Pradel, E., Pearson, N.B., Sanderson, K.E., and Maclachlan, P.R. (1991) Physical map of the rfa loci of *Escherichia coli* K-1 and *Salmonella typhimurium. J. Bacteriol.* **173:** 7410-7411.

Schulz, V.P., and Reznikoff, W.S. (1991) Translation initiation of IS*50*R readthrough transcript. *J. Mol. Biol.* **221:** 65-80.

Simon, R., Priefer, U., and Pühler, A. (1983) A broad host range mobilisation system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Biotech.* **1:** 784-791.

Simon, M., Jofre, J., and Blanch, A.R. (1998) Evaluation of the immunospecificity of the porin Om1 of *Vibrio anguillarum* serotype O1. *J. Appl. Microbiol.* **84:** 709-14.

Simon, M., Mathes, A., Blanch, A., and Engelhardt, H. (1996) Characterization of a porin from the outer membrane of *Vibrio anguillarum. J. Bacteriol.* **178:** 4182-8.

Singer, J.T. (1989) Molecular cloning of the *recA* analog from the marine fish pathogen *Vibrio anguillarum* 775. *J. Bacteriol.* **171:** 6367-71.

Singer, J.T., Choe, W., Schmidt, K.A., and Makula, R.A. (1992) Virulence plasmid pJM1 prevents the conjugal entry of plasmid DNA into the marine fish pathogen *Vibrio anguillarum* 775. *J. Gen. Microbiol.* **138:** 2485-90.

Singer, J.T., Schmidt, K.A., and Reno, P.W. (1991) Polypeptides p40, pOM2, and pAngR are required for iron uptake and for virulence of the marine fish pathogen *Vibrio anguillarum* 775. *J. Bacteriol.* **173:** 1347-52.

Sirisena, D.M., Maclachlan, P.R., Liu, S.-L., Hessel, A., and Sanderson, K.E. (1994) Molecular analysis of the *rfaD* gene, for heptose synthesis, and the *rfaF* gene, for heptose transfer, in lipopolysaccharide synthesis in *Salmonella* typhimurium. *J. Bacteriol.* **176:** 2379-2385.

Sirisena, D.M., Brozek, K.A., MacLachlan, P.R., Sanderson, K.E., and Raetz, C.R.H. (1992) The *rfaC* gene of *Salmonella typhimurium*: cloning, sequencing, and enzymatic function in heptose transfer to lipopolysaccharide. *J. Biol. Chem.* **267:** 18874-18884.

Sheeran, B., Drianan, E., and Smith, P.R. (1983) Preliminary studies on the role of extracellular proteolytic enzymes in the pathogenessis of furniculosis, p89-100. In: Fish Diseases. Acuigrup, Madrid.

Skurnik, M., Venho, R., Toivanen, P., and Al-Hendy, A. (1995) A novel locus of *Yersinia enterocolitica* serotype O:3 involved in lipopolysaccharide outer core biosynthesis. *Mol. Microbiol.* **17:** 575-594.

Slauch, J.M., Lee, A.A., Mahan, M.J., and Mekalanos, J.J. (1996) Molecular characterisation of the *oafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: OafA is a member of a family of integral membrane trans-acylases. *J. Bacteriol.* **178:** 5904-5909.

Smith, G.C., and Merkel, J.R. (1982) Collagenolytic activity of *Vibrio vulnificus*: potential contribution to its invasiveness. *Infect. Immun.* **35:** 1155-1156.

Smith, A.N., Boulnois, G.J., and Roberts, I.S. (1990) Molecular analysis of the *Escherichia coli* K5 *kps* locus: identification and characteristic of an inner-membrane capsular polysaccharide transport system. *Mol. Microbiol.* **4:** 1863-1869.

Smith, P.D. (1988) Vaccination against vibriosis. In: Fish vaccination. Ellis, A.E. ed. Academic Press, London. pp. 67-84.

So, M., Heffron, F., and McCarthy, B.J. (1979) The *E. coli* gene encoding heat stable toxin is a bacterial transposon flanked by inverted repeats of IS1. *Nature* **277:** 453-456.

Sorensen, U.B., and Larsen, J.L. (1986) Serotyping of *Vibrio anguillarum*. *Appl. Environ. Microbiol.* **51:** 593-7.

Southern, E. (1975) Gel electrophoresis of restriction fragments. *Methods Enzymol.* **68:** 152-176.

Stocker, B.A.D., and Måkelå, P.H. (1971) Genetic aspects of biosynthesis and structure of Salmonella lipopolysaccharide. In: *Microbial Toxins.* Vol 4 Weinbaum, G. *et al.* (eds) New York: Academic Press, pp 369-438.

Strain, S.M., Fesik, S.W., and Armitage, I.M. (1983a) Structure and metal-binding properties of lipopolysaccharide from heptoseless mutants of *Escherichia coli* studied by ^{13}C and ^{31}P nuclear magnetic resonance. *J. Biol. Chem.* **258:**13466-13477.

Strain, S.M., Fesik, S.W., and Armitage, I.M. (1983b) Characterisation of lipopolysaccharide from a heptoseless mutant of *Escherichia coli* by carbon 13 nuclear magnetic resonance. *J. Biol. Chem.* **258:** 2906-2910.

Stroeher, U.H., Karageorgos, L.E., Morona, R., and Manning, P.A. (1992) Serotype conversion in *Vibrio cholerae* O1. *Proc. Natl. Acad. Sci. USA.* **89:** 2566-2570.

Stroeher, U.H., Lech, A.J., and Manning, P.A. (1994) Gene sequence of *recA*$^+$ and construction of *recA* mutants of *Vibrio cholerae*. *Mol. Gen. Genet.* **244**: 295-302.

Stroeher, U.H., Jedani, K.E., Dredge, B.K., Morana, R., Brown, M.H., Karageorgos, L.E., Albert, J.M., and Manning, P.A. (1995a) Genetic rearrangement of the *rfb* regions of *Vibrio cholerae* O1 and O139. *Proc. Natl. Acad. Sci. USA.* **92**:10374-10378.

Stroeher, U.H., Karageorgos, L.E., Brown, M.H., Morona, R., and Manning, P.A. (1995b) In *Vibrio cholerae* O1, *rfaD* is closely linked to the *rfb* operon. *Gene* **166:** 67-72.

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

Stroeher, U. H., Jedani, K. E., and Manning, P. A. (1998) Genetic organization of the regions associated with surface polysaccharide synthesis in *Vibrio cholerae* O1, O139 and *Vibrio anguillarum* O1 and O2: a review. *Gene* **223:** 269-82.

Stroeher, U.H., Parasivam, G., Dredge, B.K., and Manning, P.A. (1997a) Novel *Vibrio cholerae* O139 genes involved in lipopolysaccharide biosynthesis. *J. Bacteriol.* **179:** 2740-2747.

Stroeher, U.H., and Manning, P.A. (1997b) *Vibrio cholerae* O139 serotype O139: swapping genes for surface polysaccharide biosynthesis. *T.I.M.* **5:** 178-180.

Sussman, M. (1974) Iron and Infection. In: Iron in biochemistry and medicine. Jacobs, A., and Worwood, M. (eds), Academic Press, Inc., New York. pp. 649-679.

Suzuki, S., Kuroe, K., and Kusuda, R. (1994) Characteristics of porin-like major outer membrane proteins of *Listonella anguillara* serotypes J-O-1, -2 and -3. *Biochem. Mol Biol Int.* **32:** 605-613.

Suzuki, S., Kuroe, K., Yasue, K., and Kusuda, R. (1996) Antigenicity and N-terminal amino acid sequence of a 35 kDa porin-like protein of *Listonella* (*Vibrio*) *anguillarum*: comparison among different serotypes and other bacterial species. *Lett. Appl. Microbiol.* **23:** 303-6.

Tabor, S., and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/ promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA.* **82:** 1074-1076

Takagi, M., Takada, H., and Imanaka, T. (1990) Nucleotide sequence and cloning in *Bacillus subtilis* of the *Bacillus stearothermophilus* pleiotropic regulatory gene *degT. J. Bacteriol.* **172:** 411-418.

Takayama, M., Oureshi, N., and Mascagni, P. (1983) Complete structure of lipid A obtained from the lippolysaccharide of the heptoseless mutant of *Salmonella typhimurium. J. Biol. Chem.* **258:** 12801-12803.

Taylor, R.K., Manoil, C., and Mekalanos, J.J. (1989) Broad-host range vectors for delivery of Tn*phoA*: use in genetic analysis of secreted virulence determinants of *Vibrio cholerae. J. Bacteriol.* **171:** 1870-1878

Terai, A., Shirai, H., Yoshida, O., Takeda, Y., and Nishibuchi, M. (1990) Nucleotide sequence of the thermostable direct haemolysin gene (*tdh* gene) of *Vibrio mimicus* and its evolutionary relationship to the *tdh* genes of *Vibrio parahaemolyticus*. *FEMS Microbiol. Lett.* **71:** 319-324.

Thornson, J.S., Lo, S.F., Ploux, O., He, X., and Lui, H-W. (1994) Studies of the biosynthesis of 3,6-dideoxyhexoses: Molecular cloning and characterisation of the *asc* (ascarylose) region from *Yersinia pseudotuberculosis* serogroup VA. *J. Bacteriol.* **176:** 5483-5493.

Tiainen, T., Pedersen, K., and Larsen, J.L. (1997) Immunological reactivity of *Vibrio anguillarum* sero-subgroups O2a and O2b, and comparison of their lipopolysaccharide profiles. *Curr. Microbiol.* **34:** 38-42.

Tolmasky, M.E., Actis, L.A., and Crosa, J.H. (1988a) Genetic analysis of the iron uptake region of the *Vibrio anguillarum* plasmid pJM1: molecular cloning of genetic determinants encoding a novel trans activator of siderophore biosynthesis. *J. Bacteriol.* **170:** 1913-9.

Tolmasky, M.E., Salinas, P.C., Actis, L.A., and Crosa, J.H. (1988b) Increased production of the siderophore anguibactin mediated by pJM1-like plasmids in *Vibrio anguillarum. Infect. Immun.* **56:** 1608-14.

Tolmasky, M.E., Actis, L.A., and Crosa, J.H. (1993) A single amino acid change in AngR, a protein encoded by pJM1-like virulence plasmids, results in hyperproduction of anguibactin. *Infect. Immun.* **61:** 3228-3233.

Tolmasky, M.E., Wertheimer, A.M., Actis, L.A., and Crosa, J.H. (1994) Characterisation of the *Vibrio anguillarum* fur gene: role in the regulation of expression of the FatA outer membrane protein and catechols. *J. Bacteriol.* **176:** 213-220.

Tolmasky, M.E., Actis, L.A., and Crosa, J.H. (1995) A histidine decarboxylase gene encoded by the *Vibrio anguillarum* plasmid pJM1 is essential for virulence: histamine is a precursor in the biosynthesis of anguibactin. *Mol. Microbiol.* **15:** 87-95.

Tolmasky, M.E., and Crosa, J.H. (1984) Molecular cloning and expression of genetic determinants for the iron uptake system mediated by the *Vibrio anguillarum* plasmid pJM1. *J. Bacteriol.* **160:** 860-6.

Tolmasky, M.E., Gammie, A.E., and Crosa, J.H. (1992) Characterization of the recA gene of *Vibrio anguillarum. Gene* **110:** 41-8.

Toranzo, A.E., Barja, J.L., Colwell, R.R., Hetrick, F.M., and Crosa, J.H. (1983) Haemagglutinating, haemolytic and cytotoxic activities of *Vibrio anguillarum* and related vibrios isolated from striped bass on the Atlantic coast. *FEMS Microbiol. Lett.* **18:** 257-262.

Towbin, H., Staehlin, T., and Gordon, J. (1979) Electrophoretic transfer of proteins from polysaccharide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA.* **76:** 4350-4354.

Trust, T.J., Courtice, I.D., Khouri, A.G., Crosa, J.H., and Schiewe, M.H. (1981) Serum resistance and hemagglutination ability of marine vibrios pathogenic for fish. *Infect. Immun.* **34:** 702-7.

Tsai, C.M., and Frasch, C.E. (1982) A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Anal. Biochem.* **119:** 115-119.

Tubiash, H.S., Colwell, R.R., and Sakazaki, R. (1970) Marine vibrios associated with bacillary necrosis, a disease of larval and juvenile bivalve mollusks. *J. Bacteriol.* **103:** 272-273.

Turner, R., and Tjian, R. (1989) Leucine repeats and an adjacent DNA binding doamin mediate the formation of functional cFos-cJun hetrodimers. *Science* **243:** 1689-1694.

Umelo, E., and Trust, T.J. (1997) Identification and molecular characterisation of two tandemly located flagellin genes from *Aeromonas salmonicida* A449. *J. Bacteriol.* **179:** 5292-5299.

Vachon, V., Laprade, R., and Coulton, J.W. (1986) Properties of the porin of *Haemophilus influenzae* type b in planar lipid bilayer membranes. *Biochem. Biophys. Acta.* **86:** 74-82.

Van den Bosch, L., Manning, P.A., and Morona, R. (1997) Regulation of O antigen chain length is required for *Shigella flexneri* virulence. *Mol. Microbiol.* **23:** 765-775.

van der Ley, P., de Graaff, P., and Tommassen, J. (1986) Shielding of *Escherichia coli* outer membrane proteins as receptors for bacteriophages and colicins by O-antigenic chains of lipopolysaccharide. *J. Bacteriol.* **168:** 449-451.

Verma, N.K., and Reeves, P.R. (1989) Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D *Salmonella*. *J. Bacteriol.* **171:** 5694-5701.

Vinogradov, E.V., Bock, K., Holst, O., and Brade, H. (1995) The structure of the lipid A-core region of the lipopolysaccharides from *Vibrio cholerae* O1 smooth strain 569B (Inaba) and rough mutant strain 95R (Ogawa). *Eur. J. Biochem.* **233:** 152-158.

Vinson, C., Sigler, P., and McKnight, L. (1989) Scissor-grip modle for DNA recognition by a family of leucine sipper proteins. *Science* **249:** 911-916.

Waldbeser, L.S., Chen, Q., and Crosa, J.H. (1995) Antisense RNA regulation of the fatB iron transport protein gene in *Vibrio anguillarum*. *Mol. Microbiol.* **17:** 747-56.

Waldbeser, L.S., Tolmasky, M.E., Actis, L.A., and Crosa, J.H. (1993) Mechanisms for negative regulation by iron of the *fatA* outer membrane protein gene expression in *Vibrio anguillarum* 775. *J. Biol. Chem.* **268:** 10433-9.

Waldor, M.K., Colwell, R., and Mekalanos, J.J. (1994) The *Vibrio cholerae* O139 serogroup antigen includes an O-polysaccharide capsule and lippolysaccharide virulence determinant. *Proc. Natl. Acad. Sci. USA.* **91:** 11388-11392.

Walker, J.E., Sarste, M., Runswick, M.J., and Gay, N.J. (1982) Distantly related sequences in the α- and β-subunits of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1:** 945-951.

Wang, X.H., Oon, H.L., Ho, G.W., Wong, W.S., Lim, T.M., and Leung, K.Y. (1998) Internalization and cytotoxicity are important virulence mechanisms in *Vibrio*-fish epithelial cell interactions. *Microbiology.* **144:** 2987-3002.

Ward, H.M., Morelli, G., Kamke, M., Morona, R., Yeadon, J., Hackett, J.A., and Manning, P.A. (1987) A physical map of the chromosomal region determining O-antigen biosynthesis in *Vibrio cholerae* O1. *Gene* **55:** 197-204.

Waxin, H., Virlogeux, I., Kolyua, S., and Popoff, M.Y. (1993) Identification of six open reading frames in the *Salmonella enterica* subsp. enterica ser. Typhi *viaB* locus involved in Vi antigen production. *Res. Microbiol.* **144:** 363-371.

Weisgerber, C., and Jann, K. (1982) Glucosyldiphoundecaprenol, the mannose acceptor in the synthesis of the O9 antigen of *Escherichia coli.* Biosynthesis and characterisation. *Eur. J. Biochem.* **127:** 165-168.

Wierenga, R.K., Terpstra, P., and Hol, W.G. (1986) Prediction of the occurrence of the ADP-binding beta alpha beta fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187:** 101-107.

Weiner, I.M., Higuchi, T., Rothfield, L., Saltmarsh-Andrew, M., Osborn, M.J., and Horecker, B.L. (1965) Biosynthesis of bacterial lipopolysaccharide. V. Lipid-linked intermediates in the biosynthesis of the O-antgien groups of *Salmonella typhimurium. Proc. Natl. Acad. Sci. USA.* **54:** 228-235.

Whitfield, C. (1995) Biosynthesis of lipopolysaccharide O antigens. *T.I.M.* **3:** 178-185.

Whitfield, C., Amor, P.A., and Köplin, R. (1997) Modulation of the surface architecture of gram negative bacteria by the action of surface polymer:lipid A-core ligase and by detrminants of polymer chain length. *Mol. Microbiol.* **23:** 629-638.

Whitfield, C., and Valvano, M.A. (1993) Biosynthesis and expression of cell surface polysaccharide in gram negative bacteria. *Adv. Micro. Physiol.* **35:** 135-246.

Whitfield, C., and Roberts, I.S. (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli. Mol. Microbiol.* **31:** 1307-1319.

Wigley, D.B., Derrick, J.P., and Shaw, W.V. (1990) The serine acetyltransferase from *Escherichia coli.* Over-expression, purification and preliminary crystallographic analysis. *FEBS Lett.* **17:** 267-271.

Wyk, P., and Reeves, P.R. (1989) Identification and sequence of the gene for abequose synthestase, which confers antigenic specificity on group B *Salmonella*: homology with galactose epimerase. *J. Bacteriol.* **171:** 5687-5693.

