Haemagglutinins of *Vibrio cholerae* O1: studies on the organisation of the genes encoding the mannose-fucose-resistant haemagglutinin (MFRHA)

Andrew Barker, B.Sc. (Hons) (Adelaide)

Department of Microbiology and Immunology

The University of Adelaide

Adelaide, S.A., 5005

Australia

A thesis submitted for the degree of Doctor of Philosophy

December, 1993.

Awarded 1994

Acknowledgements

I am grateful to my family for their support and encouragement throughout the years, in particular during this last and most difficult year. Finally, to Anke, thanks for being there. I hope one day to be able to repay the debt I owe you.

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 my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text. The author consents to the thesis being made available for photocopying and loan, if applicable, but only if accepted for the award of the degree.

Andrew Barker

List of abbreviations

A: adenine

ACF: accessory colonisation factor

ADA: N-[2-acetamido]-2-iminodiacetic acid

Ap: ampicillin

ATCC: American type culture collection

ATP: adenosine 5'-triphosphate

BIME: bacterial interspersed mosaic element

bp base pair

BSA: bovine serum albumin

C: cytosine

c.i.p.: calf intestinal phosphatase

Cm: chloramphenicol

CTP: cytidine 5'-triphosphate

dATP: 2'-deoxy adenosine 5'-triphosphate

dCTP: 2'-deoxy cytidine 5'-triphosphate

ddATP: 2',3'-dideoxy adenosine 5'-triphosphate

ddCTP: 2',3'-dideoxy cytidine 5'-triphosphate

ddGTP: 2',3'-dideoxy guanidine 5'-triphosphate

ddTTP: 2',3'-dideoxy thymidine 5'-triphosphate

DEPC: diethyl pyrocarbonate

dGTP: 2'-deoxy guanidine 5'-triphosphate

DIG: digoxigenin

DIG-dUTP: digoxigenin-11-[2'-deoxy uridine 5'-triphosphate]

dITP: 2'-deoxy inosine 5'-triphosphate

DNA: 2'-deoxyribonucleic acid

DNase: Deoxyribonuclease

dNTP: 2'-deoxy nucleotide 5'-triphosphate

DTT: 1,4-dithiothreitol

dTTP: 2'-deoxy thymidine 5'-triphosphate

EDTA: ethylene-diamine-tetraacetic acid

ELISA: enzyme-linked-immunosorbent assay

EMS: ethyl methyl sulphonic acid

FBS: foetal bovine serum

FSHA: fucose-sensitive haemagglutinin

G: guanine

Gm: gentamycin

GTP: guanidine 5'-triphosphate

HA/protease: haemagglutinin/protease

HGT: high gelling temperature

IPTG: isopropyl-β-D-galactopyranoside

kb: kilobase pairs

kDa: kiloDaltons

Km: kanamycin

KR: Krebs-Ringer buffer

LB: Luria broth

LGT: low gelling temperature

LPS: lipopolysaccharide

MEM: minimal Eagle medium

MFRHA: mannose-fucose-resistant haemagglutinin

M-MLV: Moloney murine leukaemia virus

MOPS: 3-[N-morpholino]-propanesulphonic acid

mRNA: messenger ribonucleic acid

MSHA: mannose-sensitive haemagglutinin

NA: nutrient agar

Nal: nalidixic acid

NB: nutrient broth

NNMG: N-methyl-N'-nitro-N-nitrosoguanidine

NTP: nucleotide 5'-triphosphate

oligo: oligodeoxynucleotide

ORF: open reading frame

ORT: oral rehydration therapy

PAGE: polyacrylamide gel electrophoresis

PBS: phosphate buffered saline

PEG: polyethylene glycol 6000

PIPES: piperazine-N,N'-bis[2-ethanesulphonic acid]

PMSF: phenyl-methyl-sulphonyl fluoride

psi: pounds *per* square inch

PU: palindromic unit

R: resistant

REP: repetitive extragenic palindrome

RF: replicative form

Rif: rifampicin

RNA: ribonucleic acid

RNase: ribonuclease

rpm: revolutions *per* minute

S: sensitive

SDS: sodium dodecyl sulphate

Sm: streptomycin

Sp: spectinomycin

T: thymine

TB: terrific broth

Tc: tetracycline

TCBS: thiosulphate-citrate-bile-sucrose

TCP: toxin co-regulated pilus

TEMED: N,N,N',N'-tetramethyl-ethylenediamine

Tn: transposon

Tris: tris[hydroxymethyl]amino methane

Triton X-100: α-[4-(1,1,3,3-tetramethylbutyl)phenyl]-ω-hydroxypoly(oxy-1,2-ethanediyl)

TSA: trypticase soy agar

TSB: trypticase soy broth

Tween 20: polyoxyethylenesorbitan monolaurate

U: uracil

UTP: uridine 5'-triphosphate

UV: ultraviolet

VCR: *Vibrio cholerae* repetitive DNA sequence

v/v: volume *per* volume

w/v: weight *per* volume

X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside

X-pho: 5-bromo-4-chloro-3-indolyl-phosphate

Abstract

Previous studies on the Mannose-Fucose Resistant Haemagglutinin (MFRHA) of *Vibrio cholerae* O1 have implicated it as a virulence determinant. The initial aim of this study was to characterise the region of the chromosome associated with the gene encoding the MFRHA. It had been hypothesised that the MFRHA forms a fimbrial-type structure, in which case the structural gene would be expected to be part of a fimbrial biosynthetic operon. To this end the nucleotide sequence of the entire insert of the original MFRHA clone, pPM471 has been determined. The clone contains 10 complete open reading frames. Two of these have been previously associated with MFRHA activity and are described further below. Of the other 8, only three show any significant similarity to entries in the PIR and Swissprot databases: a 9 kDa ORF, similar to the N-terminal portion of the DnaT protein of *Escherichia coli*; an 11 kDa ORF, similar to the *relE* gene product of *E. coli*, that is predicted to be translationally coupled with the 9 kDa ORF; and a 19 kDa ORF, similar to the α_2-microglobulin superfamily. The sequence of these surrounding ORFs makes it seem most unlikely that the MFRHA forms a fimbrial-type structure, because of the lack of similarity to the genetic organisation seen with fimbrial biosynthetic operons from a variety of bacteria.

The most striking feature of the sequence presented is the occurence of nine copies of an (imperfect) 124 bp direct repeat, here named VCR, that are found outside of the reported ORF's. The nine copies of VCR show an overall similarity of 92%, and are all in the same orientation relative to one another. Analysis of DNA gel blots probed with VCR-specific probes indicate that this sequence occurs at least 60 to 80 times in the *V. cholerae* O1 chromosome and that homologous sequences also occur in non-O1 *V. cholerae* but not in the other members of the genera *Vibrio* or *Aeromonas* tested. Analysis of pulsed field gel blots probed with a VCR-specific probe indicate that most, if not all, copies of VCR lie within approximately 10% of the *V. cholerae* O1 chromosome. This situation differs from that found with other bacterial repetitive sequences such as REP, which are distributed at random through the chromosome. The implications of these results, along with the work of other authors, have led to several proposals of functions for VCR, the relative merits of which are discussed.

As mentioned above, a 19 kDa ORF linked to the MFRHA locus shows sequence similarity to the α_2-microglobulin superfamily. These proteins are all soluble transporters of small, hydrophobic molecules, that are found in a variety of eukaryotes. The best matches are to members of a subset of this superfamily which have been shown to interact with porphyrin ring structures. This similarity is of particular interest as the 19 kDa ORF, here named *vlpA*, was predicted, and subsequently demonstrated, to be a lipoprotein. VlpA was demonstrated to bind haemin, as well as the related compounds haematoporphyrin, protoporphyrin IX (to a lesser extent) and Congo red, a dye which has been shown in other systems to bind to haemin-binding proteins. Construction of a *vlpA*::Km^R insertion mutant indicated that VlpA is not associated with the MFRHA phenotype, at least in *E. coli* clones. A series of *vlpA-phoA* fusions has been generated and analysis of the haemin binding activity of these fusion proteins supports the notion that VlpA is involved in haemin binding, as progressive deletion of VlpA from the carboxy-terminal end leads to an eventual loss of haemin binding activity. However, conditions under which *vlpA* is transcribed in *V. cholerae* have not been identified. DNA gel blots indicate that there are at least two copies of *vlpA* hybridising sequences in most strains of *V. cholerae* O1, the one exception being a strain in which the genes encoding the MFRHA have also been deleted, in which a single cross hybridising band was detected.

Two of the ORFs sequenced, that encode a 7 kDa and 25 kDa protein have been linked with MFRHA activity. These genes are here designated *mrhA* and *mrhB* respectively, and their expression has been investigated. Extensive biochemical and genetic analyses reported here show that the two genes are transcribed as an operon, from a single promoter upstream of *mrhA* that shows similarity to the consensus for *E. coli* σ^{70} promoters. Efficient transcription of *mrhA* occurs, but largely terminates close to the 3' end of a predicted stem loop structure that lies between the two genes. Transcription of *mrhB* occurs *via* a 'read-through' of this terminator and *mrhAB* transcripts terminate within a copy of VCR lying 3' to the *mrh* operon. It is unlikely that differential stability of *mrhA* and *mrhB* mRNA is responsible for the difference observed in the abundance of the two messages. This mechanism, in the absence of any data on the translational efficiencies of the two genes, explains the difference in abundance of the two proteins, as observed previously, and is contrasted with the mechanisms of differential mRNA stability leading to a similar difference

in protein abundance in the *E. coli* fimbrial operons analysed to date and predicted to occur in the *V. cholerae tcp* gene cluster by others.

A deletion mutant covering the distal portion of *mrhB* and most of a downstream gene was previously constructed and analysis suggested that *mrhB* encodes the MFRHA and is a virulence determinant for *V. cholerae* O1. Further analysis of this mutant supported this assertion. However, some controversy remains as to whether *mrhA* or *mrhB* encodes the MFRHA. To resolve this, further mutants in either *mrhA* or *mrhB* or both were constructed, and their effects on haemagglutination were examined in the heterologous host *E. coli* K-12. Under the conditions used here, either *mrhA* or *mrhB* is sufficient to mediate haemagglutination in the heterologous host. Attempts to introduce a $\Delta mrhAB$ mutation into both classical and El Tor strains of *V. cholerae* O1, in order to conduct a similar analysis in the homologous host, were unsuccessful. Comparison of the predicted sequence of MrhA and MrhB with entries in the databases showed that MrhA is not similar to any entries currently in the databases. However, MrhB is similar to several DNA methylases associated with type II restriction endonucleases, and the similarity centres on the two regions that form the active site and the S-adenosyl methionine binding site, of these proteins. Analysis of transcription of mutant *mrhAB* alleles in *E. coli* clones suggests that the ORF immediately downstream of *mrhAB* may also be involved in the regulation of expression from p_{mrh}. Other attempts to resolve the controversy of whether *mrhA* or *mrhB* is the structural gene for the MFRHA were unsuccessful, and the implications of the results presented above in resolving the controversy are discussed.

Contents

Chapter 2:
Materials and methods

Chapter 3:
Nucleotide sequence of pPM471, a mannose-fucose-resistant haemagglutinin expressing clone

Chapter 4:
Mapping of VCR, a repeated sequence present in the *Vibrio cholerae* chromosome

Chapter 5:
Characterisation of a haemin-binding protein gene, linked to genes encoding the mannose-fucose-resistant haemagglutinin of *Vibrio cholerae* O1

Chapter 6:
Transcription organisation of the *mrh* locus encoding the mannose-fucose-resistant haemagglutinin of *Vibrio cholerae* O1

Chapter 7:
Examination of the role of the mannose-fucose-resistant haemagglutinin of *Vibrio cholerae* O1 in bacterial adherence and virulence

Chapter 8:
Perspectives

Bibliography

Chapter 1

Vibrio cholerae O1: The nature of the beast

1.1 Introduction

Of the many infectious diseases that still plague mankind, cholera is viewed by modern science and medicine as one of the simplest and easiest to treat. Prevention by adequate hygiene has been abundantly demonstrated and treatment via oral rehydration therapy is successful. The disease process is viewed as straightforward, an uncomplicated event in which bacteria are ingested, colonise the small intestine and secrete a powerful enterotoxin. However, this simplistic view of cholera does not adequately reflect the disease process, which is better viewed as one step in the life cycle of a remarkable parasite.

1.2 The Disease

1.2.1 A Historical Aspect

Asiatic cholera has been endemic in the Indian sub-continent, in a region centred on the Ganges and Brahmaputra rivers, since antiquity (Pollitzer, 1959). Cholera epidemics occurred in southern India following pilgrimages to the Ganges (McNeill, 1976) and since the early nineteenth century have spread to the rest of the world. In all, seven pandemics have occurred since 1817 (Rabbani, 1986).

The appearance of cholera in western Europe generated as much horror and revulsion as had bubonic plague in previous centuries (Spangler, 1992). Similar responses were observed in North America during the 1832 epidemic, exacerbated no doubt by the utter failure of even the strictest quarantine measures to stem the tide of the disease (Rosenberg, 1962). Morality and faith were seen as the best preventative measures for the disease and the few medical practitioners who advocated public health measures such as removing the filth from the streets of New York went unheeded (Rosenberg, 1962).

Despite the initial panic, much progress was made in understanding the disease during the nineteenth Century. The recognition that the major effect of the disease was loss of fluids and electrolytes, occurred within months of the introduction of cholera to England in 1831 (Carpenter, 1992). Intravenous fluid infusion was used successfully as early as 1831 (Rabbani, 1986; Barua, 1988; Carpenter, 1992) but was not readily accepted by the majority of physicians, many of whom remained committed to treatments such as purging with

calomel, blood-letting or opium. This situation is understandable given the high mortality rates associated with air embolism and septicaemia following intravenous infusion, and the best argument in favour of intravenous rehydration was that "patients recovered sufficiently to make their wills" (Barua, 1988)!

The most significant advance in prevention of cholera centred on the epidemiological studies of John Snow (Snow, 1855). Working in London, he was able to demonstrate the link between contaminated water drawn from the Broad Street pump and disease, and also differences in incidence of cholera between purchasers of water from the Southwark and Vauxhall company, which drew water from polluted parts of the Thames, and the Lambert Company, whose source was upstream of the point of sewerage influx. New ideas in prevention seemed to take hold quicker than ideas in treatment, so that the American communities who had espoused morality and godliness as a prevention in 1832 were, by the onset of the last cholera epidemic in North America in 1866, advocating sanitation and other public health measures as the best prevention (Rosenberg, 1962).

The 'last' great discovery in the understanding of cholera during the nineteenth Century was the description of the causative organism. An Italian physician, Filippo Pacini, observed large numbers of curved bacteria in the intestinal contents of cadavers of cholera victims, which he named *Vibrio cholera* (Barua, 1988). Thirty years later, Robert Koch independently observed the same "Kommabazillen" in stool and intestinal contents, but not in other tissues, during post mortem examination of cholera victims, first in Alexandria (Koch, 1883) then in Calcutta (Koch, 1884a), thus establishing the non-invasive character of the bacteria. More importantly, he was subsequently able to obtain pure cultures of the bacteria (Koch, 1884b). The final step in establishing the link between bacterium and disease was the demonstration of a cholera-like disease following intra-duodenal inoculation of opium treated guinea pigs (Koch, 1886), although these results were not universally accepted at the time (Barua, 1988; Finkelstein, 1988).

Following Koch's isolation of the causative organism, attempts were quickly initiated to immunise against the disease: both oral (live attenuated and killed organisms) and

parenteral vaccine types were investigated (reviewed in: Barua, 1988; Finkelstein, 1988), albeit with equivocal results. The similarities between these attempts and modern attempts to produce an effective vaccine are intriguing.

