MOLECULAR CHARACTERIZATION OF THE *RFB* REGION OF *VIBRIO CHOLERAE* O1 (OGAWA)

Melissa H. Brown, B.Sc. (Hons.) (Adelaide)

Department of Microbiology and Immunology,

The University of Adelaide,

Adelaide, 5000

AUSTRALIA

A thesis submitted for the degree of Doctor of Philosophy.

April, 1991.

ABSTRACT

The genes (*rfb*) encoding the biosynthesis of the O-antigen of the lipopolysaccharide (LPS) of *V. cholerae* O1 have previously been cloned and expressed in *E. coli* K-12 (Manning *et al.*, 1986). This study has sought to further analyze the region of DNA encoding these genes.

Construction of a comprehensive restriction map of the 20 kb region encoding O-antigen biosynthesis, facilitated the generation and characterization of subclones, deletion derivatives and transposon Tn*1725* insertion mutants. Analysis of these plasmids enabled definition of a minimal coding region to 16.5 kb. Complementation analyses with the subclones in *E. coli* K-12 suggested two, or possibly three, regions were involved in serotype specificity.

Electron microscopy of RNA polymerase molecules bound *in vitro* to the purified plasmid DNA of clones harbouring the *rfb* region was used as a basis for localizing areas possessing potential promoter activity. Restriction fragments were subcloned into the promoter detection vectors pKC86 and pKC87, screened on indicator plates and the relative promoter strength was determined by assaying for galactokinase activity. These studies suggested the presence of a large transcriptional unit being involved in O-antigen expression. Promoters identified for open reading frames on the complementary strand were localized outside of the *rfb* cluster. Cloning of some DNA fragments resulted in the generation of artifical promoters. This limited the usefulness of this aspect of the study.

As a collaborative effort, the entire 20 kb *rfb* region has been sequenced, and a detailed analysis of four regions, totalling 7.5 kb has been undertaken. The sequences of these nine *rfb* encoded proteins were computer analyzed and compared to other known protein sequences.

Analysis of gene expression of the *V. cholerae rfb* region was undertaken using RNA methodology, such as dot blots and northern transfers. However, due to low amounts of transcribed message, by comparison to that for other *V. cholerae* proteins, these experiments did not yield definitive results with respect to the operon structure.

In summary, the *V. cholerae rfb* region appears to exist as one major operon encoding the majority of proteins required for biosynthesis and assembly of the O-antigen. A gene encoding the genetic determinant for the Ogawa specific antigen is located adjacent to the major operon. The promoters are weak and this has hampered direct detection of the *rfb* specific mRNA levels.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr. Renato Morona for his supervision, enthusiasm and helping me retain my sanity.

I am also indebted to Professor Paul Manning for his supervision and english prowess.

For his patience and time I would like to thank Dr. Connor Thomas.

I also must thank Briony Forbes, Sarah Robertson, and Dr. Vicki Franzon for their friendship and understanding throughout my study.

To my family, my thanks for their support and encouragement.

To Christopher Correll, thankyou for your loving understanding, without which I would not have completed this work.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and to the best of knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis. The author consents to the thesis being made available for photocopying and loan, if applicable and if accepted for the award of the degree.

Melissa H. Brown

LIST OF ABBREVIATIONS

A	: adenine
aa	: amino acid
ACL	: antigen carrier lipid/undecaprenol phosphate/bactoprenol
Ap	: ampicillin
Ara	: arabinose
ATP	: adenosine 5'-triphosphate
bp	: base or nucleotide pair
BSA	: bovine serum albumin
CAT	: chloramphenicol acetyltransferase
C	: cytosine
Cm	: chloramphenicol
cpm	: counts per minute
CT	: cholera toxin
DNA	: deoxyribonucleic acid
DNase	: deoxyribonuclease
dNTP	: deoxyribonucleoside triphosphate
ddNTP	: dideoxyribonucleoside triphosphate
DDT	: dithiothreitol
EDTA	: ethylene-diamine-tetra-acetic acid
ELISA	: enzyme linked immunosorbent assay
Etn	: ethanolamine
EtBr	: ethidium bromide
G	: guanine
Gal	: galactose
g/l	: grams/litre
Glc	: glucose
GlcNAc	: N-acetyl-glucosamine
GM_1	: galactosyl-N-acetyl-galactosaminyl-sialosyl-lactosyl ceramide

HA	: haemagglutinin
Hep	: L-glycero-D-mannoheptose
HIA	: haemagglutination inhibition assay
Hly	: haemolysin
IM	: inner membrane
IPTG	: isopropyl-ß-D-thiogalactopyranoside
kb	: kilobase or 1000 base pairs
kDa	: kilodalton
KDO	: 3-deoxy-D-manno-octulosonic acid
Km	: kanamycin
LB	: Luria broth
LPS	: lipopolysaccharide
LT	: heat labile toxin
MAb	: monoclonal antibody
Man	: mannose
MDO	: membrane derived oligosaccharide
mRNA	: messenger ribonucleic acid
NA	: nutrient agar
NAG	: non-agglutinable
NB	: nutrient broth
N.M.R.	: nuclear magnetic resonance
nt	: nucleotide
NTG	: N-metyl-N'-nitro-N-nitrosoguanidine
O-Ag	: O-antigen
O-PS	: O-specific side chain polysaccharide
OD	: optical density
ORF	: open reading frame
PAGE	: polyacrylamide gel electrophoresis
PBS	: phosphate buffered saline
PEG	: polyethylene glycol-6000
perosamine	: 4-amino-4,6-dideoxy-D-mannose
quinovosamine	: 2-amino-2,6-dideoxy-D-glucose
R	: resistant
RBC	: red blood cell

RBC	: red blood cell
RF	: replicative form
Rha	: rhamnose
RNA	: ribonucleic acid
rpm	: revolutions per minute
s	: sensitive
SD	: Shine-Dalgarno
SDS	: sodium dodecyl sulphate
SRBC	: sheep red blood cells
s/s	: single stranded
str	: streptomycin
T	: thymine
Tc	: tetracycline
TEMED	: N,N,N',N'-tetramethyl-ethylene-diamine
Tn	: transposon
Tris	: Tris-hydroxymethyl-aminomethane
U	: uracil
UV	: ultraviolet
VcRfb	: *Vibrio cholerae rfb* encoded protein
VcOag	: *Vibrio cholerae* O-antigen
v/v	: volume per volume
w/v	: weight per volume
X-gal	: 5-bromo-4-chloro-3-indoyl-β-D-galacto-pyranoside

CONTENTS

Chapter 1: Introduction

Chapter 2: Materials and Methods

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Cholera

1.1.1 Introduction

The genus Vibrio belongs to the family Vibrionaceae, which also includes the genera *Plesiomonas*, *Photobacterium* and *Lucibacterium* (Bauman and Schubert, 1984; Shewan and Veron, 1975). *Vibrio cholerae* is a Gram-negative curved rod, 0.3 - 0.4 μm wide and 1.5 - 2.0 μm long, with a single polar sheathed flagellum. The organism is facultatively aerobic, preferring alkaline growth conditions as it is extremely acid sensitive (Davis *et al.*, 1980).

V. cholerae can be subdivided into six O groups based on the O-antigen of the lipopolysaccharide (LPS). Cholera vibrios belong in O subgroup 1, as defined by Gardner and Venkatraman (1935), within which are two biotypes, Classical and El Tor (Bauman *et al.*, 1984; Freeley, 1965), both capable of causing human cholera. The Classical biotype is thought to have been responsible for all the major epidemics until the onset of the seventh pandemic initiated in South East Asia, when El Tor *V. cholerae* were found to be the causative agents (Kamal, 1974).

The two biotypes were originally differentiated by the ability of El Tor strains to produce a soluble haemolysin. However considerable variability in this property is now seen. A number of other more reliable differentiating characteristics are now used. El Tor strains may be differentiated from Classical strains by:

(1) their resistance to the antibiotic polymyxin B (Roy *et al.*, 1965),

(2) the ability to agglutinate chicken erythrocytes when grown on solid media (Gan and Tjia, 1963; Gangarosa *et al.*, 1967) and

(3) resistance to the group IV phages of Mukerjee (Monsur *et al.*, 1965; Mukerjee and Roy, 1961).

Recently, Alm and Manning (1990b) have developed a DNA probe based on the nucleotide sequence differences in the *hlyA* gene of *V. cholerae* O1 that can distinguish Classical and El Tor vibrios.

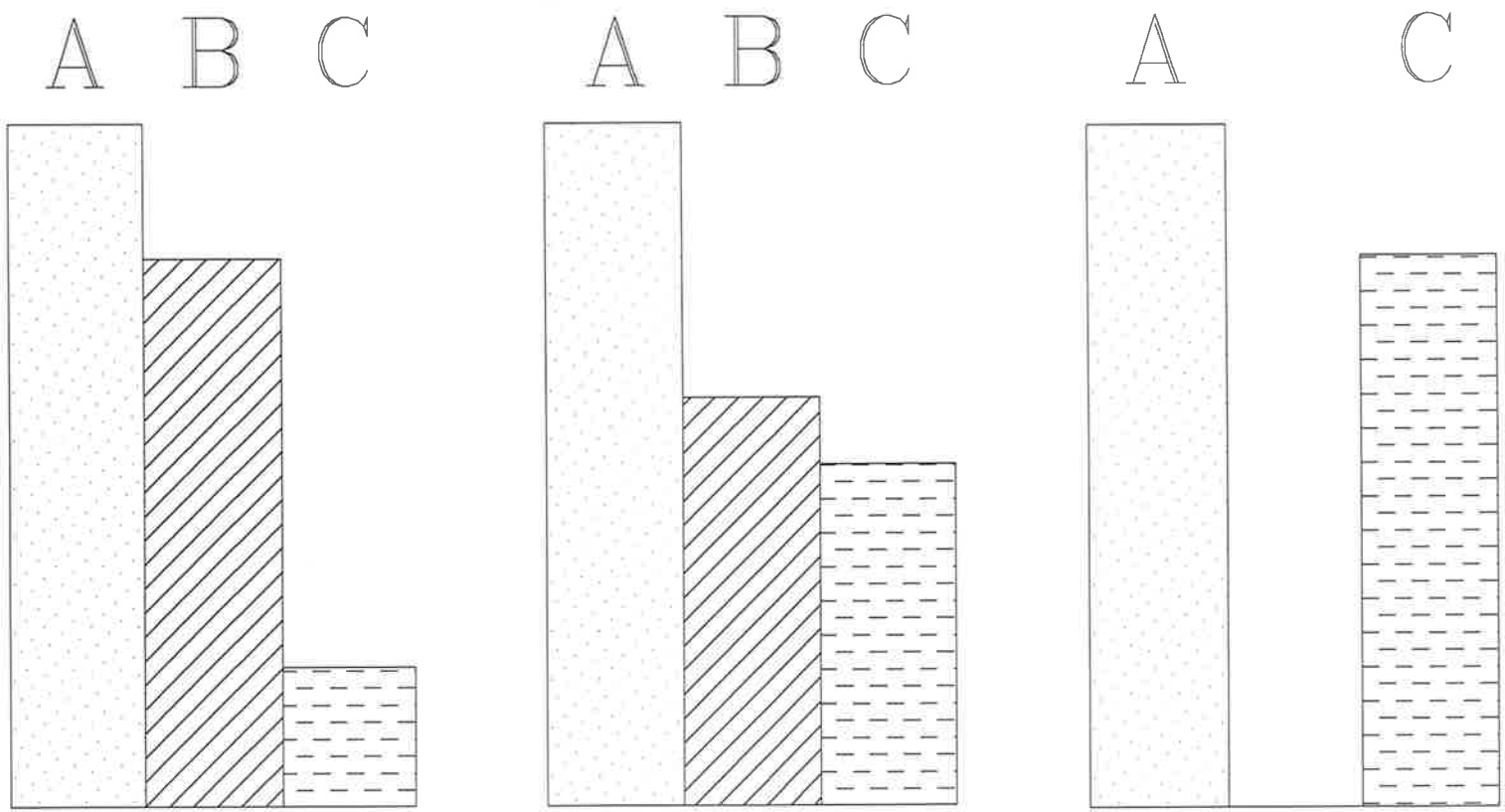
V. cholerae O1 strains of both biotypes can be further subdivided into three serotypes, Inaba, Ogawa and Hikojima. Differentiation of these subtypes is based on the presence of three antigenic factors designated A, B, and C (Figure 1.1). Strains of the Inaba serotype express the A and C antigens while the Ogawa serotype possess the antigens A, B and low amounts of C (Burrows *et al.*, 1946a, 1946b; Redmond, 1979; Redmond *et al.*, 1973; Sakazaki and Tamura, 1971). The third less common and apparently unstable serotype, Hikojima, expresses all three antigenic factors A, B and C (Bhaskaran, 1959; Bhaskaran and Sinha, 1971; Burrows *et al.*, 1946a, 1946b).

1.1.2 Pathogenesis

Cholera is a human infection inducing acute diarrhoea, usually as a consequence of ingesting contaminated food or drinking water. Gastric acid constitutes quite an effective barrier to infection, as neutralization of stomach acidity (by concomitant ingestion of sodium bi-carbonate) reduces the infectious dose from 10^8 to 10^4 organisms (Cash *et al.*, 1974a, 1974b; Davis *et al.*, 1980; Nalin *et al.*, 1978). Once *V. cholerae* has successfully passed through the gastric acid barrier of the stomach and entered the small intestine, it utilizes an array of virulence properties to overcome the nonspecific defense mechanisms of the host, namely small intestine peristalsis and the mucous coating layer. The final diarrhoeal response is multifactorial and involves motility and chemotaxis to penetrate the mucous layer (Freter and O'Brien, 1981; Guentzel and Berry, 1975; Yancey *et al.*, 1978), production of proteases, neuraminidases and DNAses to degrade this

Figure 1.1 Serotypic determinants of *Vibrio cholerae*.

Diagrammatic representation of the relative amounts of A, B, and C antigens in the different *V. cholerae* serotypes. Reproduced from Sakazaki and Tamura (1971).


A B C
A B C
A C
Ogawa
Hikojima
Inaba

mucous layer (Schneider and Parker, 1978); pili and haemagglutins to facilitate adherence to the intestinal epithelial cells (Finkelstein and Hanne, 1982; Manning, 1987), and allow close association with the gut epithelium for effective delivery of the bacterial toxins (Peterson *et al.*, 1972; Pierce *et al.*, 1985).

The identity of the *V. cholerae* adhesin(s) remains to be elucidated, although various surface structures have been implicated; LPS, the flagellar sheath, outer membrane proteins, fimbriae and haemagglutins (Attridge and Rowley, 1983a, 1983b; Chitnis *et al.*, 1982; Ehara *et al.*, 1986, 1987; Hanne and Finkelstein, 1982; Jones *et al.*, 1976; Jones and Freter, 1976; Kabir, 1983; Kabir and Showkat, 1983).

Since 1959, when De showed that the symptoms of cholera were due to the production of a cholera enterotoxin (CT), extensive studies have been performed in relation to its structure, function, biological activity and regulation (De, 1959; Holmgren, 1981; Mekalanos, 1985; Van Heyningen, 1977). All pathogenic strains of *V. cholerae* O1 seem to produce immunologically identical CTs (Evans and Richardson, 1968; Finkelstein, 1969; Holmgren *et al.*, 1971). Production of CT or a cholera-like enterotoxin is not serotype restricted. Biologically and immunologically related enterotoxins have been isolated from both non-O1 *V. cholerae* and non-cholera vibrios (Craig *et al.*, 1981; Yamamoto *et al.*, 1983a, 1983b; Zinnaka and Carpenter, Jr., 1972).

CT is a heat labile, multimeric molecule with an approximate size of 85 kDa, composed of five identical B subunits that enable the toxin to bind to GM_1 ganglioside receptors on the intestinal cells, and a toxic component, the A subunit, which triggers cyclic-AMP formation and fluid secretion by enteric cells (Cuatrecasas, 1973a, 1973b; Gill, 1976; Lospalluto and Finkelstein, 1972; Ludwig *et al.*, 1986; Mekalanos *et al.*, 1979).

1.2 Cholera vaccines

Protective immunity has been shown to be stimulated by infection with *V. cholerae* O1 (Holmgren and Svennerholm, 1977), as in cholera endemic areas there is

a fall in the incidence of cholera with age, the prevalence of serum vibriocidal antibody increases, and recurrence of clinical cholera infections are rare (Sears *et al.*, 1984).

The vibriocidal antibodies are directed towards the bacteria as well as the enterotoxin, both being assigned a protective role, as demonstrated in experimental animal models. The relative importance of each component has been the subject of considerable debate (Holmgren and Svennerholm, 1977).

Cash *et al.*,(1974a, 1974b) demonstrated that antibacterial mechanisms in the gut were more important than antitoxin mechanisms. This was supported by Levine *et al.* (1988), who provided further evidence that antitoxin immunity was not an absolute prerequisite of protective immunity against cholera in humans.

While the neutralizing effect of the antitoxin antibodies is known to be directed against the B subunit of CT (Holmgren, 1981), there is no agreement on the precise nature of the protective antigens involved in antibacterial immunity (Holmgren and Svennerholm, 1977; Levine *et al.*, 1983; Neoh and Rowley, 1970, 1972). The Vibrio outer membrane is composed of a number of proteins and LPS. Neoh and Rowley (1970), found both anti-LPS and anti-protein antibodies to be involved in vibriocidal activity. Antibodies to the outer membrane proteins have been shown to be protective in the infant mouse model (Attridge and Rowley, 1983b) and in the rabbit model, where antibodies directed against the outer membrane proteins were detected by ELISA, and immunoblotting (Cryz *et al.*, 1982; Kabir, 1980; Manning and Haynes, 1984; Sears *et al.*, 1984).

LPS is a more potent immunogen than the protein antigens, and antibodies to these components may protect by inhibiting motility, adherence and colonization (Attridge and Rowley, 1983b; Chitnis *et al.*, 1982; Manning *et al.*, 1986; Neoh and Rowley, 1970). Svennerholm (1975), found that the entire protective capacity of a hyperimmune serum could be accounted for by antibodies directed against LPS.

Antibacterial and antitoxic immunities have been experimentally shown to act synergistically in mediating resistance to cholera (Holmgren *et al.*, 1977; Pierce *et al.*,

1982). This has enticed scientists to develop combined vaccines containing bacterial and toxoid components.

Manning and coworkers (1986) reported the cloning of the genes responsible for the production of *V. cholerae* O1 O-antigen of both Ogawa and Inaba serotypes. The O-antigen clone expressing Inaba in *E. coli* K-12 was used to generate an antiserum which was then analyzed for protective activity in the infant mouse model. It was shown that the O-antigen produced by the *E. coli* K-12/*V. cholerae* Inaba hybrid organism provided the same degree of protection against challenge with a virulent *V. cholerae* Inaba 569B in the infant mouse protection test, as did an antiserum raised by using the parent *V. cholerae* Inaba strain.

The cloned Inaba genes have been used to construct an oral cholera vaccine candidate (Tacket *et al.*, 1990) using Ty21a (Germanier and Fürer, 1975), an attenuated strain of *Salmonella typhi*, as a carrier. Ty21a was chosen as a vector because;

(1) there were no side affects (Formal *et al.*, 1981),

(2) live *Salmonella* colonize Peyer's patches leading to development of an IgA response (Srisart *et al.*, 1985) and

(3) the use of Ty21a expressing protective antigens from other pathogens could produce a bivalent vaccine (Clements and El-Morshidy, 1984; Mills *et al.*, 1988). Such a bivalent vaccine could confer resistance to both typhoid and to the pathogen of interest by stimulating immunity to the cloned antigen.

The hybrid typhoid/cholera vaccine has been shown to elicit antibody responses in humans to both the *S. typhi* and *V. cholerae* O-antigens (Forrest *et al.*, 1989; Tacket *et al.*, 1990). Since the only cholera antigen in the vaccine is the Inaba O-antigen, the 25% protective efficacy and the reduction in excretion of *V. cholerae* challenge organisms by the vaccinees is due to anti-Inaba activity alone.

This particular vaccine is still not ideal. Future work needs to be done, such that the cholera genes are incorporated into the chromosome and not on a plasmid. A salmonella carrier with precisely defined attenuating mutations such as nutritional auxotrophies should also be used, since Ty21a was constructed by nonspecific

mutagenesis with NTG and probably contains numerous unidentified mutations (Edwards and Stocker, 1988; Stocker, 1988). A more immunogenic vaccine is required such that a single dose of a smaller number of organisms would produce good protection.

There are still problems effecting O-antigen expression in heterologous hosts, such as growth media effects (Manning *et al.*, 1986). A better understanding of the genetics of O-antigen expression may therefore be beneficial to vaccine construction.

1.3 Genetic analysis and gene regulation in *Vibrio cholerae*

The three methods of introducing genetic material; transformation, transduction, and conjugation have been studied in *V. cholerae*. Until recently, transformation of *V. cholerae* has not been very successful, mainly due to the production of extracellular DNAses (T. Focareta, PhD thesis, University of Adelaide, 1989; Focareta and Manning, 1987; Newland *et al.*, 1985). *V. cholerae* mutants constructed by deleting the genes encoding both DNAses, and introduction of these constructs into the *V. cholerae* chromosome has enabled transformation to be achieved (T. Focareta, PhD thesis, University of Adelaide, 1989). The technique of electroporation, using *V. cholerae* directly for genetic engineering is becoming a more feasible proposition. Transfection of *V. cholerae* with phage DNA has been reported (Balgenesh and Das, 1979), but it is not reproducible.

Transduction has also been shown by several workers (Ogg *et al.*, 1981). Even though generalized transducing phages have been identified, a cholera phage capable of specialized transduction has not been isolated.

A number of phages capable of lysogeny in both biotypes of *V. cholerae*, integrate randomly into the host chromosome making them genetically useful in creating auxotrophic mutants (Goldberg and Murphy, 1983) and cholera toxin deletion strains (Mekalanos *et al.*, 1982).

CP-T1 was the first generalized transducing phage to be described (Ogg *et al.*, 1981), and has since been examined in detail with respect to its packaging mechanism,

which is by a "headful packaging" method, commencing at a specific site designated *pac* (Guidolin *et al.*, 1984), and its cell surface receptor which has been identified as *V. cholerae* O1 O-antigen (Guidolin and Manning, 1985).

The *E. coli* K-12 phage lambda (λ), which uses the LamB protein as its receptor, can also infect *V. cholerae*, when the cloned *lamB* gene is introduced into *V. cholerae* (Galen *et al.*, 1988).

1.3.1 Chromosomal mapping

The *V. cholerae* conjugation system resembles the F system of *E. coli* K-12, as it is mediated by a naturally occurring conjugative plasmid, P (Bhaskaran, 1960). Donor cells possessing P factor (P^+) are capable of derepressed self transfer to P^- recipients (Parker and Romig, 1972). Transfer of the P plasmid produces "lacunae" due to zones of inhibited growth probably as a consequence of lethal zygosis (Parker and Romig, 1972). The P plasmid has a length of 68 kb and does not belong to any of the common incompatability groups (Bartowsky *et al.*, 1987; Manning, 1988). Unlike F, P does not stably integrate into the bacterial chromosome to form Hfr strains, possibly due to differences in the GC ratio of P, relative to *V. cholerae* chromosomal DNA and/or a lack of suitable insertion sequences (Datta *et al.*, 1973; Sublett and Romig, 1981; Wohlhieter *et al.*, 1975). Vibrio recombination frequencies are lower, and non-selected markers are poorly linked to selected markers (Bhaskaran, 1960; Parker *et al.*, 1979).

Mapping experiments in *V. cholerae* were performed by Bhaskaran in 1960, who used auxotrophic strains derived from the Classical strain 162. Seventeen markers were ordered, one being the *oag* locus encoding serotype specificity, placing it between *ilv-1* and *arg-1*. Parker and co-workers (1979) expanded the linkage map using both linkage and crossover class analyses.

In 1979, Johnson and Romig developed a transposon-facilitated recombination (Tfr) system in *V. cholerae* El Tor to produce donor strains that could transfer chromosomal genes at a high frequency. From their studies a circular genetic map of the

El Tor chromosome was conceived and the gene order obtained agreed closely with that previously described for the Classical 162 strain. In 1983, the Tfr system was applied by Green *et al.* (1983) to recent clinical isolates of *V. cholerae*, resulting in the mapping of the El Tor biotype markers: haemolysin production (*hly*), polymyxin B resistance (*pmx*) and chicken erythrocyte haemagglutinin (*cha*), identifying a locus closely linked to *leu*. This region of the chromosome, designated the biotype locus, was found to be inverted with respect to the map of *V. cholerae* Classical strain 162.

Ward and Manning (1989) mapped the *rfb* region of *V. cholerae* using strains which have transposon insertions in the regions associated with LPS biosynthesis. This confirmed that the *oag* region associated with serotype specificity of *V. cholerae* LPS mapped by Bhaskaran (1960), is the same as the *rfb* region. Another locus although separate from *rfb* but involved in LPS expression was mapped near *ilv-1*. It has been postulated that this region may correspond to the *rfa* locus responsible for the core biosynthesis (H. Ward, PhD thesis, University of Adelaide 1988) (Figure 1.2).

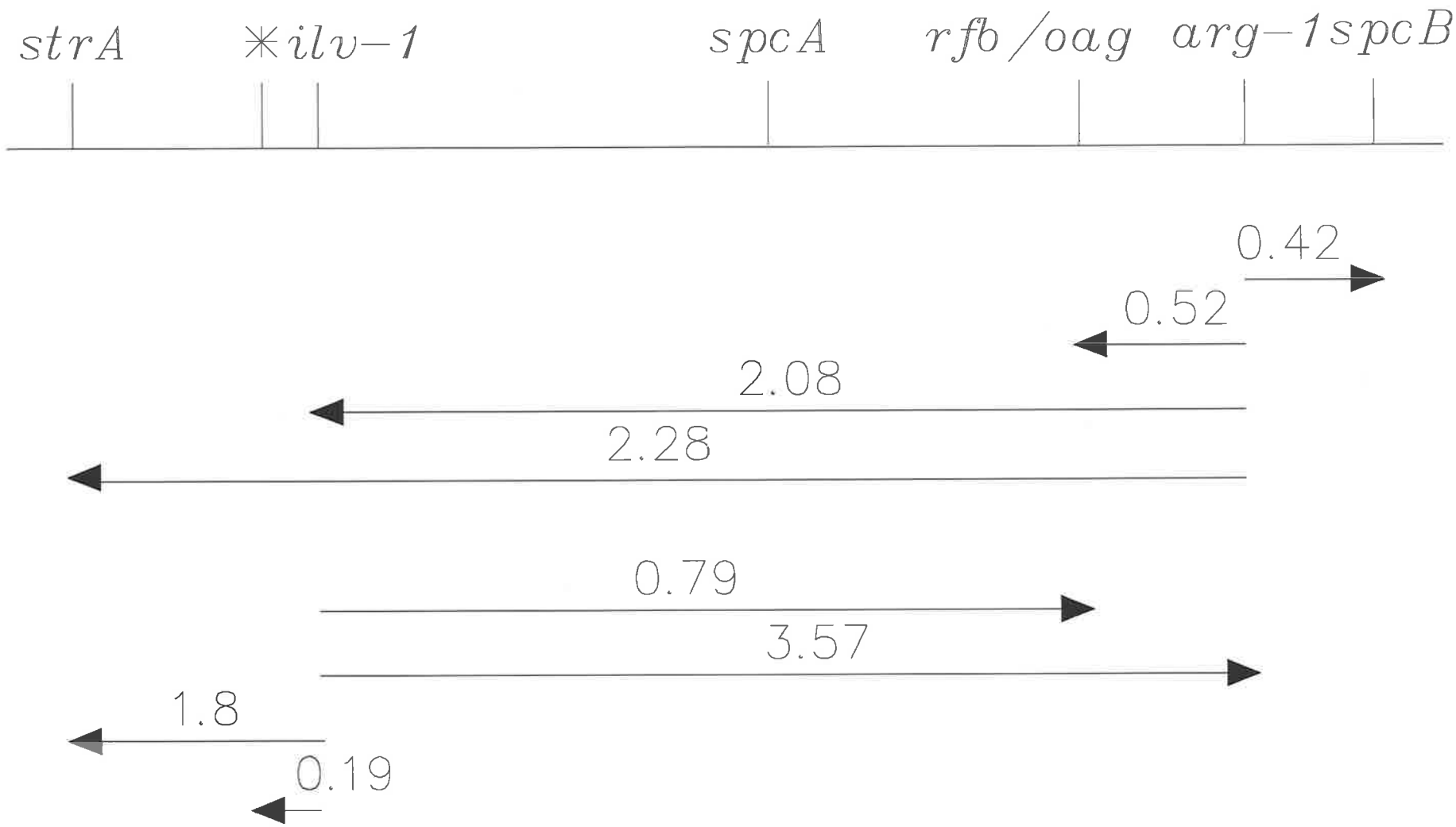
1.3.2 ToxR regulation

The genes encoding the cholera toxin (CT), *ctxA* and *ctxB* are arranged in one transcriptional unit. The *ctxAB* operon has been shown to resemble a complex transposon, and is subject to gene amplification during intestinal passage (Mekalanos *et al.*, 1983).

When the *ctxAB* operon was cloned in *E. coli*, expression was greatly reduced compared to that seen in *V. cholerae* (Pearson and Mekalanos, 1982). This was due to either *E. coli* not efficiently transcribing/translating the vibrio *ctxAB* genes, or the lack of a specific regulatory gene product. In 1984, Miller and Mekalanos, using strain 569B and *tox* mutants, showed via Northern blot analysis that mutations in a locus called *tox* decreased the level of *ctx* mRNA. Genetic evidence has now shown that the *toxR* gene is the *tox* locus previously described by Miller and Mekalanos (1985).

Figure 1.2 Chromosomal location of the *rfb* gene cluster in *V. cholerae* O1.

This genetic map of the *V. cholerae* O1 chromosome shows the linkage of *rfb* to *ilv*-1 and *arg*-1. Values represent relative distances. The star denotes the position of a second region involved in LPS expression, postulated to be the *rfa* locus. The figure is reproduced from Ward and Manning (1989).


strA
*ilv-1
spcA
rfb/oag
arg-1spcB
0.42
0.52
2.08
2.28
0.79
3.57
1.8
0.19

Analysis of the *ctx* promoter region in different strains has shown the presence of three to eight directly repeated copies of the sequence TTTTGAT, located 77 bp upstream of the initiation site for transcription of the *ctxAB* operon. These repeats are recognized by ToxR and are required in *cis* for ToxR mediated activation of *ctx* transcription (Miller *et al.*, 1987). The nucleotide sequence of *toxR* was determined and the gene product is a 32.5 kDa transmembrane DNA-binding protein, that may have the ability to sense a variety of enviromental signals including pH, osmolarity, temperature and the concentration of certain amino acids (Miller *et al.*, 1987). *toxR* shows considerable sequence homology to a set of five transcriptional prokaryotic activators: the *virG* gene of *Agrobacterium tumefaciens*; the *bvgA* gene of *Bordetella pertussis*; the *fur* gene of *E. coli*; the *lcr* locus of *Yersiniae* and the *agr* gene of *Staphylococcus aureus*. (Miller *et al.*, 1987; Peterson *et al.*, 1988).

A second regulatory element *toxS* has been found to lie 4bp downstream and in the same transcriptional unit, as *toxR* (Miller *et al.*, 1989; Peterson *et al.*, 1988). ToxS has been proposed to act in *trans*, not by directly activating the *toxR* promoter, nor by causing *toxR* to activate its own promoter, but rather to modify ToxR to make it competent for transcriptional activation of the *ctxAB* promoter (Miller *et al.*, 1989). Another element *toxT* is thought to be involved in regulation. This newly identified protein acts not on the *ctxAB* promoter but on promoters for several other cholera proteins (R.K.Taylor, personal communication).

Data compiled by a number of workers have shown that the ToxR protein may regulate the transcription of a number of genes involved in the virulence of *V. cholerae*. The transposon Tn*phoA*, has been used to generate in frame gene fusions between *phoA*, encoding alkaline phosphatase, and the genes for a variety of exported proteins of *V. cholerae*. This has revealed a number of ToxR regulated elements: the biosynthetic genes for the TCP pilus, (involving at least 7 genes); the structural genes for two outer membrane proteins, OmpU and OmpT; and a cluster of 4 genes for ACF (accessory colonization factor) (Miller, 1985; Miller and Mekalanos, 1984, 1988; Peterson and

Mekalanos, 1988; Taylor *et al.*, 1987a, 1987b). Except for OmpT which is negatively controlled, the other genes are activated by ToxR.

1.4 Biosynthesis of lipopolysaccharide

The cell walls of Gram-negative bacteria consist of two electron microscopically discernable entities, the outer membrane and the peptidoglycan layer. The outer membrane is an asymmetric lipid bilayer containing LPS in the outer leaflet and phospholipid in the inner leaflet, interspersed with major and minor protein species (Haller and Henning, 1974) (Figure 1.3).

LPS from members of the family Enterobacteriaceae have been structurally analyzed, but it is the LPS structure from *S. typhimurium* that has been most thoroughly described and provides a basis for comparison with other species.

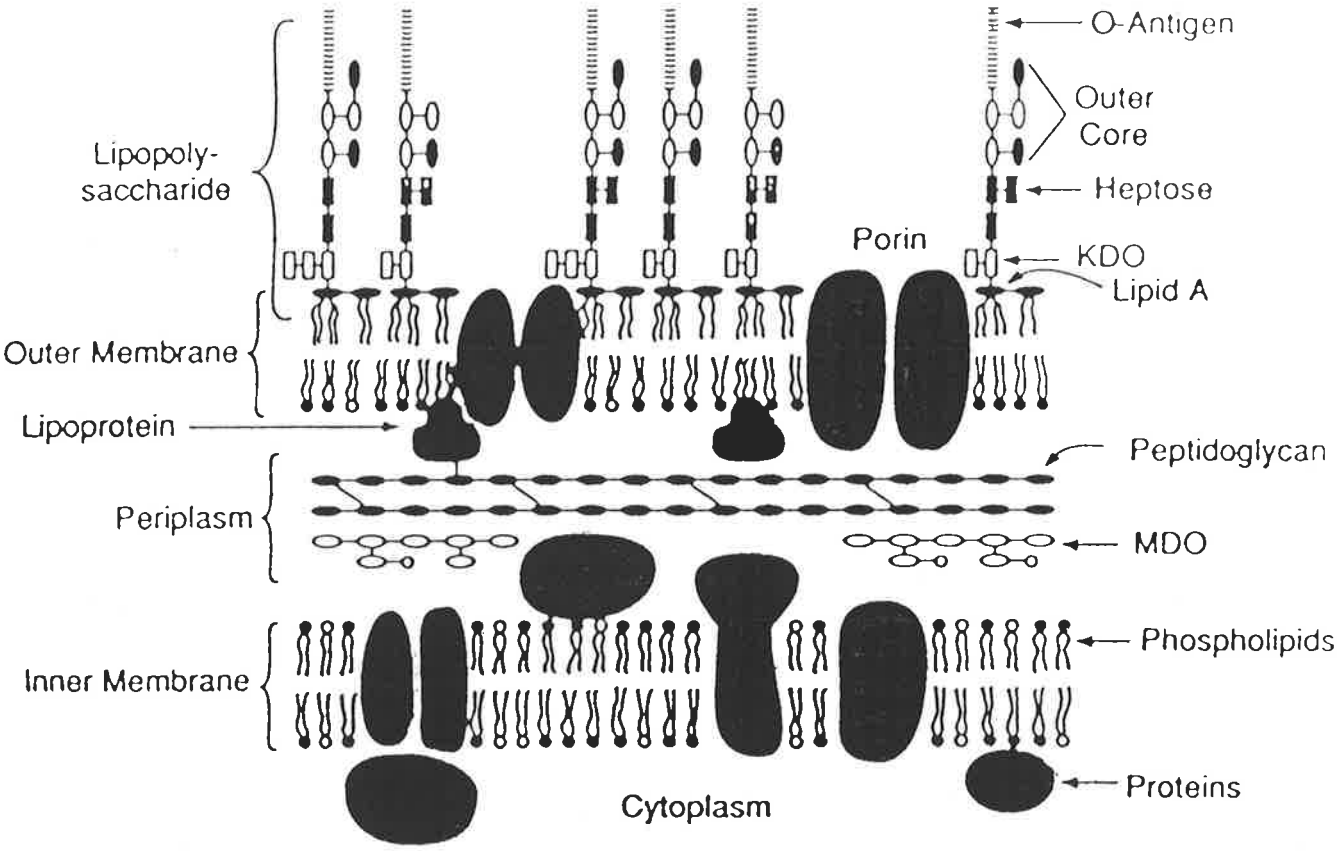
The LPS molecule consists of three covalently linked structural regions; the lipophilic lipid A moiety embedded in the outer membrane, and the two component polysaccharide composed of the core oligosaccharide and the O-somatic (O-antigenic) side-chains. The core can be further subdivided into an inner core or backbone region, and an outer core region (Figure 1.3).

The integrity of the cell wall is also dependent upon LPS (Hemming, 1970). This is most pronounced in deep rough mutants which have a lesion in the more proximal regions of the core oligosaccharide. These mutants are hypersensitive to fatty acids, phenol, polycyclic hydrocarbons, and also to detergents, antibiotics and hydrophobic dyes (Nikaido and Vaara, 1985).

The term endotoxin is synonymously used for LPS, because of the reaction induced when purified LPSs are, or bacteria producing LPS is, injected in mice (Kass and Wolff, 1973; Schlessinger, 1977). LPS is an important molecule in the physiology of the bacterium-host relationship as it appears to have the ability to both activate and suppress lymphocyte functions (Koenig and Hoffmann, 1979; Melchers, 1980; McGhee *et al.*,

Figure 1.3 *Salmonella typhimurium* envelope.

Schematic representation of the *S. typhimurium* envelope. Ovals, and rectangles represent sugar residues within the envelope. KDO is 3-deoxy-D-manno-octulosonic acid, and MDO are membrane derived oligosaccharides. Figure is reproduced from Raetz (1990).


O-Antigen
Outer Core
Heptose
KDO
Lipid A
Porin
Lipopoly-saccharide
Outer Membrane
Lipoprotein
Peptidoglycan
Periplasm
MDO
Phospholipids
Inner Membrane
Proteins
Cytoplasm

1980). Components of LPS can activate complement (Joiner *et al.*, 1984; Morrison and Kline, 1977; ~~Takasaki *et al.*, 1983~~).

1.4.1 Lipid A

The lipid A component of LPS represents the minimum structure necessary for pyrogenicity and endotoxicity (Westphal and Lüderitz, 1954). A number of investigators have shown that lipid A is an amphiphile consisting of a hydrophilic and a hydrophobic region (Munson *et al.*, 1978; Raetz, 1987; Walenga and Osborn, 1980a, 1980b). The development of procedures for the selective degradation of LPS was an important step in the isolation and structural analysis of the hydrophilic region of lipid A (Hase and Rietschel, 1976), revealing that the Salmonella lipid A contains a backbone consisting of a phosphorylated β(1-6) linked D-glucosamyl disaccharide (Gmeiner *et al.*, 1971; Lüderitz *et al.*, 1968). This particular hydrophilic backbone is common amongst various Gram-negative bacteria. Ester-, amide-, and diester-linked fatty acids, where the type and position of fatty acids vary between species, and phosphate, 4-amino-arabinose substituents, or both, can be found on the reducing and non reducing sugars (Hase and Rietschel, 1977; Rietschel *et al.*, 1983; Rosner *et al.*, 1979) (Figure 1.4).

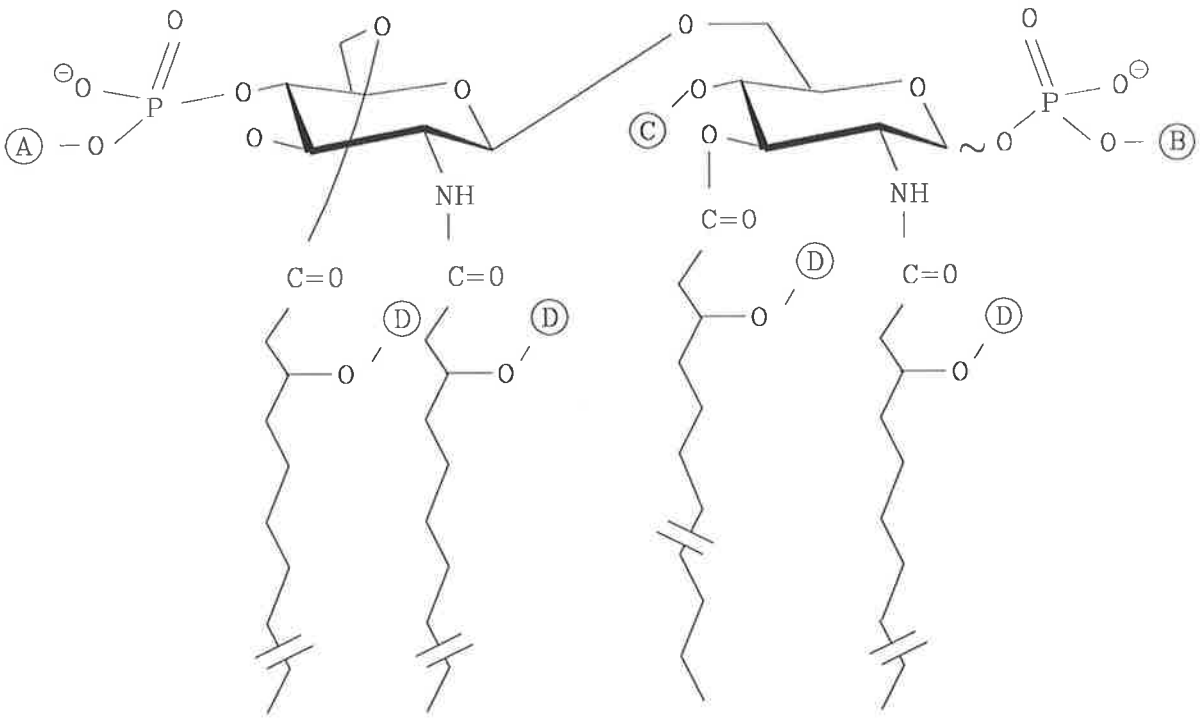
Study of lipid A biosynthesis was made possible by the isolation of conditional lethal mutants in *E. coli* (Raetz, 1987). Synthesis starts in the cytoplasm with a primary molecule of UDP-N-acetylglucosamine (UDP-N-GlcNAc), which is acylated, then hydrolyzed to produce 2,3 diacyl-Glc-1-P catalysed by disaccharide synthetase (Crowell *et al.*, 1986). The genes responsible for both the lipid A disaccharide synthetase and UDP-N-GlcNAc (*lpxB* and *lpxA* respectively), have been cloned and mapped in an operon near 4 minutes on the *E. coli* chromosome (Crowell *et al.*, 1986).

Figure 1.4 Structure of lipid A.

An overall view of the structure of lipid A form different bacterial origins. The circled letters indicate points of structural variation between different lipid As. The lipid A molecule differs between species by:

1) the presence and nature of residues attached at the A and B positions to the backbone phosphate groups
2) differing substituents of the glucosaminyl-saccharide present at position C
3) the type and chain length of fatty acids, and
4) 3-O-acetylation of the 3-hydroxy fatty acids at position D.

Diagram taken from Rietschel *et al.* (1983).


O
⊖O
P
O
A
O
O
O
O
O
NH
C=O
C=O
D
D
O
O
O
O
O
C
O
O
O
O
O
P
O⊖
O
B
NH
C=O
C=O
D
D
O
O

1.4.2 Lipolysaccharide core region

Core oligosacchrides are structurally less diverse than O-chains. The family Enterobacteriaceae has 6 different core polysaccharide structures (Jansson *et al.*, 1981). The five basal sugars of the core oligosaccharide are conserved in *Salmonella* (Hellerquist and Lindberg, 1971). Adjacent to the O-chain is a branched oligosaccharide with N-acetyl-D-glucosamine, D-glucose and D-galactose units, and an inner lipid A proximal region of a branched oligosaccharide with L-glycero-D-manno-heptose and KDO units (Figure 1.5). Except for some uncertainty about the KDO molecules, the linkages have been determined (Rick, 1987).

The structures of the *E. coli* and *Shigella* core oligosaccharides differ slightly from *Salmonella*, with *E. coli* having at least five different core types (Lüderitz *et al.*, 1971; Nikaido, 1970; Schmidt *et al.*, 1969, 1970).

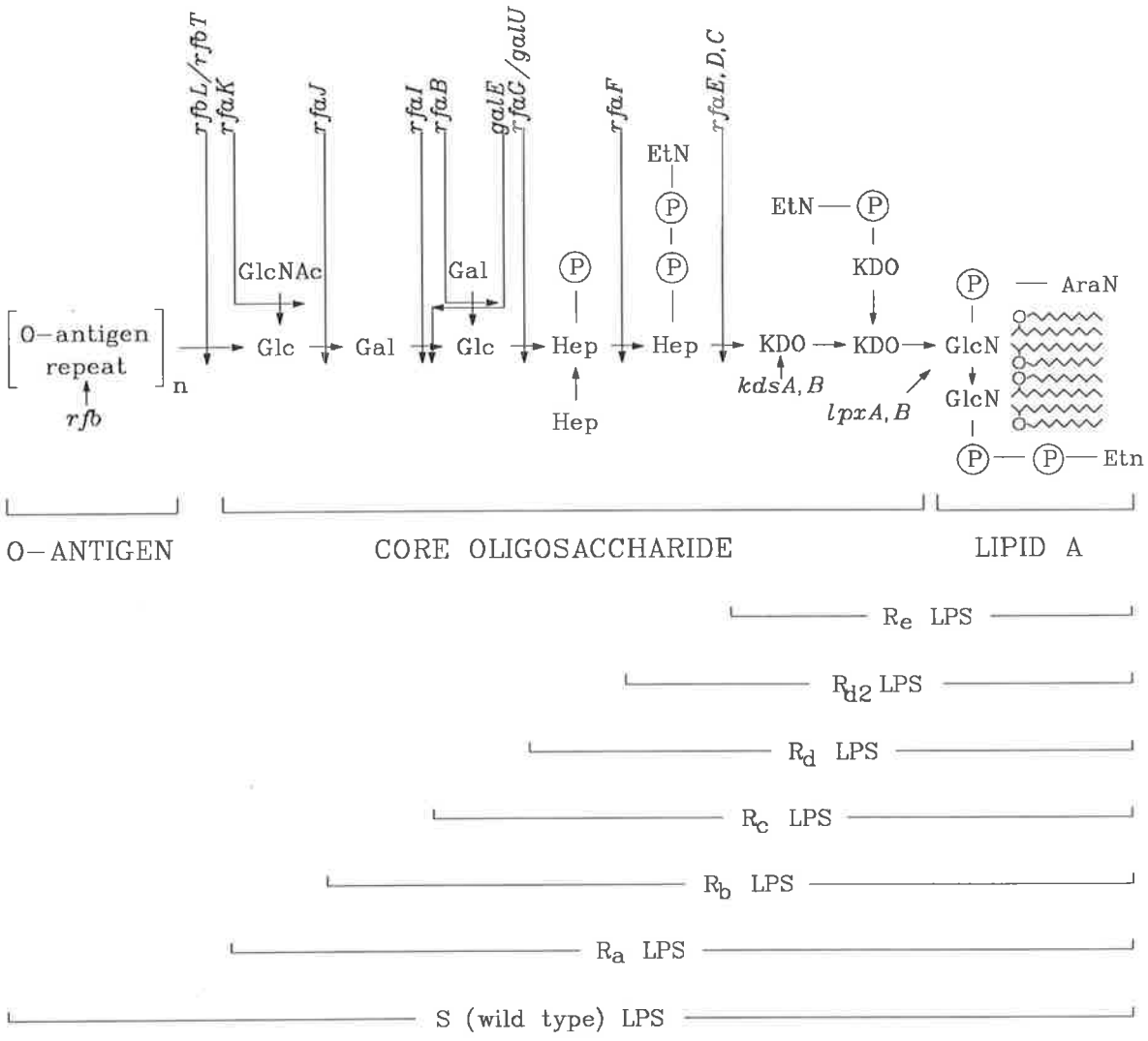
1.4.2.1 Inner core

An important tool in understanding LPS structure and biosynthetic pathways in *Salmonella typhimurium*, was the isolation of specific LPS mutant strains, lacking the ability to produce certain nucleotide sugar precursors necessary for LPS synthesis (Mäkelä and Stocker, 1984). The incomplete LPS of mutant strains has proven useful for studying *in vitro* biosynthesis and structure since it can serve as an acceptor for the addition of individual carbohydrate residues (Figure 1.5). The mutations occur primarily in two gene clusters, rough A (*rfa*), map position 79 min, and rough B (*rfb*), map position 42 min. There are also general "housekeeping" genes involved in synthesis of both the core and the O-antigen region (Kuo and Stocker, 1972; Sanderson and Saeed, 1972a,b).

3-deoxy-D-manno-2-octolosonic acid (KDO) residues play an important role in the synthesis and assembly of LPS by serving as a structural bridge between the lipid A and polysaccharide components. This acidic sugar was first detected by Levin and Racker (1959), and has since been shown by many workers to be an integral part of LPS in different bacteria (Ellwood, 1970; Osborn, 1963).

Figure 1.5 Definition of chemotype variants and genetic loci involved in the biosynthesis of the *S. typhimurium* LPS molecule.

Schematic diagram of the smooth LPS of *S. typhimurium*. Key: Glc, D-glucose; Gal, D-galactose; GlcN, D-glucosamine; GlcNAc, N-acetyl-D-glucosamine; Hep, L-glycero-D-manno-heptose; KDO, 2-keto-3-deoxy-D-manno-octulosonic acid; AraN, 4-amino-L-arabinose; P, phosphate; EtN, ethanolamine; ⩙, hydroxy and nonhydroxy fatty acids; Ra to Re indicate incomplete R from lipopolysaccharides. The arrows with the *rfa*, *rfb*, *lpxA, B*, *kdsA, B* and *gal* genes show the location of the enzymatic steps encoded by these genes. Diagram adapted from Hitchcock *et al.*, (1986); Lüderitz *et al.*, (1983); Raetz (1990).


rfbL/rfbT
rfaK
rfaJ
rfaI
rfaB
galE
rfaC/galU
rfaF
rfaE,D,C
EtN
GlcNAc
Gal
KDO
AraN
O-antigen repeat
n
rfb
Glc
Gal
Glc
Hep
Hep
Hep
KDO
KDO
GlcN
GlcN
Etn
kdsA,B
lpxA,B
O-ANTIGEN
CORE OLIGOSACCHARIDE
LIPID A
Re LPS
Rd2 LPS
Rd LPS
Rc LPS
Rb LPS
Ra LPS
S (wild type) LPS

KDO occurs as a trisaccharide in which the terminal KDO residue can be substituted with heptose (Brade *et al.*, 1983; Brade and Rietschel, 1984) (Figure 1.5). KDO is synthesized *in vivo* by three enzymatic steps:

(1) D-ribulose-5-phosphate → D-arabinose-5-phosphate

(2) D-arabinose-5-phosphate + phophoenolpyruvate → KDO-8-phosphate + P_i

(3) KDO-8-phosphate → KDO + P_i

The reactions are catalyzed by D-arabinose-5-phosphate isomerase, KDO-8-P-synthetase and KDO-CMP-synthetase, respectively.

The genes encoding KDO-8-P-synthetase (*kdsA*) and KDO-CMP-synthetase (*kdsB*) have been identified using temperature sensitive mutants defective in KDO synthesis (Rick and Osborn, 1977; Rick and Young, 1982). A *kdsA* mutant has been used by complementation, to clone and analyze the genetic structure of the *kdsA* gene from *E. coli* (Woisetschläger and Högenauer, 1986, 1987).

Eidels and Osborn (1971, 1974) have studied the biosynthesis of L-glycero-D-manno-heptose, using *S. typhimurium* mutants. The following pathway has been established:

(1) Sedoheptulose-7-phosphate ↔ D-glycero-D-manno-heptose-7-phosphate

(2) D-glycero-D-manno-heptose-7-phosphate ↔ D-glycero-D-manno-heptose-1-phosphate

(3) D-glycero-D-manno-heptose-1-phosphate + NTP → NDP-D-glycero-D-manno-heptose + P-P_i

(4) NDP-D-glycero-D-manno-heptose → NDP-L-glycero-D-manno-heptose

(Rick, 1987).

A number of *rfa* genes have been described which are associated with the biosynthesis of the heptose region of the inner core. Strains producing a mutant Rd2 type LPS, were identified as having a mutation in the *rfaF* gene, implying that it is the heptosyltransferase II enzyme (Figure 1.5) (Kuo and Stocker, 1972; Sanderson *et al.*, 1974; Sanderson and Saeed, 1972a, 1972b).

Heptoseless (Re) LPS can be produced by mutations in any of the genes *rfaC*, *rfaD* and *rfaE*. The following roles have been assigned; *rfaC* - encodes the heptosyltransferase I enzyme; *rfaD* - encodes a ADP-heptose epimerase (recently cloned on a 1.3 kb fragment and mapped to 80 min on the *E. coli* chromosome) (Pegues *et al.*, 1990); and *rfaE* - postulated to encode a second heptosyltransferase (Coleman and Deshpande, 1985; Mäkelä and Stocker, 1984; Sanderson *et al.*, 1974). The product of another locus, *rfaP*, is involved in the addition of the first glucose residue to the outer core (Hämmerling *et al.*, 1973; Jousimies and Mäkelä, 1974; Mühlradt, 1970) (Figure 1.5).

1.4.2.2 The outer core

Most *Salmonella* possess an outer core consisting of 5 saccharide units (Jann and Jann, 1984; Jansson *et al.*, 1981), catalyzed by 5 transferases mapping in the *rfa* gene cluster; *rfaB*, *rfaG*, *rfaI*, *rfaJ*, *rfaK* (shown in Figure 1.5). The biosynthetic precursors are common cellular compounds such as UDP-glucose (UDP-Glc), UDP-galactose (UDP-Gal) and UDP-N-GlcNAc all of which have "housekeeping" functions. Therefore it is not surprising that the structural genes for the enzymes associated with their synthesis are not closely linked to genes involved in LPS core synthesis (*galE*, *galU* and *pmi*) (Osborn, 1968; Wollin *et al.*, 1983). In the model for core synthesis, the LPS molecules are anchored via lipid A in the cytoplasmic membrane, on which they move along, facilitating contact with the glycosyl-transferases which transfer each sugar from its UDP precursor. Four enzymes, glucosyl-transferase I (*rfaG*) (Rothfield *et al.*, 1964), α-1,3-galactosyl-transferase (*rfaB*) (Osborn *et al.*, 1962), glucosyl-transferase II (*rfaJ*) (Osborn and Rothfield, 1971) and N-acetylglucosaminyl-transferase (*rfaK*) (Osborn and D'Ari, 1964), have been identified in this process by studying the incomplete LPS produced by various core mutant strains.

Apart from *rfaK*, these *rfa* genes have been cloned from *E. coli* K-12 and *S. typhimurium*. Although the gene order has been determined, the operon structure still

remains undefined (Creeger and Rothfield, 1982; Kadam *et al.*, 1985). Mutants in *rfaI* are deficient in the addition of the first galactose unit of the core oligosaccharide, which therefore encodes galactosyl-transferase I. The *rfaL* gene encodes a component of the O-antigen translocase/ligase system (Creeger and Rothfield, 1979; Kadam *et al.*, 1985; Mäkelä and Stocker, 1984; Rick, 1987).

The expression of this region appears to be regulated by the *rfaH* gene which is not linked to the *rfa* gene cluster and maps at 84 min on the *S. typhimurium* chromosome. This gene is associated with the regulation of not only the *rfa* genes but also the *tra* genes of the F factor (Sanderson and Stocker, 1981). *rfaH* is homologous to the *sfrB* gene of *E. coli* K-12, which acts as a transcription anti-terminator on the *tra* operon (Beutin *et al.*, 1981). Creeger and coworkers (1984) have shown that the *rfaH* gene acts as a positive regulatory element, an antiterminator, for the expression of the transferase genes *rfaJ*, *rfaK*, and also for *rfaL*. The termination in this operon is not complete, as read through from the termination loop occurs even in the absence of *rfaH*.

1.4.3 O-antigen structure and biosynthesis

The O-antigen is a linear polymer of a number of repeating oligosaccharide units. Individual repeat units may be either linear trisaccharides, pentasaccharides, or branched oligosaccharides with 4 to 6 sugars. Bacteria can be distinguished immunologically by differences in their O-antigenic structure, which can be between serogroups, or within a given O serogroup.

O-antigen synthesis has been best studied in *Salmonella* and differs markedly to that described above for the core region. The cyclical formation of the O-antigen repeat from a sugar nucleotide precursor, and its polymerization requires two molecules of bactoprenyl-pyrophosphate (antigen carrier lipid: ACL), a 55-carbon polyisoprenoid that is also essential for synthesis of peptidoglycan and teichoic acids (Goldmann and Strominger, 1972; Scher *et al.*, 1968; Wright *et al.*, 1967). When the O-unit is completed,

it is transferred to the lipid A core acceptor, with ACL being recycled. The final step is the translocation of the core-O-antigen oligosaccharide to the outside of the cell.

The *S. typhimurium* O side chain is composed of four sugars, D-galactose, D-mannose, and L-rhamnose with abequose attached to the mannose in an α-1,2 linkage. The formation of this O-unit uses the di-nucleotide sugar precursors; UDP-Gal, GDP-mannose (GDP-Man), TDP-rhamnose (TDP-Rha) and CDP-abequose (CDP-Abe) respectively (Kauffman, 1961; Nikaido, 1973). The genes coding for the biosynthesis and assembly of O-antigen are located mostly in the *rfb* gene cluster between *his* at 44 min and *metG* at 46 min (Sanderson and Hurley, 1987; Sanderson and Roth, 1983). This locus encodes 9 of the enzymes involved in the synthesis of GDP-Man, TDP-Rha, and CDP-Abe including the required glycosyl-transferases (Brahmbatt *et al.*, 1986, 1988; Nikaido *et al.*, 1967; Rick, 1987). The *galU* and *galE* genes needed for UDP-Gal synthesis and *pmi*, encoding phosphomannose-isomerase, are located outside this region.

Genetic analysis of the *rfb* region has been limited by the instability of point mutants making complementation analysis not feasible. Using the *his* deletion mutations of Hartman *et al.* (1960), Nikaido and coworkers (1967) established a partial gene order for this region, which was subsequently used as a basis for the cloning of the *rfb* region by Brahmbatt and coworkers (1986). DNA analyses of this region have shown that the *rfb* gene cluster may be between 20 to 21 kb in length (H.N. Brahmbatt, PhD thesis, University of Adelaide, 1989), and the direction of transcription of this operon is towards *his*. Analysis of the promoters of this region has revealed one major promoter, however, the possibility of weak internal promoters and terminators cannot be ruled out. It is assumed that the *rfb* region is a single operon (H.N. Brahmbatt, PhD thesis, University of Adelaide, 1989). The *rfb* region was subdivided into three segments for preliminary protein analysis, using *E. coli* minicells, enabling the identification of 16 proteins, the functions of which have yet to be defined.

Variation in O-antigen is thought largely to be due to the result of variation in the *rfb* gene cluster (Mäkelä and Stocker, 1984). The O polysaccharide of *S. typhi*, *S. paratyphi*, and *S. typhimurium* have identical trisaccharide subunit backbones

(mannosyl-rhamnosyl-galactose), however these chains may be distinguished, as each has a different 3,6-dideoxyhexose: tyvelose, paratose or abequose, attached to the mannosyl residues as a side branch (Jann and Jann, 1984; Mäkelä and Stocker, 1984; Verma *et al.*, 1988; Verma and Reeves, 1989; Wyk and Reeves, 1989). The biosynthetic pathways for these three sugars diverge only at the last step.

Thus to summarize the names and functions of the genes involved in O-antigen synthesis:

(1) TDP-Rha synthesis is associated with the genes; *rfbA*, *rfbB*, *rfbD*.

(2) GDP-Man synthesis and incorporation is encoded by genes; *rfbL*, *rfbK*, *rfbM*.

(3) CDP-Abe synthesis involves five genes; *rfbF*, *rfbG*, *rfbH*, *rfbI*, *rfbJ*.

(4) Paratose synthase which replaces abequose synthase in *S. paratyphi* is encoded by *rfbS*.

(5) CDP-tyvelose-2-epimerase, an additional enzyme present in *S. typhi*, is encoded by *rfbE*. This gene is present in a mutant, non-active form in *S. paratyphi*.

1.4.4 Polymerization and translocation

After the individual O-antigen units are synthesized on the ACL, they are then polymerized into a long polysaccharide chain while still remaining attached to the carrier lipid (Robbins *et al.*, 1966; Weiner *et al.*, 1965). This occurs at the surface of the cytoplasmic membrane, with both the intermediates and the enzymes being membrane bound.

There are two different polymerase systems described for O-antigen polymerization. The first is the *rfe* dependent system of *Salmonella* group C and L, and *E. coli* 08 and 09 (Mäkelä *et al.*, 1970; Stocker and Mäkelä, 1978). *Salmonella* hybrids constructed by transferring the group B *rfb* region into group C_1 or C_2 produced O-antigen units that were not polymerized (Hämmerling *et al.*, 1971; Mäkelä, 1966; Naide *et al.*, 1965). This type of polymerization does not use ACL, but α-glucosyl-diphospho-undecaprenol as an acceptor for mannose in direct transfer from GDP-Man

(Jann *et al.*, 1982). The *rfe* gene which maps close to the *ilv* locus on the *Salmonella* chromosome is also involved in biosynthesis of the Enterobacterial common antigen (ECA) (Mäkelä and Mayer, 1974; Mäkelä and Stocker, 1969; Schmidt *et al.*, 1976).

The second system, the *rfe* independent system, requiring the *rfc* gene, is found in *Salmonella* groups A, B, D and E. The above experiment was reversed such that the group C_1 *rfb* region was transferred to group B. Surprisingly, this resulted in synthesis of complete long chain LPS. The interpretation of these results was that the polymerase of group C_1 is encoded by a gene in the *rfb* cluster, and in the group B *Salmonella* a different polymerase gene, the *rfc* gene, is required. The *rfc* gene has been mapped between *gal* at 18 min and *trp* at 34 min on the *Salmonella* chromosome (Naide *et al.*, 1965; Sanderson and Hurley, 1987) and has recently been cloned and sequenced (L.V. Collins, PhD thesis, University of Adelaide, 1990).

The next stage of biosynthesis, the transfer of the O-polysaccharide to the lipid-A-core region involves the two genes *rfaL* and *rfbT*. It is proposed that the *rfaL* gene encodes an enzyme that recognizes the core in *Salmonella* LPS and therefore common to all *Salmonellae*, whereas the *rfbT* gene which encodes an enzyme to bind the completed O-units, would be specific for each O group (Mäkelä and Stocker, 1984; Rick, 1987). When O-antigen polymerase is rendered defective by an *rfc* mutation, the ligase attaches a single O-antigen repeat to the outer core (Naide *et al.*, 1965; Rick, 1987).

The degree of O-antigen polymerization is dependent on growth temperature and carbon source (McConnell and Wright, 1979; Nikaido, 1979). This is shown by O-antigen side chain length heterogeneity in purified LPS preparations (Munford *et al.*, 1980; Pavla and Mäkelä, 1980). The coordination of enzymes involved in O-antigen polymerization, ligation to the core and translocation is poor. Mutants producing a defective core still produce O-antigen linked to ACL (Kent and Osborn, 1968a, 1968b), and *rfb* mutants, still transport normal amounts of LPS core, even in the absence of synthesis of O-repeat units (Smit *et al.*, 1975).

The final step in the biosynthetic sequence of LPS is catalyzed by the enzyme O-antigen ligase or translocase. This enzyme catalyzes the transfer of the O-repeat unit

from the ACL to the lipid A core linking the terminal reducing galactose to the nonreducing terminal glucose of the core via an α-1,4 linkage.

1.4.5 Topology of LPS biosynthesis

The topography of LPS biosynthesis is still poorly understood. In 1972, Osborn and colleagues, showed by both *in vivo* pulse chase experiments and isolated membrane fractions, that the enzymes that synthesize core and O-antigen, and transfer the O-antigen to the core, were localized to the cytoplasmic membrane (Osborn *et al.*, 1972). These enzymes must function on the inner surface of the inner membrane, as sugar nucleotides, ATP, and ACL are required. This step was shown to be rapid and irreversible. Eleven years later Mulford and Osborn (1983), using immunoelectron microscopy, in core defective mutants, demonstrated the presence of newly synthesized LPS, at the periplasmic surface of the inner membrane before translocation to the outer membrane.

Four mechanisms for LPS translocation have been considered by Osborn (1979) and Raetz (1990), of which two require a flipping of O-antigen, core-lipid A, and/or mature LPS across the lipid bilayers, another involves budding off from the inner membrane to periplasmic space, culminating in fusion with the outer membrane, and the last alternative, involves Bayer bridges (Bayer, 1975). However, none of these mechanisms have been proven.

1.5 Lipopolysaccharides of other bacteria

Although most of the genetic and structural studies relating to LPS biosynthesis have been done using *Salmonella* as the role organism, LPS from other bacteria have been isolated and a limited number of *rfa* and *rfb* regions identified, cloned and analyzed.

1.5.1 *Escherichia coli*

Five different core oligosaccharide structures are found in *E. coli* strains. These differ from core structures found in *Salmonellae* predominantly in the hexose region, (Jansson *et al.*, 1981; Mäkelä and Stocker, 1984), however, the locus can complement certain *rfa* genes of *S. typhimurium* (Mayer *et al.*, 1976). The core region of *E. coli* is encoded by the *rfa* gene cluster that maps at 81 minutes near *pyrE* and *cysE* (Bachmann and Low, 1980).

E. coli O-antigen has many structural and genetical similarites to that of *Salmonella.* The *E. coli rfb* region also maps at the same relative position on the chromosome - linked to the *his* region at 44.5 minutes (Bachmann and Low, 1980; Ørskov and Ørskov, 1962; Schmidt *et al.*, 1969; Valvano and Crosa, 1989). The *rfb* regions from a number of serotypes have been cloned into *E. coli* K-12. This strain is a good background strain, since it does not synthesize an O-polysaccharide but can still supply the lipid-A-core component, and thereby support the synthesis of a complete LPS molecule if functional *rfb* genes are provided (Mäkelä and Stocker, 1984; Manning *et al.*, 1986; Valvano and Crosa, 1989).

In this way the *rfb* region from the following *E. coli* serotypes have been cloned:

(1) *E. coli* O9 (Kido *et al.*, 1989). Deletion analysis of this clone produced a strain expressing antigenically distinct O-antigen, as the deleted fragment of the clone was responsible for the synthesis of α-1,3 linkages of the O9-specific polysaccharide. This resulted in an oligosaccharide unit combined through α-1,2 linkages (Kido *et al.*, 1989).

(2) *E. coli* O7 (Gupta *et al.*, 1984; L'vov *et al.*, 1984; Valvano and Crosa, 1989). A 17 kb clone was isolated which contained DNA that was essential but not sufficient for full expression of O7 LPS (a branched structure of five different sugars) in *E. coli* K-12.

(3) *E. coli* O4 (Haraguchi *et al.*, 1989). This resulted in a 6 kb clone encoding 8 proteins, 39 kDa to 13 kDa. As this was a relatively small clone for O-antigen expression, it is thought that *E. coli* K-12 itself provides some *rfb* functions.

(4) *E. coli* O1O1 (Heuzenroder *et al.*, 1989). Obtaining this clone enabled hybridizations between different serotype variants of O2 and O18 *E. coli* as well as *Shigella flexneri* to be performed, which revealed homologous DNA sequences.

Unlike other *E. coli* strains which possess a chromosomally encoded *rfb* region, the O-antigen produced by *E. coli* O111 requires the presence of a 54 MDa plasmid (Riley *et al.*, 1987).

1.5.2 *Shigella*

Involvement of plasmid encoded factors in the expression of LPS has also been seen in *Shigellae*. *Shigella sonnei* strains produce only a single O-antigen type, however, its LPS can exist in 2 forms. Form I *S. sonnei* synthesize O-antigen and are smooth, virulent strains. Form II derivatives are rough avirulent strains (Keene *et al.*, 1980; Seid *et al.*, 1984).

Production of the Form I antigen is encoded on a large non-self-transmissible plasmid (Formal *et al.*, 1981; Sansonetti *et al.*, 1981). This plasmid is unstable due to its size and may be lost at frequencies ranging from 1% to 50% depending upon the strain.