Xiang, S.H., Haase, A.M., and Reeves, P.R. (1993) Variation of the *rfb* gene clusters in *Salmonella enterica. J. Bacteriol.* **171:** 5687-5693.

Xiang, S-H., Hobbs, M., and Reeves, P.R. (1994) Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence mediated recombiation event between group E and D1 strains. *J. Bacteriol.* **175**: 4357-4365.

Xue, Y., Zhao, L., Liu, H.W., and Sherman, D.H. (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA.* **95:** 12111-12116.

Yamamoto, K., Wright, A.C., Kaper, J.B., and Morris Jr, J.G. (1990) The cytolysin gene of *Vibrio vulnificus*: sequence and relationship to the *Vibrio cholerae* El Tor haemolysin gene. *Infect. Immun.* **58:** 2706-2709.

Yao, Z., and Valvano, M.A. (1994) Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (*rfb*) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a. *J. Bacteriol.* **176:** 4133-4143.

Yao, R., Burr, D.H., Doig, P., Trust, T.J, Nui, H., and Guerry, P. (1994) Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. *Mol. Microbiol.* **14:** 883-893.

Yokota, S., Kaya, S., Kawamura, T., Araki, Y., and Ito, E. (1986) The structure of the O-specific chain of lipopolysaccharide from *Pseudomonas aeruginosa* IID 1008 (ATCC 27584). *I. Biochem.* **99:** 1551-1561.

Yother, J., Ambrose, K.D., and Caimano, M.J. (1997) Association of a partial H-rpt element with the type 3 capsule locus of *Streptococcus pneumoniae*. *Mol. Microbiol.* **25**: 201-204.

Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P., and Skurnik, M. (1997) Molecular and chemical characterisation of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O:8. *Mol. Microbiol.* **23**: 63-76.

Zhang, L., Al-Hendy, A., Toivanen, P., and Skurnik, M. (1993) Genetic organisation and sequence of the *rfb* gene cluster of *Yersinia enterocolitica* serotype O:3: similarities to the dTDP-L-rhamnsoe biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. *Mol. Microbiol.* **9:** 309-321.

Zhao, S., Sandt, C.H., Feuler, G., Valzany, D.A., Gray, J.A., and Hill, C.W. (1993) *Rhs* elements of *Escherichia coli* K-12: Complex composites of shared and unique components that have different evolutionary history. *J. Bacteriol.* **175**: 2799-2808.

Zhao, Z., White, K.A., Polissi, A., Georgopoulos, C., and Raetz, C.R. (1998) Function of *Escherichia coli* MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *J. Biol. Chem.* **273:** 12466-12475.

Appendix I

APPENDIX 1:

The nucleotide and deduced amino acid sequence of the O-antigen biosynthesis locus (*wbh*) *of V. anguillarum* O1 (85-3954-2)

The nucleotide sequence is numbered in accordance with Genbank accession number AF025396 and is shown from nt 1 to nt 30627 with restriction sites as reference. The amino acid translation for each ORF is represented by single letter code below the first nucleotide of each codon. Possible ribosome binding sites are bolded. The putative JUMPstart sequence is underlined and potential promoters are highlighted and the -35 and -10 boxes indicated. The inverted repeat sequences of IS*1358* are underlined with a double line. Terminator structures are boxed.

```
          10        20        30        40        50        60        70        80        90       100       110       120
  HindIII
5'AAGCTTATTGCGTGAGATCTCTTTACGCTGTTCAGATTCACTGAGTTTGCTGAGAAATGGATGAGTCATTGTATGACCCCCAATTTCAACTAGCCCTGAATCGTGTAATGCTTTGACTTG

         130       140       150       160       170       180       190       200       210       220       230       240
  CTGACGAGACATTAAAGGCACTGGTTTTTCTGGGTTGCTCGGTACTTCAACATCCCAACGGTTAAAATCCTCACCCGTTACTACATAGACAACCGCTTTAAAACCGTATTTTTGTAATAG

         250       260       270       280       290       300       310       320       330       340       350       360
  CGGCAACATCAATTCGTAGTTATCTTGATAGCCGTCATCAACCGTTAAGATGATATAGCGCCTGCCTGATTCAAGGCGGTGAATAAGCCCTTTATCCGCTAAATCTTTAAAGGTTAGCGT

         370       380       390       400       410       420       430       440       450       460       470       480
  TTCAAACCCCATTTTTTTTAGCAGGTTGAAGTGCTTCTCAAGTATGTCAATATGCAGATAGGTACCATGTACGCCCTTTTCACTTTCATCCTTAATAAAGCGATGATACATAATAATCGG

         490       500       510       520       530       540       550       560       570       580       590       600
  CATCTCTTTACGCAGCTTTTCAACATAAACCGTTTGATAAATATCTTCTATTTGATCGACCACATTGGTTAAGCTGTAGCTGTGCTGAATTATTTCACTGACTTGTGGCGAGCAGTGTGG

         610       620       630       640       650       660       670       680       690       700       710       720
  CGTTTTTAACCCTTGTTCTACATATTGACTAATTTTTGAAAAATTAATATCGAGATCTTTAGGCCCGATGTCACCGAAGTTGTTTGCCATGGCCGTAGCGATATTATCGTCATCGACAAT

         730       740       750       760       770       780       790       800       810       820       830       840
  ACCAATTGCACTTGCCTCACCGATTGCCAGTGTAGGCCGTCCACACAACAAAGATTCCATTGCCACTCTGCCAGCCCCAATGATTAAGTCTGATTGAGCAAGTATCGAGACGACATCTTG

         850       860       870       880       890       900       910       920       930       940       950       960
  TGAATAACCAAGAAAGTGAACGTCAGAAGTAAACTTATTAAAACGCGCTTCAATTTTAGATCCAGATATGACCTGAATGTGGTATTTCGCTGAGTCTAAACACTCATCGAGCAACCGATA
```

```
       970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
ACAAAGATCCCCTTTGGGGCCGGTTAATCGACCAATAATACTGACAACAGGTTTGCTATTTTGAGGAGCATTTTGCCAAGCGAATTCTTGCGTTTCAATTCCGTTTCGGCTTACGACGAG

      1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
CGAACTTTGTGGAACATTCAATTCAGTCATGAGTTGTTGTTCAATGGCCTCACAAACCGGCAGCGCTTTGTCACCCATTGCGTGAAATTTTTTACGAGAGGCGTGTACAGGCTGACGGCC

      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
ATGTACTGTGGTGACCATAGGCGTATTTGTTAATTTGCAAGCAATGTGACAACTCCAACTGGAAGCGCGAGAGTGAGCATGCACTAGCTGGATATGGTGTTTCTTAATCAGATAAATAAG

      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
ATAACCCACATGCCAAAAGCGGCGAAGAATACTGCGTTTATTAAAACGGAGTTTATAAAATAGCCCTTGGTGAAGTTTTGTTAGCGTATCGGAAACATAAAATACCTTATGCCCACGCTT

      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
TGTCAGTTCATCGCCAGCGGTTGTGGCATAAACCTCTGCCCCAGTGACCTCAAGTTGAGATAAGGCCATTAATATGTTCATCAATACTCCACGAGTTAGCAGCCAATGCTTTTCATATGT
                     ←orf2

      1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
CTAAGAAACATGATCAATTTTTCATTCCTCCCTCTTCCAATCGTCACTCCATATTAGTCTTTCATTGAGAGGAGAGATCTATTCGTTACTGCCTTTTAGGGAGAGGTGTTTACTAGTTCC

      1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
TCCCCCTTTTAGGGGGAGGTTAGGAGGGGGTGATTACAGCCATCACTACTTATTAATCAACGCCATATACTCAGCCACACCTTCAGCCACCGTCTTAAACTGATGCTGGCAACCTGCGGC

      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920
GCGCAATTTCGTTAAATCGGCCTCAGTGAACTCTTGGTACGCACCTTTTAAGTGATCTGGGAATGGAATGGTTTCAACGTCACCTTTACCATGGTGTTTGATAACGGCTTTAGCGACTTC

      1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040
TTGGAAAGATTCTGCGTTGCCGGTACCCAAGTTATAAATACCGGATACTCCGCTGTTCATAAACCATAAATTCACAGCCGCAACGTCGCCCACATAAACGAAATCACGCTGGAAGGTTTC

      2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160
ACTACCTGCAAACAGTTTCGGGTTTTCACCCGCATTCATTTGATTATTTAAATGAAAAGCAACAGAAGCCATTGAGCCTTTGTGCTGTTCACGCGGGCCATAGACATTGAAATAACGAAA

      2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280
ACCGGTAATTTGGGAAAGAGTTTCACCGTGTGTTTGTGCTTCTTGCCAAATACGGCGCACATAGTTATCAAACTGCTGCTTGGAATAGCCATACACATTCAGCGCACCTTCATACTCTTT

      2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
TTCTTCAATGAAGGTATCGGTATCGCCATAGGTTGCGGCTGATGAAGCATACAAAAATGGAATGTCGCGCTCTAAGCAGTAGTGCAGTAGCTCTTTAGAATACTCATAGTTATTAAGCAT

      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
CACATACTTGCCATCCCATTCCGTGGTGGCAGAGCAAGCGCCTTCATGAAATACGGCTTCGATAGAACCAAAATCCTCACCAGACATGATCTGCACCAAGAAATCGTCTCGGTCCATGTA
```

2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640
ATCGGCAATGTCCAGATCGGCTAAGTTTTTGAATTTACGACCATTTTTTAAATTATCGACAACAAGAATATCACTAATGCCTTGTTCATTTAGCGCTTTGATAATGTTGCTACCAATCAT
←***rfaD***

2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760
-35 -10
GCCAGCACCACCCGTTACGATGATCATACAAAATTCCTATGGTTATAATACGAACAATTACGCGCATTATAGCGGCTTAGCTAGTGGCTGTAATCACAAAATCTAATCTAGGCTAGCGAC

2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880
AGTCTCGCTCTATCTATGGTTTTGATGTAGTATTTCTGCCCAATAAGCTCGCACCGATTTCATAAGGGATGAGTTCGTTTATATGGAATAGCCGCTTGTAGGAGTTAGACCTAAGGGCGG

2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
TAGCATACCTAAAAAACATTAATAATTAATAT**TATTCGGGTAA**CAATATGAAGATTTTAGTTACTGGCGGTGCTGGTTTTATCGGTTCGGCAGTTGTTCGCCATATTATTCAAAACACAC
rmlB M K I L V T G G A G F I G S A V V R H I I Q N T Q

3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120
AAGATAGTGTGATCAATATTGATAAGTTGACCTATGCGGGGAACTTGGAATCATTACAAGGTATTGACAGTAGTGACCGTTATGCTTTCGAACCAGTCGATATCTGTAACCGTTCAGAAT
D S V I N I D K L T Y A G N L E S L Q G I D S S D R Y A F E P V D I C N R S E L

3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240
TAGATCGTGTGTTTTATGTGCATCAACCAGATGCGATTATGCACTTGGCCGCTGAATCCCATGTGGACCGTTCTATTGATGGCCCTGCCGCGTTTATTGAGACCAATATCGTTGGCACTT
D R V F Y V H Q P D A I M H L A A E S H V D R S I D G P A A F I E T N I V G T Y

3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360
ATACCTTGCTTGAAGCTGCGCGCAGTTATTGGAACACGCTTGAGAGCAATAAAAAATCGGCATTCCGCTTCCATCATATTTCAACCGATGAAGTGTATGGTGATTTGGAAGGTACAGATG
T L L E A A R S Y W N T L E S N K K S A F R F H H I S T D E V Y G D L E G T D D

3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480
ACCTCTTTACCGAAGAGACTCCATATTCACCTTCTTCACCTTATTCTGCATCGAAAGCTTCTAGCGATCACTTATTGCGAGCTTGGTTACGTACCTATGGTTTTCCTACCATAGTCACTA
L F T E E T P Y S P S S P Y S A S K A S S D H L L R A W L R T Y G F P T I V T N

3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
ACTGCTCTAACAATTATGGACCATACCATTTTCCTGAAAAGCTGATCCCATTAATGATTCTGAACGCTCTGGACGGTAAACCACTTCCTGTCTATGGCGATGGAATGCAAATTCGCGATT
C S N N Y G P Y H F P E K L I P L M I L N A L D G K P L P V Y G D G M Q I R D W

3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720
GGTTGTTTGTCGAAGATCACGCTCGAGCGCTTTATAAAGTGGTAACAGAAGGCCAAGTAGGTGAAACCTACAACATTGGTGGTCATAACGAAAAAGCTAACATTGAAGTGGTGAACACAC
L F V E D H A R A L Y K V V T E G Q V G E T Y N I G G H N E K A N I E V V N T L

3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840
TCTGTGCATTACTTGAAGACTTAGTTCCGAATAAACCTGCCAACATTACCCATTATGCAGACCTTATTACTTACGTTAAAGACCGCCCTGGACATGATGTGCGTTACGCGATAGATGCCT
C A L L E D L V P N K P A N I T H Y A D L I T Y V K D R P G H D V R Y A I D A S

3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960
CTAAAATCGAGCGAGAGCTAGGCTGGAAGCCAGAAGAGACGTTTGAATCAGGTATTCGAAAAACAGTGGAATGGTATTTGAATAACCGCGAATGGTGGAGCCGAGTATTAGATGGTTCAT
K I E R E L G W K P E E T F E S G I R K T V E W Y L N N R E W W S R V L D G S Y

3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080
ACAGCTGTGAGCGTTTAGGCTCTGGCAATAACAATCAAT**TAAGTTGGCGAT**TTCATGAAAGGCATTATCTTAGCGGGTGGCTCTGGCACTCGTCTTTACCCCATTACTCGTGGTGTATC
S C E R L G S G Q * ***rmlA*** M K G I I L A G G S G T R L Y P I T R G V S

4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
AAAACAACTTCTTCCCATTTACGACAAGCCGATGATCTACTACCCGCTTTCGACATTAATGTTGGCGGGGATTCGCGACATTCTGATTATCACCACTCCTGAAGATAACGAAAGCTTTAA
K Q L L P I Y D K P M I Y Y P L S T L M L A G I R D I L I I T T P E D N E S F K

4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320
ACGCTTATTGGGGGATGGCAGTGACTTCGGTATTCATCTGCAATACGCGATTCAACCTAGCCCAGACGGTTTAGCGCAAGCCTTCCTTATCGGTGAAGAATTTATTGGCAATGATTCAGT
R L L G D G S D F G I H L Q Y A I Q P S P D G L A Q A F L I G E E F I G N D S V

4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440
CTGCTTAGTGTTAGGTGATAACATTTTCTATGGTCAATCATTTAGTAAAACACTGAAAAGCGCCGCTTCACGTGAGCATGGCGCAACAGTTTTTGGTTACCAAGTGAAAGACCCTGAGCG
C L V L G D N I F Y G Q S F S K T L K S A A S R E H G A T V F G Y Q V K D P E R

4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560
TTTTGGCGTAGTGGAGTTTGATGAGCAGATGCGCGCTATCTCAATCGAAGAAAAGCCCCTTAAGCCCAAGTCAAACTATGCCGTCACAGGACTGTATTTCTATGACAACCGAGTGGTGGA
F G V V E F D E Q M R A I S I E E K P L K P K S N Y A V T G L Y F Y D N R V V E

4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680
GCTTGCCAAGCAGGTCAAACCGTCGGCGCGCGGTGAACTAGAAATTACCACCCTTAATGAAATGTATTTAAACGATGGCTCACTCAATGTGGAGCTGCTTGGCCGTGGGTTTGCTTGGTT
L A K Q V K P S A R G E L E I T T L N E M Y L N D G S L N V E L L G R G F A W L

4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
AGATACGGGTACCCATGAAAGCCTGCATGAAGCCTCTTCTTTTGTTGAAACCGTGCAGCATATTCAAGGCTTGAAGATAGCGTGTTTAGAAGAGATTGCGTGGCGAAATGGATGGCTCTC
D T G T H E S L H E A S S F V E T V Q H I Q G L K I A C L E E I A W R N G W L S

4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920
TTCAGGACAACTTCTTGAGTGCGCTAAACCTATGCTTAAGAATGACTATGGTCAATATTTGGCGGGGTTAGCTAA**AGAGGAATCCC**ATTCATGATGCGTGTTCTAGTGACGGGTAGTCAT
S G Q L L E C A K P M L K N D Y G Q Y L A G L A K E E S H S * M R V L V T G S H
rmlD→

4930 4940 4950 4960 4970 4980 4990 5000 5010 5020 5030 5040
GGGCAAGTCGGCTATTGTTTAGCCCATCAACTCAATTCCATGGTTGATGTTGAGTTTTTAGCGGTAGACCGCGAGCAATTGGATATTACCAATCCGTGTCACGTCAATTCGGTGGTTAAC
G Q V G Y C L A H Q L N S M V D V E F L A V D R E Q L D I T N P C H V N S V V N

5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150 5160
GAGTTCAAACCAAACATCATCATTAATGCTGCGGCGCACACCGCGGTTGATAGAGCGGAACAAGAGGCAGAGCTCTCTTACGCGATTAACAGAGATGGCCCTAAACATCTAGCCCAAGCT
E F K P N I I I N A A A H T A V D R A E Q E A E L S Y A I N R D G P K H L A Q A

5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280
GCACATAAGGTTGGTGCGGCTATATTGCACATCTCGACAGACTATGTGTTTTCCGGTGATAAACCGAGCGCATATATAGAAGCCGATTCTACATCGCCTCAGGGGGTTTATGGCAAAAGC
A H K V G A A I L H I S T D Y V F S G D K P S A Y I E A D S T S P Q G V Y G K S