1.2.2 Current Wisdom

It is now widely accepted that the disease Asiatic cholera is caused by *Vibrio cholerae* (Baumann *et al.*, 1984; Rabbani, 1986; Barua, 1988; Finkelstein, 1988). The bacterium is Gram negative and a member of the γ subdivision of the purple bacteria (Woese, 1987). Cell shape is characteristic, being a "bent rod" or "comma", and cells are motile by virtue of a single polar, sheathed flagellum (Baumann *et al.*, 1984). Modern classification schemes have recognised numerous subdivisions of *V. cholerae* on the basis of the O-Antigen of the lipopolysaccharide (LPS). The most recent typing scheme recognises 139 O-serotypes of *V. cholerae* (Shimada *et al.*, 1993), but all true cholera vibrios fall within the O1 serotype. All other serotypes are collectively referred to as non-O1 and are not attributed with epidemic potential (Baumann *et al.*, 1984; but see Section 1.5 for recent developments). The O1 serotype can be subdivided into three sub-serotypes on the basis of antigenic differences, known as Inaba, Ogawa and Hikojima (Gustafsson and Holme, 1985). *V. cholerae* O1 can also be differentiated on the basis of biotype, which are referred to as classical and El Tor (Baumann *et al.*, 1984). This distinction is based on a number of biochemical characteristics, as well as on patterns of phage sensitivity (see below). The properties associated with biotype seem to be important as the first six pandemics were due to strains exclusively of the classical biotype (Finkelstein, 1973), while the El Tor biotype (so named as it was originally isolated at the El Tor quarantine station in Egypt), which was originally thought to be a harmless "cousin" of *V. cholerae* as it was isolated from asymptomatic carriers (Pollitzer, 1959), has been responsible for the current, seventh pandemic (Finkelstein, 1973; Finkelstein, 1988). The picture of the current pandemic has been clouded somewhat by the recent re-emergence of classical strains in southern Bangladesh (Siddique *et al.*, 1991). The appearance of a new serotype of *V. cholerae* with epidemic potential will be discussed in Section 1.5.

The clinical course of cholera is well recognised. Onset is rapid, often striking the victim unawares (Rosenberg, 1962; Carpenter, 1992) and is accompanied by dizziness,

vomiting, stomach cramps and profuse watery (rice-stool) diarrhoea (Finkelstein, 1973; Barua and Burrows, 1974). This rapidly leads to dehydration and electrolyte depletion with cyanosis in severe cases (Carpenter, 1992, and references therein). If left untreated, death results in approximately 50% of cases (Rosenberg, 1962; Carpenter, 1992), a figure which in nineteenth Century North America rose to over 70% with treatment (Rosenberg, 1962).

Treatment methods now advocated by the World Health Organisation are based on oral rehydration therapy (ORT) which has much reduced the medical and logistical problems associated with intravenous therapy (Greenough and Molla, 1988; Carpenter, 1992) and has also led to drastically reduced mortality rates. This was spectacularly demonstrated during what could have been a catastrophic outbreak of cholera in South America in 1991, when the use of ORT led to a mortality rate of less than 1% (reviewed in Carpenter, 1992). Bearing this in mind, it is interesting to note that an ORT similar to modern therapies was advocated for cholera in India some 3 000 years BC (Greenough and Molla, 1988).

Of the numerous bacterial factors implicated in the disease process (reviewed in detail below) by far the best known is the secreted enterotoxin. First postulated by Koch in 1884 (Barua, 1988; Finkelstein, 1988) the toxin was identified in 1959 (De, 1959; Dutta et al., 1959) and purified a decade later (Finkelstein and LoSpalluto, 1969). Much subsequent research has shown that this toxin initiates the massive fluid efflux from the intestinal epithelium, although the exact mechanism by which this occurs remains to be elucidated (reviewed in Spangler, 1992). Pathologic investigations show that the rice-water stool symptomatic of cholera is due to epithelial desquamation and necrosis as well as fluid efflux and excessive mucous production (Asakura et al., 1973), so factors other than cholera toxin must also contribute to diarrhoea.

Colonisation is of paramount importance in the disease process, being a necessary prerequisite (Schrank and Verwey, 1976; Srivastava et al., 1980). Therefore, factors involved in colonisation have also been the focus of much attention, usually with the ultimate goal of effective vaccine development in mind (Manning, 1987; Taylor, 1989; Manning, 1992).

Despite this intense research, colonisation remains a complex and poorly understood process, as will be seen below.

1.3 The life cycle of *V. cholerae* O1

1.3.1 The aquatic phase

The role of water in the transmission of cholera was well established by Snow (1855). However, a natural reservoir for *V. cholerae* O1 was, for a long time, not identified and was assumed to be *Homo sapiens* (Barua and Burrows, 1974). More specifically, various studies over the years reached the conclusion that *V. cholerae* does not survive for long periods in saline waters and this has been generally accepted (Grimes *et al.*, 1986). Studies by Colwell and co-workers have challenged this view by demonstrating (*via* immunofluorescent techniques) that *V. cholerae* O1 enters a non-culturable but viable state in both microcosm and field studies, rather than dying as was concluded from plate viable counting (Colwell *et al.*, 1985; Grimes *et al.*, 1986). More importantly, non-culturable *V. cholerae* from a microcosm (the classical strain CA401) retained virulence, at least in a rabbit ligated ileal loop model (Colwell *et al.*, 1985). This indicates that *V. cholerae* can indeed survive for long periods in aquatic environments and retain its pathogenic potential throughout. Such a result immediately begs the question as to how important life in an aquatic environment is for *V. cholerae*.

Interactions between *V. cholerae* O1 and a number of other occupants of the aquatic environment have been reported. The first reported an association with planktonic copepods (a sub-class of the Crustacea) in laboratory conditions (Huq *et al.*, 1983), which appears to be maximal under mildly alkaline and saline conditions: that is, conditions similar to those found in an estuarine environment (Huq *et al.*, 1984). This association with copepods is interesting in the light of their abundance at the time of seasonal cholera outbreaks in Bangladesh (Huq *et al.*, 1984), but also because of their role as a major food source for larger marine animals, and the ability of some copepods to parasitise other marine animals, crabs in particular (Barnes, 1987). Subsequently, attachment of *V. cholerae* to diverse zooplankton and phytoplankton species has been demonstrated in estuarine environments in Bangladesh (Tamplin *et al.*, 1990) indicating that plankton could be significant for the ecology and

transmission of *V. cholerae*. Independently, adherence in laboratory microcosms was demonstrated to: five freshwater plant species and a filamentous green alga, *Rhizoclonium fontanum* (Islam *et al.*, 1989); cyanobacteria, to which persistence for more than fifteen months after inoculation was observed (Islam *et al.*, 1990a); and duckweed, where adherence led to increased survival as determined by plate viable count (Islam *et al.*, 1990b). Significantly, increased cholera toxin production was observed with *V. cholerae* adherent to *R. fontanum* (Islam, 1990), a finding that I will return to in Section 1.4.1. Association with amoebae may also be beneficial, as pre-inoculation of microcosms with trophozoites of two amoebae species led to multiplication of *V. cholerae* O1 whereas a decrease in numbers was observed in the absence of amoebae, as determined by plate viable count (Thom *et al.*, 1992). However, *V. cholerae* have also been observed to be vulnerable to predation by protozoa (untyped) present in sub-tropical river water in Taiwan (Chao *et al.*, 1988) so interactions with other microorganisms are likely to be complex. *V. cholerae* O1 (cholera toxin positive) and non-O1 have been isolated from the cloacae and faeces of aquatic birds trapped on inland waters in the United States (Ogg *et al.*, 1989), adding a new dimension to the transmission of cholera.

How important are interactions in the environment to the life cycle of *V. cholerae*? Huq *et al.* (1983) suggested that adherence to copepods (and presumably to other organisms) might provide an over-wintering substrate for the bacteria, when water conditions (especially temperature) are less favourable for survival. Bacterial replication also seems to be slow, with numerous authors cited above mentioning "survival" rather than "replication". Virtually nothing is known about gene expression in environmental conditions, or the bacterial factors required for survival specifically in the aquatic environment, beyond the intriguing observation that cholera toxin expression is increased in bacteria adherent to an alga (Islam, 1990), and this remains a sadly neglected aspect of the life cycle of *V. cholerae*. In summary, it would seem that the aquatic phase is a relatively dormant one, with bacterial functions geared to maximising survival until encountering a susceptible host.

1.3.2 Ingestion

V. cholerae O1 is transferred to its primary host, *H. sapiens*, by ingestion (Pollitzer, 1959). This has traditionally been held to occur *via* the faecal-oral route, but due to the unlikelihood of estuarine water being used for drinking purposes, a scheme involving primary and secondary modes of transmission has been proposed (Miller *et al.*, 1985). Briefly, primary transmission is associated with the ingestion of incompletely cooked contaminated seafood (in particular crabs and oysters), whereas secondary transmission occurs via the faecal-oral route of contaminated water or food (as reviewed in Miller *et al.*, 1985). In the light of the study by Ogg *et al.* (1989), malicious behaviour on the part of aquatic birds must also be considered under primary transmission, often surprisingly distant from areas where *V. cholerae* is endemic.

Having arrived in its primary host, a considerable barrier faces *V. cholerae*, namely the stomach. The mildly alkalophilic nature of *V. cholerae* has been well documented (Baumann *et al.*, 1984; Carpenter, 1992; Section 1.3.1), as has its extreme sensitivity to low pH. Concomitant administration of sodium bicarbonate led to a drastic reduction of the minimum effective dose in studies with human volunteers, from 10^8 in the absence of bicarbonate, to 10^4 in the presence of bicarbonate (Hornick *et al.*, 1971). Several studies in human volunteers have shown that the number of cells required to infect half of the volunteers is large, around 10^6 (Rabbani, 1986). Whether such a large dose is required for infection under natural circumstances is a matter for some conjecture: adherence of *V. cholerae* to planktonic organisms in contaminated waters (Tamplin *et al.*, 1990) may well provide significant protection against stomach acid in the manner generally ascribed to biofilms in a wide variety of situations for a wide variety of bacteria (Blenkinsopp *et al.*, 1992), and Levine *et al.* (1983) estimated that the 'natural' infective dose may be as low as 10^2–10^3 in endemic areas. Protection against gastric acid would also be expected if contaminated food were ingested (Miller *et al.*, 1985). Thus, one may consider passage through the gastric barrier to be the critical point in the life cycle of *V. cholerae*.

Changes in gene expression in *V. cholerae* in response to ingestion have not been investigated, but experiments with various Enterobacteriaceae give a guide as to what might

be expected. An inducible acid tolerance response has been identified in *Escherichia coli* (Goodson and Rowbury, 1989) and *Salmonella typhimurium* (Foster and Hall, 1990) in which exposure to mildly acidic conditions (at least *in vitro*) enables survival in much harsher conditions with which the constitutive, allosteric pH homeostasis system could not cope (Foster and Hall, 1991). Interestingly, this inducible system is under genetic control of the ferric uptake regulator (*fur*, which will be discussed in detail in Section 1.4.2.1), but in a manner apparently independent of iron concentration (Foster and Hall, 1992). However, the existence or otherwise of a similar system in *V. cholerae* must be established before indulgence in speculation regarding its role in the life-cycle.

1.3.3. Colonisation

Having survived the low pH of the stomach, the bacteria pass on to the small intestine, where another potentially lethal secretion is encountered, namely bile (Drasar and Barrow, 1985). A detergent-shock response has been identified in Enterobacteriaceae, an energy-dependent tolerance to anionic and neutral (but not cationic) detergents that presumably underlies their tolerance to bile salts (reviewed in Nickerson and Aspedon, 1992). The expression of a number of proteins is induced upon growth of *E. coli* in sodium dodecyl sulphate (SDS) and the expression of a number of others repressed (Adamowicz *et al.*, 1991). The phospholipid composition of the *Enterobacter cloacae* outer membrane also alters on growth in SDS, although the link between membrane composition and detergent shock is not clear (Nickerson and Aspedon, 1992). The mechanism by which detergent resistance acts is also not clear, but seems to require periplasmic and cytoplasmic membrane adaptations (including possible induction of periplasmic chaperone proteins), as SDS has been shown to be present in the periplasm of resistant bacteria and might inactivate proteins exposed to the periplasm (Nickerson and Aspedon, 1992). These authors hypothesise that this explains why Enterobacteriaceae are oxidase-negative (i.e. do not possess cytochrome C) as this enzyme is exposed to the periplasm and is therefore "incompatible with high-detergent environments" (Nickerson and Aspedon, 1992).

As *V. cholerae* must also be resistant to bile salts, due to its ability to survive in the small intestine and also due to its ability to grow on MacConkey and thiosulphate-citrate-bile-

sucrose (TCBS) agars (which contain 0.15% bile salts and 0.8% oxgall respectively: Difco Manual, 1984), the existence of a similar system would be expected. However, the data available are contradictory and extremely confusing. *V. cholerae* cells are much more sensitive (two to three orders of magnitude) than *E. coli* to neutral and anionic detergents *in vitro* and the outer membrane contains exposed phospholipids (Paul *et al.*, 1992). This observation is significant as the absence of exposed phospholipids in the outer membrane is cited as a major reason for the detergent resistance of Enterobacteriaceae (Nikaido and Vaara, 1985; Sukupolvi and Vaara, 1989). *V. cholerae* cells are also much more sensitive than Enterobacteriaceae to protein denaturants such as urea (Lohia *et al.*, 1984) and have a markedly less ionic LPS (Paul *et al.*, 1992). Furthermore, detergent resistant derivatives of *V. cholerae* 569B have been isolated following chemical mutagenesis, and the effect of the mutation shown to be defective acylation of Lipid A (Paul *et al.*, 1990). Confronted with this evidence, it seems that either the membrane composition of *V. cholerae* varies drastically with growth conditions, or that a completely different system of detergent resistance is in place, that is presumably inducible, but not under the growth conditions thus far examined (Paul *et al.*, 1990; Paul *et al.*, 1992), as the oxidase positive (cytochrome C positive) status of *V. cholerae* (Baumann *et al.*, 1984) suggests. Further research, including the analysis of the nature of detergent resistance in Lipid A acylation mutants, should hopefully resolve this conundrum.

The mere presence of *V. cholerae* in the small intestine is inadequate for the next stage of their life-cycle: adherence to the intestinal epithelium, or colonisation, is essential, as has been established in animal models (Schrank and Verwey, 1986; Srivastava *et al.*, 1980) and now widely accepted (Rabbani, 1986; Manning, 1987; Taylor, 1989). Consequently the factors mediating colonisation have been the subject of much interest, not just for the sake of knowledge, but also because of the vaccine potential of colonisation factors as established in other systems, for example, K88 fimbriae of enterotoxigenic *E. coli* (Rutter *et al.*, 1976).

Freter (1980) defined eleven steps in the association of bacteria with small intestinal mucosa: five of these related to motility and chemotaxis, one to avoidance of cell clumping, three to adherence and one each to multiplication of mucosa-associated bacteria and to

modification of host cells. Such a scheme provides a useful framework in which to examine the colonisation of the small intestine by *V. cholerae* O1.

1.3.3.1 Motility and chemotaxis

It is intuitively obvious that, if adherence to the intestinal epithelium is a prerequisite for a successful infection, then factors that increase access to the epithelium will be advantageous. Therefore, it is not surprising that a chemotactic *V. cholerae* (but not a non-chemotactic mutant) demonstrated the ability *in vitro* to actively penetrate the mucous gel of slices of intestinal tissue (Freter *et al.*, 1981a) and also outcompeted a non-chemotactic mutant in both rabbit and mouse models of cholera (Freter *et al.*, 1981b). Similarly, the presence of a functional flagellum is greatly advantageous in the infant mouse model, but the picture is somewhat complicated by the apparent involvement of the flagellum as the carrier of adhesins, as well as being the mediator of motility (Attridge and Rowley, 1983a). These results have been replicated recently in two rabbit models as well as the infant mouse model for strains of both biotypes (Richardson, 1991).

1.3.3.2 Cell clumping

In vitro growth under some conditions (in liquid media) can lead to the macroscopic clumping of *V. cholerae* cells, much to the annoyance of laboratory workers attempting to obtain values for cell densities (numerous unpublishable observations). This process could also conceivably facilitate successful passage of the gastric barrier, with the external cells shielding the internal ones from the harsh conditions. Clumping was proposed to be one of the mechanisms by which protective antibodies inhibit bacterial association with the mucosa (Bellamy *et al.*, 1975; Steele *et al.*, 1975), leading Freter (1980) to assert that clumping is an undesirable event *in vivo*. However, the identification of an antiserum that mediated agglutination of vibrios, but was not protective, demonstrated that this is not the case (Attridge and Rowley, 1983c). In addition, the number of vibrios observed in microcolonies in the early stages of an experimental infection were much greater than could be explained by division of a single adherent bacterium in the elapsed time, suggesting that clumping may also help bacterial adherence (Nelson *et al.*, 1976). Thus, clumping may be an important facet in the early stages of the colonisation process.