A structural difference between the cores of *Salmonella* and *S. sonnei* was shown by transferring the Form I O-antigen into *S. typhi* strain Ty21a. Although the Form I antigen was expressed in Ty21a, it was not attached to the core, but existed as a polymer on the cell surface (Formal *et al.*, 1981; Seid *et al.*, 1984).

The biosynthesis of O-polysaccharide of *S. dysenteriae* 1 was found to be determined by both plasmid (pHW400) and chromosomal genes (containing the *his/rfb* region) (Hale *et al.*, 1984; Sturm and Timmis, 1986a; Watanabe and Timmis, 1984). At least one step of O-polysaccharide biosynthesis in this serogroup is specified by a 9 kb plasmid, which carries a gene called *rfp*. This gene catalyzes the transfer of the first O-repeat unit to the LPS core (Mills *et al.*, 1988; Sturm *et al.*, 1986a). Analysis of the chromosomal segment which has been localized to a minimal size of 6.4 kb, has revealed 6-7 genes organized as two transcriptional units (Sturm *et al.*, 1986b).

Construction of a hybrid clone based on pACYC184 containing the *rfp* gene together with the chromosomal region and its subsequent transfer to *E. coli* K-12, enabled the production of *S. dysenteriae* 1 LPS (Mills *et al.*, 1988). However, as expression of *S. dysenteriae* 1 LPS in *E. coli* K-12 revealed LPS intermediates on the cell surface, a different method of O-antigen synthesis in *S. dysenteriae* 1 to that of *Salmonella* and *E. coli* may be present, in which repeat units are polymerized on the ACL and then transferred to the lipid A-core (Sturm *et al.*, 1986b).

The genetics of *S. flexneri* have been compared with both *E. coli* and *Salmonella* revealing a common mode of synthesis. The *S. flexneri* genes for LPS biosynthesis are chromosomally encoded (Formal *et al.*, 1970; Simmons, 1971). However, *S. flexneri* differs in that after the O-side chain is polymerized, it may be modified by either trans-glucosylation and/or trans-O-acetylation. These modifications are induced by temperate bacteriophages that cause lysogenic conversion of the O-antigen (Lindberg *et al.*, 1978). A locus that maps at 6 minutes near *lac* on the chromosome, controls the glucosylation of the basic Y antigen, to form type specific antigens I, II, IV, V and antigen 7 and 8. Another unmapped locus is involved with the addition of O-acetyl residues to produce antigens III and 6.

1.5.3 Lipooligosaccharides

Interest in the role of LPS in virulence of non enteric pathogens such as *Haemophilus influenzae*, *Bordetella pertussis*, *Neisseria meninigitidis* and *Neisseria gonorrhoeae*, has provided structural and genetical data on the LPS molecule. These LPSs are quite distinct from those previously mentioned, as they lack O-antigen (Inzana, 1983; Le Dur *et al.*, 1980; Schneider *et al.*, 1984).

The LPS or lipooligosaccharide (LOS) of *H. influenzae*, is composed of a short chain of covalently linked neutral sugars, a structure analagous to the core oligosaccharide of enteric LPS molecules (Kimura and Hansen, 1986). Thus, unlike the

LPS of the Enterobacteriaceae, *H. influenzae* LOS does not possess any repeating oligosaccharide units.

The surface exposed portion of the *H. influenzae* LOS may demonstrate both interstrain and intrastrain variabilty (Tolan *et al.*, 1986; Zamze and Moxon, 1987), studied using MAbs (Weiser *et al.*, 1989b). Phase variation seen with this LOS has been shown to be independent of DNA recombinational events, and since the switching is between three discrete levels of expression (++++, +, and -), it was thought to be more complex than an on/off switch (Weiser *et al.*, 1989b).

Using MAbs specific for oligosaccharide epitopes, Weiser and colleagues (1989a) were able to clone a chromosomal locus (*lic-1*), on a 5.6 kb DNA fragment, responsible for expression of at least three LOS epitopes that displayed phase variation (Weiser *et al.*, 1989a).

The *lic-1* locus appears to be one transcriptional unit, encoding four proteins (LIC A-D) required for expression of at least two LOS structures that arise by phase variation (Weiser *et al.*, 1989b). The DNA has been sequenced, revealing multiple (~30) tandem repeats of a tetramer, CAAT, at the 5' end of the first ORF. Changing the number of copies of CAAT creates a translational shift by shifting three possible initiation codons in and out of frame with respect to the *licA* ORF. The cloning of two more regions *lic2* and *lic3* was achieved using a synthetic oligonucleotide to the repeat region (Weiser *et al.*, 1990). In contrast to *lic1* and *lic3* controlled epitopes, *lic2* controlled epitopes may or may not be expressed. This is explained by the discovery of only one translational phase with respect to an upstream initiation codon (Weiser *et al.*, 1990). However, a recently isolated clone expressing a *H. influenzae* LOS epitope in *E. coli* bears no resemblance to the previously described *lic* loci, as no CAAT repeats could be detected.

The structure of the LOS of *Neisseria* is similar to that of *H. influenzae*, however, the regulation of LOS expression in *Neisseria* is not as well understood.

A recombinant plasmid containing 9 kb of *N. gonorrhoeae* DNA has been isolated, which encodes an enzyme that modifies the *E. coli* core to produce a gonococcal-like epitope (Palermo *et al.*, 1987).

Another group, Petricoin and Stein (1989), isolated a clone reactive to LOS-specific MAbs which complemented *N. gonorrhoeae* strain FA5100, containing a LOS biosynthetic defect, such that reactivity to LOS MAbs was restored. It was postulated that the DNA encodes either a transferase to add the first sugar to the KDO, or a regulatory gene for LOS biosynthesis.

1.6 *Vibrio cholerae* lipopolysaccharide

1.6.1 Serology

As mentioned earlier (1.12) strains of both biotypes, El Tor and Classical of *V. cholerae* O1, can be subdivided into three serotypes, Inaba, Ogawa and Hikojima. This division was first proposed in 1918, by Kabeshima, who classified *V. cholerae* O1 into original (Inaba) and variant (Ogawa) types based on agglutination tests. Five years later Nobechi (1923) identified the Hikojima serotype as being a "middle type". The serotypes were designated the antigenic formulae of Inaba-AX, Ogawa-B(X) and Hikojima-A(B)X. This scheme was modified by Gardner and Venkatraman (1935) to Inaba-AX, Ogawa-BCX and Hikojima-ABX. This postulated the antigenic difference between the Ogawa and Inaba variant of *V. cholerae* to be quantitiative rather than qualitative. The introduction of a new antigen "C" was confirmed by Heiberg (1935), and Burrows *et al.* (1946a, 1946b), who also changed the antigenic formula to the one currently used today - Ogawa (AB), Inaba (AC), and Hikojima (ABC).

The chemical nature of the serotype specific antigens was first studied by Shousha in 1931. He showed that the antigens were associated with an acid and heat stable polysaccharide fraction, from which he concluded that the immunological specificity of *V. cholerae* serotypes lies in the O-antigen region of LPS. The same conclusion was also reached by Sakazaki and Tamura (1971), who used absorbed sera, and cross agglutination tests. Thus, it was proposed that the Ogawa cell has all the factors present in Inaba plus an unique factor. This implies that Ogawa cells have AB and a small amount of C

antigen, whereas Inaba cells have A, and a higher level of the C antigen (see Figure 1.1). The observation that absorption of an Ogawa antiserum with Inaba cells still leaves 5-10% of the activity against Ogawa cells, also lent credibility to this theory (Redmond *et al.*, 1973). The production of MAbs has and will help immensely in determining the serology of *V. cholerae* O1. Gustafsson and Holme (1985), have created a bank of MAbs, with the expected properties for reacting with A, B or C antigens as judged by ELISA inhibition and slide agglutination assays. Another set of MAbs (Ito and Yokota, 1987) raised against *V. cholerae* O1 Ogawa, contain antibodies specific for Ogawa determinants, but some MAbs react to an Ogawa-Inaba common antigenic determinants, presumably the A antigen.

1.6.2 Antigenic variation

In the original studies of *V. cholerae* serotypes (Kabeshima, 1918) "mutations" were seen in the serotypes of stored strains. Later Shrivastava and White (1947) were able to isolate Inaba variants from Ogawa strains, when they were grown in the presence of anti-Ogawa serum. As this could not be reproduced in the opposite direction, that is from Ogawa to Inaba, antigenic conversion was thought to occur only in one direction, due to the loss of the B antigen.

A detailed study using 13 different *V. cholerae* strains and monospecific sera was performed, and again it was suggested that *in vitro* variation occurred only from Ogawa to Inaba (Sakazaki and Tamura, 1971).

In vivo observations have also been reported in humans and germ-free mice. Gangarosa *et al.* (1967), reported that a human patient infected with Ogawa *V. cholerae*, expressed on different days Inaba and Hikojima serotypes, usually preceding a clinical relapse. However, subsequent analyses showed the Hikojima strain to be in fact, Ogawa.

Sack and Miller (1969), using germ free mice demonstrated the possibility of reciprocal conversions between Ogawa and Inaba serotypes, and also that smooth to rough changes can occur during infection. By using cyclophosphamide to suppress

antibody formation, this study inferred that *in vivo* serotype conversions were related to the presence of serotype specific antibodies.

Phage conversion leading to changes in the serotype of *V. cholerae* has been reported (Ogg *et al.*, 1978, 1979). Infection of the *V. cholerae* Classical Ogawa strain O29 with CP-T1 converted the Ogawa serotype to Hikojima, shown by increased levels of C expression using agglutination tests. Converted isolates were subcultured and the stability of these phage-converted properties examined. Loss of C antigen by unstable isolates was associated with alteration from lysogenic to a non-lysogenic state. Those isolates which maintained C antigen expression, released phage which had the same host range as CP-T1, and showed immunity to infection by CP-T1 (Ogg *et al.*, 1978).

Phage conversion was also studied with *V. cholerae* in El Tor Ogawa strains (Ogg *et al.*, 1979), and differences were noted. Phage induced seroconverted El Tor strains were less stable, resulting in a large proportion agglutinating in anti-rough sera. By comparison with Classical Ogawa strains, phage converted El Tor Ogawa strains which maintained the converted characteristics, although resistant to infection by CP-T1, could not be induced to release phage (Ogg *et al.*, 1979).

Analyses of the phage inactivating capacity of purified LPS from Inaba and Ogawa of both Classical and El Tor biotypes, resulted in the identification of the CP-T1 receptor to be O-antigen, complicating the previous findings (Guidolin and Manning, 1985, 1987). This implies that phage treatment may be selecting resistant mutants and not convertants.

Guidolin and Manning (1985) also described the inability to demonstrate lysogens of CP-T1. Using Southern hybridization, analyses of the "lysogenic" strain 1633 described by Ogg *et al.* (1978) failed to reveal any CP-T1 sequence within its genome. This strain was also not able to be induced to release CP-T1 phage (Guidiolin and Manning, 1985).

1.6.3 Lipid A

The structure and toxicity of the lipid A portion of *V. cholerae* LPS has been studied by a number of workers. In 1981, Broady and coworkers described the structure of lipid A (Figure 1.6), which showed both common and distinct features when compared to lipid A from species of Enterobacteriaceae.

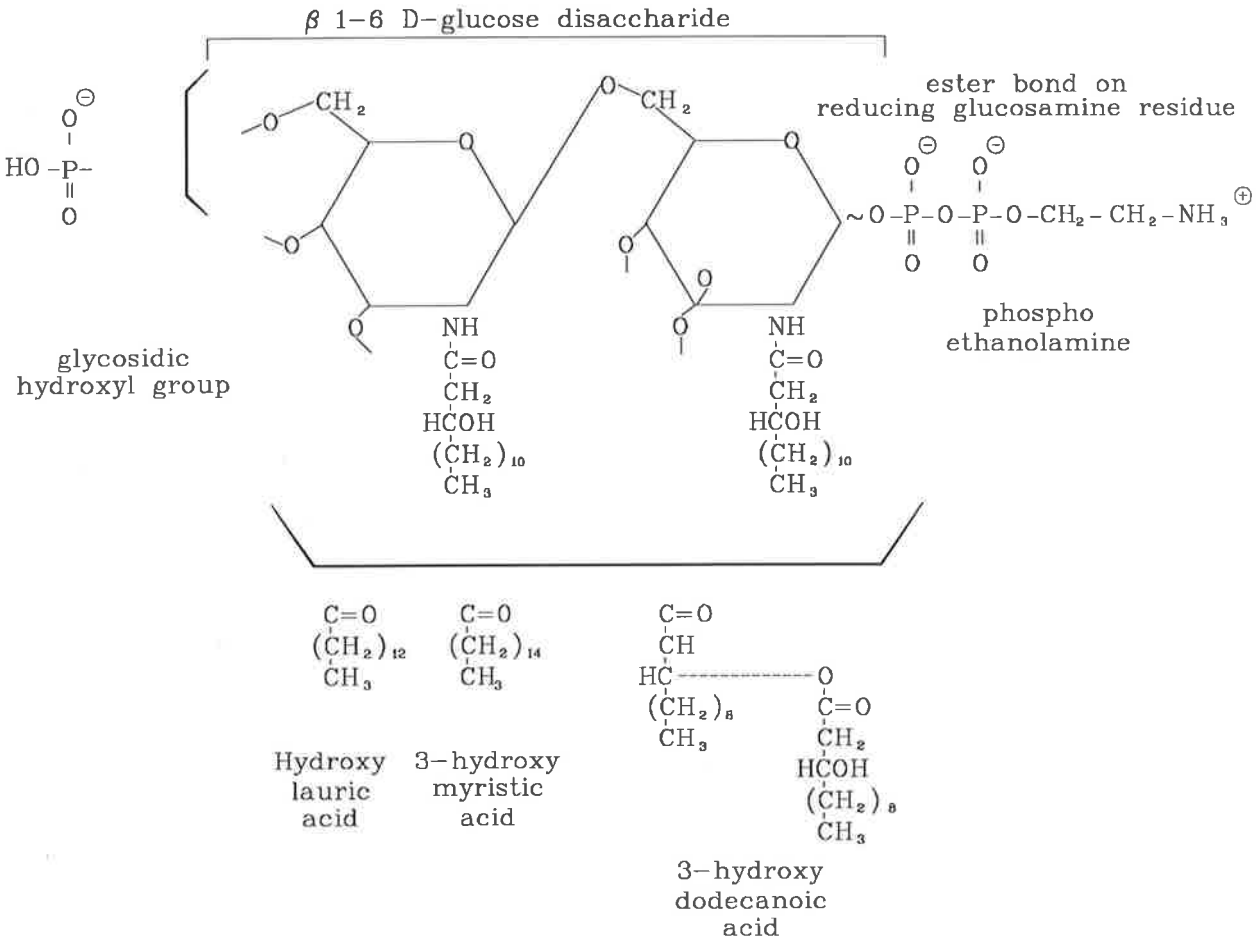
The lipid A component has been found to be similar to other Gram-negative lipid A moieties in that it consists of a β-1,6-linked D-glucosamine disaccharide that has an ester linked phosphate group on its nonreducing glucosamine residue, and another phosphate residue on the glycosidic hydroxyl group (Broady *et al.*, 1981; Raziuddin, 1977; Rietschel *et al.*, 1984). The phosphate group bound to the C-1 of the reducing glucosamine residue may be substituted by phosphorylethanolamine (Broady *et al.*, 1981). The possibility that phosphate bridges linked the lipid A units was rejected, by using ^{31}P nuclear magnetic resonance spectroscopy (Broady *et al.*, 1981).

V. cholerae lipid A has been shown to contain tetradecanoic, hexadecanoic and 3-D-hydroxydodecanoic (3-hydroxylauric) acids in ester linkages, and 3-D-hydroxytetradecanoic acid (3-hydroxymyristic acid) in amide linkages (Hase and Rietschel, 1976; Rietschel, 1976). The 3-hydoxymyristic acid which is the longest 3-hydroxy fatty acid in *V. cholerae* LPS has been found in the LPS of all Gram-negative organisms analyzed with the exception of *Brucella* (Hisatsune *et al.*, 1979). There is also less of hexadecenoic, stearic, and oleic acids in ester linkages. Removal of the ester-linked fatty acids by alkaline treatment of LPS, showed that these fatty acids play a specific role in toxicity for mice and in anti-complement activity (Lüderitz *et al.*, 1973; Raziuddin, 1979). The biological activity of the lipid A has also been studied with endotoxins that have been chemically modified by succinic or phthalic anhydrides (Raziuddin, 1980).

Hisatsune and colleagues (1979), found an unusually high amount of both odd numbered nonhydroxy and 3-hydroxy fatty acids in the LPS of an Inaba strain, especially when compared to the amount detected in Ogawa and NAG strains.

Figure 1.6 *V. cholerae* lipid A.

Chemical structure of lipid A from *V. cholerae* LPS. Figure is modified from Broady *et al.* (1981).


β 1–6 D–glucose disaccharide
ester bond on reducing glucosamine residue
glycosidic hydroxyl group
phospho ethanolamine
Hydroxy lauric acid
3–hydroxy myristic acid
3–hydroxy dodecanoic acid

1.6.4 Core oligosaccharide

Although numerous studies have been performed on the chemical structure of *V. cholerae* O1 LPS, neither the core oligosaccharide nor O-antigen side chain have been assigned a definitive structure.

The LPS of *V. cholerae* O1 has been separated into two regions consisting of the lipid A, and the polysaccharide region, by various chemical methods. In 1977, Raziuddin was able to chromatographically separate LPS into two fractions, I and II. Analysis of these two fractions revealed the presence of glucose, mannose, rhamnose, fructose, glucosamine and D-perosamine. Fraction I did not contain heptose and phosphorous which were detected in fraction II, which also had higher levels of glucose and fructose. It was proposed that the fraction I was the side chain polysaccharide whereas fraction II was the core oligosaccharide region.

Kenne and coworkers (1982), also separated the LPS fraction by acid hydrolysis into lipid A and polysaccharide. The polysaccharide fraction was then further separated by gel filtration into a M_r 9000 (major) and M_r 900 (minor) fraction. The use of MAbs in an ELISA inhibition test made it possible to determine that the M_r 900 fragment represented the core region, while the M_r 9000 fragment corresponded to the complete polysaccharide chain and core region (Gustaffson and Holme, 1985).

In 1973, Shimada and Sakazaki studied the R antigen of *V. cholerae*, comparing the serological properties and colony morphology of R strains and their parent S strains. They concluded that identification of rough compared to smooth strains with R antiserum was possible, however, it was impossible to identify the original serotypes of the R cultures. Antisera raised against various rough strains, showed cross reactivity to other R strains, indicating a common "rough" antigen (Hisatsune *et al.*, 1978).

There is still conjecture over the involvement of KDO in the LPS of *V. cholerae.* Originally it was thought that there was no KDO in the composition of *V. cholerae* O1 LPS (Raziuddin, 1977, 1980). Kabir (1982), and Kenne *et al.* (1982), postulated that the link between lipid A and the core region was via a fructose moiety, even though acid hydrolysis of LPS, which normally occurs by cleaving the glycosidic linkages of KDO,

splits *V. cholerae* O1 LPS into lipid A and polysaccharide. It was later shown by quantitative fructose release from LPS, that fructose is not involved in the linkage between core and lipid A (Kaca *et al.*, 1986). In 1985, Brade demonstrated the presence of KDO in *V. cholerae* LPS. It has also been shown that an antigen with common LPS specificity has been found in many of the Gram-negative bacteria including *V. cholerae* (Brade and Galanos, 1983). The determinant contains both KDO and a sugar.

Recently, it has been postulated that *V. cholerae* may possess two different core structures, one bearing a KDO molecule, and the other lacking KDO (E.Th. Rietschel, personal communication). If correct, this would explain the incongruity seen over core structures (Brade, 1985; Kabir, 1982; Raziuddin, 1977, 1980; Redmond, 1976).

1.6.5 O-antigen

The O-specific side chain of *V. cholerae* LPS has been shown to possess several unusual amino sugars, including 2-amino-2,6-dideoxy-D-glucose (quinovosamine), 4-amino-4-deoxy-L-arabinose, and 4-amino-4,6-dideoxy-D-mannose (perosamine) (Figure 1.7). A number of common sugars can also be found in the polysaccharide portion of LPS, such as mannose, glucose and heptose, but of note is the absence of galactose. Using an Inaba strain, Sen *et al.* (1980) showed that the major heptose was D-glycero-L-manno-heptose and the minor one, D-glycero-L-gluco-heptose. However, Majumdar *et al.* (1983), showed D-glycero-L-gluco-heptose to be the major heptose in an Ogawa strain G-2102. Guhathakurta *et al.* (1986), proposed the involvement of these two sugars in determining the subtype of *V. cholerae* O1, in conjunction with glucuronic acid and glucosamine.

There are still discrepancies about some elements of the *V. cholerae* LPS, but the presence of perosamine as a major constituent has been agreed upon. A high molecular weight polymer was identified in both Inaba and Ogawa serotypes of *V. cholerae* O1 and shown by NMR spectroscopy to consist of about 60 perosamine residues in a regular α-1,2-linked linear chain (Jackson and Redmond, 1971; Redmond, 1975, 1979; Redmond

Figure 1.7 Structural formulae for the amino sugars in *V. cholerae* LPS.

A) Perosamine - 4-NH_2-4,6-dideoxy-mannose

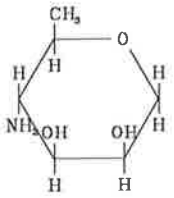
B) Quinovosamine - 2-NH-2,6-dideoxy-D-glucose

C) 4-NH_2-4-deoxy-L-arabinose

D) Repeat unit of *V. cholerae* O-antigen, comprised of (1→2) linked 4-amino-4,6-dideoxy-α-D-manno-pyranosol residues. The amino groups are acylated with 3-deoxy-L-tetronic acid. Figure is reproduced from Kenne *et al.* (1982).

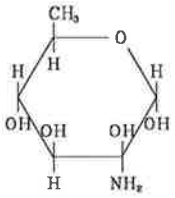
A

PEROSAMINE
4−NH_2 −4,6−DIDEOXY−MANNOSE



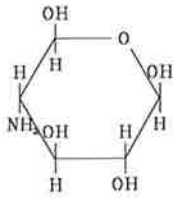
B

QUINOVOSAMINE
2−NH−2,6−DIDEOXY−D−GLUCOSE



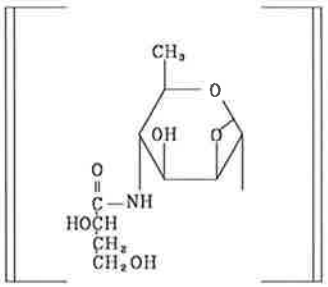
C

4−NH_2 −4−DEOXY−L−ARABINOSE



D

L−GLYCEROL TETRONIC ACID
AND PEROSAMINE



n

et al., 1973). This "backbone" structure was confirmed by both Kenne *et al.* (1979, 1982), and by Hisatsune *et al.* (1985), who showed that the amino groups of the perosamine residues are acylated with 3-deoxy-L-glycerotetronic acid (Figure 1.7).

Earlier Sen *et al.* (1979, 1980), had also postulated the perosamine polymer with glucose and heptose residues as terminal sugars at the nonreducing end. Using lectins, Kabir (1982), showed the linkage of the glucose residues to be α-linked, and that N-acetyl-D-glucose was a terminal sugar linked to the polysaccharide backbone by an α1,3 bond.

Comparisons of the LPSs produced by rough and smooth strains showed the total elimination of two of the amino sugar components, perosamine and quinovosamine (Hisatsune and Kondo, 1980). Quinovosamine was first isolated and identified by Jann *et al.* (1973). By fractionating the polysaccharide portion into core polysaccharide and O-antigen fractions it was shown that quinovosamine was present only in the O-antigen fraction of LPS (Kabir, 1982; Raziuddin, 1980). Studies using both Ogawa and Inaba serotypes have not revealed any significant differences in the amounts of both perosamine and quinovosamine in the LPS (Hisatsune *et al.*, 1985). Hisatsune *et al.* (1985), proposed a structure for the *V. cholerae* O-antigen, consisting of repeating units of perosamine, connected to the core via a quinovosamine disaccharide unit.

Redmond (1978) first detected 4-amino-4-deoxy-L-arabinose in the LPS of Ogawa cells by use of NMR spectroscopy. The presence of this particular sugar in *V. cholerae* LPS has not been confirmed by experiments performed by Hisatsune and Kondo (1980), however other studies have confirmed this particular sugar to be unique to the Ogawa cell (Kabir, 1982; Kenne, *et al.*, 1982). Therefore it has been proposed that this sugar is associated with the Ogawa specific B antigenic determinant.

The use of MAbs produced against *V. cholerae* LPS has helped in serotyping strains, and can now be used in identifying determinants specific for each serotype (B. Gustafsson, PhD thesis, Stockholm, 1985).

1.6.6 Lipopolysaccharide from other *Vibrio* species

In addition to *V. cholerae* O1 which causes the disease cholera in man, there are at least 72 additional serogroups collectively known as either nonagglutinating (NAG), non-cholera vibrios (NCV), or non O1 (Ansari *et al.*, 1986). Several studies have analyzed the LPSs of these other vibrios. Guhathakurta *et al.* (1980), examined strain NAG 108, and compared its LPS constituents with those present in Inaba and Ogawa strains. The NAG LPS contained high levels of rhamnose, glucosamine and lower levels of mannose, whereas the Ogawa strain had less, and the Inaba, contradictory to other groups as previously mentioned, possessed no rhamnose. The Ogawa and Inaba strains both had moderate amounts of mannose and glucose.

Sen and Mukerjee (1978), have studied another isolate, *V. cholerae* type *NAG NV* 384. This strain produced a mannose rich LPS, which comprised a phosphoric diester-linked polysaccharide containing mainly (1→2)-linked mannopyranose residues, highly substituted with other sugar residues.

Ansari *et al.* (1986), have analyzed a non O1 LPS produced by *V. cholerae* O:21. It was shown to be composed of tetrasaccharide repeating units each comprising one residue of 2-acetamido-2-deoxy-D-glucose, 2-acetamido-2-deoxy-D-galactose, L-rhamnose and D-glycero-D-manno-heptose. The latter sugar has been reported previously in core oligosaccharides but never in the O-antigen unit of LPS.

LPS from the strains *V. cholerae* bio-serogroup Hakata, *V. fluvialis* Kobe, and *Vibrio* bio-serogroup 1875 have been analyzed and a number of common and unique determinants have been identified (Histsune *et al.*, 1986, 1987). *V.* bio-serogroup 1875 which agglutinated with *V. cholerae* O1 antiserum and Ogawa specific antiserum can undergo antigenic conversion from "original" to "variant". Identification of antigenic factors can be summarized as follows:

(1) *V.* bio-serogroup 1875 original- B, C, D, and E
variant- C, D, and E

(2) *V. cholerae* bio-serogroup Hakata- C, D, and F

(3) *V. fluvialis* Kobe C, D, and E. (Hisatsune *et al.*, 1986, 1987).

Hisatsune *et al.* (1987), postulated that perosamine and quinovosamine are associated with antigenic factors A and B, mannose with D, unknown neutral sugars (I and II) with E and an unknown amino sugar with F.

Shewan and Véron (1974), included two additional biotypes *V. albensis* and *V. proteus* in the classification of *V. cholerae*. The latter organism is now recognized as a separate species, *V. metschnikovii* (Baumann *et al.*, 1984). Several differences were detected in the sugar content of these three biotypes, with respect to the glucose, galactose, fructose, D-glycero-D-manno-heptose, perosamine and quinovosamine levels (Kondo *et al.*, 1988).

V. ordalii is one of the causative agents of vibriosis in salmonids. The core polysaccharide portion of its LPS has been investigated using a number of techniques, and the structure has been reported (Banoub and Hodder, 1985). This identification was hampered by the presence of an amino sugar with a non acylated amino group, and resistance of the glycosidic bond between the amino sugar and the contiguous L-glycero-D-manno-heptose to acidic hydrolysis. However the presence of D-glucose, heptose, 2-amino-2-deoxy-D-glucose and KDO were shown to be present (Banoub and Hodder, 1985).

Another marine vibrio, *V. parahaemolyticus* can be divided into 12 serogroups, based on agglutination tests. Chemical studies on the LPS profile of *V. parahaemolyticus* have revealed various combinations of galactose, fucose, arabinose, D-glycero-D-manno-heptose, galactosamine, and three unidentified amino sugars (P_1, P_2, and P_3) in each of the 12 different serotypes (Hisatune *et al.*, 1980a, 1981). Except for serotype O6, the LPS of *V. parahaemolyticus* contains no KDO (Hisatune *et al.*, 1980a). *V. parahaemolyticus* O7 and O12 LPS had been found to contain a KDO-like substance, which recently Kondo *et al.* (1989) identified to be 3-deoxy-D-threo-hexulosonic. On the basis of the sugar composition, the LPS of the 12 serogroups can be classified into 9 chemotypes, as O3, O5 and O11 LPS belong to the same chemotype, and O7, O12 to another (Hisatune *et al.*, 1980a). Studies performed on *V. alginolyticus*, another marine vibrio, have shown a

similar LPS chemical composition. Members of this species also contain this KDO-like substance (Hisatune *et al.*, 1981).

The group F vibrios were once classified as *Aeromonas*. Sugar composition of the LPS of this group was investigated and it was found that this group could be separated into two subdivisions. These analyses also showed the LPS structure to be distinct from that of *Aeromonas* (Kondo *et al.*, 1980).

1.7 Aims of this study

The study of *V. cholerae* at the molecular level has led to a dramatic increase of our knowledge of this organism. Most of the studies have focussed on the multifactorial virulence factors (outer membrane proteins, LPS and secreted proteins) elaborated by *V. cholerae* and their function in the infectious process.

The genetics and biochemistry of Gram-negative LPS and specifically O-antigen is still not fully elucidated. Mapping of chromosomal regions and analysis of sugar biosynthetic pathways have been developed using *E. coli* and *S. typhimurium* as model organisms.

The objectives of this study are to characterize the 20 kb *V. cholerae* Ogawa *Sac*I fragment encoding the biosynthetic enzymes for the production of O-antigen, at the DNA level. Using a variety of techniques, identification of regions containing promoter activity, and analysis of proteins by homology studies will be attempted to postulate the organization of the *V. cholerae rfb* region.

CHAPTER 2

MATERIALS AND METHODS

2.1 Growth media

The following nutrient media were used for bacterial cultivation. Nutrient broth (NB) (Difco), prepared at double strength (16 g/l) with added sodium chloride (NaCl) (5 g/l) or Luria broth (LB), were the general growth medium for both *V. cholerae* and *E. coli* K-12 strains. Luria broth (LB) is composed of bacto-tryptone (10 g/l) (Difco), bacto-yeast (5 g/l) (Difco) and NaCl (5 g/l). 2x TY medium was prepared as described by Miller (1972). M9 minimal media (Miller, 1972) was supplemented with histidine, tryptophane and lysine for the growth of JC3272 used in the galactokinase assays. Minimal A medium (M13 minimal media) was also prepared as described by Miller (1972) and supplemented prior to use with $MgSO_4$, glucose and thiamine-HCl to concentrations of 0.2 mg/ml, 2 mg/ml and 50 µg/ml, respectively.

NA is nutrient agar is composed of Lab-Lemco powder (Oxoid) (10 g/l), peptone (Oxoid) (10 g/l), NaCl (5 g/l) and Agar (Media Makers) (15 g/l). Soft agar contains equal volumes of NB and NA. H agar consisted of bacto-tryptone (16 g/l) (Difco), NaCl (8 g/l) and bacto-agar (12 g/l) (Difco). H top agar was like H agar but contained 8 g/l bacto-agar. Galactose-tetrazolium (TZGal) agar was Luria agar supplemented with 2,3,5 triphenyltetrazolium (50 mg/ml) and galactose to a final concentration of 1%.

Antibiotics were added to broth and solid media at the following final concentrations: ampicillin (Ap) 25 µg/ml; chloramphenicol (Cm) 25 µg/ml; tetracycline (Tc) 10 µg/ml for *E. coli* and 4 µg/ml for *V. cholerae* strains.

Incubations were at 37°C unless otherwise specified. Normally, liquid cultures were grown in 20 ml McCartney bottles.

2.2 Chemicals and reagents

Chemicals were Analar grade. Phenol, polyethylene glycol-6000 (PEG), sodium dodecyl sulphate (SDS) and sucrose were from BDH Chemicals. Tris was Trisma base from Boehringer Mannheim. Caesium chloride (Cabot) was technical grade. Ethylene-diamine-tetra-acetic-acid, disodium salt (EDTA) was Analar analytical grade from Ajax Chemicals. Nitrilotriacetic acid (NTA) was from Sigma.

Antibiotics were purchased from Sigma (ampicillin, kanamycin sulphate), and Calbiochem (tetracycline, chloramphenicol). All other anti-microbial agents (dyes, detergents and antibiotics) 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).

The four deoxyribonucleotide triphosphates (dATP, dCTP, dGTP and dTTP) and their corresponding dideoxy-ribonucleotide triphosphate homologues (ddATP, ddCTP, ddGTP and ddTTP), were obtained from Boehringer-Mannheim. Adenosine-5'-triphosphate, sodium salt (ATP), herring sperm DNA and dithiothreitol (DTT) were obtained from Sigma. X-gal (5-Bromo-4-chloro-3-indolyl-ß-D-galacto-pyranoside) and IPTG (isopropyl-ß-D-thiogalacto-pyranoside) were purchased from Boehringer-Mannheim.

[^{14}C]-galactose # CFA 435 was purchased from Amersham. β-D(-)fructose (# F-0127), D(+)galactose (# G-0750) and 2,3,5-Triphenyltetrazolium chloride (T8877) were from Sigma.

M13 sequencing primer and [^{32}P]-dCTP, at a specific activity of 1,700 Ci/mMole were obtained from BRESATEC (Adelaide). The -35 sequencing primer was obtained

from New England Biolabs. [^{35}S]-Methionine (1,270 Ci/mmole), and [^{35}S]-dATP (>1000 Ci/mmole) were purchased from Amersham. Sequenase^{TS} was purchased from I.U.B. Phosphorylated *Sac*I (d[pCGAGCTCG]) and *Sal*I (d[GGTCGACC]) linkers were purchased from New England Biolabs. Linkers were obtained in a lypholized form and resuspended in 0.1 ml of TE buffer, pH 8.0 and stored frozen at -20°C.

2.3 Enzymes

Deoxyribonuclease I (DNase I) and lysozyme were obtained from Sigma. Pronase and proteinase K were from Boehringer-Mannheim.

Restriction endonucleases *Bam*HI, *Bgl*II, *Cla*I, *Eco*RV, *Eco*RI, *Hin*dIII, *Kpn*I, *Mlu*I, *Nru*I, *Pvu*I, *Pvu*II, *Pst*I, *Sac*I, *Sal*I and *Sma*I were purchased from either Boehringer-Mannheim, New England Biolabs, Pharmacia or Amersham.

DNA modifying enzymes were purchased from the following suppliers: New England Biolabs (T4 DNA ligase), Amersham (T4 DNA polymerase) and Boehringer-Mannheim (DNA polymerase I, Klenow fragment of DNA polymerase I, and molecular biology grade alkaline phosphatase).

Double stranded nested deletion kit was purchased from Pharmacia, and the Cyclone System kit was purchased from Integrated Sciences.

2.4 Synthesis of oligodeoxynucleotides

Oligodeoxynucleotides (oligos) were synthesized using reagents purchased from Applied Biosystems or Ajax Chemicals (acetonitrile). Synthesis was performed on an Applied Biosystems 381A DNA synthesizer. Oligos were routinely of a purity that no further purification was required.

2.5 Maintenance of bacterial strains

For long term storage, all strains were maintained as lyophilized cultures, stored *in vacuo* in sealed glass ampoules. When required, an ampoule was opened and its contents suspended in several drops of the appropriate sterile broth. Half the contents were then transferred to a 10 ml bottle of NB and incubated with shaking overnight at the appropriate temperature. The other half was streaked onto two nutrient agar plates and incubated overnight at the appropriate growth temperature. Antibiotics were added to the media when appropriate. If the colony form was uniform, single colonies were selected and picked off plates for subsequent storage or use. Short-term storage of strains in routine use was as a suspension of freshly grown bacteria in glycerol (32% v/v) and peptone (0.6% w/v) at -70°C. Fresh cultures from glycerols were prepared by streaking a loopful of the glycerol suspension onto a nutrient agar plate (with or without antibiotic as appropriate) followed by incubation overnight just prior to use.

Bacterial strains were prepared for long-term storage by suspension of several colonies in a small volume of sterile skimmed milk. Approximately 0.2 ml aliquots of this thick bacterial suspension were dispensed into sterile 0.25 x 4 in freeze drying ampoules and the end of each ampoule was plugged with cotton wool. The samples were then lyophilized in a freeze drier. After the vacuum was released, the cotton wool plugs were pushed well down the ampoule and a constriction was made just above the level of the plug. The ampoules were evacuated to a partial pressure of 30 microns and then sealed at the constriction without releasing the vacuum. Finally the ampoules were labelled and stored at 4°C.

2.6 Bacterial strains and plasmids

The *Vibrio cholerae* strains used are listed in Table 2.1. Strains of the El Tor biotype were distinguished from the Classical biotype by resistance to the antibiotic polymyxin B (50 units/ml) and sensitivity to biotype specific typing phages. Table 2.2 describes the *Escherichia coli* K-12 strains used in this study.

Table 2.1: *Vibrio cholerae* strains

Strain	Biotype/Serotype	Genotype/phenotype	Source
O17	El Tor, Ogawa	*str*	K. Bhaskaran
569B	Classical, Inaba	*str*, nonmotile	K. Bhaskaran

Table 2.2: *Escherichia coli* strains

Strain	Genotype/Phenotype	Source
DH1	F^-, *gyrA*-96, *recA*-1, *relA*-1, *endA*-1, *thi*-1, *hsdR*-17, *supE*-44,λ^-	B.Bachmann
DS410	F^-, *minA*, *minB*, *rpsL*	D. Sherratt
EX100	DH1, *thyA*	R. Morona
JM101	F', [*traD*-36, *proA,B*, $lacI^q$, *lacZ*, ΔM15], *supE*, *thi*-1, Δ[*lac-proA,B*]	A. Sivaprasad
LE392	F^-, *supF*, *supE*, *hsdR*, *galK*, *trpR*, *metB*, *lacY*	L. Enquist
C600	F^-, *thr*-1, *leu*-6, *tonA*1, *lacY*1, *supE*44, *thi*-1	J. Reeve
JC3272	*his*, *lys*, *trp*, *lacY*, *galK*, *strA*	R. Skurray

The plasmids cloning vectors which were used in this study are listed in Table 2.3.

2.7 Transformation procedure

Transformation was performed essentially according to the method described by Brown *et al.* (1979). *E. coli* K-12 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 shaking until the culture reached an OD of 0.6 (4x 10^8 cells/ml). The cells were chilled on ice for 20 min, pelleted at 4°C in a bench centrifuge, resuspended in half volume of cold 100 mM $MgCl_2$, centrifuged again and resuspended in a tenth volume of cold 100 mM $CaCl_2$. This was allowed to stand for 60 min on ice before addition of DNA. Competent cells (0.2 ml) were then mixed with DNA (volume made to 0.1 ml with 1x TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and left on ice for a further 30 min. The cell/DNA mixture was heated at 42°C for 2 min and then 3 ml NB was added followed by incubation with shaking at 37°C for 1-2 hr. The culture was then plated onto selection plates directly or concentrated by centrifugation and plated. Cells with sterile buffer were included as a control.

2.8 Bacterial conjugation

Overnight broth cultures grown in NB or LB were diluted 1:20 and grown to early exponential phase with slow agitation. Donor and recipient bacteria were mixed at a ratio of 1:1 and the cells pelleted by centrifugation (5000 rpm, 5 min, bench centrifuge). The pellet was gently resuspended in 200 µl of broth and spread onto a cellulose acetate membrane filter (0.45 µm, type HA, Millipore Corp.) on a NA plate. This plate was incubated for 6-16 hr at 37°C. The cells were then resuspended in 10 ml NB and samples plated onto selective agar and incubated overnight at 37°C.

Table 2.3: Plasmids and cloning vectors

Plasmid	Antibiotic marker	Reference
pBR322	Ap, Tc	Bolivar *et al.* (1977)
pUC18	Ap	Vieira and Messing (1980)
pUC19	Ap	Vieira and Messing (1980)
M13mp18		Messing and Vieira (1982)
M13mp19		Messing and Vieira (1982)
pACYC184	Cm, Tc	Chang and Cohen (1978)
pRU669	Cm,Km, $R_{ts}1$::Tn*1725*	Ubben and Schmitt (1986)
pKC86	Ap	Chak and James (1985)
pKC87	Ap	Chak and James (1985)
pOmpV500	Tc	Morona *et al.* (1990)
pEVX6	Ap	Ward *et al.* (1987)
pEVX7	Ap	Ward *et al.* (1987)

2.9 Plasmid DNA extraction procedures

Plasmid DNA was isolated by one of the three following procedures:

Method 1: Triton X-100 cleared lysates were prepared from 10 ml overnight cultures by a modification of the procedure of 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 0.25M EDTA, pH 8.0 were added to cells in Eppendorf 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% Triton X-100) was added followed by a brief mixing by inversion of the tubes. The chromosomal DNA was then pelleted by centrifugation (20 min, 4°C, Eppendorf). 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 an equal weight of propan-2-ol and allowed to stand at -70°C for 30 min. The precipitate was collected (10 min, Eppendorf), washed once with 1 ml 70% (v/v) ethanol, dried *in vacuo* and resuspended in 50 µl 1X TE buffer.

Method 2: Large scale plasmid purification was performed by the three step alkali lysis method (Garger *et al.*, 1983). Cells from an one litre culture were harvested (6,000 rpm, 15 min, 4°C, GS-3, Sorvall) and resuspended in 24 ml solution 1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). Freshly prepared lysozyme (4 ml of 20 mg/ml in solution 1) was mixed with the cell suspension and incubated at room temperature for 10 min. Addition of 55 ml of solution 2 (0.2 M NaOH, 1% (w/v) SDS), followed by a 5 min incubation on ice resulted in total lysis of the cells. After the addition of 28 ml solution 3 (5M potassium acetate, pH 4.8) and incubation on ice for 15 min, protein, chromosomal DNA and high molecular weight RNA were removed by centrifugation (8,000 rpm, 20 min, 4°C, GSA, Sorvall). The supernatant was then extracted with an equal volume of a TE saturated phenol, chloroform, isoamyl alcohol

mixture (25:24:1). Plasmid DNA from the aqueous phase was precipitated with 0.6 volume of 100% (v/v) propan-2-ol at room temperature for 10 min and collected by centrifugation (10,000 rpm at 4°C, 35 min, GSA, Sorvall). After washing in 70% (v/v) ethanol, the pellet was dried *in vacuo* and resuspended in 4.8 ml TE. Plasmid DNA was purified from contaminating protein and RNA by centrifugation on a two step CsCl ethidium bromide gradient according to Garger *et al.* (1983). The DNA band was removed by side puncture of the tube with a 19 gauge needle attached to a 1 ml syringe. The ethidium bromide was extracted using isoamyl alcohol. CsCl was then removed by dialysis overnight against three changes of 2 litres 1x TE at 4°C. DNA was stored at 4°C.

Method 3: Small scale plasmid purification was performed by the three step alkali lysis method using a modification of Garger *et al.* (1983). Overnight bacterial cultures (1.5 ml) were transferred to a microfuge tube, harvested by centrifugation (45 sec, Eppendorf), and resuspended in 0.1 ml solution 1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). The addition of 0.2 ml solution 2 (0.2 M NaOH, 1% (w/v) SDS) followed by a 5 min incubation on ice resulted in cell lysis. After the addition of 0.15 ml solution 3 (5M potassium acetate, pH 4.8) and a 5 min incubation on ice, protein, chromosomal DNA and high molecular weight RNA were collected by centrifugation (90 sec, Eppendorf). The supernatant was transferred to a fresh tube and extracted once with TE-equilibrated phenol and once with diethyl ether. Plasmid DNA was precipitated by the addition of 2x volume 100% ethanol and a 2 min incubation at room temperature. The DNA was collected by centrifugation (15 min, Eppendorf), washed with 70% (v/v) ethanol and dried *in vacuo*. The pellet was resuspended in 40 µl 1x TE.

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 µg DNA/ml (Miller, 1972).

2.10.2 Restriction endonuclease digestion of DNA

Cleavage reactions with the restriction enzymes *Sac*I and *Cla*I were performed using EB buffer (10x: 100 mM Tris-HCl pH 7.5, 60 mM $MgCl_2$, 10 mM dithiothreitol), *Hin*dIII, *Kpn*I, *Pst*I, *Pvu*II, *Xba*I, *Mlu*I, used EBS buffer (10x: EB buffer with 500 mM NaCl). The remaining restriction digests were carried out using these two buffers as a basis, with either the addition of NaCl, KCl or Tris-HCl as described by the manufacturers. 0.1-0.5 µg of DNA or purified restriction fragments were incubated with 2 units of each restriction enzyme in a final volume of 20 µl, at 37°C, for 1-2 hr. The reactions were terminated by heating at 65°C for 10 min. 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 calculated sizes of the SPP1 *Eco*RI standard fragments used differ from those published (Ratcliff *et al.*, 1979) and were calculated with the program DNAFRAG (Rood and Gawthorne, 1984) using bacteriophage lambda and plasmid pBR322 as standards. The sizes (kilobases, kb) used were: 8.0; 7.1; 6.0; 4.78; 3.44; 2.77; 1.93; 1.88; 1.55; 1.43; 1.2; 1.03; 0.7; 0.48; 0.38 (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, 0.6%, 0.8% or 1% (w/v) agarose gels (Seakem HGT), 13 cm long, 13 cm wide and 0.7 cm thick. Gels were run at 100V for 4-5 hr 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 acetate, 40 mM Tris and 2 mM EDTA). After electrophoresis the gels were stained in distilled water containing 2 µg/ml ethidium bromide. DNA bands were visualized by trans-illumination with UV light and photographed using either Polaroid 667 positive film or 665 negative film.

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 3M sodium acetate, pH 5.0. DNA was collected by centrifugation (15 min, Eppendorf), washed once with 70% (v/v) ethanol and dried *in vacuo* before being resuspended in 1x 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 1/2x TAE buffer. A current was applied causing the DNA to move out of the gel and into the buffer contained in the dialysis tubing. The DNA was then precipitated with two volumes of ethanol and one tenth volume of 3M sodium acetate, pH 5.0.

Method 3: After electrophoresis the required DNA bands were excised and placed in an Eppendorf tube containing siliconized glass wool covering a hole pierced in the bottom of the tube. This was inserted into another

Eppendorf tube and centrifuged for 15 min at half the maximum speed. The resulting solution DNA contained in the solution was collected by precipitation as previously described.

2.10.5 Dephosphorylation of DNA using alkaline phosphatase

Restriction enzyme digested DNA was alkaline phosphatase treated by either of the following two methods:

Method 1: 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 hot (56°C) 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 3M sodium acetate pH 8.0. The precipitate was collected by centrifugation (15 min, Eppendorf), washed once with 1 ml 70% (v/v) ethanol, dried *in vacuo* and dissolved in 1x TE buffer.

Method 2: The digested DNA was placed at 65°C for 10 min to heat inactivate the restriction enzyme. Then the 20 µl aliquot was converted to a final volume of 29 µl containing 35 mM Tris HCl pH 9.0, 0.17 mM $ZnCl_2$, 1.7 mM spermine and 1 unit of CIP. After vortexing, this was incubated at 37°C for 1 hr. Inactivation of CIP was achieved by increasing the reaction volume to 75 µl with the addition of SDS and nitrilotriacetic acid (NTA) to give final concentrations of 0.005% and 0.8 mM respectively. This was incubated at 68°C for 15 min, after which the CIP was removed by phenol extraction and the DNA recovered by ethanol precipitation, as previously described.

2.10.6 End-filling with Klenow fragment

Protruding ends created by cleavage with restriction endonucleases were filled in using the Klenow fragment of *E. coli* DNA polymerase I. Typically, 1 µg of digested DNA, 2 µl of 10x nick-translation buffer (Maniatis *et al.*, 1982), 1 µl of each dNTP (2 mM) and 1 unit Klenow fragment were mixed and incubated for 30 min. The reaction was stopped by heating at 65°C for 10 min, followed by removal of unincorporated dNTPs and enzyme by centrifugation through a Sepharose CL-6B column.

Sepharose CL-6B columns were prepared by placing glass wool into an Eppendorf tube which has had a hole punched in the bottom. 1 ml of Sepharose CL-6B (equilibrated with 1x TE buffer) was added, and the tube within another carrier tube was centrifuged at 2000 rpm for 2 min (Hermle bench centrifuge). This column was washed 2x with H_2O before use.

2.10.7 End-filling with T4 DNA polymerase

Plasmid DNA was cleaved and cohesive ends converted to blunt ends with T4 DNA polymerase in a final volume of 25 µl containing 2µg DNA, 2 units T4 DNA polymerase, 1µl of each dNTP (2mM) and 1µl of 10x T4 DNA polymerase buffer (Maniatis *et al.*, 1982). After a 5 min incubation at 37°C, the reaction was stopped by heating at 65°C for 10 min. Salt, unincorporated nucleotides and enzyme were removed by passage through a Sepharose CL-6B column, as described above.

2.10.8 Ligation of linkers to blunt DNA ends

Phosphorylated linkers were ligated to blunt ends generated by Klenow fragment of *E. coli* DNA polymerase I by overnight incubation of 1 µg plasmid DNA with 3 µl linkers and 4 units T4 DNA ligase in a final volume of 10 µl of 1x linker-kinase buffer (Maniatis *et al.*, 1982).

2.10.9 *In vitro* cloning

DNA to be subcloned (3 µg) was cleaved in either single or double restriction enzyme digests. This was combined with 1 µg of similarly cleaved vector DNA, then ligated with 2 units of T4 DNA ligase in a volume of 50 µl in a final buffer concentration of 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 0.6 mM ATP for 16 hours at 4°C. The ligated DNA was then used directly for transformation of *E. coli* strains. Transformants were screened for insertional inactivation of the appropriate drug resistance (Ap or Tc), wherever possible, prior to plasmid DNA isolation.

2.10.10 Nested deletions

This was performed essentially as described by Pharmacia. 2.5 µg of test DNA, (pRMB134 or pRMB200) were linearized with the appropriate restriction enzymes (pRMB200: *Xba*I, *Sph*I; pRMB134 *Xba*I, then after the addition of thionucleotides, *Bam*HI), in a final volume of 25 µl. pRMB134 required the addition of thionucleotides to ensure deletions were constructed from only one end. This was achieved by combining the *Xba*I digested DNA, Klenow buffer, dNTPαS mix and Klenow enzyme, in a volume of 10 µl. This was incubated at 37°C for 15 min, after which the enzyme was heat inactivated by a 20 min incubation at 65°C.

The DNA was precipitated by the addition of 20 µl of NaCl/glycogen and 75 µl of ethanol, and subsequent incubation at -70°C for 10 min. The precipitated DNA was collected by centrifuging for 10 min, washed with 70% (v/v) ethanol and dried *in vacuo*. After dissolving the DNA pellet in 10 µl of water, it was digested with *Bam*HI in a final volume of 20 µl. These digestions were monitored by analyzing an aliquot on a 0.8% agarose minigel.

The next step was to produce the deletions by the addition of Exonuclease III. The conditions chosen were; pRMB200 was digested at a NaCl concentration of 75 mM, at 25°C with samples being removed at 2 min intervals, and pRMB134 digestion was performed at a NaCl concentration of 75 mM, at 30°C with time samples being removed at 2 min intervals. Thus, in an Eppendorf tube, 20 µl of the appropriate 2x Exo III buffer

was combined with 10 µl of the double digested DNA. A 2 µl time-0 sample was removed and placed in a tube containing 3 µl of S1 nuclease buffer on ice. 1 µl of Exonuclease III was added to the mix, vortexed, and incubated at the required temperature. At 2 min intervals, samples were taken which were immediately mixed with 3 µl of S1 nuclease/buffer mix. After the last samples were taken, they were incubated simultaneously at room temperature for 30 min. 1 µl of S1 stop solution was added and the tubes were incubated at 65°C for 10 min.

To analyze the digestion rates, half of each sample was used for electrophoretic analysis, and the other half for recircularized by ligation. After ligation the samples were transformed into JC3272. DNA was made from random transformants, which were analyzed by agarose gel electrophoresis, tested for their colour reaction on TZGal plates and assayed for galactokinase activity.

2.10.11 Transposition with Tn*1725*

Tn*1725* (Cm^R) transposition to plasmid DNA was performed in the following manner: Plasmid pRU669 ($R_{ts}1$::Tn*1725*) (Ubben and Schmitt, 1986) was transferred by conjugation, into an *E. coli* K-12 derivative harbouring the target plasmid (pRMB2-Tc^R or pEVX7-Ap^R). This was achieved by mating for 3 hours at 30°C in a standing culture which consisted of 0.1 ml of an overnight culture of C600 [$R_{ts}1$::Tn*1725*] with 0.9 ml of the culture to be mutagenized and 1 ml of NB broth. Following plating of 0.1 ml of mating mix on NA containing Cm plus Ap/Tc, independent exconjugants were purified and used for growing up overnight cultures at 37°C in NB containing both antibiotics to select for the transposon (Cm) and the plasmid (Ap/Tc). Triton X-100 lysates (10 ml) prepared from these cultures were used to transform C600, again selecting for both the plasmid and the transposon. Following overnight incubation at 37°C, transformants were randomly chosen for analysis of their plasmid DNA.

2.10.12 Kinasing single stranded DNA and hybridization

Single stranded DNA (primers) were kinased using γ-[^{32}P]-dATP. The reaction mix consisted of 100 mM DTT, 1 µl 10x kinase buffer (10x: 500 mM Tris pH7.4 and 100 mM $MgCl_2$), 3 units of polynucleotide kinase, 10 µl of γ-[^{32}P] and 60 µg of primer. This reaction mix was made up to 10 µl in water and incubated at 37°C for 30 min. After incubation the labelled oligonucleotide was ethanol precipitated, dried *in vacuo* and resuspended in water. Before use, the oligonucleotide was heated to 65°C for 10 min. Filters to be hybridized with the labelled oligonucleotide were first prehybridized for 4 hr at 37°C in a prehybridization solution containing 0.9 mM sodium chloride, 90 mM Tris-HCl pH 7.6, 0.009 mM EDTA, 5x Denhardt's reagent (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% fractionV BSA), 83 µg/ml single stranded herring sperm DNA (Sigma) and 0.0005% SDS.

Filters were washed twice with shaking at room temperature for 20 minute in 5x SSC (0.34 M NaCl, 75 mM sodium citrate, pH 7.0), followed by 2x 20 min at 10°C below T_m for the oligo. The filters were then covered with plastic film and placed on film for autoradiography at -70°C with intensifying screens.

2.11 *In vitro* RNA polymerase binding

E. coli RNA polymerase was purified as described by Burgess and Jendrisak (1975). The purified enzyme was >95% pure and contained an equivalent amount of σ^{70} factor as judged by SDS-polyacrylamide gel electrophoresis.

RNA polymerase-DNA complexes were obtained essentially as described by Sogo *et al.* (1979) using a modification (Morelli *et al.*, 1981) of procedures described previously (Koller *et al.*, 1974; Portmann and Koller, 1976). The incubation mixture contained in a final volume of 30 µl: 30 mM EDTA buffer pH 7.9, 50 mM KCl, 8 mM Mg-acetate, 1 µg of either pEVX6 or pEVX7, and 2 µg of RNA polymerase. After 10 min incubation at 37°C the complexes were fixed for 10 min at 37°C with 0.1%

glutaraldehyde. Free enzyme was separated from the complexes on a column of Sepharose 4B in 30 mM EDTA buffer pH 7.9, 8 mM Mg-acetate, 0.1% gluteraldehyde.

The two samples were adjusted to a concentration of 1 µg/ml with 30 mM EDTA bufer pH 7.9, 8 mM Mg-acetate, adsorbed to mica, washed in H_2O for 2 hr, stained with 2% uranylacetate, dried with 80% ethanol, and shadowed with Pt/Pd. Carbon replicas were floated off the mica onto the surface of distilled water. Electron micrographs were taken at 11,000x magnification in a Philips EM301. The electron micrographs were enlarged 6x and the contour lengths of the molecules and the positions of polymerase molecules were measured with a digitizer (Numonics Corp., North Wales, PA., USA).

2.12 ^{14}C Galactokinase assay

This assay was performed as described by McKenney *et al.* (1981). An overnight culture of JC3272 harbouring plasmid was subcultured into M9 minimal media using fructose as a carbon source, and grown at 37°C shaking to OD_{650} = 0.6. A 1 ml sample was removed to a test tube to which 40 µl of lysis buffer (100 mM EDTA, 100 mM DTT, 50 mM Tris-HCl pH 8.0), and 1 drop of toluene were added. After vortexing for 10 sec, the tubes were incubated at 37°C for 1 hr. From this lysis mix a 20 µl sample was removed to a clean test tube to which included: 20 µl of solution 1 (5 mM DTT, 16 mM NaF), 50 µl of solution 2 (8 mM $MgCl_2$, 200 mM Tris-HCl pH 7.9, 3.2 mM ATP) and 10 µl ^{14}C-galactose (diluted to 4.5 x 10^6 dpm/µmole). This mix was incubated at 32°C, during which, 3 time course samples of 25 µl, at 10 min, 20 min, and 30 min were removed and spotted onto 23 mm DE81 Whatman filters, which were subsequently air dried. Two samples were also spotted onto separate discs and remained unwashed. The other discs were batch washed 3x 5 min in 300 mls of water, after which they were dried 65°C for 10 min, and counted with 3 mls of Optiphase "Hisafe II" (LKB) scintillation fluid.

2.13 Analysis and manipulation of RNA

2.13.1 RNA preparation

RNA was prepared by a modified method described by Aiba *et al.* (1981). Overnight (NB) cultures of either *V. cholerae* or *E. coli* harbouring a plasmid of interest were subcultured 1:10 and grown to $OD_{650} = 1$. Five mls of culture were centrifuged and the pellet was resuspended in 0.5 ml of solution A (0.02 mM NaAc pH 5.5, 0.5% SDS, 1 mM EDTA). This was extracted 3 to 4 times with hot (65°C) phenol (equilibrated with a solution containing; 0.02 mM NaAc, 0.02 mM KCl, 0.01 mM $MgCl_2$ at a pH of 5.2). The resulting solution was then precipitated with two times volume of ethanol and one tenth volume of NaAc. To remove contaminating DNA the precipitate obtained was resuspended in water and incubated 37°C for 10-15 min with DNAse buffer (10x: 200 mM Tris-HCl pH 7.6, 50 mM $MgCl_2$) and 1 µl of DNAse enzyme (10 u/µl, BRESATEC). This was reextracted with phenol, the RNA precipitated, dried *in vacuo* and resuspended in water.

2.13.2 Dot Blots with single stranded RNA

RNA preparations (50 µg in 100 µl H_2O) were denatured by the addition of 300 µl of 6.15 M formaldehyde/10x SSC. The reaction was then incubated at 65°C for 10 min. A nitrocellulose filter (Schleicher and Schüll) equilibrated with 10x SSC was placed in the Biorad Dot Blot apparatus, with the denatured RNA samples under *vacuo*. Appropriate denatured DNA controls were also dotted onto the nitrocellulose sheet.

The filter was then air dried, baked *in vacuo* at 80°C for 2 hr, and subjected to hybridization with either labelled oligonucleotides (2.10.12), or labelled RNA from SP6/T7 vector clones (see below).

2.13.3 SP6/T7 RNA labelling and hybridization

Labelled RNA transcripts were made from selected *rfb* regions. In an Eppendorf tube 100 µCi of α-^{32}P-UTP was dried down. To this tube were added; 2 µg of linearized DNA template, 2 µl of 10x nucleotide/ buffer cocktail, 2 µl of 100 mM DTT and 4 units of SP6 RNA-Polymerase. This was incubated at 40°C for 1 hr. The DNA template was removed by the addition of 1 µl of RNase-free DNase I to the reaction and the mixture was incubated at 37°C for 10 min. The synthesized RNA was the precipitated with 2x volume of ethanol, subjected to a 70% (v/v) ethanol wash, dried *in vacuo* and resuspended in H_2O.

Prior to hybridization with these probes, filters were incubated for 4 hr at 42°C in a pre-hybridization solution containing 50% (v/v) formamide, 50 mM sodium phosphate buffer, pH 6.4, 5x SSC (0.34 M NaCl, 75 mM sodium citrate, pH 7.0), 5x Denhardt's reagent (0.1% Ficoll, 1% polyvinylpyrrolidone, 0.1% fractionV BSA) and 83 µg/ml single stranded herring sperm DNA (Sigma) (Maniatis *et al.*, 1982). The probe was added and hybridization allowed to occur for 16-24 hr at 42°C.

Filters were washed twice with shaking at 37°C for 30 min in 2x SSC, containing 0.1% (w/v) SDS. This was followed by two further washes in 0.2x SSC plus 0.1% (w/v) SDS at 65°C. The filters were then covered in plastic film and placed on film for autoradigraphy at -70°C with intensifying screens.

2.14 Protein analysis

2.14.1 Minicell procedures

Minicells were purified and the plasmid-encoded proteins labelled with [^{35}S]-methionine as described by Kennedy *et al.* (1977) and modified by Achtman *et al.* (1979). This involved separation of minicells from whole cells (500 ml overnight culture in LB) by centrifugation through two successive sucrose gradients, pre-incubating in minimal medium to degrade long lived mRNAs corresponding to chromosomally encoded genes,

then pulse labelling with [^{35}S]-methionine in the presence of methionine assay medium. Minicells were subsequently solubilized by heating at 100°C in 100 µl of 1x sample buffer (Lugtenberg *et al.*, 1975) and analysed by SDS-PAGE.

2.14.2 Bacterial cell-free coupled transcription-translation

The prokaryotic DNA-directed translation kit was obtained from Amersham, and this system was essentially as described by DeVries and Zubay (1967), and modified by Collins (1979), and Zubay (~~1980~~ 1973). Reactions were carried out according to manufacturers' specifications. This procedure involved reacting together cell extract (S30), test DNA (CsCl purified at 1 mg/ml and RNA free), amino acids (excluding methionine), L-[^{35}S]-methionine and nucleotides at 37°C for 1 hour. The reaction was chased by the addition of a methionine chase solution, and the reaction terminated by placing the reaction tubes into an ice bath. Samples were diluted 1:1 (v/v) with loading buffer (0.08 M Tris-HCl, pH6.8, 0.1 M DTT, 2% (w/v) SDS, 10% (v/v) glycerol, 0.1 mg/ml bromophenol blue) heated to 100°C for 5 min and then analyzed by SDS-PAGE (see below) and subjected to autoradiography.

2.14.3 SDS-Polyacrylamide Gel Electrophoresis

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on either 11-20% polyacrylamide gradients (for proteins) or straight 20% polyacrylamide gels (for 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 prior to loading. Gels were generally electrophoresed at 100 V for 5 hr (11-20% gradient gels) or 10mA constant current for 16 hr (20% PAGE gels). Proteins were stained with gentle agitation overnight at room temperature 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 24 hr.

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 α-Lactalbumin (14.4 kDa).

2.14.4 Autoradiography

SDS-PAGE gels were dried on Whatman 3MM chromatography paper at 60°C for 2 hr on a Bio-Rad gel drier. [^{35}S]-methionine and [^{35}S]d-ATP autoradiography was performed at room temperature for 1-7 days without intensifying screens using Kodak XR-100 film. For autoradiography with [^{32}P]-phosphate labelled DNA, the gels were exposed to film for 6 - 72 hr at -70°C, using intensifying screens.

2.14.5 Preparation of whole cell lysates (WCL)

Whole cell lysates (WCL) were prepared by the method of Hitchcock and Brown (1983). Cells were grown overnight in NB and 1.5 ml was spun down in an Eppendorf centrifuge for 5 min. The pellets were solubilized in 50 µl of lysing buffer containing 2% SDS, 4% β-mercaptoethanol, 10% glycerol, 1 M Tris-HCl pH 6.8, and 0.1% bromophenol blue. Lysates were heated at 100°C for 10 min. 2.5 µg of Proteinase K solubilized in 10 µl of lysing buffer was added to each sample and incubated at 60°C for 2-4 hr. Samples were stored at -20°C.

2.14.6 LPS silver staining

Silver staining of LPS in polyacrylamide gels was performed using the method of Tsai and Frasch (1982). The following procedure was used: i) fixation overnight in 40% ethanol, 10% acetic acid; ii) oxidation for 5 min with 0.7% periodic acid in 40% ethanol, 10% acetic acid; iii) 4 washes with water at 30 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% $AgNO_3$

in a total volume of 150 ml; v) developing in a solution of 50 mg citric acid and 0.5 ml formaldehyde in 1 litre. The citric acid was dissolved in water and heated to 37°C and formaldehyde added just before use. Distilled, deionized water which had been passed through a series of Millipore filters and had a conductivity of not less than 18 mega ohms/cm was used to rinse all glassware and in preparation of solutions.

2.15 Preparation of formalin fixed cells

Overnight cultures were subcultured and grown to a density of $2x10^9$ cells/ml. A sample was kept for a viable count and 1% formalin was added to the cells. The cells were incubated at 37°C for 60 min with occasional shaking. The cells were then washed 3 to 4 times with saline and a sample plated out to check the proportion of viable cells remaining. The OD_{650} of all samples was measured and cells were adjusted to the same concentration ($2x10^{10}$ cells/ml).

2.16 Haemagglutination inhibition assay (HIA)

A haemagglutination assay was performed by adding 25 µl of sheep red blood cells (SRBC), sensitized with alkali treated, purified LPS, to 25 µl of 2-fold dilutions of the appropriate antiserum. *V. cholerae* Ogawa LPS was obtained from S. Attridge. The monoclonal antibodies αA (20B) and αC (13B) were as described in Ward *et al.* (1987). Monoclonal antibody αB (VCO8) was obtained from Wellcome Diagnostics, Kent U.K. Trays were incubated at 37°C for 60 min and the haemagglutination end point determined. Four haemagglutinating units of antibody were used in the HIA.

The HIA was performed as follows. The antigen being tested (in this case formalin-fixed cells) was diluted out serially in 25 µl volumes. 25 µl of anti-serum (4 haemagglutinating units, as described above), was then added to each well. The trays were incubated at 37°C for 60 min and then an equal volume, i.e. 50 µl, of sensitized

SRBC was added to each well. The trays were incubated for a further 60 min at 37°C and the end points determined.

2.17 Colony transfer and blotting with antiserum

A nitrocellulose disc (9 cm diameter) was placed onto agar plates containing the colonies to be screened. Once the colonies had adhered to the disc (3 min), the cell debris was removed from the nitrocellulose with a jet of saline (0.9% (w/v) NaCl). The nitrocellulose sheet was incubated for 30 min in 5%(w/v) skim milk powder in TTBS (0.05%(v/v) Tween 20, 20mM Tris-HCl, 0.9%(w/v) NaCl) to block non-specific protein binding sites. The antiserum was diluted 1/1000 in TTBS, 0.02%(w/v) skim milk powder (unless stated otherwise) added to the filters and incubated with gentle agitation at room temperature for 2-16 hours. The antibody was removed by washing the nitrocellulose disc three times for 10 min in TTBS with shaking. Detection of bound antibody was achieved by incubating for 2-16 hours (gentle agitation) with goat anti-rabbit IgG coupled with horseradish peroxidase (KPL) at a dilution of 1/5000 in TTBS plus 0.2%(w/v) skim milk powder. The nitrocellulose sheet was then washed four times (5 min intervals) with TTBS, followed by two 5 min washes in TBS (20mM Tris-HCl, 0.9%(w/v) NaCl). To detect the presence of the antigen-antibody complexes peroxidase substrate (9.9 mg 4-chloro-1-napthol dissolved in 3.3 ml -20°C methanol added to 16.5 ml TBS containing 15µl hydrogen peroxide) was then added and allowed to incubate 10-15 min with shaking, as described by Hawkes *et al.* (1982).