5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
AAATTAGCGGGAGAAATCGCGGTTGCGCAAGCTTGCCCTAGGCACATTATCTTGCGCACGGCTTGGGTGTTTGGTGAGCATGGCAATAACTTTGTGAAAACCATGCTCCGTTTGGCACAA
K L A G E I A V A Q A C P R H I I L R T A W V F G E H G N N F V K T M L R L A Q

```
          5410      5420      5430      5440      5450      5460      5470      5480      5490      5500      5510      5520
    ACACGCGATGCATTAGGTGTTGTGTCAGACCAGTATGGTGGCCCGACGTATGCGGGAGATATTGCTAACGCGTTACTGATAATGGCGAAAGCGATTGTTGAAGGGAAGTCGACGGCCTTT
    T  R  D  A  L  G  V  V  S  D  Q  Y  G  G  P  T  Y  A  G  D  I  A  N  A  L  L  I  M  A  K  A  I  V  E  G  K  S  T  A  F

          5530      5540      5550      5560      5570      5580      5590      5600      5610      5620      5630      5640
    GGTGTGTACCATTTTTCAGGGACACCCCATGTCAGTTGGTATGACTTTGCTTGTGCCATTTTTGATAAAGCCCGTGAACAAGGATTATTAAATAAACCCTTGCAAGTCAATGCGATAACC
    G  V  Y  H  F  S  G  T  P  H  V  S  W  Y  D  F  A  C  A  I  F  D  K  A  R  E  Q  G  L  L  N  K  P  L  Q  V  N  A  I  T

          5650      5660      5670      5680      5690      5700      5710      5720      5730      5740      5750      5760
    ACCAAAGATTACCCGACTCCGGCGAAGAGGCCGGCTAACTCCAAGTTAGAGACCCAGAAAATCCATCAACACTTTGGTGTGGTTGCCAGCGATTGGCGGGCGGCGTTGAACCATCTCATC
    T  K  D  Y  P  T  P  A  K  R  P  A  N  S  K  L  E  T  Q  K  I  H  Q  H  F  G  V  V  A  S  D  W  R  A  A  L  N  H  L  I

          5770      5780      5790      5800      5810      5820      5830      5840      5850      5860      5870      5880
    GACTATAAGAGTGAATAGCAATGAAAGTGATAGAGACAGACATTCCTGATGTCAAAATTATTGAGCCGACCGTGTTTGGTGATGAGCGTGGTTTTTTTATGGAAACCTGGAACCAACAGC
    D  Y  K  S  E  *    M  K  V  I  E  T  D  I  P  D  V  K  I  I  E  P  T  V  F  G  D  E  R  G  F  F  M  E  T  W  N  Q  Q  R
                 rmlC→

          5890      5900      5910      5920      5930      5940      5950      5960      5970      5980      5990      6000
    GCTTTGAAGAACTGGTAACAGGTAAGCCAACGCAGTTTGTGCAAGATAATCACTCAAAATCCAAAAAAGGAATTTTACGTGGCTTGCACTACCAGACAGAAAACACGCAAGGTAAATTGG
      F  E  E  L  V  T  G  K  P  T  Q  F  V  Q  D  N  H  S  K  S  K  K  G  I  L  R  G  L  H  Y  Q  T  E  N  T  Q  G  K  L  V

          6010      6020      6030      6040      6050      6060      6070      6080      6090      6100      6110      6120
    TTAGAGTGGTATCTGGCGAAGTGTTTGATGTGGCAGTAGATATTCGAAAAGATTCACCGACGTTTGGCAAATGGGTTGGTGTGCACCTGTCAGCAGAGAACAAGCGTCAGTTATGGGTAC
      R  V  V  S  G  E  V  F  D  V  A  V  D  I  R  K  D  S  P  T  F  G  K  W  V  G  V  H  L  S  A  E  N  K  R  Q  L  W  V  P

          6130      6140      6150      6160      6170      6180      6190      6200      6210      6220      6230      6240
    CTGAAGGATTTGCGCATGGTTTTTATGTAACCAGTGAAGAAGCGGAATTTGTGTATAAATGTACGAACTATTATAACCCAGAGGCTGAGCACACTTTGCTTTGGAATGATGAATCCATAA
      E  G  F  A  H  G  F  Y  V  T  S  E  E  A  E  F  V  Y  K  C  T  N  Y  Y  N  P  E  A  E  H  T  L  L  W  N  D  E  S  I  N

          6250      6260      6270      6280      6290      6300      6310      6320      6330      6340      6350      6360
    ATATTGATTGGCCAAATAAAACGTGCCCATTACTTTCCGAGAAAGATGTTCATGGTAAACATCTTTCTGATATTATTTGGCTTTAATTATATTTTAAGTCTTTGGTTTGGGCGTTGCTCA
      I  D  W  P  N  K  T  C  P  L  L  S  E  K  D  V  H  G  K  H  L  S  D  I  I  W  L  *

          6370      6380      6390      6400      6410      6420      6430      6440      6450      6460      6470      6480
    AATAATATCTGCTTTTAGTTTGGTGATAAAATTTAATTGATAGACAATATCATAAAAAGCATATTTTTAACTATATTATGTTTTGATTGTAAAAAATTTAGAATATCCTAAGGTAGCGAA

          6490      6500      6510      6520      6530      6540      6550      6560      6570      6580      6590      6600
    AAATGTATAAATTACCGTCCAAGTTTGTCCATGATGTTTTTAGCAGTTATGAGCTGTTGTTTCCGATGGCTAAACGAGATATTAGGAGCCGCTACTTGGGTTCTCTCTTTGGAGGTGTCT
wzm   M  Y  K  L  P  S  K  F  V  H  D  V  F  S  S  Y  E  L  L  F  P  M  A  K  R  D  I  R  S  R  Y  L  G  S  L  F  G  G  V  W

          6610      6620      6630      6640      6650      6660      6670      6680      6690      6700      6710      6720
    GGGCATTCATTCAACCTCTTGTTACCATTCTTGTTATATGGTTCGTTTTTCAGGTCGGATTTAAAGCTCAGGCAACGGATGATGGTGTTCCTTTTACACTTTGGTTGGTTTCTGCAATGA
      A  F  I  Q  P  L  V  T  I  L  V  I  W  F  V  F  Q  V  G  F  K  A  Q  A  T  D  D  G  V  P  F  T  L  W  L  V  S  A  M  I

          6730      6740      6750      6760      6770      6780      6790      6800      6810      6820      6830      6840
    TACCATGGTTTTTCGTTGCTGAAATGTTGTCGAGCGGAACTAACTCAATTATAGAGCAAGCTAATATTGTAAAGAAGATAGTATTCAAAGTTAGCTTGCTACCAATCGTGAAAATAATAT
      P  W  F  F  V  A  E  M  L  S  S  G  T  N  S  I  I  E  Q  A  N  I  V  K  K  I  V  F  K  V  S  L  L  P  I  V  K  I  I  S
```

```
      6850      6860      6870      6880      6890      6900      6910      6920      6930      6940      6950      6960
CAGCATTGATGATCCATATATTTTTCATTGTGATTTTATTTACGGTTTCTATCGCTTACGGTTACTATCCTAAAGTGGGTTGGTTACAAGTTCCATATTATATTTTATGTTCGTTGGTAT
   A  L  M  I  H  I  F  F  I  V  I  L  F  T  V  S  I  A  Y  G  Y  Y  P  K  V  G  W  L  Q  V  P  Y  Y  I  L  C  S  L  V  L

      6970      6980      6990      7000      7010      7020      7030      7040      7050      7060      7070      7080
TGATGCTAGGTATTTCTTGGATTACTTCTTCAATAATTGTTTTCTTCCGAGATGTCGGACAAATTATTGCGGTGTTGATACAGTTAGGTTTCTGGGCAACGCCCATCTTTTGGAATATAA
   M  L  G  I  S  W  I  T  S  S  I  I  V  F  F  R  D  V  G  Q  I  I  A  V  L  I  Q  L  G  F  W  A  T  P  I  F  W  N  I  T

      7090      7100      7110      7120      7130      7140      7150      7160      7170      7180      7190      7200
CGATGGTTCCAATTGAATACCAATGGGTATTGAAACTAAATCCGGTATTTTATATTACGGAAGGCTATAGAAATACAATAACAACAGATCTTTGGTTCTGGGAGTCTTTTTTATGGACTG
   M  V  P  I  E  Y  Q  W  V  L  K  L  N  P  V  F  Y  I  T  E  G  Y  R  N  T  I  T  T  D  L  W  F  W  E  S  F  L  W  T  A

      7210      7220      7230      7240      7250      7260      7270      7280      7290      7300      7310      7320
CATACTATTGGATTTTTACTTCTTTTACTTTATTAGTCGGAATTGTTTGCTTTAAAAAACTAAGACCACATTTTGCTGATGTACTATAGGGAAGAAATATTTGATGAATGATGTAGTTAT
   Y  Y  W  I  F  T  S  F  T  L  L  V  G  I  V  C  F  K  K  L  R  P  H  F  A  D  V  L  *        wzt     M  N  D  V  V  I

      7330      7340      7350      7360      7370      7380      7390      7400      7410      7420      7430      7440
TAGTTGTGAGAATATAACTAAAATATATCCTATGTATGATGATCATCGGGATAGGTTTAAAGAGGTCTTTCACCCGTTCAGAAAAGTTTATCACAAAAAGTTTCACGCTCTTGATGATGT
  S  C  E  N  I  T  K  I  Y  P  M  Y  D  D  H  R  D  R  F  K  E  V  F  H  P  F  R  K  V  Y  H  K  K  F  H  A  L  D  D  V

      7450      7460      7470      7480      7490      7500      7510      7520      7530      7540      7550      7560
TAGTTTTGAGGTGAAAAGAGGTGAGACTGTCGGAATTGTGGGAAAAAATGGTGCCGGTAAATCAACGCTTCTTAAAATTATTACTGGAGTCTTGACTCCAACATCTGGCTCGGTAAAGCT
  S  F  E  V  K  R  G  E  T  V  G  I  V  G  K  N  G  A  G  K  S  T  L  L  K  I  I  T  G  V  L  T  P  T  S  G  S  V  K  L

      7570      7580      7590      7600      7610      7620      7630      7640      7650      7660      7670      7680
CAATGGAGTTGTTTCATCATTACTTGAGTTAGGAACTGGATTTAACCCCGACTTAACAGGTATCGAAAATATTTATTTGAATAGCTCTTTGATGGGAATTCAGAAAGAAGATATAGATAA
  N  G  V  V  S  S  L  L  E  L  G  T  G  F  N  P  D  L  T  G  I  E  N  I  Y  L  N  S  S  L  M  G  I  Q  K  E  D  I  D  K

      7690      7700      7710      7720      7730      7740      7750      7760      7770      7780      7790      7800
AAAACTGGCTCAAGTAATAGCCTTTGCTGATATTGGAGAGCATATCCACCAACCGGTTAGAGGATATTCATCAGGAATGTTTGCCAGACTAGCATTTTCTGTAGCGATCAGTGTTGAACC
  K  L  A  Q  V  I  A  F  A  D  I  G  E  H  I  H  Q  P  V  R  G  Y  S  S  G  M  F  A  R  L  A  F  S  V  A  I  S  V  E  P

      7810      7820      7830      7840      7850      7860      7870      7880      7890      7900      7910      7920
TGATATATTAATTGTCGATGAAGCATTAGCAGTTGGCGACGCCGCTTTTGTAAATAAATGCTATGGAAAAATTAATGAACTAAAAAGTAAGGGTATGACACTCTTATTTGTAAGTCACAG
  D  I  L  I  V  D  E  A  L  A  V  G  D  A  A  F  V  N  K  C  Y  G  K  I  N  E  L  K  S  K  G  M  T  L  L  F  V  S  H  S

      7930      7940      7950      7960      7970      7980      7990      8000      8010      8020      8030      8040
CTTGGGAGCTGTCTCGGAGTTATGTACGAAAGCGATTTTAATTGACAATGGGAAATGTCTGTTGGTTTCTGATGTTGAGACTGTTGTTAACCAGTATAATAGGATGATTAGGTCTTCTTT
  L  G  A  V  S  E  L  C  T  K  A  I  L  I  D  N  G  K  C  L  L  V  S  D  V  E  T  V  V  N  Q  Y  N  R  M  I  R  S  S  F

      8050      8060      8070      8080      8090      8100      8110      8120      8130      8140      8150      8160
TGATAAAAATTTACAAGAAAATGAAAATATAGAAGAAACTACATTCGACAGCAGTGAAAAAGATACTTCAGAAATCTCTCTGATTAGTAATGTTAAATCGAAAGATATATTTAATATTAA
  D  K  N  L  Q  E  N  E  N  I  E  E  T  T  F  D  S  S  E  K  D  T  S  E  I  S  L  I  S  N  V  K  S  K  D  I  F  N  I  N

      8170      8180      8190      8200      8210      8220      8230      8240      8250      8260      8270      8280
TAATGAGTTTATGAATAAGCATTCAAAGACTAGATATGGGAAGGGCGGGGCTAGGATTGTCAATGTAGAGATCATTGAGGGAGACGACAATAAGAATAGGTTGTTTTCGTATAATACAAA
  N  E  F  M  N  K  H  S  K  T  R  Y  G  K  G  G  A  R  I  V  N  V  E  I  I  E  G  D  D  N  K  N  R  L  F  S  Y  N  T  K
```

```
      8290      8300      8310      8320      8330      8340      8350      8360      8370      8380      8390      8400
AATAAAATTAAGGGTATATATAAAATCGGATCAAAATCTGAATAAATTGAACTGTGGATATTTCATTAGAACGGATAAGGGATTGTCTATTGTTGGTAACAACTTAGAGTCTTCAAAGTT
  I  K  L  R  V  Y  I  K  S  D  Q  N  L  N  K  L  N  C  G  Y  F  I  R  T  D  K  G  L  S  I  V  G  N  N  L  E  S  S  K  F

      8410      8420      8430      8440      8450      8460      8470      8480      8490      8500      8510      8520
CGAATTGAGAGATCTAAAGGCTAATCGAAAAGTTATTATAGATTTTGATATCCCATTGACTATCAAGTCTGGTATCTATTCGTTAACGGTTGTTTTGGGTGCCAGCGATTCTGAAAATAA
  E  L  R  D  L  K  A  N  R  K  V  I  I  D  F  D  I  P  L  T  I  K  S  G  I  Y  S  L  T  V  V  L  G  A  S  D  S  E  N  N

      8530      8540      8550      8560      8570      8580      8590      8600      8610      8620      8630      8640
TGAGTTTATAGTCGATTGGGTTGATTTGGCTGACTCATTTGAAGTTTTCCATTCAAAAGACCAATACTCTTTTGAATATATGGTACATATTCAAAATAAAATTTCAATGAATTATATCGA
  E  F  I  V  D  W  V  D  L  A  D  S  F  E  V  F  H  S  K  D  Q  Y  S  F  E  Y  M  V  H  I  Q  N  K  I  S  M  N  Y  I  D

      8650      8660      8670      8680      8690      8700      8710      8720      8730      8740      8750      8760
CTAAATATTTTTTACAAAGGTTCTTATATATGTTTAATAACAAAATATTCCAAACGATTAAAAGAGATTCTGAATTAGAAGAAAAAATAACACAAGATATGAAAGAAAATTCGGGTGTAT
  *                     wbhA  M  F  N  N  K  I  F  Q  T  I  K  R  D  S  E  L  E  E  K  I  T  Q  D  M  K  E  N  S  G  V  F

      8770      8780      8790      8800      8810      8820      8830      8840      8850      8860      8870      8880
TTGCTCCACATTCTCACTGGGATAAGGTCAATACTACATACTGCACTTCAGATGAGGAATCGCTGACCTCTCTTTTGGATGATTTTATTGGTTCTGAAGGTTTAAAACCAATTGACCAGC
   A  P  H  S  H  W  D  K  V  N  T  T  Y  C  T  S  D  E  E  S  L  T  S  L  L  D  D  F  I  G  S  E  G  L  K  P  I  D  Q  Q

      8890      8900      8910      8920      8930      8940      8950      8960      8970      8980      8990      9000
AGCAACGAGCATTGTATGAAAATTTTTTTGAGGCAATAGATAACTTGAATTTGAAAAATAATTTGAATGATAGTGCTATCGCAACAGATAACGGAAGTGTAATGGATATTAACCTGATTA
   Q  R  A  L  Y  E  N  F  F  E  A  I  D  N  L  N  L  K  N  N  L  N  D  S  A  I  A  T  D  N  G  S  V  M  D  I  N  L  I  N

      9010      9020      9030      9040      9050      9060      9070      9080      9090      9100      9110      9120
ATATTTTTAGTAAGAATTCTCGAGATAAAACCTTTAATATTTGTGAGGTTGGCGGAGGCTACGGGCGTCTGGCTAGAGTTTTTTCTAATTTTTATAAGGGTTCCGTAAAATACGTATTGG
   I  F  S  K  N  S  R  D  K  T  F  N  I  C  E  V  G  G  G  Y  G  R  L  A  R  V  F  S  N  F  Y  K  G  S  V  K  Y  V  L  V

      9130      9140      9150      9160      9170      9180      9190      9200      9210      9220      9230      9240
TTGACTCTGTGCCTGTGAGTATTATGTTTTCTTACCAATATTTGGTTGATCAATGCCCTGAAGCTAAAATCGGTTTTTACTATAACGGTGATGAGTTTGATTTAGATAAGTATGATATAT
   D  S  V  P  V  S  I  M  F  S  Y  Q  Y  L  V  D  Q  C  P  E  A  K  I  G  F  Y  Y  N  G  D  E  F  D  L  D  K  Y  D  I  Y

      9250      9260      9270      9280      9290      9300      9310      9320      9330      9340      9350      9360
ATATCGTACCATCTTGGCACTTTGAAAAAATGAATAATGTAAAGTATGATATTGCCATTAATATAGAAAGTATGCAGGAAATGAACTTTGAGGAAATTAAAAGGTTCATGAATATCTTTG
   I  V  P  S  W  H  F  E  K  M  N  N  V  K  Y  D  I  A  I  N  I  E  S  M  Q  E  M  N  F  E  E  I  K  R  F  M  N  I  F  D

      9370      9380      9390      9400      9410      9420      9430      9440      9450      9460      9470      9480
ATAGATGTGTGAAAGATAATGGTATTATTTACTTATCTAACTCGAAAGAATATGTTTACAAAGGAGAGTGGCCATATCCAGATAACTGGGAATGCCAATTTAGACAAAATACCCCTCGTT
   R  C  V  K  D  N  G  I  I  Y  L  S  N  S  K  E  Y  V  Y  K  G  E  W  P  Y  P  D  N  W  E  C  Q  F  R  Q  N  T  P  R  S

      9490      9500      9510      9520      9530      9540      9550      9560      9570      9580      9590      9600
CATGGACGAAAAAACATCCGACAGAGATCTTTAAAAAATCAGATGGTTCATTTTCGGATAAGGAATCTTTGTTTGAATTTTACTCAAATGTCGAAGAGTTATATGAAATGGGATTCGATG
   W  T  K  K  H  P  T  E  I  F  K  K  S  D  G  S  F  S  D  K  E  S  L  F  E  F  Y  S  N  V  E  E  L  Y  E  M  G  F  D  G

      9610      9620      9630      9640      9650      9660      9670      9680      9690      9700      9710      9720
GTAATTCAACACAATATTTGAGTAAAAATAGCTTAGAAGAATCATTAAAAAAAGCGTATGAAAACAATTCCTATTTGACAGGAAAAGTTTCTCATTTGACAGAAGAAAAAGAAAAAGCTG
   N  S  T  Q  Y  L  S  K  N  S  L  E  E  S  L  K  K  A  Y  E  N  N  S  Y  L  T  G  K  V  S  H  L  T  E  E  K  E  K  A  E
```

```
      9730      9740      9750      9760      9770      9780      9790      9800      9810      9820      9830      9840
AGCAGTATATTAAAGAGATTGAATCATCAAGGCTATACAAATTAAGGTCTTTTCTAAAAAGATAATGGTATTTTAGAGCAGGGAAACCGCATGAAGGTCGTTGATATATAAGGGCTGCAA
  Q  Y  I  K  E  I  E  S  S  R  L  Y  K  L  R  S  F  L  K  R  *         IS1358→

      9850      9860      9870      9880      9890      9900      9910      9920      9930      9940      9950      9960
AGATGCTAAGAATCCGCTTTGAACCTATCACTAATCCCAACTATTGAATGTAACCACATGAAAATAAAGATCTGCTTGTAGATCATTTTCGTTGTTTTCTAATTGCTCCTAGACAATATA

      9970      9980      9990     10000     10010     10020     10030     10040     10050     10060     10070     10080
GGAGCCTAAATATGAGCGAGTTAACCAACCCATTTATGCATTTCCAAATCATTAAAGATTATCGACAGGAAAATAAAATAGCACACAAATTATCAGACATTATTTTGCTGACAATTTGCG

     10090     10100     10110     10120     10130     10140     10150     10160     10170     10180     10190     10200
GCGTCCTGTCGGGTCATGATGGCTGGGATGGCATTATCGATTTTGGCAATGCTCGCTTAGATTTCCTCAAACGATATGGTCACTTTGACGCTGGTATTCCTTCTGCAGATACGCTGTCTC

     10210     10220     10230     10240     10250     10260     10270     10280     10290     10300     10310     10320
GTGTGATGGGCATGATGAATCCGGCAGCGTTGCAAAGAAGCTTTATAGCTTGGATGAAGGACTGCCATACACTGACGGATGGAGAAGTCATCGCTATCGACGGTAAAACATTACGTGGCT

     10330     10340     10350     10360     10370     10380     10390     10400     10410     10420     10430     10440
CTTATGACCGCTCGAAGGGTAAAGGAACAATTCACATGGTGAATGCTCTTGCTACTGCAAACGGAATGAGTATCGGCCAACAGAAGGTTGATAGTAAAAGTAATGAAATTACCGCGATCC

     10450     10460     10470     10480     10490     10500     10510     10520     10530     10540     10550     10560
CCAAGCTACTTGACTTGCTAGAGGTAAAAGGCTGCTTGGTTACGATTGATGCAATGGATGCCAAAAGAAAATAGCGCAAGAAATTCTCGATAAAGAAGCGGATTATTTGTTAGCAGTGA

     10570     10580     10590     10600     10610     10620     10630     10640     10650     10660     10670     10680
AAGGGAATTAGGGAATGCTTGAGCAAGCCTTTGATGATTATTTTCGAATGGACATGCTTCAAGACTTTGACGGCAGTTCTTACAGTACTCAAGAAAAAAGTCACGGAAGAACGGAAACGA

     10690     10700     10710     10720     10730     10740     10750     10760     10770     10780     10790     10800
                                                                                        -35
GAGTGGCTTTAGTGAGCCGTGATTTGTCGGTTTTAGGTGATATTGAGCATGAATGGCCCGAGCTAAAATCAATGGGCATCGTCGCTTCGATTCGCCAAGAATCGGCTGTAGCAACAGAGC

     10810     10820     10830     10840     10850     10860     10870     10880     10890     10900     10910     10920
                -10
AAGATGTGAGTATTCGTTACTATATTTGTTCTAAAGAATAAGAAGCCCAAACTCTGCTTGAAGCGACGCGTTCTCATTGGGGCGTAGAGGCCATGCATTGGTCCCTTGATACCGCATTTA

     10930     10940     10950     10960     10970     10980     10990     11000     11010     11020     11030     11040
GTGAGGACAACTCTCGTATTAGAGCCGATGATCGCGCAGAGGCTTTAGCAAGGATCAGGCAGATGTGTTTGAACCTATTAAAGAGTGAAACCACCTTTAAAGGTGGCATTAAACGCAAAC

     11050     11060     11070     11080     11090     11100     11110     11120     11130     11140     11150     11160
GGATGAACTGCGCAATGGACGAAAACTACCTAAGTAAGGTTCTCGAAAGCCTTATCGGGCGGTGATGTTCATGCGGTTTCCGTGAACTAACGGAGTTGAAATGGTAAATAATAAACTTAA
                                                                        ← IS1358             wbhB    M  V  N  N  K  L  K

     11170     11180     11190     11200     11210     11220     11230     11240     11250     11260     11270     11280
GGAAGTAGAGTTTCATAATATGATTCTGCTTTTTTTGCTTTACACGTTTGTGTATGCAATTTCAGTTGTTACGAGAGTAAGTTCATCTGAGTTTATGGGGATTGATCCTCAATCAATAGT
  E  V  E  F  H  N  M  I  L  L  F  L  L  Y  T  F  V  Y  A  I  S  V  V  T  R  V  S  S  S  E  F  M  G  I  D  P  Q  S  I  V

     11290     11300     11310     11320     11330     11340     11350     11360     11370     11380     11390     11400
TACTGCTATTAATGAGTTAACAACACCTCCTTACTATAATATGAATGATTCATATCACTCAAAGTATTATGGATGGACGTATTTTTCTCTGAACTTTCTGGTTGTTTTAATAGCAAAATG
  T  A  I  N  E  L  T  T  P  P  Y  Y  N  M  N  D  S  Y  H  S  K  Y  Y  G  W  T  Y  F  S  L  N  F  L  V  V  L  I  A  K  C
```

```
     11410     11420     11430     11440     11450     11460     11470     11480     11490     11500     11510     11520
TATGGGGTTTAATTCAGAGTTTGAAATTAATGTGCTTATTCGTAGTGTTGTTTATATAATTGGCGCATGTTTAGTTCTATCCTTATACATTTTCTCTCGAGAGATATTCTCGAAGTTTGT
 M  G  F  N  S  E  F  E  I  N  V  L  I  R  S  V  V  Y  I  I  G  A  C  L  V  L  S  L  Y  I  F  S  R  E  I  F  S  K  F  V

     11530     11540     11550     11560     11570     11580     11590     11600     11610     11620     11630     11640
ATCTTTTGTTTAGTGTTGTATTTTATGTTTGACCCTGTTGTCTCTCATTATATAACATTAATCCACCCTGAAGCGTTAGGTATGACCTTGCAAATACTAGGTTGTTATTTCCTTATTAG
  S  F  V  L  V  L  Y  F  M  F  D  P  V  V  S  H  Y  I  T  L  I  H  P  E  A  L  G  M  T  L  Q  I  L  G  C  Y  F  L  I  R

     11650     11660     11670     11680     11690     11700     11710     11720     11730     11740     11750     11760
ATTCTATAAGGGTGGAGGGGAAAATAGCTGGATTTTTTATTTTTCTATAATATTACTTTCTCTATCCTCTTTAGCAAAGCAACCATTTTTTATAATTAATTTTTTTATTGGATGTATTTA
  F  Y  K  G  G  G  E  N  S  W  I  F  Y  F  S  I  I  L  L  S  L  S  S  L  A  K  Q  P  F  F  I  I  N  F  F  I  G  C  I  Y

     11770     11780     11790     11800     11810     11820     11830     11840     11850     11860     11870     11880
CCTCAAGTTTCTTGCGGACAAACTTAATTTAAGCTATAGAAATATAATTAGTTTTTTTCTCCATTCTGTAATAATTTTCCTTGCCTGCTTCTTAATTATTCATCCTTATGCTTTTATTGA
  L  K  F  L  A  D  K  L  N  L  S  Y  R  N  I  I  S  F  F  L  H  S  V  I  I  F  L  A  C  F  L  I  I  H  P  Y  A  F  I  E

     11890     11900     11910     11920     11930     11940     11950     11960     11970     11980     11990     12000
GTTTGATAGGTTTATTTTGGCTCAGTCAGAACTGTCATCAGGCCATTCGTCAGGAAGTATGAGTGAAGTTTTAAATATTTGGTTTTCTGAATATTCAAAAAGCTTACTTTTTTTCTTCCA
  F  D  R  F  I  L  A  Q  S  E  L  S  S  G  H  S  S  G  S  M  S  E  V  L  N  I  W  F  S  E  Y  S  K  S  L  L  F  F  F  H

     12010     12020     12030     12040     12050     12060     12070     12080     12090     12100     12110     12120
TACAGTTGTTTTGGTATTAGTTACATTTTCTAGGGATAAAAATAAGTACCACTTTATATCTCTAGTTACGGTATCATTGATAGTTGTTGTGTTTATGTATAAGTCTAGGATTTTCATTAA
  T  V  V  L  V  L  V  T  F  S  R  D  K  N  K  Y  H  F  I  S  L  V  T  V  S  L  I  V  V  V  F  M  Y  K  S  R  I  F  I  N

     12130     12140     12150     12160     12170     12180     12190     12200     12210     12220     12230     12240
CCTTGGATATTTATTCCCTTTATATTTATTGGCTTTTGTAAATATTACTTATTTTTTTGTTAAATATTTAAAAATAAATAATATATATAATAAATCTCTAGTCTCCATTATTTTAGTCTT
  L  G  Y  L  F  P  L  Y  L  L  A  F  V  N  I  T  Y  F  F  V  K  Y  L  K  I  N  N  I  Y  N  K  S  L  V  S  I  I  L  V  L

     12250     12260     12270     12280     12290     12300     12310     12320     12330     12340     12350     12360
AGTTCCCCTTAATTTTTTATCTAATTTTTTCTTCAGCGTATTTGAAACCCAGCATAAGTACTATATTGATGGCTTGGCAACGAAAAACTCAATATGGAACTATATAAAAACTCTTCCTGA
  V  P  L  N  F  L  S  N  F  F  F  S  V  F  E  T  Q  H  K  Y  Y  I  D  G  L  A  T  K  N  S  I  W  N  Y  I  K  T  L  P  E

     12370     12380     12390     12400     12410     12420     12430     12440     12450     12460     12470     12480
AAATACAAAAATTGCATATTCACCAAACATTGCAGTACCAAATCCATACAAGTCAATAGGTTGCCATGCTTGGCAAGGCTGCGCAAAATCAACCGATTTAAGAAAGTATAATCCTGATGT
  N  T  K  I  A  Y  S  P  N  I  A  V  P  N  P  Y  K  S  I  G  C  H  A  W  Q  G  C  A  K  S  T  D  L  R  K  Y  N  P  D  V

     12490     12500     12510     12520     12530     12540     12550     12560     12570     12580     12590     12600
GATTGTTTATAGTCCTAAATATACCTTCTTCGAATCAGATGAATATACAAACTATATTAATAAATATGGTTATATTCTTGTGAGTACTGTATCACCAGCACCTGAAGTGAATTATACGTG
  I  V  Y  S  P  K  Y  T  F  F  E  S  D  E  Y  T  N  Y  I  N  K  Y  G  Y  I  L  V  S  T  V  S  P  A  P  E  V  N  Y  T  C

     12610     12620     12630     12640     12650     12660     12670     12680     12690     12700     12710     12720
TAGTTCAACATCTTCCGTAGGGGAGGGGAAAGGGGTAGAGAATACTTTCTACTTTTTTAATATACCTCGTCTTGTCAGTAATATTTCGAGATGTATTGATAGTTATAAGCTCTCAATACT
  S  S  T  S  S  V  G  E  G  K  G  V  E  N  T  F  Y  F  F  N  I  P  R  L  V  S  N  I  S  R  C  I  D  S  Y  K  L  S  I  L

     12730     12740     12750     12760     12770     12780     12790     12800     12810     12820     12830     12840
CCAAAAGAAAAAATCAAGCTTAACGGGATTAGAATCTATGTATATGAGAGGTAAAGTTATTGAAACCTAATGATTAGTGGATACGGTGAGTTTATATTATTCCCTCTTTACAATAAAATT
  Q  K  K  K  S  S  L  T  G  L  E  S  M  Y  M  R  G  K  V  I  E  T  *
```