1.3.3.3 Adherence

1.3.3.3.1 Establishing the importance of adherence

Having arrived in the upper small intestine, *V. cholerae* must adhere in order to avoid being swept away by gut peristalsis (Schrank and Verwey, 1976; Drasar and Barrow, 1985). Correspondingly, this step has long been recognised as the most important in the colonisation process and studies on colonisation have focused on the factors influencing adherence (Freter, 1980).

Experiments with pathogenic *E. coli* established that many colonisation factors are proteinaceous, filamentous surface appendages called fimbriae (reviewed in Mooi and DeGraaf, 1985). Similar appendages have been observed to mediate attachment of a number of other pathogenic bacteria to mucosal surfaces (Finlay and Falkow, 1989). However, an initial suggestion that *V. cholerae* might also bear fimbriae (Tweedy *et al.*, 1968) was not widely accepted (Nelson *et al.*, 1976; Hanne and Finkelstein, 1982). Consequently, many studies on the colonisation factors of *V. cholerae* focused on various *in vitro* adherence assays, such as haemagglutination, as first reported by Bales and Lankford (1961), or adherence to isolated intestinal tissue (Jones *et al.*, 1976).

Much of this work, as reported, is contradictory. By defining conditions under which *V. cholerae* could adhere to both fresh and formalin-fixed isolated rabbit brush border membranes and human erythrocytes (Jones *et al.*, 1976), it was possible to demonstrate inhibition of adherence by L-fucose and to a lesser extent by D-mannose, suggesting some receptor specificity (Jones and Freter, 1976). However, when intact intestinal tissue slices were used rather than isolated brush border membranes, significantly different results were obtained, which was interpreted as indicating at least two receptors in the rabbit small intestine (Freter and Jones, 1976). On the other hand, in a study examining adherence to mouse small intestinal tissue slices and guinea pig erythrocytes, no receptor specificity was observed. Furthermore, *in vitro* adherence in the model systems used was shown to be mediated by a previously described slime agglutinin which was shown to be unimportant for *in vivo* adherence (Attridge and Rowley, 1983b, and references therein). This slime agglutinin has been hypothesised to be important outside of the host (Attridge, 1979).

The reasons for these discrepancies are not clear, but could originate in the use of different model systems, strains and growth conditions by the various authors, a problem that still plagues cholera research. The one matter these authors agree on is the importance of the flagellum as a carrier of an adhesin (Jones and Freter, 1976; Freter and Jones, 1976; Attridge and Rowley, 1983a). In spite of this role of the flagellum, the initial association of *V. cholerae* with the intestinal mucosa, at least in the rabbit model, appears to be with the non-flagellar pole of the cell (Nelson *et al.*, 1976; Teppema *et al.*, 1987) while subsequent attachment appears to involve the longitudinal surfaces of the cell (Teppema *et al.*, 1987).

The subsequent characterisation of the bacterial factors involved in adherence was initiated by the purification of the "cholera lectin" or haemagglutinin/protease (Finkelstein *et al.*, 1978; Finkelstein and Hanne, 1982) and the description of four distinct haemagglutinins and the phase of their production during *in vitro* growth (Hanne and Finkelstein, 1982). More recently, a fimbrial adhesin has been identified, called the Toxin-coregulated pilus (TCP) because expression is under control of the *toxR* locus and hence TCP production is activated under *in vitro* conditions that also activate cholera toxin expression (Taylor *et al.*, 1987). Subsequently two other fimbrial types were identified on strains of both biotypes in a morphological study (Hall *et al.*, 1988) and haemagglutinating fimbriae were observed on both adherent and non-adherent El Tor strains (Iwanaga *et al.*, 1989), thus finally confirming the existence of fimbriae on *V. cholerae*. Three other colonisation factors have also been identified: the accessory colonisation factor (*acf*), a group of four linked genes that, when disrupted by insertional mutation, show a reduced ability to colonise in the infant mouse model (Peterson and Mekalanos, 1988); a second (non-proteolytic) soluble haemagglutinin (Banerjee *et al.*, 1990); and a recently described homologue of an *Aeromonas hydrophila* flexible pilin gene that is associated with the cholera toxin structural genes (Pearson *et al.*, 1993). This plethora of adhesins and potential adhesins will be discussed in detail below.

1.3.3.3.2 The haemagglutinin/protease

At first glance the haemagglutinin/protease (HA/protease, also termed soluble haemagglutinin or cholera lectin by various workers) appears to be the universal virulence determinant. First identified by McFarlane Burnet in 1947 as a mucinase (Finkelstein *et al.*,

1983), the purification of this protein has enabled the assignation to it of a number of properties. Partially purified protein possessed the ability to prevent attachment of vibrios to the infant rabbit intestine, when used as a pre-treatment (Finkelstein et al., 1978). The purified protein forms filamentous polymers and while an antiserum raised against the purified protein inhibits haemagglutination, the purified protein does not, a property attributed to its polymerised state (Finkelstein and Hanne, 1982). A cell-associated form of the protein has also been detected (Finkelstein and Hanne, 1982; Hanne and Finkelstein, 1982). The purified protein was also shown to possess protease activity (Finkelstein and Hanne, 1982) and has been shown to be a Zn^{2+}-dependent metallo-protease (Booth et al., 1983) which is involved in the proteolytic nicking of the A subunit of cholera toxin that leads to its activation (Booth et al., 1984), although this role could equally be played by other proteases. In addition to this role, the HA/protease also appears to enhance activity of cholera toxin in rat intestinal loops by eroding the protective layer of epithelial mucous (Crowther et al., 1987). The protein is related to a family of metallo-proteases, most notably the elastase of *Pseudomonas aeruginosa* (Häse and Finkelstein, 1990; Häse and Finkelstein, 1991) and shows similar substrate specificity and immunological cross-reactivity with the elastase (Häse and Finkelstein, 1990). Finally, two monoclonal antibodies raised against the HA/protease inhibit both haemagglutination and protease activities (including cholera toxin activation) but not mucinase activity, and one of these is protective in a rabbit model (Wikstrom et al., 1991).

In the light of this evidence, it is surprising to learn that a HA/protease (*hap*) null mutant, constructed by introduction of an inactivated cloned gene into the chromosome (Häse and Finkelstein, 1991) is fully virulent in infant rabbits and is therefore not directly involved in virulence (Finkelstein et al., 1992)! Instead Finkelstein et al. (1992) proposed that the protein functions as a "detachase", noting that pre-treatment of cultured human intestinal epithelial cells with HA/protease inhibited attachment of vibrios cultivated under conditions to stimulate production of three different haemagglutinins, in a dose-dependent manner. Furthermore, the attachment and subsequent detachment observed by Jones et al. (1976) could be explained by HA/protease activity as a similar detaching phenomenon is observed with adherence of HA/protease-positive, but not HA/protease-negative strains to cultured human intestinal epithelial cells (Finkelstein et al., 1992). Thus, it seems that the major

contribution of this protein to the life cycle of *V. cholerae* is leading to detachment, and thus causing greater shedding of progeny bacteria by the host (Finkelstein *et al*, 1992), apart from its postulated role in the removal of the protective mucous layer (Crowther *et al*, 1987) and in proteolytic activation of the cholera toxin A subunit (Booth *et al.*, 1984).

1.3.3.3.3 Fucose sensitive haemagglutinin (FSHA)

Adherence as observed by Jones *et al.*, (1976), using *V. cholerae* grown for 18 hours, was inhibited by the sugar L-fucose. However, Hanne and Finkelstein (1982) observed FSHA activity only transiently during early exponential-phase growth in classical strains, and in El Tor strains defective for another haemagglutinin, the mannose sensitive haemagglutinin (see Section 1.4.3.3.5). FSHA expression seems to be easier to demonstrate on strains that are old isolates, as many recent isolates seem to have "lost" FSHA activity and it has been speculated that this acquired expression may be a result of long term culturing under laboratory conditions (Jonson *et al.*, 1989a). However, strains that do express FSHA *in vitro* also express it *in vivo*, at least in a rabbit model, so the FSHA must be considered as a potential adhesin (Jonson *et al.*, 1990). The nature of the FSHA is not known: fimbriae have not been observed that correlate with FSHA expression (Yamamoto *et al*, 1988; Jonson, 1991) and conditions that often lead to the release of fimbriae from bacterial cell surfaces fail to separate FSHA activity from the cell (Jonson, 1991). Haemagglutination titres of *in vivo* grown FSHA⁺ cells are lower than *in vitro* grown cells, leading to speculation that the growth phase dependent expression observed *in vitro* (Hanne and Finkelstein, 1982) might also occur *in vivo* (Jonson *et al.*, 1990). Despite this, a clear role for the FSHA in pathogenesis is unclear.

1.3.3.3.4 Toxin coregulated pilus (TCP)

TCP was originally identified after Tn*phoA* mutagenesis of the classical strain O395 gave a class of mutants that were defective in colonisation, and in which the *phoA* reporter gene was expressed only under growth conditions that also led to activation of cholera toxin gene expression (Taylor *et al.*, 1987). These mutants showed a dramatic increase in LD_{50} in infant mice and were defective in a fucose resistant haemagglutinin (Taylor *et al.*, 1987). TCP was also identified as one of the targets of a protective antiserum directed against "non-

LPS protective antigens" (Sharma *et al.*, 1989a, Sharma *et al.*, 1989b). The significance of TCP as a colonisation factor has been well demonstrated in the infant mouse model (Taylor *et al*, 1987) and a non-motile variant of 569B was more virulent if it was fed to mice after growth under conditions conducive to TCP production than not, although this could also be due to the expression of other factors (Sharma *et al*, 1989b). TCP is also a significant protective antigen in the infant mouse model (Sharma *et al.*, 1989a; Sharma *et al.*, 1989b; Sun *et al.*, 1990a) and TCP remains the only adhesin demonstrated to be required for colonisation in *H. sapiens* (Herrington *et al.*, 1988).

TCP is now a well defined structure, forming characteristic thick bundles readily seen on examination with the electron microscope (Taylor *et al.*, 1987; Hall *et al.*, 1988; Sharma *et al.*, 1989a,b; Jonson *et al.*, 1992; Voss and Attridge, 1993). Sequencing studies have revealed that the major subunit, TcpA, is related to type 4 pilins, which have been shown to be important in the pathogenesis of a number of Gram negative bacteria (Taylor *et al.*, 1987; Faast *et al.*, 1989; Shaw and Taylor, 1990). This relatedness is sufficiently close for (denatured) TcpA to share antigenic determinants, apparently localised to the highly conserved amino-termini, with several other type 4 pilins (Patel *et al.*, 1991). Two protective monoclonal antibodies have been isolated (Sun *et al.*, 1990b) which recognise epitopes in the carboxy-terminal portion of TcpA, indicating that TCP also shows similarity to type 4 pilins in quaternary structure (Sun *et al.*, 1991).

The TCP fimbria has multiple components: *tcpA* is one of a number of a cluster of genes that appear to be co-transcribed. These include: the gene encoding a lipoprotein, TcpC (Parsot *et al.*, 1991; Ogierman and Manning, 1992b) that has been hypothesised to act as the membrane anchor for TCP (Ogierman and Manning, 1992b) and mutants in which show a marked increase in sensitivity to antibody-dependent complement-mediated lysis (Parsot *et al.*, 1991); a number of genes that show significant homology to genes involved in secretion of proteins in other Gram negative bacteria and are presumably involved in pilin export and fimbrial assembly (Kaufman *et al.*, 1993; Ogierman *et al.*, 1993); a signal peptidase, TcpJ, required for correct processing of the atypical TcpA pre-protein (Kaufman *et al.*, 1991); and a number of regulatory genes, including *toxT*, a component of the ToxR regulatory cascade

(Ogierman and Manning, 1992a; Higgins *et al.*, 1992). This clustering of biosynthetic genes is unique among the clustering of type 4 fimbrial systems characterised to date (Ogierman *et al.*, 1993). An unlinked gene, *tcpG*, was originally thought to encode a tip adhesin for the TCP molecule, as *tcpG* mutants produced morphologically unaltered but defective TCP (Taylor *et al.*, 1988; Shaw *et al.*, 1989). However, subsequent analysis of *tcpG* has shown that it encodes a homologue of an *E. coli* periplasmic protein, DsbA (Bardwell *et al.*, 1991), that is required for correct disulphide bond formation in exported proteins (Peek and Taylor, 1992). The same gene is also required for the secretion of active cholera toxin (Peek and Taylor, 1992; Yu *et al.*, 1992). The biogenesis of TCP is obviously complex, and remains an active field of research.

The regulation of expression of *tcp* genes is also complex. First identified as under the control of the regulator ToxR (Taylor *et al.*, 1987), additional factors have since been identified that regulate *tcp* expression. Shaw *et al.* (1989) proposed that *tcpI* and *tcpH* (*tcpH* is now called *tcpP*: Iredell and Manning, 1994) might encode negative and positive regulators of TCP expression, respectively, and analysis of reporter fusion constructs has demonstrated that TcpP activates expression of both *tcpP* and *tcpI* (Thomas, S., Williams, S.G. and Manning, P.A., manuscript in preparation). TcpI appears to be an outer-membrane protein and neither TcpI nor TcpP shows appreciable sequence similarity to known regulatory proteins, so an understanding of their role in *tcp* expression awaits further study (Thomas *et al.*, *op. cit.*). ToxR is known to play a role, either directly or *via* ToxT, in the activity of these and the *tcpA* promoters (DiRita *et al.*, 1991; Thomas *et al.*, *op. cit.*) and ToxT has been shown to be required for *tcp* expression (DiRita *et al.*, 1991). The ToxR,S,T proteins have been proposed to form a regulatory cascade for the expression of virulence genes (DiRita *et al.*, 1991; DiRita, 1992) and are described in detail in Section 1.4.1.

Expression of *tcp* also appears to be regulated differently in the two biotypes, as TCP expression has been readily detected in classical strains (Taylor *et al.*, 1987; Hall *et al.*, 1988; Sharma *et al.*, 1989a,b; Jonson *et al.*, 1990) but until recently only TcpA, not polymerised TCP, has been detected in El Tor strains (Jonson *et al.*, 1990; Jonson *et al.*, 1992). This appears to be due to the use of inappropriate antisera, as when an El Tor specific antiserum is

used, TCP bundles become obvious on TcpA producing El Tor strains (Voss and Attridge, 1993). El Tor *tcpA* mutant strains are markedly attenuated in the infant mouse model, confirming that TCP is important in the pathogenesis of strains of both biotypes (Attridge *et al.*, 1994).

1.3.3.3.5 Mannose sensitive haemagglutinin (MSHA)

The MSHA was first identified as an El Tor specific chicken erythrocyte haemagglutinin and has been one of the factors used in biotyping *V. cholerae* O1 (Finkelstein and Mukerjee, 1963). It is readily expressed under *in vitro* growth conditions and the isolation of spontaneously arising MSHA negative variants was necessary to detect other haemagglutinins expressed by El Tor strains (Hanne and Finkelstein, 1982). The MSHA mediates attachment of vibrios to rabbit intestinal slices *in vitro* (Bhattachargee and Srivastava, 1978). A Tn5-generated mutant deficient in MSHA has a significantly reduced ability to colonise in the rabbit ileal loop model (Finn *et al.*, 1987) and both polyclonal and monoclonal antisera raised against the MSHA are protective in infant mouse and rabbit ileal loop models (Osek *et al.*, 1992). The MSHA has been shown to be fimbrial in nature and a protective monoclonal antibody binds along its length, indicating that a tip adhesin is probably not part of the fimbrial structure (Jonson *et al.*, 1991b). Jonson *et al.* (1991b) have detected the major pilin subunit in classical strains but have not been able to detect assembled MSHA fimbriae. As earlier work has shown that agglutination of human and chicken erythrocytes by some classical strains is mannose sensitive (Jonson *et al.*, 1989a), it is tempting to propose that this failure is due to the lack of antisera of appropriate specificity, as has been detailed for the detection of TCP expression by El Tor strains in Section 1.3.3.3.4.