2.18 M13 cloning and sequencing procedures

2.18.1 Preparation of M13 replicative form (RF) DNA

Fresh 2x TY broth (10 ml) was inoculated with 10 µl of an overnight culture of JM101 (in M13 minimal medium). A single plaque of M13mp18 or M13mp19 picked from an H agar plate with a sterile toothpick was added to this bottle. The culture was grown at 37°C with vigorous shaking for 6 hr. Bacterial cells were removed by

centrifugation (5,000 rpm, 10 min, bench centrifuge) and the supernatant added to 1 litre NB containing 10 ml of a shaken overnight culture of JM101. Following incubation for 14 hr at 37°C with shaking, replicative form (RF) DNA was prepared as described above for plasmid DNA purification.

2.18.2 Cloning with M13mp18 and M13mp19

The M13 vectors, M13mp18 and M13mp19 (Messing and Vieira, 1982; Vieira and Messing, 1982) were used for selective cloning of restriction enzyme generated DNA fragments. Stocks of M13 vectors cleaved with various enzyme combinations and alkaline phosphatase treated were stored at 4°C, after heat inactivation of enzymes. Plasmid DNA was cut with the appropriate enzyme combinations for subcloning into the M13 vectors. The ligation conditions used for blunt ends and cohesive ends were identical. The reaction mixtures consisted of the DNA to be cloned (100 ng) and the DNA vector (20 ng) in a final volume of 10 µl of ligation buffer. Ligation with T4 DNA ligase was carried out overnight at 4°C.

2.18.3 Generation of stepwise deletions

Deletions of increasing size were made in the *Kpn*I - *Bam*HI fragment of the *V. cholerae rfb* region using a M13mp19 clone harbouring this fragment. The "Cyclone System" (#77200) available from Integrated Sciences, was employed to construct these overlapping deletions.

Once a single stranded template had been purified, a short synthetic oligodeoxynucleotide was annealed to the 3' end of the cloned insert at the end of the M13 polylinker generating a restriction site (which in this case was *Eco*RI). Once the DNA had been digested by the enzyme (*Eco*RI), single-stranded 3' to 5' exonuclease digestion was performed by T4 DNA polymerase for various lengths of time depending on the size of deletion required. After heat-inactivation, the deletion products were tailed

with dG's using terminal deoxyribonucleotidyl transferase thus producing a short homopolymer tail at the 3' end. Fresh oligodeoxynucleotide was annealed to the deletion products joining the two ends of the molecule. The remaining nick was sealed with T4 DNA ligase. The product was then used to transfect competent *E. coli* JM101 cells and individual plaques picked and screened for the size of insert. A suitable set of deleted clones were then subjected to DNA sequencing.

2.18.4 Transfection of JM101

The *E. coli* strain JM101 was made competent for transformation/transfection as described in section 2.7. Competent cells (0.2 ml) were added directly to the ligation mixes and incubated on ice for 30 min. This was followed by a 2 min heat shock at 42°C. Cells were then transferred to sterile test tubes to which was added a mixture of JM101 indicator cells (200 µl), 100 mM IPTG (40 µl) and 2% (w/v) X-gal in N, N'-dimethyl formamide (40 µl) and finally 4 ml H top agar. The mixture was poured as an overlay onto an H agar plate and incubated overnight at 37°C.

2.18.5 Screening M13 vectors for inserts

White plaques were picked from X-gal/IPTG plates with sterile toothpicks and added to 1 ml 2x TY broth in microfuge tubes containing a 1:100 dilution of an overnight culture of JM101. These tubes were incubated for 5 hr at 37°C. The cells were pelleted by centrifugation (30 sec, Eppendorf). RF DNA, suitable for restriction analysis, was prepared by the miniprep method (section 2.9, method 3). After restriction enzyme digestion, DNA was examined on 1% (w/v) agarose gels.

2.18.6 Purification of single stranded template DNA

M13 RF DNA containing appropriate inserts were reintroduced into JM101 and single white plaques from this transfection picked with sterile toothpicks to inoculate 2 ml 2x TY broth containing 20 µl of an overnight culture of JM101. After vigorous shaking at 37°C for 6 hr, the culture was transferred to Eppendorf tubes and centrifuged for 10 min. The supernatant was transferred to clean tubes and recentrifuged for 5 min. Three methods of lysing phage and collecting single stranded phage were employed. They are as follows:

Method 1: A 1 ml aliquot of the supernatant from each tube was withdrawn and mixed in a fresh tube with 0.27 ml 20% (w/v) polyethylene glycol (PEG), 2.5 M NaCl. These tubes were then incubated at room temperature for 15 min. The phage were pelleted by centrifugation for 5 min in an Eppendorf centrifuge and the supernatant discarded. Following another short spin (10 sec), the remainder of the PEG/NaCl supernatant was removed with a drawn out Pasteur pipette. The pellets were resuspended in 0.2 ml TE buffer. Redistilled TE saturated phenol (0.1 ml) was then added to the phage suspension and the tubes were briefly vortexed. After standing for 15 min at room temperature, the tubes were centrifuged for 2 min and 0.15 ml of the top phase transferred to clean tubes. To the aqueous phase 6 µl of 3 M sodium acetate pH 5.0 and 400 µl absolute ethanol was added. Single stranded DNA was precipitated at -20°C overnight, followed by centrifugation for 15 min in an Eppendorf centrifuge. DNA pellets were washed once with 1 ml 70% (v/v) ethanol followed by centrifugation. After drying *in vacuo* the pellets were resuspended in 25 µl 1x TE buffer and stored at -20°C until required.

Method 2: This method is the same as method 1 to precipitate the phage, which was then resuspended in 300 µl of TE buffer (110 mM Tris HCl, 0.1 mM EDTA, pH8.0). 300 µl of TE-saturated phenol was added and the mix was vortexed sporadically for 10 min. The tubes were centrifuged for 3 min,

and then the extraction was repeated with chloroform:isoamyl alcohol (24:1). The liberated single stranded DNA was precipitated by the addition of 1/10 volume 5 M $NaClO_4$ and 1 volume of isopropanol. Single stranded DNA was precipitated at -20°C overnight, followed by centrifugation for 15 min in an Eppendorf centrifuge. DNA pellets were washed once with 1 ml 70% (v/v) ethanol followed by centrifugation. After drying *in vacuo* the pellets were resuspended in 50 µl 1x TE buffer and stored at -20°C until required.

Method 3: 1 ml of the supernatant was added to a microfuge tube containing 250 µl of a 20% PEG/3.5 M ammonium acetate solution. This was vortexed and incubated on ice for 30 min. The phage were collected by centrifugation (15 min, Eppendorf) and all the supernatant carefully removed. The pellet was dissolved in 100 µl 1x TE buffer. 50 µl redistilled phenol was added and the tube vortexed for 2 min followed by incubation at room temperature for 5 min. 50 µl chloroform was then added, vortexed for 2 min, spun in a microfuge for 5 min and the upper aqueous phase was transferred to a fresh tube. The phenol/chloroform phase was extracted with 100 µl 1x TE buffer, spun as before and the aqueous phases combined. The combined aqueous phases were then extracted with an equal volume phenol/chloroform (three more times). This was then extracted with chloroform and 250 µl of the supernatant transferred to a microfuge tube containing 125 µl 7.5 M ammonium acetate. After addition of 0.75 ml 95% (v/v) ethanol the tubes were stored overnight at -20°C. The DNA was collected by centrifugation (15 min, Eppendorf), and the pellet washed twice with 95% (v/v) ethanol. The pellet was dried *in vacuo* before resuspending the DNA in 20 µl distilled water.

2.18.7 Dideoxy sequencing protocol with Klenow fragment

The method was based on that described by Sanger *et al.* (1977, 1980). Stock solutions of the four dNTPS and ddNTPs were 10 mM in 1x TE buffer and stored frozen at -20°C. Working stocks of the dNTPs were made by diluting to 0.5 mM with 1x TE. Working stocks of the ddNTPs were diluted to the following concentrations in 1x TE: ddATP (0.1 mM), ddCTP (0.1 mM), ddGTP (0.3 mM) and ddTTP (0.5 mM). The deoxynucleotide mixes (A, C, G, T) were made for each of the four sequencing reactions, with [^{32}P]-dCTP, as follows:

Components	Mixes			
	A°(µl)	C°(µl)	G°(µl)	T°(µl)
0.5 mM dATP	4	40	40	40
0.5 mM dCTP	--	5	--	--
0.5 mM dGTP	40	40	6	40
0.5 mM dTTP	40	40	40	6
10x TE buffer	10	10	10	10

Mixes of N° and working solutions of ddNTPs were made by the addition of the following combination of components:

Components	Mixes			
	A°+ddA	C°+ddC	G°+ddG	T°+ddT
N°	7 µl	7 µl	7 µl	7 µl
ddNTP	14 µl	14 µl	14 µl	14 µl

The mixes were stored at -20°C until required for later use in sequencing reactions. The annealing of synthetic primer to template was achieved by incubating 6 µl template, 1 µl

M13 primer, 1 µl 10x TM buffer (100 mM Tris-HCl, pH 8.0, 50 mM $MgCl_2$) and 2 µl water. The mixture was heated at 65°C for 60 min and then allowed to cool at room temperature. Rows of four microfuge tubes (one tube for each sequencing reaction) were prepared containing 2 µl of annealed DNA. 5 µCi of [^{32}P]-dCTP were dispensed into each of four tubes marked A, C, G and T and dried *in vacuo*. The solution of appropriate N°/ddN mix was used to resuspend the dried label. The N°/ddN label mix (2 µl) was aliquoted into each of four tubes (one for each sequencing reaction) containing 2 µl of annealed DNA. To the side of each tube was added .25 units of Klenow fragment. These components were simultaneously brought together by a brief spin in an Eppendorf centrifuge and the reaction mixes incubated at 37°C for 15 min. Chase solution (2 µl), consisting of 0.25 mM of each dNTP and 0.025 units/µl Klenow, were added to the side of each tube and the chase reaction started by another brief spin. After 15 min at 37°C, 4 µl formamide dye mix (95% (w/v) formamide, 0.1% (w/v) xylene cyanol, 0.1% (w/v) bromophenol blue, 10 mM EDTA pH 8.0) were added to stop the reaction. Reaction mixes were heated in a 100°C heating block for 2.5 min and immediately 0.5-1.0 µl loaded onto 6% polyacrylamide denaturing gels (see below). For re-running, these samples were boiled for 60 sec immediately prior to loading.

2.18.8 Dideoxy sequencing protocol with Sequenase™

The dideoxy chain termination procedure of Sanger *et al.* (1977) was modified to encompass the use of Sequenase™ (modified T7 DNA polymerase) in place of Klenow (Tabor and Richardson, 1987). All reagents were stored at -20°C. Two types of labelling and termination mixes were used, namely the dGTP mixes and the dITP mixes. The contents of the dGTP mixes are as follows :

Labelling Mix (dGTP):	7.5 µM dGTP, dCTP and dTTP
ddG Termination Mix (dGTP):	80 µM dNTP, 8 µM ddGTP, 50 mM NaCl
ddA Termination Mix (dGTP):	80 µM dNTP, 8 µM ddATP, 50 mM NaCl
ddC Termination Mix (dGTP):	80 µM dNTP, 8 µM ddCTP, 50 mM NaCl
ddT Termination Mix (dGTP):	80 µM dNTP, 8 µM ddTTP, 50 mM NaCl

The dITP mixes were used to reduce gel artifacts due to secondary structures in DNA synthesized in the sequencing reaction (Barnes *et al.*, 1983; Gough and Murray, 1983). The dITP mixes were as follows :

Labelling Mix (dITP):	15 µM dITP, 7.5 µM dCTP, 7.05 µM dTTP
ddG Termination Mix (dITP):	160 µM dITP, 80 µM dATP, dCTP dTTP, 1.6 µM ddGTP, 50 mM NaCl
ddA Termination Mix (dITP):	160 µM dITP, 80 µM dATP, dCTP dTTP, 8 µM ddATP, 50 mM NaCl
ddC Termination Mix (dITP):	160 µM dITP, 80 µM dATP, dCTP dTTP, 8 µM ddCTP, 50 mM NaCl
ddT Termination Mix (dITP):	160 µM dITP, 80 µM dATP, dCTP dTTP, 8 µM ddTTP, 50 mM NaCl

Normally the labelling mix was diluted 1:5 with water to obtain the working concentration, however, to read long sequences in a single reaction, a dilution of 1:2 was used. The synthetic primer was annealed to the template by incubating 7 µl template (5-10 nM), 1 µl primer (500 nM) and 2 µl 5x Sequenase buffer (200 mM Tris-HCl

pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl). The mixture was heated in a metal block at 65°C for 3 minutes and then the block containing the tubes was allowed to cool to room temperature. To the annealed mixture, 2 µl of the appropriately diluted labelling mix, 1 µl DTT (0.1 M), 0.5 µl [α-^{35}S]-dATP (1000 Ci/mmol) and 2 µl of diluted Sequenase™ (1:8 dilution in 1x TE buffer) was added, spun, mixed, resuspended and then incubated for 5 minutes at room temperature. 3.5 µl of this mix was then aliquoted into four microfuge tubes, prewarmed to 37°C, labelled A, C, G and T, each containing 2.5 µl of the corresponding termination mix, then spun briefly to start the termination reaction. After 5 minutes at 37°C, 4 µl Stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) was added to each of the reactions. Reaction mixes were heated to 100°C for 2 min and immediately 1.2 µl loaded onto the sequencing gel. For re-running, these samples were kept at -20°C for up to 2 weeks and heated to 100°C for 3 min prior to loading.

2.18.9 DNA sequencing gels

Polyacrylamide gels for DNA sequencing were prepared using glass plates 33 x 39.4 cm and 33 x 42 cm. Spacers and combs were high density polystyrene (0.25 mm thick). The gel mix contained 70 ml acrylamide stock [5.7% (w/v) acrylamide, 0.3% (w/v) bis-acrylamide, 8M urea in 1x TBE buffer (89 mM Tris base, 89 mM boric acid, 2.5 mM EDTA, pH 8.3)], plus 420 µl 25% ammonium persulphate and 110 µl TEMED (N,N,N',N'-tetramethyl-ethylene-diamine, Sigma). After thorough mixing the gel mix was poured into a clean gel sandwich and the comb inserted. Polymerization took place for 60 min, with the gel in a horizontal position. The gel was mounted onto the sequencing apparatus and a waterjacket was attached to the outside plate of the gel. This consisted of a plastic bag wedged between two 0.3 cm thick spacers and a third plate. The plastic bag was filled with 1x TBE buffer and this was sufficient to evenly distribute heat throughout the gel. Gels were pre-electrophoresed at 700 V for 30 min. After the samples had been loaded the gel was electrophoresed at a constant voltage (700 V) for

15 min, and then increased to 1200 V (33 mA). After 4 hr the samples were reloaded into a second set of wells on the same gel. The gel was further electrophoresed, initially at 700 V, then 1200 V for 2.5 hr by which time the bromophenol blue dye front from the second loading, had reached the bottom of the gel. Plates were separated and tissue paper was used along the borders of the gel to hold it to the plate during the fixation procedure which involved slowly washing the gel using 2 litres of 10% (v/v) acetic acid, 20% (v/v) ethanol in a 60 ml syringe. The gel was then dried at 100°C for 20 min. Plastic wrap was used to cover the gel before placing on film for autoradiography. Autoradiography was performed at room temperature, without the use of intensifying screens, for 16-24 hr.

2.18.10 Analysis of DNA sequences

Sequencing data was analysed using the following computer programs: Nucleic Acids Analysis System, version 1.7; the IBI Pustell Sequence Analysis Program version 4.0, the LKB DNA and protein analysis programs, DNASIS and PROSIS, the MailfastA programme from EMBL, the GCG programme package of the Wisconsin Genetics Computer Group, (Devereaux *et al.*, 1986), and CLUSTAL (Higgins and Sharp, 1988). The protein bank screened was Swissprot (January, 1990).

CHAPTER 3

CHARACTERIZATION OF THE *RFB* REGION OF *VIBRIO CHOLERAE* O1

3.1 Introduction

The O-antigen region of the *Vibrio cholerae* chromosome of both serotypes has been cloned and expressed in *E. coli* K-12 (Manning *et al.*, 1986). These clones, initially obtained from cosmid banks, were used to localize the *V. cholerae rfb* region to a 20 kb *Sac*I fragment, which was subcloned into the vector pUC18, yielding pEVX6 or pEVX7 (depending upon the orientation of the cloned segment) (Ward *et al.*, 1987). Analysis of *V. cholerae* O-antigen (VcOAg) expression by pEVX6 and pEVX7 using silver stained polyacrylamide gels, revealed an instability problem, as VcOAg expression from pEVX6 or pEVX7 was not reproducible. For this reason the 20 kb *Sac*I fragment from pEVX7 was subcloned into a pBR322 based vector pOmpV500 (Morona *et al.*, manuscript in preparation). pOmpV500 is a pBR322 vector with the *V. cholerae* outer membrane protein OmpV cloned into the *Sca*I site. Cloning of the 20 kb *rfb* region into the *Sac*I site positions it in the terminator of the *ompV* gene (Morona *et al.*, manuscript in preparation).

Although some preliminary restriction data were compiled on the O-antigen clone, a more detailed restriction map was required (Ward *et al.*, 1987). To accomplish this, subclones of the 20 kb fragment into both pUC18 (Yanisch-Perron *et al.*, 1985) and pACYC184 (Chang and Cohen, 1978) were obtained. Additionally deletion derivatives were constructed which were used for mapping, defining a minimal coding region, and for identifying the region(s) involved in serotype specificity via complementation analysis.

Definition of the minimal coding region of VcOAg would hopefully reduce the size of the plasmid expressing VcOAg, enabling easier manipulation. The *rfb* genes have been defined to 14 kb with *E. coli* O7 (Valvano and Crosa, 1989) and require a minimum of 8.9 kb and a maximum of 11.8 kb with *E. coli* 0101 (Heuzenroeder *et al.*, 1989). Thus, it seemed feasible that the *V. cholerae rfb* coding region could be reduced from 20 kb.

Transposon insertion mutants were also used to help define the coding region. This, in principle, would enable one to define different gene clusters corresponding to distinct functions.

3.2 Results

3.2.1 Cloning of VcOAg into a stable vector - pOmpV500

The cloning procedure to produce the plasmids pRMB1 and pRMB2 is summarized in Figure 3.1. Plasmid DNA of pOmpV500 was digested with the restriction enzyme *Sac*I and treated with alkaline phosphatase. This was ligated to *Sac*I digested pEVX7 and transformed into the *E. coli* strain DH1. Screening the transformants revealed plasmids possessing the 20 kb *rfb* region in both orientations. Whole cell lysates prepared from these clones (pRMB1 and pRMB2), were subjected to SDS-PAGE which was then silver stained (Figure 3.2) to detect VcOAg production. Using this method LPS with VcOAg from cells of either pRMB1 or pRMB2 was detected in comparable amounts. This implied that the *V. cholerae rfb* region was being transcribed from a (some) naturally derived promoter(s) , and not exclusively using promoters present on the *E. coli* vector. These plasmids proved to give a reproducible high level of VcOAg production when compared to the VcOAg production from cells containing pEVX6 or pEVX7.

Figure 3.1 Construction of pRMB2.

The plasmid pRMB2 was constructed by the ligation of pEVX7 restricted with *Sac*I, to similarly restricted pOmpV500, which had also been treated with alkaline phosphatase (AP). The hatched region in pEVX7 symbolizes DNA originating from pUC18, and the sotid line in pRMB2 indicates DNA originating from pOMPV500.

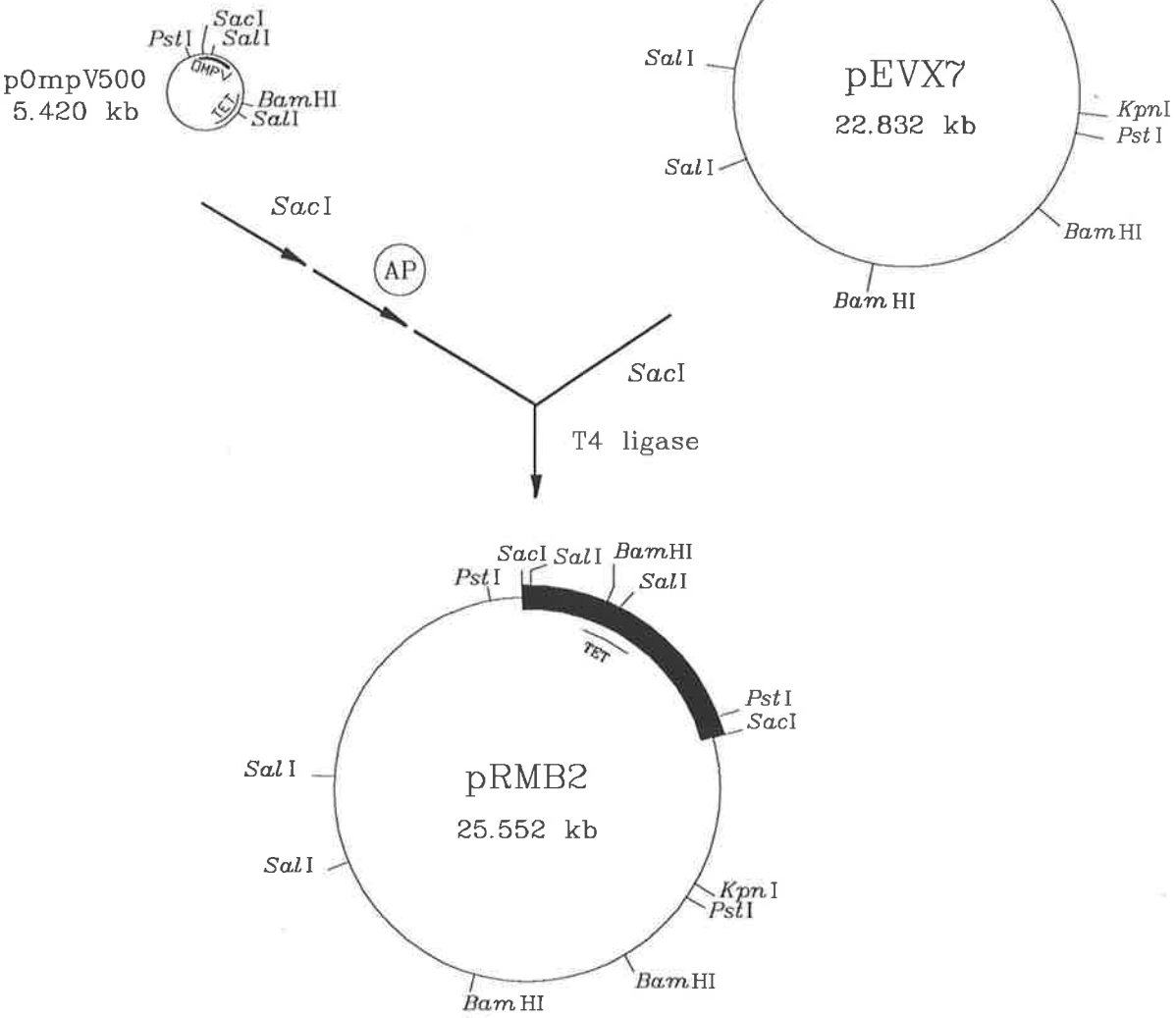
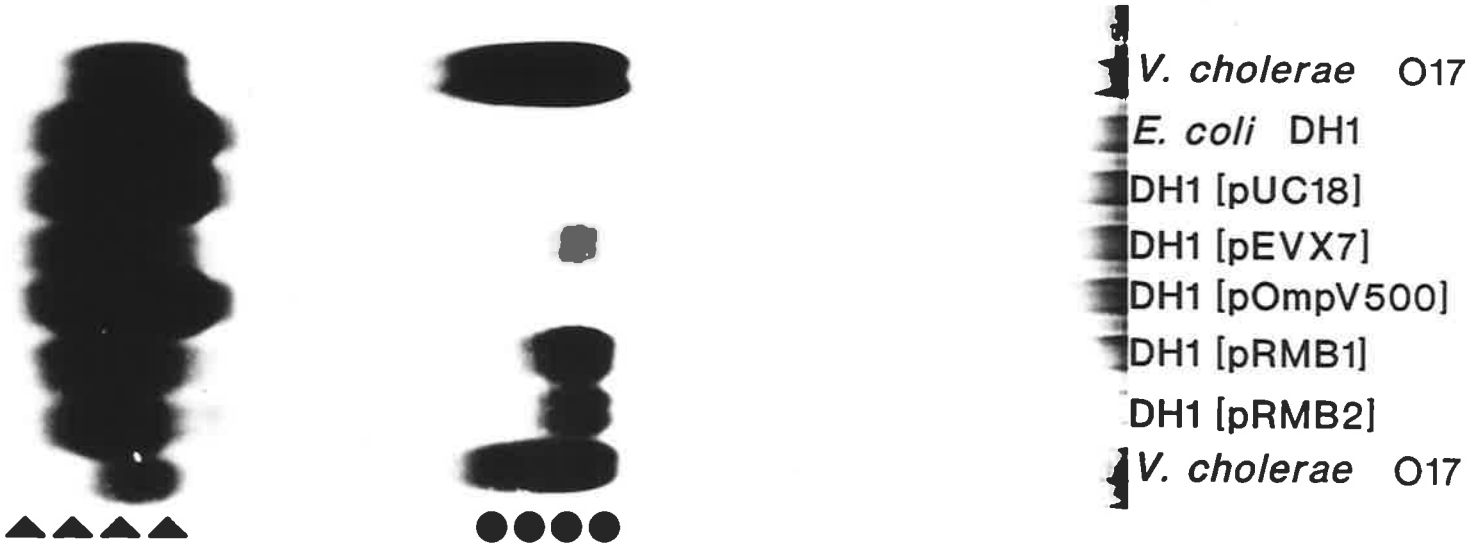

p0mpV500
5.420 kb
SacI
PstI
SalI
OMPV
ICT
BamHI
SalI
PstI
SacI
SacI
AMP
SalI
pEVX7
22.832 kb
KpnI
PstI
SalI
BamHI
BamHI
SacI
AP
SacI
T4 ligase
SacI SalI
BamHI
PstI
SalI
TET
PstI
SacI
SalI
pRMB2
25.552 kb
SalI
KpnI
PstI
BamHI
BamHI

Figure 3.2 Analysis of LPS production by *E. coli* K-12 strains harbouring pEVX6, pRMB1 and pRMB2.

Bacterial cultures were grown in NB. Samples of whole cell lysates prepared by proteinase K treatment were electrophoresed in a 20% SDS-polyacrylamide gel which was then silver stained. Symbols: solid circles indicate the position of LPS with core sugars substituted with long *V. cholerae* O-side chains; solid triangles indicate the position of LPS with core sugars only.


V. cholerae O17
E. coli DH1
DH1 [pUC18]
DH1 [pEVX7]
DH1 [pOmpV500]
DH1 [pRMB1]
DH1 [pRMB2]
V. cholerae O17

3.2.2 Restriction analysis of the *V. cholerae rfb* region

Subclones in pUC18 and pACYC184 were constructed using either the 20 kb *Sac*I fragment derived from pEVX7 or pRMB2, or pEVX10 (a pACYC184 derivative containing a *Sac*I-*Sal*I fragment of *V. cholerae* DNA which overlaps and is contiguous to the 20 kb *rfb* region) (Ward *et al.*, 1987) (Figure 3.3). These clonings were performed using the restriction enzyme *Eco*RI as this gave a number of smaller sized subclones, enabling the restriction map to be developed more precisely (Figure 3.4). The restriction endonuclease cleavage sites that were mapped were confirmed later by nucleotide sequence analysis of the *rfb* region. There are no cleavage sites within the cloned *V. cholerae* DNA for the following enzymes: *Eag*I, *Nae*I, *Nar*I, *Not*I, *Sfi*I, *Sma*I and *Xma*I. The 6 base pair restriction enzymes *Ase*I, *Eco*RV, *Hpa*II, *Dra*I, and *Hae*I cut the cloned DNA at numerous sites resulting in these sites being identified by sequence analysis only.

3.2.3 Deletion derivatives of pRMB2

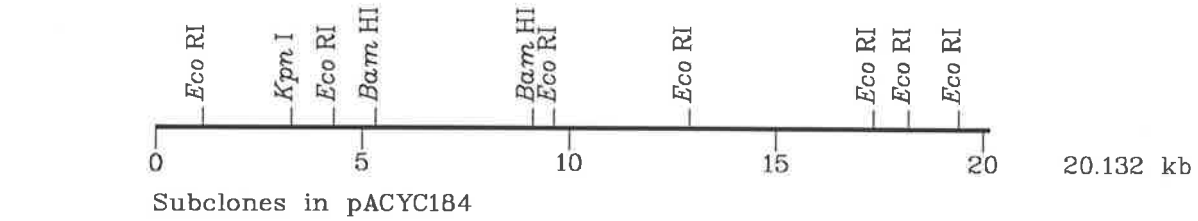
To help define the minimal amount of DNA required for VcOAg expression, a number of deletion derivatives of pRMB2 were created by partial restriction endonuclease digestion followed by dilution and ligation, as described in Chapter 2 (Figure 3.5, 3.6). The deletions in Figure 3.5 were constructed using *Eco*RI, whereas Figure 3.6 shows deletion derivatives made using the restiction enzymes *Cla*I, *Bam*HI and *Pst*I.

All of the deletion derivatives and the previously described subclones were analyzed for VcOAg production by SDS-PAGE of proteinase K treated whole cell lysates followed by silver staining, and by colony immunoblotting with a polyclonal rabbit anti-*V. cholerae* O17 serum, which had been absorbed with *E. coli* to remove any cross reacting antibodies. None of the constructs showed any VcOAg production.

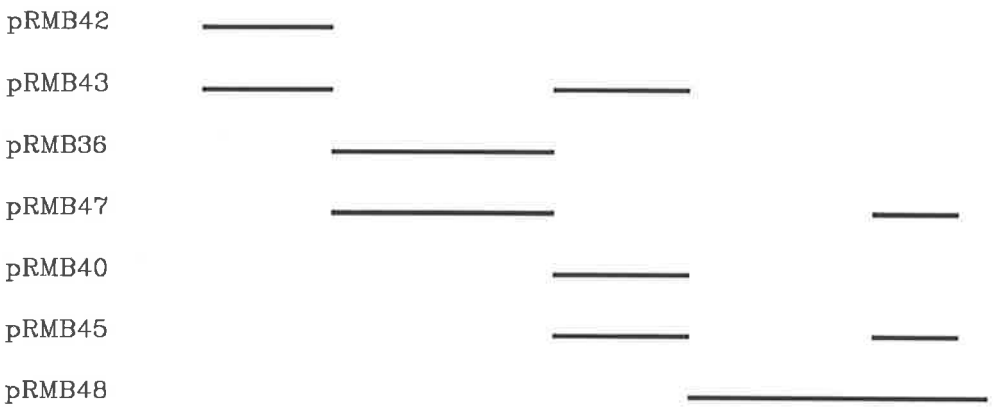
From these data it appeared that the whole 20 kb region was required for VcOAg expression in *E. coli* K-12, however, as described in Ward *et al.* (1987) comparison of

Figure 3.3 Construction of subclones.

The 20 kb *V. cholerae rfb* region isolated from either pRMB2 or pEVX7, was digested with *Eco*RI and the resulting fragments ligated to either pACYC184, or pUC18, which had been cleaved with *Eco*RI. The region of the cloned *V. cholerae* DNA is indicated.



Subclones in pACYC184



Subclones in pUC18

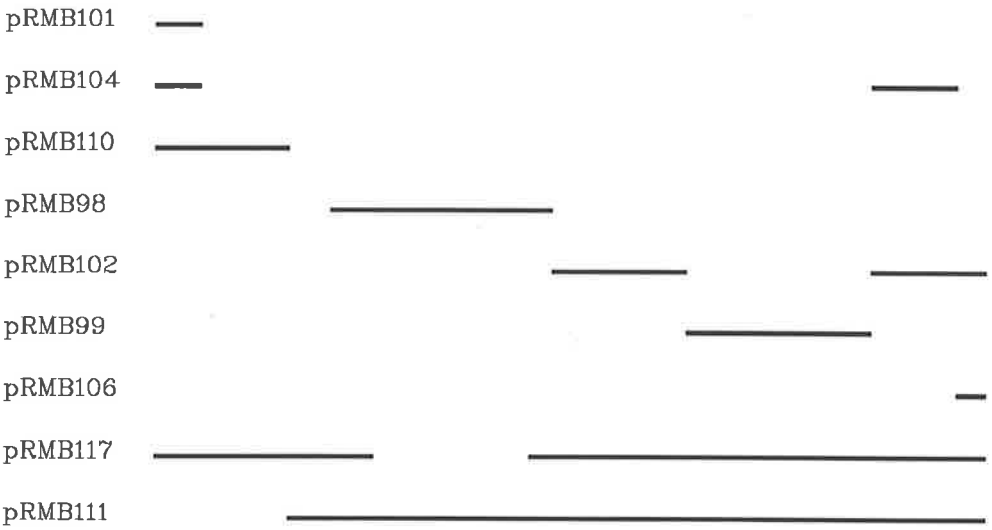


Figure 3.4 Restriction endonuclease cleavage map of the 20 kb *V. cholerae rfb* region.

The restiction map depicted here was constructed using the restriction digestion pattern obtained from incubating pRMB2, pEVX7, and many of the previously described subclones with the different restriction enzymes.

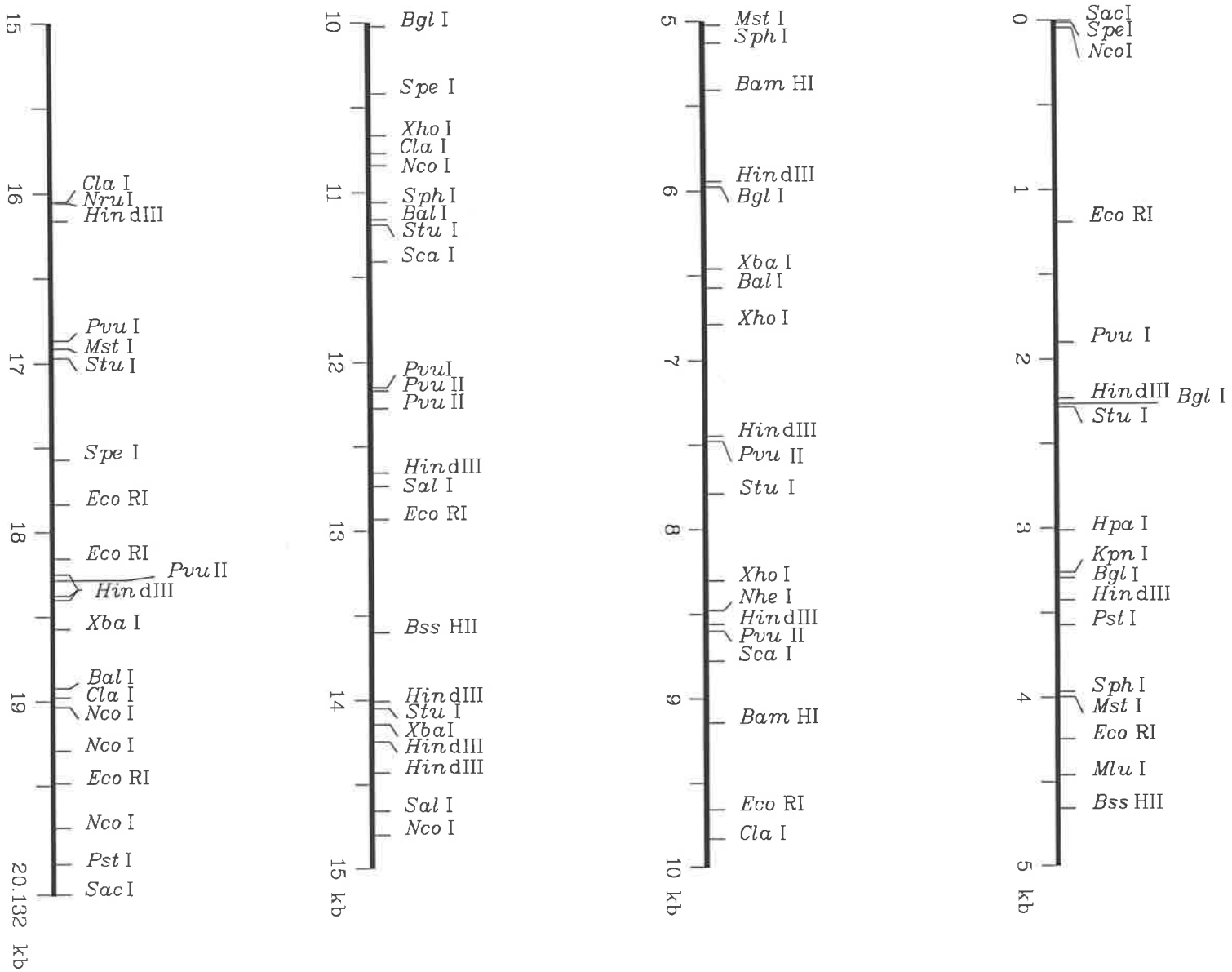

15
16
17
18
19
20.132 kb
Cla I
Nru I
Hin dIII
Pvu I
Mst I
Stu I
Spe I
Eco RI
Eco RI
Pvu II
Hin dIII
Xba I
Bal I
Cla I
Nco I
Nco I
Eco RI
Nco I
Pst I
Sac I
10
11
12
13
14
15 kb
Bgl I
Spe I
Xho I
Cla I
Nco I
Sph I
Bal I
Stu I
Sca I
PvuI
Pvu II
Pvu II
Hin dIII
Sal I
Eco RI
Bss HII
Hin dIII
Stu I
Xba I
Hin dIII
Hin dIII
Sal I
Nco I
5
6
7
8
9
10 kb
Mst I
Sph I
Bam HI
Hin dIII
Bgl I
Xba I
Bal I
Xho I
Hin dIII
Pvu II
Stu I
Xho I
Nhe I
Hin dIII
Pvu II
Sca I
Bam HI
Eco RI
Cla I
0
1
2
3
4
5 kb
Sac I
Spe I
Nco I
Eco RI
Pvu I
Hin dIII
Bgl I
Stu I
Hpa I
Kpn I
Bgl I
Hin dIII
Pst I
Sph I
Mst I
Eco RI
Mlu I
Bss HII

Figure 3.5 *Eco*RI deletion derivatives of pRMB2.

The plasmid pRMB2 was used to construct *Eco*RI deletion derivatives. These plasmids were obtained by partially digesting pRMB2 with *Eco*RI, diluting the reaction, and subsequently ligating. The filled area represents the vector used (pOmpV500).

pRMB75
pRMB69
pRMB68
pRMB70
pRMB66
pRMB93
pRMB92
pRMB78
pRMB96
pRMB77
pRMB74
pRMB71
pRMB2
Sac I
Sal I
Cla I
Eco RI
Bam HI
Pst I
Kpn I
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
25.552 kb

Figure 3.6 *Cla*I, *Bam*HI and *Pst*I deletion derivatives of pRMB2.

The plasmid pRMB2 was used to construct *Cla*I (pRMB80, pRMB81, pRMB83, pRMB85, pRMB87, pRMB90), *Bam*HI (pRMB61) and *Pst*I (pRMB63, pRMB64) deletion derivatives. These plasmids were obtained by partially digesting pRMB2 with either *Cla*I, *Bam*HI or *Pst*I, diluting the reaction, and subsequent ligating. The filled area represents the vector used (pOmpV500).

pRMB64
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
PstI
SacI
pRMB63
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
PstI
SacI
pRMB61
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
ClaI
SalI
EcoRI
SalI
ClaI
EcoRI
EcoRI
ClaI
EcoRI
PstI
SacI
pRMB90
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
ClaI
SalI
EcoRI
SalI
ClaI
ClaI
EcoRI
PstI
SacI
pRMB87
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
ClaI
EcoRI
EcoRI
ClaI
EcoRI
PstI
SacI
pRMB85
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
ClaI
EcoRI
PstI
SacI
pRMB83
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
ClaI
pRMB81
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
pRMB80
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
pRMB2
SacI SalI
ClaI
EcoRI
ClaI BamHI
SalI
PstI
SacI
EcoRI
KpnI
PstI
EcoRI
BamHI
BamHI
EcoRI
ClaI
ClaI
SalI
EcoRI
SalI
ClaI
EcoRI
EcoRI
ClaI
EcoRI
PstI
SacI
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
25.552 kb

endpoints of certain cosmid clones constructed earlier reveal a requirement of between 16 and 19 kb (see Figure 3.9).

The pRMB2 deletions have also been used to investigate another aspect of *V. cholerae* O-antigen production, that of serotype specificity. No difference in the DNA structure between Inaba or Ogawa could be detected by restriction enzyme mapping with enzymes having a 6 bp recognition sequence (Manning *et al.*, 1986). Thus, it seemed appropriate to use the deletion derivatives to help localize the region(s) involved in determining the serotype of VcOAg.

The pRMB2 deletions, and pRMB2 itself (as a control) were used in complementation analysis in an *E. coli* K-12 *recA* strain habouring pEVX22 (Inaba derived pSC101 based plasmid) (Morona *et al.*, 1990). Transformants obtained were analyzed for production of either C (Inaba), B (Ogawa), or both antigens by slide agglutination with specific typing sera. These experiments demonstrated dominance of the Ogawa phenotype over the Inaba phenotype in *E. coli* K-12.

By performing complementation analyses on all of the constructed deletions, two separate regions were identified on the 20 kb fragment (as summarized in Figure 3.7). Two plasmids pRMB61 and pRMB89, containing *Bam*HI and *Cla*I deletions respectively, were found to induce the serotypic change from Inaba to Ogawa. Construction of a *Cla*I deletion of pRMB61 resulted in the plasmid pRMM1, containing both deletions present in pRMB61 and pRMB89. Analysis of this plasmid showed it was also able to seroconvert the *E. coli* strain harbouring pEVX22. To confirm that the region *Cla*I (16029 bp) to *Sac*I (20132 bp) was required, the *Cla*I region was deleted from pRMB2 to produce pRMB90. This plasmid lost the ability to seroconvert, thereby giving rise to the confines of region 1 as being *Cla*I (16029 bp) to *Sac*I (20132 bp).

The left hand end of pRMB2 was also analyzed using the same methodology. pRMM2 which had been constructed as a *Pst*I deletion of pRMB89, was unable to serotype convert *E. coli* harbouring pEVX22, resulting in the identification of another region involved in serotype conversion - region 2 (1 bp to 5339 bp).

Figure 3.7 Summary of deletion derivatives of pRMB2 and the phenotype of pEVX22 harbouring these plasmids.

The plasmids of the pRMB series were enzymatic deletion derivatives of pRMB2. The plasmid pRMM1 was made by *Cla*I digestion of pRMB61. In a similar fashion pRMM2 was constructed by *Pst*I digestion of pRMB89. The filled area represents the vector used (pOmpV500). pOmpV500 had no effect on the phenotype of EX100 [pEVX22] (not shown). The plasmids were transformed into strain EX100 [pEVX22]. The phenotype of the transformants was scored with Ogawa or Inaba specific typing antisera (Wellcome). In, Inaba; Og, Ogawa; +, positive for indicated serotype; -, negative for indicated serotype.

In Og
pRMB2
pRMB61
pRMB89
pRMM1
pRMM2
pRMB90
pRMB77
pRMB92
SacI SalI
Cla I
Eco RI
Cla I Bam HI
Sal I
Pst I
Sac I
Kpn I
Bam HI
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
25.552 kb
REGION 2
REGION 1

As can be seen in Figure 3.7, the gap in region 2 corresponds to an area which also appears to be important for this phenomenon as shown by pRMB77 and pRMB92. These plasmids produce antigens from both serotypes, so they could be labelled as having a Hikojima serotype (Morona *et al.*, 1990). The region affected by these two deletions corresponds to an area in the vicinity of the *Eco*RI (4230 bp) site.

3.2.4 Transposon Tn*1725* mutagenesis

Both pEVX7 and pRMB2 were mutagenized with the Cm^R transposon Tn*1725* (Ubben and Schmitt, 1986) (Figure 3.8). Tn*1725* was transposed into pEVX7 and pRMB2 from plasmid pRU669 ($R_{ts}1$::Tn*1725*) selecting for Cm^R derivatives. A total of 50 independent insertions were isolated and mapped by restriction analysis with *Eco*RI to give the insertion point (Tn*1725* has *Eco*RI sites 15 base pairs from each end) and *Pst*I for orientation (due to the asymmetric position of *Pst*I sites). 82% of the resulting transposon derivatives contained gross deletions, leaving only 9 transposon insertions to be further characterized. Of these 9 transposon insertions, 2 were mapped to the vector, which had no effect on VcOAg production. The 7 remaining independent transposon insertions were mapped by restriction mapping and analyzed for VcOAg production.

The transposon insertions resulting in the plasmids; pRMB32 (insertion site at 17072 bp), pRMB15 (16370 bp), pRMB58 (4881 bp), pRMB20 (2840 bp), and pRMB56 (1973 bp) have not only been positioned on the 20 kb fragment by restriction mapping, but also by sequencing the insertion site (Figure 3.9). This was accomplished by isolating the *Eco*RI fragment which had been altered in size by the transposition event, such that the fragment contained not only DNA from the *rfb* region but also the terminal 15bp from Tn*1725*. These fragments were isolated and then cloned into M13mp18 and sequenced, as described in 2.18.8. pRMB50 has not been sequenced, but has been mapped to within the *Pst*I (3577 bp) - *Eco*RI (4230 bp) fragment. pRMB49 is more complex as this particular derivative has not only a Tn*1725* insertion mapping between the insertions of

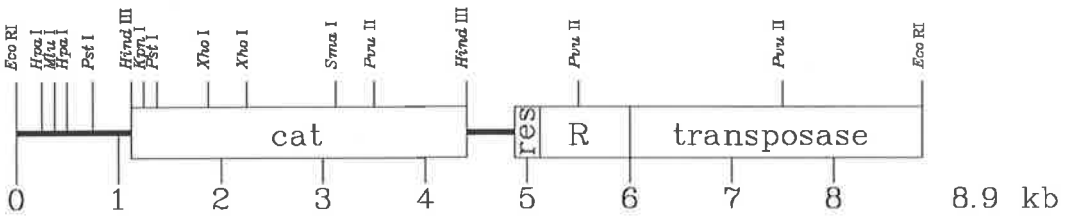
Figure 3.8 Restriction map and partial sequence of Tn*1725*.

A) The res, (resolution site), R, (resolvase gene), and transposase regions contain genes required for transposition. Tn*1725* contains the chloramphenicol transacetylase gene (cat). Relevant restriction sites and their coordinates (kb) are given with reference to the *Eco*RI site located in the left inverted repeat.

B) The sequence of the left inverted repeat, indicating the position of the *Eco*RI site. This sequence was used to identify the position of insertion of the transposon by sequencing adjacent to the insertion site (see text). Figure adapted form Ubben and Schmitt (1986). Sequence obtained from R. Schmitt, personal communication.

Tn *1725*

A


Eco RI
Hpa I
Mlu I
Hpa I
Pst I
Hind III
Kpn I
Pst I
Xho I
Xho I
Sma I
Pvu II
Hind III
Pvu II
Pvu II
Eco RI
cat
res
R
transposase
0
1
2
3
4
5
6
7
8
8.9 kb

B

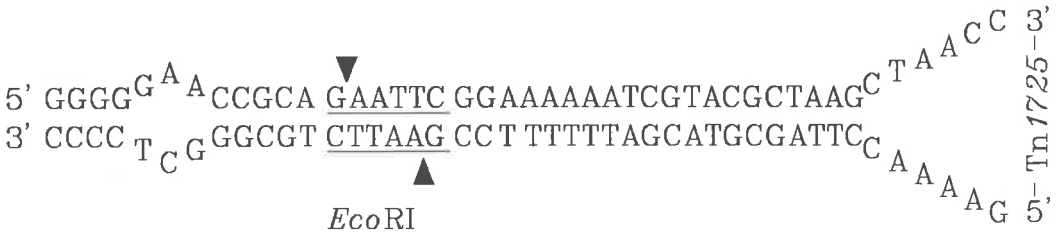
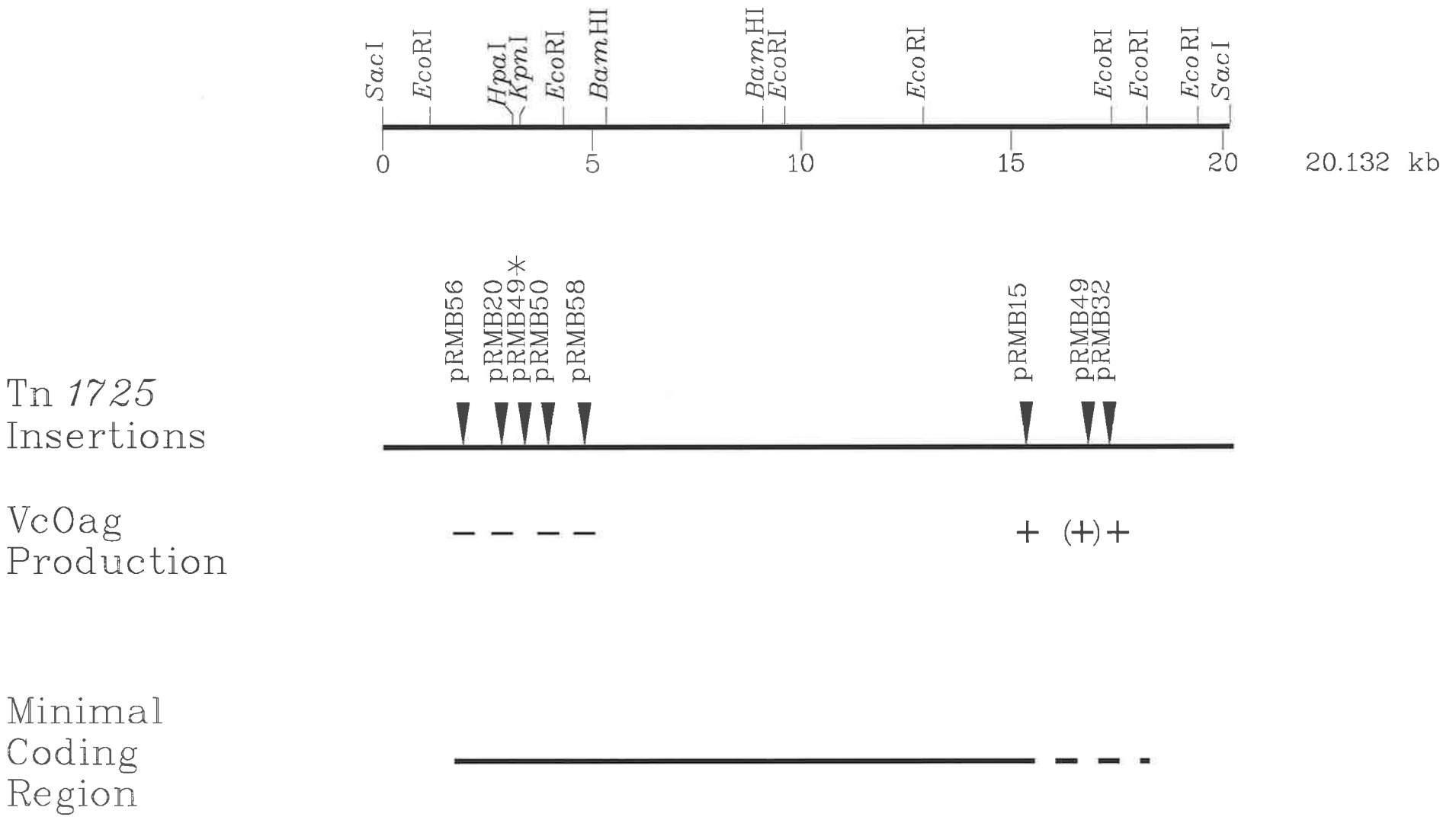

5' GGGG G A A CCGCA GAATTC GGAAAAAATCGTACGCTAAG C T A A C C 3'
3' CCCC T C G GGCGT CTTAAG CCT TTTTTAGCATGCGATTC C A A A A G 5'
EcoRI
3'- Tn1725 -5'

Figure 3.9 Transposon Tn*1725* mutagenesis of the *rfb* region

The plasmids pRMB2 (pOmpV500 based) or pEVX7 (pUC18 based) were mutagenized with the transposon Tn*1725*. The insertion points of this transposon within these plasmids were identified by restriction analysis. pRMB20, pRMB15 and pRMB32 were produced by pEVX7, whereas pRMB56, pRMB49 and pRMB50 were produced from mutagenesis of pRMB2. pRMB49* indicates that this plasmid contains not only a Tn*1725* insertion but an unknown insertion sequence.

The transposon mutants were analyzed for VcOAg production by SDS-PAGE and silver staining, as shown by the symbols + and -. (+); indicates that *E. coli* K-12 harbouring pRMB49 produced bands corrresponding to O-antigen units when analyzed by PAGE, however thase bands seem to be of a shorter length.

The minimal coding region has also been indicated.


SacI
EcoRI
HpaI
KpnI
EcoRI
BamHI
BamHI
EcoRI
EcoRI
EcoRI
EcoRI
SacI
0
5
10
15
20
20.132 kb
pRMB56
pRMB20
pRMB49*
pRMB50
pRMB58
pRMB15
pRMB49
pRMB32
Tn 1725 Insertions
VcOag Production
- - - -
+ (+) +
Minimal Coding Region

pRMB32 and pRMB15, but during the transposition event it also picked up an insertion element mapping between *Pst*I (3577 bp) and *Eco*RI (4230 bp).

The Tn*1725* insertion plasmids were all analyzed for VcOAg production by:

(1) Colony immunoblotting

(2) Silver staining following SDS-PAGE (Figure 3.10)

(3) Haemagglutination inhibition assays (HIA) (Table 3.1).

It appears that all of the insertions in the left hand side of the *rfb* clone, with the exception of pRMB49, result in VcOAg expression being eliminated.

The insertions on the right hand side are more interesting. For example, pRMB32 produces as much O-antigen as the parent clone, as determined by silver staining of SDS-PAGE. Haemagglutination inhibition assays indicated expression of the LPS determinants, as characterized by αA and αC MAbs, however, αB did not show any reactivity. Analysis of this transposon insertion in the light of the DNA sequence and predicted protein data revealed Tn*1725* inserted at 17072 bp on the *Sac*I fragment. This positioned the insertion between the *rfbS* and *rfbT* genes (see Figure 5.15). Functions for all of the proteins encoded by the *rfb* region have not yet been ascertained, however, it appears that the *rfbT* gene encodes a protein involved in serotype specificity, related to the Ogawa serotype (U.H. Stroeher, personal communication). Thus, the insertion in pRMB32 disrupted the production of the B antigen, while producing a reaction with αA and αC MAbs, and expressing O-antigen as shown by silver stained polyacrylamide gels, thereby mediating a serotype conversion. In pRMB15, although the transposon insertion maps at 16370 bp, placing it in the middle of the *rfb*R gene (16087 bp to 16788 bp), VcOAg expression could still be detected, not only by silver staining, but also by reactivity with all three O-antigen specific MAbs. This infers that for *V. cholerae* Ogawa O-antigen expression in *E. coli* this particular *rfb* gene is not required. The presence of pRMB49 leads to the production of bands on silver stained polyacrylamide gels corresponding to LPS substituted with O-antigen units, however it appears that these bands are of a shorter chain length when compared with cells harbouring the parent plasmid. Haemagglutination assays showed reactions with the αA MAb, but no reactivity

Figure 3.10 Silver stained polyacrylamide gel analysis of *E. coli* K-12 strains harbouring transposon derivatives.

Bacterial cultures were grown in NB. Samples of whole cell lysates prepared by proteinase K treatment were electrophoresed in a 20% SDS-polyacrylamide gel which was then silver stained. Symbols: circles, indicate the position of LPS with long *V. cholerae* O-side chains; triangles, indicate the position of LPS with core sugars only.

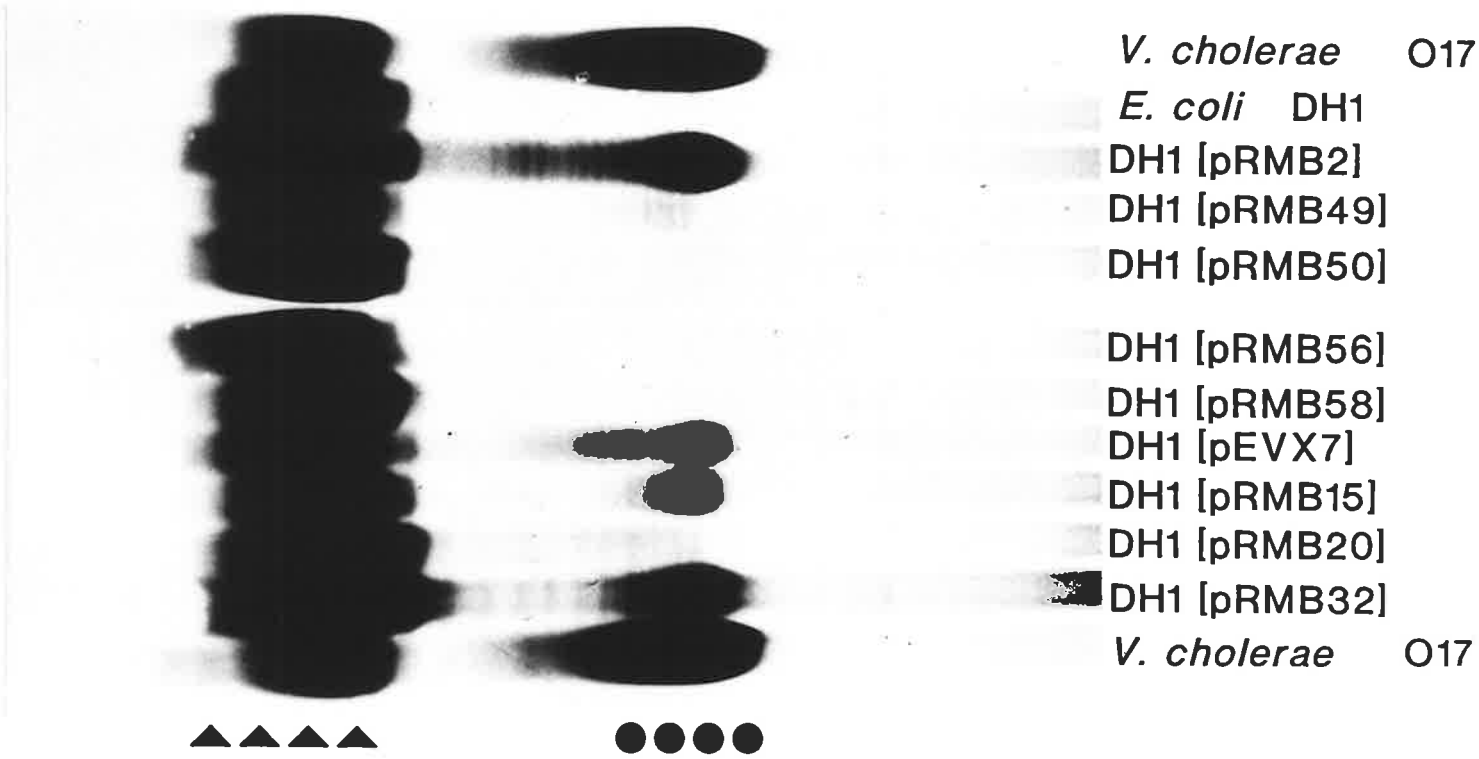

V. cholerae O17
E. coli DH1
DH1 [pRMB2]
DH1 [pRMB49]
DH1 [pRMB50]
DH1 [pRMB56]
DH1 [pRMB58]
DH1 [pEVX7]
DH1 [pRMB15]
DH1 [pRMB20]
DH1 [pRMB32]
V. cholerae O17

TABLE 3.1: Analysis of transposon mutants

STRAIN	PRODUCTION OF VcOAg BY SLVERSTAIN	COLONY IMMUNOBLOT[c]	WELLS OF INHIBITION[a]		
		antiserum[b]:	αA	αB	αC
V. cholerae O17	++++	+	10	11	11
DH1 [pOmpV500]	-	-	-	-	-
DH1 [pRMB2]	+++	+	10	10	10
DH1[pUC18]	-	-	-	-	-
DH1[pEVX7]	++	+	4	?	5
DH1[pRMB15]	++	+	8	5	10
DH1[pRMB20]	-	-	-	-	-
DH1[pRMB32]	+++	+	7	-	9
DH1[pRMB49]	+	+	3	-	-
DH1[pRMB50]	-	-	-	-	-
DH1[pRMB56]	-	-	-	-	-
DH1[pRMB58]	-	-	-	-	-

Footnote:

[a] Bacterial cells were adjusted to $1x10^{10}$ cells/ml. Sheep red blood cells were coated with *V. cholerae* O17 LPS. The data indicate the numbers of wells in which inhibition of haemagglutination was observed on addition of bacterial suspensions.

[b] The antisera used were: anti-O17, rabbit polyclonal; anti-A, mouse monoclonal 20B; anti-B, mouse monoclonal VC08; anti-C, mouse monoclonal 110-1-10A.

[c] The antiserum used was anti-O17 rabbit polyclonal.

with the αB or αC MAbs could be detected. Unfortunately, it is impossible to directly correlate these results to the presence of the transposon, due to the additional insertion element present.

3.3 Summary and conclusions

Early studies performed in *E. coli* K-12 of the *rfb* region of *V. cholerae* serotype Ogawa, used cosmid clones. Analysis of these clones showed a minimum requirement of 16 kb-19 kb, for VcOAg expression, contained within a 20 kb *Sac*I fragment. This region was subsequently subcloned into pUC18 yielding the plasmids pEVX6 and pEVX7. Although these plasmids produced high DNA yields, they were unstable within the *E. coli* strain. Thus, a different plasmid vector, pOmpV500, was chosen to overcome this problem.

Restriction mapping of the 20 kb *rfb* region was performed and later confirmed by DNA sequence data. The mapping of the 20 kb region proved invaluable for the construction of deletion derivatives, subclones and characterization of transposon insertion mutants. In turn, these constructs also helped confirm the finer mapping detail of the *rfb* region, and were very useful in DNA sequencing of the *rfb* region.

As *rfb* clones from *V. cholerae* serotype Ogawa and Inaba express their corresponding serotype in *E. coli* K-12, the genes encoding the serotype specific determinants must also be present on the 20 kb *rfb* fragment (Manning *et al.*, 1986; Ward *et al.*, 1987). Localization of these areas were attempted using deletion derivatives of the Ogawa *rfb* region in pRMB2, and analyzing them via complementation analysis in *E. coli* K-12, using the Inaba *rfb* region in a compatible vector. These studies demonstrated the requirement for 2 regions involved in serotype specificity; region 1 (16029 bp to 20132 bp) and region 2 (1 bp to 5339 bp). Analysis of the transposon derivative pRMB32, whose parent plasmid pEVX7 expresses the Ogawa phenotype, revealed a serotype conversion event, detected by haemagglutination inhibition asssay with O-antigen specific MAbs. This particular Ogawa transposon insertion, lost the ability to

produce factor B, while retaining expression of the A and C determinants. Nucleotide sequence analysis revealed this insertion to be at 17072 bp, which is between the *rfbS* and *rfbT* genes, correlating with region 1. The *rfbT* gene product has been shown to be involved in the Ogawa phenotype (U.H. Stroeher, personal communication). Thus, the Tn*1725* insertion in pRMB32 may seroconvert the *E. coli* K-12 strain by blocking expression of the *rfbT* protein.

It has still not been resolved as to why two regions are required for serotype specificity. One may postulate that the promoter for *rfbT* may require activation by an element encoded in region 2.

CHAPTER 4

Analysis of potential promoter domains of the *rfb* region of *Vibrio cholerae* O1

4.1 Introduction

A **promoter** is that sequence of DNA containing signals for the proper binding and subsequent activation of RNA polymerase holoenzyme, to a form capable of initiating transcription. The biochemical events in transcription involve a series of highly specific interactions between regulatory sequences in DNA and the cellular enzyme RNA polymerase that catalyzes this reaction. Transcription is the first stage of gene expression in both prokaryotic and eukaryotic organisms, and is the major point of regulation.

In some cases, there is an additional sequence of DNA containing signals for the specific binding of repressor or activator proteins/complexes, that can modulate the activity of the promoter.

In addition to the most abundant sigma (σ) subunit of RNA polymerase (which in *E. coli* is σ^{70}), both Gram-negative and Gram-positive bacteria employ alternative σ factors that confer different promoter specificties on the core form of RNA polymerase (Doi and Wang, 1986).

Sigma factors are used to regulate genes with diverse biological roles. The first alternative σ factors were identified in *B. subtilis* and its phages (Fox *et al.*, 1976; Tijian *et al.*, 1975), involving sporulation. Recent studies have also shown flagellin genes to be transcribed by alternate σ factors. In *E. coli*, *B. subtilis* and *Rhizobium melilotti*, σ^{28} in conjunction with core RNA polymerase, is required for recognition of flagella promoters (Arnosti and Chamberlin, 1989; Helmann *et al.*, 1988).

σ^{54}, a minor σ subunit in *E. coli* (although under certain growth conditions it can play a vital role) enables RNA polymerase to recognize an unique set of promoter sequences (CTGGYAYR-N_4-TTGCA) with the invariant GG and GC sequences centered at -24 and -12 from the transcription start point (Hunt and Magasanik, 1985). This particular σ factor was originally identified by its involvment in transcription of genes encoding enzymes of nitrogen assimilation and nitrogen fixation (Kustu *et al.*, 1989). This σ factor has since been shown to be involved in transcription of genes whose products have diverse physiological roles (Birkmann *et al.*, 1987; Dixon, 1986; Johnson and Downie, 1984). For example, σ^{54} is involved in transcription of genes: involved in nitrogen fixation (Haselkorn, 1986); encoding amino acid transport components (Magasanik, 1982); involved in toluene catabolism (Dixon, 1986); and encoding pili in *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* (Johnson *et al.*, 1986; Meyer *et al.*, 1984). However, for transcription to occur using σ^{54}, an activator protein (NtrG, or NrI) is required to produce open promoter complexes (see below).

The σ^{70} holoenzyme is the major *E. coli* RNA polymerase species, and the one which has been studied in most detail.

The nucleotide sequences of many different promoters transcribed by σ^{70}, have been compiled and many similarites have become apparent;

(1) The promoter sequence is asymmetric, thereby ensuring mRNA synthesis is initiated in the correct direction and on the sense strand,

(2) Two highly conserved regions have been identifed at the -10 and -35 region on the promoter (Pribnow, 1975; Schaller *et al.*, 1975; Siebenlist *et al.*, 1980). They are located ~35 bp and 10 bp upstream from the first nucleotide of the transcript, and have consensus sequences of TTGaca and TAtaaT; the capital letters denoting highly conserved bases and those in lower case show larger variation (Hawley and McClure, 1983; Jaurin *et al.*,1982; Siebenlist *et al.*, 1980; von Hippel *et al.*, 1982). The sequences upstream and downstream from the -35 regions are often AT rich (Siebenlist *et al.*, 1980),

(3) The -10 and -35 promoter regions are generally separated by 16-19 bp with 17 bp being the most frequent and resulting in maximum promoter strength (Hawley and McClure, 1983; Siebenlist *et al.*, 1980; Stefano and Gralla, 1982), where the term promoter strength refers to the relative rate of synthesis of full length RNA product from a given promoter, and

(4) The starting nucleotide (+1) is usually a purine with A being predominant (Siebenlist *et al.*, 1980).

A promoter encodes a program for a complex process, which can be subdivided into 4 major stages;

(1) recognition of the promoter sequence by RNA polymerase to produce a stable or closed promoter complex,

(2) isomerization of this complex into a conformation capable of initiation, termed the "open promoter polymerase complex",

(3) initiation of mRNA synthesis and the first phase of elongation,

(4) transition into an elongation complex and promoter clearance. (Bujard *et al.*, 1987; Kammerer *et al.*, 1986; McClure, 1985; Tachibana and Ishihama, 1985).

Each of these stages may be rate limiting for the overall function of a promoter, implying that promoters of identical strength may differ in their structure due to alternate functional optimizations (Brunner and Bujard, 1987; Bujard, 1980; Deuschle *et al.*, 1986).

Although most of the information essential for the function of an unregulated σ^{70} *E. coli* promoter is stored between the -35 region and the +1 position, sequences flanking this area have also been shown to have some affect on the promoter. This is not surprising, since when the *E. coli* RNA polymerase is bound to the promoter, it can cover approximately 70 bp, from -45 to +20 (Kammerer *et al.*, 1986; Schmitz and Galas, 1979; Siebenlist, 1980; Simpson, 1982). Conserved sequences have been found in some strong promoters both upstream and downstream of the 35 bp "core" region (Bujard, 1980; Bujard *et al.*, 1983; Gentz and Bujard, 1985; Kammerer *et al.*, 1986).