12850 12860 12870 12880 12890 12900 12910 12920 12930 12940 12950 12960
-35 -10
TTATGAAAAAGTTCCAGGTTTTATTTAAATATTAATCGTACGAAGCATTTTTAGATGCGTACTAAGTCATATAT**TCAGAAATA**TAGTTTATGGAATTTAAAGTAATAGTTAAACCGTCAG
ddhD M E F K V I V K P S G

12970 12980 12990 13000 13010 13020 13030 13040 13050 13060 13070 13080
GTGTTGAATATCAAAGTGGAAGAAATATTCTTGATGATGCCTTTGCAAGCTCCATATCATTAGAGCACAGTTGTAAAACAGGTGACTGTGGAGTTTGTTGTGCTGAGGTAATTTCTGGTT
V E Y Q S G R N I L D D A F A S S I S L E H S C K T G D C G V C C A E V I S G L

13090 13100 13110 13120 13130 13140 13150 13160 13170 13180 13190 13200
TGGTAGAAAACGAGAATGGTGAATTGGTAACTCAAGGACATATCCTTACTTGCCAATCAAAAGCAAAATCAGATGTTGTTTTAAAGGCTAATTATTATCCTGAACTTGTAGATATAAAGC
V E N E N G E L V T Q G H I L T C Q S K A K S D V V L K A N Y Y P E L V D I K Q

13210 13220 13230 13240 13250 13260 13270 13280 13290 13300 13310 13320
AGCAAACAATATCATGTAAAGTTGCAAGCTTTGAATTTGTCACTAAAGATATTGTTTCGATTAGGTTTCGATTTCACCAAAAACTATATTCAATTATTTGCCAGGCCAGTATGTTGATT
Q T I S C K V A S F E F V T K D I V S I R F R F S P K T I F N Y L P G Q Y V D L

13330 13340 13350 13360 13370 13380 13390 13400 13410 13420 13430 13440
TAAGTTTTAGGGGAGTAAAGCGTAGTTACTCAATAGCGAATGCTAAAAGTAAGTCTAATGAATTAGAATTGCATATCCGTAAAGTTCCAAATGGGGAAATGTCAGAGCTATTATTTGAAC
S F R G V K R S Y S I A N A K S K S N E L E L H I R K V P N G E M S E L L F E H

13450 13460 13470 13480 13490 13500 13510 13520 13530 13540 13550 13560
ATTTAAAAGAAAATCAGTTAATGCGAATAGAAGGTCCGAAAGGCACATTTTTTGTAAGGGATAATATTAAACCATTAATTTTTATTGCAACAGGTACTGGTATCGCTCCAATAAAAGCCA
L K E N Q L M R I E G P K G T F F V R D N I K P L I F I A T G T G I A P I K A I

13570 13580 13590 13600 13610 13620 13630 13640 13650 13660 13670 13680
TAGTTGAAGAACTTATTGCAAAAGAAGATAAACGTAATGTTTATATCTATTGGGGTATGCGATACAAAAATGAAATATATTGTGATGAATTATCTTTGTTAGCTGCAGAAAACCAGAACA
V E E L I A K E D K R N V Y I Y W G M R Y K N E I Y C D E L S L L A A E N Q N I

13690 13700 13710 13720 13730 13740 13750 13760 13770 13780 13790 13800
TTTTCTTTAATTTAGTATTGTCGCGTGAGTTTGAGGTTTCGCCTGACTATAAAAAAGGTTATGTTCAAGACGCAGTAATTCGCGATTTTAATTCATTAAAGGACATTGAAGTTTACGCTT
F F N L V L S R E F E V S P D Y K K G Y V Q D A V I R D F N S L K D I E V Y A C

13810 13820 13830 13840 13850 13860 13870 13880 13890 13900 13910 13920
GTGGTTCTTCCAAAATGATAGAATGTGCCAAAGCCTTATTGTTACAGCATCAACTCCCGAACGATGCATTTTTCTCAGATGCATTCACACCCGCAAAGTAGTTTAATAGTTA**GTTGGAGT**
G S S K M I E C A K A L L L Q H Q L P N D A F F S D A F T P A K *