Although amino-terminal sequence of the purified protein suggests that the MSHA is a true type 4 pilin (Ehara *et al.*, 1991), data on the genetic organisation, relatedness to other fimbrial systems and control of MSHA gene expression are currently unavailable. The only clue on this matter is that the expression of MSHA under a wide variety of *in vitro* growth conditions (Hanne and Finkelstein, 1982; Jonson *et al.*, 1989a; Jonson *et al.*, 1990) make it seem unlikely that MSHA expression is linked with the ToxRST regulon (Jonson *et al.*, 1990).

1.3.3.3.6 Mannose fucose resistant haemagglutinin (MFRHA)

The MFRHA is another haemagglutinin whose expression was "unmasked" in a MSHA mutant, being expressed in late-exponential and stationary phase *in vitro* (Hanne and Finkelstein, 1982). A MFRHA clone was originally detected with an antiserum raised against the HA/protease (Franzon and Manning, 1986) and also by virtue of the haemagglutinating phenotype that positive clones confer on *E. coli* (van Dongen and DeGraaf, 1986). A mutation that abolishes MFRHA activity has been introduced into a classical strain and attenuates it (Franzon, 1988). The nature and expression of the MFRHA are the subjects of this thesis and so will be discussed in subsequent Chapters.

1.3.3.3.7 Other adhesins

In addition to those described above, reports of other adhesins have appeared. The first of these is the accessory colonisation factor, ACF, encoded by four linked genes which, when disrupted by Tn*phoA* insertion, lead to a decrease in ability to colonise the intestine in the infant mouse model (Peterson and Mekalanos, 1988). No subsequent characterisation of the *acf* genes has been reported, beyond that one, *acfD*, encodes a lipoprotein (Parsot *et al.*, 1991). Expression is regulated by ToxR (Peterson and Mekalanos, 1988) but seems to be more complex in that DNA gyrase inhibitors also affect *acf* expression (Parsot and Mckalanos, 1992). However, the significance of this observation is obscure in that the effect of DNA gyrase inhibitors on other promoters subjected to ToxR control has not been reported.

The purification of a second soluble haemagglutinin, initially from non-O1 strains but also present in O1 strains, has been reported (Banerjee *et al.*, 1990). This molecule differs from the HA/protease in molecular weight, and shows no protease activity, and its activity is inhibited by glycoproteins but not by free proteins or simple sugars (Banerjee *et al.*, 1990). A membrane-bound form of the protein was also detected and electron microscopy of the purified protein revealed globular aggregates rather than the filaments observed for the HA/protease (Banerjee *et al.*, 1990). The role this protein plays in infection remains to be elucidated.

A recent report (Pearson *et al.*, 1993) has identified an ORF (designated *cep*) that shows a high degree of similarity to the *A. hydrophila* flexible pilin major subunit gene (Ho *et al.*, 1992). This ORF lies between the RS1 sequences flanking *ctx* and other toxin genes (see Sections 1.3.3.4.2, 1.3.3.4.5, and 1.3.3.4.6) and deletions that remove *cep* are defective for intestinal colonisation (Pearson *et al.*, 1993). Data on the control of *cep* expression are currently unavailable.

1.3.3.3.8 The targets of adherence

In spite of the keen interest in the adhesins utilised by *V. cholerae* O1, minimal interest has been displayed in their targets, the receptors on the intestinal epithelium. The intestinal epithelium consists of a heterogenous mixture of cell types including: the predominant enterocytes, epithelial cells responsible for the bulk of intestinal nutrient uptake (Lobley, 1991); goblet cells, which are the site of mucous secretion (Lencer *et al.*, 1990); and lymphoid follicles, where specialised absorptive cells (M cells) allow sampling of intestinal contents by the immune system (Owen *et al.*, 1986). Furthermore, the cells of the intestinal epithelium turn over at a phenomenal rate - each villous is estimated to shed up to 2,500 cells per day - with new cells being introduced at the crypts surrounding each villous base (the site of cell division) and then migrating to the tip of the villous where cell death and extrusion occur (Potten and Allen, 1977). Changes in enterocyte surface glycoconjugates (Etzler and Branstrator, 1979) and brush border enzymes (Nordström *et al.*, 1967; Smith, 1985) are observed as the cells differentiate and migrate from the crypts to the villi. Changes in surface glycoconjugates are also observed along the length of the small intestine (Etzler and Branstrator, 1979) and there are significant differences between individuals in the glycosylation patterns of enterocyte surface proteins (Lobley, 1991). Migration of enterocytes over lymphoid follicles also appears to affect their differentiation, leading to differences in brush border hydrolase expression in comparison to non-Peyer's patch-associated enterocytes (Smith, 1985) and there is a strong indication that the surface markers (measured as receptors for *V. parahaemolyticus*) are different on M cells and enterocytes associated with lymphoid follicles (Yamamoto and Yokota, 1989b). Hormonal and dietary changes also affect the expression of brush border enzymes (reviewed in Smith, 1985), and the expression of surface glycoconjugates (Etzler and Branstrator, 1979) and receptors for

enteropathogenic *E. coli* (Cheney *et al.*, 1981) change with the age of the host. Thus, there is a large number of potential receptors present in the small intestine.

In terms of adherence, the scant data available are contradictory. Experiments measuring adherence to formalin fixed tissue (of both human and rabbit origin) show that vibrios adhere more or less uniformly to the epithelium, regardless of cell type (Yamamoto *et al.*, 1988) and subsequently that the main adherence targets on lymphoid follicles were M cells (Yamamoto and Yokota, 1989a). However, following the course of *in vivo* adherence in a rabbit model, a different outcome is observed, as some (albeit unquantitated) difference in the sites of adherence was observed, with a haemagglutinin-positive strain showing some preference for more protected sites, such as villous clefts, than haemagglutinin negative strains (Teppema *et al.*, 1987). This result is intriguing given the suggestion that cholera stool production might be predominantly from crypt, rather than villous cells, in a canine cholera model (Elliott *et al.*, 1970). It also offers an explanation for the plethora of 'adhesins' described above, if the presumably different receptors have different spatial distribution and the receptor(s?) recognised by the non-haemagglutinating strains used by Teppema *et al.*, (1987) is (are?) expressed preferentially on the older cells of the villous tip. The use of biopsy tissue as a model for adherence *in vitro* has also been questioned, because of highly variable results due to antigenic differences between individual donors (Darfeuille-Michaud *et al.*, 1990). These authors have shown a correlation between adherence of enterotoxigenic *E. coli* to a differentiated colonic carcinoma cell line, Caco-2, and adherence to intestinal epithelium. Similarly, the ability to adhere to Caco-2 cells has been correlated with virulence in non-O1 *V. cholerae* (Panigrahi *et al.*, 1990). However, in the light of the variations in enterocyte surface antigens *in vivo* (see above), a combination of *in vitro* and animal assays seems the best future approach to studying the process of adherence, given the difficulties associated with studies in the primary host.

Significant attachment to and entrapment in the mucous layer, using both formalin fixed and unfixed tissue, has also been observed *in vitro* for both *V. cholerae* O1 (Yamamoto *et al.*, 1988; Yamamoto and Yokota, 1988a) and non-O1 (Yamamoto and Yokota, 1988b) strains, leading these authors to claim the mucous layer as a primary adherence target for *V.*

cholerae. This attachment has been shown to be sensitive to sugars, indicating the presence of specific receptors in the mucous blanket (Yamamoto and Yokota, 1988a). Numerous authors have emphasised the possible interaction between *V. cholerae* and the mucous blanket (for example; Schrank and Verwey, 1976; Jones *et al.*, 1976; Freter *et al.*, 1981a; Freter *et al.*, 1981b), but more as a transient association with a traversible barrier. Freter *et al.* (1981b) argue that penetration is rapid, and could occur *via* capilliary channels present along stress planes in the mucous gel. This flowing, stressed structure is expected to be destroyed by formalin fixing, which could explain why Yamamoto and Yokota (1988a) observed entrapment in the mucous layer. It should also be noted that the formalin fixed tissue used by Jones *et al.* (1976) was isolated brush border membrane and not intact intestinal mucosa, and thus did not possess a mucous coat. Following *in vivo* adherence, Teppema *et al.* (1987) found minimal association with the mucous blanket, but emphasised that further study was essential.

Clearly, much remains to be elucidated about the receptors for *V. cholerae* O1. Beyond the sugar sensitivity of several haemagglutinins, the chemical and biological nature of receptors is unknown: they are generally assumed to be surface glycoproteins because of their apparent sensitivity to purified HA/protease (Rabbani, 1986; Manning, 1987; Finkelstein *et al.*, 1992) and because of the inhibition of a recently described haemagglutinin by glycoproteins but not by free protein or simple sugars (Banerjee *et al.*, 1990). Another possibility is that cell surface glycolipids may be important, as has been described for uropathogenic *E. coli* (Lund *et al.*, 1988) and *N. gonorrhoeae* (Strömberg *et al.*, 1988; Paruchuri *et al.*, 1990)

1.3.3.3.9 Non-specific adherence?

The evidence presented strongly argues in favour of specific receptors, but the possibility that *V. cholerae* adherence to the intestinal epithelium might occur *via* non-specific means has also been raised. Surface hydrophobicity and ionic factors have been proposed as potential adherence mechanisms (Kabir and Ali, 1983) and non-specific adherence has been proposed as a general means by which pathogenic bacteria might avoid host resistance due to the loss of a single receptor (Attridge and Rowley, 1983b). Variations

in the glycosylation patterns of enterocyte glycoproteins between individuals have been clearly demonstrated (Lobley, 1991), adding to the attractiveness of the idea of non-specificity of adherence, but it is also conceivable that the large number of adhesins possessed by *V. cholerae* offer an alternative strategy to non-specific adherence for circumventing this problem. A similar strategy is again observed with uropathogenic *E. coli*, which express adhesins of differing receptor specificity that are presumably required for colonisation of different regions of the urinary tract (Lund *et al.*, 1988).

1.3.3.4 Modification of host cells

1.3.3.4.1 General considerations

The consequences of bacterial adherence for the host tissues are poorly understood. Studies with atoxigenic *E. coli* have shown that colonisation *per se* can cause diarrhoea in rabbits (Wanke and Guerrant, 1987; Schlager *et al.*, 1990). The possibility of a similar effect attributable to *V. cholerae* O1 adherence causing diarrhoea in humans has been raised (Levine *et al.*, 1988). Colonisation of the upper small intestine is not normal in a 'disease-free' state (Drasar and Barrow, 1985), but the effects of colonisation are generally assumed to be limited to those mediated by secreted toxins. This is largely due to the ability of purified toxin to elicit a cholera-like diarrhoea (Levine *et al.*, 1983; Spangler, 1992). However, microbial adhesins have been shown in other systems to affect host (and microbial) cell function and although no such effects have yet been identified for *V. cholerae* they are proving an expanding area of interest (reviewed in Wick *et al.*, 1991; Hoepelman and Tuomanen, 1992).

Nutrients, most notably iron, can be limiting in the small intestine, despite it being the site of most nutrient uptake (Crosa, 1989; Williams and Roberts, 1989). Therefore, the secretion of toxins and other extracellular factors can be seen as a way of altering this situation to establish conditions more favourable for bacterial growth. *V. cholerae* is a prolific producer of extra-cellular proteins (Manning, 1988) many of which are implicated in the process of nutrient acquisition.

1.3.3.4.2 Cholera enterotoxin

Cholera toxin has been subjected to more scrutiny than almost any other bacterial toxin. As noted above, cholera-like disease can be induced by the purified toxin (Levine *et al.*, 1983; Spangler, 1992). The atomic structure of a related heat-labile toxin (LT) from a porcine strain of *E. coli* has recently been elucidated confirming the long-predicted structure of the toxin, in which a single active subunit is bound to a "doughnut shaped" pentamer of binding subunits (Sixma *et al.*, 1991; van Heyningen, 1991). An exhaustive review on the structure and function of cholera toxin has recently appeared (Spangler, 1992) which details the process by which the holotoxin binds to enterocytes *via* GM1 ganglioside, the active subunit is internalised and ADP-ribosylates a G protein associated with the adenylate cyclase complex, thus triggering (by as yet unclear means) the metabolic cascade that leads to hypersecretion of fluids and electrolytes.

The identity of GM1 ganglioside as the receptor for cholera toxin has been recognised for about twenty years and, unlike the related LT toxin from *E. coli*, cholera toxin does not bind to other gangliosides (reviewed in Spangler, 1992). However, interactions between cholera toxin and both blood group A active glycolipids (Bennun *et al.*, 1988) and glycopeptides (Monferran *et al.*, 1990) isolated from pig intestinal mucosa have been demonstrated and in the latter case this interaction was shown to inhibit cholera toxin binding to erythrocyte ghosts. This inhibitory effect is also expected to occur in individuals who are secretors (Monferran *et al.*, 1990) and offers a rational explanation for the observation that cholera is most severe in blood group O individuals (Barua and Paguio, 1977; Chaudhuri and De, 1977) and that group O individuals are more susceptible to cholera than group A or B positive individuals (Glass *et al.*, 1985). It seems that this relative severity of disease has profoundly affected the population in the Ganges delta, which has the lowest percentage of group O individuals of anywhere in the world (Glass *et al.*, 1985).

Purified toxin also causes many of the pathological changes associated with experimental cholera including partial villous denudation and epithelial cell extrusion (Teppema *et al.*, 1987). Changes in cellular morphology are detected after intoxication *in vitro*, at least with a human colonic carcinoma cell line that forms "crypt like" organoids, with

cells forming large membrane vesicles and some cells swelling and losing microvilli (Barkla *et al.*, 1992). The organoid lumen also dilates rapidly, in association with fluid outpouring from the cells, but whether similar changes occur *in vivo* has yet to be determined (Barkla *et al.*, 1992). Examination *in vitro* of an immortalised goblet cell line has revealed that while cholera toxin affects goblet cells, intoxication does not lead to hypersecretion of mucous, raising the possibility that the hypersecretion observed *in vivo* is the result of an indirect effect of cholera toxin on goblet cells (Lencer *et al.*, 1990), or alternatively may be the result of some other *V. cholerae* factor. An effect of cholera toxin on epithelial cells of the large intestine, the site of fluid uptake in the normal state (Drasar and Barrow, 1985), has not been documented, although it is known to be exposed as concentrations of cholera toxin of greater than 100 ng/ml in diarrhoeal fluid have been recorded (Holmgren, 1981).

Despite our knowledge of the effects of cholera toxin, the role of the toxin in the life cycle of *V. cholerae* is not clear. Some evidence has been presented that suggests cholera toxin expressing strains colonise the intestinal epithelium more efficiently than non-toxigenic strains (Mekalanos, 1985), but this could be due to loss of *cep* (Pearson *et al.*, 1993) in the (Pierce et al., 1985) *ctx⁻* strains used in the earlier study. It is possible that the elicited outpouring of electrolytes creates conditions that are optimal for the growth of the moderately halophilic and alkalophilic *V. cholerae*. As conditions are optimal for growth of *V. cholerae* but not for *E. coli*, it has been proposed that this might explain why cholera is a much more severe disease than diarrhoea caused by enterotoxigenic *E. coli*, which are equipped with a similar toxin (Abdulkadir, 1992). Cholera toxin does not cause malabsorption - potassium uptake is accelerated by crypt-like cells *in vitro* after intoxication (Barkla *et al.*, 1992) - and the whole rationale behind oral rehydration therapy is that absorption will continue despite the hypersecretion (Drasar and Barrow, 1985). Another possible role for cholera toxin in the life cycle of *V. cholerae* might be that it is analogous to a bacteriophage lysis gene - enabling the expulsion and spread from the host of progeny cells (as originally suggested by Finkelstein, 1973). It has been noted that cholera toxin expressing strains are excreted in higher amounts by the host than non-expressing strains (Levine *et al.*, 1988).