The location of promoters can also play a regulatory role in transcription initiation. Closely spaced promoters can be geometrically divided into three classes:

(1) $\rightarrow\rightarrow$ **tandem** - two or more promoters are orientated in the same direction and transcribe the same gene or operon;

(2) $\leftarrow\rightarrow$ **divergent** - two RNA polymerases can bind within a common region and transcribe in opposite directions into separate genes or operons. This class of promoters can be further subdivided;

(a) back to back promoters with intervening DNA between the promoters

$\leftarrow P_1P_2\rightarrow$

(b) promoters can overlap $\leftarrow P_1$

$P_2\rightarrow$

(c) promoters can be face to face $\leftarrow P_1$

$P_2\rightarrow$

(Beck and Warren, 1988);

(3) $\rightarrow\leftarrow$ **convergent** - RNA polymerases oppose one another and transcribe both strands of the DNA over a common interval (Hawley and McClure, 1983; Horowitz and Platt, 1982; McClure, 1985).

Gene fusion techniques have proved extemely useful in the study of prokaryotic gene regulation. This has been accomplished using several specially developed vector systems (Adams and Hatfield, 1984; Casadaban and Cohen, 1980; Close and Rodriguez, 1982; Chak and James, 1985; Gentz *et al.*, 1981; Hirano *et al.*, 1987; Koop *et al.*, 1987; McKenney *et al.*, 1981; Rosenberg *et al.*, 1983). These plasmid vectors each contain a gene that codes for an assayable product but does not include the promoter for that gene. DNA fragments to be analyzed for promoter activity can then be inserted upstream from this gene and measurements of the amount of gene product are used to determine the level of transcription from the promoter(s) within the fragment. Reporter genes such as; galactokinase (*galK*) (Adams and Hatfield, 1984; Chak and James, 1985; McKenney *et al.*, 1981), ß-galactosidase (*lacZ*) (Evans and Dennis, 1985; Koop *et al.*, 1987), chloramphenicol acetyl-transferase (*cat*) (Close and Rodriguez, 1982), tetracycline

resistance (*tet*) (Thomas and Drabble, 1985), aminoglycoside phosphotransferase (*aph*) (Bibb *et al.*, 1985), and alkaline phosphatase (*phoA*) (Schneider and Beck, 1986) have been used in this manner to analyze transcriptional signals. Both high and low copy number vectors have been constructed (Hirano *et al.*, 1987; Koop *et al.*, 1987).

This chapter describes the identification of putative promoter regions on the *V. cholerae rfb* region by RNA polymerase binding data. These regions were then further localized using one of the described vector systems, having a promoterless *galK* gene. Constructs obtained were then analyzed for the relative promoter strength using a [^{14}C]-galactose galactokinase assay.

4.2 Results

4.2.1 RNA polymerase binding data

Sites for *in vitro* binding of *E. coli* σ^{70} RNA polymerase on the 20 kb *Sac*I fragment were determined by electron microscopy. Plasmids pEVX6 and pEVX7 which have the 20 kb *rfb* region, encoding O-antigen production, cloned in opposite orientations in pUC18 (possessing an unique *Sma*I site within the pUC18 polylinker DNA) were compared in this study. Figure 4.1 shows the DNA with the RNA polymerase molecules bound along its length.

Figure 4.2 summarizes the data obtained by computer analysis using the location of the known pUC18 (*lac* and *bla*) promoters as references. Such a procedure gave excellent correlation when applied to bacteriophage lambda (Vollenweider and Szybalski, 1978), the minichromosome pCM959 (Morelli *et al.*, 1981) and the F factor of *E. coli* (Manning *et al.*, 1984).

It can be seen from the relative number of RNA polymerase molecules bound to the DNA length, that there are four major regions of RNA polymerase binding, (A, B, C, and D), suggesting that promoters responsible for the transcription of the *rfb* region are located within these areas. However, each region may contain more than one binding

Figure 4.1 RNA polymerase binding to pEVX6.

An electron micrograph showing *E. coli* RNA polymerase molecules bound on DNA (shown by arrows), using 11,000x magnification. A is pUC18, and B is pEVX6.

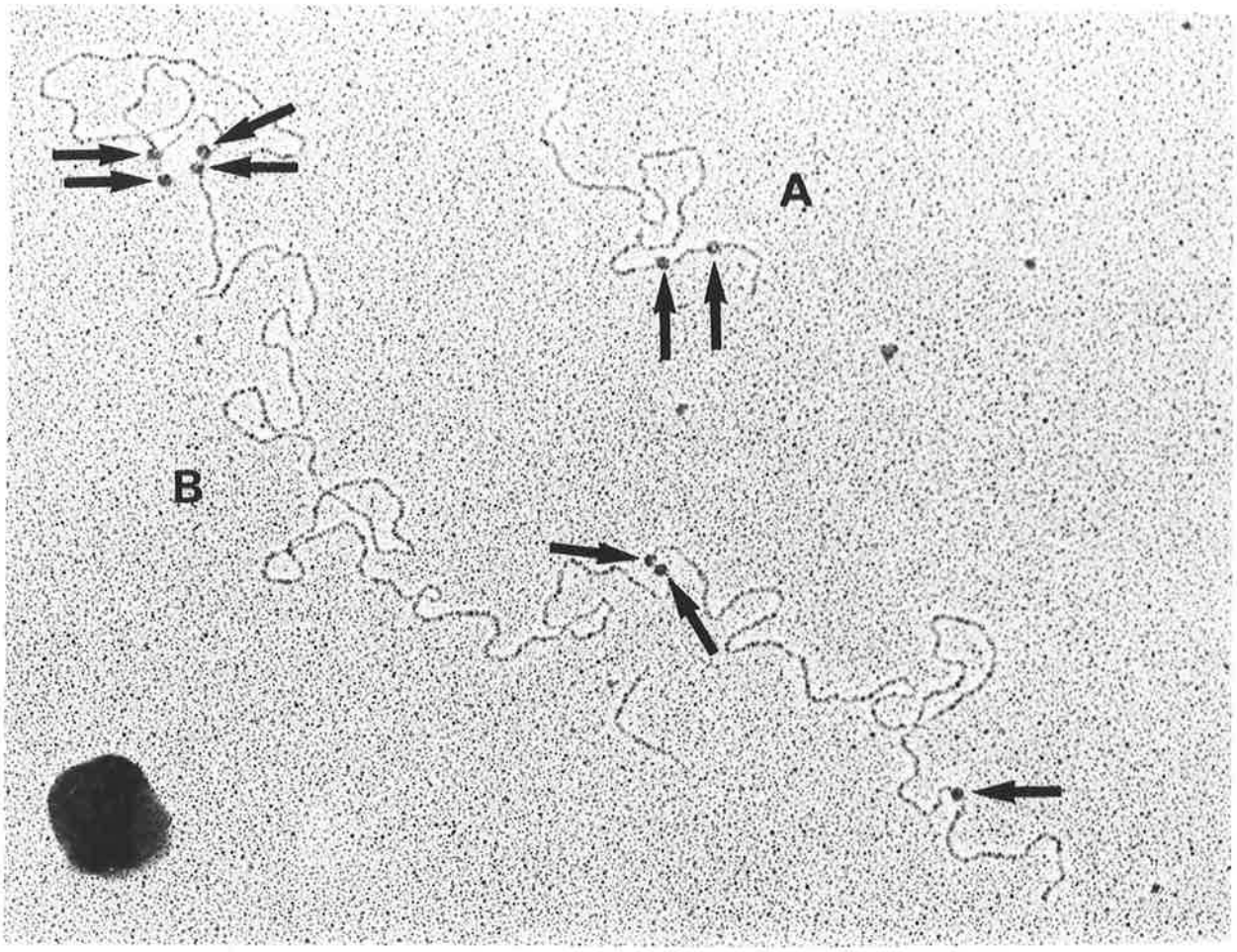
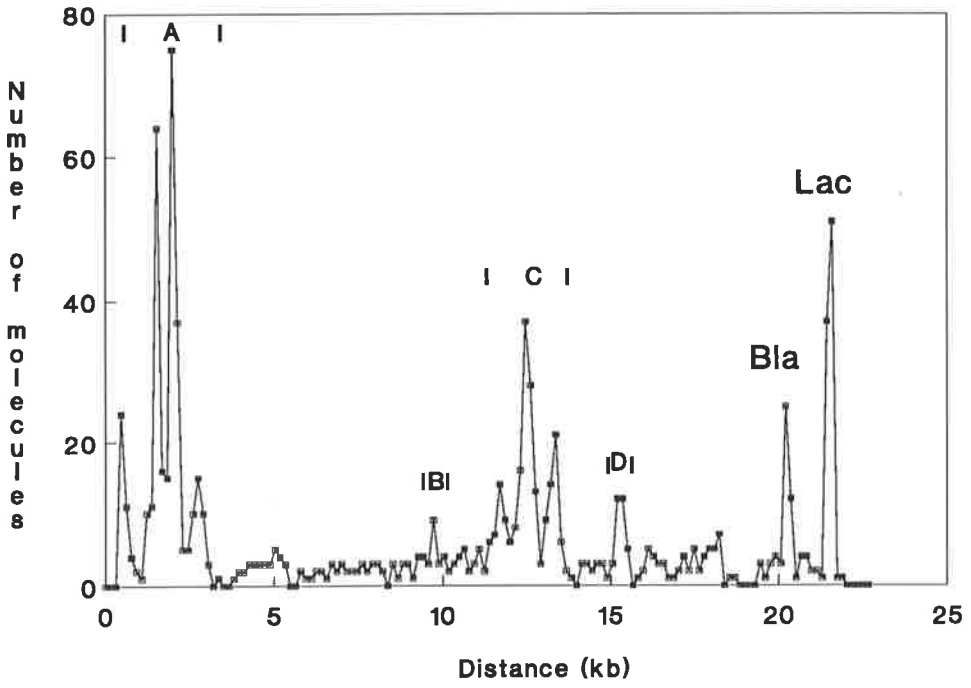

A
B

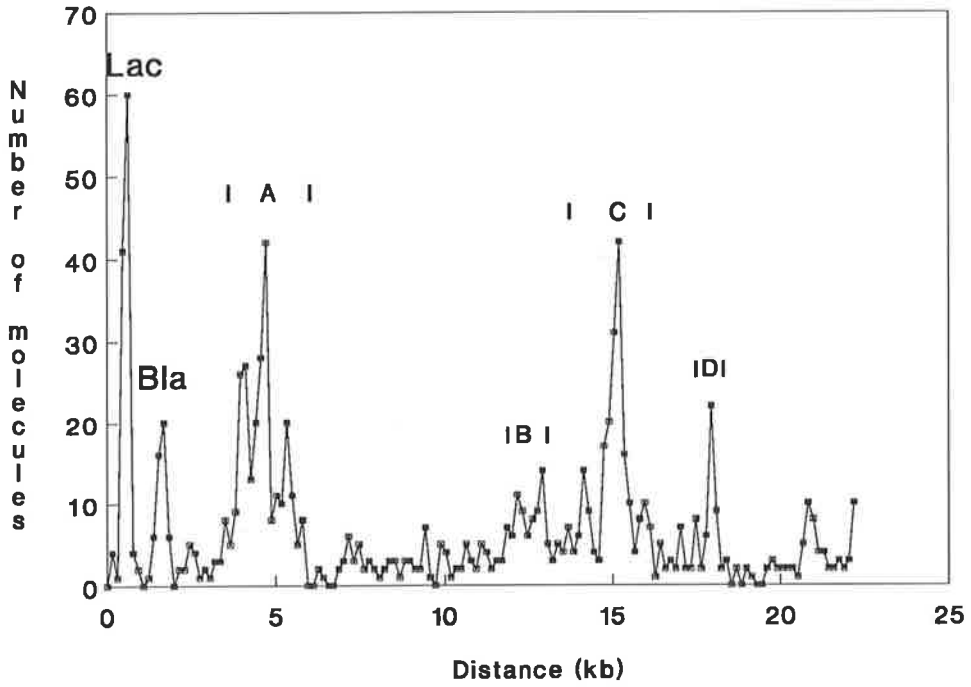
Figure 4.2 Graphic representation of RNA polymerase binding to pEVX6 and pEVX7.

The plasmids pEVX6 and pEVX7 were used to study areas where RNA polymerase bound. There are four main areas of binding as denoted by the letters A, B, C, and D. Peaks of RNA polymerase binding seen in the vector pUC18 are marked accordingly (Bla and Lac).

RNA polymerase binding within pEVX6



RNA polymerase binding within pEVX7



domain, as seen in peak A, and by the close proximity of RNA polymerase molecules in Figure 4.1. The first lies between .03 kb and 3.8 kb, (corresponding roughly to be between *Sac*I and *Sph*I), the second major binding area lies between 9.2 kb and 10.5 kb, (*Bam*HI and *Cla*I), region C is placed between 11.5 kb and 13.8 kb (*Sph*I and *Hin*dIII), and region D is between 14.8 kb to 15.8 kb (*Sal*I to *Cla*I).

Compared to other systems studied by the same methodology (Morelli *et al.*, 1981; Vollenweider and Szybalski, 1978) a very high level of background binding can be seen in the graphs (Figure 4.2). This may result in the masking of some areas of weak RNA polymerase binding. One must also remember that *V. cholerae* DNA is being analyzed using an *E. coli* system, which may also limit the detection of RNA polymerase binding regions, especially if the *rfb* gene promoters are transcribed using different σ factors.

4.2.2 Analyses using promoter detection vectors

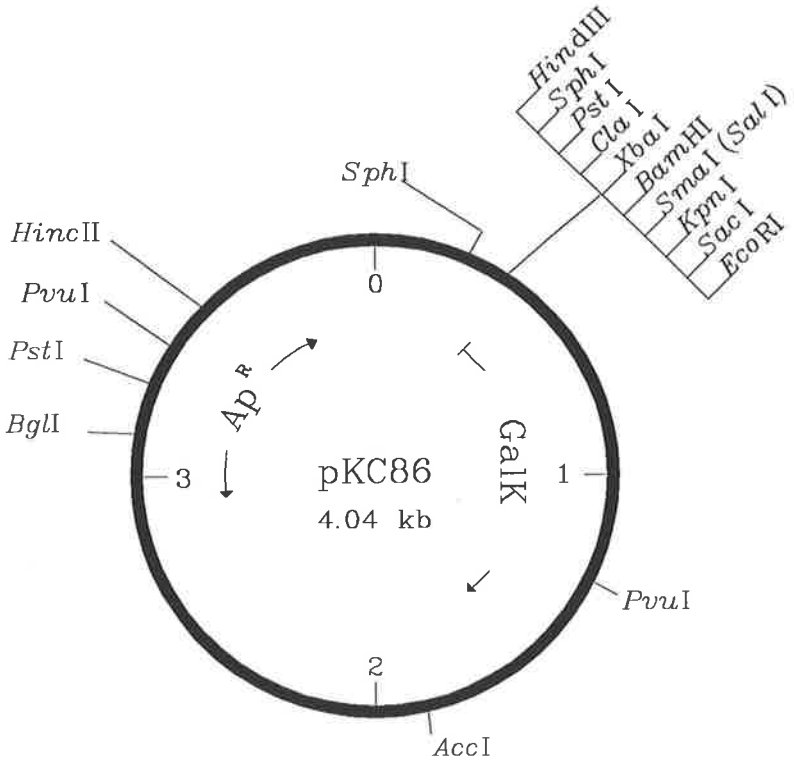
The promoter detection vectors pKC86 and pKC87 were constructed by insertion of the pUC19 polylinker proximal to a promoterless galactokinase gene (*galK*), present in either pED101 or pKO1 respectively (Chak and James, 1985) (Figure 4.3). To facilitate clonings, a *Sal*I linker was inserted into the *Sma*I site of both pKC86 and pKC87, thereby increasing the versatility of the system. Analyses of the 20 kb *Sac*I fragment were performed utilizing constructs in all four vectors.

Figure 4.4 shows analysis of the region *Sac*I (0 kb) to *Bam*HI (5.3 kb). Analyses of the plasmid pRMB264, and its (*Hin*dIII or *Eco*RI) deletion derivatives have indicated the presence of a promoter between the *Eco*RI (1.2 kb) and the *Hin*dIII (2.2 kb) sites, referred to as P_2 (Fig. 4.8). This promoter present in the plasmid pRMB278, which is the smallest clone containing this promoter, is orientated towards the right hand side of the *rfb* region, as conventionly drawn in this thesis.

Analysis of clones containing this region in the opposite direction, as depicted by plasmids pRMB268, pRMB270, and pRMB276, resulted in the identification of the

Figure 4.3 Promoter detection vectors pKC86 and pKC87.

The promoter detection vectors pKC86, and pKC87 were used to define areas of promoter activity within the 20 kb *rfb* region. These plasmids contain the *bla* gene for antibiotic selection and a promoterless *galK* gene for identifying promoter activity. Diagram adapted from Chak and James (1985).


HindIII
SphI
PstI
ClaI
XbaI
BamHI
SmaI (SalI)
KpnI
SacI
EcoRI
SphI
HincII
PvuI
PstI
BglI
ApR
pKC86
4.04 kb
GalK
PvuI
AccI
0
1
2
3

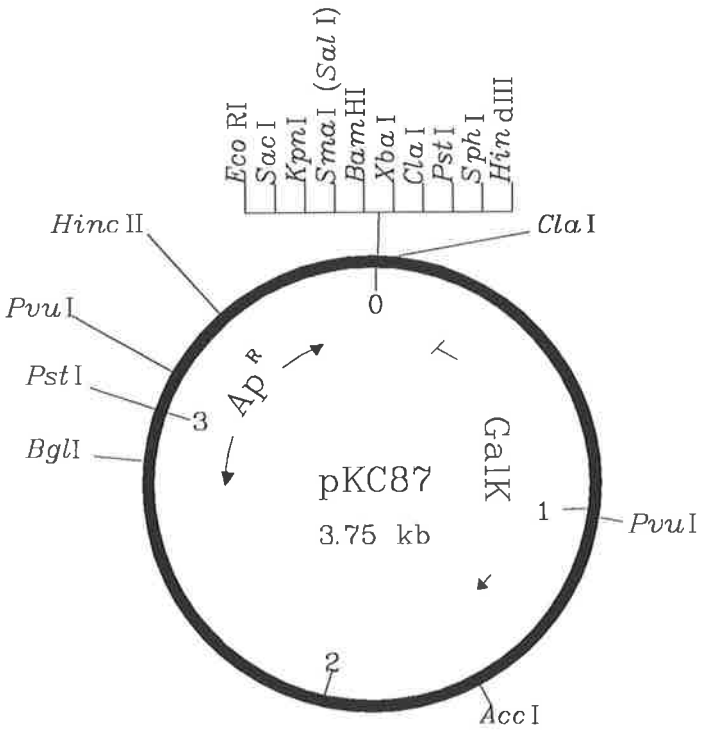
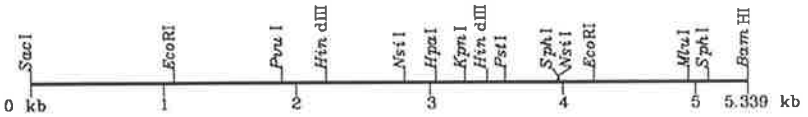
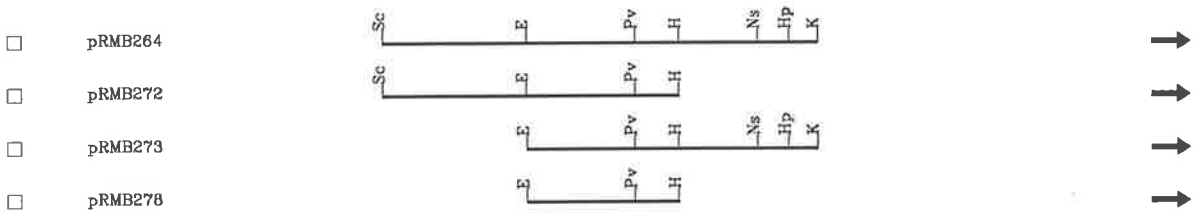

EcoRI
SacI
KpnI
SmaI (SalI)
BamHI
XbaI
ClaI
PstI
SphI
HindIII
ClaI
HincII
PvuI
PstI
BglI
ApR
pKC87
3.75 kb
GalK
PvuI
AccI
0
1
2
3

Figure 4.4 Subclones of the region *Sac*I (0 kb) to *Bam*HI (5.3 kb) in the promoter detection vectors pKC86 and pKC87.

Symbols: triangle, designates clones in pKC86; square, designates clones based on pKC87; arrows denote position of *galK* gene; open arrows indicate a negative phenotype on TZGal plates; solid arrows indicate a positive phenotype on TZGal plates, (*) indicates end point of deletion created by nested deletion analysis.

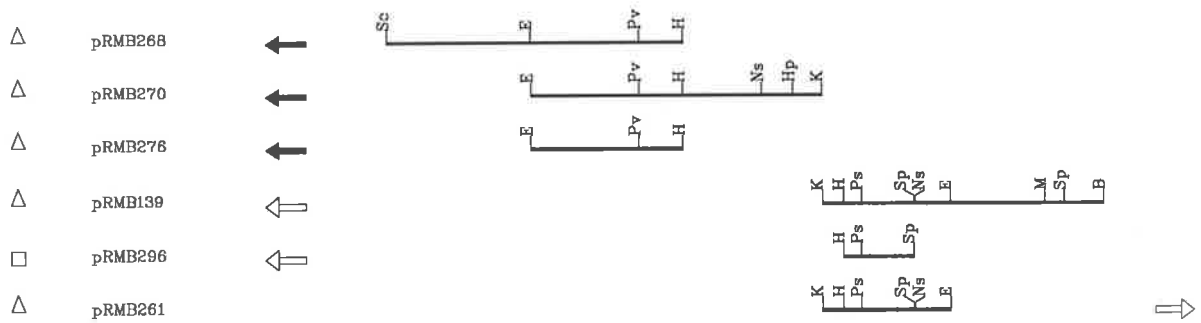


Derivatives of pRMB264



Derivatives of pRMB134

Independent subclones



promoter P_1. This promoter is also placed between the *Eco*RI (1.2 kb) and the *Hin*dIII (2.2 kb) sites, however, it would transcribe in the opposite direction to P_2. The locations of these promoters, P_1 and P_2, correlate with two of the peaks of RNA polymerase binding seen in region A of Figure 4.2.

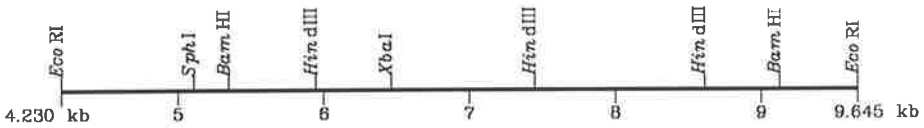
Derivatives of pRMB134, may localize yet another promoter (P_3), between the *Sph*I (3.9 kb) and *Bam*HI (5.3 kb) sites, as seen in Figure 4.4, by the plasmids pRMB134, and pRMB203. The nested deletion derivatives pRMB228 and pRMB231 (see 2.10.10) are negative, as is pRMB226, a *Sph*I deletion of pRMB134. This corresponds to placement of the promoter between the *Sph*I and *Bam*HI site. As seen with pRMB139, there is no promoter activity from this region in the opposite orientation. However, when interpreting the results seen in Figure 4.5, i.e. pRMB159, pRMB179, pRMB180, and pRMB181, it first appears that everything is in accordance with the location of P_3, except for pRMB180, which shows no *galK* activity. Therefore, it may be that when the *Bam*HI region is placed next to the *galK* gene in the vector, a promoter detected with pRMB203, pRMB134, and pRMB181, is constructed. Although pRMB180 contains this region (P_3), the *Bam*HI site at 5.3 kb is not placed in the same juxtaposition to the *galK* gene.

Figure 4.5 also depicts the localization of a promoter (P_4) between *Bam*HI (5.3 kb) and *Xba*I (6.4 kb) shown in pRMB200, although no peak was seen with RNA polymerase binding. The *Hin*dIII deletion pRMB204, and the nested deletion derivative pRMB238, confirm these data (as do pRMB190 and pRMB298), and also position the promoter between the *Bam*HI site and about 150 bp before the *Hin*dIII site. No promoter activity was detected in the opposite orientation, as demonstrated by the negative phenotypes of pRMB197, pRMB288, pRMB210, pRMB212, and pRMB209.

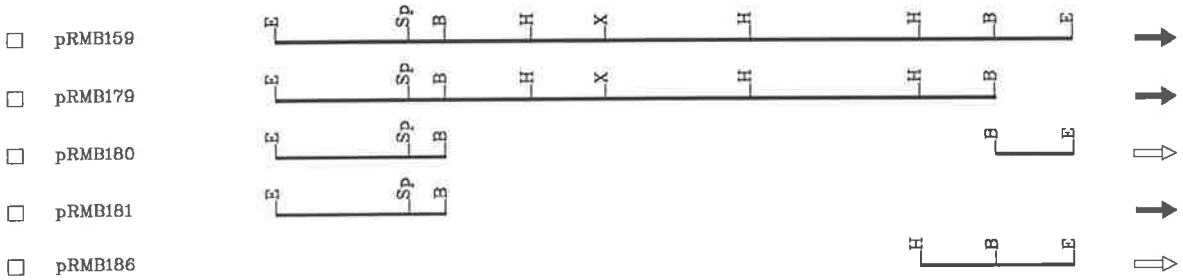
Figure 4.6 shows 2 subclones, pRMB136 and pRMB300, derived from the region 8.6 kb to 14.6 kb. pRMB300 is negative whereas pRMB136, which possesses the *Sal*I (12.7 kb) to *Sal*I (14.6 kb) region has a positive phenotype. This positions the promoter designated as P_5. Unfortunately, no subclones were constructed in this region to localize the promoter activity more accurately. However, as described in 4.2.4, this promoter

Figure 4.5 Subclones of the region *Eco*RI (4.2 kb) to *Eco*RI (9.6 kb) in the promoter detection vectors pKC86 and pKC87.

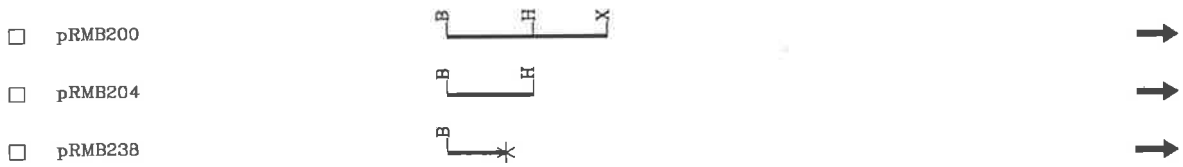
Symbols: triangle, designates clones in pKC86; square, designates clones based on pKC87; arrows denote position of *galK* gene; open arrows indicate a negative phenotype on TZGal plates; solid arrows indicate a positive phenotype on TZGal plates, (*) indicates end point of deletion created by nested deletion analysis.



Derivatives of pRMB159



Derivatives of pRMB200



Independent subclones

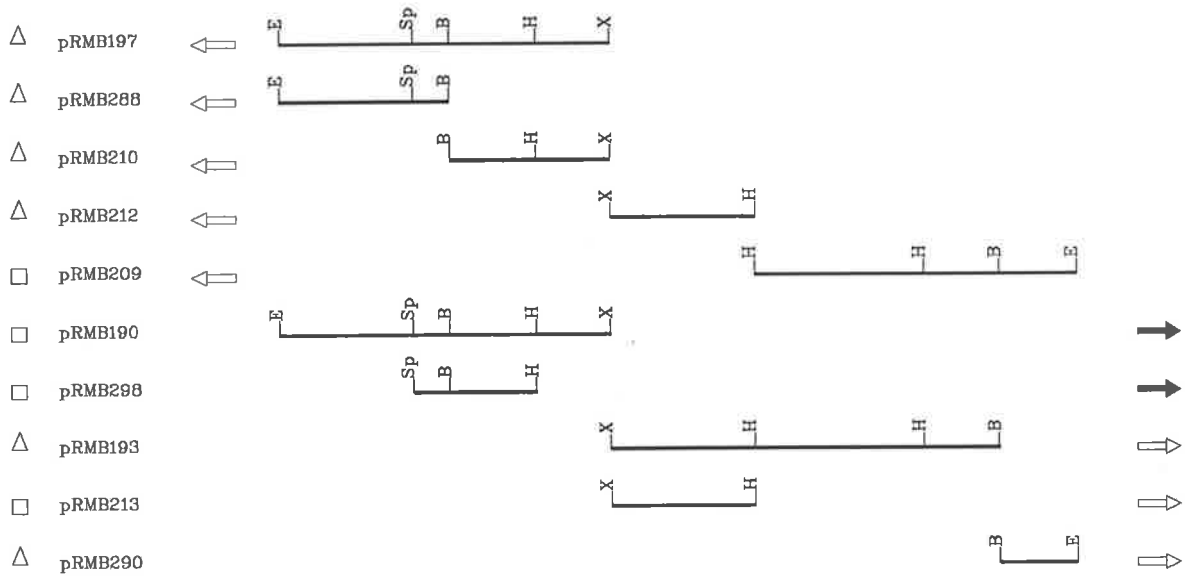
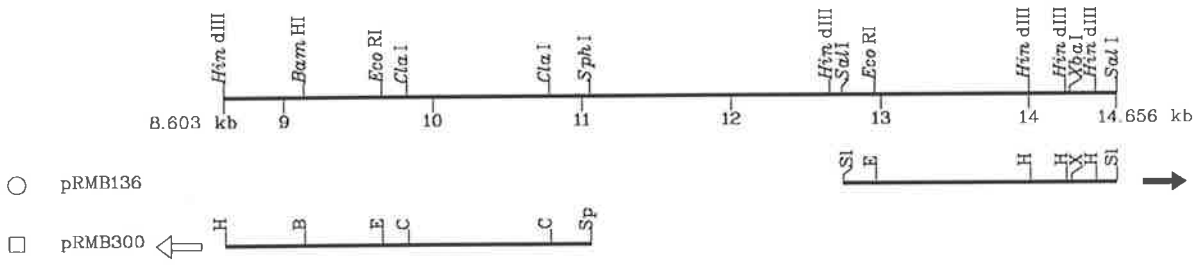


Figure 4.6 Subclones of the region *Hin*dIII (8.6 kb) to *Sal*I (14.6 kb) in the promoter detection vectors pKC87 and the *Sal*I linker modified pKC86.

Symbols: square, designates clones based on pKC87; circle, designates clones in *Sal*I linker modified pKC86; arrows denote position of *galK* gene; open arrows indicate a negative phenotype on TZGal plates; solid arrows indicate a positive phenotype on TZGal plates.


Hin dIII
Bam HI
Eco RI
Cla I
Cla I
Sph I
Hin dIII
Sal I
Eco RI
Hin dIII
Hin dIII
Xba I
Hin dIII
Sal I
8.603 kb
9
10
11
12
13
14
14.656 kb
Sl
E
H
H
X
H
Sl
pRMB136
pRMB300
H
B
E
C
C
Sp

bears no relation to the predicted protein map, and therefore may be a constructed promoter, although this region correlates with region C seen in Figure 4.2.

Figure 4.7 indicates the presence of 2 more promoters. The first, P_6, as shown by pRMB151 and its derivatives, lies between *Sal*I (14.6 kb) and *Cla*I (16 kb), and is oriented in the opposite direction to the transcription of the *rfb* region, as determined from analysis of the DNA sequence and potential ORFs. This may also be a constructed promoter, as analysis of the sequence of the region does not indicate any open reading frames in this direction.

Analysis of pRMB133 and its *Eco*RI deletion derivatives pRMB171, pRMB173, pRMB160, its *Cla*I deletion derivatives pRMB162, pRMB163 and the *Hin*dIII deletion derivative pRMB168, have localized another region of promoter activity (P_7) between *Eco*RI (19.4 kb) and *Sac*I (20.1 kb), reading rightwards out from the *rfb* region. This promoter, which has the highest GalK activity by comparison with the other promoters (see Table 4.1), is not involved in expression of the *rfb* operon, as shown in Fig. 3.9. P_7 appears to be involved in expression of DNA distal to ORF2 and ORF3. However, there is still the possibility that P_7 is a constructed promoter, as a subclone of this promoter region without using the *Sac*I site to position the DNA in front of the *galK* gene was not constructed.

Promoter locations are summarized in Figure 4.8.

4.2.3 Galactokinase assays

The galactose operon (*gal*) of *E. coli* is known to consist of 3 structurally contiguous genes which specify the enzymes required for the metabolism of galactose; *galE* (uridine diphosphogalactose-4-epimerase), *galT* (galactose-1-phosphate uridyltransferase) and *galK* (galactokinase). The product of the *galK* gene catalyzes the first reaction of galactose catabolism:

galactose + ATP → galactose-1-phosphate + ADP.

Figure 4.7 Subclones of the region *Sal*I (14.6 kb) to *Sac*I (20.1 kb) in the promoter detection vectors pKC86, pKC87 and their *Sal*I linker modified derivatives.

Symbols: triangle, designates clones in pKC86; square, designates clones based on pKC87; circle, designates clones in *Sal*I linker modified pKC86; diamond, designates clones in *Sal*I linker modified pKC87; arrows denote position of *galK* gene; open arrows indicate a negative phenotype on TZGal plates; solid arrows indicate a positive phenotype on TZGal plates.

Sal I
EcoRV
EcoRV
ClaI
HindIII
PvuI
EcoRV
EcoRI
EcoRI
HindIII
HindIII
HindIII
XbaI
ClaI
EcoRV
EcoRI
PstI
SacI
14.656 kb
15
16
17
18
19
20
20.131 kb
Derivatives of pRMB151
pRMB151
pRMB178
pRMB177
pRMB175
pRMB165
pRMB207
pRMB205
Derivatives of pRMB133
pRMB133
pRMB171
pRMB173
pRMB162
pRMB163
pRMB168
pRMB160
Independent subclones
pRMB215
pRMB152
pRMB126
pRMB221
pRMB219
pRMB258
pRMB128
pRMB255
pRMB218
pRMB224

TABLE 4.1 Promoter activities of plasmid constructs in pKC vectors

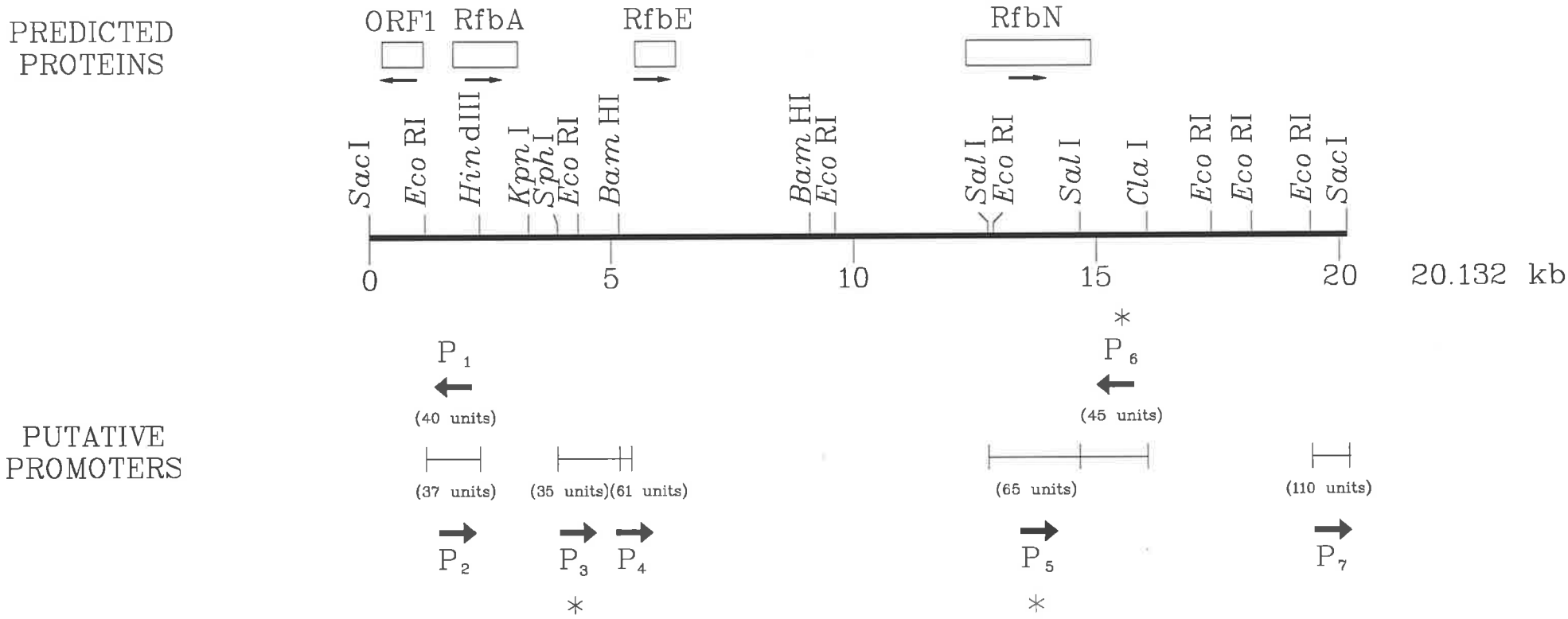
Plasmid no.	TZ colour	GalK units	Plasmid no.	TZ colour	GalK units	Plasmid no.	TZ colour	GalK units
pKC86	red	10.5 ± 11.1	pRMB159	white	13.4 ± 8.7	pRMB151	white	41.4 ± 20.4
pKC87	red	15.3 ± 6.5	pRMB179	white	36.2 ± 18.2	pRMB178	white	52.7 ± 9.1
pKC86 (*SalI*)	red	9.3 ± 3.9	pRMB180	red	12.3 ± 3.5	pRMB177	white	46.2 ± 12.4
pKC87 (*SalI*)	red	3.8 ± 2.9	pRMB181	white	35.3 ± 5.2	pRMB175	white	42.8 ± 18.9
Lac promoter	white	626.0 ± 57.4	pRMB186	red	6.5	pRMB165	white	45.3 ± 19.8
pRMB264	white	58.6	pRMB200	white	122.8 ± 19.7	pRMB207	red	4.3
pRMB272	white	31.5	pRMB204	white	87.1	pRMB205	red	4.6
pRMB273	white	50.4	pRMB238	white	61.1	pRMB133	white	123.4 ± 97.9
pRMB278	white	37.2	pRMB209	red	3.9	pRMB171	white	106.2 ± 47.1
pRMB268	white	76.1	pRMB197	red	14.9	pRMB173	red	11.5 ± 5.2
pRMB270	white	63.9	pRMB288	red	17.3	pRMB162	white	117.6 ± 49.4
pRMB276	white	40.7	pRMB210	red	24.7	pRMB163	white	111.7 ± 72.5
pRMB134	white	51.2 ± 24.3	pRMB212	red	9.4	pRMB168	white	110.3 ± 44.4
pRMB228	red	6.2	pRMB190	white	70.1	pRMB160	white	110.8 ± 23.5
pRMB231	red	2.7	pRMB298	white	44.4	pRMB215	red	21.3
pRMB203	white	23.2	pRMB193	red	13.7	pRMB152	red	1.9 ± 1.6
pRMB226	red	11.9	pRMB213	red	25.3	pRMB126	red	1.9 ± 1.5
pRMB139	red	0.8 ± 0.5	pRMB290	red	22.3	pRMB221	red	7.6
pRMB296	red	1.6	pRMB136	white	65.2 ± 1.9	pRMB219	red	24.6
pRMB261	red	1.2	pRMB300	red	7.4	pRMB258	white	96.4
						pRMB128	red	1.1±0.2
						pRMB255	red	5.6
						pRMB218	red	27.1
						pRMB224	red	9.4

Footnote: TZ colour - appearance of colonies on TZGal agar plates; LB agar containing 1% galactose, Ap (25 μg/μl) and 2,3,5 triphenyltetrazolium (50 mg/ml). Red denotes no promoter activity, white denotes promoter activity.

GalK units - galactokinase units expressed as nmol. of galactose phosphorylated per minute per ml of cells at OD_{650} = 1. Standard deviation has been calculated when constructs were assayed twice or more.

Figure 4.8 Summary of regions of promoter activity.

The regions of the 20 kb *rfb* region that have been found to have promoter activity have been summarized here, with their corresponding GalK activity levels. (*) denotes region of promoter activity which are probably due to cloning artifacts. The relevant proteins described in the text are included in this diagram. Only the relevant restriction sites are shown


PREDICTED PROTEINS
ORF1
RfbA
RfbE
RfbN
Sac I
Eco RI
Hin dIII
Kpn I
Sph I
Eco RI
Bam HI
Bam HI
Eco RI
Sal I
Eco RI
Sal I
Cla I
Eco RI
Eco RI
Eco RI
Sac I
0
5
10
15
20
20.132 kb
PUTATIVE PROMOTERS
P 1
(40 units)
P 6
(45 units)
(37 units)
(35 units)
(61 units)
(65 units)
(110 units)
P 2
P 3
P 4
P 5
P 7

This reaction is readily monitored by a simple and sensitive assay that utilizes [^{14}C]-galactose as substrate (described in 2.12).

This assay, was performed on all of the previously described subclones and the appropriate controls, shown in Table 4.1. The *lac* promoter was obtained by cloning a 350 bp *Pvu*I-*Sma*I fragment from pK18 into pKC87 (Pridmore, 1987). This construct was used as a positive control, producing 626 GalK units. This is in correlation with results described by McKenney *et al.* (1981), who obtained 520 GalK units.

Rating the strength of the promoter in GalK units identifies P_7 as the strongest (110 units) followed by P_5 (65 units), P_4 (61 units), P_6 (45 units), P_2 (40 units), P_1 (37 units) and P_3 (35 units). Shown in Figure 4.8. Thus, by comparison with the *lac* promoter and other promoters described by workers using the same assay system, the *rfb* promoters appear to be very weak (Jalajakumari *et al.*, 1987; McKenney *et al.*, 1981). The reults are near the limit of sensitivity of the assay with the vectors used.

4.2.4 Sequence and analysis of promoter regions

Localization of areas of RNA polymerase binding followed by analysis of transcription signals using a promoterless *galK* gene has resulted in several areas being identified as possibly containing a promoter region involved in the expression of genes for *V. cholerae* O-antigen biosynthesis.

The entire 20 kb *rfb* region sequence has been determined, and promoter sequences searched for using DNASIS.

P_1 was previously localized between *Eco*RI (1.2 kb) and *Hin*dIII (2.2 kb). A putative promoter can be identified in this region, where the bold letters correspond to identity with the consensus sequence and the number in brackets is the number of base pairs between the -35 and -10 regions:

$_{(1235\ bp)}$ ATTACA (17 bp) $_{(1213\ bp)}$ TATAAT

-35 -10

When one studies this region in the light of the translated DNA sequence (described more fully in the next chapter), this promoter may be involved with the expression of ORF1, as it is within the confines of the restriction sites (*Hin*dIII and *Eco*RI, shown in Figure 4.4) and situated before ORF1 (as shown in Figure 4.8).

The other promoter shown in this region (P_2) could be the promoter necessary for the beginning of the *rfb* operon, as the RfbA protein starts at 1733 bp. Two potential promoter sequences can be identified in this region:

(1612 bp) GTGAGA (17 bp) (1634 bp) TGTAAT

-35 -10

or

(1579 bp) TTGAGA (17 bp) (1602 bp) ACTAAT

-35 -10

The data on P_3 is more difficult to interpret. Analysis of the region between *Sph*I (3.9 kb) and *Bam*HI (5.3 kb) does not reveal a promoter sequence. As this region is internal to the putative RfbD protein, as a consequence of cloning this region, a promoter may have been constructed. Analysis of the sequence preceeding the *Bam*HI site (5.3 kb) has revealed 19 bp upstream, a sequence which may correspond to a -35 region - GTGACA. Thus, if the cloning procedure has produced a sequence which may function as a -10 region, promoter activity would be detected.

Such a phenomena has been seen by Rosenberg *et al.* (1983) when using the pKO system characterized promoter signals created by mutations from DNA sequences that previously had no promoter function. The promoters studied were constructed by a deletion in the bacterial chromosome. The promoter P_{482} was constructed by fusing the sequence of the λN gene to the middle of the *galT* gene. This resulted in construction of a promoter where the -10 region is derived from the λN sequence (TATAAT), and the -35 region from the *galT* sequence (CTGCCA), producing a positive phenotype when plated on indicator agar.

Localization of P_4 between *Bam*HI (5.3 kb) and 5.7 kb, implies it may be involved in the expression of *rfbE*, (5547 bp to 6647 bp). Inspection of this region identified a potential promoter sequence at 5.495 kb.

(5495 bp) TTGAAA (16 bp) (5517 bp) TAAATG

-35 -10

P_5 between *Sal*I (12.7 kb) and *Sal*I (14.6 kb), may be similar to P_3 in being a constructed promoter, as this places it internal to the *rfbN* gene (see Fig. 4.8). A -35 region can be identified 19 bases upstream from the *Sal*I site TTGATA.

Another *Sal*I cloning (14.6 kb) resulting in P_6, may also produce a constructed promoter, although in this case the -35 region is not as strong (TGGATA). However, this may still be sufficient to produce a promoter sequence. This sequence would also direct transcription in the opposite strand to the rest of the *rfb* operon, and from computer analyses it appears that there are no significant open reading frames in this direction.

The last and strongest promoter P_7, as shown in the previous chapter, is not involved in *rfb* transcription. Placement of this promoter betweeen *Eco*RI (19.4 kb) and *Sac*I (20.1 kb) puts it outside the minimal coding region (Fig. 3.9), past the known Rfb protein coding domains, and predicted *rfb* region (Ward *et al.*, 1987). Analysis of the sequence preceeding this promoter has revealed a putative terminator sequence from 19975 bp to 20021 bp, with a ΔG of -12.1 kcal/mol (Figure 4.9).

(20046 bp) CTGCAA (18 bp) (20071 bp) TTAAAT

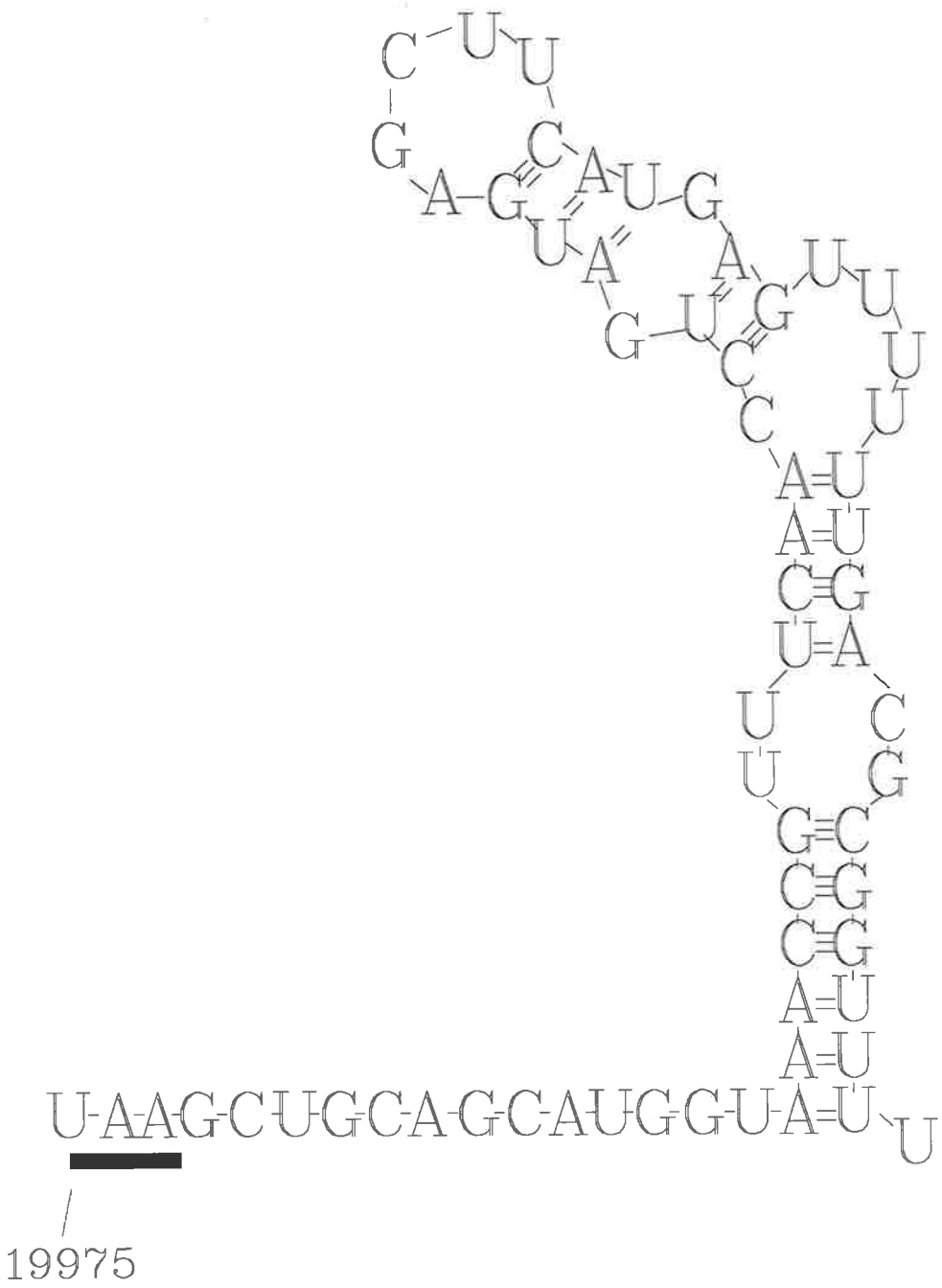
-35 -10

As previously mentioned, promoter activity in this region may be due to construction of a promoter. A sequence which may act as a -35 region is positioned at 20124 bp (TTGAGC).

The whole 20 kb *rfb* region and in particular the DNA sequences corresponding with regions containing promoter activity, were also analysed by DNASIS for the presence of the -24, -12 promoter consensus sequence associated with the alternative

Figure 4.9 Transcriptional terminator identified in the *rfb* region.

A transcriptional terminator can be identified at position 19975 bp. It has a potential free energy (ΔG) of -12.1 kCal/mol (Tinoco *et al.*, 1973). The UAA termination codon of *rfbT* is underlined.



△G= −12.1 kCal/mol

sigma factor, σ^{54}. This analysis did not reveal any promoter sequences containing the invariant GC, GG bases (see 4.1).

4.2.5 RNA analyses

To analyze transcription of the *rfb* region more fully, RNA was isolated (see 2.13.1) and characterized. Attempts were made to identify the transcribed strand by performing RNA dot blot hybridization, using either oligonucleotides from either strand of the DNA, or RNA probes from either DNA strand using subclones in SP6/T7 vectors (described in 2.13.3). Results obtained using this methodology were ambiguous, due to a high background resulting in a low signal to noise ratio. From these studies it appeared that low levels of O-antigen specific mRNA were being produced. To prove this hypothesis, RNA dot blots (using 50 µg of RNA, see 2.13.2) of the *rfb* clone in *E. coli*, *V. cholerae* of both serotypes, and appropriate controls were subjected to hybridization with an oligonucleotide specific for *rfb* mRNA, or with an oligonucleotide specific for the *V. cholerae* haemolysin structural gene (*hlyA*) (S.G. Williams, personal commumication). These hybridizations (shown in Figure 4.10) prove that by comparison with *hlyA* specific mRNA, *rfb* mRNA is produced in considerably lower amounts.

Primer extension analysis to confirm the suggested location of promoter regions was attempted, as was Northern transfer analysis, but were again unsuccessful due to low levels of *rfb* specific mRNA.

4.3 Summary and Conclusions

The RNA polymerase binding data together with analysis of subclones in a *galK* promoter detection system, have provided evidence for the location of putative promoter regions within *rfb*. DNA sequencing of the regions has enabled the potential location of the promoters to be predicted. Attempts were made to confirm the location of these

Figure 4.10 RNA analyses of the *rfb* region of *V. cholerae*.

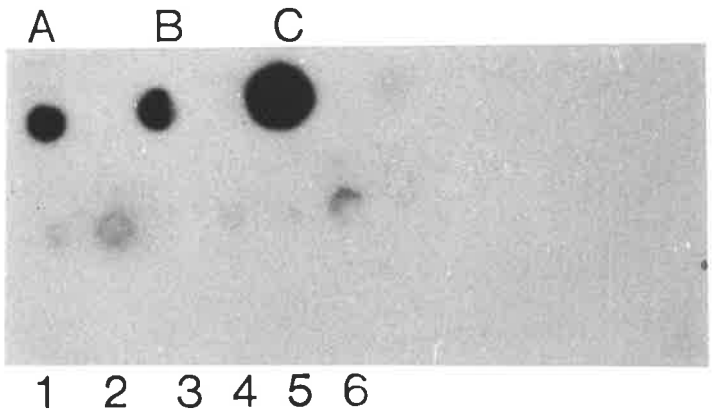
RNA dot blot hybridizations were performed using 50 µg of denatured RNA. DNA controls were included to assess the efficiency of radioactive labelling of the specific probes.

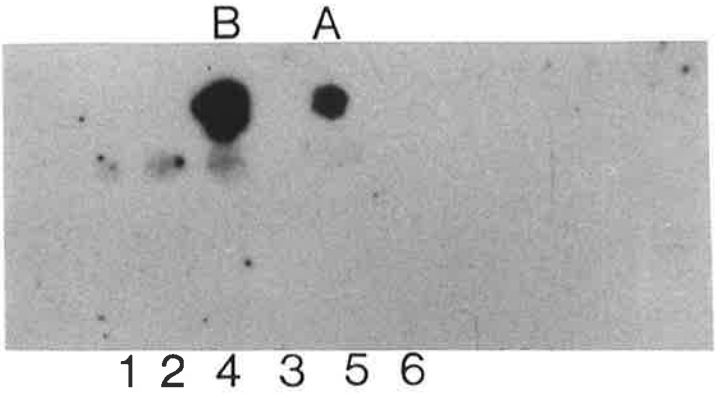
Panel 1: RNA dot blot hybridization performed using as a probe the ^{32}P labelled RNA transcript derived from a SP6 clone containing a 315 bp *Kpn*I-*Pst*I fragment from within the *rfbB* gene.

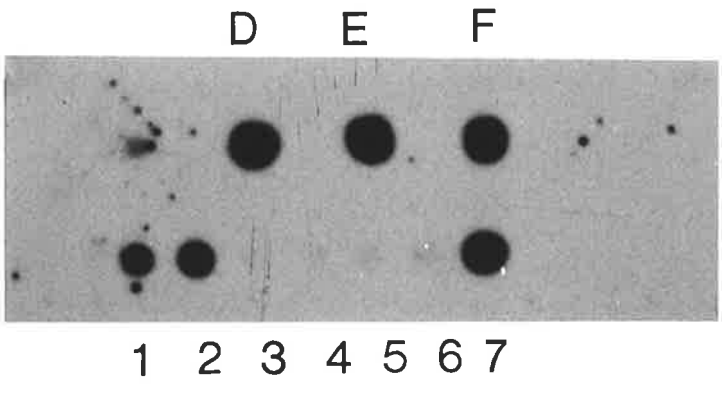
Panel 2: RNA dot blot hybridization performed using a [^{32}]P labelelled oligonucleotide Og 295, which binds from nt 1781 to nt 1803. This is complementary to the 5' end of the *rfbA* gene.

Panel 3: RNA dot blot hybridization performed using a [^{32}P] labelled oligonucleotide (265), which is specific for the *hlyA* gene of *V. cholerae* (S.G. Williams, personal communication).

Symbols: A - 0.25 µg of denatured pRMB1 DNA; B - 0.5 µg of denatured pRMB1 DNA; C - 5 µg of denatured pRMB1 DNA; D - 0.25 µg of denatured pRIC1 DNA (a pSUP201-1 derivative containing the *hlyA* gene cloned into the *Pst*I site); E - 0.5 µg of denatured pRIC1 DNA; F - 5 µg of denatured pRIC1 DNA; 1 - *V. cholerae* Ogawa (O17) RNA; 2 - *V. cholerae* Inaba (569B) RNA; 3 - RNA from *E. coli* harbouring pRMB1; 4 - RNA from *E. coli* harbouring pOmpV500; 5 - RNA from *E. coli* harbouring pEVX374, (corresponds to the same clone as pRMB1, except it contains the *rfb* region from *V. cholerae* of the Inaba serotype); 6 - *E. coli* K-12 RNA; 7 - RNA from *E. coli* harbouring pRIC1.


A
B
C
1 2 3 4 5 6


B
A
1 2 4 3 5 6


D
E
F
1 2 3 4 5 6 7

promoter regions using mRNA analyses, however, due to the low level of *rfb* specific mRNA, these experiments were inconclusive.

These studies have resulted in the identification of 7 potential promoter sequences (Figure 4.8), which can be summarized as follows:

P_1 (*Eco*RI - *Hin*dIII) $_{(1213\ bp)}$ TAATAT (17 bp) ACATTA producing 40 GalK units. This promoter is probably involved in the transcription of ORF1.

P_2 (*Eco*RI - *Hin*dIII) $_{(1612\ bp)}$ GTGAGA (17 bp) TGTAAT or $_{(1579)}$ TTGAGA (17 bp) ACTAAT producing 37 GalK units. This promoter sequence is postulated to be the first one involved in expression of the *rfb* region, i.e. it is required for transcription of *rfb*A.

P_3 (*Sph*I - *Bam*HI) has GTGACA 19 bp from the *Bam*HI site. It yields 35 GalK units. The cloning of P_3 may have resulted in construction of a promoter sequence. Sequence analysis positions this promoter within the open reading frame for *rfbD*.

P_4 (*Bam*HI - 5.7 kb) $_{(5495)}$ TTGAAA (16 bp) TAAATG producing 61 GalK units, may be involved in the expession of *rfbE* .

P_5 (*Sal*I - *Sal*I) producing 35 GalK units. A -35 region TTGATA was identifed 19 bp from the *Sal*I site. This promoter may also have been constructed by the cloning procedure, especially since it lies in the middle of the open reading frame for a 91.2 kDa predicted protein (encoded by *rfbN*).

P_6 (*Sal*I - *Cla*I) producing 45 GalK units. A -35 region TGGATA can be seen 20 bp from the *Sal*I site. Similar to P_3 and P_5, P_6 may be a cloning artifact, especially as it is in the opposite orientation to transcription of the rest of the *rfb* region and no significant open reading frames have been identified.

P_7 (*Eco*RI - *Sac*I) $_{(20046)}$ CTGCAA (18 bp) TTAAAT (110 GalK units). This promoter is the strongest promoter, and is not involved in the production of VcOAg in *E. coli* cells harbouring *rfb* clones (Ward *et al.*, 1987).

CHAPTER 5

SEQUENCE ANALYSIS OF A PORTION OF THE *RFB* REGION OF *VIBRIO CHOLERAE* OGAWA

5.1 Introduction

The analysis of prokaryotic nucleotide sequences has until recently predominantly involved that of *E. coli* and its bacteriophages. In contrast, the nucleotide sequences of very few *V. cholerae* genes have been determined. These include the genes that encode the following proteins: cholera enterotoxin (Lockman and Kaper, 1983; Mekalanos *et al.*, 1983); OmpV, and OmpW two outer membrane proteins (Jalajakumari and Manning, 1990; Pohlner *et al.*, 1986); the haemolysin locus from both El Tor and Classical *V. cholerae* (Alm *et al.*, 1988; Alm and Manning, 1990; Yamamoto *et al.*, 1990); an extracellular DNAse (Focareta and Manning, 1987); ToxR and ToxS proteins involved in co-ordinate regulation of virulence determinants (Di Rita and Mekalanos, 1991; Miller and Mekalanos, 1984; Miller *et al.*, 1987; Miller *et al.*, 1989); TcpA, the major pilin for the TcpA locus (Faast *et al.*, 1989; Shaw and Taylor, 1990), and TcpH, B, I, and X, proteins involved in the assembly of the pilin; the region involved in packaging CP-T1 DNA (Guidolin, A. and Manning P.A., 1988) and a group of genes involved in the production of the mannose-fucose resistant haemagglutinin (MFRHA) (Barker, A., Clark, C.A., Williams, S.G., Franzon, V.L., and Manning, P.A. manuscript in preparation; Van Dongen *et al.*, 1987). In the light of the data obtained from nucleotide sequence determination, a better understanding of how *V. cholerae* can regulate and express some

of its gene products has been reached (Mekalanos *et al.*, 1983; Miller *et al.*, 1987; Taylor *et al.*,1987a). This chapter describes the salient features of the DNA sequences thought to have promoter activity within regions of the *Vibrio cholerae* 20 kb *rfb* locus.

5.2 Results

5.2.1 DNA sequencing strategy

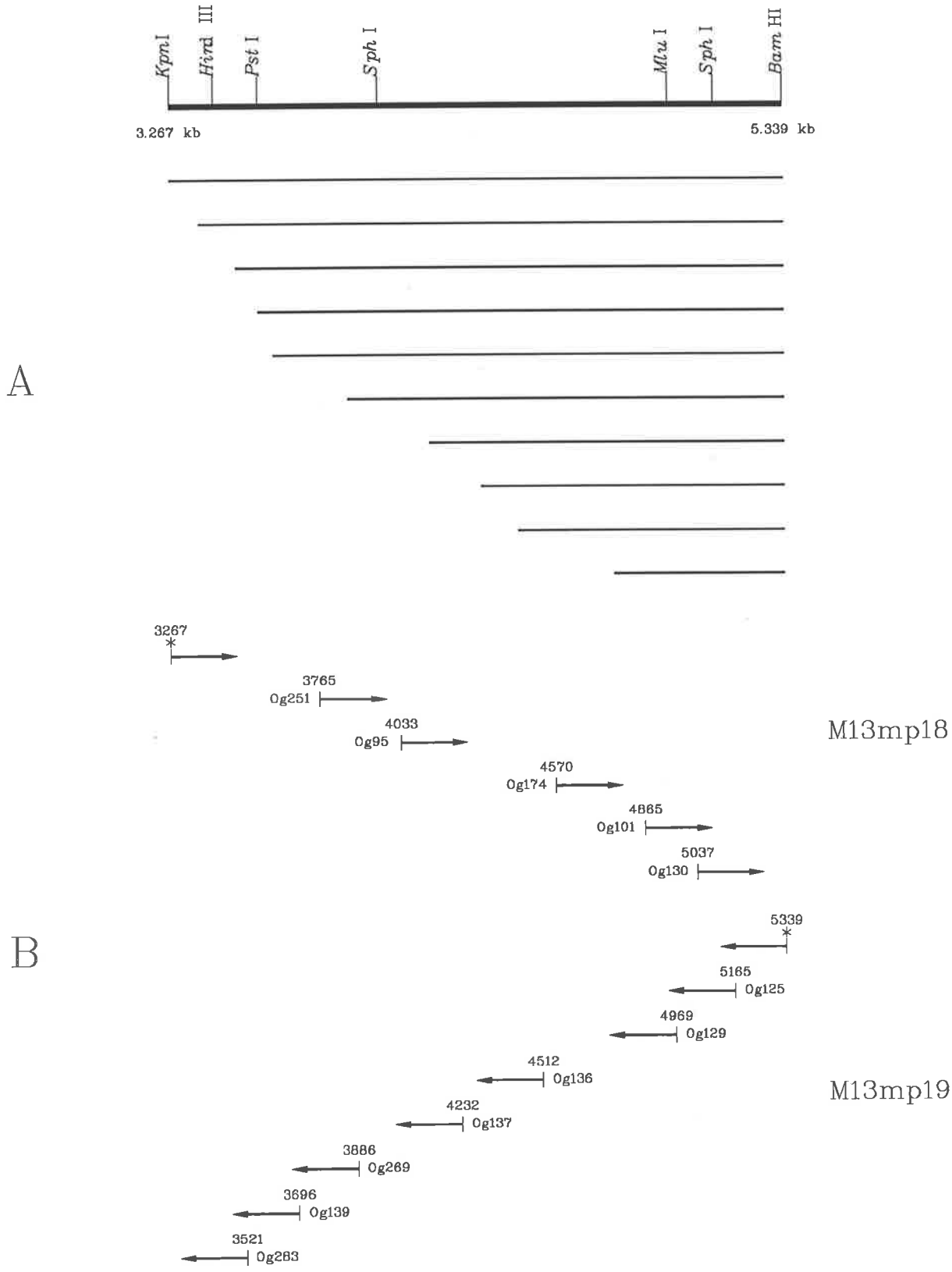
As descibed in the previous chapter various regions of the 20 kb *rfb* region were studied for promoter activity. To enable a more detailed assessment of the promoter activity of these regions the DNA sequence was required. As it was not possible in the time frame allowed to sequence the whole 20 kb myself, I was allocated the following four regions to sequence to identify promoter sequences: Region1: *Kpn*I (3267 bp) to *Bam*HI (5339 bp); Region2: *Bam*HI (5339 bp) to *Xba*I (6465 bp); Region3: *Sal*I (14657 bp) to *Hin*dIII (16150 bp); and Region4: *Cla*I (16029 bp) to *Cla*I (19010 bp), although in this last region only 16029 bp to 17101 was sequenced.

Region1 was sequenced by firstly cloning a fragment containing *Kpn*I (3267 bp) to *Bam*HI (5339 bp) into the multiple cloning site polylinker of both M13mp18 and M13mp19 (Vieira and Messing, 1982). The resulting clone in M13mp19 was used to make seqential overlapping deletions, using the Cyclone kit, (Section 2.18.4), employing the exonuclease activity of T4 DNA polymerase and varying the incubation times with the enzyme prior to heat inactivation. Sequencing was performed using the chain termination method (Sanger *et al.*, 1977, 1980) using both Klenow enyzme (with ^{32}P) and Sequenase^{TS} (^{35}S) (described in 2.18.8 in more detail). Synthetic olignucleotides were used to confirm the sequence data obtained both in this direction, and it was rechecked using the complementary strand cloned into M13mp18. A map of the sequencing strategy is shown in Figure 5.1.

Figure 5.1 Strategy used for dideoxy sequencing of Region1: *Kpn*I to *Bam*HI.

A) Overlapping deletions were made in the 2.1 kb *Kpn*I - *Bam*HI fragment cloned into M13mp19. The fragment was digested at one end with T4 DNA polymerase, and universal primer used to generate the nucleotide sequence. The -40 sequencing primer (5' GTTTTCCCAGTCACGAC 3') was also employed to accurately determine the extreme ends of sequence at the junction of the polylinker site in both clones (M13mp18 and M13mp19), denoted by (*).

B) Synthetic oligonucleotides, designated by Og numbers, were constructed to extend the sequence in the directions shown by the arrows. The location of these 18-mer oligonucleotides are denoted by numbering the 5' end of the oligo.


Kpn I
Hind III
Pst I
Sph I
Mlu I
Sph I
Bam HI
3.267 kb
5.339 kb
A
3267
3765
0g251
4033
0g95
4570
0g174
4865
0g101
5037
0g130
M13mp18
B
5339
5165
0g125
4969
0g129
4512
0g136
4232
0g137
3886
0g269
3696
0g139
3521
0g283
M13mp19

Regions 2, 3 and 4 were sequenced by cloning the appropriate fragment (*Bam*HI-*Xba*I), (*Sal*I-*Hin*dIII), and (*Cla*I-*Cla*I), respectively, into both M13mp18 and M13mp19. Sequencing was carried out using the chain termination method of Sanger *et al.* (1977, 1980), using Sequenase^{TS} enzyme (described in 2.18.8). Universal primer was employed to sequence from both directions and then synthetic oligonucleotides were made to extend the sequence. A map of the sequencing strategy used for these three areas, showing the oligonucleotides used can be seen in Figures 5.2, 5.3, and 5.4.

5.2.2 Nucleotide sequence determination and analysis of the *V. cholerae rfb* region

The whole 20 kb *Sac*I region involved in the biosynthesis of the *V. cholerae* Ogawa O-antigen has been sequenced in this laboratory. Figures 5.6 to 5.14 show the nucleotide sequence of the *V. cholerae rfb* region from base 2977 to 6647, and from 14929 bp to 17007 bp.

Figure 5.5 shows the proteins predicted from the DNA sequence using the DNA program DNASIS. These open reading frames (ORFs) have been given gene designations based on their location and not on the predicted molecular weight of the corresponding proteins. *rfbA* to *rfbT* are transcribed in the same direction, rightwards, as are ORF2 and ORF3. ORF1 and *ompX* are both transcribed in the opposite direction to the previously mentioned genes.