13930 13940 13950 13960 13970 13980 13990 14000 14010 14020 14030 14040
CATATAATGAAAGTAGTCATTTTAGCTGGTGGTTTAGGTACACGCTTAAGTGAAGAGACATCAGTGAAACCAAAGCCAATGGTTGAAATTGGTGGTAAGCCCATCTTATGGCATATCATG
ddhA M K V V I L A G G L G T R L S E E T S V K P K P M V E I G G K P I L W H I M

14050 14060 14070 14080 14090 14100 14110 14120 14130 14140 14150 14160
AAGCAATATTCTGCTCACGGTATTAACGATTTCATTATTTGTTGTGGTTATAAGGGGTATATCATTAAGGAGTATTTTGCTAATTATTTCTTACATATGTCTGATGTTACCTTTGACATG
K Q Y S A H G I N D F I I C C G Y K G Y I I K E Y F A N Y F L H M S D V T F D M

14170 14180 14190 14200 14210 14220 14230 14240 14250 14260 14270 14280
AAAGAAAATAAAATGGAAGTGCATCATAAGAGAGCAGAACCTTGGACCGTAACATTAGTTGATACTGGTGATAACTCAATGACTGGTGGCCGGTTAGCTCGTGTTGCTGATTATGTAAAA
K E N K M E V H H K R A E P W T V T L V D T G D N S M T G G R L A R V A D Y V K

```
     14290     14300     14310     14320     14330     14340     14350     14360     14370     14380     14390     14400
GATGAGGAAGCTTTCTGTTTTACTTATGGTGATGGCGTCAGTGATATAGATATTACTAAGTCAATAGAGTTTCACCAAGCGCATGGTAAACAAGCTACATTGACAGCAACATTTCCGCCA
D  E  E  A  F  C  F  T  Y  G  D  G  V  S  D  I  D  I  T  K  S  I  E  F  H  Q  A  H  G  K  Q  A  T  L  T  A  T  F  P  P

     14410     14420     14430     14440     14450     14460     14470     14480     14490     14500     14510     14520
GGTCGTTTTGGAGCTCTAGATATTACTTCAGGTAAGGTGGATAACTTTAAAGAAAAGCCTCGTGGTGATGGTGCTATGATTAATGGCGGTTTTTTTGTTCTTTCGCCAAGAGTATTGCAG
G  R  F  G  A  L  D  I  T  S  G  K  V  D  N  F  K  E  K  P  R  G  D  G  A  M  I  N  G  G  F  F  V  L  S  P  R  V  L  Q

     14530     14540     14550     14560     14570     14580     14590     14600     14610     14620     14630     14640
CTTATCGACAGTGATAGTTGCATTTGGGAACAGTACCCTCTGAATAGATTAGCAGATGATGGTGAGTTAATGGCCTATGAGCACAATGGTTTTTGGCAACCAATGGATACATTGCGAGAT
L  I  D  S  D  S  C  I  W  E  Q  Y  P  L  N  R  L  A  D  D  G  E  L  M  A  Y  E  H  N  G  F  W  Q  P  M  D  T  L  R  D

     14650     14660     14670     14680     14690     14700     14710     14720     14730     14740     14750     14760
AAGCTATACCTAGATGAATTATGGCAAGCTGGTAAAGCGCCTTGGAAAATTTGGGAGTAATAATGAATCCAACATTTTGGCACGGTAAAAAAGTGTTTATTACTGGACATACTGGCTTTA
K  L  Y  L  D  E  L  W  Q  A  G  K  A  P  W  K  I  W  E  *    M  N  P  T  F  W  H  G  K  K  V  F  I  T  G  H  T  G  F  K
                                                  ddhB

     14770     14780     14790     14800     14810     14820     14830     14840     14850     14860     14870     14880
AAGGTGGCTGGCTTTCACTATGGCTACAGGAAATGGGTGCTATAGTGAAAGGATATTCTCTTCCGGCTCCTACAACTCCCAGTTTATTTGAGCAAGGGAAAGTATGGGCTGGAATGAGAA
  G  G  W  L  S  L  W  L  Q  E  M  G  A  I  V  K  G  Y  S  L  P  A  P  T  T  P  S  L  F  E  Q  G  K  V  W  A  G  M  R  T

     14890     14900     14910     14920     14930     14940     14950     14960     14970     14980     14990     15000
CAGAAGAAGGCGATATCCGCGATTTTACTCACATGCGACAAAGTATGTATGAGTTTAAACCAGAAATAGTTTTTCATATGGCGGCCCAGCCATTAGTACGCCTTTCATACCATGAGCCAA
  E  E  G  D  I  R  D  F  T  H  M  R  Q  S  M  Y  E  F  K  P  E  I  V  F  H  M  A  A  Q  P  L  V  R  L  S  Y  H  E  P  I

     15010     15020     15030     15040     15050     15060     15070     15080     15090     15100     15110     15120
TCGAAACATACTCAACGAATGTTATGGGTACGGTCTATTTGTTAGAAGCTGTTAAGCAAGTTGGTGGCGTTAAGGCTGTAGTTAATATTACATCCGATAAATGTTACGAAAACCGAGAGT
  E  T  Y  S  T  N  V  M  G  T  V  Y  L  L  E  A  V  K  Q  V  G  G  V  K  A  V  V  N  I  T  S  D  K  C  Y  E  N  R  E  W

     15130     15140     15150     15160     15170     15180     15190     15200     15210     15220     15230     15240
GGGTTTGGGGATACCGAGAAGACGAACCAATGGGTGGTTATGATCCATATAGCAACAGTAAAGGATGTGCTGAGTTGGTTGCTTCATCTTATCGCCAGTCCTTTTTTAATAAAGAAAATT
  V  W  G  Y  R  E  D  E  P  M  G  G  Y  D  P  Y  S  N  S  K  G  C  A  E  L  V  A  S  S  Y  R  Q  S  F  F  N  K  E  N  Y

     15250     15260     15270     15280     15290     15300     15310     15320     15330     15340     15350     15360
ACCATCAGCATAGATGTGCTTTAGCATCAGTTCGTGCAGGTAACGTTATTGGTGGTGGCGATTGGGCAGAAGATCGATTGATTCCAGATATGTTAAAAGCATTTTCAAATAATCAAACAG
  H  Q  H  R  C  A  L  A  S  V  R  A  G  N  V  I  G  G  G  D  W  A  E  D  R  L  I  P  D  M  L  K  A  F  S  N  N  Q  T  V

     15370     15380     15390     15400     15410     15420     15430     15440     15450     15460     15470     15480
TAGAAATTCGAAGTCCTCATGCAATTCGTCCTTGGCAGCATGTTCTTGAACCTTTGTCTGGTTACCTAACTATAGCAGAAAAGCTTTATACTGTAGGTCCCGGCTTTGCTGAAGGATGGA
  E  I  R  S  P  H  A  I  R  P  W  Q  H  V  L  E  P  L  S  G  Y  L  T  I  A  E  K  L  Y  T  V  G  P  G  F  A  E  G  W  N

     15490     15500     15510     15520     15530     15540     15550     15560     15570     15580     15590     15600
ATTTTGGTCCTCGAGATGAAGATGCAAAGCCTGTTGATTGGATAGTTAATCGATTAACTGAACTTTGGGGGGGAGGGGCTCGCTGGTCTTTGTGTGATGGGGAACATCCACATGAGGCTC
  F  G  P  R  D  E  D  A  K  P  V  D  W  I  V  N  R  L  T  E  L  W  G  G  G  A  R  W  S  L  C  D  G  E  H  P  H  E  A  H

     15610     15620     15630     15640     15650     15660     15670     15680     15690     15700     15710     15720
ATTATCTTAAGCTTGATTGTTCGAAAGCAAAAATGCGACTAGATTGGCAGCCTGTTTGGGATGTAGATACTACACTGGAAAAAATTGTCTCTTGGCATAAAGCTTGGCTTGCAAAACAAG
  Y  L  K  L  D  C  S  K  A  K  M  R  L  D  W  Q  P  V  W  D  V  D  T  T  L  E  K  I  V  S  W  H  K  A  W  L  A  K  Q  D
```

15730 15740 15750 15760 15770 15780 15790 15800 15810 15820 15830 15840
ATATGCACCAATATACAATTAACGAAATTAAACAATATATGACCGC**ACGAGCAGGT**AAAGAGTAATGTCAAAAGAACAATTAAGAAGTCAAATTGCAGAATTAGTAGCACAATATGCGGT
M H Q Y T I N E I K Q Y M T A R A G K E *M S K E Q L R S Q I A E L V A Q Y A V
ddhC

15850 15860 15870 15880 15890 15900 15910 15920 15930 15940 15950 15960
CATTGAATACGCGCCTAAGATGTTTATCGGAGGAGAAAGCGTTGTTCCGCCATCGGGTAAAGTTCTTGGAGCAAAAGAACTACAACTTATGGTAGATGCTTCTCTTGATGGATGGTTAAC
I E Y A P K M F I G G E S V V P P S G K V L G A K E L Q L M V D A S L D G W L T

15970 15980 15990 16000 16010 16020 16030 16040 16050 16060 16070 16080
TACTGGTCGCTTTAATGATGCTTTTGAAAAACGCTTAGGTGAGTACCTCGGTGTTCCATTTGTGTTGACTACCACTTCAGGTTCCTCTGCTAACCTATTAGCACTGACAACGTTAACTTC
T G R F N D A F E K R L G E Y L G V P F V L T T T S G S S A N L L A L T T L T S

16090 16100 16110 16120 16130 16140 16150 16160 16170 16180 16190 16200
TCCAAAACTTGGTGATCGTCAATTGAAACCGGGTGATGAAGTGATCACTGTAGCTGCAGGTTTTCCAACGACAGTTAACCCTACGATACAAAACGGTTTAATTCCCGTATTTGTGGATGT
P K L G D R Q L K P G D E V I T V A A G F P T T V N P T I Q N G L I P V F V D V

16210 16220 16230 16240 16250 16260 16270 16280 16290 16300 16310 16320
TGATATTCCAACTTATCAAATCAAGCCAGAAATGATTGAAGCTGCTGTATCTGAAAAAACTAAAGCAATAATGGTTGCACATACATTAGGCAATACGTTTGACCTTACCGAAGCTCGAAG
D I P T Y Q I K P E M I E A A V S E K T K A I M V A H T L G N T F D L T E A R R

16330 16340 16350 16360 16370 16380 16390 16400 16410 16420 16430 16440
AGTTGCTGATAAGTACAACCTTTGGTTAATTGAAGACTGTTGTGATGCACTAGGTTCAACTTACAATGGCAAAATGGTTGGTACCATTGGTGATATTGCTACAGTAAGTTTCTATCCAGC
V A D K Y N L W L I E D C C D A L G S T Y N G K M V G T I G D I A T V S F Y P A

16450 16460 16470 16480 16490 16500 16510 16520 16530 16540 16550 16560
TCATCATATTACAATGGGTGAGGGCGGTGCAGTATTCACAAAAGATAAAGAACTACGTAAACTTATTGAATCTTTCAGAGACTGGGGACGTGATTGTTATTGTGCTCCTGGTTGTGACAA
H H I T M G E G G A V F T K D K E L R K L I E S F R D W G R D C Y C A P G C D N

16570 16580 16590 16600 16610 16620 16630 16640 16650 16660 16670 16680
TACTTGTGGAAAACGCTTCGACCAACAACTAGGATCTTTACCTCAAGGATATGACCATAAATATACTTATTCTCATCTAGGATATAATTTAAAAATTACAGATATGCAAGCTGCTTGTGG
T C G K R F D Q Q L G S L P Q G Y D H K Y T Y S H L G Y N L K I T D M Q A A C G

16690 16700 16710 16720 16730 16740 16750 16760 16770 16780 16790 16800
TCTGGCTCAAATGGATCGCGTTGAAGAGTTTGTTCAAGCGCGTAAAGAAAACTTCGCCTACCTTAAAAATGGGTTAGCTTCATGCGAAGAGTTCATTATCTTACCAGAGGCGACAGAAAA
L A Q M D R V E E F V Q A R K E N F A Y L K N G L A S C E E F I I L P E A T E N

16810 16820 16830 16840 16850 16860 16870 16880 16890 16900 16910 16920
TTCAGAACCATCTTGGTTTGGTTTCCCCATTACAATAAAAGATGATTCTGGCATTAGCCGTGTAGATTTGTTGAAATTTATGGATCAGCACAAAATAGGTACACGCTTGTTATTTGCTGG
S E P S W F G F P I T I K D D S G I S R V D L L K F M D Q H K I G T R L L F A G

16930 16940 16950 16960 16970 16980 16990 17000 17010 17020 17030 17040
TAATTTGACACGTCAACCATATTTTGAGCATGTAAAATATAGAGTAGTGGGTGAACTCACTAATACTGATTTAATCATGAATAATACGTTTTGGATTGGTGTATACCCAGGTTAACTCT
N L T R Q P Y F E H V K Y R V V G E L T N T D L I M N N T F W I G V Y P G L T L

17050 17060 17070 17080 17090 17100 17110 17120 17130 17140 17150 17160
AGCTCATTTAGATTTTGTGCTTGAAAAATTTGAAGAGTTCTTTGGAGTAAATTTTTAAA**TTTAGGATGT**AATATATGAAATCCTCTGATGCAATAGCACACCTGCTCTCACTAAATAATG
A H L D F V L E K F E E F F G V N F * ***wbhC*** M K S S D A I A H L L S L N N V

```
     17170     17180     17190     17200     17210     17220     17230     17240     17250     17260     17270     17280
TTACAGTAGGCTTCGAGTTAATTGGTGGTATGATCACACATCTAGTGGACAGTATTAACGAGCTAGGCAAGACGAAGCTTATTTCTCTCCACCATGAGCAGGCAGCTGCATTTGCTGCTG
  T  V  G  F  E  L  I  G  G  M  I  T  H  L  V  D  S  I  N  E  L  G  K  T  K  L  I  S  L  H  H  E  Q  A  A  A  F  A  A  G

     17290     17300     17310     17320     17330     17340     17350     17360     17370     17380     17390     17400
GCGGTGTTGCCCGTGCCACTAATAATGAACAAGTTGGACTTGCTTTAGGTACTAGTGGGCCTGGAGCTACTAACCTAATAACTGGCATTGCAGACTGTTGGCTGGATAGTTATCCATGTA
  G  V  A  R  A  T  N  N  E  Q  V  G  L  A  L  G  T  S  G  P  G  A  T  N  L  I  T  G  I  A  D  C  W  L  D  S  Y  P  C  I

     17410     17420     17430     17440     17450     17460     17470     17480     17490     17500     17510     17520
TTTTCATTACAGGTCAAGTTAATACTTATGAGCTAAAAGATAAAAGACCAATTCGTCAGCAAGGTTTTCAAGAATTGGATATTGTAAGTTTAGTCGATTCTATTACAAAGTATAGCATCC
  F  I  T  G  Q  V  N  T  Y  E  L  K  D  K  R  P  I  R  Q  Q  G  F  Q  E  L  D  I  V  S  L  V  D  S  I  T  K  Y  S  I  Q

     17530     17540     17550     17560     17570     17580     17590     17600     17610     17620     17630     17640
AAGTTAAAACGGTTGAACAGTTGCTTATTGAAATACAGAAAGCAATTTCTATTGCGAGAAGTGGTCGACCTGGGCCCGTACTTATTGATATCCCAATGGATCTACAGCGTAAAGAGCTTG
  V  K  T  V  E  Q  L  L  I  E  I  Q  K  A  I  S  I  A  R  S  G  R  P  G  P  V  L  I  D  I  P  M  D  L  Q  R  K  E  L  D

     17650     17660     17670     17680     17690     17700     17710     17720     17730     17740     17750     17760
ACATTACTTTTGATGATATCGCTAGATTAGTGGTGCCGTCAGCAGAAGAGGAAATGAATAGTGGATTTTTTTCAGTGGATAACGCATTGAAAGAAGCACAAAAACCACTTTTTATTATTG
  I  T  F  D  D  I  A  R  L  V  V  P  S  A  E  E  E  M  N  S  G  F  F  S  V  D  N  A  L  K  E  A  Q  K  P  L  F  I  I  G

     17770     17780     17790     17800     17810     17820     17830     17840     17850     17860     17870     17880
GAGGAGGGGCATGTGCCGAAGTGCAATTTTCAGCGTGGCAAAAGAAAATTTCTTCGCTAGGTATTCCACATGTTTCGAGTCTTAAAGGAAGTGAGAGAACCTCAAACTATCCAGAATATT
  G  G  A  C  A  E  V  Q  F  S  A  W  Q  K  K  I  S  S  L  G  I  P  H  V  S  S  L  K  G  S  E  R  T  S  N  Y  P  E  Y  L

     17890     17900     17910     17920     17930     17940     17950     17960     17970     17980     17990     18000
TAGGAATGATTGGGGCGTACGGTACTCGAGCCGCTAATTATGCTGTCCAAAATGCAGATATAATCATTGTATTAGGAAGTAGATTAGATATTCGCCAAACTGGAGCTAATGTTGCTGACT
  G  M  I  G  A  Y  G  T  R  A  A  N  Y  A  V  Q  N  A  D  I  I  I  V  L  G  S  R  L  D  I  R  Q  T  G  A  N  V  A  D  F

     18010     18020     18030     18040     18050     18060     18070     18080     18090     18100     18110     18120
TCGCAAGAAATGCTAAGAAAATTATTCAAATAGATGTTGATAAAGGGCAAATTGATAACCGTATAACTACGCACCTGAATATCGTATCAAAATGCAATAGCTACTTTGAACATTTTTTAT
  A  R  N  A  K  K  I  I  Q  I  D  V  D  K  G  Q  I  D  N  R  I  T  T  H  L  N  I  V  S  K  C  N  S  Y  F  E  H  F  L  S

     18130     18140     18150     18160     18170     18180     18190     18200     18210     18220     18230     18240
CTGAAGATTATATAATTAATTGTAGCTTGTGGCGTGAAAAATTAAAAGAAACTTTCCGTAAAAAATTCATTGATGAGTACGAAGCATATAGATTTAGCCCGTTTAAAATTATGCAAACTT
  E  D  Y  I  I  N  C  S  L  W  R  E  K  L  K  E  T  F  R  K  K  F  I  D  E  Y  E  A  Y  R  F  S  P  F  K  I  M  Q  T  L

     18250     18260     18270     18280     18290     18300     18310     18320     18330     18340     18350     18360
TAAGTGAGAAATTTTCTGGTAAGATAGTCCATTATATTCCTGATGTTGGAAATCATCAAATGTGGTTAGCCCATTCTTTATTTATCGAACCACAACAGAAAATTCATCATTCAGGAGGGC
  S  E  K  F  S  G  K  I  V  H  Y  I  P  D  V  G  N  H  Q  M  W  L  A  H  S  L  F  I  E  P  Q  Q  K  I  H  H  S  G  G  L

     18370     18380     18390     18400     18410     18420     18430     18440     18450     18460     18470     18480
TTGGTGCTATGGGTTTTTCTTTACCTACTGCTATAGGTGTTAGAGTGGTAACGGGAAATTATGTTGTTTCTATCTCAGGTGATGGTGGATTTCAGCTTAATATTCAAGAATTGGATGTAA
  G  A  M  G  F  S  L  P  T  A  I  G  V  R  V  V  T  G  N  Y  V  V  S  I  S  G  D  G  G  F  Q  L  N  I  Q  E  L  D  V  I

     18490     18500     18510     18520     18530     18540     18550     18560     18570     18580     18590     18600
TAAATAGAGATAAAATACCTATTCTTATAATTATCCTTAATAATAAGTCTTTAGGAATGGTTAAGAATTTCCAAGATATGTATTTTAATGGTCGAAATAAACCAACCTATTGGGGGGGGT
  N  R  D  K  I  P  I  L  I  I  I  L  N  N  K  S  L  G  M  V  K  N  F  Q  D  M  Y  F  N  G  R  N  K  P  T  Y  W  G  G  Y
```