Cholera toxin gene expression is regulated by the ToxRST regulatory proteins (DiRita, 1992) that will be described in detail in Section 1.4.1. This appears to be the predominant control mechanism as strains differing in the amount of cholera toxin produced show a corresponding difference in *ctx* mRNA levels (Mishra and Holmes, 1987). The genes for the two subunits overlap but a translational coupling effect is not seen: the distal *ctxB* cistron is translated at a much higher efficiency than the proximal *ctxA* cistron, leading to the observed stoichiometry of subunit production (Mekalanos *et al.*, 1983). Cholera toxin expression is also subject to copy number control, as the genes are flanked by copies of a 2.7 kb repetitive sequence called RS1 (Mekalanos, 1983) and gene dosage can be altered by recombination between these RS1 elements, leading to amplification or, at lower frequency, deletion of the intervening sequences (Goldberg and Mckalanos, 1986). This process was initially claimed to be dependent on a functional *recA* gene (Goldberg and Mekalanos, 1986) but has subsequently been described as *recA* independent, although it occurs at a lower frequency in *recA* strains (Pearson and Mekalanos, 1989; Pearson *et al.*, 1993). Similarly, amplification of sequences between RS1 elements has been observed when cloned on a multicopy plasmid in *E. coli* (Filkova *et al.*, 1988), again in a *recA* independent manner (Filkova *et al.*, 1989). RS1 appears to encode a site specific recombination system, reminiscent of a bacteriophage integration system, which is predicted to be capable of efficient conversion of non-toxigenic *V. cholerae* into toxigenic *V. cholerae* via integration at an 18 bp (*attRS1*) sequence (Pearson *et al.*, 1993). This interconversion appears to occur naturally, at least in the Gulf of Mexico (Pearson and Mekalanos, 1989), as the only differences between toxigenic and non-toxigenic strains isolated there appear to be the presence or absence of *ctx* genes (Almeida *et al.*, 1989). In addition to this role in propagation of *ctx* in *V. cholerae* populations, a role for *ctx* gene dosage during the course of an infection has also been demonstrated, with the detection of amplification of *ctx* during intestinal passage of experimental animals (Mckalanos, 1983). However, this seems to be a secondary role for RS1 in pathogenesis, given that RS1 associated deletion events also occur, albeit at a lower frequency than amplification events (Goldberg and Mekalanos, 1986).

1.3.3.4.3 Other factors affecting cholera toxin expression

A number of loci have been identified that are vital for production of active cholera toxin in the extracellular milieu. Proteolytic activation of the A subunit has been discussed in section 1.3.3.3.2. Correct formation of enterotoxin requires the catalysed formation of an intramolecular disulphide bond in the B subunit (Yu *et al.*, 1992) by a homologue of the *E. coli* DbsA protein (Bardwell *et al.*, 1991; Peek and Taylor, 1992; Yu *et al.*, 1992). The possibility that other secreted factors, in addition to TCP, require this activity for correct folding has been raised (Peek and Taylor, 1992) – this seems likely to be a general phenomenon of protein export – as DbsA is required for disulphide bond formation in several periplasmic and outer membrane proteins (Bardwell *et al.*, 1991) and for the LT toxin of *E. coli* (Yu *et al.*, 1992).

A gene essential for secretion of cholera toxin through the outer membrane of *V. cholerae* has also been identified. A *V. cholerae* mutant that accumulates cholera toxin in the periplasm was identified (Holmes *et al.*, 1975; Hirst and Holmgren, 1987) and shown to be defective in a protein homologous with a family of genes required for extracellular secretion of proteins in a number of Gram negative bacteria (Sandkvist *et al.*, 1993). This protein is encoded by one of a cluster of genes, designated *eps*, that are homologous to genes associated with protein secretion in a number of other Gram negative bacteria (Overbye *et al.*, 1993). Proteins homologous to other members of export gene families (Pugsley, 1992) are involved in TCP biogenesis (Kaufman *et al.*, 1993; Ogierman *et al.*, 1993) and the question of how *V. cholerae* actively secretes the number of proteins it does (Manning, 1987; Manning, 1988) is an area of current interest.

Another factor implicated in cholera toxin activity is a secreted neuraminidase, which has been proposed to convert enterocyte polysialogangliosides to GM1, thus increasing the number of available receptor sites for cholera toxin (Kabir *et al.*, 1984). The enzyme is encoded by a gene, *nanH*, which appears to be part of an operon whose expression is controlled by sialic acid availability (Vimr *et al.*, 1988). The construction of isogenic *nanH*[+] and *nanH* strains has enabled the demonstration of increased cholera toxin binding to mouse fibroblasts *in vitro* in the presence of neuraminidase: similar effects were not apparent in a

sealed mouse model (Galen *et al.*, 1992). It is not clear if the sialic acid level *in vivo* is sufficient to activate *nanH* expression (Vimr *et al.*, 1988), but in the light of the mouse experiment, the importance of an active neuraminidase to cellular sensitivity to cholera toxin remains unclear.

1.3.3.4.4 Haemolysin

The ability or otherwise to lyse sheep red blood cells was the first "biotyping" characteristic recognised for *V. cholerae* O1, with the isolation of haemolytic *V. cholerae* from asymptomatic carriers at the El Tor quarantine station in Egypt (Pollitzer, 1959), although the picture has since been complicated by the isolation of non-haemolytic El Tor strains (De Moor, 1963) and an unrelated haemolysin from classical strains (Richardson *et al.*, 1986). The El Tor haemolysin has been shown to be a potent toxin, possessing both enterotoxic and cytotoxic activity (Honda and Finkelstein, 1979; Hall and Drasar, 1990; Alm *et al.*, 1991). The protein, designated HlyA, is produced as an 82 kDa pre-pro-protein with a signal peptide that is cleaved on secretion to give a 79 kDa pro-protein (Goldberg and Murphy, 1985; Rader and Murphy, 1988; Alm *et al.*, 1988; Yamamoto *et al.*, 1990a). This product is then proteolytically processed after secretion to give a 65 kDa mature form in *V. cholerae* (Alm *et al.*, 1988; Yamamoto *et al.*, 1990a) that is not readily detected in *E. coli* clones (Mercurio and Manning, 1985). Amino-terminal sequencing of the processed protein has demonstrated that the amino-terminal portion of the pro-protein is cleaved, but has led to conflict over the site of processing, with sizes for the mature form of 65 kDa (Yamamoto *et al.*, 1990a) or 69 kDa (Hall and Drasar, 1990) being proposed. Processing seems to be important for haemolytic activity as the unprocessed (79 kDa) form purified from *E. coli* has less than 5% of the activity of the processed form purified from *V. cholerae*, although the possibility of another post-translational modification required for full activity has not been discounted (Yamamoto *et al.*, 1990a). Classical strains possess the *hlyA* gene (Goldberg and Murphy, 1984; Brown and Manning, 1985) which appears to be 'defective' in the case of the strain 569B, in that an 11 bp deletion leads to production of a 26 kDa protein called HlyA* (Rader and Murphy, 1988; Alm *et al.*, 1988) which is haemolytic on rabbit red blood cells but not sheep red blood cells (Richardson *et al.*, 1986) and possesses cytotoxic activity (Alm *et al.*, 1991). All classical strains examined to date appear to bear an identical lesion (Alm and

Manning, 1990b). Two linked genes also appear to be involved in the haemolytic phenotype: *hlyB*, which appears to be required for maximal production and secretion of HlyA (Alm and Manning, 1990a) and which shares significant sequence similarity with the methyl chemotactic proteins (MCP) Trg, Tsr, Tar and Tap (Manning, 1994) involved in chemotaxis in *E. coli* (Bollinger *et al.*, 1984; Ames and Parkinson, 1988); and *hlyC*, which encodes a 15 kDa protein (Alm, 1989) that shows similarity to a *P. fragi* phospholipase (Kugimiya *et al.*, 1986).

Regulation of *hlyA* expression is genetically complex. The level of available iron in the growth medium affects *hlyA* expression, *via* the *fur* regulon (Stoebner and Payne, 1988), with *hlyA* transcription being derepressed under conditions of iron starvation (S.G. Williams, unpublished; Chapter 6). Expression also requires the presence of a transcriptional activator protein HlyU (Williams and Manning, 1991), although the environmental signals regulating HlyU activity are not clear (S.G. Williams, personal communication). A third locus, *hlyR* has been identified by mutation that is closely linked to, but distinct from, *toxR* (von Mechow *et al.*, 1985). However, an *hlyR* strain (Hly⁻) can be restored to a haemolytic phenotype by transduction with an *hlyA* transducing phage (Varcoe, 1992) so the role of this locus remains unclear. Transcription of *hlyA* initiates from a promoter 430 nucleotides upstream of the translation start site (Williams and Manning, 1991) and a putative Fur binding site is over 100 nucleotides upstream from this point (Alm, 1989; Williams and Manning, 1991). Furthermore, non-haemolytic El Tor strains produce essentially identical amounts of *hlyA* mRNA (Rader and Murphy, 1988; S.G. Williams, unpublished) indicating that post-transcriptional control mechanisms are also important. Transcription of *hlyA* appears to be unaffected in a *hlyR* mutant, but is markedly decreased (but not absent) in an *hlyU* mutant, which shows an intermediary haemolytic phenotype (S.G. Williams, unpublished). Transcription of *hlyB* appears to occur from the upstream *hlyA* promoter (S.G. Williams, unpublished)

A postulated role for the haemolysin in infection is to provide iron for the growth of *V. cholerae* (Stoebner and Payne, 1988). There is also some evidence for a role as an enterotoxin. Cholera toxin negative strains still cause diarrhoea and the introduction of a *hlyA*

lesion did not affect this reactogenicity, leading to the suggestion that *hlyA* is unimportant for pathogenesis (Levine *et al.*, 1988). However the mutation used in that study is similar to the classical mutation (*i.e.* still allows production of the amino-terminal portion of the protein) and inactivation of *hlyA** leads to a reduction of the virulence of the classical strain 569B in the infant mouse model (Alm *et al.*, 1991).

1.3.3.4.5 Zonula occludens toxin (Zot)

The residual diarrhoea associated with the cholera toxin delete strain mentioned above (Levine *et al.*, 1988) led to the identification of another toxin, Zot, which markedly affects the intercellular tight junctions (zonula occludens) of enterocytes, and thus is another candidate for the observed reactogenicity of *ctx*-negative strains (Fasano *et al.*, 1991). The gene encoding the toxin has been cloned and shown to lie adjacent to *ctx*, within the RS1 repeats, and encodes a 45 kDa polypeptide (Baudry *et al.*, 1992). However, as the active polypeptide appears to be less than 30 kDa in size (Fasano *et al.*, 1991) it appears that proteolytic processing is important for Zot activity as well (Baudry *et al.*, 1992). Control of *zot* gene expression has not been closely examined to date, but appears to be independent of *toxR* (Baudry *et al.*, 1992).

1.3.3.4.6 Accessory cholera enterotoxin (Ace)

Yet another enterotoxin, termed Ace, has been identified in *V. cholerae* O1, with its structural gene also lying within the region between RS1 elements along with *ctxAB*, *zot* and *cep* (Baudry *et al.*, 1991). Ace causes an increase in potential difference across epithelial cell membranes *in vitro*, and fluid efflux in rabbit ileal loops (Trucksis *et al.*, 1993). Sequencing of the structural gene has revealed significant similarity to a family of ion-transporting ATPases, for example the cystic fibrosis gene, suggesting a mechanism by which Ace causes fluid efflux (Trucksis *et al.*, 1993). Information on the control of expression of *ace* is currently unavailable.

1.3.3.4.7 Other toxins

A second haemolysin has also been detected in classical strains (Richardson *et al.*, 1986) but has not been characterised further. Toxins more familiar in other contexts have

also been identified. In a survey of a large number of fresh isolates, a single *V. cholerae* O1 strain was identified that expressed a heat stable (ST) toxin (Takeda *et al.*, 1991a; Takeda *et al.*, 1991b) common to a number of non-O1 *V. cholerae* (Ogawa *et al.*, 1990). *V. cholerae* O1 strains that produce a cytotoxin that is partially neutralised by Shiga antitoxin have also been identified (O'Brien *et al.*, 1984). Possession of such toxins would no doubt contribute to the diarrhoeagenicity (Levine *et al.*, 1988) and nutrient scavenging ability of host strains, but precise roles for these accessory toxins in the disease process remain to be elucidated.

1.3.3.4.8 Other factors

In addition to the secreted toxins described above, other proteins have been detected in culture supernatants of *V. cholerae* O1. Two secreted DNases have been identified (Newland *et al.*, 1985; Focareta and Manning, 1987). As DNA is a significant component of the mucous blanket (Ferencz *et al.*, 1980) it has been hypothesised that DNase production could aid penetration of this barrier, and also aid in the provision of nutrients. However, when strains deficient in both DNases were constructed, no difference was observed in the virulence of DNase$^+$ and DNase$^-$ strains in the infant mouse model, indicating that any role played by these enzymes in pathogenesis is, as for the neuraminidase, subtle (Focareta and Manning, 1991).

Iron acquisition is of extreme nutritional importance in pathogenesis (Williams and Roberts, 1989) and many pathogenic bacteria secrete iron binding compounds called siderophores to facilitate iron uptake (Crosa, 1989). *V. cholerae* is no exception, secreting a phenolate-type siderophore called vibriobactin when grown in iron deficient media (Sigel and Payne, 1982) and expressing a specific receptor protein for iron-vibriobactin in the outer membrane (Stoebner *et al.*, 1992). Uptake of iron-vibriobactin presumably occurs by a mechanism analogous to siderophore uptake in other Gram negative bacteria (Sigel and Payne, 1982), in which a specific outer membrane receptor protein binds and transports the iron-siderophore complex, using energy provided by the TonB protein (Crosa, 1989). A specific receptor protein, ViuA, has been identified (Stoebner *et al.*, 1992), the expression of which is negatively regulated by Fur (Butterton *et al.*, 1992). However, vibriobactin production is not essential for *V. cholerae* virulence, so other iron acquisition systems must

also be in place (Sigel *et al.*, 1985). A possible role for *hlyA* in iron acquisition has already been mentioned (Section 1.3.3.4.4), but since the iron released from lysed cells is presumably still complexed to proteins, other factors must also be required for its utilisation. *V. cholerae* can utilise iron from haemin and haemoglobin (Stoebner and Payne, 1988) and at least two proteins are associated with haem utilisation: a 77 kDa outer membrane protein that is presumably a haem receptor; and a 26 kDa cytoplasmic membrane protein that could conceivably serve TonB-like functions or act as an inner membrane receptor (Henderson and Payne, 1993). In addition, some *V. cholerae* O1 strains show specific lactoferrin binding properties (Ascencio *et al.*, 1992). A second 77 kDa iron-regulated outer membrane protein, IrgA, has been identified that is essential for virulence. This was originally proposed to be the vibriobactin receptor (Goldberg *et al.*, 1990a; Goldberg *et al.*, 1990b), but *irgA* mutants are unaffected in vibriobactin transport and also in the utilisation of iron from a number of other sources, including haemin and haemoglobin (Goldberg *et al.*, 1992). The sequence of *irgA* reveals homology with several TonB-dependent proteins, but a function for IrgA remains to be determined (Goldberg *et al.*, 1992).

A problem in the utilisation of haem from haemoglobin is the removal of the porphyrin moiety from protein. *V. cholerae* mutants that can utilise iron from haem but not haemoglobin have not been isolated (Henderson and Payne, 1993) but secreted protease-deficient mutants of *V. vulnificus* are unable to utilise haemoglobin as a sole source of iron (Nishima *et al.*, 1992). This protease has been partially purified and has been shown to be a neutral metalloprotease reminiscent of, but biochemically distinct from, the HA/protease of *V. cholerae* (Kothary and Kreger, 1985). Protease-deficient mutants of *V. cholerae* have been isolated that have lost virulence (Schneider and Parker, 1978) and although the protease identified is apparently the HA/protease (Finkelstein and Hanne, 1982), the involvement of proteases in iron utilisation is possibly of great importance. Three other biochemically distinct proteases that are insensitive to chelating agents have also been identified (Young and Broadbent, 1982), but their role in infection has not been examined. Proteolytic processing of cholera toxin (Booth *et al.*, 1984), haemolysin (Hall and Drasar, 1990; Yamamoto *et al.*, 1990a) and Zot (Baudry *et al.*, 1992), presumably all by *V. cholerae* secreted proteases, has already been mentioned.