5.2.2.1 Nucleotide sequence of four regions of the *V. cholerae rfb* operon

The regions 1, 2, 3, and 4 previously specified, were sequenced and the ORFs in these regions predicted and analyzed using programs from the DNA and protein analysis packages DNASIS, PROSIS and GCG, shown in Figures 5.6 to 5.14 and summarized in Table 5.1, together with possible Shine-Dalgarno sequences.

Figure 5.2 Strategy used for dideoxy sequencing of Region2: *Bam*HI to *Xba*I.

The 1.1 kb *Bam*HI - *Xba*I fragment was cloned into both M13mp18 and M13mp19. This fragment was sequenced with universal primer and the -40 sequencing primer (5' GTTTTCCCAGTCACGAC 3') was employed to accurately determine the extreme ends at the junction of the polylinker, shown by (*). Synthetic oligonucleotides, designated by Og numbers, were used to extend the sequence in the directions shown by the arrows. The location of these 18-mer oligonucleotides are denoted by numbering the 5' end of the oligo.

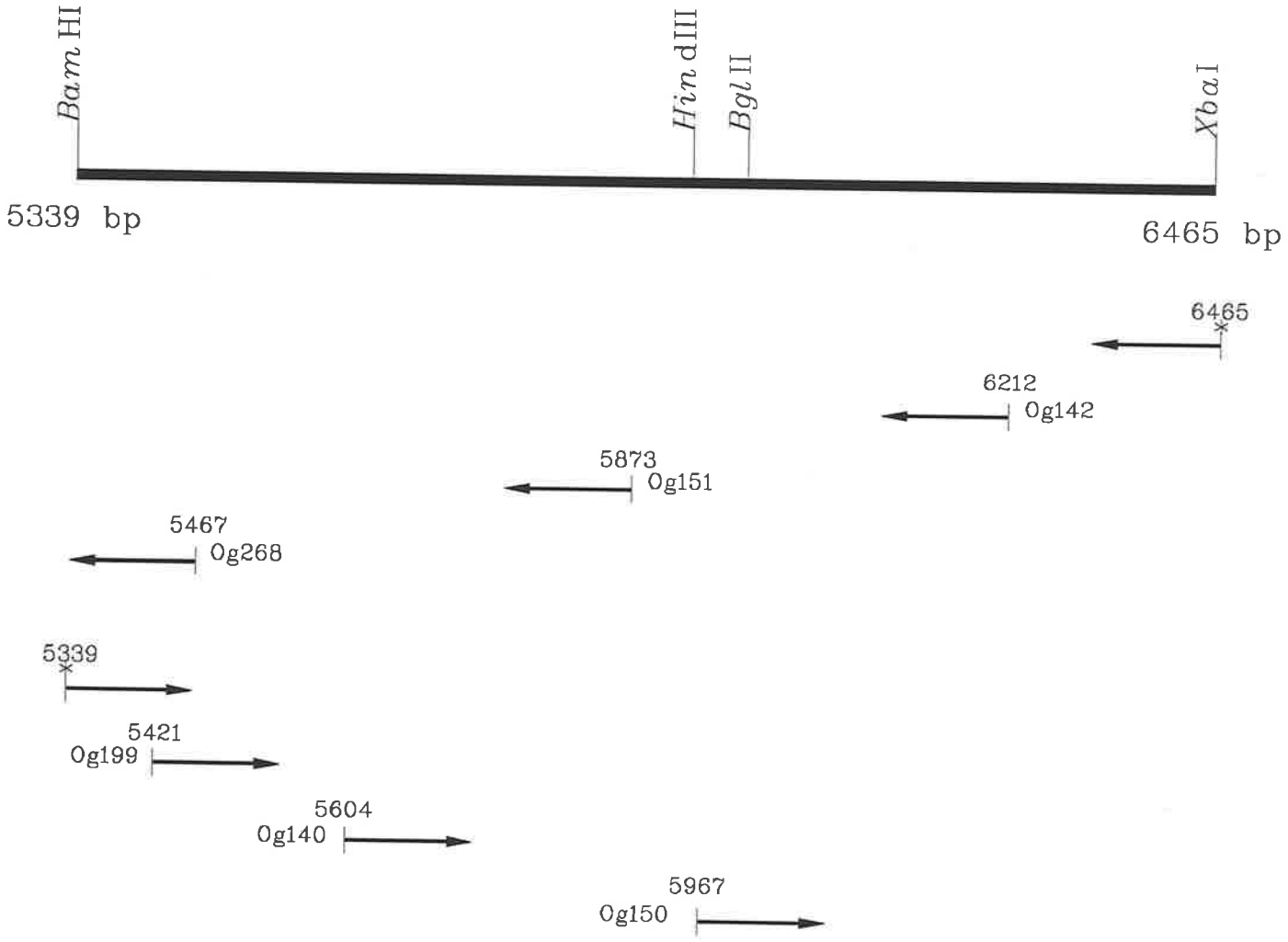

Bam HI
Hin dIII
Bgl II
Xba I
5339 bp
6465 bp
6465
6212
0g142
5873
0g151
5467
0g268
5339
5421
0g199
5604
0g140
5967
0g150
M13mp18
M13mp19

Figure 5.3 Strategy used for dideoxy sequencing of Region3: *Sal*I to *Hin*dIII.

The 1.5 kb *Sal*I - *Hin*dIII fragment was cloned into both M13mp18 and M13mp19. This fragment was sequenced with universal primer and the -40 sequencing primer (5' GTTTTCCCAGTCACGAC 3') was employed to accurately determine the extreme ends at the junction of the polylinker, shown by (*). Synthetic oligonucleotides, designated by the Og numbers, were constructed to extend the sequence in the directions shown by the arrows. The location of these 18-mer oligonucleotides are denoted by numbering the 5' end of the oligo.

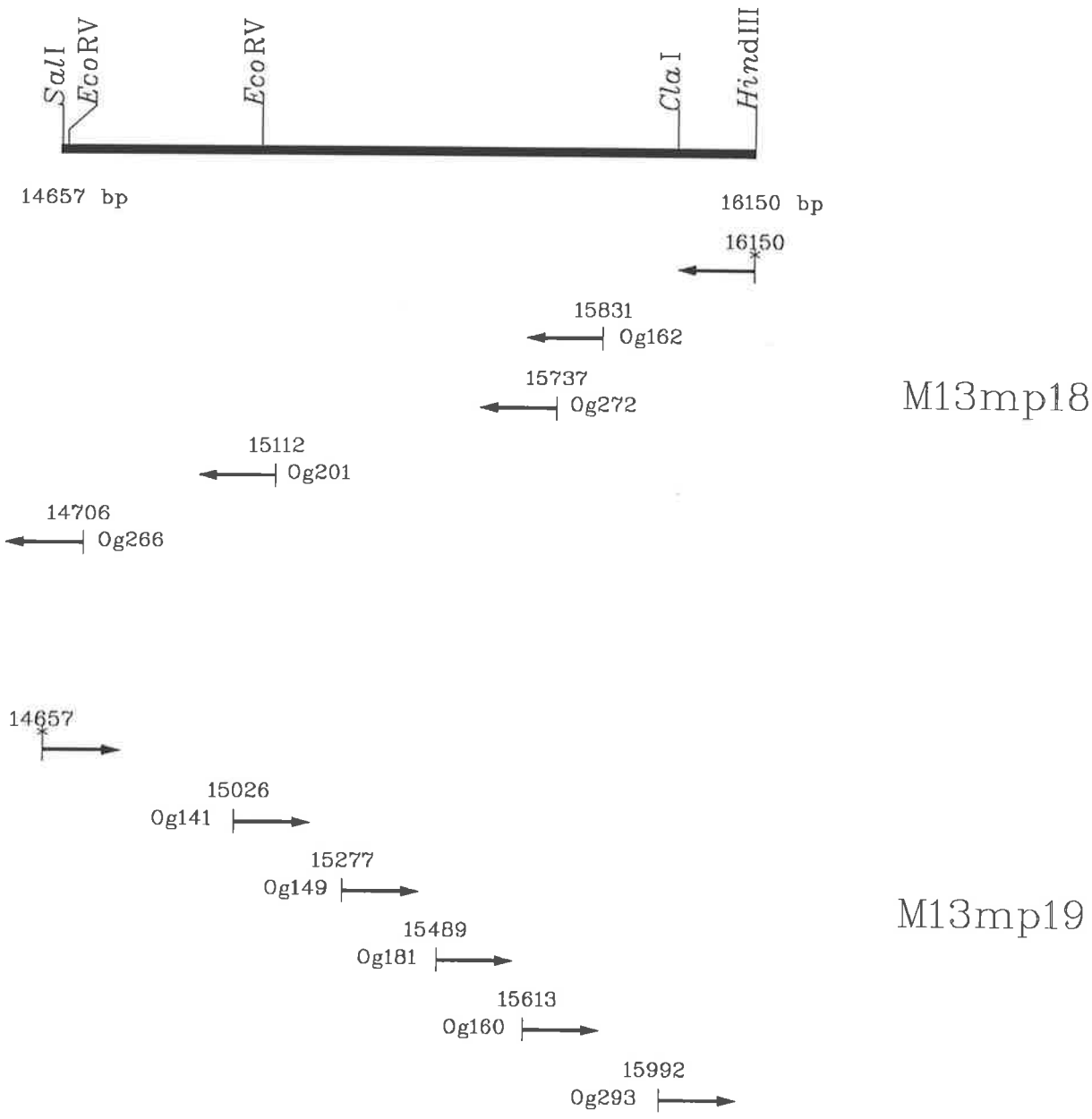

SalI
EcoRV
EcoRV
ClaI
HindIII
14657 bp
16150 bp
16150
15831
0g162
15737
0g272
M13mp18
15112
0g201
14706
0g266
14657
15026
0g141
15277
0g149
M13mp19
15489
0g181
15613
0g160
15992
0g293

Figure 5.4 Strategy used for dideoxy sequencing of Region4: *Cla*I to *Cla*I.

The 3 kb *Cla*I - *Cla*I fragment was cloned into both M13mp18 and M13mp19. This fragment was sequenced with universal primer and the -40 sequencing primer (5' GTTTTCCCAGTCACGAC 3') was employed to accurately determine the extreme ends at the junction of the polylinker, shown by (*). Synthetic oligonucleotides, designated by Og numbers, were used to extend the sequence in the directions shown by the arrows. The location of these 18-mer oligonucleotides are denoted by numbering the 5' end of the oligo. This diagram only shows the extent of the clone personally sequenced.

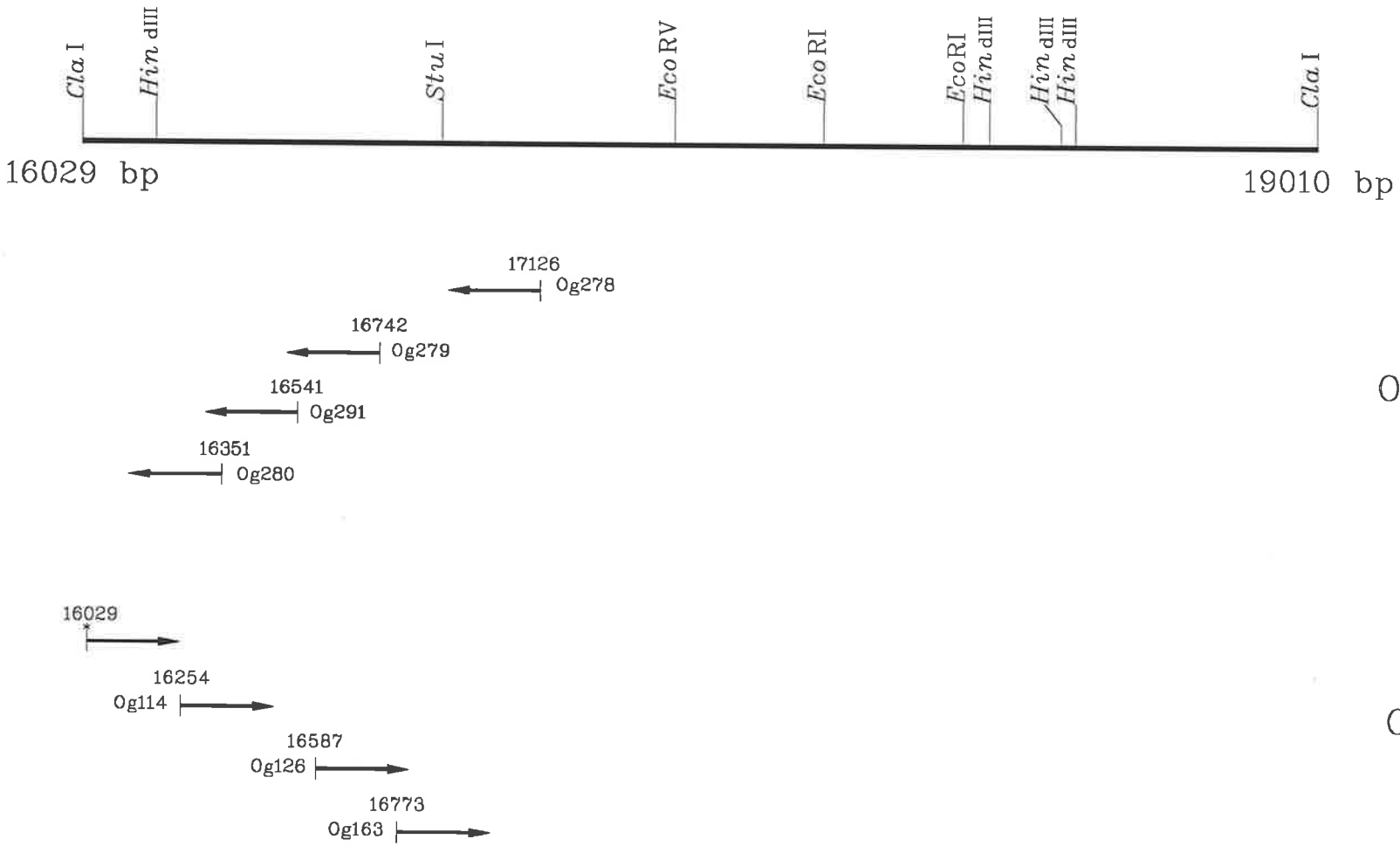

Cla I
Hin dIII
Stu I
Eco RV
Eco RI
EcoRI
Hin dIII
Hin dIII
Hin dIII
Cla I
16029 bp
19010 bp
17126
0g278
16742
0g279
16541
0g291
16351
0g280
Orientation 1
16029
16254
0g114
16587
0g126
16773
0g163
Orientation 2

Figure 5.5 Map of open reading frames of the *rfb* region.

The nucleotide sequence of the 20 kb *rfb* region was determined and analyzed for open reading frame locations using the programme DNASIS. The *rfb* region predicted proteins are labelled A through to T. ORF1, 2, and 3, and *ompX*, refer to open reading frames not involved in the biosynthesis of the O-antigen. The direction of transcription of the genes is shown under the genes.

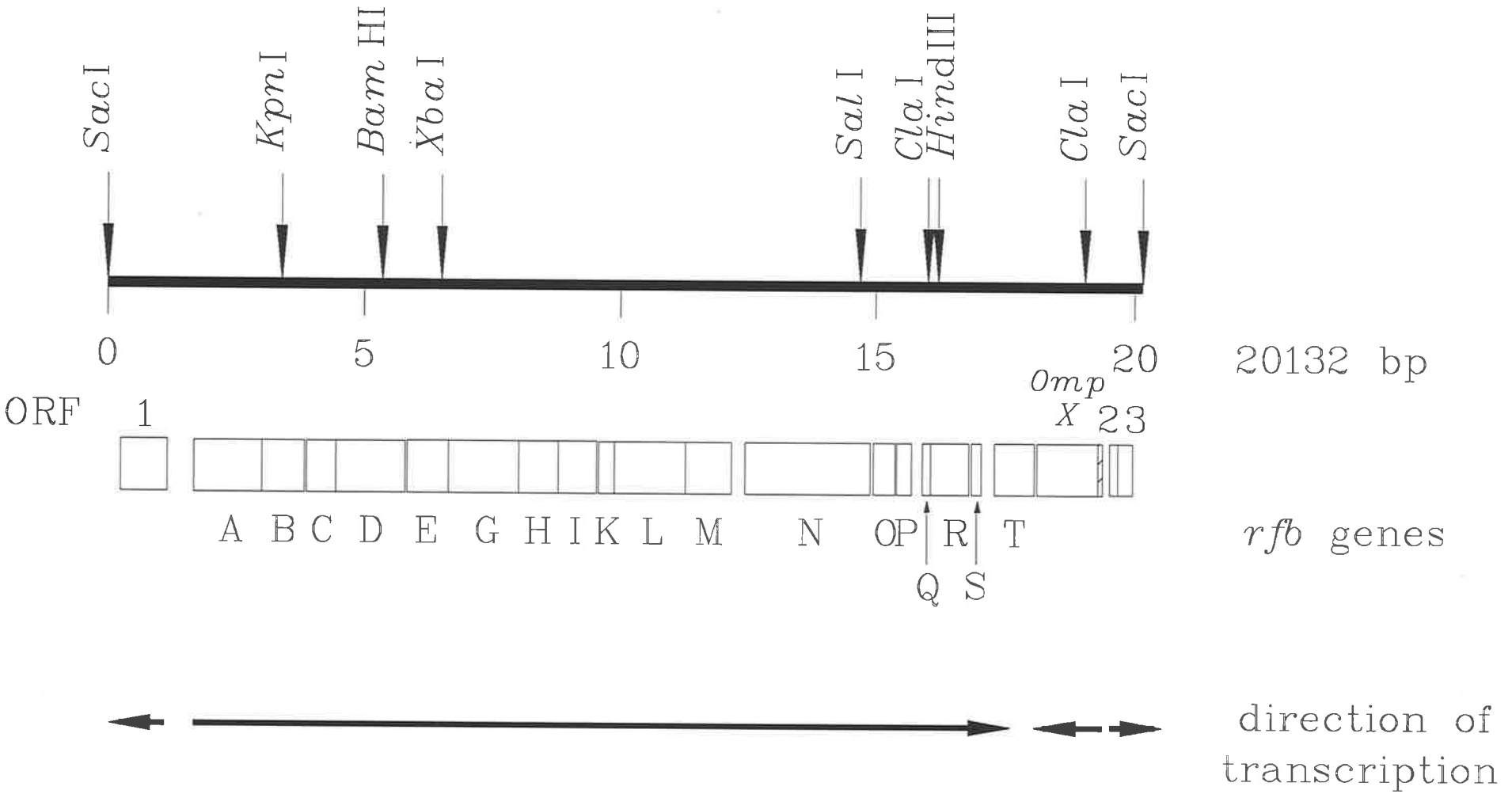

SacI
KpnI
BamHI
XbaI
SalI
ClaI
HindIII
ClaI
SacI
0
5
10
15
20
20132 bp
ORF 1
Omp X
23
A B C D E G H I K L M N O P R T
Q S
rfb genes
direction of transcription

Figure 5.6 Nucleotide sequence of the *V. cholerae rfb* region 2977 bp - 3840 bp.

The nucleotide sequence of the 863 base-pair region containing the *rfbB* gene (nt 3022 to 3828). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbB (translation region 3022 bp - 3828 bp)

```
2977  ATA CCT AAA CGA AGA TGA TGT AGT ACG TTT CGA GGA TAA ATA AGA GTG   3024
   1                                                          Met      1

3025  AAA GAG TTA ACT TGT TTT AAA GCC TAC GAT ATT CGC GGA CAA TTA GGG   3072
   2  Lys Glu Leu Thr Cys Phe Lys Ala Tyr Asp Ile Arg Gly Gln Leu Gly    17

3073  AGT GAG CTT GAT AAT GAA ATT GCC TAC AGA ATT GGT CGC TCG TAT GGG   3120
  18  Ser Glu Leu Asp Asn Glu Ile Ala Tyr Arg Ile Gly Arg Ser Tyr Gly    33

3121  CAG TTT TTA AAA AGC GAA AAT GAC GCA GAT AAA ACA GTT GTG GTA GGT   3168
  34  Gln Phe Leu Lys Ser Glu Asn Asp Ala Asp Lys Thr Val Val Val Gly    49

3169  GGT GAT GTT CGT TTA ACC AGC GAA GCG CTT AAA CAA GCA CTA GCT AAT   3216
  50  Gly Asp Val Arg Leu Thr Ser Glu Ala Leu Lys Gln Ala Leu Ala Asn    65

3217  GGC TTA ATG GAT GCA GGT ATT AAC GTT ATT GAT ATT GGC GTT ACT GGT   3264
  66  Gly Leu Met Asp Ala Gly Ile Asn Val Ile Asp Ile Gly Val Thr Gly    81

3265  ACC GAA GAG ATC TAT TTT GCA ACC TTC TAC CTT GGT GTT GAT GGT GGT   3312
  82  Thr Glu Glu Ile Tyr Phe Ala Thr Phe Tyr Leu Gly Val Asp Gly Gly    97

3313  ATT GAA GTT ACC GCA AGT CAT AAC CCA ATG GAT TAC AAT GGT ATG AAG   3360
  98  Ile Glu Val Thr Ala Ser His Asn Pro Met Asp Tyr Asn Gly Met Lys   113

3361  CTG GTT CGT GAA GGT TCA AAA CCC ATT AGT GGT GAT ACT GGT TTA CGA   3408
 114  Leu Val Arg Glu Gly Ser Lys Pro Ile Ser Gly Asp Thr Gly Leu Arg   129

3409  GAA ATT CAA GCT TTG GCT GAA AAG AAT GAA TTT ATG GAC GTT GAA GTT   3456
 130  Glu Ile Gln Ala Leu Ala Glu Lys Asn Glu Phe Met Asp Val Glu Val   145

3457  AAA GGT AAC TAC AAA AAG GTT TCT TTA CTT CCA GAA TAT GTA GAT CAC   3504
 146  Lys Gly Asn Tyr Lys Lys Val Ser Leu Leu Pro Glu Tyr Val Asp His   161

3505  TTA ATA TCT TAT ATT ACA CCC GCG AAA ATT AAA CCA ATG AAG TTG GTC   3552
 162  Leu Ile Ser Tyr Ile Thr Pro Ala Lys Ile Lys Pro Met Lys Leu Val   177

3553  ATC AAT TCA GGT AAC GGA GCT GCA GGC CAC GTG ATT GAT GAG TTG GAA   3600
 178  Ile Asn Ser Gly Asn Gly Ala Ala Gly His Val Ile Asp Glu Leu Glu   193

3601  AAA CGA TTT ATT GAA TTG AGT ATC CCG CTA GAA ATT ATC AAA GTG CAT   3648
 194  Lys Arg Phe Ile Glu Leu Ser Ile Pro Leu Glu Ile Ile Lys Val His   209

3649  CAT GAA GAA GAT GGT AAC TTC CCG AAT GGT ATT CCT AAC CCA TTG TTA   3696
 210  His Glu Glu Asp Gly Asn Phe Pro Asn Gly Ile Pro Asn Pro Leu Leu   225

3697  CCT GAA TGC CGT GCT GAT ACC GCA AAC GCA GTA AAG GAG CAC AAG GCT   3744
 226  Pro Glu Cys Arg Ala Asp Thr Ala Asn Ala Val Lys Glu His Lys Ala   241

3745  GAT ATG GGC ATT GCT TTG ATG GTG ACT TTG ATC GCT GCT TTT TGT TTG   3792
 242  Asp Met Gly Ile Ala Leu Met Val Thr Leu Ile Ala Ala Phe Cys Leu   257

3793  ATG AAA ATG GCG ATT TTA TTG AAG GTT ACT ACA TCG TAG GTT TAT TGG   3840
 258  Met Lys Met Ala Ile Leu Leu Lys Val Thr Thr Ser                   269
```

Figure 5.7 Nucleotide sequence of the *V. cholerae rfb* region 3792 bp - 4415 bp.

The nucleotide sequence of the 623 base-pair region containing the *rfbC* gene (nt 3837 to 4409). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbC (translation region 3837 bp - 4409 bp)

```
3792  GAT GAA AAT GGC GAT TTT ATT GAA GGT TAC TAC ATC GTA GGT TTA TTG   3839
   1                                                          Met       1

3840  GCA GAA GCG TTT TTA CAG AAA GAG CAA GGT GCC AAA ATT ATT CAT GAT   3887
   2  Ala Glu Ala Phe Leu Gln Lys Glu Gln Gly Ala Lys Ile Ile His Asp    17

3888  CCA CGC TTA AGC TGG AAC ACC ATT GAT GTG GTG ACT AAG TCG GGT GGC   3935
  18  Pro Arg Leu Ser Trp Asn Thr Ile Asp Val Val Thr Lys Ser Gly Gly    33

3936  GTG CCT GTA ATG TCT AAA ACA GGG CAT GCA TTT ATA AAA GAG CGT ATG   3983
  34  Val Pro Val Met Ser Lys Thr Gly His Ala Phe Ile Lys Glu Arg Met    49

3984  CGC AAA GAA GAT GCT ATC TAC GGT GGT GAA ATG AGT GCC CAT CAC TAC   4031
  50  Arg Lys Glu Asp Ala Ile Tyr Gly Gly Glu Met Ser Ala His His Tyr    65

4032  TTC CGT GAT TTC GGT TAC TGT GAT TCG GGG ATG ATT CCG TGG CTT CTG   4079
  66  Phe Arg Asp Phe Gly Tyr Cys Asp Ser Gly Met Ile Pro Trp Leu Leu    81

4080  ATT ACA GAG TTA CTT TCC TTA GCC CCG GAT ATT AGC TTA TCT AAG CTG   4127
  82  Ile Thr Glu Leu Leu Ser Leu Ala Pro Asp Ile Ser Leu Ser Lys Leu    97

4128  ATA TCT GCA AAA AGA TTT TTG TTT CCA TGT AGC GGT GAA ATT AAT TTT   4175
  98  Ile Ser Ala Lys Arg Phe Leu Phe Pro Cys Ser Gly Glu Ile Asn Phe   113

4176  AAG GTC AAG CAG GCA AAG CTA ATA ATG GAG CAA GTT TAT TTA CAT TAT   4223
 114  Lys Val Lys Gln Ala Lys Leu Ile Met Glu Gln Val Tyr Leu His Tyr   129

4224  TAC GAG AAT TCT ATA CAT TTC TCT GCC ATT GAT GGT ATC TCT CTT GAA   4271
 130  Tyr Glu Asn Ser Ile His Phe Ser Ala Ile Asp Gly Ile Ser Leu Glu   145

4272  TTT GAA GGA TGG CGT TTT AAT CTT AGA GAT TCA AAT ACT GAG CCC TTA   4319
 146  Phe Glu Gly Trp Arg Phe Asn Leu Arg Asp Ser Asn Thr Glu Pro Leu   161

4320  TTA AGA TTG AAT GTG GAA TCT AAA CAA AAT ATC GCG TTG ATG AAT GAT   4367
 162  Leu Arg Leu Asn Val Glu Ser Lys Gln Asn Ile Ala Leu Met Asn Asp   177

4368  AAA GTT GAA GAG CTT ACT AAA TTA ATT AAG AAA TTG GAT ATT TAA CAA   4415
 178  Lys Val Glu Glu Leu Thr Lys Leu Ile Lys Lys Leu Asp Ile          191
```

Figure 5.8 Nucleotide sequence of the *V. cholerae rfb* region 4367 bp - 5566 bp.

The nucleotide sequence of the 1199 base-pair region containing the *rfbD* gene (nt 4415 to 5533). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold. The arrow head denotes the position of the Tn*1725* insertion, (nt 4881), sequenced from the plasmid pRMB58.

Vibrio cholerae RfbD (translation region 4415 bp - 5533 bp)

```
4367  TAA AGT TGA AGA GCT TAC TAA ATT AAT TAA GAA ATT GGA TAT TTA ACA   4414

4415  ATG AAT AAA AAA GTT GCG TTA ATC ACA GGC ATT ACT GGA CAA GAT GGT   4462
   1  Met Asn Lys Lys Val Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly     16

4463  TCT TAT CTA GCG GAG TTT TTG CTA GAA AAA GGG TAT GAA GTA CAT GGA   4510
  17  Ser Tyr Leu Ala Glu Phe Leu Leu Glu Lys Gly Tyr Glu Val His Gly     32

4511  ATT AAA AGA CGT TCA TCT TTA TTC AAT ACA CAA CGT GTT GAT CAT CTT   4558
  33  Ile Lys Arg Arg Ser Ser Leu Phe Asn Thr Gln Arg Val Asp His Leu     48

4559  TAC AAA GAC CCA CAT GAG GAA GAT GTA AAT TTC AAA CTA CAT TAC GGA   4606
  49  Tyr Lys Asp Pro His Glu Glu Asp Val Asn Phe Lys Leu His Tyr Gly     64

4607  GAC CTT ACG GAT TCC TCT AAC TTG ACT CGT ATA TTA GCG GAA GTA CAG   4654
  65  Asp Leu Thr Asp Ser Ser Asn Leu Thr Arg Ile Leu Ala Glu Val Gln     80

4655  CCA GAT GAA GTC TAT AAC TTG GGC GCG CAG TCC CAT GTG GCG GTT AGT   4702
  81  Pro Asp Glu Val Tyr Asn Leu Gly Ala Gln Ser His Val Ala Val Ser     96

4703  TTC CAG TCA CCG GAA TAT ACA GCG GAC GTA GAT GCT ATT GGA ACT TTA   4750
  97  Phe Gln Ser Pro Glu Tyr Thr Ala Asp Val Asp Ala Ile Gly Thr Leu    112

4751  AGG CTA CTA GAA GCA ATC AGA TTT TTA GGT TTG ACG AAA AAA ACT AAG   4798
 113  Arg Leu Leu Glu Ala Ile Arg Phe Leu Gly Leu Thr Lys Lys Thr Lys    128

4799  TTT TAT CAA GCC TCA ACT TCA GAG CTT TAT GGC TTA GTA CAA GAA ATA   4846
 129  Phe Tyr Gln Ala Ser Thr Ser Glu Leu Tyr Gly Leu Val Gln Glu Ile    144

4847  CCT CAG AAA GAA ACT ACC CCT TTT TAT CCT AGA AGC CCT TAT GCT GTG   4894
 145  Pro Gln Lys Glu Thr Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Ala Val    160

4895  GCA AAG ATG TAT GCG TAT TGG ATA ACA ATT AAT TAT AGA GAA TCG TAC   4942
 161  Ala Lys Met Tyr Ala Tyr Trp Ile Thr Ile Asn Tyr Arg Glu Ser Tyr    176

4943  GGA ATA TAC GCG TGT AAT GGC ATA CTT TTT AAC CAC GAA TCC CCT AGA   4990
 177  Gly Ile Tyr Ala Cys Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg    192

4991  CGA GGA GAA ACT TTT GTA ACT CGC AAA ATA ACT CGG GGG ATG GCA AAC   5038
 193  Arg Gly Glu Thr Phe Val Thr Arg Lys Ile Thr Arg Gly Met Ala Asn    208

5039  ATT GCG CAA GGG TTA GAA AAA TGC CTT TTT ATG GGA AAT TTG GAC GCT   5086
 209  Ile Ala Gln Gly Leu Glu Lys Cys Leu Phe Met Gly Asn Leu Asp Ala    224

5087  CTT CGA GAT TGG GGG CAT GCT AAA GAT TAC GTT AAG ATG CAA TGG ATG   5134
 225  Leu Arg Asp Trp Gly His Ala Lys Asp Tyr Val Lys Met Gln Trp Met    240

5135  ATG TTA CAG CAA GAT GAA CCT AGA GAT TTT GTT ATT GCT ACA GGT GTA   5182
 241  Met Leu Gln Gln Asp Glu Pro Arg Asp Phe Val Ile Ala Thr Gly Val    256

5183  CAG TAT AGC GTA AGA GAG TTT ATT GAT ATG TCA GCA CGA GAA CTA GGC   5230
 257  Gln Tyr Ser Val Arg Glu Phe Ile Asp Met Ser Ala Arg Glu Leu Gly    272
```

```
5231  ATT GAG TTA GAA TTT GTC GGT AAG GGT GTA GAT GAA AAA GCT GTA GTG   5278
 273  Ile Glu Leu Glu Phe Val Gly Lys Gly Val Asp Glu Lys Ala Val Val    288

5279  AAG TCA GTT ATT GGA ACG AAA GCT CCT GCA ATC AAG GTT GGT GAC ATT   5326
 289  Lys Ser Val Ile Gly Thr Lys Ala Pro Ala Ile Lys Val Gly Asp Ile    304

5327  ATT GTA GCT GTG GAT CCT GCA TAT TTT AGA CCC GCA GAG GTA GAA ACG   5374
 305  Ile Val Ala Val Asp Pro Ala Tyr Phe Arg Pro Ala Glu Val Glu Thr    320

5375  CTG TTA GGT GAT CCT AGC CTA GCT AAG AAG GAG CTT GGA TGG GTG CCA   5422
 321  Leu Leu Gly Asp Pro Ser Leu Ala Lys Lys Glu Leu Gly Trp Val Pro    336

5423  GAA ATT ACT TTA CAA CAG ATG GTT TCT GAA ATG GTA GCT TCT GAT TTA   5470
 337  Glu Ile Thr Leu Gln Gln Met Val Ser Glu Met Val Ala Ser Asp Leu    352

5471  GAA CAA GCT CAA AGT CAT GCA CTA TTG AAA AAA CAT GGC TAT AAC GTA   5518
 353  Glu Gln Ala Gln Ser His Ala Leu Leu Lys Lys His Gly Tyr Asn Val    368

5519  AAT GTA TCT GTA GAG TGA GGT CCT TTA AAT GAT TCC TGT ATA CGA ACC   5566
 369  Asn Val Ser Val Glu                                              373
```

Figure 5.9 Nucleotide sequence of the *V. cholerae rfb* region 5520 bp - 6670 bp.

The nucleotide sequence of the 1150 base-pair region containing the *rfbE* gene (nt 5547 to 6647). A putative promoter sequence is double underlined, however, no terminator sequences could be identified by DNASIS. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbE (translation region 5547-6647 bp)

```
                                                  -35                     -10
5472  AAC AAG CTC AAA GTC ATG CAC TAT TGA AAA AAC ATG GCT ATA ACG TAA    5519

5520  ATG TAT CTG TAG AGT GAG GTC CTT TAA ATG ATT CCT GTA TAC GAA CCA    5567
   1                                        Met Ile Pro Val Tyr Glu Pro      7

5568  AGT TTG GAT GGA AAT GAG CGT AAA TAT CTA AAC GAT TGC ATT GAT TCC    5615
   8  Ser Leu Asp Gly Asn Glu Arg Lys Tyr Leu Asn Asp Cys Ile Asp Ser     23

5616  GGT TGG GTA TCC TCA AGG GGG AAA TAT ATT GAT CGC TTC GAA ACT GAG    5663
  24  Gly Trp Val Ser Ser Arg Gly Lys Tyr Ile Asp Arg Phe Glu Thr Glu     39

5664  TTT GCG GAG TTT TTA AAA GTA AAG CAC GCC ACA ACA GTA TCT AAT GGA    5711
  40  Phe Ala Glu Phe Leu Lys Val Lys His Ala Thr Thr Val Ser Asn Gly     55

5712  ACA GTT GCG CTA CAT TTG GCA ATG AGC GCG TTG GGA ATA ACT CAA GGC    5759
  56  Thr Val Ala Leu His Leu Ala Met Ser Ala Leu Gly Ile Thr Gln Gly     71

5760  GAT GAA GTG ATT GTA CCA ACA TTC ACT TAT GTT GCC TCG GTT AAT ACC    5807
  72  Asp Glu Val Ile Val Pro Thr Phe Thr Tyr Val Ala Ser Val Asn Thr     87

5808  ATA GTC CAG TGT GGT GCG TTA CCC GTT TTT GCT GAA ATC GAA GGT GAG    5855
  88  Ile Val Gln Cys Gly Ala Leu Pro Val Phe Ala Glu Ile Glu Gly Glu    103

5856  TCT CTA CAA GTG AGC GTA GAG GAC GTT AAA CGT AAA ATA AAT AAA AAG    5903
 104  Ser Leu Gln Val Ser Val Glu Asp Val Lys Arg Lys Ile Asn Lys Lys    119

5904  ACA AAA GCA GTT ATG GCC GTT CAC ATA TAT GGA CAA GCT TGC GAT ATT    5951
 120  Thr Lys Ala Val Met Ala Val His Ile Tyr Gly Gln Ala Cys Asp Ile    135

5952  CAA TCT TTA AGA GAT CTG TGT GAT GAG CAC GGC CTC TAT TTA ATT GAA    5999
 136  Gln Ser Leu Arg Asp Leu Cys Asp Glu His Gly Leu Tyr Leu Ile Glu    151

6000  GAC TGT GCA GAA GCA ATA GGT ACT GCT GTT AAT GGC AAA AAA GTA GGT    6047
 152  Asp Cys Ala Glu Ala Ile Gly Thr Ala Val Asn Gly Lys Lys Val Gly    167

6048  ACA TTT GGC GAT GTG TCA ACG TTT AGT TTC TTT GGA AAT AAA ACC ATC    6095
 168  Thr Phe Gly Asp Val Ser Thr Phe Ser Phe Phe Gly Asn Lys Thr Ile    183

6096  ACA TCG GGC GAG GGT GGT ATG GTT GTG TCA AAC TCA GAC ATA ATT ATA    6143
 184  Thr Ser Gly Glu Gly Gly Met Val Val Ser Asn Ser Asp Ile Ile Ile    199

6144  GAT AAA TGT TTA CGC TTA AAA AAC CAA GGG GTA GTG GCA GGG AAG CGC    6191
 200  Asp Lys Cys Leu Arg Leu Lys Asn Gln Gly Val Val Ala Gly Lys Arg    215

6192  TAC TGG CAT GAC CTA GTT GCG TAT AAC TAT AGA ATG ACG AAT CTA TGT    6239
 216  Tyr Trp His Asp Leu Val Ala Tyr Asn Tyr Arg Met Thr Asn Leu Cys    231

6240  GCT GCG ATT GGT GTT GCT CAG CTT GAA AGA GTT GAT AAA ATT ATA AAG    6287
 232  Ala Ala Ile Gly Val Ala Gln Leu Glu Arg Val Asp Lys Ile Ile Lys    247
```

```
6288  GCC AAG CGT GAT ATA GCT GAA ATT TAT AGG TCT GAA CTC GCT GGA CTA   6335
 248  Ala Lys Arg Asp Ile Ala Glu Ile Tyr Arg Ser Glu Leu Ala Gly Leu    263

6336  CCA ATG CAA GTG CAC AAA GAG TCT AAT GGA ACT TTC CAC TCA TAT TGG   6383
 264  Pro Met Gln Val His Lys Glu Ser Asn Gly Thr Phe His Ser Tyr Trp    279

6384  TTG ACG TCT ATT ATA TTA GAC CAA GAG TTT GAA GTG CAC CGC GAT GGG   6431
 280  Leu Thr Ser Ile Ile Leu Asp Gln Glu Phe Glu Val His Arg Asp Gly    295

6432  TTG ATG ACC TTC TTA GAA AAT AAT GAC ATC GAG TCT AGA CCA TTT TTC   6479
 296  Leu Met Thr Phe Leu Glu Asn Asn Asp Ile Glu Ser Arg Pro Phe Phe    311

6480  TAT CCT GCT CAC ACG CTG CCA ATG TAC GAG CAT TTG GCA GAA AAG ACA   6527
 312  Tyr Pro Ala His Thr Leu Pro Met Tyr Glu His Leu Ala Glu Lys Thr    327

6528  GCT TTT CCG TTA TCA AAT AGC TAC TCT CAT CGA GGT ATT AAT TTA CCC   6575
 328  Ala Phe Pro Leu Ser Asn Ser Tyr Ser His Arg Gly Ile Asn Leu Pro    343

6576  AGT TGG CCA GGG TTA TGC GAT GAT CAA GTA AAA GAA ATT TGC AAC TGT   6623
 344  Ser Trp Pro Gly Leu Cys Asp Asp Gln Val Lys Glu Ile Cys Asn Cys    359

6624  ATT AAA AAT TAT TTC AAC TGT ATT TGA ATG TCA TGA CAG AAA AAA AGA   6671
 360  Ile Lys Asn Tyr Phe Asn Cys Ile                                   367
```

Figure 5.10 Nucleotide sequence of the *V. cholerae rfb* region 14881 bp - 15504 bp.

The nucleotide sequence of the 623 base-pair region containing the *rfbO* gene (nt 14929 to 15492). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbO (translation region 14929 bp - 15492 bp)

```
14881  AGA AAA ATT CGT TTC AAT AAG TCG ATG TTC CAA ACT TAG GGT GTA ATT  14928

14929  ATG AGC GGT AAC GGT TAC TAC AGT GAA GAT GTA TTG AAA CAA ATG GGT  14976
    1  Met Ser Gly Asn Gly Tyr Tyr Ser Glu Asp Val Leu Lys Gln Met Gly     16

14977  TTT TCT TCT TTA GGA AAG AAT GTA AAA ATT TCT GAA AAG GCT TCA CTT  15024
   17  Phe Ser Ser Leu Gly Lys Asn Val Lys Ile Ser Glu Lys Ala Ser Leu     32

15025  TAT GGC ATC AGT CGA ATT TCT ATA GGT TCA AAC GTT AGA ATT GAT GAT  15072
   33  Tyr Gly Ile Ser Arg Ile Ser Ile Gly Ser Asn Val Arg Ile Asp Asp     48

15073  TAT GTC ACG ATA TCA GCA GGT GTC GGC GGC GTC GAA ATT GGC TCT CAC  15120
   49  Tyr Val Thr Ile Ser Ala Gly Val Gly Gly Val Glu Ile Gly Ser His     64

15121  GTA CAT ATT GGA GTA TAC AGC AGC TTG ATA GGC GCT GGA AAA ATA ACC  15168
   65  Val His Ile Gly Val Tyr Ser Ser Leu Ile Gly Ala Gly Lys Ile Thr     80

15169  CTA GAA GAT TTT GTA GGT GTA TCA GGA AGA GTA TCT ATT TAC TCT TCC  15216
   81  Leu Glu Asp Phe Val Gly Val Ser Gly Arg Val Ser Ile Tyr Ser Ser     96

15217  TCT GAT GAC TAC ACC GGA ATG GCT ATG AGT AAC CCA ACT GTT CCA GAA  15264
   97  Ser Asp Asp Tyr Thr Gly Met Ala Met Ser Asn Pro Thr Val Pro Glu    112

15265  GAG CTT ACG AAA GTT ACG TCA CTC CCA GTT TTG ATC AAA AAG CAC TCA  15312
  113  Glu Leu Thr Lys Val Thr Ser Leu Pro Val Leu Ile Lys Lys His Ser    128

15313  ATA CTT GGG GCT GGA TGT GTT GTT TTA CCT AAA GTC ATT GTG GGA GAA  15360
  129  Ile Leu Gly Ala Gly Cys Val Val Leu Pro Lys Val Ile Val Gly Glu    144

15361  GGT GTT TCA GTT GGA GCA TTA TCG TTG GTG AAT AAG TCT TTG GAT GAT  15408
  145  Gly Val Ser Val Gly Ala Leu Ser Leu Val Asn Lys Ser Leu Asp Asp    160

15409  TGG CAT ATA TAT TCT GGT AAT CCA ATA CAG AAA TTT ATC AGA AAG GCC  15456
  161  Trp His Ile Tyr Ser Gly Asn Pro Ile Gln Lys Phe Ile Arg Lys Ala    176

15457  CGT AAA CCA TTG GAG CTT GAA AAG AAA CTT ATT TTG TAG AAA AGT ATC  15504
  177  Arg Lys Pro Leu Glu Leu Glu Lys Lys Leu Ile Leu                    188
```

Figure 5.11 Nucleotide sequence of the *V. cholerae rfb* region 15505 bp - 15744 bp.

The nucleotide sequence of the 239 base-pair region containing the *rfbP* gene (nt 15514 to 15732). Using DNASIS no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbP (translation region 15514 bp - 15732 bp)

```
15505  AGA AGA TAC TTG ATT TTT ATT GTC ATC CCT AAA CCA CCG CTT TTA GCG   15552
    1             Met Ile Phe Ile Val Ile Pro Lys Pro Pro Leu Leu Ala      13

15552  GTG GTG ATT GTT CCT AGG GGC ATT TTT CCC GAA AAT GCG CCC ATG TTA   15600
   14  Val Val Ile Val Pro Arg Gly Ile Phe Pro Glu Asn Ala Pro Met Leu      29

15600  GAA GAC AAA CTC TTA TTC ACC ATA AGT AAG AGG ATT CAA ATA ACA TGG   15648
   30  Glu Asp Lys Leu Leu Phe Thr Ile Ser Lys Arg Ile Gln Ile Thr Trp      45

15649  GCG ACT ACA GAA GTT CAT CAC ACG TCT ATT GGC GTT GCA AAT ATC ATA   15696
   46  Ala Thr Thr Glu Val His His Thr Ser Ile Gly Val Ala Asn Ile Ile      61

15697  TCG TTT GGA CAC CCA GGG AAA CCG CAT GAA GGT CGT TGA TAT ATA AGG   15744
   62  Ser Phe Gly His Pro Gly Lys Pro His Glu Gly Arg                      73
```

Figure 5.12 Nucleotide sequence of the *V. cholerae rfb* region 15841 bp - 16080 bp.

The nucleotide sequence of the 239 base-pair region containing the *rfbQ* gene (nt 15883 to 16077). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbQ (translation region 15883 bp - 16077 bp)

```
15841  TGT TGT TTT GTA ATA GCT CCT AGA CAA TAT AGG AGC CTA AAT ATG AGC   15888
    1                                                      Met Ser       2

15889  GAG TCA ATC AAC CCA TTT ATG CAT TTC CAA ATC ATT AAA GAC TAT CGA   15936
    3  Glu Ser Ile Asn Pro Phe Met His Phe Gln Ile Ile Lys Asp Tyr Arg     18

15937  CAA GAA AGC AAA GTA GAA CAC AAA TTA TCA GAC ATT ATT TTG CTG ACA   15984
   19  Gln Glu Ser Lys Val Glu His Lys Leu Ser Asp Ile Ile Leu Leu Thr     34

15985  ATA TGT GGT GTT TTG TCG GGT CAC GAT GGC TGG GAT GGC ATT ATC GAT   16032
   35  Ile Cys Gly Val Leu Ser Gly His Asp Gly Trp Asp Gly Ile Ile Asp     50

16033  TTT GGT CAT GCT CGC TTA GAT TTC CTT AAA CGA TAT GGT CAC TTT TAA   16080
   51  Phe Gly His Ala Arg Leu Asp Phe Leu Lys Arg Tyr Gly His Phe         65
```

Figure 5.13 Nucleotide sequence of the *V. cholerae rfb* region 16033 bp - 16799 bp.

The nucleotide sequence of the 766 base-pair region containing the *rfbR* gene (nt 16087 to 16788). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold. The arrow head denotes the position of the Tn*1725* insertion, (nt 16370), sequenced from the plasmid pRMB15.

Vibrio cholerae RfbR (translation region 16087 bp - 16788 bp)

```
16033  TTT GGT CAT GCT CGC TTA GAT TTC CTT AAA CGA TAT GGT CAC TTT TAA   16080

16081  GCT GGA ATG CCT TCT GCG GAT ACG CTT TCT CGT GTG ATG GGT ATG ATT   16128
    1          Met Pro Ser Ala Asp Thr Leu Ser Arg Val Met Gly Met Ile      14

16129  AAT CCT GTA GCT TTG CAA AGA AGC TTC ATT ACC TGG ATG AAG GAC TGC   16176
   15  Asn Pro Val Ala Leu Gln Arg Ser Phe Ile Thr Trp Met Lys Asp Cys      30

16177  CAT ACA CTA ACG GAT GGT GAA GTT ATT GCC ATC GAC GGT GAA ACA TTA   16224
   31  His Thr Leu Thr Asp Gly Glu Val Ile Ala Ile Asp Gly Glu Thr Leu      46

16225  CGC GGC TCT TAT GAC CGC TCG AAA GGC AAA GGA ACA ATC CAC ATG GTG   16272
   47  Arg Gly Ser Tyr Asp Arg Ser Lys Gly Lys Gly Thr Ile His Met Val      62

16273  AAC GCT CTT GCT ACA GCA AAT GGA ATG AGC ATT GGG CAA CTG AAG GTT   16320
   63  Asn Ala Leu Ala Thr Ala Asn Gly Met Ser Ile Gly Gln Leu Lys Val      78

16321  GAT TCT AAG AGC AAC GAG ATT ACC GCG ATC CCC AAG CTA CTT GAC TTA   16368
   79  Asp Ser Lys Ser Asn Glu Ile Thr Ala Ile Pro Lys Leu Leu Asp Leu      94

16369  CTA GAT GTA AAA GGC TGC TTG ATT ACG ATT GAT GCA ATG GGC TGC CAA   16416
   95  Leu Asp Val Lys Gly Cys Leu Ile Thr Ile Asp Ala Met Gly Cys Gln     110

16417  AAG AAA ATA GCA CAG AAA ATT CTT GAT AAA GAA GCG GAT TAT TTA TTG   16464
  111  Lys Lys Ile Ala Gln Lys Ile Leu Asp Lys Glu Ala Asp Tyr Leu Leu     126

16465  GCG GTC AAA GGT AAT CAG GGA ATG CTT GAG CAA GCC TTT GAT GAT TAT   16512
  127  Ala Val Lys Gly Asn Gln Gly Met Leu Glu Gln Ala Phe Asp Asp Tyr     142

16513  CTT CGA ATG GAC ATG CTT CAC GAC TTC GAC GGT AGT TCT TAT AGT ACA   16600
  143  Leu Arg Met Asp Met Leu His Asp Phe Asp Gly Ser Ser Tyr Ser Thr     158

16561  CAA GAA AAA AGT CAC GGA AGA ATA GAA ACG AGA GTG GCT TTA GTG AAT   16608
  159  Gln Glu Lys Ser His Gly Arg Ile Glu Thr Arg Val Ala Leu Val Asn     174

16609  CGC GAT TTG TCG GTT TTG GGT GAT ATT GAA CAT GAA TGG CCT GAG CTT   16656
  175  Arg Asp Leu Ser Val Leu Gly Asp Ile Glu His Glu Trp Pro Glu Leu     190

16657  AAA TCA ATG GGC ACC GTG GCT TCA ATT CGA CAA GAA TCG GCA GTC GCA   16704
  191  Lys Ser Met Gly Thr Val Ala Ser Ile Arg Gln Glu Ser Ala Val Ala     206

16705  ACA GAG CAA GAT GTG AGT ATT CGT TAC TAC ATA TGC TCT AAA GAA TTG   16752
  207  Thr Glu Gln Asp Val Ser Ile Arg Tyr Tyr Ile Cys Ser Lys Glu Leu     222

16753  GAA GCT CAA ACG TTA CTT GAA GCG ACT TGT TCT CAC TAG GGG GTA GAG   16800
  223  Glu Ala Gln Thr Leu Leu Glu Ala Thr Cys Ser His                     234
```

Figure 5.14 Nucleotide sequence of the *V. cholerae rfb* region 16753 bp - 17040 bp.

The nucleotide sequence of the 287 base-pair region containing the *rfbS* gene (nt 16804 to 17007). Using DNASIS, no potential promoter or terminator sequences could be identified. The potential ribosome binding site is underlined. Both the nucleotides and the amino acids are numbered. The stop codon is in bold.

Vibrio cholerae RfbS (translation region 16804 bp - 17007 bp)

```
16753  GAA GCT CAA ACG TTA CTT GAA GCG ACT TGT TCT CAC TAG GGG GTA GAG   16800

16801  GTC ATG GAT TGG TCA CTT GAT ACC GCA TTT TGT GAG GAC AAT TCA CGC   16848
    1      Met Asp Trp Ser Leu Asp Thr Ala Phe Cys Glu Asp Asn Ser Arg      15

16849  ATT AGA GCG GAC GAT CGA GCA GAG GCC TTT GCA AGG ATC AGG CAG ATA   16896
   16  Ile Arg Ala Asp Asp Arg Ala Glu Ala Phe Ala Arg Ile Arg Gln Ile      31

16897  TGT TTG AAC CTA TTA AAG AGC GAA CCC ACG TTT AAA GGT GGT ATC AAA   16944
   32  Cys Leu Asn Leu Leu Lys Ser Glu Pro Thr Phe Lys Gly Gly Ile Lys      47

16945  CGT AAA CGG ATG AAC TGC GCA ATG AAC GAA AAG TAC CTA AGT AAG GTT   16992
   48  Arg Lys Arg Met Asn Cys Ala Met Asn Glu Lys Tyr Leu Ser Lys Val      63

16993  CTT GAA AGC CTT ACG TGA CGG TGA TGT TCA TGC GGT TTC CGT GCT AAA   17040
   64  Leu Glu Ser Leu Thr                                              68
```

Table 5.1 Summary of features of *rfb* encoded ORFs.

ORF	TRANSLATION REGION (bp)	SHINE-DALGARNO	M_r	AMINO ACIDS	PREDICTED P_I
RfbB	3022-3828	TCGAGGA	29,384	269	5.79
RfbC	3837-4409	CGTAGGT	19,235	169	9.28
RfbD	4415-5533	TGGATAT	42,031	373	6.14
RfbE	5547-6647	GTGAGGT	40,989	367	6.06
RfbO	14929-15492	AGGGTGT	19,872	188	9.83
RfbP	15514-15732	AGAAGAT	8,073	73	10.19
RfbQ	15883-16077	AGGAGCT	7,558	65	6.47
RfbR	16087-16788	TTAAGCT	25,826	234	5.79
RfbS	16804-17007	AGGGGGT	7,881	68	10.19

A ribosome binding site (RBS) is required for the efficient initiation of protein translation. The RBS consists essentially of an initiation codon (most commonly AUG), and the Shine-Dalgarno sequence which is a sequence upstream displaying homology with the 3' end of the 16S rRNA (Shine and Dalgarno, 1974). The Shine-Dalgarno sequence is a 3-9 bp long, purine-rich sequence (5'AGGAGGU3'). Mutations in this region have been shown to drastically affect the levels of initiation of translation (Gold *et al.*,1981; Kozak, 1983). It has been shown that the distance between the initiation codon and the Shine-Dalgarno sequence, usually ranging fom 3 to 11 nt, optimally 7 nt, can effect the translational efficiency (Itoh *et al.*, 1984; Shepard *et al.*, 1982). For example the Shine-Dalgarno sequence for *rfbB* UC**GAGGA** displays excellent homology to the consensus but is located 9nt from the initiation codon. These sequences are summarized in Table 5.1.

Also of note is the fact that not all of the genes have an AUG initiation codon. *rfbB* has GUG (which is employed as the initiation codon in about 10% of known *E. coli* genes) (Niedhardt, 1987), and *rfbC* and *rfbP* have UUG codons (which is used less frequently).

5.2.2.2 Translational coupling

Bacterial genes that code for functionally related products are often co-transcribed as a polycistronic mRNA that is then translated into protein products. The different enzymes encoded by a single mRNA need not necessarily be synthesized at the same rate since factors such as the sequence of the Shine-Dalgarno region and codon usage will affect both efficiency of translation and rate of polypeptide elongation. This phenomenon, called translational coupling was first observed in the tryptophane operon of *E. coli* (Oppenheim and Yanofsky, 1980), and has been observed in many other systems since (Aksoy *et al.*, 1984; Lindahl *et al.*, 1989, Little *et al.*,1989). When two

genes are translationally coupled, efficient translation of a downstream gene depends on the translation of the adjacent upstream gene. Translationally coupled genes are often separated by intercistronic regions which are 3 bases or less (Aksoy *et al.*, 1984; Brot *et al.*, 1986; Little *et al.*, 1989; Oppenheim and Yanofsky, 1980; Schumperli *et al.*, 1982). In several of these cases the termination codon of the upstream gene even overlaps with the initiation codon of the downstream gene (Schumperli *et al.*, 1982). It has been speculated that the short distance between the genes is essential for efficient translational coupling. However, ribosomal protein operons have been shown to contain genes separated by longer intercistronic regions (Lindahl *et al.*, 1989; Mattheakis and Nomura, 1988; Oppenheim and Yanofsky, 1980), and in the *E. coli* α operon it has been reported that the ribosomal protein gene L17, is translationally coupled to upstream ribosomal protein genes, despite the presence of an intervening, separately regulated gene (Thomas *et al.*, 1987).

Oppenheim and Yanofsky (1980) speculated on two theories in relation to translational coupling:

1) that translation of the promoter-proximal gene ensures a high localized concentration of ribosomes around the RBS of the second gene, or
2) that the polycistronic message in the absence of translation, forms a secondary structure which sequesters the RBS of the coupled gene.

Analysis of the translation signals involved in the expression of the *rfb* region, revealed:

1) no putative terminator regions proximal to each gene,
2) no terminator regions distal to each gene, and
3) close association between the Shine-Dalgarno sequence and the termination codon of the preceeding gene.

With these observations in mind it is proposed that the expression of the *V. cholerae rfb* region is translationally coupled as shown in Table 5.2. This, however,

Table 5.2 Translational coupling of the *rfb* genes

rfbA *rfbB*

GAG GAT AAA TAG A **GTG** AAA
Gly Asp Lys *** Met Lys

rfbB *rfbC*

ACA TCG TAG GTT TA **TTG** GCA
Thr Ser *** Met Ala

rfbC *rfbD*

TTG GAT ATT TAA CA **ATG** ATT
Leu Asp Ile *** Met Asn

rfbD *rfbE*

GAG TGA GGT CCT TTA A **ATG**
Glu *** Met

rfbE *rfbG*

TGT ATT TGA ATG TC **ATG** GAC
Cys Ile *** Met Thr

rfbG *rfbH*

GGG AAT GAA A**AT** **G**CT TAA A GAT
Gly Asn Glu Asn Ala ***
Met Leu Lys Asp

rfbH *rfbI*

TGG GTG TAG T **ATG** ATA GAG
Trp Val *** Met Ile Glu

rfbI *rfbK*

38 bp

rfbK *rfbL*

AAC GAG TAA ATA TT **ATG** CTA
Asn Glu *** Met Ser

rfbL *rfbM*

GAA TAG GAA CAT TC **ATG** CTA
Glu *** Met Leu

rfbM *rfbN*

133 bp

rfbN *rfbO*

69 bp

rfbO *rfbP*

18 bp

rfbP *rfbQ*

147 bp

rfbQ *rfbR*

CAC TTT TAA GCT TGG A **ATG** CCT
His Phe *** Met Pro

rfbR *rfbS*

CAC TAG GGG GTA GAGA GTC **ATG**
His *** Met

rfbS *rfbT*

135 bp

Footnote: Bases in bold indicate start codon; bases underlined show position of potential Shine-Dalgarno sequences; (*) denotes position of stop codons.

appears only to be the case with expression of *rfbA* to *rfbM* after which the coupling seems to break down, as *rfbM* and *rfbN* are separated by 133 bp.

5.2.2.3 Codon usage

A summary of the codon usage within the coding region of the VcRfb proteins is shown in Table 5.3. Table 5.3 shows the putative *rfb* protein codon usage as compared to the average usage in other sequenced *V. cholerae* genes. These genes include *toxR* (Miller *et al.*, 1987), *toxS* (Di Rita and Mekalanos, 1991), *ctxA, B* (Mekalanos *et al.*, 1983), *pac* (Guidolin and Manning, 1988), *ompV* (Pohlner *et al.*, 1986), *ompW* (Jalajakumari and Manning, 1990), DNAse (Focareta and Manning, 1987), *mutL* (Bera *et al.*, 1989), *tcpA* (Faast *et al.*, 1989), *tcpB, H, I, X* (Ogierman, M., Meahney, C., and Manning, P.A., manuscript in preparation), *hlyA, B, C* (Alm, R., personal communication; Alm *et al.*, 1988; Alm and Manning, 1990) and a locus of nine genes including the MFRHA (van Dongen *et al.*, 1987; Barker, A., Clark, C.A., Williams, S.G., Franzon, V.L., and Manning P.A., manuscript in preparation). Unfortunately this codon usage may be slightly biased as these genes represent a select group of proteins, namely outer membrane and secreted proteins. There are not a sufficient number of sequenced *rfb* genes from other organisms to enable the construction of a "*rfb*" codon usage table.

As shown in Table 5.3, the preferred codon usage within the *rfb* genes analyzed, conforms well with the other *V. cholerae* genes.

5.2.2.4 G+C content of the *rfb* region

The G+C content of the open reading frames contained in the 20 kb *Sac*I fragment were analyzed using the DNASIS programme, and shown in Table 5.4. The G+C content of the *rfb* genes was found to be significantly lower than the average G+C content of *V.*

Table 5.3 Comparison of *rfb* gene codon usage with the codon usage amongst sequenced *V. cholerae* genes

Codon	*%V. cholerae*	*%rfbB*	*%rfbC*	*%rfbD*	*%rfbE*	*%rfbO*	*%rfbP*	*%rfbQ*	*%rfbR*	*%rfbS*
UUU-Phe	70.5	75.0	70.0	78.6	56.3	100.0	75.0	60.0	33.3	100.0
UUC-Phe	29.5	25.0	30.0	21.4	43.7	0.0	25.0	40.0	66.7	0.0
UUA-Leu	28.0	38.5	40.9	35.3	39.3	17.6	50.0	33.3	20.8	14.3
UUG-Leu	19.5	34.6	22.7	17.6	21.4	41.2	16.7	33.3	25.0	14.3
CUU-Leu	18.4	15.4	22.7	21.2	3.6	29.4	16.7	16.7	37.5	42.9
CUC-Leu	8.1	0.0	0.0	0.0	7.1	5.9	16.7	0.0	0.0	0.0
CUA-Leu	13.8	7.7	4.5	24.2	21.4	5.9	0.0	0.0	12.5	28.6
CUG-Leu	12.2	3.8	9.1	3.0	7.1	0.0	0.0	16.7	4.2	0.0
AUU-Ile	50.7	73.9	58.8	57.1	53.6	44.4	54.5	50.0	64.7	25.0
AUC-Ile	26.2	21.7	17.6	14.3	10.7	16.7	18.2	37.5	17.6	50.0
AUA-Ile	23.1	4.3	23.5	28.6	35.7	38.9	27.3	12.5	17.6	25.0
AUG-Met	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
GUU-Val	36.2	55.0	25.0	25.8	41.4	38.1	50.0	50.0	23.1	100.0
GUC-Val	15.0	5.0	12.5	6.5	3.4	19.0	16.7	0.0	15.4	0.0
GUA-Val	27.1	15	12.5	51.6	31.0	33.3	0.0	50.0	15.4	0.0
GUG-Val	21.7	25	50.0	16.1	24.1	9.5	33.3	0.0	46.2	0.0
UCU-Ser	22.1	16.7	46.7	28.6	33.3	40.0	33.3	0.0	36.8	0.0
UCC-Ser	7.7	0.0	6.7	14.3	8.3	4.0	0.0	0.0	0.0	0.0
UCA-Ser	21.0	16.7	6.7	28.6	25.0	28.0	0.0	40.0	10.5	40.0
UCG-Ser	11.8	16.7	13.3	4.7	8.3	4.0	33.3	20.0	15.8	0.0
AGU-Ser	22.1	33.3	6.7	9.5	12.5	12.0	33.3	0.0	21.1	20.0
AGC-Ser	15.4	16.7	20.0	14.3	12.5	12.0	0.0	40.0	15.8	40.0
CCU-Pro	29.6	20.0	16.7	64.3	18.2	16.7	25.0	0.0	75.0	0.0
CCC-Pro	11.0	20.0	16.7	7.1	18.2	0.0	25.0	0.0	25.0	100.0
CCA-Pro	38.5	40.0	33.3	21.4	54.5	83.3	25.0	100.0	0.0	0.0
CCG-Pro	20.9	20.0	33.3	7.1	9.1	0.0	25.0	0.0	0.0	0.0
ACU-Thr	29.7	38.5	50.0	50.0	32.5	16.7	20.0	0.0	6.7	0.0
ACC-Thr	23.4	38.5	16.7	5.0	15.0	33.3	20.0	0.0	20.0	33.3
ACA-Thr	28.3	23.1	33.3	25.0	40.0	0.0	40.0	100.0	40.0	0.0
ACG-Thr	18.7	0.0	0.0	20.0	20.0	50.0	20.0	0.0	33.3	66.7
GCU-Ala	34.8	34.6	9.1	37.9	36.0	57.1	0.0	100.0	33.3	0.0
GCC-Ala	15.2	9.1	36.4	3.4	16.0	14.3	0.0	0.0	11.1	16.7
GCA-Ala	29.7	36.4	36.4	27.6	24.0	28.6	25.0	0.0	27.8	66.7
GCG-Ala	20.2	11.5	18.2	31.0	24.0	0.0	75.0	0.0	27.8	16.7

Table 5.3 cont.

Codon	%*V. cholerae*	%*rfbB*	%*rfbC*	%*rfbD*	%*rfbE*	%*rfbO*	%*rfbP*	%*rfbQ*	%*rfbR*	%*rfbS*
UAU-Tyr	64.0	44.4	33.3	73.7	73.3	37.5	0.0	100.0	66.7	0.0
UAC-Tyr	36.0	55.6	66.7	26.3	26.7	62.5	0.0	0.0	33.3	100.0
CAU-His	67.0	50.0	83.3	88.9	36.4	50.0	50.0	50.0	33.3	0.0
CAC-His	33.0	50.0	16.7	11.1	63.6	50.0	50.0	50.0	66.7	0.0
CAA-Gln	68.1	75.0	60.0	58.8	80.0	50.0	100.0	100.0	80.0	0.0
CAG-Gln	31.9	25.0	40.0	41.2	20.0	50.0	0.0	0.0	20.0	100.0
AAU-Asn	61.5	50.0	87.5	58.3	68.4	50.0	100.0	0.0	66.7	25.0
AAC-Asn	38.5	50.0	12.5	41.7	31.6	50.0	0.0	100.0	33.3	75.0
AAA-Lys	70.3	65.0	62.5	66.7	72.7	60.0	75.0	100.0	66.7	50.0
AAG-Lys	29.7	35.0	37.5	33.3	27.3	40.0	25.0	0.0	33.3	50.0
GAU-Asp	76.4	87.5	100.0	75.0	71.4	87.5	0.0	66.7	63.5	60.0
GAC-Asp	23.6	12.5	0.0	25.0	28.6	12.5	100.0	33.3	36.5	40.0
GAA-Glu	67.1	77.3	53.3	72.4	56.0	77.8	100.0	66.7	73.3	60.0
GAG-Glu	32.9	22.7	46.7	27.6	44.0	22.2	0.0	33.3	26.7	40.0
UGU-Cys	65.2	66.7	100.0	50.0	63.4	100.0	0.0	100.0	25.0	66.7
UGC-Cys	34.7	33.3	0.0	50.0	36.4	0.0	0.0	0.0	75.0	33.3
UGG-Trp	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
CGU-Arg	27.5	37.5	37.5	17.6	21.4	20.0	33.3	0.0	20.0	14.3
CGC-Arg	17.6	25.0	25.0	5.9	28.6	0.0	0.0	33.3	30.0	14.3
CGA-Arg	15.7	25.0	0.0	17.6	7.1	20.0	0.0	66.7	20.0	14.3
CGG-Arg	6.5	0.0	0.0	5.9	0.0	0.0	0.0	0.0	0.0	14.3
AGA-Arg	23.6	12.5	37.5	47.1	28.6	60.0	0.0	0.0	30.0	14.3
AGG-Arg	9.0	0.0	0.0	5.9	14.3	0.0	66.7	0.0	0.0	28.6
GGU-Gly	43.3	66.7	63.7	26.9	34.6	36.4	20.0	66.7	37.5	100.0
GGC-Gly	24.2	16.7	9.1	23.1	19.2	22.7	40.0	33.3	31.3	0.0
GGA-Gly	18.9	8.3	9.1	34.6	26.9	36.4	20.0	0.0	25.0	0.0
GGG-Gly	13.6	8.3	18.2	15.4	19.2	4.5	20.0	0.0	6.3	0.0
UAA-STOP	43.8	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
UAG-STOP	29.2	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0
UGA-STOP	27.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0

TABLE 5.4 G+C CONTENT OF THE *RFB* GENES

Rfb gene	G+C ratio
A	42.1%
B	39.5%
C	37.7%
D	39.6%
E	40.1%
G	35.2%
H	35.1%
I	36.5%
K	34.2%
L	37.7%
M	43.3%
N	42.4%
O	39.0%
P	42.9%
Q	37.9%
R	43.3%
S	44.1%
T	31.1%
ORF1	46.2%
ORF2	45.8%
ORF3	44.3%
ompX	35.3%

cholerae (47-49%) (Baumann and Schubert, 1984). Comparison of the G+C content of the *rfb* encoded proteins revealed a series of blocks along the 20 kb length. This could have some bearing on the arrangement of functional units within the *rfb* operon.