18610 18620 18630 18640 18650 18660 18670 18680 18690 18700 18710 18720
ATTCATGTTCATTCTCTCAAGTTGGAGAAGCATATGGTATTGAATCACATTTAATAAAAAATGATCTTGAGTTTTCAGCCTTAGTGGAAAGTTATGTTAAAAATGCACGTCCATTGCTAA
S C S F S Q V G E A Y G I E S H L I K N D L E F S A L V E S Y V K N A R P L L I

18730 18740 18750 18760 18770 18780 18790 18800 18810 18820 18830 18840
TAGAAGTATCTTTAGAGGATGTAACTGTTTGCAAGCCAAGATTAGTATATGGTAAGTCTATTGATGAGCAATACCCATTTGACGAGCAATAAATACGGCGTAAATTACATTAATAAAATT
E V S L E D V T V C K P R L V Y G K S I D E Q Y P F D E Q *

18850 18860 18870 18880 18890 18900 18910 18920 18930 18940 18950 18960
TTAGTTTTTAGAATGGAAAGTTCTAAAGAGATATTATAACAACAT**TAAGATAAGG**CCAGATAATGAGTGATTTTCAAAATACCTCTGAAAAAATCATTGAAAATTATAAGAACAGCCCAG
wbhD M S D F Q N T S E K I I E N Y K N S P E

18970 18980 18990 19000 19010 19020 19030 19040 19050 19060 19070 19080
AAATTAAAGTTAGAACTGGTAGACTTGGCGTGAATTTAACTCCAATCACTTATAATGATCTCTCCAAATCAGAAAACATTAGACTTCTCTCTGAGTGGAGAAAAGCATCGGAACATGCTT
I K V R T G R L G V N L T P I T Y N D L S K S E N I R L L S E W R K A S E H A F

19090 19100 19110 19120 19130 19140 19150 19160 19170 19180 19190 19200
TCTTGAAAGTGTTTGAAGTGACCGATACAGGAACAGCCAAATGGTTGGAAAGTGGAGTTCTGCATAACCCTTTAAGATTAATGTTTTGGGTGGAAGACTCAAAAGGAAGGAAACTAGGCC
L K V F E V T D T G T A K W L E S G V L H N P L R L M F W V E D S K G R K L G H

19210 19220 19230 19240 19250 19260 19270 19280 19290 19300 19310 19320
ATATAGGCGTATCTAGCTTTGATTCCGAAGAATTATCTTGTGAGGTAGATAATGTGATTAAATCTCCATTGTGTAATGATAAAGGCATATTCACCGATGTGCTTGAATGTTTAATTAATT
I G V S S F D S E E L S C E V D N V I K S P L C N D K G I F T D V L E C L I N L

19330 19340 19350 19360 19370 19380 19390 19400 19410 19420 19430 19440
TAGTGAAATTAGAGTTTTCTCCGAAAAAAATTAAACTGAGGGTTTTCTCTGAAAATATTAAAGCAATTTCACTTTATGATAGATTGGGCTTTAAACCTATTGATATTGTTACTTTTACAA
V K L E F S P K K I K L R V F S E N I K A I S L Y D R L G F K P I D I V T F T K

19450 19460 19470 19480 19490 19500 19510 19520 19530 19540 19550 19560
AAGTAGTTGGTAAAGACTATATCGAATGGATCGAATCAAAATCTGATATAGATCGATGTTTTTTAAT**TATGGAGCT**AAAAAATGATAAATGAGATGATACTTACGGCTGGGCCAAGTATT
V V G K D Y I E W I E S K S D I D R C F L I M E L K N D K *
wbhE M I N E M I L T A G P S I

19570 19580 19590 19600 19610 19620 19630 19640 19650 19660 19670 19680
ACCGAAAAAGAGATCAGTTATGTTACCGATGCTGTAAAAAATGGATGGAATAACAATTGGAATAACTACCTACTAAAGTTTGAAAAATCATTGGCGGATTATGTTGGAGTTAAACATAGT
T E K E I S Y V T D A V K N G W N N N W N N Y L L K F E K S L A D Y V G V K H S

19690 19700 19710 19720 19730 19740 19750 19760 19770 19780 19790 19800
TTATCGACATCTAGTGCAACAGGTGCATTACATTTATCTATGCTTGCATGTGGCATTGGACCTGGGGATGAGGTTATTGTTCCGGAGATAAGTTGGGTAGCCTCGGCTTCCGCAGTGGCT
L S T S S A T G A L H L S M L A C G I G P G D E V I V P E I S W V A S A S A V A

19810 19820 19830 19840 19850 19860 19870 19880 19890 19900 19910 19920
TATGTCGGTGCAACTCCTGTTTTTTGTGATATAGATCCAGTAAGTTGGTGTTTGGATATTGAATCAGCTGCAAGATTATTAACGCCTAAAACAAAGGCTATCCTTCCGGTTCATATATAT
Y V G A T P V F C D I D P V S W C L D I E S A A R L L T P K T K A I L P V H I Y

19930 19940 19950 19960 19970 19980 19990 20000 20010 20020 20030 20040
GGCCATCCAGCCAATATGCCTGCAATCATGGAATTTGCGAGAGCTAATAATATTTTGATAATCGAAGATGCGGCTCCTTCTATTGGTGCTGAGGTTGATGGTAAGAAAACAGGTAGCTTT
G H P A N M P A I M E F A R A N N I L I I E D A A P S I G A E V D G K K T G S F

```
     20050     20060     20070     20080     20090     20100     20110     20120     20130     20140     20150     20160
GGTGATGCTGCTGCTTTCAGTTTTCAAGGAGCTAAAATACTATCTACAGGCGAGGGGGCATGTTCGTTTCTAATAACGATGAAATATTTAATAGAGTTAAGTCATTGAATGATCATGGG
G  D  A  A  A  F  S  F  Q  G  A  K  I  L  S  T  G  E  G  G  M  F  V  S  N  N  D  E  I  F  N  R  V  K  S  L  N  D  H  G

     20170     20180     20190     20200     20210     20220     20230     20240     20250     20260     20270     20280
CGTGATCCTAGTCAACCATTTGCTTCTGTTGAGGTTGGTTATAAGTATAAAATGTCTAATTTACAAGCTGCAATGGGCTTGGCCCAGATAGAGCGTGTAGAGGAGCTTGTTAATAAAAAA
R  D  P  S  Q  P  F  A  S  V  E  V  G  Y  K  Y  K  M  S  N  L  Q  A  A  M  G  L  A  Q  I  E  R  V  E  E  L  V  N  K  K

     20290     20300     20310     20320     20330     20340     20350     20360     20370     20380     20390     20400
CGTGAAATTAATTCCATATACCAAGAGCTTCTAAAAGATTGCACTGCAGTCAAGGTAACGACCGAGCTGCCTGATTGTAAAAGTATTCACTGGATGACATCAGTGGAGCTACTTGGTTTT
R  E  I  N  S  I  Y  Q  E  L  L  K  D  C  T  A  V  K  V  T  T  E  L  P  D  C  K  S  I  H  W  M  T  S  V  E  L  L  G  F

     20410     20420     20430     20440     20450     20460     20470     20480     20490     20500     20510     20520
GATTATGATAAACGCCAGAGATTTATGGGTAAACTTAGAGAAAACTTAGTTGATTCACGTCCAGTATTTTCACCGTTGAGTTCTTTACCTATGTTTGAACCACGAGTAAAAAATCCGGTT
D  Y  D  K  R  Q  R  F  M  G  K  L  R  E  N  L  V  D  S  R  P  V  F  S  P  L  S  S  L  P  M  F  E  P  R  V  K  N  P  V

     20530     20540     20550     20560     20570     20580     20590     20600     20610     20620     20630     20640
GCTCTACGTATTGGTCAAAGTGCAATCAACTTACCGAGTGGGCACAATTTAATATTAGAACAACTTGAACATGTTGCAACAACGATTAAAAAAACTCGTATAGCAATCTAGATTACGCCT
A  L  R  I  G  Q  S  A  I  N  L  P  S  G  H  N  L  I  L  E  Q  L  E  H  V  A  T  T  I  K  K  T  R  I  A  I  *

     20650     20660     20670     20680     20690     20700     20710     20720     20730     20740     20750     20760
GGTTAACTGCTAGGCGTTAATTTTCATACATTCTATCTAATAGTGAATATGAATATTATATCAAGGTTACACAATCCCATTATTTTTAAATATTTATTTGCAGGTGTAATTAATACTATT
                                          wbhF  M  N  I  I  S  R  L  H  N  P  I  I  F  K  Y  L  F  A  G  V  I  N  T  I

     20770     20780     20790     20800     20810     20820     20830     20840     20850     20860     20870     20880
TTTGGGTATTCATTATTTTCTTTTTTAATATTTATAAATATTGATCATAAAATATCTATAACTATATCTACAATAATAGGTGTGGTTTTTAATTACTTCAATTTTGGTAGAAACGTTTTT
F  G  Y  S  L  F  S  F  L  I  F  I  N  I  D  H  K  I  S  I  T  I  S  T  I  I  G  V  V  F  N  Y  F  N  F  G  R  N  V  F

     20890     20900     20910     20920     20930     20940     20950     20960     20970     20980     20990     21000
AAAATAGAAAGTAACCAGTTTACATTTTATAGGTTTTTTTTTGTATATATACTATCATATTTAATTAATTTGGTTTTTATAACTTTGCTGGTGGATTGTCTTAAATATAATCCATATTTA
K  I  E  S  N  Q  F  T  F  Y  R  F  F  F  V  Y  I  L  S  Y  L  I  N  L  V  F  I  T  L  L  V  D  C  L  K  Y  N  P  Y  L

     21010     21020     21030     21040     21050     21060     21070     21080     21090     21100     21110     21120
TCTCAAGGTGTATGTATTTTTTCATTGATTTTAATTAATTGGTTTTTATTTAATTTTTGGGTTTTTAAAAAATGAAAAAAAAGAAAATAAGCATTGTGACACCATGCTTTAATGAAGAAG
S  Q  G  V  C  I  F  S  L  I  L  I  N  W  F  L  F  N  F  W  V  F  K  K  *
                                                      wbhG  M  K  K  K  K  I  S  I  V  T  P  C  F  N  E  E  D

     21130     21140     21150     21160     21170     21180     21190     21200     21210     21220     21230     21240
ATAATGTAGAGTTATTATATGAGAAGGTTAAAGAGGAATTTGAGAAACTAGATGCTTATTTATATGAACATATTTTTATTGATAATAACTCTGTAGATAAAACGGTTGAGAAATTAAGAA
  N  V  E  L  L  Y  E  K  V  K  E  E  F  E  K  L  D  A  Y  L  Y  E  H  I  F  I  D  N  N  S  V  D  K  T  V  E  K  L  R  S

     21250     21260     21270     21280     21290     21300     21310     21320     21330     21340     21350     21360
GTATCGCCAAAAGGGATTGTAATGTTAAAGTAATATTAAATAGCCGAAACTTTGGGCCTGTGCGTTCTCCTCACTATGGTCTTTTACAAGGCTCAGGAGATGCTACAATGCTAGTTGTTG
  I  A  K  R  D  C  N  V  K  V  I  L  N  S  R  N  F  G  P  V  R  S  P  H  Y  G  L  L  Q  G  S  G  D  A  T  M  L  V  V  A

     21370     21380     21390     21400     21410     21420     21430     21440     21450     21460     21470     21480
CAGATCTTCAAGATCCACCAGAGCTCATACCAGAGTTTATTGCAAAATGGGAAGATGGAAATGATATCGTAATAGGAGTAAAGAGTGAAAGTGATGAATCTCCTGCCATGTATATGATAA
  D  L  Q  D  P  P  E  L  I  P  E  F  I  A  K  W  E  D  G  N  D  I  V  I  G  V  K  S  E  S  D  E  S  P  A  M  Y  M  I  R
```

21490 21500 21510 21520 21530 21540 21550 21560 21570 21580 21590 21600
GAAAAGTATATTATAATTTTAGTAACAAATTATCTGAGTCAAAGCTAGTAAAAAATTATTACGGTTTTGGTCTCTATGATAAAAAGATAATCGAAATATTAGATTCAATTGATGATCCAT
K V Y Y N F S N K L S E S K L V K N Y Y G F G L Y D K K I I E I L D S I D D P F

21610 21620 21630 21640 21650 21660 21670 21680 21690 21700 21710 21720
TTCCTTATACTAGGGGTATGATTATGGATCTTGGATTTAAATTGGAGAAAATATACTACCGTCAACCAATACGTCGAAGAGGTGTTACTTCTACTAATTTTTTATCGCTTTATGATATTG
P Y T R G M I M D L G F K L E K I Y Y R Q P I R R R G V T S T N F L S L Y D I A

21730 21740 21750 21760 21770 21780 21790 21800 21810 21820 21830 21840
CTATGCTTGGTATTTGTAGTCACTCAAAGATTCCTCTACGCATTGCGACCATATCAGGATTTATACTCAGTCTAATAAGTTTTACTCTTGCTATCATATTTCTAGCCCTCAAAATTATTT
M L G I C S H S K I P L R I A T I S G F I L S L I S F T L A I I F L A L K I I Y

21850 21860 21870 21880 21890 21900 21910 21920 21930 21940 21950 21960
ATTGGGATGATTTCCCTATGGGAACTGCACCAATCATTATTGGTATGTTTTTCTTAGGTTCTGTCCAAATATTTTTTATAGGGCTGGTTGGTGAGTATGTTGGTCATCTAGTAGCTAAAC
W D D F P M G T A P I I I G M F F L G S V Q I F F I G L V G E Y V G H L V A K L

21970 21980 21990 22000 22010 22020 22030 22040 22050 22060 22070 22080
TTTCCAAGTTTCCACTTGTTATAGAAAAAGAAAGGATAAATTTTGACTAGCTTGTAATTACAACTTATTATTTAATAACAA**CAATGATGTAAT**ATTCTTACATGCGAGATAGTTATATAT
S K F P L V I E K E R I N F D * ***wbhH*** M R D S Y I F

22090 22100 22110 22120 22130 22140 22150 22160 22170 22180 22190 22200
TTATTATGAATAAACTGAAAATTGTGTCAATTTTAAAAGTTTTCTACTATATGTTGCCTAAATTTATTAGAGTTCAGTACTCTGATAGTGTTACAGTTCTCCGTAATATGGCTTTAGGTG
I M N K L K I V S I L K V F Y Y M L P K F I R V Q Y S D S V T V L R N M A L G D

22210 22220 22230 22240 22250 22260 22270 22280 22290 22300 22310 22320
ATAAAAGCTTACATAATGATTTTTTTGTGGAAAATAAAATTAAACTAGTAAAAGAACTAGAATCATTGGGCGCTACATTTGATATTATTACCCATATTATAGGGCGTAAGAATATCATAG
K S L H N D F F V E N K I K L V K E L E S L G A T F D I I T H I I G R K N I I G

22330 22340 22350 22360 22370 22380 22390 22400 22410 22420 22430 22440
GTGCTCCTGCTATTTACATAGACAATGTTAGGTGCTTTCGACCTTCTCTCGAAGGAGGCTTGCTTGATGTGAAAGTTATTCATTTATTTCATGTCACTCTGATTGGATCTACAGATGGAG
A P A I Y I D N V R C F R P S L E G G L L D V K V I H L F H V T L I G S T D G V