1.3.3.5 Multiplication

Having survived the rigours of ingestion, successfully colonised the small intestine, and modified its environment to provide optimal conditions, the most important step of the disease process now occurs - bacterial multiplication. The growth of $V. cholerae$ O1 is quite spectacular $in\ vivo$: in an experiment where volunteers were fed 10^6 cells of an El Tor strain (N16961) 92% of volunteers developed diarrhoea, excreting on average 3×10^7 cells/g of stool and 4.2 l of stool, representing an average increase of approximately 10^5-fold (Levine $et\ al.$, 1988)! Considering that El Tor strains are generally considered to be less virulent than classical strains (Finkelstein, 1973; Barua and Burrows, 1974), that infectious doses in natural infections are 10^3-10^4 lower than that administered in this experiment (Levine $et\ al.$, 1983) and that stool volumes of greater than 30 litres have been recorded (Carpenter, 1992), increases of up to $\approx 10^{10}$ in cell numbers are possible. Thus, the ability to cause diarrhoea is of great reproductive advantage in comparison to vibrios which remain in the estuarine environment (Section 1.3.1).

1.3.4 Other bacterial factors important in disease

The factors described in Section 1.3.3 contribute to the persistence of vibrios in the small intestine, or to their nutritional state there, but are not the sole bacterial or host factors involved in pathogenesis. Host immunity, biotype and cell surface also influence the process, being in some ways interdependent, and will be discussed here.

1.3.4.1 Host immunity

Studies with human volunteers (Cash $et\ al.$, 1974; Levine $et\ al.$, 1983) and the epidemiology of cholera in endemic areas (Glass $et\ al.$, 1985) have shown that the disease induces protective immunity against subsequent infection and it is on this protection that hopes of effectively vaccinating against cholera have been pinned (Attridge and Rowley, 1983c; Manning, 1987; Levine $et\ al.$, 1988; Pierce $et\ al.$, 1988). A secretory antibody response is both necessary and sufficient for protection (Levine $et\ al.$, 1983; Holmgren, 1991; Winner $et\ al.$, 1991) and seems to be induced by vibrios adhering to M cells (Yamamoto and Yokota, 1989a) followed by their efficient uptake as intact cells and transfer to the Peyer's patches (Owen $et\ al.$, 1986). Antigens against which an elicited immune response is

protective fall into two classes, LPS and non-LPS (Neoh and Rowley, 1972; Holmgren *et al.*, 1977; Attridge and Rowley, 1983c), although the relative importance of these to immunity remains to be determined accurately (Levine *et al.*, 1983). LPS will be discussed in Section 1.3.4.2. Non-LPS protective antigens seem to be shared by both biotypes, at least in animal models (Sharma *et al.*, 1987). One of these has been shown to be TCP (Sharma *et al.*, 1989a; Sharma *et al.*, 1989b) and the identity of others remains an area of active research interest because of the potential applications in vaccine development (Manning, 1992). Cholera toxin also plays a role in host immunity, being the most powerful mucosal adjuvant known (Holmgren, 1991). As both anti-toxin and anti-bacterial immunity appear to be important, there has been considerable controversy as to which is more important for protection against disease. Anti-bacterial immunity alone appears capable of protecting against infection, whereas similar results have not been obtained after immunisation with purified cholera toxin, or toxoids (reviewed in Levine and Pierce, 1992). Thus, the short term gain of *V. cholerae* O1 is apparently to its long term detriment as an infected host is not available for reinfection, due either to death, or an efficient immune response against the factors necessary for establishing an infection.

1.3.4.2 Cell surface

The outer membrane of Gram negative bacteria is an asymmetric structure in which the inner sheath is composed of phospholipids but the outer sheath is composed of Lipid A, in which a disaccharide unit is attached to six or seven fatty acid residues, to which is linked, *via* a core oligosaccharide, a long chain of repeating sugar units termed the O-antigen, the whole structure being referred to as lipopolysaccharide, or LPS (Nikaido and Vaara, 1985; Manning *et al.*, 1993). LPS is highly charged and plays a major role in the permeability barrier that the outer membrane poses towards hydrophobic molecules (Nikaido and Vaara, 1985; Sukupolvi and Vaara, 1989). As noted in Section 1.3.3, *V. cholerae* appears to be an exception to the rule on the membrane composition of enteric bacteria, as topological probes detect phospholipids in the outer sheath of the outer membrane and the LPS appears to be markedly less ionic than that of the Enterobactcriaceae (Paul *et al.*, 1992). The structure of *V. cholerae* LPS is far from clear, but the O-antigen seems to consist of a homopolymer of D-perosamine that is substituted to give the sub-serotypes observed (Kenne *et al.*, 1979; Redmond, 1979;

Kenne *et al.*, 1982). LPS contains at least three antigens which are expressed in differing amounts on the three sub-serotypes: Ogawa strains express high levels of A and B antigens, and a small amount of C; Inaba strains express A and C antigens only; and Hikojima strains express high levels of all three antigens (Sakazaki and Tamura, 1971; Redmond *et al.*, 1973). Serotype conversion has been demonstrated in *V. cholerae* O1 and is the result of expression, or lack thereof, of a single gene *rfbT*, that encodes "Ogawaness" (Stroeher *et al.*, 1992). The Ogawa character is also dominant in *E. coli* clones, although a second region also appears to be required for Ogawa expression in *E. coli* (Morona *et al.*, 1990). Gene sequencing has not revealed any obvious explanation for this result (Manning *et al.*, 1993). Serotype switching seems to be a mechanism by which *V. cholerae* O1 might evade the anti-LPS immune response (Stroeher *et al.*, 1992) and has been observed in clinical cholera (Sheehy *et al.*, 1966; Gangarosa *et al.*, 1967) although some cross serotype protection, primarily due to the A antigen, has been observed in an infant mouse model (Manning *et al.*, 1986). Strains of Hikojima serotype are unstable and appear to be intermediates of serotype conversion (Manning *et al.*, 1993).

A role for LPS in outer membrane integrity is apparent, as rough strains are defective in motility, due to flagellar defects (Manning *et al.*, 1993) and also appear to be defective in appropriate surface localisation of a number of proteins, including TCP (Manning, 1991). Rough mutants are markedly attenuated, which is probably due to outer membrane defects (Manning *et al.*, 1993). LPS has also been proposed as a *V. cholerae* adhesin (Freter and Jones, 1976; Chitnis *et al.*, 1982) but this appears not to be the case (Manning, 1991; Manning *et al.*, 1993).

The expression of membrane proteins may also affect the virulence of *V. cholerae* O1. The presence of the conjugative plasmid P (Bartowsky *et al.*, 1987) has been shown to lead to the reduction of virulence of the hypertoxigenic classical strain 569B, possibly by causing a reduction in TCP expression, although a mechanism by which this occurs has not been proposed (Bartowsky *et al.*, 1990). The presence of V, an unrelated conjugative plasmid originally isolated from a non-O1 strain (Bhaskaran and Sinha, 1971) does not affect virulence (Bartowsky *et al.*, 1990), but the presence of pPH1JI, a conjugative plasmid used in

this laboratory in intermediate stages in the construction of mutant *V. cholerae* strains (Alm and Manning, 1990a; Focareta and Manning, 1991) is also attenuating (Focareta, 1989). It is possible that attenuation in the latter case is due to the cost of gentamycin resistance encoded by the plasmid, but further work is needed to unravel the underlying causes of attenuation in these cases.

The expression of at least two outer membrane proteins, OmpU and OmpT, has been shown to be regulated, at least in part, by ToxR (Taylor *et al.*, 1987; Miller and Mekalanos, 1988). It is conceivable that these and/or other outer membrane proteins (other than those involved in iron acquisition) are also involved in pathogenesis, but none have been implicated to date. Attempts to isolate Tn*phoA* insertions in *ompT* and *ompU* have been unsuccessful, possibly because the desired mutations are lethal (Peterson and Mekalanos, 1988).

1.3.4.3 Biotype

V. cholerae O1 exists as two biotypes, classical and El Tor, of which strains of classical biotype generally cause more severe disease (Finkelstein, 1973; Barua and Burrows, 1974). The first six pandemics are presumed to have been ~~were~~ caused by classical strains (Finkelstein, 1973; Finkelstein, 1988) and El Tor strains first emerged as a major cause of cholera with the seventh pandemic (Finkelstein, 1988). The issue has been further complicated by the re-emergence of classical strains in Bangladesh (Siddique *et al.*, 1991) and by the appearance of a new serotype (Section 1.5). An adequate explanation for this switch in biotype has yet to be offered, although a link between biotype and the relative severity of disease has been proposed as an underlying cause (Ewald, 1993).

Consequently, the basis for biotype differences remains an interesting problem. Of the original distinguishing features, the sensitivity to various vibriophages and polymyxin B resistance remain the most reliable means of determining biotype (Baumann *et al.*, 1984). Biotype characteristics, however, may be due to one or a few genes. Mitra (1989) identified mutants resistant to an El Tor-specific vibriophage, PS166, which were smooth, but appeared to have switched biotype, in that they had lost both polymyxin B resistance and the ability to agglutinate chicken erythrocytes, and had gained sensitivity to a classical group 4 phage. As

these mutants arose spontaneously (Mitra, 1989), it seems reasonable to assume that a single gene or biochemical pathway is affected. Also, expression of a 14 kDa cytoplasmic membrane protein in El Tor but not classical strains (although both appear to have the gene) confers resistance to the Group IV (classical) phage φ149, presumably by inhibiting binding of replication intermediates to the cell membrane (Biswas *et al.*, 1992). This phenomenon may also involve the *sulA* gene product, an inhibitor of cell division whose gene partially overlaps the 14 kDa protein gene at the distal end but is encoded on the opposite strand (Biswas *et al.*, 1992). Unfortunately, these authors have not monitored polymyxin B resistance, or haemagglutination. It is of interest to determine whether these independently described effects are due to the same gene or not, in particular if the 14 kDa protein is involved in polymyxin B resistance. The mechanism of polymyxin B resistance is unclear, but mutations that lead to a decrease in LPS acidity confer polymyxin B resistance on *S. typhimurium* (Sukupolvi and Vaara, 1989). The cause of polymyxin B resistance in El Tor *V. cholerae* has not been investigated, but biotype-specific antigenic changes in LPS are not apparent (Manning *et al.*, 1993). Purification of LPS from El Tor strains has been complicated by the copurification of uncharacterised material (R. Morona, personal communication) and the possibility that this material, which may be a polysaccharide, is responsible for biotype polymyxin B resistance should be investigated. However, as the ultimate target of polymyxin B is the cytoplasmic membrane (Sukupolvi, 1989) it is also possible that differences in the cytoplasmic membrane, for example the presence or absence of the 14 kDa protein, are responsible.

Given that the phenotypically complex biotype differences may well result from simple genetic differences, what role do they play in the pathogenesis of *V. cholerae* O1? The greater severity of disease as caused by classical strains has been recognised for as long as biotype differences have been known (Pollitzer, 1959). More recently, epidemiological studies in Bangladesh have suggested that episodes of classical cholera will elicit protective immunity against reinfection, but that episodes of El Tor cholera may not, although the underlying cause is not clear (Clemens *et al.*, 1991). DNA-DNA hybridisation studies have not revealed large differences between classical and El Tor strains (Baumann *et al.*, 1984). Studies characterising differences in virulence determinants between biotypes have shown

clear results. Sequencing of *clxB* from 45 isolates (representing 26 countries and a 70 year period) has revealed three alleles in *V. cholerae* O1: one found in all classical strains and El Tor strains isolated in the Gulf of Mexico; one found in El Tor strains isolated in Australia; and one found in other El Tor strains, including those responsible for the current pandemic (Olsvic *et al.*, 1993). Sequencing of *tcpA* from numerous strains of both biotypes has revealed two alleles that are biotype-specific (Iredell, J. and Manning, P.A., manuscript in preparation). In addition, all classical strains examined to date have the same *hlyA* lesion (Alm and Manning, 1990b), although El Tor strains are more diverse (Section 1.3.3.4.4). There are also clear biotype differences in restriction/modification systems (Imbesi and Manning, 1982). Thus, biotypes appear to be clonal, with classical strains showing less diversity than El Tor strains in the factors thus far examined. Because it appears possible that classical strains can arise from El Tor through mutation (Mitra, 1989), it is proposed that classical strains represent a clone of *V. cholerae* O1 that has arisen because of its greater ability to cause severe disease, with the concomitant reproductive advantage (Section 1.3.3.5). Rigorous analysis of the clonal structure of *V. cholerae* O1 populations would hopefully also lead to an understanding of the reasons behind the appearance of El Tor strains in the current pandemic.

1.4 Regulation of virulence genes

Virulence determinants are defined as those factors contributing to infection (and disease) with the exception of housekeeping functions (Mekalanos, 1992). In the case of *V. cholerae* O1 these can conveniently be defined as those factors described in Section 1.3.3. Some discussion of the regulation of expression of these factors has already been undertaken, but in a factor by factor manner. The expression of the regulatory factors involved will be considered in this section.

1.4.1 The ToxRST regulon

Examination of cholera toxin gene expression was initiated by the isolation of regulatory mutants in distinct loci designated: *tox*, mutations in which cause a dramatic decrease in toxinogenicity (Vasil *et al.*, 1975; Baine *et al.*, 1978); and *htx*, mutations in which cause hypertoxigenicity (Mekalanos *et al.*, 1979; Mekalanos and Murphy, 1980). Strains that

bear the *htx* lesion show reduced growth rate and accumulate revertants, one class of which (designated *ltx*) appears to be allelic with *htx* (Mekalanos and Murphy, 1980). Both loci are unlinked to *ctx* (Sporecke *et al.*, 1984).

Subsequent analysis has shown that *tox* consists of two genes that are co-transcribed, *toxR* (Miller and Mekalanos, 1984; Miller *et al.*, 1987), and *toxS* (Miller *et al.*, 1989). The ToxR protein has unusual features for a regulatory protein, being associated with the cytoplasmic membrane with both cytoplasmic and periplasmic domains (Miller *et al.*, 1987). The sequence of the amino-terminal cytoplasmic domain shows similarity with a number of transcriptional activators that are members of the two-component regulator family (Miller *et al.*, 1987; Otteman *et al.*, 1992). Two-component regulators are involved in responses to a variety of environmental stimuli in a number of bacteria and consist of a sensor protein (a histidine protein kinase) which, in the presence of an appropriate environmental signal, phosphorylates the effector protein, leading to its activation (reviewed in: Stock *et al.*, 1989; Stock *et al.*, 1990). However, ToxR shows similarity only with the DNA binding and activation domain of these effector proteins, not with the modification (phosphorylation) domain (Miller *et al.*, 1987; Otteman *et al.*, 1992). The carboxy-terminal periplasmic domain of ToxR is required for the dimerisation of monomers that leads to activity (Miller *et al.*, 1987). In wild type *toxR* strains, this dimerisation (and hence full ToxR activity) requires the ToxS protein, a second cytoplasmic membrane protein that resides "almost completely" in the periplasm, and which interacts with the periplasmic domain of ToxR to confer activity (DiRita and Mekalanos, 1991). The hypertoxigenic classical strain 569B is a *toxS* deletion mutant in which a (presumably) *toxR* mutation has led to *toxS* independence (Miller *et al.*, 1989). This is possibly due to over-expression of *toxR* in 569B obviating the ToxS requirement for ToxR dimerisation (Miller *et al.*, 1989). Details of the interaction between ToxR and ToxS have been reviewed recently (DiRita, 1992).

ToxR has been reported to bind to a heptameric sequence (TTTTGAT) that occurs three to eight times immediately upstream of the -35 region of the *ctx* promoter (Miller *et al.*, 1987; Otteman *et al.*, 1992) but which has not been identified in association with any other *toxR*-dependent gene. Furthermore, ToxR does not directly activate other members of the

ToxR regulon characterised to date (DiRita, 1992). This activation instead occurs *via* ToxT, identified as a gene that activates *toxR*-regulated genes in *E. coli* in the absence of other *V. cholerae* regulatory factors (DiRita *et al.*, 1991). ToxT is a member of the AraC family of transcription activators (Ogierman and Manning, 1992a; Higgins *et al.*, 1992) and acts in a regulatory cascade: ToxR (in an as yet undefined manner) activates expression of *toxT*, the product of which then activates *toxR* dependent genes (DiRita *et al.*, 1991; DiRita, 1992). Although complementation experiments analogous to those showing that *toxRS* form the *tox* locus (Miller and Mekalanos, 1984) have not been reported, it is possible that *toxT* is equivalent to the *htx* locus (Mekalanos *et al.*, 1979; Mekalanos and Murphy, 1980). This premise is based on the observation that ToxT can activate *ctx* expression in the absence of ToxR (DiRita *et al.*, 1991) and that the phenotypes of *htx* and *ltx* mutants (Mekalanos and Murphy, 1980) could be explained as *toxR* independent and down mutations in *toxT*, respectively.