ORF1, which is not involved in O-antigen synthesis, has the highest value of 46.2%. A block of genes containing *rfbA* to *rfbE* have an average G+C content of 39.7%. Two other blocks can be created from *rfbG* to *rfbL* (average G+C of 35.7%), and from *rfbM* to *rfbS* (42.3% G+C). *rfbT* stands out alone as its G+C content is only 31.1%. This may be a recent acquisition to the *rfb* region, derived from a different organism, or perhaps a defective phage. ORFs 2 and 3 also have a G+C content with an average of 45.0%, which together with ORF1 is not too dissimilar to the average G+C content for *V. cholerae*.

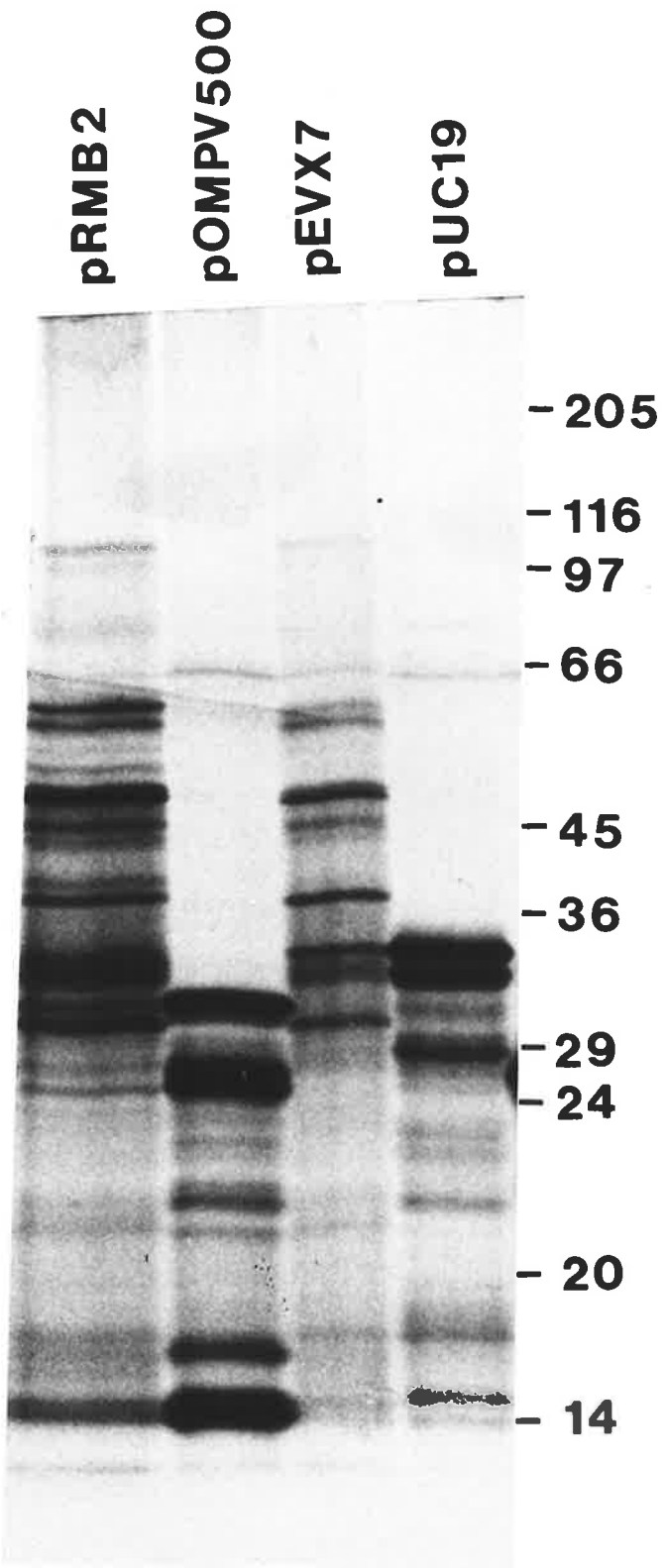
5.2.3 Identification of potential ORFs by DNA sequence analysis

5.2.3.1 Identification of the proteins encoded by the *rfb* region

The 20 kb *Sac*I fragment encoding the *rfb* region of *V. cholerae* Ogawa has been cloned into two different plasmid vectors; pUC18, resulting in the plasmids pEVX6 and pEVX7, and in a pBR322 derivative pOmpV500 (previously described in 3.2.1), to produce pRMB1, and pRMB2. These constructs, together with the appropriate controls, were analyzed in an *in vitro* transcription/translation system for preliminary identification of the proteins encoded by this region. Figure 5.15 shows an autoradiograph of the SDS-PAGE gel of this procedure. As can be seen, a large number (17 identifiable) of proteins appear to be encoded by the 20 kb fragment. The putative sizes of these proteins are (in kDa); 105, 52.8, 52.5, 47.9, 46.8, 42, 36, 35.2, 34.7, 32.5, 29.9, 29.4, 20, 18.9, 16.3, 12.2, and 8.8. This represents 88% of the total coding capacity of the *Sac*I region. To enable one to positively identify the proteins encoded by each putative gene, subclones need to

Figure 5.15 Autoradiograph of *in vitro* translation studies of the 20 kb *Sac*I region.

Plasmid encoded proteins were labelled with [^{35}S]-methionine using an *in vitro* transcription/translation kit (Amersham). The plasmid encoded proteins were visualized by autoradiography after electrophoretic separation on an 11-20% polyacrylamide gradient gel. The size standards (kDa) are shown on the right hand side.


pRMB2
pOMPV500
pEVX7
pUC19
205
116
97
66
45
36
29
24
20
14

be analyzed using this same system. Attempts to identify the proteins using *E. coli* minicells did not produce any results due to an instability problem seen with the whole 20 kb region in the minicell producing strain of *E. coli* (DS410). Various constructed subclones described in Figure 3.3, were employed in an attempt to identify the proteins using the minicell system (data not shown). The data obtained were ambiguous primarily due to insufficient data to select the appropriate subclones for study. This is now straightforward with the availability of the DNA sequence complete restriction data, and consequently a detailed approach to protein identification could now be undertaken.

5.2.3.2 Hydropathic profile and predicted secondary structure of 9 VcRfb proteins

Figures 5.16 to 5.24 show the hydropathic plot of the 9 VcRfb proteins, (VcRfbB, C, D, E, O, P, Q, R, S) according to the methods of Kyte and Doolittle (1982), superimposed on the predicted secondary structure (Chou and Fasman, 1974, 1974a, 1978).

All of the proteins to be discussed have been analyzed for the presence of a signal sequence, required for protein secretion (Inouye and Halegoua; 1980, von Heijne, 1983), however, none could be identified. It was therefore assumed that these proteins are either found in the cytoplasmic membrane or present within the cytoplasm itself. An attempt to decide between these two locations was made by using the hydropathic profile in conjunction with the predicted secondary structure.

Integral membrane proteins can be classified on the basis of the number of times the polypeptide spans the membrane into;

(a) monotopic proteins - which are rare (one example is cytochrome b5) (Arinc *et al.*, 1987). These proteins are hydrophobically associated with the membrane but do not pass all the way across the bilayer,

(b) biotopic proteins - are quite common (eg. oncogene products), and cross the membrane once (Yarden and Ullrich, 1988), and,

(c) polytopic - are a large group of proteins which cross the membrane more than once (Blobel, 1980).

A sequence of hydrophobic amino acid residues is an energetically favourable structure for a membrane spanning alpha helix (Jennings, 1989). The length required to span a lipid bilayer if the helix is orientated perpendicular to the bilayer plane, is 20 residues in an alpha helix formation, however, this region may be smaller if the amino acids are in a beta sheet formation.

VcRfbB (Figure 5.16) has a mean hydrophilic index of 0.0. As can be seen there are a large number of hydrophilic regions interspersed between hydrophobic domains. From aa 238 to 269 is a very hydrophobic region, comprising residues in an alpha helical arrangment, which may serve as a membrane anchoring domain. Using the secondary structure prediction and the hydropathy plot, it can be postulated that this protein is a cytoplasmic membrane protein.

VcRfbC (Figure 5.17) has a mean hydrophilic index of -0.2, indicating its overall hydrophilic nature. The hydrophobic domains of this protein are not large enough to span the membrane, therefore this protein may be located in the cytoplasm.

VcRfbD (Figure 5.18) is also a hydrophilic protein with a mean hydrophilic index of -0.3. This protein (discussed in more detail in 5.2.3.3) although possessing a strong but short hydrophobic region at aa 301 to 310, is thought to reside in the cytoplasm.

VcRfbE (Figure 5.19) has a mean hydrophilic index of 0.1 and possesses a strong hydrophobic region at aa 65 - aa 100, comprised of amino acids mainly in an alpha helical formation, which may function as a membrane anchor, thereby indicating the likelihood that VcRfbE is a cytoplasmic membrane protein. Although other regions of hydrophobicity are shown, as previously mentioned, these regions are not sufficient in length to span the membrane.

Figure 5.16 Secondary structure prediction and hydrophobicity of VcRfbB.

The protein sequence of VcRfbB was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chǒu and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix: 

random coil:

β sheet: 

β turn: where the structure rotates by 90°

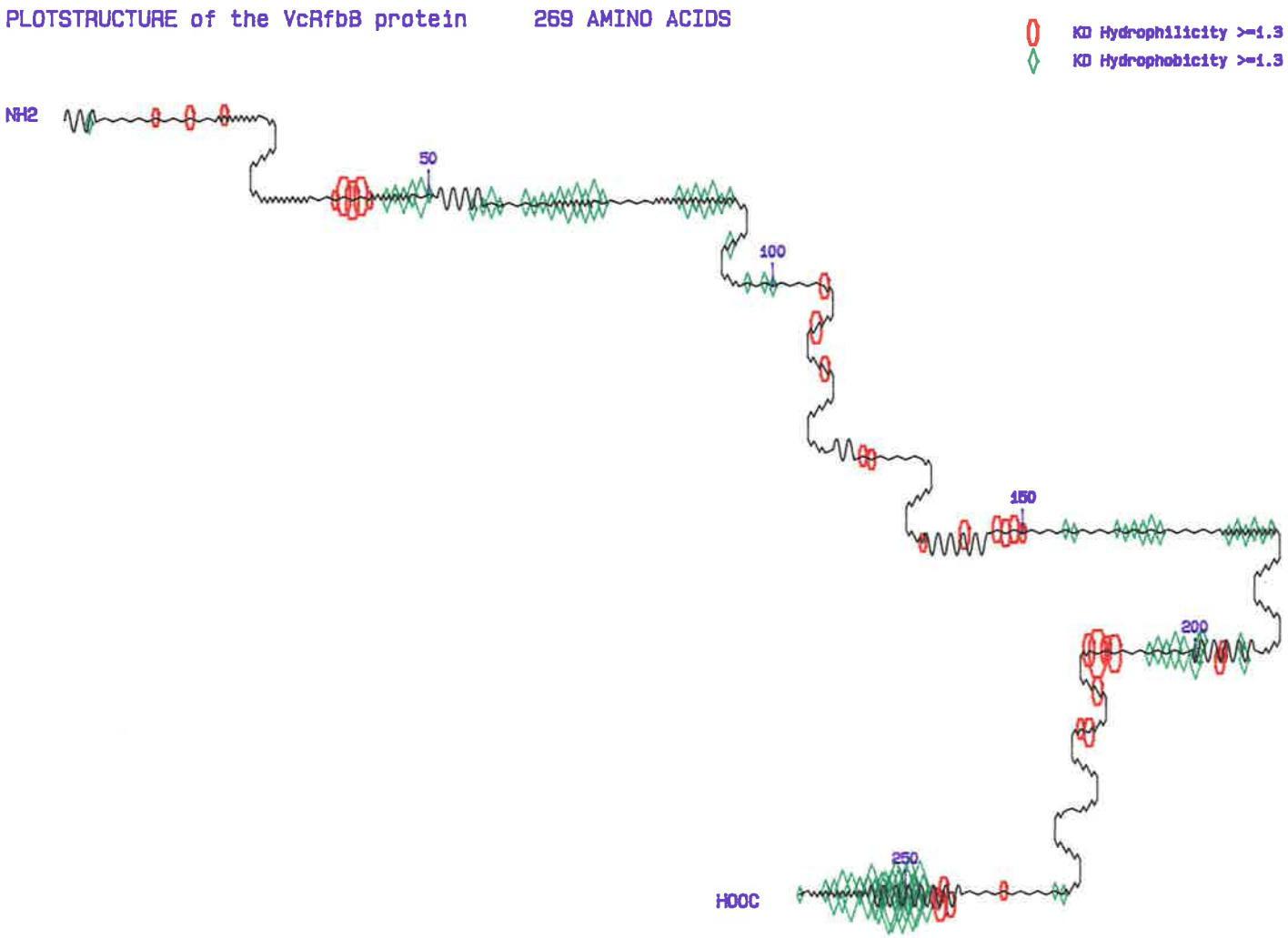

PLOTSTRUCTURE of the VcRfbB protein 269 AMINO ACIDS
KD Hydrophilicity >=1.3
KD Hydrophobicity >=1.3
NH2
50
100
150
200
250
HOOC

Figure 5.17 Secondary structure prediction and hydrophobicity of VcRfbC.

The protein sequence of VcRfbC was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix: 

random coil:

β sheet: 

β turn: where the structure rotates by 90°

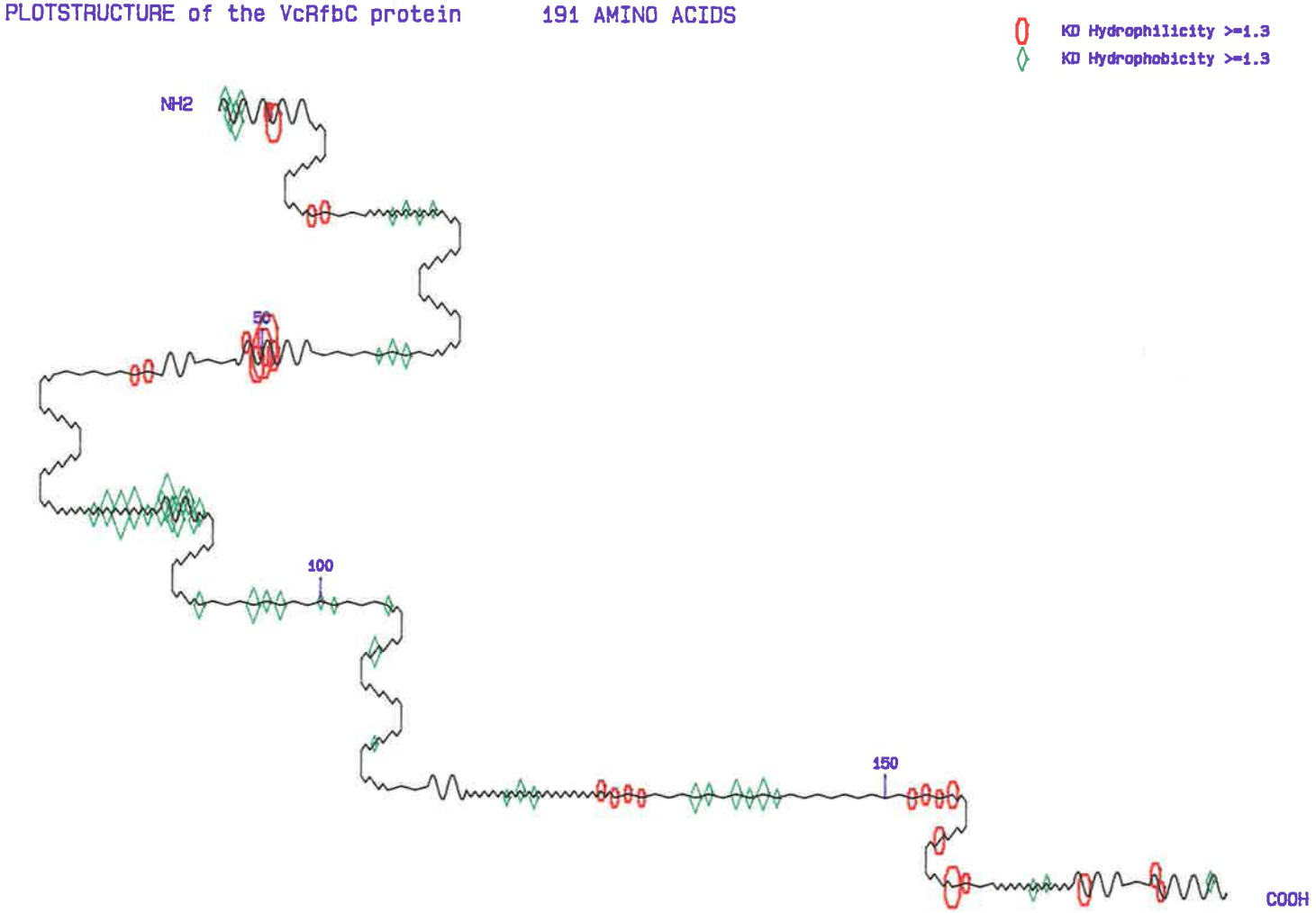

PLOTSTRUCTURE of the VcRfbC protein
191 AMINO ACIDS
KD Hydrophilicity >=1.3
KD Hydrophobicity >=1.3
NH2
50
100
150
COOH

Figure 5.18 Secondary structure prediction and hydrophobicity of VcRfbD.

The protein sequence of VcRfbD was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix: 

random coil:

β sheet: 

β turn: where the structure rotates by 90°

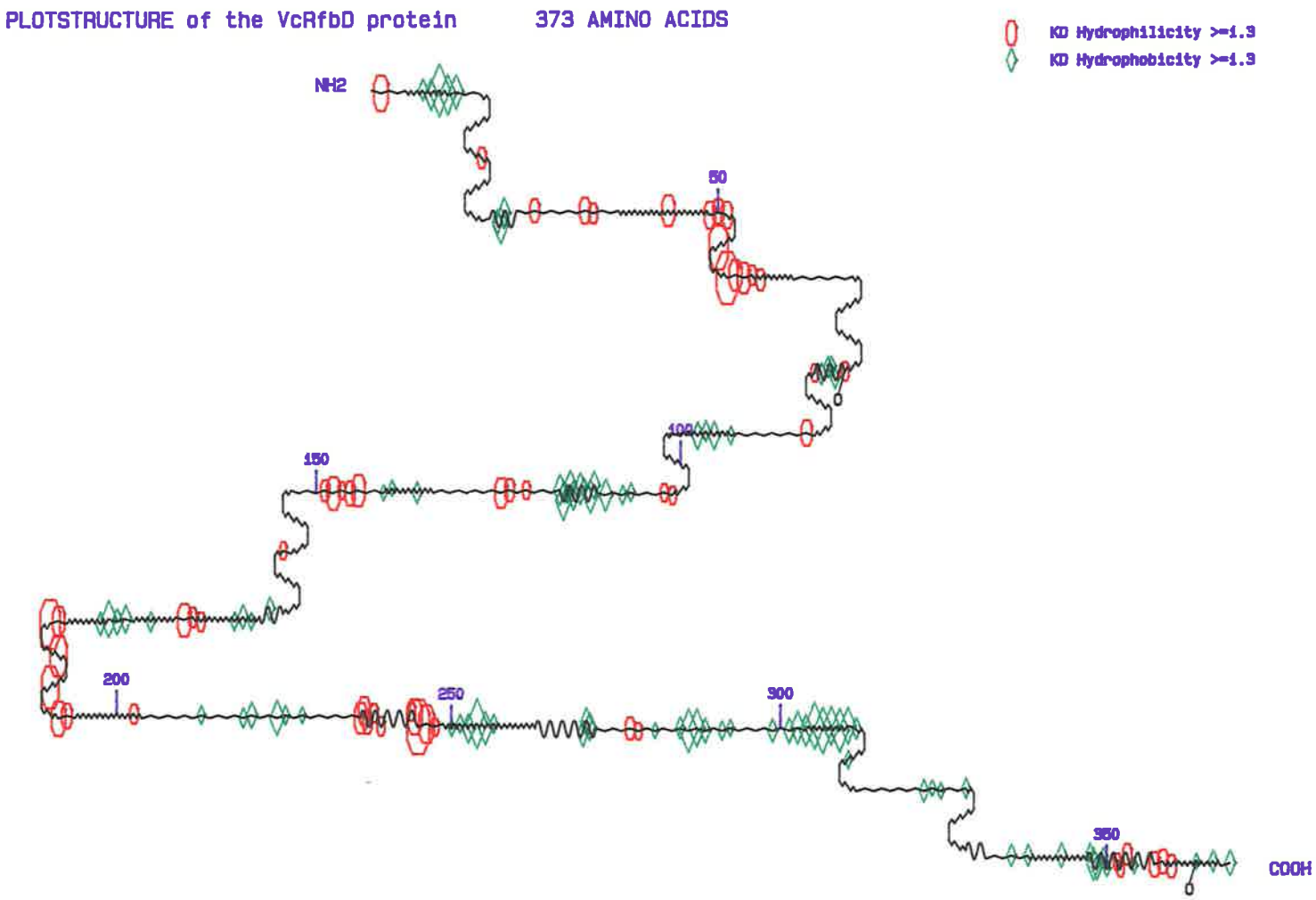

PLOTSTRUCTURE of the VcRfbD protein
373 AMINO ACIDS
KD Hydrophilicity >=1.3
KD Hydrophobicity >=1.3
NH2
50
100
150
200
250
300
350
COOH

Figure 5.19 Secondary structure prediction and hydrophobicity of VcRfbE.

The protein sequence of VcRfbE was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

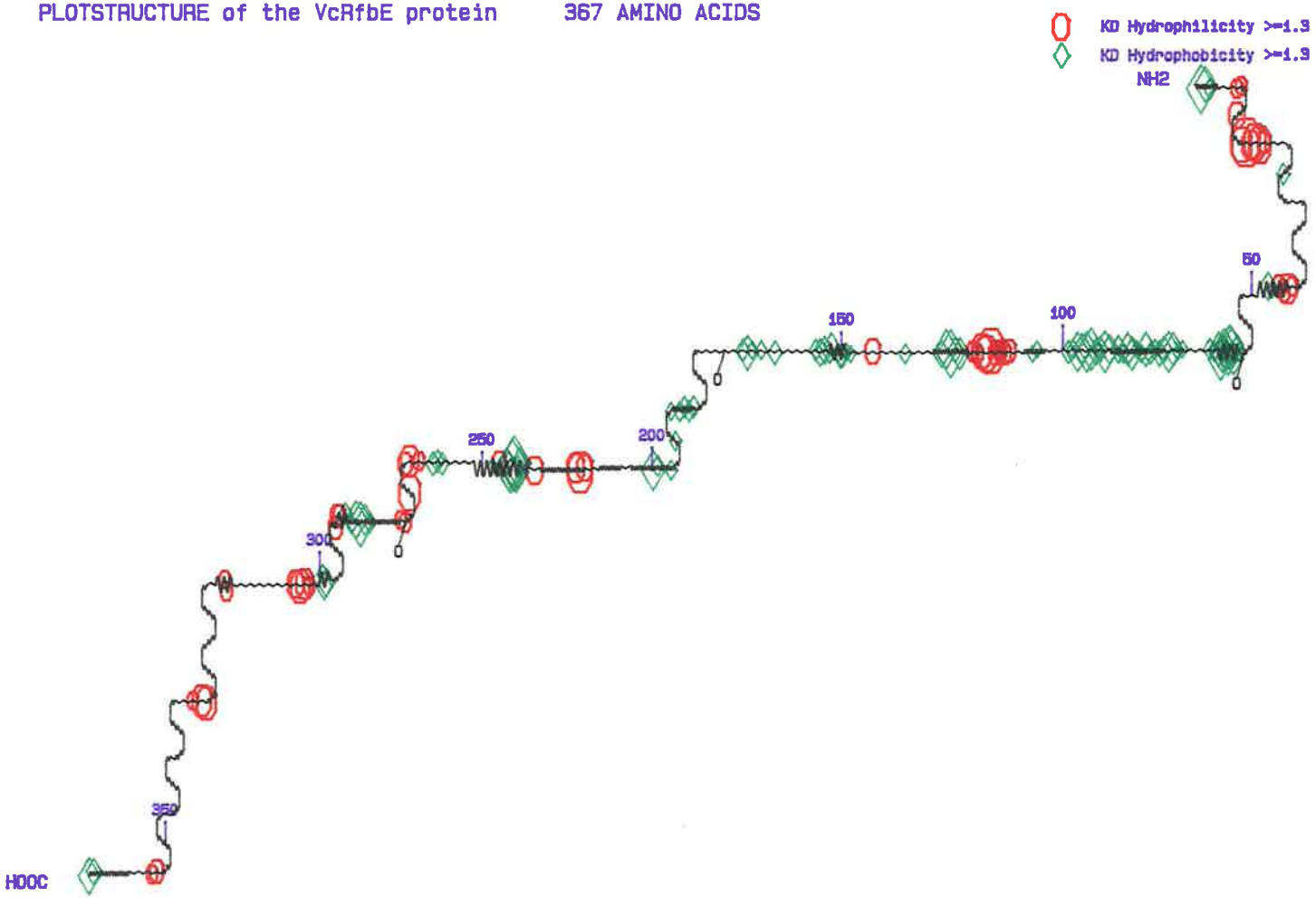
α helix: 

random coil:

β sheet: 

β turn: where the structure rotates by 90°

PLOTSTRUCTURE of the VcRfbE protein 367 AMINO ACIDS



VcRfbO (Figure 5.20) has a mean hydrophilic index of 0.2 as it has many hydrophobic domains. These domains, for example, from aa 137 to aa 155 are likely to be membrane spanning domains, indicating the posibility that this protein would probably reside in the cytoplasmic membrane.

VcRfbP (Figure 5.21) is very hydrophobic, especially at the NH_2 end, with a mean hydrophilic index of 0.4, however, as with VcRfbE, the hydrophobic regions are not of a sufficient length to be involved in spanning the membrane.

VcRfbQ (Figure 5.22) has a very hydrophobic region from aa 22 to aa 36, which may function as a membrane anchoring domain.

VcRfbR (Figure 5.23) is similar to VcRfbQ, in that a it possesses a hydrophobic region at aa 80 to aa 106 which may function as a membrane anchor.

VcRfbS (Figure 5.24) is a small, hydrophilic protein (mean index -0.6), which is probably located within the cytoplasm.

5.2.3.3 Homology profiles

Computer homology searches of the proteins VcRfbB, C, D, E, O, P, Q, R, S were performed using FASTA, and the SWISSPROT, protein sequence database. From these studies proteins with significant homology to the *Vibrio cholerae rfb* encoded proteins VcRfbB, D, E were be detected, however, no sequences significantly similar to VcRfbO, P, Q, or S were detected. The VcRfb protein sequences and the corresponding homologous sequences were aligned via the CLUSTAL program (Higgins and Sharp, 1989), shown in Figures 5.25, 5.26, 5.27, and 5.28.

Figure 5.25 depicts the homology seen between *V. cholerae* RfbB (VcRfbB) and the RfbK protein from *S. typhimurium* (StRfbK). The *S. typhimurium rfbK* gene encodes the enzyme phospho-manno-mutase, the second enzyme of the GDP-mannose pathway. The active sites within the protein have not been localized, but as can be seen, significant

Figure 5.20 Secondary structure prediction and hydrophobicity of VcRfbO.

The protein sequence of VcRfbO was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix:

random coil:

β sheet: 

β turn: where the structure rotates by 90°

PLOTSTRUCTURE of the VcRfbO protein 188 AMINO ACIDS

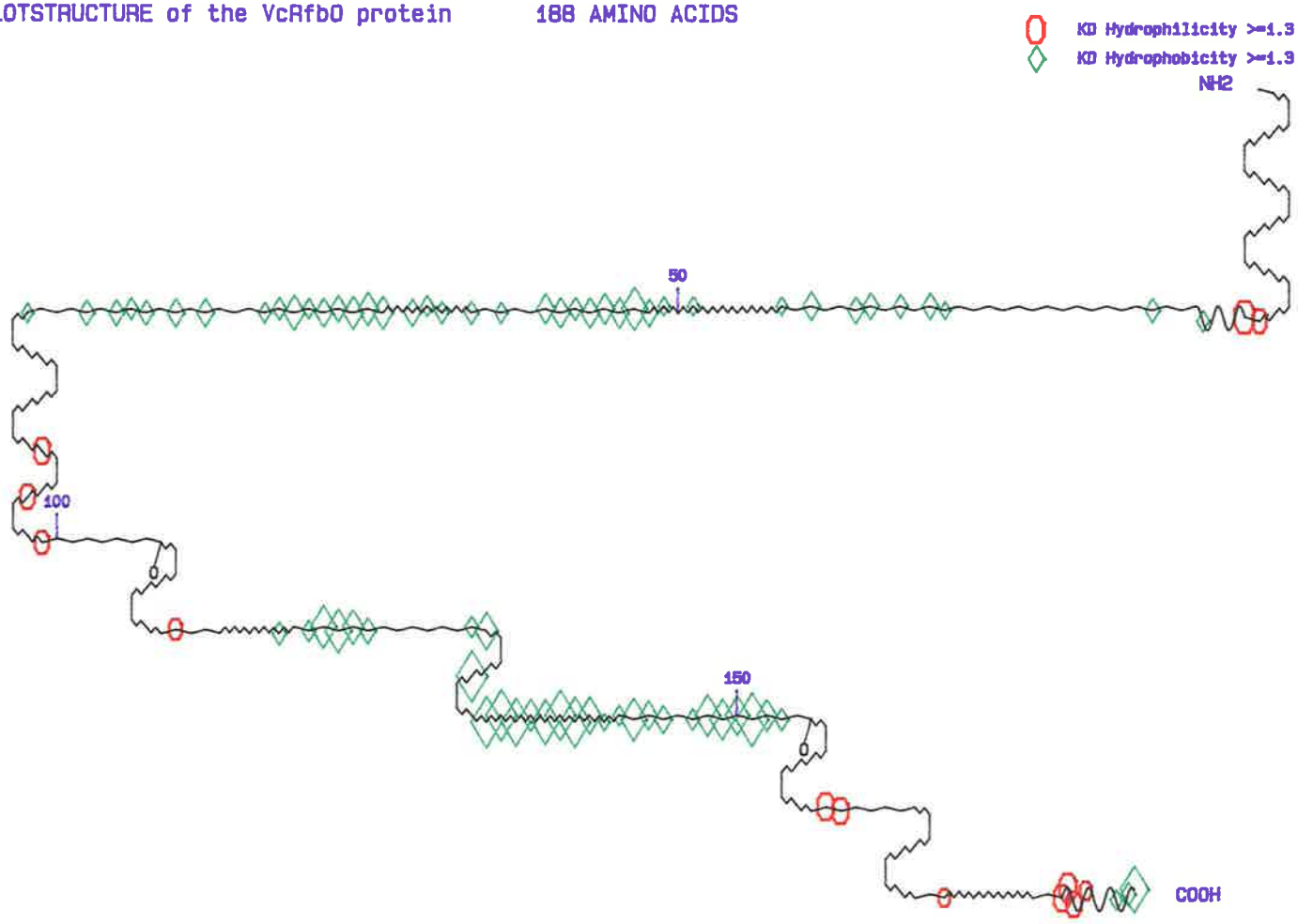


Figure 5.21 Secondary structure prediction and hydrophobicity of VcRfbP.

The protein sequence of VcRfbP was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix:

random coil:

β sheet: 

β turn: where the structure rotates by 90°

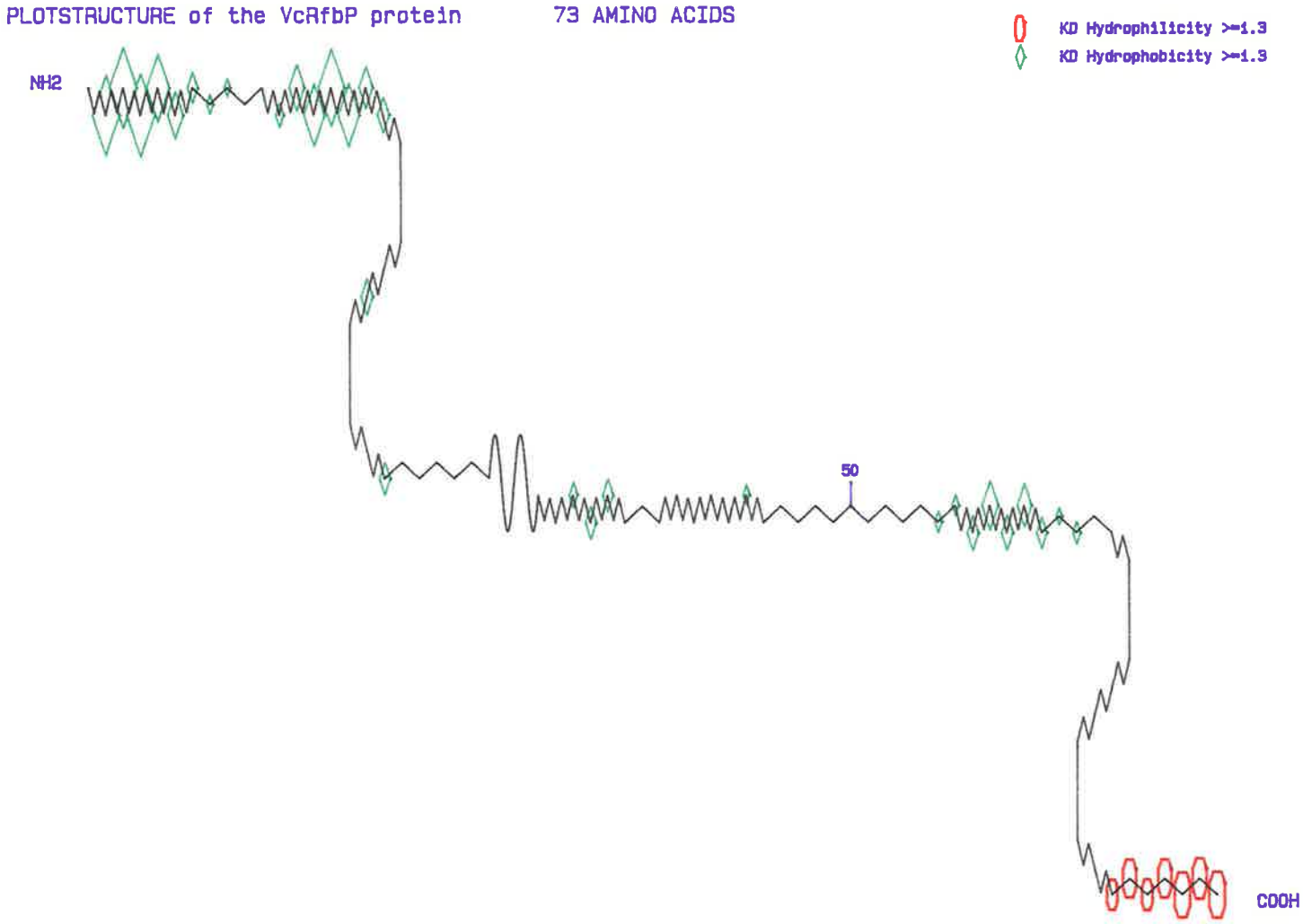

PLOTSTRUCTURE of the VcRfbP protein
73 AMINO ACIDS
KD Hydrophilicity >=1.3
KD Hydrophobicity >=1.3
NH2
50
COOH

Figure 5.22 Secondary structure prediction and hydrophobicity of VcRfbQ.

The protein sequence of VcRfbQ was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix:

random coil:

β sheet: 

β turn: where the structure rotates by 90°

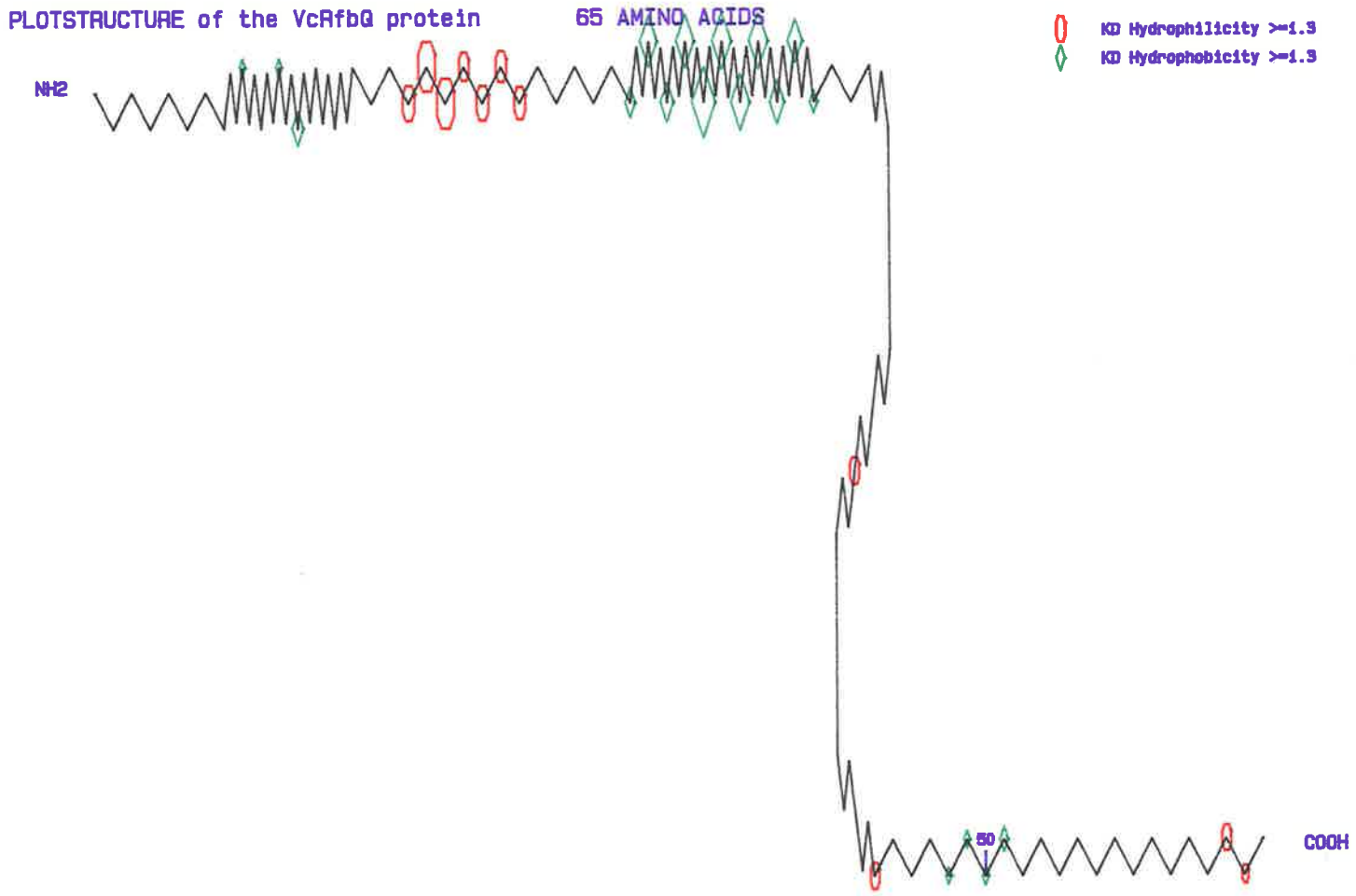

PLOTSTRUCTURE of the VcAfbQ protein
65 AMINO ACIDS
KD Hydrophilicity >=1.9
KD Hydrophobicity >=1.9
NH2
50
COOH

Figure 5.23 Secondary structure prediction and hydrophobicity of VcRfbR.

The protein sequence of VcRfbR was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix: 

random coil:

β sheet: 

β turn: where the structure rotates by 90°

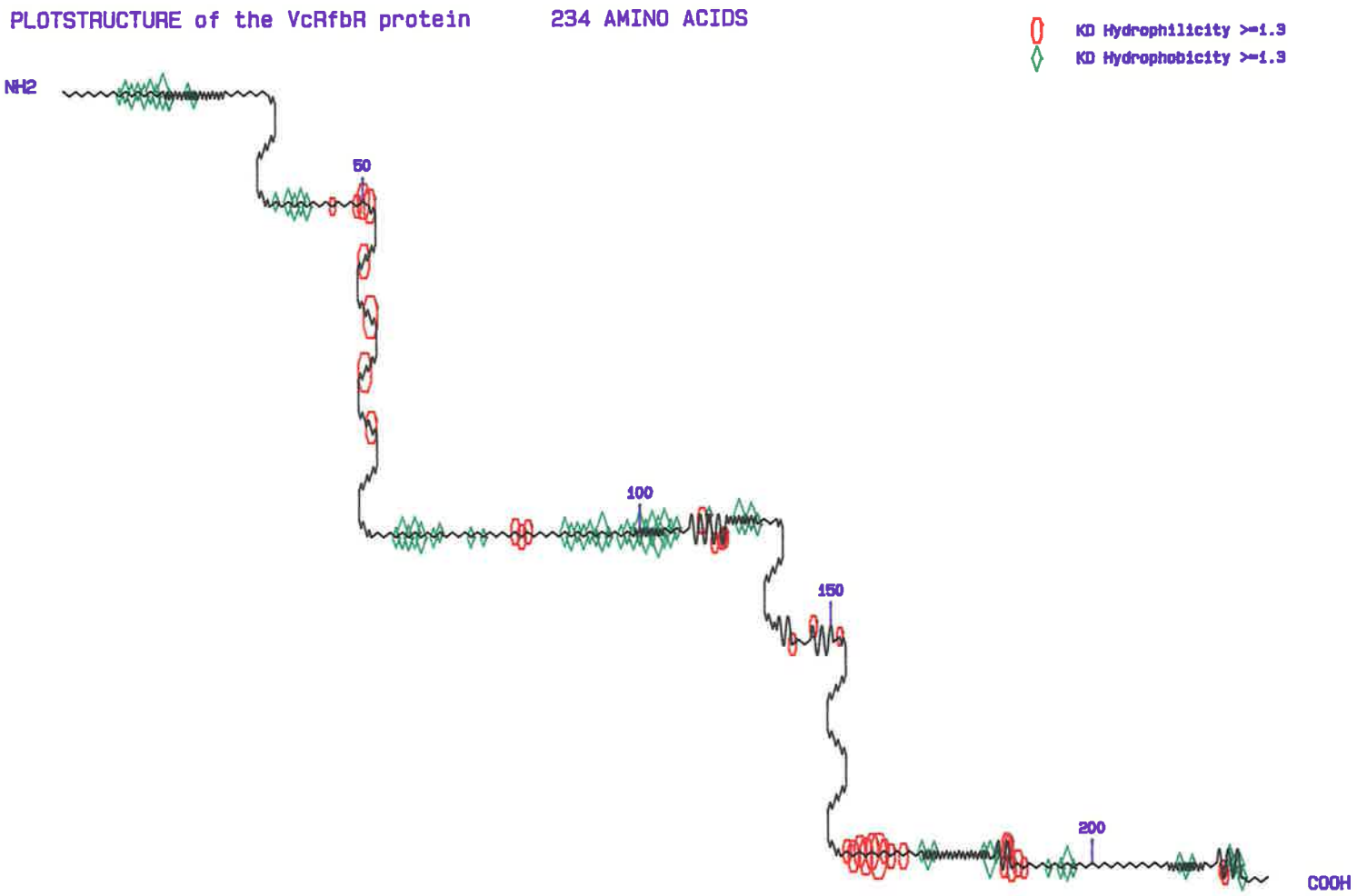

PLOTSTRUCTURE of the VcRfbR protein
234 AMINO ACIDS
KD Hydrophilicity >=1.9
KD Hydrophobicity >=1.9
NH2
50
100
150
200
COOH

Figure 5.24 Secondary structure prediction and hydrophobicity of VcRfbS.

The protein sequence of VcRfbS was subjected to the peptide structure and plotstructure programmes from the GCG package (Deveraux *et al.*, 1988), using the algorithim of Chou and Fasman (1978). The green diamonds represent regions of hydrophobicity, whereas the red ovals represent regions of hydrophilicity, both calculated according to Kyte and Doolittle (1982), using a threshold value of 1.3. The secondary structure symbols are:

α helix: 

random coil:

β sheet: 

β turn: where the structure rotates by 90°

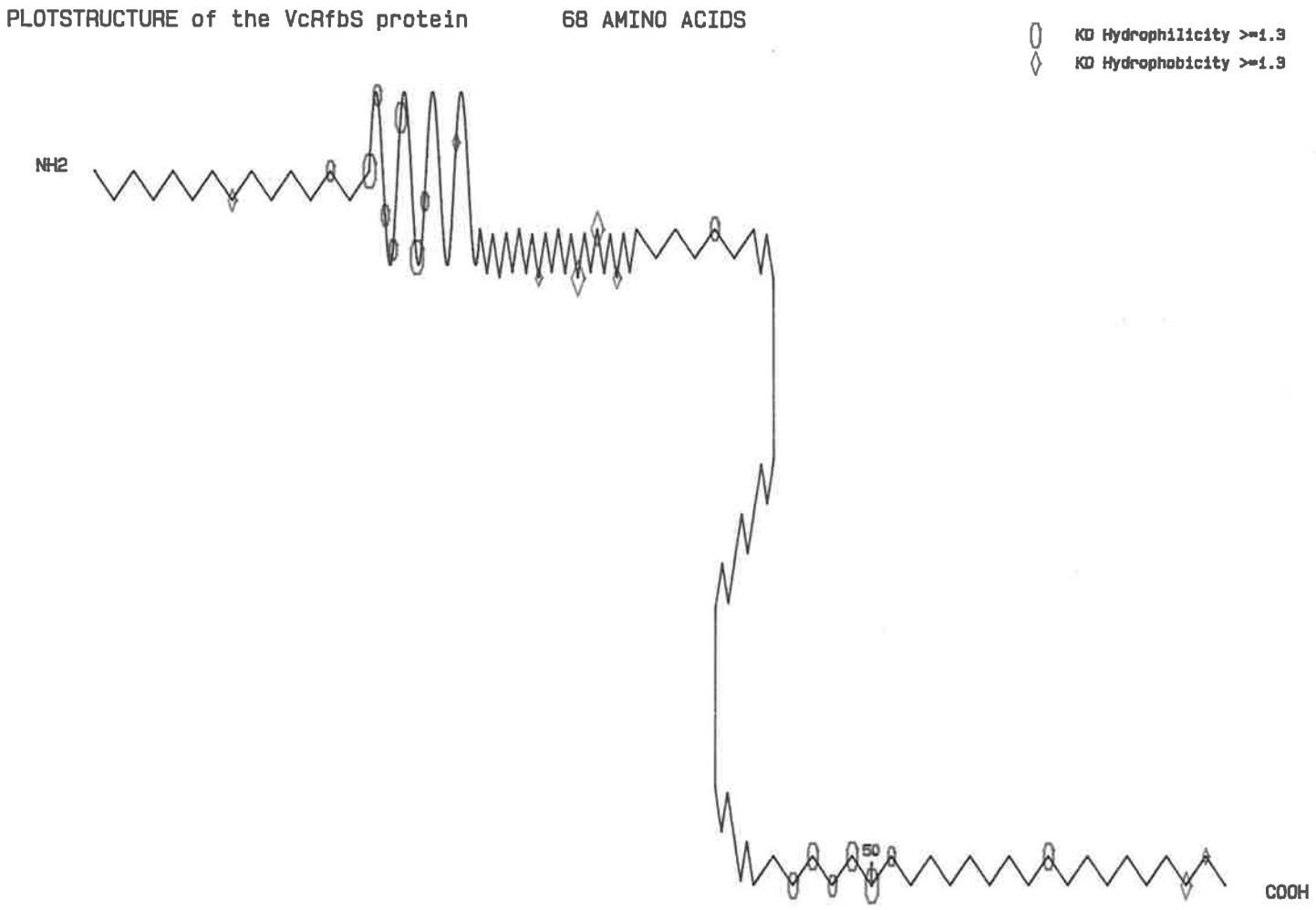

PLOTSTRUCTURE of the VcRfbS protein
68 AMINO ACIDS
KD Hydrophilicity >=1.3
KD Hydrophobicity >=1.3
NH2
50
COOH

Figure 5.25 Homology studies of VcRfbB and StRfbK

The amino sequence of VcRfbB was used to search data in the SWISSPROT bank for homologous sequences. The *S. typhimurium* RfbK sequence was found by this method to show significant homology to VcRfbB. These two proteins were then analyzed by the CLUSTAL programme for maximum alignment (Higgins and Sharp, 1988). Amino acids showing identity between proteins have been bolded. Gap fixed - is the penalty used for each gap, and Gap vary. - is the penalty used for every item in each gap. (*) denotes a match across all sequences, and (.) denotes a conservative substitution.

```
Gap fixed =  10 Gap vary. =  10
 * :=>  match across all seqs.
 . :=>  conservative substitutions

VcRfbB          MKELT-----------CFKAYDIRGQLGSELDNEIAYRIGRSYGQFLKSENDADKTVVVG
StRfbK          MNVVNNSRDVIYSSGIVFGTSGARG-LVKDFTPQVCAAFTVSFVAVMQEHFSFD-TVALA
                *. ..            * . . ** * .... ..   .. *. . .... . * **...

VcRfbB          GDVRLTSEALKQALANGLMDAGINVIDIGVTGTEEIYFATFYLGVDGGIEVTASHNPMDY
StRfbK          IDNRPSSYGMAQACAAALADKGVNCIFYGVVPTPALAFQSMSDNMPA-IMVTGSHIPFER
                 * * .* .. ** *..* * *.* *  **. * .. *...  .. . * **.** *..

VcRfbB          NGMKLVREGSKPISGDTG------------LRE-------------------------
StRfbK          NGLKFYRPDGEITKHDEAAILSVEDTCSHLELKELIVSEMAAVNYISRYTSLFSTPFLKN
                **.*. * ... .. *..             *.*

VcRfbB          -------------------------IQALAEKNEFMDVE--------------------
StRfbK          KRIGIYEHSSAGRDLYKPLFIALGAEVVSLGRSDNFVPIDTEAVSKEDREKARSWAKEFD
                                         . .*. ...*. ..

VCRFBB          ----------------------VKGN------------------------------YKK
STRFBK          LDAIFSTDGDGDRPLIADEAGEWLRGDILGLLCSLALDAEAVAIPVSCNSIISSGRFFKH
                                    ..*.                                .*.

VcRfbB          VSLL---PEYV-----------DHLISYIT--------------------PAK---IKPM
StRfbK          VKLTKIGSPYVIEAFNELSRSYSRIVGFEANGGFLLGSDICINEQNLHALPTRDAVLPAI
                *.*    . **           ..... .                      *..   . ..

VcRfbB          KLVINSGNGAAGHVIDELEKRFIELSIPLEIIKVHHEED-GNFPNGIPNPLLPECRADTA
StRfbK          MLLYKSRNTSISALVNELPTRYTH-SDRLQGITTDKSQSLISMGRENLSNLLSYIGLENE
                .*. .* *.. . ...** .*... * .*. *.......  .. ..  . **.    ...

VcRfbB          NAVKEHKAD-MGIALMVTLIAAF----------C------LMKMAILLKVTTS-------
StRfbK          GAISTDMTDGMRITLRDGCIVHLRASGNAPELRCYAEANLLNRAQDLVNTTLANIKKRCL
                .*......* * *.*. . *. .          *      * . . *...* .

VcRfbB          -
StRfbK          L
```

homology extends over the N terminal and C terminal ends of the proteins. There is a difference in size between these two proteins as VcRfbB is 269 aa and StRfbK is 477 aa. This results in the gaps seen in the alignment (Figure 5.25), however, it is still believed these proteins perform a similar function in their respective organisms (see 6.3).

Figure 5.26 reveals homologous regions between VcRfbD, GalX (UDP-glucose-4-epimerase from *Streptomyces lividians*) and StRfbE (CDP-tyvelose epimerase from *Salmonella typhimurium*) (Wyk and Reeves, 1989; Verma and Reeves, 1989). The discovery of these homologies, warranted further studies with other *S. typhimurium rfb* protein sequences to be performed: StRfbB (dTDP-D-glucose-3,5 epimerase), StRfbG (CDP-glucose 4,6-dehydratase), and StRfbJ (abequose synthetase) which have been shown previously to have homology with each other (Wyk and Reeves, 1989; Verma and Reeves, 1989) and with GalE (UDP-glucose-4-epimerase from *S. typhimurium*) (shown in Figure 5.27). The region of homology near the N-terminus seen in this composite figure shows the motif GxxGxxG, which is very similar to the NAD binding domain (GxGxxG) (Verma and Reeves, 1989; Wierenga *et al.*, 1986; Wyk and Reeves, 1989).

A "fingerprint" for ADP-binding folds using five dehydrogenase protein structures has been determined (Wierenga *et al.*, 1986). This fingerprint (comprised of between 29 and 31 residues) describes the type of amino acid that consistently occurrs at a specific position within the ADP binding βαβ fold. The "core" fingerprint demands glycine residues at position 6, 8, and 11, and that an acidic residue should occur at the end of the fingerprint (Wierenga *et al.*, 1986). Although some of these proteins (StRfbG, StRfbJ, StRfbS, and VcRfbD) do not conform completely to this motif, enough similarities exist for this to be a potential NAD binding domain as postulated by Wyk and Reeves (1989) and Verma and Reeves (1989). This hypothesis is also confirmed by the secondary structure of this region.

Figure 5.28 reveals homology between VcRfbE and two proteins, DegT in *Bacillus stearothermophilus*, (Takagi *et al.*, 1990) and ErbS, also known as EryC in

Figure 5.26 Homology studies of VcRfbD with GalX and StRfbE.

The amino sequence of VcRfbD was used to search data in the SWISSPROT bank for homologous sequences. The galactose proteins from *Streptomyces lividians* and the *S. typhimurium rfb* gene product StRfbE were found to have considerable homology. These proteins were then analyzed by the programme CLUSTAL for maximum alignment (Higgins and Sharp, 1988). Amino acids showing identity between proteins have been bolded. Gap fixed - is the penalty used for each gap, and Gap vary. - is the penalty used for every item in each gap. (*) denotes a match across all sequences, and (.) denotes a conservative substitution.

```
Gap fixed =  10 Gap vary. =  10
 * :=>  match across all seqs.
 . :=>  conservative substitutions

VcRfbD          MNKKVALITGITGQDGSYLAEFLLEKGYEV---HGIKR--RSSLFNTQRVDHLYKDPHEE
GalX            MSGKY-LVTGGAGYVGSVVAQHLVEAGNEVVVLHNLSTGFREVCRRVPRSSRRHPGRRQV
StRfbE          MK---LLITGGCGFLGSNLASFALSQGIDLIVFDNLSR--KGATDNLHWLSSL-------
                *.    *.**  *  ** .*   .. * ..   ....   ..   . .. .

VcRfbD          DVNFKLHYGDLTDSSNLTRILAEVQPDEVYNLGAQSHVAVSFQSPEYTADVDAIGTLRLL
GalX            RGRLSF-------------------DGVLHFAAFSQVGESVVKPEKYWDNNVGGTMALL
StRfbE          -GNFEFVHGDIRNKNDVTRLITKYMPDSCFHLAGQVAMTTSIDNPCMDFEINVGGTLNLL
                  ....                    *.  ....   .. *  .*    . .. **. **

VcRfbD          EAIRFLGLTKKTKFYQASTSELYGLVQEIPQKETTPFYPRSPYAVAKMYAYWITINYRES
GalX            EAMRGAGVRR---LVFSSTAATYGEPEQVPIVESAPTRPTNPYGASKLAVDHMITGEAAA
StRfbE          EAVR--QYNSNCNIIYSSTNKVYGDLEQYKYNETETRY--TCVDKPNGYDESTQLDFHSP
                **.*     .   .  .**.  **  ..    *...    .  . ..         .  ..

VcRfbD          YGIYACNGILFNHE--SPRRGETFVTRKITRGMA-NIAQGLEK--CLFMGNLDALRDWGH
GalX            HGLGAVSVPYFNVAGAYGEYGERHDPESHLIPLVLQVAQGRREAISVYGDDYPTPDRPVC
StRfbE          YGCSKGAADQYMLD--YAR---IFGLNTVVFRHS-SMYGGRQ-----FATYDQGWVCWFC
                .*    .   .  .              .       .  *       . .   .

VcRfbD          AKDYVKMQWMMLQQDEPRDFVIATGVQYSVREFIDMSARELGIELEFVGKGVDEKAVVKS
GalX            ATTSTSPTW-------PRP---------TCWPCAAAPGEHLICNLG-NGNGFSVREVVET
StRfbE          QKA------VEIKNGINKPFTISGNGK-QVRDVLHAEDM---ISLYFTA--------LAN
                ...             .           .     .     .*   .        . .

VcRfbD          VIGTKAPAI-KVGDIIVAVDPAYFRPAEVETLLGD--------PSLAKKELGWVPEITLQ
GalX            VRRVTGHPIPEIMAPRRGRDPAVLVAS----------------AGTAREKLGWNPS----
StRfbE          VSKIRGNAF-NIGGTIVNS----LSLLELFKLLEDYCNIDMRFTNLPVRKV--ISVFLLQ
                *  . . .. .. .   .     .                      .. .  ..   .

VcRfbD          QMVSEMVASDLEQAQSHALLKKHG------YNVNV---SVE
GalX            --RADLAIVSDAWELPQR------------------RAGQ
StRfbE          ILKKSLMQLTGARKSRQKMVSRKCMIGLVLYDVSARESGHN
                    ..   . .   ..                    . .
```

Figure 5.27 Homology studies of VcRfbD with GalX, GalE, StRfbE, StRfbB, StRfbG, StRfbS, StRfbJ.

The amino sequence of VcRfbD was used to search data in the SWISSPROT bank for homologous sequences. The galactose genes GalE and GalX, and the *S. typhimurium rfb* genes *rfbE, rfbB, rfbG, rfbS, rfbJ* were all found to have considerable homology. These eight proteins were then analyzed by the programme CLUSTAL for maximum alignment (Higgins and Sharp, 1988). Amino acids showing identity between several proteins have been bolded. Gap fixed - is the penalty used for each gap, and Gap vary. - is the penalty used for every item in each gaps. (*) denotes a match across all sequences, and (.) denotes a conservative substitution.

```
Gap fixed =  10 Gap vary. =  10
 * :=>  match across all seqs.
 . :=>  conservative substitutions

VcRfbD          M-------NKKVALITGITGQDGSYLAEFLLE------KGYEVHGIKRRSSLFNTQRVDH
GalX            MSG------K--YLVTGGAGYVGSVVAQHLVEAGNEVVV-LHNLSTGFREVCRRVPRSSR
GalE            M---R-------VLVTGGSGYIGSHTCVQLLQNGHDVII-LDNLCNSKRSVLPVIERLGG
StRfbE          M--------K--LLITGGCGFLGSNLASFALSQGIDLIV-FDNL--SRKGATDNLHWLSS
StRfbB          M--------K--ILITGGAGFIGSAVVRHIIKNTQDTVVNIDKL--TYAGNLESLSDISE
StRfbG          MIDKNFWQGK-RVFVTGHTGFKGSWLSLWLTEMGA-IVKGYALDAPTV-PSLFEIVRLND
StRfbS          M--------K--ILIMGAFGFLGSRLTSYFESRH-TVI-GLARK-RNNEATINNIIYTTE
StRfbJ          MTFL-----KEYVIVSGASGFIGKHLLEALKKSGISVVAITRDVIKNNSNALANVRWCS-
                *        *    . *  *  *.                                  .

VcRfbD          LYKDPHEEDVNFKLHYGDLTDSSNLTRILAEVQPDEVYNLGAQSHVAVSFQSPEYTADVD
GalX            RHP---------G-----------RRQVRGRLSFDGVLHFAAFSQVGESVVKPEKYWDNN
GalE            KHP---------TFVEGDIRNEALMTEILHDHAIDTVIHFAGLKAVGESVQKPLEYYDNN
StRfbE          LGN--------FEFVHGDIRNKNDVTRLITKYMPDSCFHLAGQVAMTTSIDNPCMDFEIN
StRfbB          SNR--------YNFEHADICDSAEITRIFEQYQPDAVMHLAAESHVDRSITGPAAFIETN
StRfbG          LMES----------HIGDIRDFEKLRNSIAEFKPEIVFHMAAQPLVRLSYEQPIETYSTN
StRfbS          NNW--------IEKILE---FEPNII--IN----------TIACYGRHNEPATALIESN
StRfbJ          --WD------NIELLVEELSIDSALIGIIH---------LA--TEYGHKTSSLINIEDAN
                                                                  . .

VcRfbD          AIGTLRLLEAIRFL------GLTKKTKFYQASTSELYGLVQEI-------------P--
GalX            VGGTMALLEAMR------GAGVR---RLVFSSTAATYGEPEQVPIVES--------APT-
GalE            VNGTLRLISAMR------AANVK---NFIFSSSATVYGDQPKIPYVES--------FPTG
StRfbE          VGGTLNLLEAVRQY--------NSNCNIIYSSTNKVYGDLEQYKYNETETRYTCVDKPNG
StRfbB          IVGTYALLEVARKYWSALGEDKKNNFRFHHISTDEVYGDLPH-----PDEVENSVTLPL-
StRfbG          VMGTVHLLETVKQV------GNIKAVV-NITSDKC-YDNREWV--------------WG
StRfbS          ILMPIRVLE-SISSLNAV-------------------------FINCGTSLP-------
StRfbJ          VIKPLKLLDLAIKYRADI-------------------------FLNTDSFFA-------
                    .  ...

VcRfbD          QKETTPFYPRSPYAVAKMYAYWITINYRESY---GIYACNGILFNHESPRRGETFVTRKI
GalX            -------RPTNPYGASKLAVDHMITG-EAA-----AHGLGAVSVPYFNVAGAY--GEYGE
GalE            -------TPQSPYGKSKLMVEQILTDLQKA-----QPDWSIALLRYFNPVGAHPSGDMGE
StRfbE          YDESTQLDFHSPYGCSKGAADQYMLDYARI------FGLNTVVFRHSSMYGGRQFATYDQ
StRfbB          FTETTAYAPSSPYSASKASSDHLVRAWRRT------YGLPTIVTNCSNNYGPYHFP---E
StRfbG          YRENEPMGGYDPYSNSKGCAELVASAFRNSFFNPANYEQHGVGLASVRAGNVIGGGDWAK
StRfbS          -PNTSLY----AYTL--QLANELAAAIIDK------VCGKYIELKLEHFYGAFP----GD
StRfbJ          -KKDFNYQHMRPYIITKRHFDEIGHYYANM------HDISFVNMRLEHVYGPGD----GE
                          .*

VcRfbD          TRGMANIAQGLEKCLFMGNLDALRDWGHAKDYVKMQWMMLQQDEPRDFVIATGVQYSVRE
GalX            -R------HDPESHLIPLVLQVAQ--GRREAISVYGDDYPTPDRPVCATTSTSPTWPRPT
GalE            DP------QGIPNNLMPYIAQVAV--GRRDSLAIFGNDYPTEDGTGVRDYIHVMDLADGH
StRfbE          ---------GWVGWFCQKAVEIKN--GINKPFTISGNGKQVRDVLHAEDMISLYFTALAN
StRfbB          ---------KLIPLVILNALE-------GKPLPIYGKGDQIRDWLYVEDHARALHMVVTE
StRfbG          DRLIPDILRSFEN----NQQVIIRN-----PYSIRPWQHVL-EPLSGYIVVAQRLYT---
StRfbS          DKFTSMV--------------IRRCLSNQP--VKLTSGLxxRDFLYIKNLLTAFDCIISN
StRfbJ          NKFIPYI--------------ID-CLNKKQSCVKCTTGEQIRDFIFVDDVVNAYLTILEN
                                                              .
```

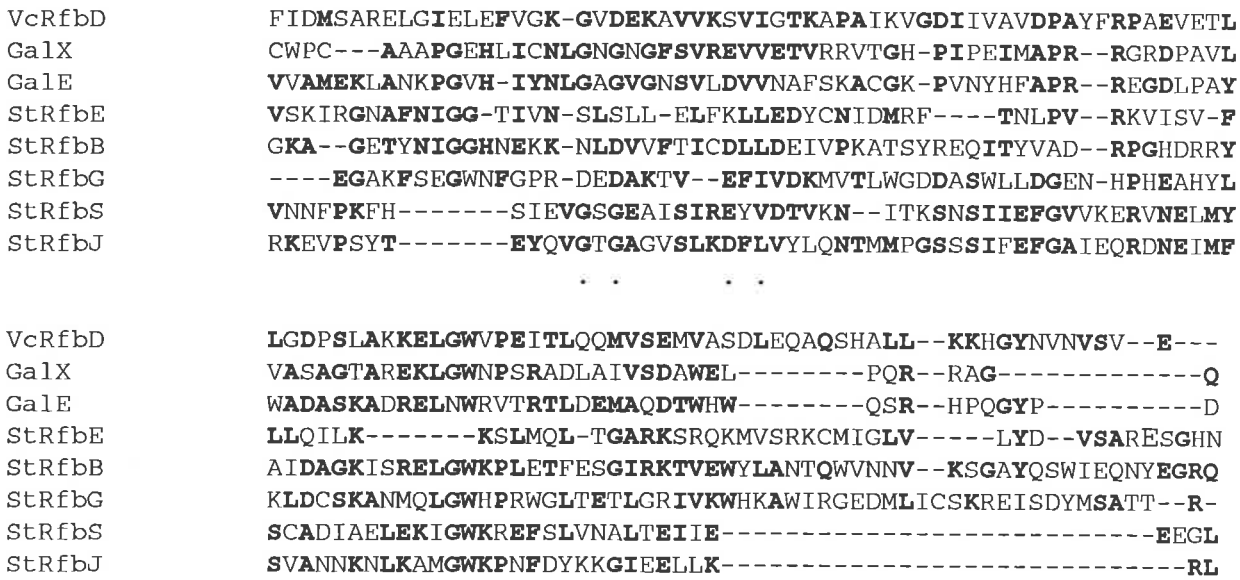

VcRfbD FIDMSARELGIELEFVGK-GVDEKAVVKSVIGTKAPAIKVGDIIVAVDPAYFRPAEVETL
GalX CWPC---AAAPGEHLICNLGNGNGFSVREVVETVRRVTGH-PIPEIMAPR--RGRDPAVL
GalE VVAMEKLANKPGVH-IYNLGAGVGNSVLDVVNAFSKACGK-PVNYHFAPR--REGDLPAY
StRfbE VSKIRGNAFNIGG-TIVN-SLSLL-ELFKLLEDYCNIDMRF----TNLPV--RKVISV-F
StRfbB GKA--GETYNIGGHNEKK-NLDVVFTICDLLDEIVPKATSYREQITYVAD--RPGHDRRY
StRfbG ----EGAKFSEGWNFGPR-DEDAKTV--EFIVDKMVTLWGDDASWLLDGEN-HPHEAHYL
StRfbS VNNFPKFH-------SIEVGSGEAISIREYVDTVKN--ITKSNSIIEFGVVKERVNELMY
StRfbJ RKEVPSYT-------EYQVGTGAGVSLKDFLVYLQNTMMPGSSSIFEFGAIEQRDNEIMF
VcRfbD LGDPSLAKKELGWVPEITLQQMVSEMVASDLEQAQSHALL--KKHGYNVNVSV--E---
GalX VASAGTAREKLGWNPSRADLAIVSDAWEL--------PQR--RAG----------Q
GalE WADASKADRELNWRVTRTLDEMAQDTWHW--------QSR--HPQGYP----------D
StRfbE LLQILK-------KSLMQL-TGARKSRQKMVSRKCMIGLV-----LYD--VSAREsGHN
StRfbB AIDAGKISRELGWKPLETFESGIRKTVEWYLANTQWVNNV--KSGAYQSWIEQNYEGRQ
StRfbG KLDCSKANMQLGWHPRWGLTETLGRIVKWHKAWIRGEDMLICSKREISDYMSATT--R-
StRfbS SCADIAELEKIGWKREFSLVNALTEIIE----------EEGL
StRfbJ SVANNKNLKAMGWKPNFDYKKGIEELLK----------RL

Figure 5.28 Homology studies of VcRfbE with DegT and ErbS.