22450 22460 22470 22480 22490 22500 22510 22520 22530 22540 22550 22560
TTATCTATAAAGATAAAATGTATCATCAAGAACTATTGTCAATGCGTCATACTAATGATTTAAAGCAACCAGATATTTTTGTTAGTCAAAATTTTGTCAAAATTCACAACTATGATTATC
I Y K D K M Y H Q E L L S M R H T N D L K Q P D I F V S Q N F V K I H N Y D Y L

22570 22580 22590 22600 22610 22620 22630 22640 22650 22660 22670 22680
TTGTCAGAAGTAGAAAAAATGTATTAAATATTGAAGGGCCTGCCATCAGCTTACTAAAAGAGCACTCTTCAAATTATTACCATTGTGTTACTGAAATTCTCCCCAAACTTCAACAAATAC
V R S R K N V L N I E G P A I S L L K E H S S N Y Y H C V T E I L P K L Q Q I L

22690 22700 22710 22720 22730 22740 22750 22760 22770 22780 22790 22800
TTAGTTGTGTACCTAGTAATAAGGATTCATTCTCAATACTTATTGATGACTGTATGCCTATTCAAGTTGTTAACATGATTAACGTTATACTGAGTGCCTATACTTGCTTTCAATATGAAC
S C V P S N K D S F S I L I D D C M P I Q V V N M I N V I L S A Y T C F Q Y E L

22810 22820 22830 22840 22850 22860 22870 22880 22890 22900 22910 22920
TTATAAAAGTTGAAAAAGGGCAGAAGGTAAATTGTAACGAAGTGATCTACTGTACTCCATTGTGGTTATCTCTTGATAATACACAGTACTTACCTGATCCCAAGAACGAGTTTTTTGTTT
I K V E K G Q K V N C N E V I Y C T P L W L S L D N T Q Y L P D P K N E F F V S

```
     22930     22940     22950     22960     22970     22980     22990     23000     23010     23020     23030     23040
CTGTAGATAGTTTAAAAAAACTTAAAGGTCATATATCGAGTGTGCTTTTTGCTCCTACAGTATCTTCTAAGTATAGGAAGGTTTATTTACAGAGACAGAATAATAGATTAAGAAAAATTA
   V  D  S  L  K  K  L  K  G  H  I  S  S  V  L  F  A  P  T  V  S  S  K  Y  R  K  V  Y  L  Q  R  Q  N  N  R  L  R  K  I  S

     23050     23060     23070     23080     23090     23100     23110     23120     23130     23140     23150     23160
GTAATATTCTTGAAGTTGAGCGTGTGTTGTACAAGCATGGTTTTGAGTTTGTTAACACAGGTACTCTTAGTTTTCAAGAGCAATATGAGTTATTTTCTGATACAGATGTTGTTATTGGAG
   N  I  L  E  V  E  R  V  L  Y  K  H  G  F  E  F  V  N  T  G  T  L  S  F  Q  E  Q  Y  E  L  F  S  D  T  D  V  V  I  G  V

     23170     23180     23190     23200     23210     23220     23230     23240     23250     23260     23270     23280
TGTCAGGCGCATCTTTTACTAATATACTGTTTATGCAAAGAAATAGTAAAGCAATCTTATTATCTCCATCGGCGCAATGTACTAATTATTATATTTTCCAGCCATTGGCAGACGTGTCTG
   S  G  A  S  F  T  N  I  L  F  M  Q  R  N  S  K  A  I  L  L  S  P  S  A  Q  C  T  N  Y  Y  I  F  Q  P  L  A  D  V  S  E

     23290     23300     23310     23320     23330     23340     23350     23360     23370     23380     23390     23400
AAGTTGAGCTGGTGCACTTACTTTCGAAACCTGATGATGACTCTAATTCATTACATGGAGATGCTAGCGTTAACGTCGAAGAGTTGGAATTATTTCTTAGTGAAATGTCGTGTGATATCT
   V  E  L  V  H  L  L  S  K  P  D  D  D  S  N  S  L  H  G  D  A  S  V  N  V  E  E  L  E  L  F  L  S  E  M  S  C  D  I  Y

     23410     23420     23430     23440     23450     23460     23470     23480     23490     23500     23510     23520
                                              -35               -10
ATAGATAGAGGCTTTATTAAGTAATAGCGATTTTACTAAGATTTAACTATTAATTTTAGGATGAAAATATAATATAATATAATTTCGCTCTTTAGCGATATAATTACGCTTCAGTGACCT
   R  *

     23530     23540     23550     23560     23570     23580     23590     23600     23610     23620     23630     23640
TTACTGTTTGATGAATGGCAGTTGGGGTGGATAGCATTTTTGCAACAAAGCCGTTGGATAGGATATAACTATTTTTTATTAACTATTAATTAGGTAAGGCAATGGACATTAAATTTATTC
                                                                                             wbhI     M  D  I  K  F  I  Q

     23650     23660     23670     23680     23690     23700     23710     23720     23730     23740     23750     23760
AGTTTCAAGCTCATGGTGATGATAGGGGCTCATTAGTCTCCTTAGAAGATGATAAAAATATCCCCTTTACTATAAAGAGAGTATATTATTTATACAATACAAAGTCTGGTGTACGCCGCG
   F  Q  A  H  G  D  D  R  G  S  L  V  S  L  E  D  D  K  N  I  P  F  T  I  K  R  V  Y  Y  L  Y  N  T  K  S  G  V  R  R  G

     23770     23780     23790     23800     23810     23820     23830     23840     23850     23860     23870     23880
GTTTTCACGCTCATAAAGAACTAAAACAGTTAGCCGTAGTTTTAAAAGGATCATGTAGATTCTTGTTAGATGACGGTAGCGAGAAGATTGAAGTTTTATTGGATAACCCTGAACAAGGCT
   F  H  A  H  K  E  L  K  Q  L  A  V  V  L  K  G  S  C  R  F  L  L  D  D  G  S  E  K  I  E  V  L  L  D  N  P  E  Q  G  L

     23890     23900     23910     23920     23930     23940     23950     23960     23970     23980     23990     24000
TATTCATCGACTCGTTTGTTTGGAGAGAAATGTTCGATTTTTCTGAAGACTGTGTACTGCTGGTATTGGCAGATAAATTCTATGATGAAGCTGACTATATCCGTGATTATGATGCTTTCC
   F  I  D  S  F  V  W  R  E  M  F  D  F  S  E  D  C  V  L  L  V  L  A  D  K  F  Y  D  E  A  D  Y  I  R  D  Y  D  A  F  L

     24010     24020     24030     24040     24050     24060     24070     24080     24090     24100     24110     24120
TGGCAGAAATTAAGGCATAAAGAAAGACAGGTTTTAAATGACATTTATTCATGAGCTTGCCGATGTTCAATCTAGTAATATTGGGATAAATACAAAAATTTGGCAGTTTAGCGTAGTTTT
   A  E  I  K  A  *            wbhJ   M  T  F  I  H  E  L  A  D  V  Q  S  S  N  I  G  I  N  T  K  I  W  Q  F  S  V  V  L

     24130     24140     24150     24160     24170     24180     24190     24200     24210     24220     24230     24240
ACCCAATGCTATCATTGGTAAAAATTGTAACATCTGTTCGCATACTTTTATTGAGAACGACGTCACTATAGGTAATAATGTTACTATCAAGTGTGGTGTACAAATTTGGGATGGTATCTT
  P  N  A  I  I  G  K  N  C  N  I  C  S  H  T  F  I  E  N  D  V  T  I  G  N  N  V  T  I  K  C  G  V  Q  I  W  D  G  I  L

     24250     24260     24270     24280     24290     24300     24310     24320     24330     24340     24350     24360
AATTGGTAACAATGTTTTTATTGGGCCCAATGCTACTTTTACAAATGATATGTACCCTCGTTCTAAGCAATACCCCGATGAATTTATGAAGACGGTTGTGTGTGATAATGCTTCTATTGG
  I  G  N  N  V  F  I  G  P  N  A  T  F  T  N  D  M  Y  P  R  S  K  Q  Y  P  D  E  F  M  K  T  V  V  C  D  N  A  S  I  G
```

```
     24370     24380     24390     24400     24410     24420     24430     24440     24450     24460     24470     24480
TGCGAATACTACAATTCTTCCAGGGGTGACTATTGGAGAGGGTGCATTAGTTGGTGCTGGCTCTGTGGTTACAAAAGACGTAAAACCTTTTACTATTGTTGCGGGAAATCCTGCTAGAGA
 A  N  T  T  I  L  P  G  V  T  I  G  E  G  A  L  V  G  A  G  S  V  V  T  K  D  V  K  P  F  T  I  V  A  G  N  P  A  R  E

     24490     24500     24510     24520     24530     24540     24550     24560     24570     24580     24590     24600
AATTGAAAAATGATATCCTTTTTAGACTTAAAGAAACTTAATGCTCAATACGAACAAGAGCTAAAAGATGCATGCACTCGAGTTATAGATTCTGGTTGGTATATACTTGGTAATGAAGTT
 I  E  K  *
    wbhK  M  I  S  F  L  D  L  K  K  L  N  A  Q  Y  E  Q  E  L  K  D  A  C  T  R  V  I  D  S  G  W  Y  I  L  G  N  E  V

     24610     24620     24630     24640     24650     24660     24670     24680     24690     24700     24710     24720
TCAGAATTTGAGAAAGAGTTTGCTGCATATTGTAATGTTGAACACTGTTTAGGTGTTGCTAACGGCTTAGATGCTTTGATACTTATACTACGCGCTTACATTGAACTAGGAGTTATGTCT
S  E  F  E  K  E  F  A  A  Y  C  N  V  E  H  C  L  G  V  A  N  G  L  D  A  L  I  L  I  L  R  A  Y  I  E  L  G  V  M  S

     24730     24740     24750     24760     24770     24780     24790     24800     24810     24820     24830     24840
AAAGGTGACGAGGTTATTGTTCCGTCAAACACCTATATTGCATCTATACTCGCAATATCTGAAGCGGGGTTAGTTCCAGTACTTGTTGAGCCTTGCGAAGTAACTTTCAATCTAGACCCT
K  G  D  E  V  I  V  P  S  N  T  Y  I  A  S  I  L  A  I  S  E  A  G  L  V  P  V  L  V  E  P  C  E  V  T  F  N  L  D  P

     24850     24860     24870     24880     24890     24900     24910     24920     24930     24940     24950     24960
AATTTAATCGAACTAGCAGTTACTACTCGAACAAAAGCAATTCTAACCGTTCATTTATACGGACAGGTTTCTGGTATGGATCAAATCCAAAAGATTGCTCGAAAACATAATTTAAAGGTA
N  L  I  E  L  A  V  T  T  R  T  K  A  I  L  T  V  H  L  Y  G  Q  V  S  G  M  D  Q  I  Q  K  I  A  R  K  H  N  L  K  V

     24970     24980     24990     25000     25010     25020     25030     25040     25050     25060     25070     25080
ATAGAAGATTGCGCACAAGCTCACGGAGCGTTGTATGATTCGAAAGAAGGAATCAAGAAGGTTGGTTCCATTGGTGATGCTGCTGGTTTTAGTTTTTATCCAGGAAAAAATCTTGGAGCA
I  E  D  C  A  Q  A  H  G  A  L  Y  D  S  K  E  G  I  K  K  V  G  S  I  G  D  A  A  G  F  S  F  Y  P  G  K  N  L  G  A

     25090     25100     25110     25120     25130     25140     25150     25160     25170     25180     25190     25200
CTTGGAGATGCTGGTGCGGTTACAACTAATGACCCAGAGCTTGCCAGTACTATCTCGGCACTCCGTAACTACGGTTCACATGAAAAATACCGAAATATATTTAAGGGGTTGAATAGTCGA
L  G  D  A  G  A  V  T  T  N  D  P  E  L  A  S  T  I  S  A  L  R  N  Y  G  S  H  E  K  Y  R  N  I  F  K  G  L  N  S  R

     25210     25220     25230     25240     25250     25260     25270     25280     25290     25300     25310     25320
TTAGATGAAATACAAGCGGCGATGCTTAGAGTCAAACTACGCTATTTAGATGACGAAATATCTCTACGACGCAAAGTTGCAAGCCGTTACTTAAATGAAGTAAGAAATCCTCTAATCAAG
L  D  E  I  Q  A  A  M  L  R  V  K  L  R  Y  L  D  D  E  I  S  L  R  R  K  V  A  S  R  Y  L  N  E  V  R  N  P  L  I  K

     25330     25340     25350     25360     25370     25380     25390     25400     25410     25420     25430     25440
GTACCTTTAGTCGAAGATGATAATGCGCATGTATGGCACCTTTTTGTGGTAGTGGTCGAGGATAGAGATAAATTAGTCGAACATCTGAGTGAAAACCAAATTCAGTCATTAATACATTAC
V  P  L  V  E  D  D  N  A  H  V  W  H  L  F  V  V  V  V  E  D  R  D  K  L  V  E  H  L  S  E  N  Q  I  Q  S  L  I  H  Y

     25450     25460     25470     25480     25490     25500     25510     25520     25530     25540     25550     25560
CCAATTCCACCACACAAGCAAGAGGCTTATATGGAGTGGAATAACAGTGCCTTTCCGTTATCAGAGAAAATGCATAAGCAAGTTTTAAGTCTTCCACTAAGCTCTGTTTTGGATAAAAGT
P  I  P  P  H  K  Q  E  A  Y  M  E  W  N  N  S  A  F  P  L  S  E  K  M  H  K  Q  V  L  S  L  P  L  S  S  V  L  D  K  S

     25570     25580     25590     25600     25610     25620     25630     25640     25650     25660     25670     25680
                                                                      -35           -10
GAAATTGACAAAGTTATAGATGTATTGAACTCTTTTCTTGGTTAATTAAGCCGTCATACATAGTTTTTCTGGAGGATGACTTTTTGGAAAAACTGTTTACTATTAAAATAATTGAGCTTC
E  I  D  K  V  I  D  V  L  N  S  F  L  G  *

     25690     25700     25710     25720     25730     25740     25750     25760     25770     25780     25790     25800
CATTACCAGGTTTACTCCTATGCACTTGTCTGACTCCAAGCCTCCTTTAGTAAGTTTATGTATACTAACTTTTAATCATGAGAAATATATAACTAAAGCAATAAATAGTTGTTTGGCGCA
             wbhL  M  H  L  S  D  S  K  P  P  L  V  S  L  C  I  L  T  F  N  H  E  K  Y  I  T  K  A  I  N  S  C  L  A  Q
```

```
     25810     25820     25830     25840     25850     25860     25870     25880     25890     25900     25910     25920
ATCTTATAGTAATATAGAGATAATTATTGTAGACAATAATTCAAGTGATGGAACTGTAAATAAGATTAGATCTGATTTTAAAAATGAGCTAGAAGTTGGGGAAATAAAGCTTTTTGATTT
 S  Y  S  N  I  E  I  I  I  V  D  N  N  S  S  D  G  T  V  N  K  I  R  S  D  F  K  N  E  L  E  V  G  E  I  K  L  F  D  L

     25930     25940     25950     25960     25970     25980     25990     26000     26010     26020     26030     26040
GGAACACAATACTTATCCTTCACATGGTTTTAACTATGCACTGAAAAAGTCTCAAGGAGAATATGTCTCTCTTTTTTCTGGTGATGATACGCTTTGTTTAAACAAAGTAGAACGGCAGAT
 E  H  N  T  Y  P  S  H  G  F  N  Y  A  L  K  K  S  Q  G  E  Y  V  S  L  F  S  G  D  D  T  L  C  L  N  K  V  E  R  Q  I

     26050     26060     26070     26080     26090     26100     26110     26120     26130     26140     26150     26160
TGGTATAATGGTTAAAGAAGGCTTATCTAATTTATTTACATGGGTAAATATAATCAATGATAAAGATGAAATTATAAAGTGTGATTACTTAGAAAGTATTTTTAATCGCAATTATAATAG
 G  I  M  V  K  E  G  L  S  N  L  F  T  W  V  N  I  I  N  D  K  D  E  I  I  K  C  D  Y  L  E  S  I  F  N  R  N  Y  N  S

     26170     26180     26190     26200     26210     26220     26230     26240     26250     26260     26270     26280
CCAACAAATAAAAGAGCATTTTATTCATAGTGGGAATATGTTATCCGCACTGTCCGTCATGCTATCTAGAGATGTATTTGATAGGTATGGTCATTTCGACGAGCGTCTTGTTCAGCTACA
 Q  Q  I  K  E  H  F  I  H  S  G  N  M  L  S  A  L  S  V  M  L  S  R  D  V  F  D  R  Y  G  H  F  D  E  R  L  V  Q  L  Q

     26290     26300     26310     26320     26330     26340     26350     26360     26370     26380     26390     26400
GGATTTTGATTTTTGGCTGCGTATGGCCTCTAATGATGATCTTAATTTATTGACCGAAAAACTATCAAACTATAGATTGAGAGATGATGGCGGAAATTTAAGTCTAGCTAATCATAAATC
 D  F  D  F  W  L  R  M  A  S  N  D  D  L  N  L  L  T  E  K  L  S  N  Y  R  L  R  D  D  G  G  N  L  S  L  A  N  H  K  S

     26410     26420     26430     26440     26450     26460     26470     26480     26490     26500     26510     26520
GCGACAGTTAAGAACTGATTTTGAAGAAGTTTATGTATATAGACATTTGTTAAATTTTGATTTAAAGACAATTCAAAGTGTTGTTGGTATGCTGAATAAAGACCAATCTATCGCTATGGC
 R  Q  L  R  T  D  F  E  E  V  Y  V  Y  R  H  L  L  N  F  D  L  K  T  I  Q  S  V  V  G  M  L  N  K  D  Q  S  I  A  M  A

     26530     26540     26550     26560     26570     26580     26590     26600     26610     26620     26630     26640
TCTGCATGGCTATTATCATAATGAGAATAAAATGAAGTTGGCCAAAGGTTTTTTGTTGTCCATATATGAAGAATTAGGGACAAATATTGTTTTCCCTTCATCTCACTATAGCTACTTCTT
 L  H  G  Y  Y  H  N  E  N  K  M  K  L  A  K  G  F  L  L  S  I  Y  E  E  L  G  T  N  I  V  F  P  S  S  H  Y  S  Y  F  F

     26650     26660     26670     26680     26690     26700     26710     26720     26730     26740     26750     26760
TGACATATATTCTAAATGTGAATTTTTCAACTCAGATGAGAATGATGAAGTATCAAAGCTGAAAGAGAAAATACACGTTTATGAAAATTCTAGGGCGATTAGATTTACTAGCTTAATCGT
 D  I  Y  S  K  C  E  F  F  N  S  D  E  N  D  E  V  S  K  L  K  E  K  I  H  V  Y  E  N  S  R  A  I  R  F  T  S  L  I  V

     26770     26780     26790     26800     26810     26820     26830     26840     26850     26860     26870     26880
TTCTTGGCTTGGTAGGTTAAAAAAATAAGTGAGAAAAATGAAGTATCCACGAAGCTTATCATGGGAAAAAATTTCTAGCAGTACTATCTCTATTAGAGAGTTTAATCGTTTTGAGAATAA
 S  W  L  G  R  L  K  K  *    wbhM   M  K  Y  P  R  S  L  S  W  E  K  I  S  S  S  T  I  S  I  R  E  F  N  R  F  E  N  K

     26890     26900     26910     26920     26930     26940     26950     26960     26970     26980     26990     27000
GAAAAAAGTAGCTATATTTTGTGGTTATGTAAGGGATAAATATGAATTTAACTATCATGATAGAACTTATCAATGGTTAAGAAACAATGACTATTACGTCATCTTAGTTATGCCTAAAAG
 K  K  V  A  I  F  C  G  Y  V  R  D  K  Y  E  F  N  Y  H  D  R  T  Y  Q  W  L  R  N  N  D  Y  Y  V  I  L  V  M  P  K  S

     27010     27020     27030     27040     27050     27060     27070     27080     27090     27100     27110     27120
TGGAATTCTACTAGAATCTGATTTATGCAAAGCTTGTGATGTTTTTATTGAAAGAGAAAACTTTGGCTACGATTTTGGATCTTATGCATGCGGATTACAATATGTTAACTTAATCGAGGG
 G  I  L  L  E  S  D  L  C  K  A  C  D  V  F  I  E  R  E  N  F  G  Y  D  F  G  S  Y  A  C  G  L  Q  Y  V  N  L  I  E  G

     27130     27140     27150     27160     27170     27180     27190     27200     27210     27220     27230     27240
CTCAGAGAGGATTGATAGGTTACTTTTCGTGAACGATAGTTTTATTGGGCCTTTCGGATACTGCAATTTAATAGAAGATTCTAGCGAATTTTGGGGAAATACAGATAGCAACCAAGTAAA
 S  E  R  I  D  R  L  L  F  V  N  D  S  F  I  G  P  F  G  Y  C  N  L  I  E  D  S  S  E  F  W  G  N  T  D  S  N  Q  V  K
```

```
     27250     27260     27270     27280     27290     27300     27310     27320     27330     27340     27350    273600
ATATCACTATCAGAGTTACCTGTTTGGATTTAATCTCGAAAAAGTAAACTTAGATATTATAAATAACTTTTTTTTTAGCCGAGGAGATATATATACAGATGATAAATCTCTAGTTATTGA
 Y  H  Y  Q  S  Y  L  F  G  F  N  L  E  K  V  N  L  D  I  I  N  N  F  F  F  S  R  G  D  I  Y  T  D  D  K  S  L  V  I  E

     27370     27380     27390     27400     27410     27420     27430     27440     27450     27460     27470     27480
AAATTTTGAGCTTTCATTATATGAATATTTTAATGGGAAAGGCCTAAGATGTAGTGTGTTACATCCCATAAGTGTTTTAAAATCGGATTTTATTAAACAGACTTTTCACTTTATTTCATA
  N  F  E  L  S  L  Y  E  Y  F  N  G  K  G  L  R  C  S  V  L  H  P  I  S  V  L  K  S  D  F  I  K  Q  T  F  H  F  I  S  Y

     27490     27500     27510     27520     27530     27540     27550     27560     27570     27580     27590     27600
TCCATATCTAACATCCAAAATTTTCTTTTATATAATGGTGATTGCACGCGACGTGAATCCAACTCATCAACTATGGTTGCAACTATTCAAAAGAGGCTTTCCTTTTATAAAGAAAGAGCT
  P  Y  L  T  S  K  I  F  F  Y  I  M  V  I  A  R  D  V  N  P  T  H  Q  L  W  L  Q  L  F  K  R  G  F  P  F  I  K  K  E  L

     27610     27620     27630     27640     27650     27660     27670     27680     27690     27700     27710     27720
TTTGAGAGATAATCCGACAGGTTACCCAGAGCTTTACAAAAAAGTAGAGGAAGTTATGGGTAGCAATGATTTTAATGGAGAGTATAAACAAATCTTCAAAAATCATTTATGAGCAATAGT
  L  R  D  N  P  T  G  Y  P  E  L  Y  K  K  V  E  E  V  M  G  S  N  D  F  N  G  E  Y  K  Q  I  F  K  N  H  L  *

     27730     27740     27750     27760     27770     27780     27790     27800     27810     27820     27830     27840
          -35                                     -10
CTAGAGTAATTTGATTTGGTATTCTATTTTCATATTTCACGTAGTTCGACTATTATATTTTTAGTGCTCTGCCCACAGATTTTTTATATAAATATTATAGTGACATTGTGAGCATTCGAT
                                                                                  wbhN    V  S  I  R  Y
     27850     27860     27870     27880     27890     27900     27910     27920     27930     27940     27950     27960
ATGTATCTTCCCCCCCCCCCAAAGTATAGGAGTATTTTTTGTCGATTTCATCCGTAATTATCACTTTTCAACCCGATCTATCGAATGTCTCCCAGTTACTATTCGCTTGTATATCCTTTG
M  Y  L  P  P  P  P  K  Y  R  S  I  F  C  R  F  H  P  *
  V  S  P  P  P  P  K  V  *  E  Y  F  L  S  I  S  S  V  I  I  T  F  Q  P  D  L  S  N  V  S  E  L  L  F  A  C  I  S  F  G
     27970     27980     27990     28000     28010     28020     28030     28040     28050     28060     28070     28080
GAAACAAGGCTGTTGTTATTGATAATGGTAGTAACAATGCGGAAGAACTACAGGATATTTGTCGGTTATTTGAGCATGTTAAGCTAATCCGTCTTGATGAAAATGTGGGTATCGCGTCGG
  N  K  A  V  V  I  D  N  G  S  N  N  A  E  E  L  Q  D  I  C  R  L  F  E  H  V  K  L  I  R  L  D  E  N  V  G  I  A  S  A

     28090     28100     28110     28120     28130     28140     28150     28160     28170     28180     28190     28200
CACAAAACATAGCAATTAGTAATTTGAATGGTAATGAAGACGATATTATTGTTTTTTTGATCAGGACTCATCTATTGACAATGGGTATTTGAGCAAGGTCGAACTTGCGTATAACCGGT
  Q  N  I  A  I  S  N  L  N  G  N  E  D  D  I  I  V  F  F  D  Q  D  S  S  I  D  N  G  Y  L  S  K  V  E  L  A  Y  N  R  L

     28210     28220     28230     28240     28250     28260     28270     28280     28290     28300     28310     28320
TGGAGAGTGATTTTGGTCGAGGCATTGTTTTAGGGCCAAGGTTTTATAATCGAGTTTCCAAGTTCGAGTATCCAGTAATAAAGTTTAATATCTTTGGTCTTAGGAGCAGAATATATCCTA
  E  S  D  F  G  R  G  I  V  L  G  P  R  F  Y  N  R  V  S  K  F  E  Y  P  V  I  K  F  N  I  F  G  L  R  S  R  I  Y  P  S

     28330     28340     28350     28360     28370     28380     28390     28400     28410     28420     28430     28440
GTGAATCTCGCTACCCGATAGAAGCTTCTTGTATTATCTCTTCCGGAATGGCAGTTAGAAAAAATATTCTTGATTCCGTTGGTGTGATGGATGACTCGTTGTTTATCGACTATGTAGATA
  E  S  R  Y  P  I  E  A  S  C  I  I  S  S  G  M  A  V  R  K  N  I  L  D  S  V  G  V  M  D  D  S  L  F  I  D  Y  V  D  T

     28450     28460     28470     28480     28490     28500     28510     28520     28530     28540     28550     28560
CAGAGTGGAGTTTAAGAGCTAGATATTTGGGGAATCTGATACTCGTCGATCCTCAGTTAGTTATGGGGCATGAGATTGGCACTGATAATCTCAAGCTTTTTAAATGGAGAGTTCCTGTCC
  E  W  S  L  R  A  R  Y  L  G  N  L  I  L  V  D  P  Q  L  V  M  G  H  E  I  G  T  D  N  L  K  L  F  K  W  R  V  P  V  H

     28570     28580     28590     28600     28610     28620     28630     28640     28650     28660     28670     28680
ACTCCGCTTCACGCCGATATTACCGAATTAGAAACTCATTTTTTCTTTTTAGGTACCCTCATATACCGAGATTAGTTTGTACACGTGAAGTAACCTTTTCAATACTTCATCAGTTATTTT
  S  A  S  R  R  Y  Y  R  I  R  N  S  F  F  L  F  R  Y  P  H  I  P  R  L  V  C  T  R  E  V  T  F  S  I  L  H  Q  L  F  L

     28690     28700     28710     28720     28730     28740     28750     28760     28770     28780     28790     28800
TAGTTTTGTTAACCAATGAAAAAAAAGCTCACTGGAAGTCACTCTGGCGAGGGATTAAAGATGGTGTTTTTTATAAAAGTTAGCCATTTGAGAACAAGCTAGCCCTAAAATTTAGAGCTT
  V  L  L  T  N  E  K  K  A  H  W  K  S  L  W  R  G  I  K  D  G  V  F  Y  K  S  *
```

```
     28810     28820     28830     28840     28850     28860     28870     28880     28890     28900     28910     28920
GATTGAAATCTTACAGAACTCTTCAAATACCAAAAAAAATGCTTTGATAGAAATTTTCTATTTTGATGTGACGCATAATGAATGGCTTTAATATCTGGAAAGTAGTAGATATTGAAGTCG

     28930     28940     28950     28960     28970     28980     28990     29000     29010     29020     29030     29040
CTTTTTTTTGCACGAGTACAAATGTCTGCATCCTCAAAGTACATAAAGTATTTTTCATCAAAGCCGCCTAACTTCTTAAAGCAACTTACTTTAAATAGTAAAAATGATCCTGCAGCCCAC

     29050     29060     29070     29080     29090     29100     29110     29120     29130     29140     29150     29160
TCTATTTTCATTGGTTGCTTTATTTGGCTTTTGTCGTAAACATCATTCCGCTTTATACCAAGTAGCGATTTAAATGGTGATAGGAGTTTAGGATATCGGCGGATAGAGTTGTCATACATA

     29170     29180     29190     29200     29210     29220     29230     29240     29250     29260     29270     29280
CTTAACGAACGGTCAGTAAATAGGTTAATAGCAGAAATATCGGCTTGATACTCTTCAGCTTGCTCGATAAGTTGAGATAGGCTGCTGGTATCGATTTCAACATCAGGGTTAAGTACTAGA

     29290     29300     29310     29320     29330     29340     29350     29360     29370     29380     29390     29400
AAATAGTCTCGCTCCGCAGGTTTGAATTCACGCATCGTATACAAAAACACTTCGTTATTATTAGCACCAAATCCTTTTTTGGTGTCTCCATGGATTAGCATAATCTCATGCTCTAAGCAG

     29410     29420     29430     29440     29450     29460     29470     29480     29490     29500     29510     29520
TAATTAGACAGTGCTTCGGAGGCGACAGTATTCGACTTCACGACAACTTTATATTGTTTGGCCAGTAATGCTAGTGTTGGATTGGTTACGATCATTTCGTCATGGGCATGATTGACGACC

     29530     29540     29550     29560     29570     29580     29590     29600     29610     29620     29630     29640
GATATGAATAGGTTCATTAACTAGTAGCCTTAATTAATTTTTCGTTAGGATTATAACCAGTTGGCGCTTCTAAAAGAAGGCGCTTAATACATTCTACATCGAAGTTATGACAGCATATAT

     29650     29660     29670     29680     29690     29700     29710     29720     29730     29740     29750     29760
CCAATTCTTGGAGTAGGGGATGCATCTCTTTCCAAGTGAGCTTATCTTCACAAGCCGTCATTATCTTTTGGTGGCTTGTGCCCTCGACATTGTCACCAATCAATAATTCCTCGTAGAGCT

     29770     29780     29790     29800     29810     29820     29830     29840     29850     29860     29870     29880
TTTCACCGGGTCTCAAACCAGTGAATTTAATTTCAATATCTCCCTCATTGCTATTGCCATCAAAGTACTCTTTCATACCCATCAGGTGGATCATCCGTTTTGCTAGTCGACAATTTTAAC

     29890     29900     29910     29920     29930     29940     29950     29960     29970     29980     29990     30000
TGGCTCGCCCATATCTAACACAAAGACTTGTCCATTATGCCCCATTGCTCCAGCCTGAATAACCAACTGGGCTGCTTTCTGGTATTAGCATAAAGTAGCGGATAATGTCAGGGTGAGTAA

     30010     30020     30030     30040     30050     30060     30070     30080     30090     30100     30110     30120
CGGTAACAGGGCCGCCCGCTTTTATCTGTTTTTTAAAGAGGGGTACAACAGAACCTGAAGAACCAAGCACATTACCAAAACGCACCATAGTAAAAATAGTTTTGGTTCCTTTATCAGCCA

     30130     30140     30150     30160     30170     30180     30190     30200     30210     30220     30230     30240
GTGCCTGTAGAACTAATTCTGCCATCCGTTTACTGGTGCCCATAATATTCGTAGGACGTACCGCTTTATCGGTTGAGATAAGTGTGAAACTTTCGACACCAGCCTCGATAGCCGCCGTAG

     30250     30260     30270     30280     30290     30300     30310     30320     30330     30340     30350     30360
CACATGCTAGGGTACCGAATACATTGTTGCGAATACCTTCGATGACGTTATCTTCTACAAGCGGTACATGCTTATAAGCTGCAGCATGGTAAATAGTGTTAACTTTATGAGCGACCATTA

     30370     30380     30390     30400     30410     30420     30430     30440     30450     30460     30470     30480
GTTTATGTAAACGGTTTTCCTTTTGCACTGAACCAAGAGCCGCGACAATTTTGGTCTCTAATTTTAGATTTTGTTTTATGGCTTGCAGTTCTTGATCCTGCCATGGTTGCTCACGGCTCA

     30490     30500     30510     30520     30530     30540     30550     30560     30570     30580     30590     30600
AAAGATTCGAACTGGGGTACTGGCTTAGAGCAAAAGCAAATTTGGTTCTTAACGTTCTGTATCAGCACGTTCACCAATCAAATTCAGCAATGCGTAAACAAAAAGCTACTTACCCCAGTG

     30610     30620
                     EcoRI
CAGAGAGTTAGCTATTACAGCGAATTC 3'
```