Thus, the regulatory cascade whereby ToxR causes the coordinated expression of a number of virulence genes has been well documented, although the details of how ToxR activates *toxT* and how the repression of some genes, such as *ompT* (Taylor *et al.*, 1987; Miller and Mekalanos, 1988) is mediated, remain to be elucidated. What is not clear, and this could have far reaching implications for our understanding of *V. cholerae* virulence, is the environmental signal(s) that the ToxRST regulon responds to *in vivo*. A series of pairs of *in vitro* growth conditions has been defined in which ToxR is "on" or "off" by virtue of changes in medium pH, osmolarity, or availability of amino acids (Peterson and Mekalanos, 1988). Similarly, specialised rich media such as CFB (Taylor *et al.*, 1987; Sharma *et al.*, 1989a; Sharma *et al.*, 1989b) or AKI broth (Iwanaga *et al.*, 1986; Jonson *et al.*, 1990; Voss and Attridge, 1993) have been identified as conducive for cholera toxin or TCP expression, albeit with differences between biotypes. However, the great paradox remains that *in vitro*, ToxR is "off" at 37°C (Miller and Mekalanos, 1984; Peterson and Mekalanos, 1988), the temperature at which *in vivo* expression must occur for ToxR regulated genes to play a role in virulence. Temperature regulation of expression of virulence genes has been documented for a number of other mammalian pathogens and in all other cases thus far described this regulation ("on" at 37°C, "off" at 30°C) also occurs *in vitro* (reviewed in Peterson *et al.*, 1988; Maurelli, 1989;

Dorman, 1991). A model has been proposed to explain this aberrant behaviour by ToxR, in which expression of *toxR* is regulated by the level of σ^{32} (RpoH), by virtue of the proximity of the *toxR* promoter to a divergent heat shock promoter (Parsot and Mekalanos, 1990). These authors report a two fold difference in activity of a *toxR-lacZ* reporter construct between 30°C and 37°C, a temperature difference that leads to a roughly ten-fold difference in cholera toxin expression (measured as "blueing units": Richardson, 1969). On the other hand, growth in liquid media at a starting pH of 6.5 or 8.4, growth conditions that also lead to an approximately ten-fold difference in the expression of a number of ToxR-regulated genes (Peterson and Mekalanos, 1988), leads to no change in activity of a *toxR-lacZ* reporter construct (Parsot and Mekalanos, 1990). This difference leads to the conclusion that the response of the ToxR regulon to changes in temperature and medium pH occur by two different mechanisms, namely, that response to temperature involves changes in the level of *toxR* transcription, whereas response to changes in pH does not. However, *toxR* transcription is still significant at 37°C (Parsot and Mekalanos, 1990). Therefore, it seems likely that the difference in transcription of *toxR* between 30°C and 37°C *in vitro* is coincidental, and that the true cause for the aberrant temperature regulation of ToxR activity *in vitro* lies elsewhere.

A clue for the resolution of this problem might come from the recent identification of *toxR* and *toxS* homologues in *V. parahaemolyticus* that activate expression of thermostable direct haemolysin genes in a manner apparently analogous to activation of *ctx* expression in *V. cholerae* O1 (Lin *et al.*, 1993). *V. parahaemolyticus* ToxR,S shows a similar response to *V. cholerae* ToxR,S to environmental signals such as medium osmolarity and pH, but does not show a similar temperature sensitivity, being active at 37°C *in vitro*. Unfortunately, the sequence reported by Lin *et al.* (1993) commences three nucleotides downstream of the nucleotide equivalent to the start of the divergent heat shock promoter in the *V. cholerae toxR* sequence (Parsot and Mekalanos, 1990), so comparison to check whether an equivalent divergent gene also exists in the *V. parahaemolyticus toxR* locus is not possible. The *V. cholerae* ToxR must play some role in virulence as a *toxR* mutant has been shown to be completely attenuated in humans (Herrington *et al.*, 1988). Seemingly contradictory to this role in coordinating expression of virulence genes has been the identification of *aldA*, the gene encoding an aldehyde dehydrogenase as being subject to ToxR control (Parsot and

Mekalanos, 1991). In addition, a group of genes (termed *tag*, for ToxR-activated genes) has been identified by Tn*phoA* mutagenesis, whose expression is controlled by ToxR but mutations in which do not affect the virulence of *V. cholerae* (Peterson and Mckalanos, 1988). Perhaps an understanding of the functions of *tag* gene products and of how ToxR control of metabolic genes such as *aldA* fits into the 'big picture' will help explain the conundrum associated with ToxR activity. An alternative explanation of ToxR control may come from the observation that vibrios adherent to filamentous algae express more cholera toxin than non-adherent vibrios in the same environment (Islam, 1990). Adherence *per se* has yet to be defined as a regulator of bacterial gene expression, but *toxR* expression has been noted to be activated during the later stages of colonisation (Mekalanos, J.J., cited in Wick *et al.*, 1991). However, the signals governing ToxR activity *in vivo* require further study for elucidation.

1.4.2 Other described virulence gene regulons in *V. cholerae*
1.4.2.1 Fur

Although subscribing to the view that iron acquisition factors can be considered to be virulence determinants, Mekalanos (1992) points out that they occupy an uncertain position in the current definition of virulence determinant, as they can also be required for growth outside of the host when free iron is scarce. However, one can safely assume that the haem uptake system (Henderson and Payne, 1993), for example, does not play a role in acquiring iron from the insoluble iron oxides in an aquatic environment (Crosa, 1989) and therefore qualifies unequivocally as a virulence determinant.

The control of iron regulated factors has been most widely studied in the Enterobacteriaceae. Here, iron acquisition genes are controlled by a repressor, Fur, that in the presence of iron forms dimers that bind to a "Fur box" associated with the promoters of Fur regulated genes: under conditions of iron limitation these dimers cannot form, leading to the derepression of iron acquisition genes (reviewed in Bagg and Neilands, 1987). Recent reports complicate the Fur regulon by showing that Fur with or without iron also activates the expression of some genes in *S. typhimurium*, including a subset of genes involved in the acid tolerance response (Foster and Hall, 1992; Foster, 1993). *V. cholerae* also possesses a *fur*

gene (Litwin *et al.*, 1992) which controls the expression of a number of virulence genes (Sections 1.3.3.3.4, 1.3.3.3.7). Fur activity in *V. cholerae* seems to be similar to that elucidated for the Enterobacteriaceae with the exception thus far described that expression of *fur* in *V. cholerae* is not autoregulatory (Litwin *et al.*, 1992). The expression of at least two Fur-regulated virulence genes requires additional regulatory proteins: *irgA* requires an upstream gene *irgB* (whose product is a member of the LysR family of transcription activators) for expression (Goldberg *et al.*, 1991); and *hlyA* requires an activator *hlyU* (Williams and Manning, 1991). However, these regulatory systems differ in that *irgB* expression is repressed by sufficient iron and thus *irgA* activation is a regulatory cascade (Goldberg *et al.*, 1991) whereas *hlyU* expression and HlyU activation of *hlyA* appears to be independent of iron concentration (S.G. Williams, unpublished). HlyU will be described further in Section 1.4.2.2. Thus, although Fur is a global regulator of virulence gene expression in *V. cholerae*, additional regulatory factors may provide more precise control of individual genes within this regulon (Butterton *et al.*, 1992).

1.4.2.2 HlyU

Expression of the toxin gene *hlyA* is activated by a locus called *hlyU* (Williams and Manning, 1991). HlyU is a 12 kDa protein that is a member of a family of small regulatory proteins that include the NolR (*Rhizobium meliloti*), SmtB (*Synechococcus*) and ArsR (*E. coli*, *Staphylococcus aureus* and *S. xylosus*) repressors (Williams *et al.*, 1993). In addition to *hlyA*, HlyU is required for expression of a 28 kDa secreted protein (Williams *et al.*, 1993) of unknown function (Varcoe, 1992). A *hlyU* mutant (in the El Tor strain O17) shows a reduced ability to compete with the parental strain in the infant mouse model, unlike an *hlyA* mutant, indicating that HlyU may be required for expression of factors that are required for maximum colonisation and growth *in vivo* (Williams *et al.*, 1993). The signal responsible for triggering HlyU activity is currently unknown.

1.4.3 Gene expression *in vivo*

By way of experimental convenience, the bulk of the work examining gene expression in pathogenic bacteria is performed *in vitro*, but it is the behaviour of pathogens *in vivo* that should be the focus of such studies (Smith, 1990). Studies of the pathogenesis of *V. cholerae*

O1 are no exception, so it remains for us to examine what role is played by the regulatory systems described above, and additional ones, during the course of an infection.

As noted previously (Section 1.4.1) the temperature sensitivity of ToxR activity seemingly precludes it from a role in virulence, yet a *toxR* mutant is attenuated in humans (Herrington *et al.*, 1988). Similarly, the expression of cholera toxin (Section 1.3.3.4.2) and TCP (Section 1.3.3.3.4) *in vivo*, both of which require ToxR for expression *in vitro* provides a clear role for the ToxRST regulon in expression of virulence genes *in vivo*. At least some iron-regulated outer membrane proteins are expressed *in vivo* (Sciortino and Finkelstein, 1983) and the scarcity of free iron in the intestine (Crosa, 1989; Williams and Roberts, 1989) must lead to the induction of Fur-repressed genes *in vivo*, at least during the initial stages of an infection. HlyU activity has not been examined *in vivo*, but the attenuation due to an *hlyU* mutation indicates that its activity is important, at least in O17 (Williams *et al.*, 1993). As noted in Section 1.3, numerous virulence determinants are not controlled by any of these regulators, so the possibility that additional regulatory factors also come into play in the intestine must not be discounted. Experiments specifically examining gene expression *in vivo* have been extremely limited to date, and have been performed mainly with vaccine development, rather than biological activity, in mind. It is known that *V. cholerae* O1 expresses cell surface antigens *in vivo* that are not expressed under a variety of *in vitro* growth conditions and that antibodies to these "*in vivo* antigens" form a significant component of anti-*V. cholerae* antisera generated by immunisation with whole cells (Richardson *et al.*, 1989; Jonson *et al.*, 1989b), but factors controlling the expression of these antigens have not been examined.

Consideration of the signals governing the activities of regulatory genes (even though they are largely unknown) and of the virulence determinants expressed by *V. cholerae* (Section 1.3) suggests that the 'linear' view of a *V. cholerae* infection currently favoured (Section 1.1) should be replaced by a life cycle model, reminiscent of those proposed for many viruses. This involves consideration of the many virulence determinants described in Section 1.3 as falling under the heading of 'early', 'middle' or 'late' functions, and their being regulated *in vivo* as such. Under this framework, we can see how iron induced factors might

be 'early' functions, namely providing a nutrient required for growth along with (at least some) colonisation factors, which are obviously required early in an infection, how ToxR,S,T-induced factors (presumably co-activated with cholera toxin, which appears approximately 9-12 hours post-infection: Teppema *et al.*, 1987) might be considered 'middle' functions and the hypothesised "detachase" role for the HA/protease (Finkelstein *et al.*, 1992) can be considered a 'late' function, allowing progeny cells to depart the host on the wave of the cholera toxin-induced diarrhoea. A candidate for an 'early' adhesin has been proposed in the FSHA (Section 1.3.3.3.3), which is expressed transiently in early-exponential growth phase *in vitro*, and possibly *in vivo* as well (Jonson *et al.*, 1990). Obviously, this hypothesis as presented here is a gross simplification, but the absence of any real data on gene expression during the course of an infection precludes anything more concrete at this stage.

Recent developments have made experiments examining gene expression *in vivo* possible. A genetic screen developed by Mckalanos and co-workers for identifying *in vivo* induced genes in *S. typhimurium* should provide a sensitive method for isolating genes whose expression is induced *in vivo* in almost any pathogenic bacterium (Mahan *et al.*, 1993). The main limitation of this system is that a bias may exist in genes identified using it toward 'early' functions, as the complementation of a metabolic mutant that it requires may need to occur before successful infection can take place, and the mutant thus be identified. Similarly, *in situ* hybridisation techniques are now sufficiently sensitive to allow the detection of single genes in eukaryotic cells (reviewed in Lawrence, 1990). The adaptation of these methods to the analysis of gene expression during the course of an infection should allow the temporal and spatial pattern of gene expression of *V. cholerae* to be followed in the intestine. However, the practice of some authors of collecting vibrios from diarrhoeal fluid and referring to them as expressing "*in vivo*" antigens should be viewed with caution, as there is no guarantee that the progeny cells being excreted by the host are expressing the same factors as those cells adherent to the intestinal epithelium and actually causing disease. Addressing the problems of expression of virulence determinants *in vivo* will hopefully lead to a far better understanding of the life cycle of *V. cholerae*.

1.5 An eighth cholera pandemic?

The straightforward picture of *V. cholerae* O1 strains causing epidemic cholera and non-O1 strains causing sporadic diarrhoeal disease was recently disturbed by the emergence of an "untypable" non-O1 *V. cholerae* epidemic, initially in Madras in southern India, in October, 1992 (Ramamurthy *et al.*, 1993). The epidemic rapidly spread to the Bay of Bengal (Ramamurthy *et al.*, 1993; Albert *et al.*, 1993b) and more recently to Thailand (Chongsa-nguan *et al.*, 1993). The strains causing the epidemic have been assigned to a new serogroup O139 synonym Bengal (Shimada *et al.*, 1993) and it has been suggested that this new epidemic may represent the beginning of a new, eighth, pandemic of cholera (Chongsa-nguan *et al.*, 1993; Swerdlow and Ries, 1993). Molecular analyses of cholera toxin and TCP produced by O139 strains (Hall *et al.*, 1993) and of general typing characteristics, such as variable haemolysis of sheep red blood cells, polymyxin B resistance and agglutination of chicken erythrocytes (Albert *et al.*, 1993a) suggest that these strains arose as an O-antigen mutant of an El Tor strain.

The new strains also show some important differences from *V. cholerae* O1: they are resistant to O1-specific vibriophages; their pattern of antibiotic resistance is different to recent isolates of O1 *V. cholerae*; and most cases in Bangladesh are in adults, suggesting that the population is immunologically naive (Albert *et al.*, 1993a). This last finding is important for a better understanding of host immunity to cholera, as although the O139 strains are hypertoxigenic (Albert *et al.*, 1993a) it seems that the antitoxin immunity prevalent in the population of Bangladesh is not protective. Similarly, although these strains readily produce TCP *in vitro* (Hall *et al.*, 1993), pre-existing immunity is not observed, suggesting that contrary to the situation in the infant mouse model (Section 1.3.3.3.4), TCP may not be a protective antigen in humans. Thus the preliminary study of this recent outbreak has shed light on some long-standing problems in cholera research, namely that an anti LPS immune response may be the best host defence. Such information will not however immediately help people in areas threatened by the epidemic, for whom the most effective preventative measures will remain the elusive improvements in water quality.

1.6 Aims of this study

Franzon (1988) undertook a characterisation of the MFRHA (Section 1.3.3.3.6) in which the structural gene was cloned and sequenced, and a mutant constructed that caused attenuation of the classical strain 569B in the infant mouse model. However, several questions as to the nature of the MFRHA and the control of its expression were left unanswered. Therefore, this study was designed initially to characterise further the cloned DNA that confers MFRHA activity on *E. coli* and to examine the expression of the genes encoding the MFRHA, with the ultimate aim of placing the MFRHA within the scheme outlined in this Chapter. Further examination of the mutant strain constructed by Franzon (1988) was also undertaken, and attempts to define the nature of the MFRHA were initiated.