The amino sequence of VcRfbE was used to search the SWISSPROT protein data base for homologous sequences. The DegT protein (in *Bacillus stearothermophilus*) and the ErbS protein (from *Saccharopolyspora erythrawa*) were found to have significant homology. These three proteins were analyzed by the programme CLUSTAL for maximum alignment (Higgins and Sharp, 1988). Amino acids showing identity between proteins have been bolded. Gap fixed - is the penalty used for each gap, and Gap vary. - is the penalty used for every item in each gap. (*) denotes a match across all sequences, and (.) denotes a conservative substitution.

```
Gap fixed =  10 Gap vary. =  10
 * :=>  match across all seqs.
 . :=>  conservative substitutions

VcRfbE          M-IPVYEPS-----LDGNERKYLNDCIDSGWVSSRGKYIDRFETEFAEFLKVKHATTVSN
BsDegT          MNVPMLDLSEQYEQLKPEIMRVLDEVMRSSRFI-LGDYVKKLEADIAAYSRAKHGIGCGN
SeErbS          MDVPFLDLQAAYLELRSDIDQACRRVLGSGWYL-HGPENEAFEAEFAAYCENAHCVTVGS
                * .*  .       *  . .     . *..   *   . .*...*..    * .. ..

VcRfbE          GTVALHLAMSALGITQGDEVIVPTFTYVASVNTIVQCGALPVFAEIEGESLQVSVEDVKR
BsDegT          GSDAIHIALQAAGVGPGDEVITTAFTFFATAGSIARAGAKPVFVDIDPVTFNIDPAQVEA
SeErbS          GCDALELSLVALGVGQGDEVIVPSHTFIATWLGVP-VGAVPVPVEPEGVSHTLDPALVEQ
                *  *..... * *...*****... *. *.  ..   ** ** .. .  .  .. . *.

VcRfbE          KINKKTKAVMAVHIYGQACDIQSLRDLCDEHGLYLIEDCAEAIGTAVNGKKVGTFGDVST
BsDegT          AVTEKTKAIIPVHLYGQMADMEAIAAIAKRHGLVVIEDAAQAIGAKYNGKCVGELGTAAT
SeErbS          AITPRTAAILPVHLYGHPADLDALRAIADRHGLALVEDVAQAVGARHRGHRVGAGSNAAA
                 .. .* *...**.**.  *.... .. . *** ..** *.*.*.  .*. **. .....

VcRfbE          FSFFGNKTITS-GEGGMVVSNSDIIIDKCLRLKNQGVVAGKRYWHDLVAYNYRMTNLCAA
BsDegt          YSFFPTKNLGAYGDGGMIITNDDELAEKCRVIRVHG--SKPKYYHHVLGYNSRLDEMQAA
SeErbs          FSFYPGKNLGALGDGGAVVTTDPALAERIRLLRNYG--SKQKYVHEVRGTNARLDELQAA
                .**. .*.... *.** .....  . ..   ..  *  .  .* *.. . * *.... **

VcRfbE          IGVAQLERVDKIIKAKRDIAEIYRSELAGLPMQ--V-HKESNGTFHSYWLTSIILDQEFE
BsDegT          ILSVKFPHLDRWTEQRRKHAATYTRLLEEAVGDLVVTPKEVDGRYH-VFHQYTIRAPK--
SeErbS          VLRVKLRHLDDWNARRTTLAQHYQTELKDVPG---ITLPETHPWADSAWHLFVLRCEN--
                .  ... ..*     . . *. *   * .       .   * .   .  .   ..   .

VcRfbE          VHRDGLMTFLENNDIESRPFF-YPAHTLPMYEHLA-EKTAFPLSNSYSHRGINLPSWPGL
BsDegT          --RDELQAFLKEQGIATMVYYPLPLHLQPVFASLGYKEGQLPEAEKAAKEALSLPMFPEL
SeErbS          --RDHLQRHLTDAGVQTLIHYPTPVHLSPAYADLGLPPGSFPVAESLAGEVLSLPIGPHL
                  ** *   *.......   .  * *  * .. *.   . .* ... .   ..**  * *

VcRfbE          CDDQVKEICNCIKNYFNCI-
BsDegT          KEEQQQYVVEKIAEFYRHFA
SeErbS          SREAADHVIATL----KAGA
                  .. . . . .    .
```

Saccharopolyspora erythraea (Dhillon *et al.*, 1989). The *degT* gene has been postulated to be a regulatory gene enhancing production of an extracellular alkaline protease from *Bacillus subtilis*, as it can induce pleiotropic phenotypes (Takagi *et al.*, 1990).

Similar to VcRfbE, DegT possesses a highly hydrophobic region between aa 55 and aa 100, which may be located in the cytoplasmic membrane. Interestingly, it has been postulated that this region may, in the case of DegT, function as a membrane sensor protein for enviromental stimuli. Between aa 160 and aa 179 of DegT is a consensus sequence for DNA binding proteins (Pabo and Sauer, 1984), however, this region is not homologous to either VcRfbE, nor to ErbS.

The *erbS* (*eryC*$_1$) gene of *S. erythraea* may encode an enzyme which is involved in the formation of the deoxy amino sugar desosamine (Dillon *et al.*, 1989). This sugar is a precursor of the antibiotic erythromycin. Although Dillon *et al.* (1989) did not predict the location of the ErbS, comparisons of the predicted secondary structure of ErbS (using the GCG package) have revealed many similarities to VcRfbE. ErbS also appears to be located in the cytoplasmic membrane, due to many regions that may function as transmembrane domains.

Analysis of this sequence for functional domains has not been undertaken, although, the function of this enzyme may bear a direct relationship to the postulated function of VcRfbE. Perosamine, the backbone of the VcOAg is a dideoxy amino sugar similar to desosamine. Therefore if VcRfbE performs a similar function to ErbS, it may be involved in perosamine synthesis.

5.3 Summary and discussion

This chapter analyzes the DNA sequence of segments from the *V. cholerae rfb* region. These sequences have been compiled and analyzed using DNASIS to study the G+C content, predict open reading frames and provide codon usage data. The open

reading fames predicted were then analyzed by a variety of computer programmes to generate hydrophobicity plots, predict secondary structure, and show translational coupling along the length of the *rfb* region.

Analyses of the DNA sequence has enabled identification of putative Shine-Dalgarno sequences. The spatial configuration of the *rfb* genes has shown the possibility of translational coupling between various genes. Since multiple promoter sequences within the 20 kb *rfb* region have not been identified, these data corresponds to the *rfb* region being composed of gene clusters.

Analysis of the G+C content of the whole 20 kb region revealed a trend, such that blocks of genes could be compiled having the same relative G+C content. Comparison of these blocks using homology studies has revealed the likelihood that they correspond to functional segments of the *rfb* operon (see Figure 6.3).

Due to the absence of signal sequences, the analyzed proteins have been classified into either cytoplasmic proteins or proteins located in the cytoplasmic membrane, based on hydrophobicity plots and the predicted secondary structure.

Further analyses confirming the open reading frames with the proteins seen in the *in vitro* transcription/translation system, and the cellular location of these proteins need to be performed.

CHAPTER 6

DISCUSSION

6.1 Introduction

The composition of the LPS of *V. cholerae* O1 has been characterized at the chemical level, but limited structures for both the O-antigen and the core oligosaccharide have been defined (Hisatsune and Kondo, 1980, Hisatsune *et al.*, 1978, 1985; Kenne *et al.*, 1979; Redmond, 1975, 1978, 1979). In 1970, Bhaskaran mapped the *oag* locus encoding serotype specificity between *ilv-1* and *arg-2*. A breakthrough in analysis of this region was the cloning of a chromosomal DNA fragment encoding the appropriate *V. cholerae* serotype in *E. coli* (Manning *et al.*, 1986). This *oag* locus was subsequently shown to be identical to the *rfb* locus (Ward and Manning, 1989).

This study has centred on characterizing the molecular basis for *Vibrio cholerae* O-antigen (VcOag) expression and the organization of the genes involved. DNA sequence analyses and comparison with proteins in the databases has enabled biosynthetic pathways for the sugar residues to be predicted.

6.2 Characterization of the 20 kb *Sac*I *rfb* region

Expression of VcOAg from *E. coli* strains harbouring pEVX6 or pEVX7 was found to be inconsistent due to plasmid instability. To alleviate this problem, the 20 kb *Sac*I fragment was cloned into pOmpV500, producing pRMB1 and pRMB2 (differing only upon orientation of the DNA insert). These two sets of clones, one pUC18 based (pEVX6, pEVX7) and the other pBR322 derived (pRMB1, pRMB2) were used: in analyses to determine the minimal coding region required for VcOAg expression; in

construction of a comprehensive restriction map; in transposon mutagenesis, to attempt to define and characterize regions within the *Sac*I fragment; and in preliminary delineation of regions involved in serotype specificty.

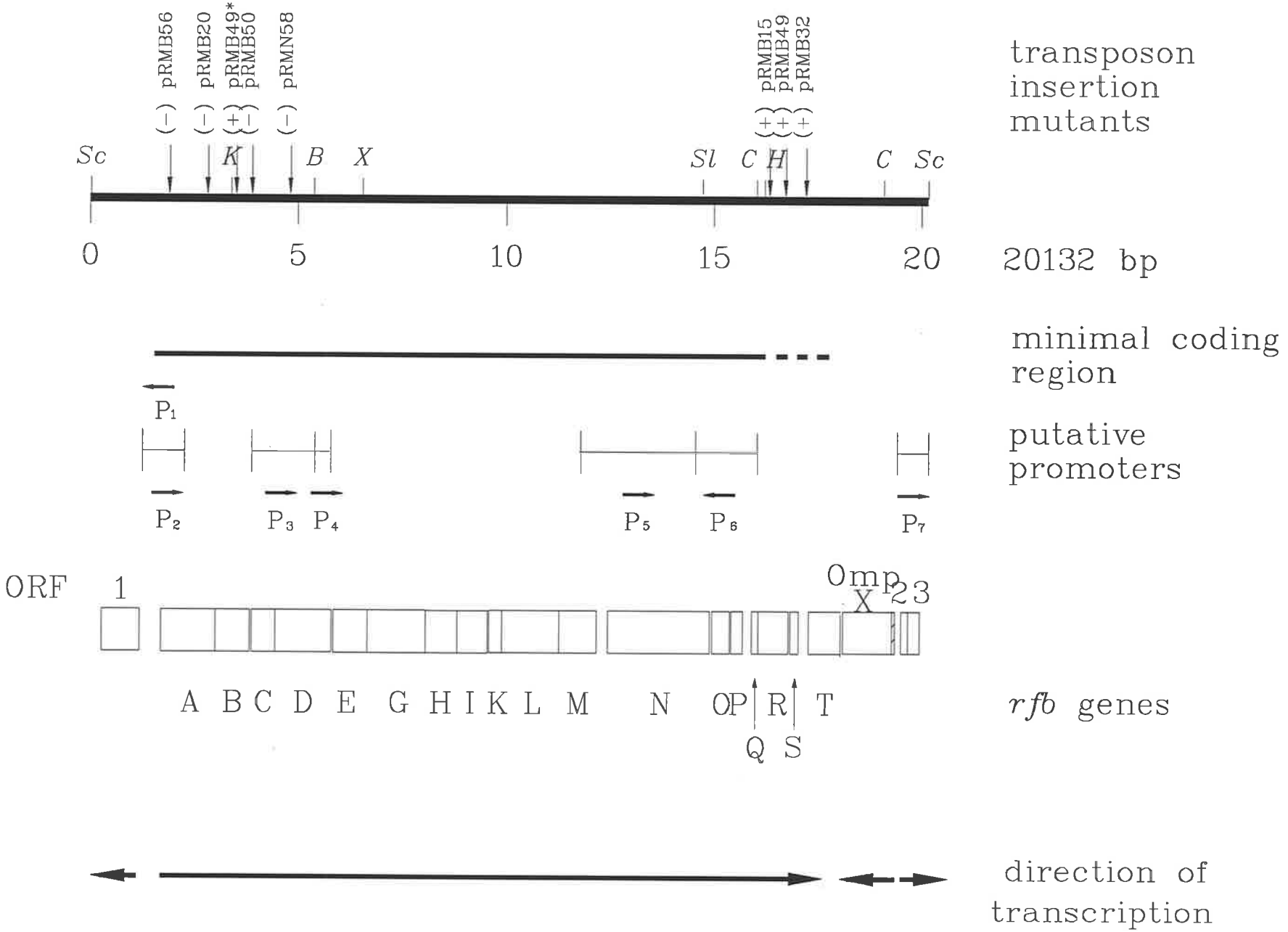
The genes encoding the K1 capsular antigen of *E. coli* have been analyzed, revealing a requirement for 15 kb of contiguous DNA (Timmis *et al.*, 1985). This region was subjected to transposon and deletion analyses, revealing 3 functional regions (Boulnois *et al.*, 1987). Likewise, Sturm *et al.* (1986b) analyzed the genetic and functional organization of the *Shigella dysenteriae rfb* gene cluster using insertion and deletion mutants. These results encouraged the use of Tn*1725* on the *V. cholerae rfb* region. However, due to the restricted ability to isolate insertions within the 20 kb region, it could not be divided into functional domains. Seven independent transposon insertions were mapped by restriction mapping, and five confirmed by DNA sequence analysis. These insertions were located in *rfbA*, *rfbD*, *rfbR*, and the intergenic region between *rfbS* and *rfbT*. *rfbA* and *rfbD* are early genes in the *rfb* operon, *rfbR* appears to be unnecessary for VcOAg production in *E. coli*, and *rfbT* is involved in the Ogawa serotype specificity (U.H. Stroeher, personal communication). These insertions in conjunction with complementation, HIA and silver stained PAGE analyses of deletion derivatives of pRMB2, resulted in the identification of 2 regions (16029 bp - 20132 bp; 1 bp - 5339 bp) involved in the Ogawa serotype specificity of *V. cholerae* O-antigen within *E. coli* K-12 (summarized in Figure 6.1).

6.3 Identification of regions of promoter activity

The initiation of transcription is often the focus for regulation of gene expression. Regions of RNA polymerase binding within the 20 kb *Sac*I fragment were identified. Analyses of complexes by electron microscopy revealed 4 major areas of binding (1: .03 kb - 3.8 kb; 2: 9.2 kb - 10.5 kb; 3: 11.5 kb - 13.8 kb; 4: 14.8 kb - 15.8 kb), however, a high background level of binding may have masked weaker RNA polymerase binding sites. Unfortunately, a system for studying RNA polymerase binding in

Figure 6.1 Summary diagram of the *rfb* region of *V. cholerae* Ogawa.

This figure is a composite drawing of Figures 3.7, 3.9, and 4.8. Shown in this diagram are the regions involved in serotype specificity, shown by the hatched areas. The position of transposon insertions are shown by the pRMB numbers. (+) and (-) relate to the production of VcOAg and (*) denotes the presence of an insertion element. The minimal coding region is represented by a solid line. Putative promoters (P_1 to P_7) are shown. The predicted open reading frames with their direction of transcription are depicted by arrows. Limited restriction sites have been shown. *Sc* - *Sac*I, *K* - *Kpn*I, *B* - *Bam*HI, *X* - *Xba*I, *Sl* - *Sal*I, *C* - *Cla*I, *H* - *Hin*dIII.


pRMB56
pRMB20
pRMB49*
pRMB50
pRMN58
pRMB15
pRMB49
pRMB32
(−)
(−)
(+)
(−)
(−)
(+)
(+)
(+)
Sc
K
B
X
Sl
C
H
C
Sc
0
5
10
15
20
transposon insertion mutants
20132 bp
minimal coding region
P1
P2
P3
P4
P5
P6
P7
putative promoters
ORF 1
Omp X
23
A B C D E G H I K L M N O P Q R T
rfb genes
direction of transcription

not been established, consequently these studies, based on *E. coli* RNA polymerase, rely on the fact that production of VcOAg occurs in *E. coli* cells harbouring plasmids containing the *rfb* region, implying *E. coli* RNA polymerase recognizes the appropriate *V. cholerae* promoter regions.

Gene fusion techniques have proved extremely useful in the study of prokaryotic gene regulation. A series of vectors pKC86, pKC87, (Chak and James, 1985), and derivatives, containing a promoterless *galK* gene, were used to analyze the *rfb* region for promoter activity. These constructs were analyzed qualitatively, on galactose-tetrazolium indicator plates (TZGal), and quantitatively using a radioactive [^{14}C]-gal galactokinase assay. These studies revealed seven regions of promoter activity. Subsequent DNA sequence analyses of these regions has revealed the likelihood that three of these regions showing promoter activity are due to construction of artificial promoters. Such a phenomenon has been seen by Rosenberg *et al.* (1983), where deleting DNA from regions of no known promoter activity resulted in construction of detectable promoters.

The regions with promoter activity have been analyzed for promoter consensus sequences with the predicted *rfb* gene protein map in mind. This has resulted in the identification of: P_1 ((-35) 1240 bp - (-10) 1213 bp), postulated to be involved in the transcription of ORF1, a protein that is transcribed in the opposite direction to the *rfb* operon; P_2 ((-35) 1612 bp - (-10) 1639 bp or (-35) 1579 bp - (-10) 1607 bp), may be involved in transcription of *rfbA*; P_4 ((-35) 5495 bp - (-10) 5522 bp) is related to *rfbE* transcription; and P_7 ((-35) 20046 - (-10) 20076) (summarized in Figure 6.1). At 19975 bp, before P_7, which is located outside the known open reading frames contained in the *Sac*I fragment, is a putative terminator, with a potential free energy (ΔG) of -12.1 kCal/mol (Figure 4.9).

The galactokinase assay was performed on all clones obtained. The GalK units obtained showed the promoters to be weak when compared to the Lac promoter in the control plasmid. This was confirmed by direct analysis of *rfb* specific RNA.

RNAs obtained from *V. cholerae* and *E. coli* harbouring the *rfb* clones were analyzed by: primer extension, to identify the P_2 promoter in front of *rfbA*; Northern

analysis, to show the number of transcripts produced from the 20 kb *rfb* region; and by dot blot hybridization, to confirm the identification of the transcription strand. Unfortunately due to a low level of mRNA present within the cells, it was not possible to obtain definitive results from any of these procedures. The amount of O-antigen specific mRNA was shown to be at levels significantly less than that for the haemolysin structural gene (*hlyA*), compared in the same RNA preparation. This, however, is not surprising, as the mRNA encodes enzymes, which are probably only required in small amounts within the cell. These observations correlate with the low level of activity seen by the promoters identified in the GalK assay. There may also be a requirement of an activating factor to bind to the *V. cholerae* DNA to produce optimal transcription. This would not occur in the GalK assay performed, as only small segments of DNA are analyzed at the one time.

The various regions have been analyzed primarily for consensus sequences involving transcription using the sigma factors σ^{70} and σ^{54}. However, one can't rule out the possibility that different sigma factors may be employed in transcription of *V. cholerae* RNA, as no studies have been published in this field.

6.4 Partial DNA sequence analysis of the *rfb* region

Various segments of the *V. cholerae rfb* region were sequenced to analyze regions possessing promoter activity. Within the 20 kb region there are 22 open reading frames of which 3 (ORFs 1, 2, 3) are presumed not to play a part in O-antigen biosynthesis. Previous studies (Manning *et al.*, 1986) using cosmid clones of the *rfb* region, showed a minimum requirement of 16-19 kb, encompassed within a 20 kb *Sac*I fragment for VcOAg production. These three ORFs are outside of this minimal coding region.

Protein analyses have been performed using an *in vitro* transcription/translation system. This allowed the identification of 17 proteins, representing 88% of the total coding capacity of the *Sac*I region. The putative sizes of these proteins are (in kDa); 105, 52.8, 52.5, 47.9, 46.8, 42, 36, 35.2, 34.7, 32.5, 29.9, 29.4, 20, 18.9, 16.3, 12.2, and 8.8. Unforunately, until mutations in each gene have been obtained, and the protein product

analyzed, it is not possible to positively identify the protein encoded by the *rfb* gene. However, the 105 kDa protein seems likely to correspond to *rfbN*, the largest *rfb* encoded protein (91 kDa).

The *rfb* region is predicted to be composed of translationally coupled segments by analysis of open reading frames, and placement of Shine-Dalgarno sequences. Translational coupling has also been seen in the *S. typhimurium rfb* operon, as the start codon of *rfbE* overlaps the stop codon of the *rfbS* gene (Verma and Reeves, 1989).

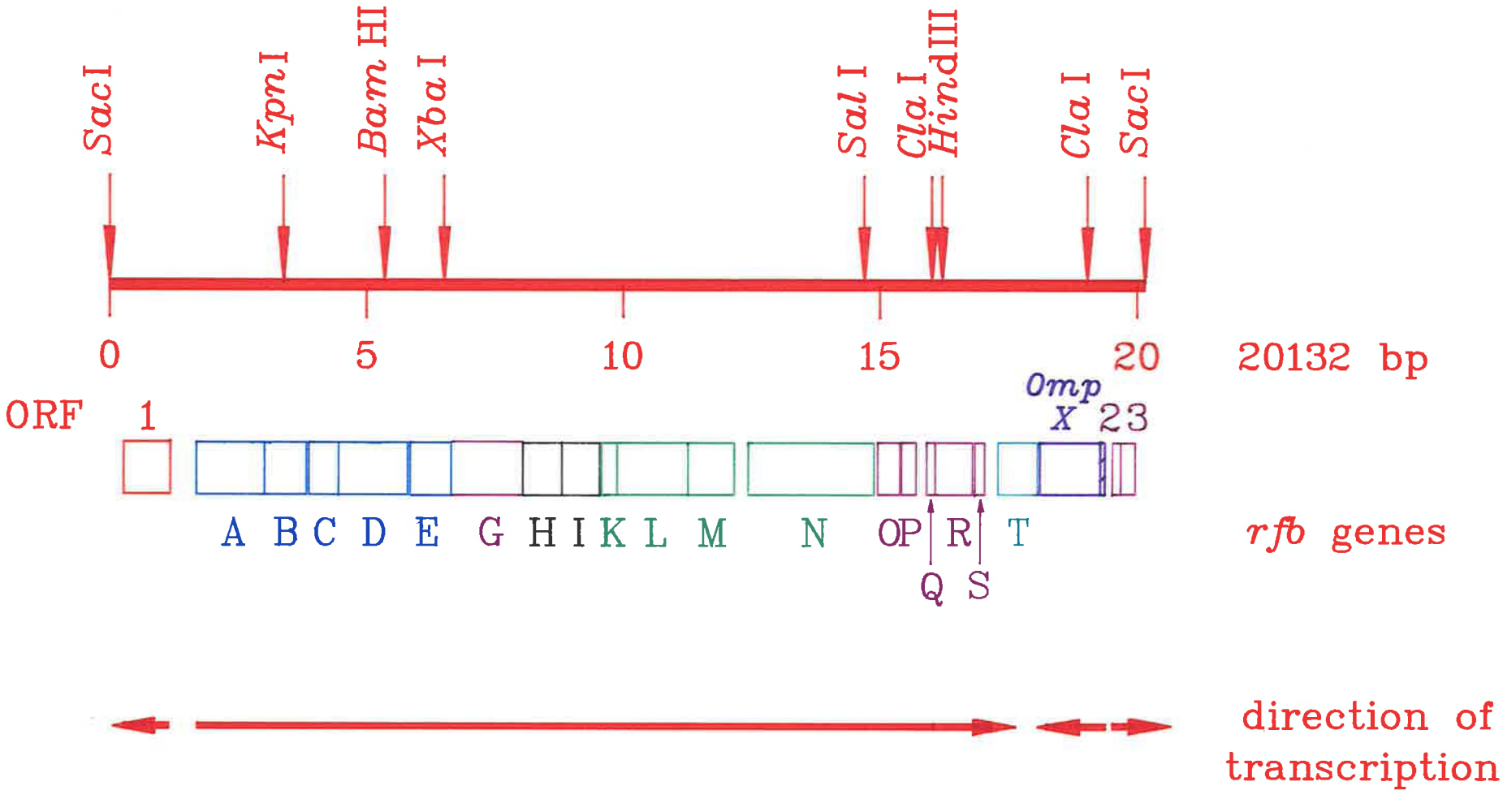
Analyses of the G+C content of the *rfb* region revealed a trend, such that blocks of genes could be compiled having a similar G+C content. This gave the idea that these blocks could correspond to functional units of the *rfb* region. To analyze this further, the amino acid sequence of the predicted proteins was used to identify homology with other proteins (Karageorgos, L., and Stroeher, U.H., personal communication). From these studies it was possible to divide the *rfb* region into functional domains involved with; perosamine synthesis (*rfbA, B, C, D, E*), 3-deoxy-L-glycero-tetronic acid synthesis (*rfbK, L, M, N*), transport (*rfbH, I*), and serotype specificity (*rfbT*) (Figure 6.2). (Manning, P.A., Morona, R., and Stroeher, U.H., personal communication) The *V. cholerae* proteins (VcRfb) VcRfbO to VcRfbS show no significant homology to any protein in the databases suggesting that their function is unique to *V. cholerae* O-antigen biosynthesis.

6.5 Proposed functional pathways of the biosynthesis of *V. cholerae* O-antigen

From both chemical analyses (Kenne *et al.*, 1979; Redmond, 1975, 1978, 1979; J. Redmond, personal communication) and the homologies seen between the VcRfb proteins and other bacterial proteins, it is possible to propose a potential pathway for O-antigen biosynthesis. As previously discussed, the *V. cholerae* O-antigen is composed of a "backbone" of α-1,2 linked perosamine (4-NH_2-4,6-dideoxy-mannose) residues, acylated with 3-deoxy-L-glycerotetronic acid. Also associated with the *V. cholerae* Ogawa O-antigen are the sugars 4-amino-4-deoxy-L-arabinose and quinovosamine (2-NH_2-2,6-dideoxy-glucose).

Figure 6.2 Division of the *V. cholerae rfb* region into functional domains.

Analyses of the G+C content of the *rfb* region combined with homology studies has enabled the 20 kb *rfb* region to be sectioned into functional areas, indicated by the different colours. ORF1 possesses protein homolgy to the *E. coli* RfaD protein, VcRfbA, B, C, and D may be involved in perosamine synthesis, VcRfbH, I may be involved in a transport function, VcRfbK, L, M, N are postulated to be associated with the production of 3-deoxy-L-glycero-tetronic acid, VcRfbT is associated with serotype specificity, and the proteins VcRfbG, O, P, Q, R, S, ORF2, and ORF3 have as yet no known function.


SacI
KpnI
BamHI
XbaI
SalI
ClaI
HindIII
ClaI
SacI
0
5
10
15
20
20132 bp
ORF 1
Omp X 23
A B C D E G H I K L M N O P Q R S T
rfb genes
direction of transcription

The biosynthetic pathway of perosamine appears to involve the *Vibrio cholerae rfb* genes *rfbA*, *rfbB*, *rfbC*, *rfbD*, and *rfbE* (summarized in Figure 6.3). This has been proposed due to proteins with similar functions in other bacteria possessing regions of significant homology with the *V. cholerae rfb* encoded proteins.

VcRfbA has significant homology to the phosphomannoisomerase enzyme from (Karageorgos, L., and Stroeher, unpublished results.)
P. aeruginosa. This enzyme isomerizes fructose-6-phosphate to mannose-6-phosphate. A fructose intermediate has been postulated to be involved in perosamine synthesis (Gabriel, 1987). The next step is catalyzed by phosphomannomutase. This protein (StRfbK) from *S. typhimurium*, has been found to possess homologous regions with the VcRfbB protein (Figure 5.25). This enzyme produces the isomer mannose-1-phosphate from mannose-6-phosphate.

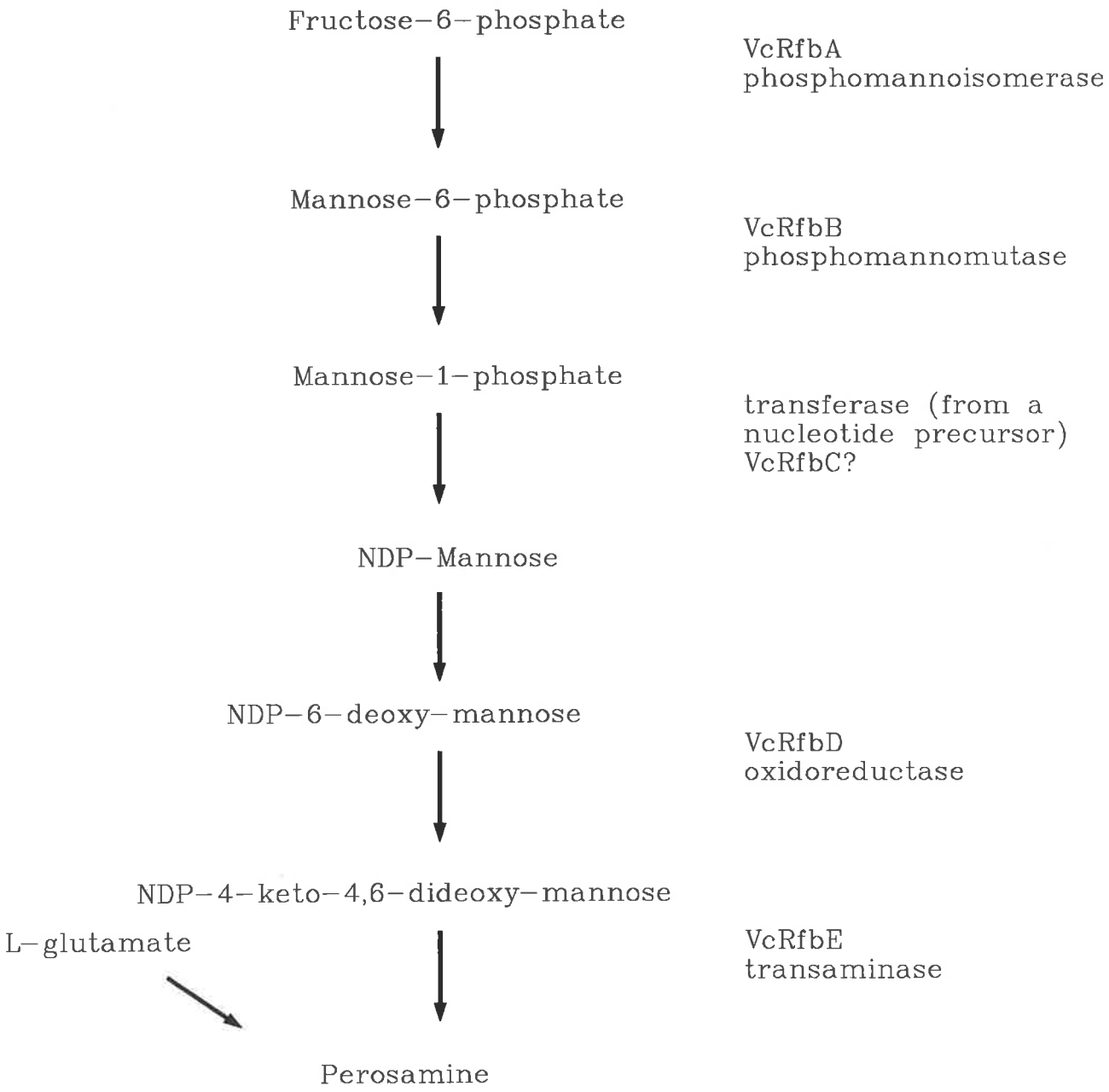
The next few steps in perosamine synthesis are not as easy to predict. At some stage it is presumed that the biosynthesis involves an activated nucleotide sugar (NDP), but it is uncertain at what stage it enters the pathway. However, this step requires a transferase, an enzyme which has not been identified amongst the predicted ORFs.

The production of 4-keto-4,6 dideoxy mannose-NDP from 6-deoxy-mannose-NDP requires an oxidoreductase, which is postulated to be the VcRfbD protein. A series of oxidoreductases involved in biosynthesis of 3,6-dideoxy sugars have been characterized in *Yersinia pseudotuberculosis* and *S. typhimurium* (Matsuhashi, 1966; Rubenstein and Strominger, 1974). Regions of homology have been shown between the RfbB, G, J, and S proteins of *S. typhimurium*, identifying a potential NADP binding domain (Figure 5.26). This step may involve an enzyme complex composed of more than one protein. Since, as yet no function has been defined for VcRfbC, this protein may also be involved in the production of 4-keto-4,6-dideoxy-mannose-NDP.

The last enzymatic step in the production of perosamine, transamination, requires L-glutamate (Gabriel, 1987). It is proposed that VcRfbE is involved, due to the similarity seen with the ErbS protein from *Saccharopolyspora erythraea* (Figure 5.27). This protein is involved in the production of the dideoxy sugar desosamine, however, the exact step it catalyzes is not known (Bibb *et al.*, 1985).

Figure 6.3 Proposed biosynthetic pathway for perosamine.

The pathway for the synthesis of perosamine has been postulated by homology studies. The VcRfb proteins thought to be involved in each reaction are shown on the right with the function they catalyze.


Fructose-6-phosphate
VcRfbA
phosphomannoisomerase
Mannose-6-phosphate
VcRfbB
phosphomannomutase
Mannose-1-phosphate
transferase (from a
nucleotide precursor)
VcRfbC?
NDP-Mannose
NDP-6-deoxy-mannose
VcRfbD
oxidoreductase
NDP-4-keto-4,6-dideoxy-mannose
L-glutamate
VcRfbE
transaminase
Perosamine

3-deoxy-L-glycero-tetronic acid production may involve the VcRfb proteins K, L, M, and N. (Karageorgos, L., Manning, P.A. and Morona, R., personal communication) This proposition is based on homology data: VcRfbK has homology to acyl carrier protein (ACP); VcRfbL is similar to enzymes synthesizing low molecular weight CoA derivatives; VcRfbM appears to encode an alcohol dehydrogenase; and VcRfbN is related to the LuxE and LuxC genes of *V. harveyii*. It is proposed that synthesis of 3-deoxy-L-glycero-tetronic acid commences with condensation of malonyl-ACP and acetyl-CoA (Alberts and Vagelos, 1972; Cronan and Rock, 1987).

Homology studies have not revealed any potential enzymes for the production of D-quinovosamine nor for the production of 4-amino-4-deoxy-L-arabinose. The most obvious route for D-quinovosamine production is via an oxido-reduction of NDP-D-glucosamine, similar to fucose biosynthesis in *S. typhimurium* (Shibaev, 1986). Synthesis of 4-amino-4-deoxy-L-arabinose is multifactorial (Figure 6.4). The precursor of arabinose may be either:

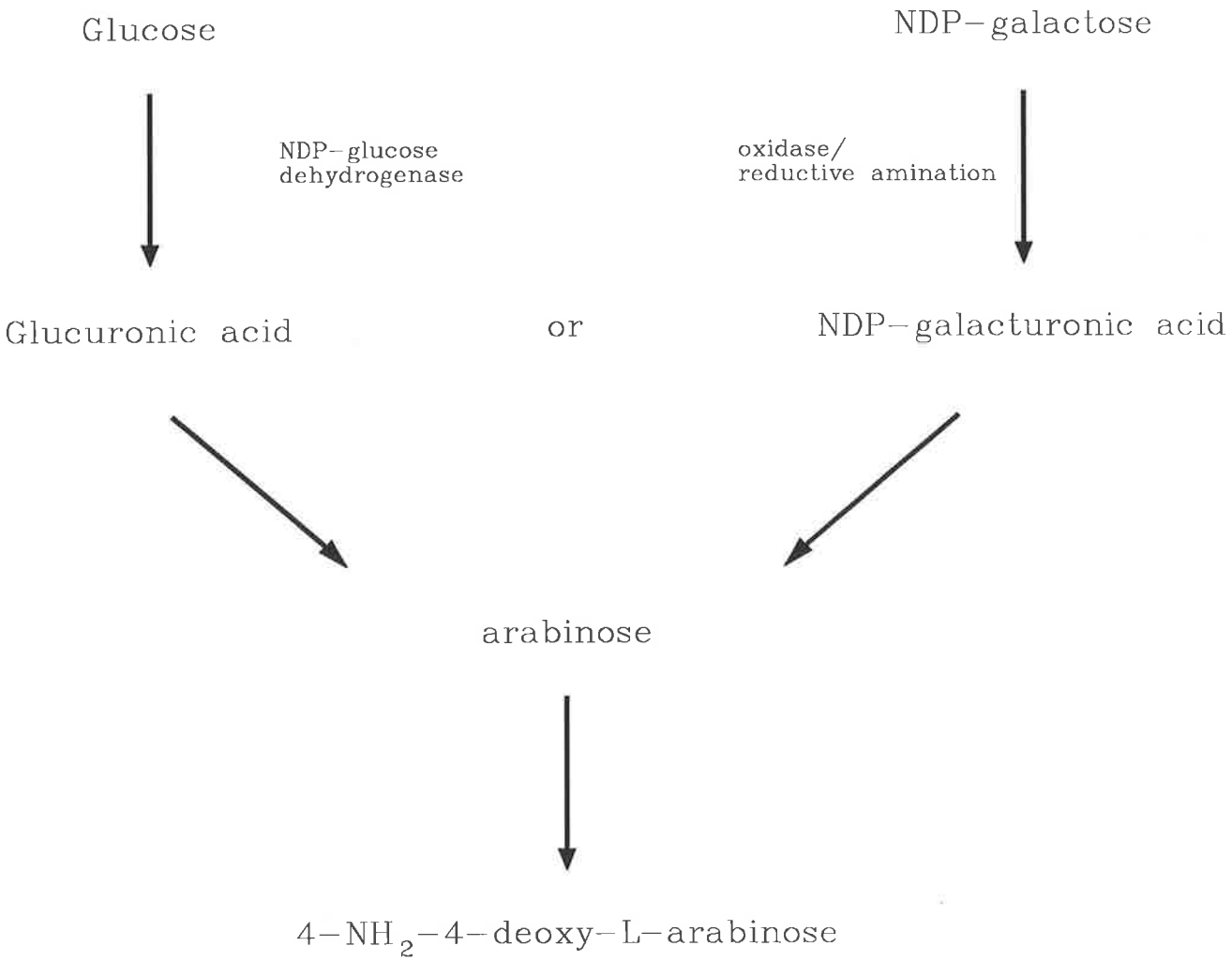
(1) glucose which is converted to glucuronic acid, catalyzed by a NDP-glucose dehydrogenase, and subsequently epimerased to form galacturonic acid. A similar step is performed by GalE in *S. typhimurium* (Rick, 1987) or

(2) NDP-galacturonic acid from NDP-D-galactose. Formation of the 4-amino group would require an oxidase followed by a reductive amination.

A surprising result seen by homology studies was the identity seen between ORF1 and RfaD of *E. coli*. The enzyme encoded by *rfaD* is involved in the formation of heptose, a component of the *E. coli* LPS core oligosaccharide. In studies centered on *E. coli* and *S. typhimurium* the *rfa* and *rfb* locus required for core and O-antigen biosyntheses, respectively, are located at different chromosomal locations. In *V. cholerae*, it appears that the *rfb* gene cluster and at least the *rfaD* gene from the *rfa* gene cluster are linked. Interestingly, Ward and Manning (1989) isolated an insertion mutant which mapped about 1 kb to the left of the *Sac*I fragment. This eliminated O-antigen biosynthesis. Thus, the *rfa* and *rfb* regions of *V. cholerae* may be contiguous but divergent.

Figure 6.4 Proposed biosynthetic pathway for 4-NH_2-4-deoxy-L-arabinose.

The pathway for the synthesis of 4-NH_2-4-deoxy-L-arabinose has been postulated by homology studies. The VcRfb proteins thought to be involved in each reaction are shown on the right with the function they catalyze.


Glucose
NDP-glucose dehydrogenase
NDP-galactose
oxidase/ reductive amination
Glucuronic acid
or
NDP-galacturonic acid
arabinose
4-NH2-4-deoxy-L-arabinose

Another block of genes which have been postulated to play a functional part in the formation of VcOAg, due to homology seen with other proteins, is *rfbH* and *rfbIA*. The corresponding proteins have been analyzed and significant homology has been found to bacterial proteins involved in membrane transport, and which possess a nucleotide binding consensus sequence, such as the HlyB protein from *E. coli*, the HlyB protein from *Proteus valgaris* and the BexA protein from *Haemophilus influenzae* (Fry *et al.*, 1986; Walker *et al.*, 1982). (Karageorgos, L., Manning, P.A. and Morona, R., personal communication)

6.6 Further experiments

The characterization of the *rfb* locus of *V. cholerae* Ogawa described in this thesis, presents a basis for future research.

Although the chemical structure of the *V. cholerae* LPS is not definitively known, using chemical (Redmond, 1975, 1978, 1979) and protein homology studies it has been possible to propose putative biochemical pathways for the biosynthesis of the O-antigen. These pathways needs to be confirmed, using both chemical and genetic means. Overexpression of VcRfb proteins would be advantageous to use with precursors (e.g. mannose-6-P) in enzyme assays, to enable a positive identification of certain enzymes.

The cellular location of the *rfb* gene products has only been identified by studying the predicted secondary structures and hydrophobicity plots. A more detailed analysis should be undertaken using genetic tools such as the transposon Tn*phoA*, (Manoil and Beckwith, 1985), fractionating *E. coli* containing clones expressing single VcRfb proteins, or sectioning cells, and analyzing the protein localization with antisera. As stated previously, the *E. coli* minicell strain harbouring the 20 kb *rfb* region was unstable. Various subclones were analyzed in this system, however, in the light of the DNA sequence data, analyses of subclones containing one *rfb* gene would produce a more definitive answer. Identification of *rfb* proteins, and N-terminal sequence analysis would help confirm the predicted location of the open reading frames.

Low levels of RNA expression hampered positive identification of promoter regions. This problem may be overcome by amplifying the mRNA in a reverse transcriptase dependent polymerase chain reaction (PCR), and thereby allowing confirmation of the direction of transcription. The methodology of hybridizing a [^{32}P]-labelled single stranded probe to ~~hybridize to~~ the mRNA, followed by S1 nuclease digestion of the single stranded region could also be used to identify transcription products. Another method to look at mRNA from each gene, is to transfer the cloned gene into *V. cholerae*. This approach has been undertaken with the *rfbT* gene (U.H. Stroeher, personal communication). Thus, if the promoters require an activating factor, then this would be present.

Once promoters have been identified, a number of genetic analyses can be performed. These include mutagenesis; to confirm and also to increase message expression, and identification of any activator or repressor proteins.

Thus, to summarize, the expression of VcOAg has been studied in *E. coli* harboring *rfb* clones. This has resulted initially in identification of the restriction endonuclease sites contained within this DNA. These data were used to construct deletion derivatives, Tn*1725* insertion derivatives, and subclones. Analyses of these constructs enabled the identification of regions involved in serotype (Ogawa) specificity, however the way in which these regions interact is stiil not fully understood.

Using a promoterless *galK* gene, areas of RNA polymerase binding were evaluated for promoter activity, and the strength of the cloned promoter deduced by a [^{14}C]-galactose assay. The DNA sequence of these regions was analyzed for promoter consensus sequences. Due to low levels of O-antigen specific mRNA precise mapping of the promoters was not achieved. This needs to be performed to enable a more detailed analysis of the genetic control of the *rfb* region to be undertaken.

Biosynthetic pathways for the production of VcOAg have been postulated based on protein homology studies. Confirmation of these pathways would be easier if the chemical structure of the O-antigen was definitive. However, by devising enyme assays, the predicted O-antigen biosynthetic pathway could be confirmed.

Genetic analysis of the *rfb* region has revealed, similar to other large operons, the presence of a large transcriptional unit. The regulation of the operon has yet to be defined, however, the analyses presented here have provided a basis for this to occur. Comparison of the genetics of the *rfb* regions of *S. typhimurium*, *E. coli* and *Shigella* has revealed similarities, however, it has also shown that the *V. cholerae rfb* region contains some novel features. This is exemplified by the presence of the *rfbH* and *rfbI* genes, possibly involved with transport of the O-antigen, within the *rfb* region. Another unique feature is the presence of ORF1, which has homology to the RfaD protein of *E. coli*. In other studied systems the *rfa* and *rfb* loci are distinct.

Thus, the analyses performed on the *V. cholerae rfb* region have provided an insight into *V. cholerae* O-antigen biosynthesis, and with the genetic construction of specific mutants, a better understanding of the chemical structure can be achieved.

BIBLIOGRAPHY

Achtman, M., Manning P.A., Edelbuth, 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.

Achtman, M., Schwuchous, S., Helmuth, R., Morelli, G., and Manning, P.A. (1978) Cell-cell interactions in conjugating *Escherichia coli*: Con$^-$ mutants and stabilization of mating aggregates. *Molec. Gen. Genet.* **164:** 171-183.

Adams, C.W., and Hatfield, G.W. (1984) Effects of promoter strengths and growth conditions on copy number of transcription-fusion vectors. *J. Biol. Chem.* **259:** 7399-7403.

Aiba, H., Adhya, S., and de Crombrugghe, B. (1981) Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* **256:** 11905-11910.

Aksoy, S., Squires, C.L., and Squires, C. (1984) Translational coupling of the *trpB* and *trpA* genes in the *Escherichia coli* tryptophane operon *J. Bacteriol.* **157:** 363-367.

Alm, R.A. and Manning, P.A. (1990a) Characterization of the *hlyB* gene and its role in the production of the El Tor hemolysin of *Vibrio cholerae* O1. *Molec. Microbiol.* **4:** 413-425.

Alm, R.A. and Manning, P.A. (1990b) Biotype-specific probe for *Vibrio cholerae* serogroup O1. *J. Clin. Microbiol.* **28:** 823-824.

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 strian O17 and characterization of the *hlyA* mutation in the non-haemolytic classical strain 569B. *Molec. Microbiol.* **2:** 481-488.

Ansari, A.A., Kenne, L., Lindberg, B., Gustafsson, B., and Holme, T. (1986) Structural studies of the O-antigen from *Vibrio cholerae* O:21. *Carbohydr. Res.* **150:** 213-219.

Arinc, E., Rzepecki, L.M., and Strittmatter, P. (1987) Topography of the C terminus of cytochrome b_5 tightly bound to dimyrisolphosphatidylcholine vesicles. *J. Biol. Chem.* **262:** 15563-15567.

Arnosti, D.N. and Chamberlin, M.J. (1989) Secondary σ factor control; σ transcription of flagellar and chemotaxis genes in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **86:** 830-834.

Attridge, S.R. and Rowley, D. (1983a) The role of the flagellum in the adherence of *Vibrio cholerae. J. Infect. Dis.* **147**: 864-872.

Attridge, S.R. and Rowley, D. (1983b) The specificity of *Vibrio cholerae* adherence and the significance of the slime agglutinin as a second mediator of *in vitro* attachment. *J. Infect. Dis.* **147**: 873-881.

Bachmann, B.J. and Low, K.B. (1980) Linkage map of *Escherichia coli* K-12., edition 6. *Microbiol. Rev.* **44:** 1-56.

Balgenesh, M. and Das, J. (1979) Transfection of *Vibrio cholerae* by bacteriophage 149 DNA. *Biochem. Biophys. Res. Comm.* **90:** 726-733.

Banoub, J.H. and Hodder, H.J. (1985) Structural investigation of the lipopolysaccharide core isolated from a virulent strain of *Vibrio ordalii. Can. J. Biochem. Cell. Biol.* **63:** 1199-1205.

Barnes, W.M., Bevan, M., and Son, P.H. (1983) Kilo-sequencing: Creation of an ordered nest of assymetric deletions across a large target sequence carried on phage M13. *Meth. Enzymol.* **101:** 98-122.

Bartowsky, E.J., Morelli, G., Kamke, M., Manning, P.A. (1987) Characterization and restriction analysis of the P sex factor and the cryptic plasmid of *Vibrio cholerae* strain V58. *Plasmid* **18:** 1-7.

Bera, T.K., Ghosh, S.K., and Das, J. (1989) Cloning and characterization of *mutL* and *mutS* genes of *V. cholerae*: Nucleotide sequence of the *mutL* gene. *Nucl. Acids. Res.* **17:** 6241-6251.

Baumann, P., Furniss, A.L., and Lee, J.V. (1984) Genus I *Vibrio*. pp. 518-538. *In:* Bergey's Manual of Systematic Bacteriology. Vol 1. (ed. Krieg, N.R.) Williams and Wilkins, Baltimore.

Bauman, P. and Schubert, R.H.W. (1984) The family Vibrionaceae. pp. 516-517. *In:* Bergey's Manual of Systematic Bacteriology. Vol 1. (ed. Krieg, N.R.) Williams and Wilkins, Baltimore.

Bayer, M.E. (1975) Role of adhesion zones in bacterial cell-surface function and biogenesis. pp. 393-427. *In:* Membrane Biogenesis. (ed. Tzagaloff, A.) Plenum Publ., New York.

Beck, C.F. and Warren, R.A.J. (1988) Divergent promoters, a common form of gene organization. *Microbiol. Rev.* **52:** 318-326.

Beutin, L., Manning, P.A., Achtman, M., and Willets, N. (1981) *sfrA* and *sfrB* products of *Escherichia coli* K12 are transcriptional control factors. *J. Bacteriol.* **145:** 840-844.

Bhaskaran, K. (1959) Observations of the nature of genetic recombination in *Vibrio cholerae*. *Indian J. Med. Res.* **47**: 253-260.

Bhaskaran, K. (1960) Recombination of characters between mutant stocks of *Vibrio cholerae*, strain 162. *J. Gen. Microbiol.* **23:** 47-54.

Bhaskaran, K. and Sinha, V.B. (1971) Transmissible plasmid factors and fertility inhibition in *Vibrio cholerae*. *J. Gen. Microbiol.* **69:** 89-97.

Bibb, M.J., Janssen, G.R., and Ward, J.M. (1985) Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. *Gene* **38:** 215-226.

Birkmann, A., Sawers, R.G., and Bock, A. (1987) Involvement of the *ntrA* gene product in the anaerobic metabolism of *Escherichia coli*. *Molec. Gen. Genet.* **210:** 535-542.

Blobel, G. (1980) Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. USA.* **77:** 1496-1500.

Brade, H. (1985) Occurrence of 2-keto-deoxyoctonic 5-phosphate in lipopolysaccharides of *Vibrio cholerae* Ogawa and Inaba. *J. Bacteriol.* **161:** 795-798.

Brade, H. and Galanos, C. (1983) Common lipopolysaccharide specificity: new type of antigen residing in the inner core region of S- and R- form lipopolysaccharides from different families of Gram negative bacteria. *Infect. Immun.* **42:** 250-256.

Brade, H. and Rietschel, E.-T. (1984) α-2-4 interlinked 3-deoxy-D-manno-octulosonic acid disaccharide. A common constituent of enterobacterial lipopolysaccharides. *Eur. J. Biochem.* **145:** 231-236.

Brahmbhatt, H.N., Quigley, N.B., and Reeves, P.R. (1986) Cloning part of the region encoding biosynthetic enzymes for surface antigen (O-antigen) of *Salmonella typhimurium*. *Molec. Gen. Genet.* **203:** 172-176.

Brahmbhatt, H.N., Wyk, P., Quigley, N.B., and Reeves, P.R., (1988) Complete physical map of the *rfb* gene region cluster encoding biosynthetic enzymes for the O-antigen of *Salmonella typhimurium* LT2. *J. Bacteriol.* **170:** 98-102.

Broady, K.W., Rietschel, E., and Lüderitz, O. (1981) The chemical structure of the lipid A component of lipopolysaccharides from *Vibrio cholerae*. *Eur. J. Biochem.* **115:** 463-468.

Brot, N., Peacock, S., and Weissbach, H. (1986) Regulation of the *Escherichia coli* L10 operon. pp. 749-765. *In:* Structure, function and genetics of ribosomes. (eds. Hardesty, B. and Kramer, G.) Springer Verlag, New York.

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.

Brunner, M. and Bujard, H. (1987) Promoter recognition and promoter strength in the *Escherichia coli* system. *Eur. J. Mol. Biol.* **10:** 3139-3144.

Bujard, H. (1980) The interaction of *E. coli* RNA polymerase with promoters. *Trends Biochem. Sci.* **5:** 274-278.

Bujard, H., Baldri, C., Brunner, M., Deuschle, U., Gentz, R., Hughes, J., Kammerer, W., and Stüber, D. (1983) pp. 65-87. *In:* Gene amplification and analysis, expression of cloned genes in procaryotic and eucaryotic cells. Vol. 3. (eds. Papas, T.S., Rosenberg, M., and Chirikijan, J.G.) Elsevier, Amsterdam.

Bujard, H., Brunner, M., Deuschle, U., Kammerer, W., and Knows, R. (1987) Structure-function relationship of *Escherichia coli* promoters. pp. 95-103. *In:* RNA polymerase and the regulation of transcription. (eds. Reznikoff *et al.*) Elsevier Science Pub Co.

Burgess, R.R. and Jendrisak, J.J. (1975) A procedure for the rapid, large-scale purification of *E. coli* DNA-dependent RNA polymerase involving Polymycin P precipitation and DNA cellulose chromatography. *Biochemistry* **14:** 4634-4638.

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.

Casadaban, M. and Cohen, S. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli. J. Mol. Biol.* **138:** 179-207.

Cash, R., Music, S., Libonati, J., Snyder, M., Wenzel, R. and Hornick, B. (1974a) Response of man to infection with *V. cholerae.* I. Clinical, serological and bacteriologic responses to a known inoculum. *J. Infect. Dis.* **129:** 45-52.

Cash, R.A., Music, S.I., Libonati, J.P. and Craig, J.P. (1974b) Response of man to infection with *Vibrio cholerae.* II. Protection from illness afforded by previous disease and vaccine. *J. Infect. Dis.* **130**: 325-333.

Chak, K-F. and James, R. (1985) Analysis of the promoters for the two immunity genes present in the ColE3-CA38 plasmid using two new promoter probe vectors. *Nucl. Acids. Res.* **13:** 2519-2531.

Chang, A.C.Y. and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid *J. Bacteriol.* **134:** 1141-1156.

Chitnis, D.S., Sharma, K.D. and Kamat, R.S. (1982) Role of somatic antigen of *Vibrio cholerae* in adhesion to intestinal mucosa. *J. Med. Microbiol.* **5:** 52-61.

Chou, P.Y. and Fasman, G.D. (1974a) Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins. *Biochemistry* **13:** 211-222.

Chou, P.Y. and Fasman, G.D. (1974b) Prediction of protein conformation. *Biochemistry* **13:** 222-245.

Chou, P.Y. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.***47:** 45-148.

Clements, S.D. and El-Morshidy, S. (1984) Construction of a potential live oral bivalent vaccine for typhoid fever and Cholera-*E. coli*-related diarrhoeas *Infect. Immun.* **46:** 564-569.

Close, T.J. and Rodriguez, R.L. (1982) Construction and characterization of the chloramphenicol-resistance gene cartridge: A new approach to the transcriptional mapping of extrachromosomal elements. *Gene* **20:** 305-316.

Coleman, W.G. Jr. and Deshpande, K.S. (1985) New *cysE-pyrE*-linked *rfa* mutation in *Escherichia coli* K-12 outer membrane. *J. Bacteriol.* **161:** 1209-1214.

Collins, J. (1979) Improved *in vitro* transcription/translation of plasmid encoded genes. *Gene* **6:** 29-33.

Collins, L.V. Attenuation of *Salmonella typhimurium.* (1990) PhD thesis. University of Adelaide.

Craig, J.P., Yamamoto, K., Takeda, Y. and Minatani, T. (1981) Production of cholera-like enterotoxin by a *Vibrio cholerae* non-O1 strain isolated from the environment. *Infect. Immunol.* **34:** 90-97.

Creeger, E.S. and Rothfield, L.I. (1979) Cloning of genes for bacterial glycosyltransferases. I. Selection of hybrid plasmids carrying genes for two glycosyltransferases. *J. Biol. Chem.* **254:** 804-810.

Creeger, E.S. and Rothfield, L.I., (1982) Cloning genes for bacterial glycotransferases. *Methods. Enzym.* **83:** 326-331.

Creeger, E.S., Schulte, T., and Rothfield, L.I. (1984) Regulation of membrane glycosyltransferases by the *sfrB* and *rfaH* genes of *Escherichia coli* and *Salmonella typhimurium. J. Biol. Chem.* **259:** 3064-3069.

Crowell, D.N., Anderson, M.S., and Raetz, C.R.H. (1986) Molecular cloning of the genes for lipid A disaccharide synthase and UDP-N-acetylglucosamine in *Escherichia coli. J. Bacteriol.* **168:** 152-159.

Cryz, S.J., Jr., Fürer, E., and Germainer, R. (1982) Development of an enzyme-linked immunosorbent assay for studying *Vibrio cholerae* cell surface antigens. *J. Clin. Microbiol.* **16:** 41-45.

Cuatrecasas, P. (1973a) Interaction of *Vibrio cholerae* enterotoxin with cell membranes. *Biochemistry* **12:** 3547-3558.

Cuatrecasas, P. (1973b) Gangliosides and membrane receptors for cholera toxin. *Biochemistry* **12:** 3558-3566.

Datta, A., Parker, C.D., Wohlheiter, J.A., and Baron, L.S. (1983) Isolation and characterization of the fertility factor P of *Vibrio cholerae. J. Bacteriol.* **113:** 763-771.

Davis, B.D., Dulbecco, R., Eisen, H.N. and Ginsberg, H.S. (1980) Microbiology. 3rd ed. Harper and Row, Publishers, Inc.

De, S.N. (1959) Enterotoxicity of bacteria-free culture filtrate of *Vibrio cholerae. Nature* **183:** 1533-1534.

Deuschle, U., Kammerer, W., Gentz, R., and Bujard H. (1986) Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *Eur. J. Molec. Biol.* **5:** 2987-2994.

DeVries, J.K. and Zubay, G. (1967) DNA-directed peptide synthesis, II. The synthesis of the α-fragment of the enzyme β-galactsidase. *Proc. Natl. Acad. Sci. USA* **57:** 1010-1013.

Dhillon, N., Hale, R.S., Cortes, J., Leadloy, P.F. (1989) Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. *Mol. Microbiol* **3:** 1405-1414.

Di Rita, V.J. and Mekalanos, J.J. (1991) Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS, results in signal transduction and transcriptional activation. *Cell* **64:** 29-37.

Dixon, R. (1986) The *xylABC* promoter from *Pseudomonas putida* TOL plasmid is activated by nitrogen regulatory genes in *Escherichia coli. Molec. Gen. Genet.* **203:** 129-136.

Doi, R.H. and Wang, L-F. (1986) Multiple procaryotic ribonucleic acid polymerase sigma factors. *Microbiol. Rev.* **50:** 227-243.

Edwards, M.F. and Stocker, B.A.D. (1988) Construction of Δ*aro his* Δ*pur* strains of *Salmonella typhi. J. Bacteriol.* **170:** 3991-3995.

Ehara, M., Ishibashi, M., Ichinose, Y., Iwanaga, M., Shimotori, S. and Naito, T. (1987) Purification and partial characterization of fimbriae of *Vibrio cholerae* O1. *Vaccine* **5:** 283-288.

Ehara, M., Ishibashi, M., Watanabe, S., Iwanaga, M., Shimotori, S. and Naito, T. (1986) Fimbriae of *Vibrio cholerae* O1: observation of fimbriae on the organisms adherent to the intestinal epithelium and development of a new medium to enhance fimbriae production. *Trop. Med.* **28:** 21-33.

Eidels, L. and Osborn M.J. (1971) Lipopolysaccharide and aldoheptose biosynthesis in transketolase mutants of *Salmonella typhimurium*. *Proc. Natl. Acad. Sci. USA* **68:** 1673-1677.

Eidels, L. and Osborn, M.J. (1974) Phosphoheptose isomerase, first enzyme in the biosynthesis of aldoheptose in *Salmonella typhimurium*. *J. Biol. Chem.* **249:** 5642-5648.

Ellwood, D.C. (1970) The distribution of 2-keto-3-deoxy-octonic acid in bacterial walls. *J. Gen. Microbiol.* **60:** 373-380.

Evans, D.J. and Richardson, S.H. (1968) In vitro production of choleragen and vascular permeability factor by *Vibrio cholerae*. *J. Bacteriol.* **96:** 126-130.

Evans, S. and Dennis, P.P. (1985) Promoter activity and transcript mapping in the regulatory region for genes encoding ribosomal protein S15 and polynucleotide phosphorylase of *Escherichia coli*. *Gene* **40:** 15-22.

Finkelstein, R.A. (1969) The role of choleragen in the pathogenesis and immunology of cholera. *Tex. Rep. Biol. Med.* **27:** S181-S201.

Finkelstein, R.A. and Hanne, L.F. (1982) Purification and characterization of the soluble hemagglutinin (cholera lectin) produced by *Vibrio cholerae*. *Infect. Immun.* **36:** 119-1208.

Focareta, T. (1989) The extracellular Dnase(s) of *Vibrio cholerae*. PhD. Thesis, University of Adelaide.

Focareta, T. and Manning, P.A. (1987) Extracellular proteins of *Vibrio cholerae*: Molecular cloning, nucleotide sequence and characterization of the deoxyribonuclease (DNase) together with its periplasmic localization in *Escherichia coli* K-12. *Gene* **53:** 31-40.

Formal, S.B., Baron, L.S., Kopecko, D.J., Washington, O., Powell, C., and Life, C.A. (1981) Construction of a potential bivalent vaccine strain: introduction of *Shigella sonnei* Form I antigen genes into the *galE Salmonella typhi* Ty21a typhoid vaccine strain. *Infect. Immun.* **34:** 746-750.

Formal, S.B., Gemski, P., Jr., Baron, L.S., and La Brec, E.H. (1970) Genetic transfer of *Shigella flexneri* antigens to *Escherichia coli* K-12. *Infect. Immun.* **1:** 279-287.

Forrest, B.D., LaBrooy, J.T., Attridge, S.R., Boehm, G., Beyer, L., Morona, R., Shearman, D.J.C., and Rowley, D. (1989) Immunogenicity of a candidate live oral typhoid/cholera hybrid vaccine in humans. *J. Infect. Dis.* **159:** 145-146.

Fox, T.D., Losick, R., and Pero, J. (1976) Regulatory gene *28* of bacteriophage SP01 codes for a phage-induced subunit of RNA polymerase. *J. Mol. Biol.* **101:** 427-433.

Franzon, V.L. and Manning, P.A. (1986) Molecular cloning and expression in *Escherichia coli* K-12 of the genes for a hemagglutinin from *Vibrio cholerae*. *Infect. Immun.* **52:** 279-284.

Franzon, V.L., Clark, C.A., Williams, S.G., and Manning P.A. (1990) Nucleotide sequence and construction of a mutant in a mannose-fucose resistant haemagglutinin (MFRHA) of *Vibrio cholerae* O1. *In:* Advances in Research on Cholera and Related Diarrhoeas. Vol 8. (eds. Takeda, Y. and Sack, R.B.). KTK Publications, Tokyo, in press.

Freeley, J.C. (1965) Classification of *Vibrio cholerae* (Vibrio comma), including El Tor vibrios, by intrasubspecific characteristics. *J. Bacteriol.* **89:** 665-674.

Freter, R. and O'Brien, P.C.M. (1981) Role of chemotaxis in the association of motile bacteria with intestinal mucosa: fitness and virulence of nonchemotactic *Vibrio cholerae* mutants in infant mice. *Infect. Immun.* **34:** 222-233.

Fry, D.C, Kuby, S.A., and Mildva, A.S. (1986) ATP-binding site of adenylate kinase: Mechanistic implications of its homology with *ras*-encoded p21, F1-ATPase, and other nucleotide-binding proteins. *Proc. Natl. Acad. Sci. USA.* **83:** 907-911.

Gabriel, O. (1987) Biosynthesis of sugar residues for glycogen, peptidoglycan, lipopolysaccharide, and related systems. pp. 504-512. *In: Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Vol I. (ed Neidhardt, F.C.) American Society for Microbiology.

Galen J.E. (1988) *In:* Advances in Cholera and related diarrhoeas. Vol VII. (eds. Zinnaka, Y. and Sack, R.B.) KTK Publications, Tokyo.

Gan, K.H. and Tjia, S.K. (1963) A new method for the differentiation of *Vibrio comma* and Vibrio El Tor. *Am. J. Hyg.* **77:** 184-186.

Gangarosa, E.J., Sonati, A., Saghari, H. and Feeley, J.C. (1967) Multiple serotypes of *Vibrio cholerae* from a case of cholera. *Lancet* **1:** 646-648.

Gardner, A.D. and Venkatramen, K.V. (1935) The antigens of the cholera group of vibrios. *J. Hyg.* **25:** 262-282.

Garger, S.J., Griffin, 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.

Gentz, R. and Bujard, H.G. (1985) Promoters recognized by *Escherichia coli* RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. *J. Bacteriol.* **164:** 70-77.

Gentz, R., Langer, A., Chang, A.C.Y., Cohen, S.N., and Bujard, H. (1981) Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. *Proc. Natl. Acad. Sci. USA* **78:** 4936-4940.

Germanier, R. and Fürer, E. (1975) Isolation and characterization of *galE* mutant Ty21a of *Salmonella typhi*: a candidate strain for a live oral typhoid vaccine. *J. Infect. Dis.* **131:** 553-558.

Gill, D.M. (1976) The arrangement of subunits in cholera toxin. *Biochemistry* **15:** 1242-1248.

Gmeiner, J., Simon, M., and Lüderitz, O. (1971) The linkage of phosphate group and of 2-keto-3-deoxyoctonate to the lipid A component in *Salmonella minnesota* lipopolysaccharide. *Eur. J. Biochem.* **21:** 355-356.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.W., and Stormo, G. (1981) Translational initiation in prokaryotes. *Annu. Rev. Microbiol.* **35:** 365-403.

Goldberg, S. and Murphy, J.R. (1983) Molecular epidemiological studies of United States Gulf Coast *Vibrio cholerae* strains: integration site for mutator vibriophage VcA-3. *Infect. Immun.* **42:** 224-230.

Goldman, R. and Strominger, J.L. (1972) Purification and properties of C_{55}-isoprenyl-pyrophoshate from *Micrococcus lysodeikticus*. *J. Biol. Chem.* **247:** 5116-5122.

Gough, J.A. and Murray, N.E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J. Mol. Biol.* **166:** 1-19.

Green, B.A., Newland, J.W., and Holmes, R.K. (1983) Mapping of chromosomal genes that determine the El Tor biotype in *Vibrio cholerae*. *Infect. Immun.* **42:** 924-929.

Guentzel, M.N. and Berry, L.J. (1975) Motility as a virulence factor for *Vibrio cholerae*. *Infect. Immun.* **11:** 890-897.

Guhathakurta, B., Majumdar, M., Sen, A.K., Sasmal, D., Mukherjee, A.K., and Datta, A. (1986) Immunochemical proerties of the lipopolysaccharide O-antigen of *Vibrio cholerae* O1 in relation to its chemical structure. *J. Gen. Microbiol.* **132:** 1641-1646.

Guhathakurta, B., Sasmal, D., Datta, A., Sean, A.K., and Mukherjee, A.K. (1980) Biochemical studies on lipopolysaccharide isolated from agglutinable and non-agglutinable strains of *V. cholerae*. *Ind. J. Biochem. Biophys.* **17:** 467-469.

Guidolin, A. and Manning, P.A. (1985) *Vibrio cholerae* bacteriophage CP-T1: characterization of the receptor. *Eur. J. Biochem.* **153:** 89-94.

Guidolin, A., and Manning P.A. (1987) Genetics of *Vibrio cholerae* and its bacteriophages. *Microbiol. Rev.* **51:** 285-298.

Guidolin, A. and Manning, P.A. (1988) Molecular analysis of the packaging signal in bacteriophage CP-T1 of *Vibrio cholerae*. *Molec. Gen. Genet.* **212:** 514-521.

Guidolin, A., Morelli, G., Kamke, M. and Manning, P.A. (1984) *Vibrio cholerae* bacteriophage CP-T1: Characterization of bacteriophage DNA and restriction analysis. *J. Virol.* **51:** 163-169.

Gupta, D.S. Jann, B., and Jann, K. (1984) *Escherichia coli* O18ac antigen: Structure of the O-specific polysaccharide moiety. *Infect. Immun.* **45:** 203-209.