Appendix II

Stroeher, U.H., Jedani, K.E., and Manning, P.A. (1998) Genetic organisation of the regions associated with surface polysaccharide synthesis in *Vibrio cholerae* O1, O139 and *Vibrio anguillarum* O1 and O2: a review. *Gene* **223:** 269-282.

Daniels, K.E., Stroeher, U.H., and Manning, P.A. (1999) Distribution of IS*1358* and linkage to *rfb*-related genes in *Vibrio anguillarum. Microbiol.* (Manuscript submitted).

Daniels, K.E., Stroeher, U.H., Milton, D., Manning, P.A., and Morona, R.M. (1999) Genetics of lipopolysaccharide biosynthesis in *V. anguillarum* O1. (Manuscript in prep.).

Corrigenda

Fig. 3.3. The protein indicated by an asterisk is abundant in the *V. anguillarum* strains tested. Evidence that it is outer membrane protein was stated on pg 79, Line 1 although the data was not shown.

Fig. 3.6. Figure legend should include: Numerous fields were examined and gold particles observed were found to be located on the body and flagella of the bacterial cells.

Fig. 3.10. Figure legend "as shown in Fig. 3.10" should read "as shown in Fig. 3.9".
Page 84, Line 12-13. Partial cleavage of DNA was ruled out as other restriction enzymes were used to determine the copy number of IS*1358* in serotype O2, however the data was not shown.

Page 84, Line 15. Sentence "Serotype O7 was found to contain one complete IS*1358* element and a partial element" should read to clarify Fig. 3.10a "Serotypes O4 and O7 were each found to contain a complete and partial IS*1358* element".

Page 87, Line 6-7. It was decided to investigate the distribution of IS*1358* in numerous Vibrionaceae to determine how widespread this element was in this genera, as stated on Page 87, line 20.

Page 91, Section 4.25 and 4.26. The 2.2 kb and 0.72 kb fragments generated were sequenced, however no sequence homology was found and the sequences and therefore were not studied further.

Page 107, Line 6. No comment, affiliation included in this reference.

Page 108, Line 4. "Immediately upstream of the start of *rmlD*" should read "Immediately upstream of the start of *rmlB*".

Page 117, Section 5.2.8. No comment required.

Page 120, Line 10. Due to the presence of numerous potential promoters it seems that the region sequenced consists of several operons.