Gustafsson, B. (1985) Detection and characterization of lipopolysaccharide antigens in *Vibrio cholerae* O:1. PhD. thesis, Karolinska Institutet, Stockholm.

Gustafsson, B. and Holme, T. (1985) Immunological characterization of *Vibrio cholerae* O1 lipopolysaccharide, O-side chain, and the core with monoclonal antibodies. *Infect. Immun.* **49:** 275-280.

Hale, T.L., Guerry, P., Seid, R.C., Kapfer, C., Wingfield, M.E., Reaves, C.B., Baron, L.S., and Formal, S.B. (1984) Expression of lipopolysaccharide O-antigen in *Escherichia coli* K-12 hybrids containing plasmids and chromosomal genes from *Shigella dysenteriae* 1. *Infect. Immun.* **46:** 470-475.

Haller, I., and Henning, U. (1974) Cell envelope and shape of *Escherichia coli* K12: crosslinking with dimethyl imidoesters of whole cell wall. *Proc. Natl. Acad. Sci. USA* **71:** 2018-2021.

Hämmerling, G., Lehmann, V., and Lüderitz, O. (1973) Structural studies on the heptose region of *Salmonella* lipopolysaccharides. *Eur. J. Biochem.* **38:** 453-458.

Hämmerling, G., Lüderitz, O., Westphal, O., and Mäkelä, P.H. (1971) Structural investigations on the core polysaccharide of *Escherichia coli* O100. *Eur. J. Biochem.* **22:** 331-344.

Hanne, L.F. and Finkelstein, R.A. (1982) Characterization and distribution of the hemagglutinins produced by *Vibrio cholerae. Infect. Immun.* **36:** 209-214.

Haraguchi, G.E., Hull, R.A., Krallmann-Wenzel, U., and Hull, S.I. (1989) Molecular cloning and expression of the O4 polysaccharide gene cluster from *Escherichia coli. Microbial. Pathogenesis.* **6:** 123-132.

Hartman, P.E., Loper, J.C., and Serman, D. (1961) Fine structure mapping by complete transduction between histidine requiring *Salmonella* mutants. *J. Gen. Microbiol.* **22:** 323-353.

Hase, S. and Rietschel, E.Th. (1976) Isolation and analysis of the lipid A backbone. Lipid A structure of lipopolysaccharides from various bacterial groups. *Eur. J. Biochem* **63:** 101-107.

Hase, S. and Rietschel, E.Th (1977) The chemical structure of the lipid A component of lipopolysaccharides from *Chromobacterium violaccum. Eur. J. Biochem.* **75:** 23-34.

Haselkorn, R. (1986) Organization of the genes for nitrogen fixation in photosynthetic bacteria and cyanobacteria. *Annu. Rev. Microbiol.* **40:** 527-547.

Hawkes, R., Niday, E., and Gordon, J. (1982) A dot-immuno binding assay for monoclonal and other antibodies. *Anal. Biochem.* **119:** 142-147.

Hawley, D.K. and McClure, W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl. Acids. Res.* **8:** 2237-2255.

Heiberg, B. (1935) On the classification of *Vibrio cholerae* and the *cholera*-like vibrios. Busk, Copenhagen: 1936, Two serlogically different groups among the true cholera vibrios. *J. Hyg.* **36:** 118-124.

Hellerqvist, C.G. and Lindberg, A.A. (1971) Structural studies of the common-core polysaccharide of the cell wall lipopolysaccharide from *Salmonella typhimurium. Carbohydr. Res.* **16:** 39-48.

Helmann J.D., Marquez, L.M., and Chamberlin, M.J. (1988) Cloning, sequencing, and disruption of the *Bacillus subtilis* σ^{28} gene. *J. Bacteriol.* **170:** 1568-1574.

Hemming, F.W. (1970) Polyprenols. pp. 105-117. *In:* (ed. Goodwin, T.W.) Vol. 29. Biochemistry Society Symposia Academic Press, London.

Heuzenroeder, M.W., Beger, D.W. Thomas, C.T., and Manning, P.A. (1989) Molecular cloning and expression in *Escherichia coli* K-12 of the O1O1 *rfb* region from *E. coli* B41 (O1O1:K99/F41) and the genetic relationship to other O1O1 *rfb* loci. *Mol. Microbiol.* **3:** 295-302.

Higgins, D.G. and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73:** 237-244.

Hirano, M., Shigesada, K., and Imai, M. (1987) Construction and characterization of plasmid and lambda phage vector systems for study of transcriptional control in *Escherichia coli*. *Gene* **57:** 89-99.

Hisatsune, K. and Kondo, S. (1980) Lipopolysaccharides of R mutants isolated from *Vibrio cholerae*. *Biochem. J.* **185:** 77-81.

Hisatsune, K., Haishima, Y., and Kondo, S. (1987) O-antigenic lipopolysaccharides isolated from *Vibrio cholerae* bio-serogroup Hakata and *Vibrio fluvialis* each possessing common antigenic factor to O1 *Vibrio cholerae*. pp. 52-53. 23rd Joint Conference U.S.-Japan Cooperative Medical Science program, Cholera and related diarrheal diseases, Tokyo.

Hisatsune, K., Kiuye, A., and Kondo, S. (1980) Sugar comosition of O-antigenic lipopolysaccharides isolated from *Vibrio parahaemolyticus*. *Microbiol. Immun.* **24:** 691-701.

Hisatsune, K., Kiuye, A., and Kondo, S. (1981) A comparative study of the sugar composition of O-antigenic lipopolysaccharides isolated from *Vibrio alginolyticus* and *Vibrio parahaemolyticus*. *Microbiol. Immun.* **25:** 127-136.

Hisatsune, K., Kondo, S., and Haishima, Y. (1986) O-antigenic lipopolysaccharide (LPS) of vibrio bio-serogroup 1875, a marine vibrio possessing common antigen factor to O1 *Vibrio cholerae*. pp. 11-12. 22nd Joint Conference U.S.-Japan Cooperative Medical Science program, Cholera and related diarrheal diseases, Tokyo.

Hisatsune, K., Kondo, S., and Iguchi, T. (1978) Further chemical and immunochemical characterization of lipopolysaccharides of *Vibrio cholerae* including their R-mutants. pp. 171-181. 14th Joint Conference U.S.-Japan Cooperative Medical Science program, Cholera panel. NIH. Tokyo.

Hisatsune, K., Kondo, S., and Kobayashi, K. (1978) Lipopolysaccharides of *Vibrio cholerae* (II) - an immunchemical study on O-antigen structure. *Japan J. Med. Sci. Biol.* **31:** 181-184.

Hisatsune, K., Kondo, S., Kawata, T., and Kishimoto, Y. (1979) Fatty acid composition of lipopolysaccharides of *Vibrio cholerae* 35A3 (Inaba), NH90 (Ogawa), and 4715 (NAG). *J. Bacteriol.* **138:** 288-290.

Hisatsune, K., Yamamoto, F., and Kondo, S. (1985) Lipopolysaccaride of *V. cholerae* - chemical and serological properties. pp. 17-24. *In:* Advances in research on cholera and related diarrheas. Vol. 2. (eds. Kuwahara, S. and Pierce, N.F.).

Hitchcock, P.J. and Brown, T.M. (1983) Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver stained polyacrylamide gels. *J. Bacteriol.* **154:** 269-277.

Holmgren, J. (1981) Actions of cholera toxin and the prevention and treatment of cholera. *Nature* **292:** 413-417.

Holmgren, J., Lônnroth, I., and Ouchterlony, O. (1971) Immunochemical studies of two cholera toxin containing standard culture filtrate preparations of *Vibrio cholerae*. *Infect. Immun.* **3:** 747-755.

Holmgren, J. and Svennerholm, A.-M. (1977) Mechanisms of disease and immunity in cholera: a review. *J. Infect. Dis.* **136:** S105-S112.

Holmgren, J., Svennerholm, A-M., Lonnroth, E., Fall, I., Persson, M., Markman, B., and Lundbeck, H. (1977) Development of an improved cholera vaccine based on subunit toxoid. *Nature* **269:** 602-604.

Horowitz, H. and Platt, T. (1982) Regulation of transcription from tandem and convergent promoters *Nucl. Acids. Res.* **10:** 5447-5465.

Hunt, T.P. and Magasanik, B. (1985) Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF, glnC,* and *glnL* *Proc. Natl. Acad. Sci. USA* **82:** 8453-8457.

Inouye, M. and Halegoua, S. (1980) Secretion and membrane localization of proteins in *Escherichia coli. CRC Crit. Rev. Biochem.* **7:** 339-371.

Inzana, T., (1983) Electrophoretic heterogeneity and interstrain variation of the lipopolysaccharide of *Haemophilus influenzae. J. Infect. Dis.* **148:** 492-499.

Ito, T. and Yokota, T. (1987) Different types of monoclonal antibodies to Ogawa-specific and group-specific antigens of *Vibrio cholerae* O1. *J. Clin. Microbiol.* **25:** 2289-2295.

Itoh, S., Mizukami, T., Masumtot, T., Nichi, T., Saito, A., Oka, T., Furuya, A., Takaoka, C., and Taniguchi (1984) Efficient expression in *Escherichia coli* of a mature and modified human interferon-b_1. *DNA* **3:** 157-165.

Jackson, G.D.F. and Redmond, J.W. (1971) Immunchemical studies of the O-antigens of *Vibrio cholerae* I. The constitution of a lipopolysaccharide from *Vibrio cholerae* 569B. *FEBS Lett.* **13:** 117-120.

Jalajakumari, M.B. and Manning, P.A. (1990) Nucleotide sequence of the gene *ompW*, encoding a 22 kDa immunogenic outermembrane protein of *Vibrio cholerae. Nucl. Acids. Res.* **18:** 2180.

Jalajakumari, M.B., Guidolin, A., Buhk, H-J., Manning, P.A., Ham, L.M., Hodgson, A.L.M., Cheah, K.C., and Skurray, R.A. (1987) Surface exclusion genes *traS* and *traT* of the F sex factor of *Escherichia coli* K-12: Determination of the nucleotide sequence and promoter and terminator activities. *J. Mol. Biol.* **198:** 1-11.

Jann, K. and Jann, B. (1984) Structure and biosynthesis of O-antigens. pp. 138-186. *In:* Handbook of Endotoxin. Vol.I. (ed. Rietschel, E.Th.) Elsevier Science Publishers, Amsterdam.

Jann, K., Goldemann, G., Weisberger, C., Wolf - Ullisch, C., and Kanegasaki, S. (1982) Biosynthesis of the 09 antigen of *Escherichia coli*. Initial reaction and overall reactions. *Eur. J. Biochem.* **127:** 157-164.

Jann, B., Jann, K., and Beyaert, G.O. (1973) 2-amino-2,6-dideoxy-D-glucose (D-Quinavosamine): a constituent of the lipopolysaccharides of *Vibrio cholerae*. *Eur. J. Biochem.* **37:** 531-534.

Jansson, P. E., Lindberg, A., Lindberg, B., and Wollin, R. (1981) Structural studies on the hexose region of the core lipopolysaccharides from *Enterobacteriaceae*. *Eur. J. Biochem.* **115:** 571-577.

Jaurin, B., Grundström, T., and Normak, S. (1982) Sequence elements determining *ampC* promoter strength in *E. coli*. *Eur. J. Mol. Biol.* **1:** 875-881.

Jennings, M.L. (1989) Topography of membrane proteins *Annu. Rev. Biochem.* **58:** 999-1027.

Johnson, A.W.B. and Downie, J.A. (1984) What is a *nif* promoter? *Trends Biochem. Sci.* **9:** 367-368.

Johnson, K., Parker, M.L., and Lory, S. (1986) Nucleotide sequence and transcriptional initiation site of two *Pseudomonas aeruginosa* mutants with altered piliation *J. Bacteriol.* **169:** 5663-5667.

Johnson, S.R. and Romig, W.R. (1979) Transposon-facilitated recombination in *Vibrio cholerae*. *Molec. Gen. Genet.* **170:** 93-101.

Joiner, K.A., Brown, E.J., and Frank, M.M. (1984) Complement and bacteria: chemistry and biology in host defense. *Annu. Rev. Immunol.* **2:** 461-491.

Jones, G.W. and Freter, R. (1976) Adhesive properties of *Vibrio cholerae*: nature of the interaction with isolated rabbit brush border membranes and human erythrocytes. *Infect. Immun.* **14**: 240-245.

Jones, G.W., Abrams, G.D., and Freter, R. (1976) Adhesive properties of *Vibrio cholerae*: adhesion to isolated rabbit brush border membranes and haemagglutinating activity. *Infect. Immun.* **14:** 232-239.

Jousimies, H. and Mäkelä, P.H. (1974) Genetic analysis of *Salmonella minnesota* R mutants with defects in the biosynthesis of the lipopolysaccharide core. *J. Bacteriol.* **119:** 753-759.

Kabeshima, T. (1918) On the serology of *Vibrio cholerae. Acta. path. microbiol. Scand.* **27:** 282.

Kabir, S. (1980) Composition and immunochemical properties of outer membrane proteins of *Vibrio cholerae. J. Bacteriol.* **144:** 382-389.

Kabir, S. (1982) Characterization of the lipopolysaccharide from *Vibrio cholerae* 395 (Ogawa). *Infect. Immun.* **38:** 1263-1272.

Kabir, S. (1983) Immunochemical properties of the major outer membrane protein of *Vibrio cholerae. Infect. Immun.* **39:** 452-455.

Kabir, S. and Showkat, A. (1983) Characterization of surface properties of *Vibrio cholerae. Infect Immun.* **39:** 1048-1058.

Kaca, W., Brade, L., Rietschel, E. Th., and Brade, H. (1986) The effect of removal of D-fructose on the antigenicity of the lipopolysaccharide from a rough mutant of *Vibrio cholerae* Ogawa. *Carbohydr. Res.* **149:** 293-298.

Kadam, S.K., Rehemtulla, A., and Sanderson, K.E. (1985) Cloning of *rfa G, B, I,* and *J* genes for glycosyltransferase enzymes for synthesis of the lipopolysaccharide core of *Salmonella typhimurium. J. Bacteriol.* **161**: 277-284.

Kahn, M., Kotler, T., Thomas, C., Figurski, D., Meyer, R., Renault, E., and Helinski, D.R. (1979) Plasmid cloning vehicles derived from plasmids ColE1, RGK and RK2. *Methods Enzymol.* **68:** 268-280.

Kamal, A.M. (1974) The Seventh Pandemic of Cholera. pp. 1-14. *In:* Cholera. (eds. Barua, D. and Burrows, W.). W.B. Saunders Co.

Kammerer, W., Deuschle, U., Gentz, R., and Bujard, H. (1986) Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. *Eur. J. Mol. Biol.* **5:** 2995-3000.

Kass, E.H. and Wolff. S.M. (eds) (1973) Bacterial Lipopolysaccharides, the Chemistry, Biology, and Clinical Significance of Endotoxins. Univ. of Chicago Press, Chicago Illinois.

Kauffman, F. (1961) Die Bakteriologie der *Salmonella* species. Munksgaard, Copenhagen.

Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E., and Romanowska, E. (1980) Structural studies of the O-specific side-chains of the *Shigella sonnei* Phase I lipopolysaccharide. *Carbohydr. Res.* **78:** 119-126.

Kenne, L., Lindberg, B., Unger, P., Holme, T., and Holmgren, J. (1979) Structural studies of the *Vibrio cholerae* O-antigen. *Carbohydr. Res.* 68: C16-C17.

Kenne, L., Lindberg, B., Unger, P., Holme, T., and Holmgren, J. (1982) Structural studies of the *Vibrio cholerae* O-antigen. *Carbohydr. Res.* **100:** 341-349.

Kennedy, N., Beutin, L., Achtman, M., Skurray, R., Rahmsdorf, U., and Herrich, P. (1979) Conjugation proteins of the F sex factor. *Nature* **270:** 580-585.

Kent, J.L. and Osborn, M.J. (1968a) Properties of the O-specific hapten formed *in vivo* by mutant strains of *Salmonella typhimurium*. *Biochemistry* **7:** 4396-4408.

Kent, J.L. and Osborn, M.J. (1968b) Haptenic O-antigen as a polymeric intermediate of *in vivo* synthesis of lipopolysaccharide by *Salmonella typhimurium*. *Biochemistry* **7:** 4419-4422.

Kido, N., Ohta, M., Iida, K., Hasegawa, T., Ito, H., Arakawa, Y., Komatsu, T., and Kato, N. (1989) Partial deletion of the cloned *rfb* gene of *Escherichia coli* O9 results in synthesis of a new O-antigenic lipopolysaccharide. *J. Bacteriol.* **171:** 3629-3633.

Kimura, A. and Hansen, E.J. (1986) Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. *Infect. Immun.* **51:** 69-79.

Koenig, S. and Hoffmann, M.K. (1979) Bacterial lipopolysaccharide activates suppressor B lymphocytes. *Proc. Natl. Acad. Sci. USA* **76:** 4608-4612.

Koller, T.H., Sogo, J.M., and Bujard, H. (1974) An electron microscopic method for studying nucleic acid-protein complexes. Visualization of RNA polymerase bound to the DNA of bacteriphages T7 and T3. *Biopolymers* **13:** 995-1009.

Kondo, S., Iguchi, T., and Hisatsune, K. (1980) Lipopolysaccharides of group F, a new group of vibrios. *Biochem. Biophys. Res. Commun.* **97:** 437-442.

Kondo, S., Iguchi, T., and Hisatsune, K. (1988) A comparative study of the sugar composition of lipopolysaccharides isolated from *Vibrio cholerae*, *"Vibrio albensis"* and *Vibrio metschnikovii*. *J. Gen. Microbiol.* **134:** 1699-1705.

Kondo, S., Zähringer, Rietschel, E.Th., and Hisatsune, K. (1989) Isolation and identification of 3-deoxy-D-threo-hexulosonic acid as a constituent of the lipopolysaccharide of *Vibrio parahaemolyticus* serotypes O7 and O12. *Carbohydr. Res.* **188:** 97-104.

Koop, A.H., Hartley, M.E., and Bourgeois, S. (1987) A low-copy-number vector utilizing β-galactosidase for the analysis of gene control elements. *Gene* **52:** 245-256.

Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol. Rev.* **47:** 1-45.

Kuo, T.-T. and Stocker, B.A.D. (1972) Mapping of the *rfa* genes in *Salmonella typhimurium* by ES18 and P22 transduction and by conjugation. *J. Bacteriol.* **112:** 48-57.

Kustu, S., Santero, E., Keener, J., Popham, D., and Weiss, D. (1989) Expression of σ^{54} (*ntrA*)-dependent genes is probably united by a common mechanism. *Microbiol. Rev.* **53:** 367-376.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105-132.

L'vov, V.-L., Shashkov, A.S., Dmitriev, B., Kochetkov, N.K., Jann, B., and Jann, K. (1984) Structural studies of the O-specific side-chain of the lipopolysaccharide from *Escherichia coli* O7. *Carbohydr. Res.* **126:** 249-259.

Le Dur, A., Chaby, R., and Szabo, L. (1980) Isolation of two protein-free and chemically different LPSs from *Bordetella pertussis* phenol-extracted endotoxin. *J. Bacteriol.* **143:** 78-88.

Levine, D.H. and Racker, E. (1959) Condensation of arabinose-5-phosphate and phosphoenolpyruvate by 2-keto-3-deoxy-8-phosphooctonic acid synthetase. *J. Biol Chem.* **234:** 2532-2539.

Levine, M.M., Kaper, J.B., Black, R.E., and Clements, M.L. (1983) New knowledge on pathogenesis of bacterial enteric infection as applied to vaccine development. *Microbiol. Rev.* **47:** 510-550.

Liang-Takasaki, C.-J., Grossman, N., and Leive, L. (1983) Salmonellae activate complement differentially via the alternative pathway depending on the structure of their lipopolysaccharide O-antigen. *J. Immun.* **130:** 1867-2215.

Lindahl,L., Archer, R.H., McCormick, J.R., Freedman, L.P., and Zengel, J.M. (1989) Translational coupling of the two proximal genes in the S10 ribosomal protein operon of *Escherichia coli*. *J. Bacteriol.* **171:** 2639-2645.

Lindberg, A.A., Wollin, R., Gemski, P., and Wohlhieter, J.A. (1978) Interaction between bacteriophage Sf6 and *Shigella flexneri* 1b, 3 and 4b. *Jpn. J. Microbiol.* **2:** 271-278.

Little, S., Hyde, S., Campbell, C.J., Lilley, R.J., and Robinson, M.K. (1989) Translational coupling in the threonine operon of *Escherichia coli* K-12. *J. Bacteriol.* **171:** 3518-3522.

Lockman, H., and Kaper, J.B. (1983) Nucleotide sequence analysis of the A2 and B subunits of *Vibrio cholerae* enterotoxin. *J. Biol. Chem.* **258:** 13722-13726.

Lospalluto, T.J. and Finkelstein, R.A. (1972) Chemical and physical properties of cholera exo-enterotoxin (choleragen) and its spontaneously formed toxoid (choleragenoid). *Biochem. Biophys. Acta.* **257:** 158-166.

Lüderitz, O., Galanos, C., Lehmann, V., Nurimen, M., Rietschel, E.T., Rosenfelder, G., Simon, M., and Westphal, O. (1973) Lipid A: Chemical structure and biological activity. *J. Infect. Dis.* **128:** S17-S29.

Lüderitz, O., Jann, K., and Wheat, R. (1968) Somatic and capsular antigens of gram negative bacteria. pp. 105-227. *In:* Comprehensive Biochemistry, extracellular and supporting structures. Vol. 26A. (eds. Florkin, M. and Stotz, E.H.).

Lüderitz, O., Westphal, O., Staub, A.M., and Nikaido, H. (1971) Isolation and chemical and immunological characterization of bacterial lipopolysaccharides. pp. 145-233. *In:* Microbial toxins. (eds. Weinbaum, G., Kadis, S., and Ajl, S.J.) Vol4. Academic Press, New York.

Ludwig, D.S., Ribi, H.O., Schoolnik, G.K., and Kornberg, R.D. (1986) Two-dimensional crystals of cholera toxin B-subunit receptor complexes: Projected structure at 17- AA resolution. *Proc. Natl. Acad. Sci. USA* **83:** 8585-8588.

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.

Magasanik, B. (1982) Genetic control of nitrogen assimilation in bacteria. *Annu. Rev. Genet.* **16:** 135-168.

Majumdar, M., Mukherjee, A.K., Guhatakurta, B., Dutta, A., and Sasmal, D. (1983) Structural investigation on the lipopolysaccharide isolated from *Vibrio cholerae* G-2102. *Carbohydr. Res.* **108:** 269-278.

Mäkelä, P.H. (1966) Genetic determination of the O-antigens of *Salmonella* groups B (4, 5, 12) and C_1 (6, 7). *J. Bacteriol.* **91:** 1115-1125.

Mäkelä, P.H., Jahkola, M., and Lüderitz, O. (1970) A new gene cluster *rfe* concerned with the biosynthesis of *Salmonella* lipopolysaccharide. *J. Gen. Microbiol.* **60:** 91-95.

Mäkelä. P.H. and Mayer, H. (1974) Enterobacterial common antigen. *Bacteriol. Rev.* **40:** 591-632.

Mäkelä, P.H. and Stocker, B.A.D. (1969) Genetics of polysaccharide biosynthesis. *Annu. Rev. Genetics* **3:** 291-322.

Mäkelä, P.H. and Stocker, B.A.D. (1984) pp. 59-137. *In:* Handbook of endotoxin. Vol. I. (ed. Rietschel, E.T.) Elsevier Science Publishers, Amsterdam.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning: - A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Manning, P.A. (1987) Involvement of cell envelope components in the pathogenesis of *Vibrio cholerae*: targets for cholera vaccine development. *Vaccine* **5:** 83-87.

Manning, P.A. (1988) Molecular genetic approaches to the study of *Vibrio cholerae*. *Microbiol. Sci.* **5:** 196-201.

Manning, P.A and Haynes, D.R. (1984) A common-immunogenic *Vibrio* outer membrane protein. *FEMS. Microbiol. Lett.* **24:** 297-302.

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 lipopolysccharide of *Vibrio cholerae* O1 and their potential for vaccine development. *Infect. Immun.* **53:** 272-277.

Manning, P.A., Morelli, G., and Fisseau, C. (1984) RNA-polymerase binding sites within the *tra* region of the F factor of *Escherichia coli* K-12. *Gene* **27:** 121-131.

Mattheakis, L.C. and Nomura, M. (1988) Feedback regulation of the *spc* operon in *Escherichia coli*: translational coupling and mRNA processing. *J. Bacteriol.* **170:** 4484-4492.

McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes *Annu. Rev. Biochem.* **54:** 171-204.

McConnell, M. and Wright, A. (1979) Variation in the structure and bacteriophage-inactivating capacity of *Salmonella anatum* lipopolysaccharide as a function of growth temperature. *J. Bacteriol.* **137:** 746-751.

McGhee, J.R. Kiyono, H., Michalek, S.M., Babb, J.L., Rosenstretch, D.L., and Mergenhagen, S.E. (1980) Lipopolysaccharide (LPS) regulation of the immune response: T lymphocytes from normal mice suppress mitogenic and immunogenic responses to LPS. *J. Immun.* **124:** 1603-1611.

McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C., and Rosenberg, M. (1981) A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase. pp. 384-415. *In:* Gene amplification and Analysis. Vol. II: Analysis of nucleic acids by enzymatic methods. (eds Chirikijian, J.C. and Papas, T.S.) Elsevier, North Holland.

Mekalanos, J.J. (1985) Cholera toxin: genetic analysis, regulation and role in pathogenesis. *Curr. Top. Microbiol. Immunol.* **18:** 97-118.

Mekalanos, J.J., Collier, R.J., and Romig, W.R. (1979) Enzymatic activity of cholera toxin. II. Relationships to proteolytic processing, disulfide bond reduction, and subunit composition. *J. Biol. Chem.* **254:** 5855-5861.

Mekalanos, J.J., Moseley, S.L., Murphy, J.R., and Falkow, S. (1982) Isolation of enterotoxin structural gene deletion mutations in *Vibrio cholerae* induced by two mutagenic vibriophages. *Proc. Natl. Acad. Sci. USA* **79:** 151-155.

Mekalanos, J.J., Swartz, D.J., Pearson, G.D.N, Harford, N., Groyne, F. and de Wilde M. (1983) Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development. *Nature* **306:** 551-557.

Melchers, F.(1980) Interactions of bacteria with the immune system. *Life Sci. Res. Rep.* **6:** 285-305.

Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. *Gene* **19:** 269-276.

Meyer, T.F., Billyard, E., Haas, R., Störzbach, S., and So, M. (1984) Pilus genes of *Neisseria gonorrhoeae*: chromosomal organization and DNA sequence. *Proc. Natl. Acad. Sci. USA* **81:** 6110-6114.

Miller J. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, New York.

Miller, V.L. (1985) Analysis of the cholera toxin positive regulatory gene, *toxR*. Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Miller, V.L., DiRita, V.J., and Mekalanos, J.J. (1989) Identification of *toxS*, a regulatory gene whose product enhances ToxR-mediated activation of the cholera toxin promoter *J. Bacteriol.* **171:** 1288-1293.

Miller, V.L. and Mekalanos, J.J. (1984) Synthesis of cholera toxin is positively regulated at the transcriptional level by *toxR*. *Proc. Natl. Acad. Sci. USA* **81:** 3471-3475.

Miller, V.L. and Mekalanos, J.J. (1985) Genetic analysis of the cholera toxin positive regulatory gene *toxR*. *J. Bacteriol.* **163:** 580-585.

Miller, V.L., Taylor, R.K., and Mekalanos, J.J. (1987) Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. *Cell* **48:** 271-279.

Mills, S.D., Sekizaki, T., Gonzalez-Carrero M.I., and Timmis, K.N. (1988) Analysis and genetic manipulation of *Shigella* virulence determinants for vaccine development. *Vaccine* **6:** 116-122.

Monsur, K.A., Rizvi, S.S.H., Huq, M.I. and Beneson, A.S. (1965) Effect of Mukerjee's Group IV phage on El Tor vibrios. *Bull. W.H.O.* **32:** 211-216.

Morelli, G., Buhk, H-J., Fisseau, C., Lother, H., Yoshinaga, K., and Messer, W. (1981) Promoters in the region of the *E. coli* replication origin. *Molec. Gen. Genet.* **184:** 255-259.

Morona, R., Matthews, M., Morona, J.K., and Brown, M.H. (1990) Regions of the cloned *Vibrio cholerae rfb* genes needed to determine the Ogawa form of the O-antigen. *Molec. Gen. Genet.* **224:** 405-412.

Morrison, D.C. and Kline, L.F. (1977) Activation of the classical and properdin pathways of complement by bacterial lipopolysaccharides (LPS). *J. Immunol.* **118:** 362-369.

Mühlradt, P.F. (1970) Biosynthesis of *Salmonella* lipopolysaccharide. Studies on the transfer of glucose, galactose and phosphate to the core in a cell-free system. *Eur. J. Biochem.* **18:** 20-27.

Mukerjee, S. and Roy, G. (1961) Evaluation of tests for differentiating *Vibrio cholerae* and El Tor vibrios. *Annu. Biochem. Exp. Med.* **21:** 129-132.

Mulford, C.A. and Osborn, M.J. (1983) An intermediate step in the translocation of lipopolysaccharide to the outer membrane of *Salmonella typhimurium. Proc. Natl. Acad. Sci. USA* **80:** 1159-1163.

Munford, R.S., Hall, C.L., and Rick, P.D. (1980) Size heterogeneity of *Salmonella typhimurium* lipopolysaccharides in outer membranes and culture supernatant membrane fragments. *J. Bacteriol.* **144:** 630-640.

Munson, R.S., Jr., Rasmussen, N.S., and Osborn, M.J. (1978) Biosynthesis of lipid A: Enzymatic incorporation of 3-deoxy-D-mannooctulosonate into a precursor of lipid A in *Salmonella typhimurium. J. Biol. Chem.* **253:** 1503-1511.

Naide, Y., Nikaido, H., Mäkelä, P.H., Wilkinson, R.G., and Stocker, B.A.D. (1965) Semirough strains of *Salmonella. Proc. Natl. Acad. Sci. USA* **53:** 147-153.

Nalin, D.R., Levine, R.J., Levine, M.M., Hoover, D., Bergquist, E., McLaughlin, J., Libonati, J., Alam, J. and Hornick, R.B. (1978) Cholera, non-vibrio cholera and stomach acid. *Lancet* **ii:** 856-859.

Niedhardt, F.C. (1987) Chemical composition of *Escherichia coli* pp. 3-7. *In: Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Vol I. (ed Neidhardt, F.C.) American Society for Microbiology.

Neoh, S.H. and Rowley, D. (1970) The antigens of *Vibrio cholerae* involved in vibriocidal action of antibody and complement. *J. Infect. Dis.* **121:** 505-513.

Neoh, S.H. and Rowley, D. (1972) Protection of infant mice against cholera by antibodies to three antigens of *Vibrio cholerae. J. Infect. Dis.* **126:** 41-47.

Newland, J.W., Green, B.A., Foulds, J., and Holmes, R.K. (1985) Cloning of extracellular DNase and construction of a DNase-negative strain of *Vibrio cholerae*. *Infect. Immun.* **47:** 691-696.

Nikaido, H. (1970) Structure of the cell wall lipopolysaccharide from *Salmonella typhimurium*: further studies on the linkage between O-side chains and R-core. *Eur. J. Biochem.* **15:** 57-62.

Nikaido, H. (1973) Biosynthesis and assembly of lipopolysaccharide and the outer membrane layer of Gram negative cell wall. pp. 131-208. *In:* Bacterial Membranes and Walls. Vol. I. (ed. Leive, L.) Marcel Dekker, New York.

Nikaido, H. (1979) Nonspecific transport through the outer membrane. pp. 361-407 *In:* Bacterial outer membranes. (ed. Inouye, M.) John Wiley and Sons, Inc., New York.

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

Nikaido, H., Levinthal, M., Nikaido, K., and Nakane, K. (1967) Extended deletions in the histidine-rough -B region of the *Salmonella* chromosome. *Proc. Natl. Acad. Sci. USA* **57:** 1825-1832.

Nobechi, K. (1923) Immunological studies upon types of *Vibrio cholerae*. *Sci. Rep. Inst. Infect. Dis.* Tokyo Univ. **2:** 43.

Ogg, J.E., Ogg, B.J., Shrestha, M.B. and Poudaryl, L. (1979) Antigenic changes in *Vibrio cholerae* biotype El Tor serotype Ogawa after bacteriophage infection. *Infect. Immun.* **24:** 974-978.

Ogg, J.E., Shrestha, M.B., and Poudaryl, L. (1978) Phage-induced changes in *Vibrio cholerae*: serotype and biotype conversions. *Infect. Immun.* **19**: 231-238.

Ogg, J.E., Timme, T.L., and Alemohammad, M.M. (1981) General transduction in *Vibrio cholerae*. *Infect. Immun.* **31:** 737-741.

Oppenheim, S.D. and Yanofsky, C. (1980) Translational coupling during expression of the tryptophane operon of *Escherichia coli*. *Genetics* **95:** 785-795.

Ørskov, F., and Ørskov, I. (1962) Behaviour of *E. coli* antigens in sexual recombination. *Acta Pathol. Microbiol. Scand.* **55:** 99-109.

Osborn, M.J. (1963) Studies on the gram-negative cell wall, I. Evidence for the role of 2-keto-3-deoxyoctonate in the lipopolysaccharide of *Salmonella typhimurium*. *Proc. Natl. Acad. Sci. USA* **50:** 499-506.

Osborn, M.J. (1968) Biochemical characterization of mutants of *Salmonella typhimurium* lacking glucosyl or galactosyl lipopolysaccharide transferases. *Nature*. **217:** 957-960.

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. and D' Ari, L. (1964) Enzymatic incorporation of N-acetylglucosamine into cell wall lipopolysaccharide in a mutant strain of *Salmonella typhimurium*. *Biochem. Biophys. Res. Commun.* **16:** 568-574.

Osborn, M.J., Gander, J.E., and Parisi, E. (1972) Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. Site of synthesis of lipopolysaccharide. *J. Biol. Chem.* **247:** 3973-3980.

Osborn, M.J. and Rothfield, L.I. (1971) Biosynthesis of the core region of lipopolysaccharide. pp. 331-350. *In:* Microbial Toxins. Vol. IV: Bacterial endotoxins. (eds. Weinbaum G., Kadis S., and Ajl, A.) Academic Press, Inc., New York.

Osborn, M.J., Rosen, S.M., Rothfield, L., and Horecker, B.L. (1962) Biosynthesis of bacterial lipopolysaccharide. I. Enzymatic incorporation of galactose in a mutant strain of *Salmonella*. *Proc. Natl. Acad. Sci. USA* **48:** 1831-1838.

Palermo, D.A., Evans, T.M., and Clark, V.L. (1987) Expression of a cloned lipopolysaccharide antigen from *Neisseria gonorrhoeae* on the surface of *Escherichia coli* K-12. *Infect. Immun.* **55:** 2844-2849.

Palva, E.T. and Mäkelä, P.H. (1980) Lipopolysaccharide heterogeneity in *Salmonella typhimurium* analyzed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis. *Eur. J. Biochem.* **107:** 137-143.

Parker, C., Gauthier, D., Tate, A., Richardson, K., and Romig, W.R. (1979) Expanded linkage map of *Vibrio cholerae*. *Genetics* **91:** 191-214.

Parker, C. and Romig, W.R. (1972) Self transfer and genetic recombination mediated by P, the sex factor of *Vibrio cholerae*. *J. Bacteriol.* **112:** 707-714.

Pearson, G.D.N., and Mekalanos, J.J. (1982) Molecular cloning of the *Vibrio cholerae* enterotoxin genes in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* **79:** 2976-2980.

Pegues, J.C., Chen, L., Gordon, A.W., Ding, L., and Coleman, W.G., Jr. (1990) Cloning, expression, and characterization of the *Escherichia coli* K-12 *rfaD* gene. *J. Bacteriol.* **172:** 4652-4660.

Peterson, J.W., Lospalluto, J.J., and Finkelstein, R.A. (1972) Localization of cholera toxin in vivo. *J. Infect. Dis.* **126:** 617-628.

Peterson, K.M. and Mekalanos, J.J. (1988) Characterization of the *Vibrio cholerae* ToxR regulon: Identification of novel genes involved in intestinal colonization. *Infect. Immun.* **56:** 2822-2829.

Peterson, K.M., Taylor, R.K., Miller, V.L., Di Rita, V.J., and Mekalanos, J.J. (1988) Coordinate regulation of virulence determinants in *Vibrio cholerae*. pp. 165-172. *In:* Molecular Biology of Infectious Diseases. (ed. Shwartz, M.) Elsevier, Paris.

Petricoin III, E.F. and Stein, D.C. (1989) Molecular analysis of lipopolysaccharide biosynthesis in *Neisseria gonorrhoeae*. *Infect. Immun.* **57:** 2847-2852.

Pierce, N.F., Cray, W.C., Jr,. and Sacci, J.B., Jr. (1982) Oral immunization of dogs with purified cholera toxin, crude cholera toxin, or B subunit: evidence for synergistic protection by antitoxic and antibacterial mechanisms. *Infect. Immun.* **37:** 687-694.

Pierce, N.F., Kaper, J.B., Mekalanos, J.J., and Cray, W.C. (1985) Role of cholera toxin in enteric colonization by *Vibrio cholerae* O1 in rabbits. *Infect. Immun.* **50:** 813-816.

Pohlner, J., Meyer, T.F., Jalajakumari, M.B., and Manning, P.A. (1986) Nucleotide sequnce of ompV, structural gene for a major *Vibrio cholerae* outer membrane protein. *Molec. Gen. Genet.* **205:** 494-500.

Portmann, R. and Koller, T. (1976) The divalent cation method for protein-free spreading of nucleic acid molecules. pp. 546-549. Sixth European Congress Electron Microscopy Jerusalem. Vol. 2. Tal International Publishing Co.

Pribnow, D. (1975) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. USA* **72:** 784-788.

Pridmore, R.D. (1987) New and versatile cloning vectors with kanamycin-resistance marker. *Gene* **56:** 309-312.

Raetz, C.R.H. (1987) Structure and Biosynthesis of Lipid A in *Escherichia coli.* In: *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Vol.I. (ed. Neidhardt, F.C.) American Society for Microbiology.

Raetz, C.R.H. (1990) Biochemistry of endotoxins *Annu. Rev. Biochem.* **59:** 1298-170.

Ratcliff, S.W., Luh, J., Ganesen, A.T., Behrens, B., Thompson, R., Montenegro, M., Morelli, G., and Trautner, T.A. (1979) The genome of *B. subtilis* phage SPP1; the arrangement of restriction endonuclease generated fragments. *Molec. Gen. Genet.* **168:** 165-172.

Raziuddin, S. (1977) Studies of the polysaccharide fraction form the cell wall lipopolysaccharides (O-antigen) of *Vibrio cholerae. Ind. J. Biochem. Biophys.* **14:** 262-263.

Raziuddin, S. (1980) Immunochemical studies of the lipopolysaccharides of *Vibrio cholerae*: constitution of O-specific side chain and core polysaccharide. *Infect. Immun.* **27:** 211-215.

Redmond, J.W. (1975) 4-amino-1,6-dideoxy-D-mannose (D-perosamine): a component of the lipopolysaccharide of *Vibrio cholerae* 569B (Inaba). *FEBS Lett.* **50:** 147-149.

Redmond, J.W. (1978) The 4-amino sugars present in the lipopolysaccharides of *Vibrio cholerae* and related Vibrios. *Biochem. Biophys. Acta.* **542:** 378-384.

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 characterization of an acid-labile antigenic determinant. *Aust. J. Exp. Biol. Med. Sci.* **51:** 229-235.

Rick, P.D. (1987) Lipopolysaccharide Biosynthesis. pp. 648-643. *In: Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Vol I. (ed Neidhardt, F.C.) American Society for Microbiology.

Rick, P.D. and Osborn, M.J. (1977) Lipid A mutants of *Salmonella typhimurium*. Characterization of a conditional lethal mutant in 3-deoxy-D-mannooctulosonate-8-phosphate synthetase. *J. Biol. Chem.* **252:** 4895-4905.

Rick, P.D. and Young, D.A. (1982) Isolation and characterization of a temperature-sensitive lethal mutant of *Salmonella typhimurium* that is conditionally defective in 3-deoxy-D-manno-octulosonate-8-phosphate synthesis. *J. Bacteriol.* **150:** 447-455.

Rietschel, E.Th (1976) Absolute configuration of 3-hydroxy fatty acids present in lipids from various bacterial groups. *Eur J. Biochem* **64:** 423-428.

Rietschel, E.Th. (ed.) (1984) Handbook of endotoxin. Vol. 1: Chemistry of endotoxin. Elsevier Biomedical Press, Amsterdam.

Rietschel, E.Th., Wollenweber, H-W., Sidorczyk, Z., Zähringer, U., and Lüderitz, O. (1983) Analysis of the primary structure of lipid A. pp. 195-219. *In:* Bacterial lipopolysaccharides. Structure, synthesis and biological activites. (ed Anderson, L. and Unger, F.M.) ACS Symposium Series 231. American Chemical Society, Washington D.C.

Riley, L.W., Junio, L.N., Libaek, L.B., and Schoolnik, G,K, (1987) Plasmid encoded expression of lipopolysaccharide O-antigenic polysaccharide in enteropathogenic *Escherichia coli. Infect. Immun.* **55:** 2052-2056.

Robbins, P.W., Wright, A., and Dankert, M. (1966) Polysaccharide biosynthesis. *J. Gen. Physiol.* **49:** 331-346.

Rood, J.I. and Gawthorne, J.M. (1984) Apple software for analysis of the size of restriction fragments. *Nucl. Acids. Res.* **12:** 689-694.

Rosenberg, M., Chepelinsky, A.B., and McKenney, K. (1983) Studying promoters and terminators by gene fusion. *Science* **222:** 734-739.

Rosner, M.R., Tana, J.Y., Barzilay, I., and Khorana, H.G. (1979) Structure of the lipopolysaccharide from an *Escherichia coli* heptose-less mutant I. Chemical degradations and identification of products. *J. Biol Chem* **254:** 5906-5917.

Rothfield, L., Osborn, M.J., and Horecker, B.L. (1964) Biosynthesis of bacterial lipopolysaccharide. II. Incorporation of glucose and galactose catalyzed by particulate and soluble enzymes. *J. Biol. Chem.* **239:** 2788-2795.

Roy, C., Mridha, K. and Mukerjee, S. (1965) Action of polymyxin on cholera vibrios. Techniques of determination of polymyxin-sensitivity. *Proc. Soc. Exp. Biol. Med.* **119:** 893-896.

Sack, R.B. and Miller, C.E. (1969) Progressive changes of vibrio serotypes in germ-free mice infected with *Vibrio cholerae. J. Bacteriol.* **99:** 688-695.

Sakazaki, R. and Tamura, K. (1971) Somatic antigen variation in *Vibrio cholerae. Jpn. J. Med. Sci. Biol.* **24:** 93-100.

Sanderson, K.E. and Hurley, J.A. (1987) Linkage map of *Salmonella typhimurium*. pp. 817-919. *In: Escherichia coli* and *Salmonella typhimurium*. Cellular and Molecular Biology. Vol. I. (ed. Neidhardt, F.C.) American Society for Microbiology.

Sanderson, K.E. and Roth, J.R. (1983) Linkage map of *Salmonella typhimurium*, edition VI. *Microbiol. Rev.* **47:** 410-453.

Sanderson, K.E. and Saeed, H. (1972a) Insertion of the F factor into the cluster of *rfa* (rough A) genes of *Salmonella typhimurium*. *J. Bacteriol.* **112:** 64-73.

Sanderson, K.E. and Saeed, H. (1972b) P22-mediated transduction analysis of the rough A (*rfa*) region of the chromosome of *Salmonella typhimurium*. *J. Bacteriol.* **112:** 58-63.

Sanderson, K.E. and Stocker, B.A.D. (1981) Gene *rfaH*, which affects lipopolysaccharide core structure in *Salmonella typhimurium*, is required also for expression of F-factor functions. *J. Bacteriol.* **146:** 535-541.

Sanderson, K.E., Van Wyngaarden, J., Lüderitz, O., and Stocker, B.A.D. (1974) Rough mutants of *Salmonella typhimurium* with defects in the heptose region of the lipopolysaccharide core. *Can. J. Microbiol.* **20:** 1127-1134.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143:** 161-178.

Sanger, F., Nickelen, S., and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463-5467.

Sansonetti, P.J., Kopecko, D.J., and Formal, S.B. (1981) *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. *Infect. Immun.* **34:** 75-83.

Schaller, H., Gray, C., and Herrman, K. (1975) Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc. Natl. Acad. Sci. USA* **72:** 737-741.

Scher, M., Lennarz, W.J., and Sweeley, C.C. (1968) The biosynthesis of mannosyl-1-phosphoryl-polyisoprenol in *Micrococcus lysodeikticus* and its role in mannan synthesis. *Proc. Natl. Acad. Sci. USA* **59:** 1313-1320.

Schlessinger, D. (1977) Bacterial antigens and host response. *Microbiology* **1977:** 219-326.

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.

Schmidt, G., Jann, B., and Jann, K. (1970) Immunochemistry of R lipopolysaccharides of *Escherichia coli*. Studies on R mutants with an incomplete core, derived from *E. coli* O8: K27. *Eur. J. Biochem.* **16:** 382-392.

Schmidt, G., Mayer, H., and Mäkelä, P.H. (1976) Presence of *rfe* genes in *Escherichia coli*: their participation in biosynthesis of O-antigen and enterobacterial common antigen. *J. Bacteriol.* **127:** 755-762.

Schmitz, A. and Galas, D.J. (1979) The interaction of RNA polymerase and lac repressor with the lac control region. *Nucl. Acids. Res.* **6:** 111-132.

Schneider, H., Hale, T.L., Zollinger, W.D., Seid, R.C., Jr., Hammack, C.A., and Griffiss, J.M. (1984) Heterogeneity of molecular size and antigenic expression within lipooligosaccharides of individual strains of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Infect. Immun.* **45:** 544-549.

Schneider, K. and Beck, C.F. (1986) Promoter-probe vectors for the analysis of divergently arranged promoters. *Gene* **42:** 37-48.

Schneider, D.R. and Parker, C.D. (1978) Isolation and characterization of protease-deficient mutants of *Vibrio cholerae*. *J. Infect. Dis.* **138:** 143-151.

Schumperli, D., McKenney, K., Sobieski, D., and Rosenberg, M. (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell* **30:** 865-871.

Sears, S.D., Richardson, K., Young, C., Parker, C.D., and Levine, M.M. (1984) Evaluation of the human immune response to outer membrane proteins of *Vibrio cholerae*. *Immun. Infect.* **44:** 439-444.

Seid, R.C., Kopecko, D.J., Sadoff, J.C., Schneider, H., Baron, L.S., and Formal, S.B. (1984) Unusual lipopolysaccharide antigens of a *Salmonella typhi* oral vaccine strain expressing the *Shigella sonnei* Form I antigen. J. Biol. Chem. **259:** 9028-9034.

Sen, A.K. and Mukherjee, A.K. (1978) Structural studies of a specific polysaccharide isolated from non-agglutinable *Vibrio*. *Carbohydr. Res.* **64:** 215-223.

Sen, A.K., Mukherjee, A.K., Guhatharkurta, B., Dutta, A., and Sasnal, D. (1979) Stuctural investigation on the lipopolysaccharide isolated from *Vibrio cholerae* Inaba 569B. *Carbohydr. Res.* **72:** 191-199.

Sen, A.K., Mukherjee, A.K., Guharthakurta, B., Dutta, A., and Sasmal, D. (1980) Studies on the partial structure of the O-antigen of *Vibrio cholerae* 569B. *Carbohydr. Res.* **86:** 113-121.

Shephard, H.M., Yelverton, E., and Goeddel, D.V. (1982) Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in ribosome-binding sites. *DNA* **1:** 125-131.

Shewan, J.M., and Véron, M. (1974) Genus I. *Vibrio*. pp. 74-77. *In:* Bergey's Manual of Determinative Bacteriology. 8th edn. (eds. Buchanan, R.E. and Gibbons, N.E.) Williams and Wilkins, Baltimore.

Shewan, J.M. and Véron. M. (1975) The genus Vibrio. pp. 340-345. *In:* Bergey's Manual of Determinative Bacteriology. 8th edn. (eds. Buchanan, R.E. and Gibbons, N.E.) Williams and Wilkins, Baltimore.

Shine, J. and Dalgarno, L. (1974) The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71:** 1342-1346.

Shousha, A.T. (1931) Group agglutination reaction in cholera. A contribution to the identification of *V. cholerae*. *J. Egypt. Med. Ass.* **14:** 438-453.

Shrivastava, D.L. and White, P.B. (1947) Note on the relationship of the so-called Ogawa and Inaba types of *V. cholerae*. *Ind. J. Med. Res.* **35:** 117-129.

Siebenlist, U., Simpson, R.B., and Gilbert, W. (1980) *E. coli* RNA polymerase interacts homologously with 2 different promoters. *Cell* **20:** 269-281.

Simmons, D.A.R. (1971) Immunochemistry of *Shigella flexneri* O-antigens: a study of structural and genetic aspects of the biosynthesis of cell surface antigens. *Bacteriol. Rev.* **35:** 117-148.

Simpson, R.B. (1982) Evidence for close contacts between RNA polymerase and promoter DNA. pp. 164-180. *In:* Promoters: structure and function. (eds. Rodriguez, R.L. and Chamberlin, M.L.) Praeger publishers, New York.

Smit, J., Kamio, Y., and Nikaido, H. (1975) Outer membrane of *Salmonella typhimurium*: chemical analysis and freeze-fracture studies with lipopolysaccharide mutants. *J. Bacteriol.* **124:** 942-958.

Sogo, J.M., Inciarte, J.C., Viñuela, E., and Salas, M. (1979) RNA polymerase binding sites and transcription map of the DNA of *Bacillus subtilis* phage ø29. *J. Mol. Biol.* **127:** 411-436.

Srisart, P., Reynolds, B.L., and Rowley, D. (1985) The correlation between serum IgA antibody levels and resistance to infection with *Salmonella typhimurium* after oral immunization with various Salmonellae. *Aust. J. Exp. Biol. Med. Sci.* **63:** 177-182.

Stefano, J.E. and Gralla, J.D. (1982) Specific and non-specific interactions at mutant *lac* promoters. pp. 69-79. In Promoters: Structure and function. (eds. Rodriguez, R.L. and Chamberlin) M.J. Praeger Publishers, New York.

Stocker, B.A.D. (1988) Auxotrophic *Salmonella typhi* as a live vaccine. *Vaccine* **6:** 141-145.

Stocker, B.A.D. and Mäkelä, P.H. (1978) Genetics of the (gram-negative) bacterial surface. *Proc. Roy. Soc. Lond. B.* **202:** 5-30.

Sturm, S. and Timmis, K.N. (1986) Cloning of the *rfb* gene region of *Shigella dysenteriae* 1 and construction of an *rfb-rfp* gene cassette for the development of lipopolysaccharide-based live anti-dysentry vaccines. *Microbial Pathogenesis*. **1:** 289-297.

Sturm, S., Jann, B., Jann, K., Fortnagel, P., and Timmis, K.N. (1986a) Genetic and biochemical analysis of *Shigella dysenteriae* 1 O-antigen polysaccharide biosynthesis in *Escherichia coli* K12: 9kb plasmid of *S. dysenteriae* 1 determines the addition of a galactose residue to the lipopolysaccharide core. *Microbial Pathogenesis*. **1:** 299-306.

Sturm, S., Jann, B., Jann, K., Fortnagel, P., and Timmis, K.N. (1986b) Genetic and biochemical analysis of *Shigella dysenteriae* 1 O-antigen polysaccharide biosynthesis of the *rfb* gene cluster. *Microbial Pathogenesis*. **1:** 307-324.

Sublett, R.D. and Romig, W.R. (1981) Transposon facilitated recombination in classical biotypes of *Vibrio cholerae*. *Infect. Immun.* **32:** 1132-1138.

Svennerholm, A.M. (1975) Experimental studies on cholera immunization. IV. The antibody response to formalinized *Vibrio cholerae* and purified endotoxin with special reference to protective capacity. *Int. Arch. Allergy. Appl. Immun.* **49:** 434-452.

Tabor, S. and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase *Proc. Natl. Acad. Sci. USA* **84:** 4767-4771.

Tachibana, H. and Ishihama, A. (1985) Correlation between the rate of transcription initiation and the strand-melting property of *Escherichia coli* promoters. *Nucl. Acids. Res.* **13:** 9031-9042.

Tacket, C.O., Forrest, B., Morona, R., Attridge, S.R., LaBrooy, J., Tall, B.D., Reymann, M., Rowley, D., and Levine, M.M. (1990) Safety, immunogenicity, and efficacy against cholera challenge in man of a typhoid-cholera hybrid vaccine derived from *Salmonella typhi* Ty21a. *Infect. Immun.* **58:** 1620-1627.

Sturm, S. and Timmis, K.N. (1986) Cloning of the *rfb* gene region of *Shigella dysenteriae* 1 and construction of an *rfb-rfp* gene cassette for the development of lipopolysaccharide-based live anti-dysentry vaccines. *Microbial Pathogenesis*. **1:** 289-297.

Sturm, S., Jann, B., Jann, K., Fortnagel, P., and Timmis, K.N. (1986a) Genetic and biochemical analysis of *Shigella dysenteriae* 1 O-antigen polysaccharide biosynthesis in *Escherichia coli* K12: 9kb plasmid of *S. dysenteriae* 1 determines the addition of a galactose residue to the lipopolysaccharide core. *Microbial Pathogenesis*. **1:** 299-306.

Sturm, S., Jann, B., Jann, K., Fortnagel, P., and Timmis, K.N. (1986b) Genetic and biochemical analysis of *Shigella dysenteriae* 1 O-antigen polysaccharide biosynthesis of the *rfb* gene cluster. *Microbial Pathogenesis*. **1:** 307-324.

Sublett, R.D. and Romig, W.R. (1981) Transposon facilitated recombination in classical biotypes of *Vibrio cholerae*. *Infect. Immun.* **32:** 1132-1138.

Svennerholm, A.M. (1975) Experimental studies on cholera immunization. IV. The antibody response to formalinized *Vibrio cholerae* and purified endotoxin with special reference to protective capacity. *Int. Arch. Allergy. Appl. Immun.* **49:** 434-452.

Tabor, S. and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase *Proc. Natl. Acad. Sci. USA* **84:** 4767-4771.

Tachibana, H. and Ishihama, A. (1985) Correlation between the rate of transcription initiation and the strand-melting property of *Escherichia coli* promoters. *Nucl. Acids. Res.* **13:** 9031-9042.

Tacket, C.O., Forrest, B., Morona, R., Attridge, S.R., LaBrooy, J., Tall, B.D., Reymann, M., Rowley, D., and Levine, M.M. (1990) Safety, immunogenicity, and efficacy against cholera challenge in man of a typhoid-cholera hybrid vaccine derived from *Salmonella typhi* Ty21a. *Infect. Immun.* **58:** 1620-1627.

Takegi, M., Takeda, H., and Imanaka, T. (1990) Nucleotide sequence and cloning in *Bacillus subtilis* of the *Bacillus stearothermophilus* pleiotrophic regulatory gene *degT*. *J. Bacteriol.* **172:** 411-418.

Taylor, R.K., Miller, V.L., Furlong, D.B. and Mekalanos, J.J. (1987a) Use of *pho*A gene fusions to identify a pilus colonization factor co-ordinately regulated with cholera toxin. *Proc. Natl. Acad. Sci. USA* **84:** 2833-2837.

Taylor, R.K., Spears, P., and Mekalanos, J.J. (1987b) Applications of the Tcp pilus in cholera vaccine development. pp. 105-106. *In:* Abstracts of the 23rd Joint Conference on Cholera. U.S.-Japan Cooperative Medical Science Program.

Thomas, M.S. and Drabble, W.T. (1985) Nucleotide sequence and organization of the *gua* promoter region of *Escherichia coli*. *Gene* **36:** 45-53.

Thomas, M.S., Bedwell, D.M., and Nomura, M. (1987) Regulation of α operon gene expression in *Escherichia coli*: a novel form of translational coupling. *J. Mol. Biol.* **196:** 333-345.

Tijian, R., Stinchcomb, D., and Losick, R. (1975) Antibody directed against *Bacillus subtilis* σ factor purified by sodium dodecyl sulphate slab gel electrophoresis: Effect on transcription by RNA polymerase in crude extracts of vegetative and sporulating cells. *J. Biol. Chem.* **250:** 8824-8828.

Tolan, R.W., Jr., Munson, R.S., and Granoff, D.M. (1986) Lipopolysaccharide gel profiles of *Haemophilus influenzae* type b are not stable epidemiologic markers. *J. Clin Microbiol.* **24:** 223-227.

Tsai, C.M. and Frasch, C.E. (1982) A sensitive silver stain for detecting lipopolysaccharides in polacrylamide gels. *Anal. Biochem.* **119:** 115-119.

Ubben, D. and Schmitt, R. (1986) Tn*1721* derivatives for transposon mutagensis, restriction mapping and nucleotide sequence analysis. *Gene* **41:** 145-152.

Valvano, M.A. and Crosa, J.H. (1989) Molecular cloning and expression in *Escherichia coli* K12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E. coli* O7:K1. *Infect. Immun.* **57:** 937-943.

Van Dongen, W.M.A.M., van Vlerken, M.M.A., and de Graaf, F.K. (1987) Nucleotide sequence of a DNA fragment encoding a *Vibrio cholerae* haemagglutinin. *Mol. Gen. (Life Sci. Adv.)* **6:** 85-91.

Van Heyningen, S. (1977) Cholera toxin. *Biol. Rev.* **52:** 509-549.

Verma, N.K. and Reeves, P.R. (1989) Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and D salmonellae. *J. Bacteriol.* **171:** 5694-5701.

Verma, N.K., Quigley, N.B., and Reeves, P.R. (1988) O-antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. *J. Bacteriol.* **170:** 103-107.

Vieira, J. and Messing, J. (1982) The pUC plasmids, a M13mp-7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19:** 258-268.

Vollenweider, H.J. and Szbalski, W. (1978) Electron microscopic mapping of RNA polymerase binding to coliphage lambda DNA. *J. Mol. Biol.* **123:** 485-498.

Von Heijne, G. (1983) Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* **133:** 17-21.

Von Hippel, P.H., Bear, D.G., Winter, R.B., and Berg O.G. (1982) Molecular aspects of promoter function: an overview. pp. 3-33. *In:* Promoters: Structure and function. (eds. Rodriguez, R.L. and Chamberlin, M.J.) Praeger Publishers, New York.

Walenga, R.W. and Osborn, M.J. (1980a) Biosynthesis of lipid A. *In vivo* formation of an intermediate containing 3-deoxy-D-mannooctulosonate in a mutant of *Salmonella typhimurium*. *J. Biol. Chem.* **255:** 4252-4256.

Walenga, R.W. and Osborn, M.J. (1980b) Biosynthesis of lipid A. Formation of acyl-deficient lipopolysaccharides in *Salmonella typhimurium* and *Escherichia coli*. *J. Biol. Chem.* **255:** 4257-4263.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) Distantly related sequences in the α- and β- subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Eur. J. Mol. Biol.* **263:** 5804-5814.

Ward, H.M. and Manning, P.A. (1989) Mapping of chromosomal loci associated with lipopolysaccharide synthesis and serotype specificity in *Vibrio cholerae* O1 by transposon mutagenesis using Tn*5* and Tn*2680*. *Molec. Gen. Genet.* **218:** 367-370.

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.

Watanabe, H. and Timmis, K.N. (1984) A small plasmid in *Shigella dysenteriae* 1 specifies one or more functions essential for O-antigen production and bacterial virulence. *Infect. Immun.* **43:** 391-396.

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-antigen groups of *Salmonella typhimurium. Proc. Natl. Acad. Sci. USA* **54:** 228-235.

Weiser, J.N., Lindberg, A.A., Manning, E.J., Hansen, E.J., and Moxon E.R. (1989a) Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in *Haemophilus influenzae. Infect. Immun.* **57:** 3045-3052.

Weiser, J.N., Love, J., and Moxon, E.R. (1989b) Molecular mechanism of phase-variation of *H. influenzae* lipopolysaccharide epitopes. *Cell* **59:** 657-665.

Weiser, J.N., Maskell, D.J., Butler, P.D., Lindberg, A.A., and Moxon, E.R. (1990) Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J. Bacteriol.* **172:** 3304-3309.

Westphal, O. and Lüderitz, O. (1954) Chemische erforschung von lipopolysaccharide gram-negative bakterien. *Anger. Chem.* **66:** 407-417.

Wierenga, R.K., Terpstra, P., and Hol, W.G.J. (1986) Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **187:** 101-107.

Wohlhieter, J.A., Datta, A., Brenner, D.J., and Baron, L.S. (1975) Homology between the deoxyribonucleic acid of fertility factor P and *Vibrio cholerae* chromosomal deoxyribonucleic acid. *J. Bacteriol.* **122:** 464-467.

Woisetschläger, M. and Högenauer, G. (1987) The *kdsA* gene coding for 3-deoxy-D-manno-octulosonic acid 8-phosphate synthetase is part of an operon in *E. coli*. *Molec. Gen. Genet.* **207:** 369-373.

Wollenweber, H.-W., Broady, K.W., Lüderitz, O., and Rietschel, E.Th. (1982) Further investigations of lipid A. Demonstration of amide-linked 3-acyl-oxyacyl residues in *Salmonella* and other lipid A's. *Eur. J. Biochem* **124:** 191-198.

Wollin, R., Creeger, E.S., Rothfield, L.I., Stocker, B.A.D., and Lindberg, A.A. (1983) *Salmonella typhimurium* mutants defective in UDP-D-galactose: lipopolysaccharide α1,6-D-galactosyltransferase. *J. Biol. Chem.* **258:** 3769-3774.

Wright, A., Dankert, M., Fennessay, P., and Robbins, P.W. (1967) Characterization of a polyisoprenoid compound functional in O-antigen biosynthesis. *Proc. Natl. Acad. Sci. USA* **57:** 1798-1803.

Wyk, P. and Reeves, P.R. (1989) Identification and sequence of the gene for abequose synthase which confers antigenic specificity on group B salmonellae: homology with galactose epimerase.*J. Bacteriol.* **171:** 5687-5693.

Yamamoto, K., Ichinose, Y., Shinagawa, H., Makino, K., Nakata, A., Iwanaga, M., Honda, T., and Miwatani, T. (1990) Two step processing for activation of the cytolysin/hemolysin of *Vibrio cholerae* O1 biotype El Tor: nucleotide sequence of the structural gene (*hlyA*) and characterization of the processed products. *Infect. Immun.* **58:** 4106-4116.

Yamamoto, K., Takeda, Y., Miwatani, T. and Craig, J.P. (1983b) Evidence that a non-O1 *Vibrio cholerae* produces enterotoxin that is similar but not identical to cholera enterotoxin. *Infect Immun.* **41:** 896-901.

Yancey, R.J., Willis, D.L., and Berr, L.J. (1978) Role of motility in experimental cholera in adult rabbits. *Infect. Immun.* **22**: 387-392.

Yanish-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC17 vectors. *Gene* **33:** 103-119.

Yarden, Y. and Ullrich, A. (1988) Growth factor receptor tyrosine kinases. *Annu. Rev. Biochem.* **57:** 443-478.

Zamze, S.E. and Moxon, E.R. (1986) Composition of the lipopolysaccharide from different capsular serotype strains of *Haemophilus influenzae. J. Gen. Microbiol.* **133:** 1443-1451.

Zinnaka, Y. and Carpenter, C.C.J., Jr. (1972) An enterotoxin produced by noncholera vibrios. *John Hopkins Med. J.* **131**: 403-411.

Zubay, G. (1973) The *in vitro* synthesis of proteins in microbial systems. *Annu. Rev. Genet.* **7:** 267-287.

Zwahlen, A., Rubin, L.G., Connelly, C.J., Inzana, T.J. and Moxon. E.R. (1985) Alteration of the cell wall of *Haemophilus influenzae* type b by transformation with cloned DNA: association with attenuated virulence. *J. Infect. Dis.* **152:** 485-492.

ADDITIONAL REFERENCES

Alberts, A. and Vagelos, P.R. (1972) Acyl-CoA carboxylases. p. 37-82. *In*: The enzymes. 3rd edition. (ed. Boyer, P.D.) Academic Press, Inc., New York.

Boulnois, G.J., Roberts, I.S., Hodge, R., Hardy, K.R., Jann, K.B., and Timmis, K.N. (1987) Analysis of the K1 capsule biosynthesis gene of *Escherichia coli*: definition of three functional regions for capsule production. *Molec. Gen. Genet.* **208:** 242-246.

Brade, H., Galanos, C. and Lüderitz, O. (1983) Differential determination of the 3-deoxy-D-mannooctulosonic acid residues in lipopolysaccharide of *Salmonella minnesota* rough mutants. *Eur. J. Biochem.* **131:** 195-200.

Brahmbatt, H. (1989) Cloning and molecular characterization of the *rfb* gene cluster of Salmonella typhimurium LT2. Ph.D. thesis. University of Adelaide.

Cronan, J.E., Jr. and Rock, C.O. (1987) Biosynthesis of membrane lipids. pp. 474-498. *In*: *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology. Vol. I. (ed. Neidhardt, F.C.) American Society for Microbiology.

Faast, R., Ogierman, M.A., Stroeher, U.H., and Manning, P.A. (1989) Nucleotide sequence of the structual gene *tcpA*, for a major pilin subunit of *Vibrio cholerae*. *Gene* **85**: 227-231.

Hitchock, P.J., Leive, L., Mäkelä, P.H., Rietschel, E.T., Strittmatter, W., and Morrison, D.C. Lipopolysaccharide nomenclature - past, present, and future. *J. Bacteriol.* **166**: 699-705.

Levine, M.M., Kaper, J.B., Herrington, D., Losonsky, G., Morris, J.G., Clements, M.L., Black, R.W., Tall, B., and Hall, R. (1988) Volunteer studies of deletion mutants of *Vibrio cholerae* 01 prepared by recombinant techniques. *Infect. Immun.* **56**: 161-167.

Manoil, C. and Beckwith, J. (1985) Tn*phoA*: A transposon probe for protein export signals. *Proc. Natl. Acad. Sci. USA.* **82**: 8129-8133.

Matsuhashi, S. (1966) Enzymatic synthesis of cytidine diphosphate 3,6-dideoxyhexoses. II. Reversible 2-epimerazation of cytidine diphosphate paratoes. *J. Biol. Chem.* **241**: 4275-4282.

Mayer, H., Rapin, A.M., Schmidt, G., and Bowman, H.G. (1976) Immunochemical studies on lipopolysaccharide from wild-type and mutants of *Escherichia coli* K-12. *Eur. J. Biochem.* **66**: 357-368.

Pabo, C.O. and Sauer, R.T. (1984) Protein-DNA recognition. *Annu. Rev. Biochem.* **53**: 293-321.

Raziuddin, S. (1979) Toxic and immunological properties of the lipopolysaccharide (O-antigen) from Vibrio el-tor. Immunochemistry. **15**: 611-614.

Rubenstein, P.A. and Strominger, J.L. (1974) Enzymatic syntheis of cytidine diphosphate 3,6-dideoxyhexoses. 8. mechanistic roles of enzyme E-1 and pridoxamine 5'-phosphate in the formation of cytidine diphosphate 4-keto-3,6-dideoxy-D-glucose from cytidine diphosphate-4-keto-6-deoxy-D-glucose. *J. Biol. Chem.* **249**: 3776-3781.

Shaw, C.E. and Taylor, R.K. (1990) *Vibrio cholerae* O395 *tcpA* pilin gene sequence and comparison of predicted protein structural features to those of type 4 pilins. *Infect. Immun.* **58**: 3042-3049.

Ward, H. (1988) Serotype variation in *Vibrio cholerae*. Ph.D. thesis. University of Adelaide.

Woisetchläger, M. and Hogenauer, G. (1986) Cloning and characterization of the gene encoding 3-deoxy-D-manno-octulosonate 8-phosphate synthetase from *Escherichia coli*. *J. Bacteriol.* **168**: 437-439.

Yamamoto, K., Takeda, Y., Miwatani, T. and Craig, J.P. (1983a) Purification and some properties of a non-O1 *Vibrio cholerae* enterotoxin that is identical to cholera enterotoxin. *Infect. Immun.* **39**: 1128-1135.