The initial characterisation involved determination of the nucleotide sequence of the MFRHA structural genes, which is reported in Chapter 3. Some interesting features of the nucleotide sequence, the occurrence of numerous copies of a 124 bp direct repeat and a gene predicted to encode a lipoprotein, are examined further in Chapters 4 and 5, respectively. The two genes proposed to encode the MFRHA, here named *mrhA* and *mrhB*, are adjacent to each other, and their transcriptional organisation is examined in Chapter 6 Finally, an analysis of the contributions of *mrhA* and *mrhB* to MFRHA activity, and the role of the MFRHA in the virulence of *V. cholerae* O1, is presented in Chapter 7.

Chapter 2

Materials and Methods

2.1 Bacterial strains and growth media

Vibrio spp. and *Aeromonas* spp. strains used are listed in Table 2.1. *Escherichia coli* K-12 strains used are listed in Table 2.2. Unless otherwise stated, bacteria were cultivated on Nutrient Broth (NB), consisting of peptone (10 g/l), lab lemco (10 g/l) (Oxoid) and NaCl (5 g/l). *V. parahaemolyticus* strains were cultivated in NB containing NaCl at 20 g/l and *Aeromonas* spp. strains were cultivated in Trypticase Soy Broth (TSB) (BBL Microbiology Systems). Luria Broth (LB) and 2xTY Medium were prepared as described by Miller (1972). ψ Broth (ψB) consisted of bactotryptone (20 g/l), yeast extract (5 g/l) (Difco) and $MgSO_4$ (5 g/l) with pH adjusted to 7.6 by the addition of KOH before autoclaving. Terrific broth (TB) was 12 g bactotryptone, 24 g yeast extract (Difco) and 4 ml glycerol made up to 900 ml with distilled water and autoclaved. Before use, one tenth volume of 0.17 M KH_2PO_4, 0.72 M K_2HPO_4 (autoclaved) was added (Tartof and Hobbs, 1987). M9 Minimal Medium (Miller, 1972) was supplemented with $MgSO_4$ (0.2 mg/ml) and glucose (2 mg/ml) prior to use, unless otherwise stated. AKI broth was prepared as described (Iwanaga *et al.*, 1986) and cultures were gassed with 5% CO_2 prior to incubation (Voss and Attridge, 1993), with incubations standing for 4 hours at 30°C, followed by 16 hours at 30°C with shaking (Iwanaga *et al.*, 1986). Methionine assay medium (Difco) was resuspended to 52.5 g/l in distilled water and autoclaved.

Nutrient Agar (NA), ψ Agar, TSA, and Minimal Agar are NB, ψB, TSB, and M9 Minimal Medium, respectively, solidified with bactoagar (15 g/l) (Difco). H-Top Agar is bactotryptone (10 g/l) (Difco) and NaCl (8 g/l) solidified with bactoagar (8 g/l). MacConkey Agar was made from MacConkey Agar base (Difco) supplemented with galactose (10 g/l) (Sigma) after autoclaving.

Antibiotics were added to broth and solid media at the following final concentrations unless otherwise stated: ampicillin (Ap), 50 μg/ml; chloramphenicol (Cm), 25 μg/ml; gentamycin (Gm), 40 μg/ml; globomycin, 100μg/ml; kanamycin (Km), 50μg/ml; nalidixic acid (Nal), 20 μg/ml; rifampicin (Rif), 400μg/ml; spectinomycin (Sp), 100 μg/ml; streptomycin (Sm), 100 μg/ml, and tetracycline (Tc), 4 μg/ml.

Table 2.1 Characteristics of *Vibrio* spp. and *Aeromonas* spp. strains

A. *V. cholerae* O1 strains

strain	Biotype	Serotype	source[a] (year of isolation)
569B	classical	Inaba	a (1946)
CA401	classical	Inaba	b (1953)
CA411	classical	Ogawa	b (1953)
C21	classical	Ogawa	c (1957)
AA14041	classical	Ogawa	d (1985)
Z17561	classical	Inaba	d (1985)
O17	El Tor	Ogawa	a (pre-1965)
C5	El Tor	Ogawa	c (1957)
C31	El Tor	Ogawa	c (1957)
AA14073	El Tor	Ogawa	d (1985)
BM69	El Tor	Inaba	e (1985)
H1	El Tor	Ogawa	d (1985)

B. *V. cholerae* O1 mutant strains

strain	parent strain	genotype	source[a] (Reference)
V761	569B	*mrhB*::Km^R [pPH1JI]	f (Franzon, 1988)
V795	O17	*toxR*::Km^R	g (S.G. Williams, unpublished)
V876	O17	*hlyU*::Km^R	g (Williams *et al.*, 1993)
V885	Z17561	*toxR*::Km^R	g (S.G. Williams, unpublished)

Table 2.1 (cont.)

C. *V. cholerae* non-01 strains

strain	isolated from	source[a]
NCV165	environmental isolate	a
N34	human isolate (septicaemia)	h
N41	human isolate (traveller's diarrhoea)	h
N50	from river water	h
N125	wallaby isolate	h
BV7	environmental isolate	i
BV22	environmental isolate	i
BV41	environmental isolate	i

D. Other *Vibrio* spp. strains

species	strain	features	source[a]
V. mimicus	V800	none described	j
V. fluvialis	V564	none described	j
V. parahaemolyticus	NCTC10884	Kanagawa$^+$	k
V. parahaemolyticus	NCTC10885	Kanagawa$^-$	k

E. *Aeromonas* spp. strains

species	strain	source[a]
A. hydrophila	A006	l
A. sobria	A187	l
A. sobria	A191	l
A. caviae	V14	l
Aeromonas sp	AB1	l

Table 2.1 (cont.)

ª sources are: a = Dr. K. Bhaskaran (Central Drug Institute, Lucknow, India); b = Dr. J. Berry (University of Texas, Austin, Texas, USA); c = Dr. P. Guinée[†] (Rijksinstituut voor Volksgesondheit en Milieuhygiene, Bilthoven, Holland); d = Dr. B. Kay (International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh); e = Dr. S. Pal (National Institute of Cholera and Enteric Diseases, Calcutta, India); f = Dr. V. Franzon (University of Adelaide, Adelaide, South Australia); g = Dr. S. Williams (University of Adelaide, Adelaide, South Australia); h = Dr. P. Desmarchelier (University of Queensland, Brisbane, Queensland); i = Dr. M. Voll (University of Maryland, College Park, Maryland, USA); j = Dr. S. Attridge (University of Adelaide, Adelaide, South Auatralia); k = Institute of Medical and Veterinary Science culture collection (Adelaide, South Australia); l = Dr. M. Atkinson (University of South Australia, Adelaide, South Australia).

[†] deceased

Table 2.2 Characteristics of *E. coli* K-12 strains

strain	genotype	source (reference)
DH5	F⁻, *deoR, recA1, endA1, hsdR17 (r_k⁻,m_k⁺), supE44, thi1, gyrA96, relA1*, λ⁻	Bethesda Research Laboratories, USA
DH5α	F⁻, *φ80dlacZΔM15, Δ(lacZYA-argF) U169, deoR, recA1, endA1, hsdR17 (r_k⁻,m_k⁺), supE44, thi1, gyrA96, relA1*, λ⁻	Bethesda Research Laboratories, USA
JM101	[F', *traD36, proAB⁺, lacI⁹ZΔM15*], *supE44, thi, Δ(lac-proAB)*, λ⁻	A.V. Sivaprasad (Yanisch-Perron *et al.*, 1985)
LE392	F⁻, *supE, supF, hsdR, galK, trpR, metB, lacY*, λ⁻	L. Endquist (Maniatis *et al.*, 1982)
HB101	F⁻, *hsdS20(r_k⁻,m_k⁻), recA13, supE44, ara14, galK2, lacY1, proA2, rpsL20, xyl5, leu, mtl1*, λ⁻	R. Morona (Boyer and Roulland-Dussoix, 1969)
AAEC189	F⁻, *endA1, hsdR17(r_k⁻,m_k⁺), supE44, thi1, Δ(lacZYA-argF)U169, Δfim, recA56*, λ⁻	I. Bloomfield (Bloomfield *et al.*, 1991a)
C75a	HfrC, *tonA22, pho64ⁿ, ompF627(T2ᴿ), relA1, pit10, spoT1*, λ⁻	B. Bachman (A. Garen, unpublished)
G206	HfrC, *tonA22, phoB62, ompF627(T2ᴿ), relA1, pit10, spoT1*, λ⁻	B. Bachman (Kreutzer *et al.*, 1975)
CC118	F⁻, *araD139, Δ(ara-leu)7697, Δlac X74, phoAΔ20, galE, galK, thi, rpsE, rpoB, argE_am, recA1*, λ⁻	J. Beckwith (Manoil and Beckwith, 1985)
CB806	F⁻, *ΔlacZ, galK, rpsL, thi, phoA8, recA56*, λ⁻	C. Beck (Schneider and Beck, 1986)

ᵃ *pho64* is is probably a *phoR* mutation (A. Garen, unpublished).

Incubations were at 37°C unless otherwise stated. Normally, liquid cultures were grown in 20 ml McCartney bottles or 100 ml conical flasks. Culture densities were measured at 650 nm (A_{650}) using an LKB Ultrospec Plus spectrophotometer. Where neccessary, solid media were supplemented with isopropyl-β-D-thiogalacto-pyranoside (IPTG) (24 μg/ml) and 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-gal) (20 μg/ml), purchased from Boehringer Mannheim.

2.2 Chemicals and reagents

Chemicals were analytical grade. Phenol, polyethylene glycol 6000 (PEG) and sodium dodecyl sulphate (SDS) were from BDH Chemicals Ltd. N-[2-acetamido]-2-iminodiacetic acid (ADA), diethyl pyrocarbonate (DEPC), ficoll 400 (approximate molecular weight 400,000) 3-[N-morpholino]propanesulphonic acid (MOPS), piperazine-N,N'-bis[2-ethanesulphonic acid] (PIPES), polyvinyl pyrrolidone (average molecular weight 360,000), tris[hydroxymethyl]amino-methane (Tris), 2,2' dipyridyl, haemin, haematoporphyrin and protoporphyrin IX were purchased from Sigma Chemical Company. Ethylene-diamine-tetraacetic acid (EDTA) and α-[4-1,1,3,3,-tetramethylbutyl)phenyl]-ω-hydroxypoly(oxy-1,2-ethanediyl) (Triton X-100) was from Ajax Chemicals. Mineral oil was Primol 352 (Esso). Adenosine-5'-triphosphate (ATP), 2'-deoxynucleotides (dNTPs), 1,4-dithiothreitol (DTT), IPTG, X-gal and 5-bromo-4-chloro-3-indolyl phosphate (X-pho) were from Boehringer Mannheim. Other chemicals were from Ajax, BDH, or Sigma. Milli Q water was water purified using a Milli Q water purification system (Millipore Corp.) with a measured resistance to conductivity of 18 MΩ/cm.

Antibiotics were purchased from Sigma (Ap, Gm, Km, Nal, Rif, Sm, Sp) or Calbiochem (Cm, Tc). Globomycin was the generous gift of Dr. Masatoshi Inukai, Fermentation Research Laboratories, Sankyo Co. Ltd., Tokyo, Japan. All other antimicrobial agents (dyes, detergents and antibiotics) were purchased from Sigma Chemical Co., BDH Chemicals Ltd., Glaxo, or Calbiochem.

Electrophoresis grade reagents were: acrylamide and N,N' methylene bis acrylamide (Boehringer Mannheim); ammonium persulphate and pulsed field certified agarose (Biorad);

ultrapure™ low gelling temperature (LGT) agarose and ultrapure™ urea (Bethesda Research Laboratories); high gelling temperature (HGT) agarose (Seakem); and N,N,N',N'-tetramethyl-ethylenediamine (TEMED) (Sigma).

Radiochemicals used were: γ-[^{32}P]-ATP (4,000 Ci/mmole), α-[^{32}P]-dCTP (3,000 Ci/mmole) and α-[^{32}P]-UTP (3,000 Ci/mmole), purchased from Bresatec; α-[^{35}S]-dATP (1,000 Ci/mmole), L-[^{35}S]-methionine (1,000 Ci/mmole) and D-[1-^{14}C]-galactose (62.5 mCi/mmole), purchased from Amersham; and [9, 10-^{3}H]- palmitic acid (60 Ci/mmole), purchased from New England Nuclear.

2.3 Enzymes and Nucleic Acids

Ribonuclease A (RNase) and lysozyme were obtained from Sigma. Glycogen, proteinase K, pronase and phenyl-methyl-sulphonyl fluoride (PMSF) were from Boehringer Mannheim.

Restriction endonucleases *AccI, AhaI, ApaI, BamHI, BgIII, BstBI, ClaI, DraI, EcoRI, EcoRV, FokI, HincII, HindIII, HpaII, MfluI, NdeI, NotI, NruI, NsiI, PstI, PvuII, SaII, SfiI, SmaI, SnaBI, SpeI, SphI, XbaI* and *XhoI* were purchased from Amersham, Bethesda Research Laboratories, Boehringer Mannheim, or New England Biolabs.

DNA modifying enzymes were purchased from: New England Biolabs (T4 DNA ligase); Boehringer Mannheim (calf intestinal phosphatase (c.i.p.), Klenow frament of DNA polymerase I, polynucleotide kinase); IBI (*Bal*31 exonuclease, slow form); Promega (Exonuclease III, S1 nuclease); United States Biochemical Corporation (Sequenase™ and Sequenase II™ modified T7 DNA polymerases).

RNase free DNaseI, SP6 and T7 RNA polymerases were purchased from Bresatec or Promega. Moloney murine leukaemia virus (M-MLV) reverse transcriptase was from Bethesda Research Laboratories. RNaseA/T1 mix and yeast RNA were from Ambion Inc.

Bacterial plasmids used are listed in Table 2.3. Oligodeoxynucleotides (oligos) are described in Table 2.4 and with the exception of M13 -20 primer (purchased from United States Biochemical Corporation) were synthesised on an Applied Biosystems model 381A DNA synthesiser, using chemicals purchased from Applied Biosystems or Ajax Chemicals (acetonitrile). Oligos were purified after synthesis using oligo purification cartridges purchased from Applied Biosystems, according to manufacturers specifications. Phosphorylated *Sal*I (5' GGTCGACC 3') and *Xba*I (5' CTCTAGAG 3') linkers were purchased from New England Biolabs.

2.4 Maintenance of bacterial strains

For long term storage, all strains were maintained as lyophilised 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 grown with shaking overnight at the appropriate temperature. The other half was streaked onto two nutrient agar plates and incubated overnight at the appropriate 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 peptone (1% w/v) and glycerol (25% v/v) for *E. coli* strains, or in peptone (1% w/v) and glycerol (30% v/v) for *Vibrio* and *Aeromonas* strains, all at -80°C. Fresh cultures from glycerol stocks were prepared by streaking a loop-full of the glycerol suspension onto a NA plate (with or without antibiotics 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 inch freeze drying ampoules and the end of each ampoule was plugged with cotton wool. The samples were then lyophilised in a freeze dryer. 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

Table 2.3 Bacterial plasmids and cloning vectors

plasmid	markers	reference
pBR322	Ap^R, Tc^R	Bolivar *et al.* (1977)
pACYC177	Ap^R, Km^R	Chang and Cohen (1978)
pACYC184	Cm^R, Tc^R	Chang and Cohen (1978)
pUC19	Ap^R	Yanisch-Perron *et al.* (1985)
pK18	Km^R	Pridmore (1987)
pGEM3Zf$^+$	Ap^R	Promega
pGEM5Zf$^+$	Ap^R	Promega
pGEM7Zf$^+$	Ap^R	Promega
pCH39	Ap^R	Hoffman and Wright (1985)
pPM3024	Ap^R	Williams and Manning (1991)
pKL600	Ap^R	K. McKenney, unpublished
pGP1-2	Km^R	Tabor and Richardson (1985)
pCACTUS	Cm^R, *rep*^{ts}	C.A. Clark, unpublished
pPHIJI	Gm^R, Sm^R, Sp^R, IncP	Beringer *et al.* (1978)
pME305	Ap^R, Tc^R, *trfA*^{ts}, *tra*$^+$, IncP	Rella *et al.* (1985)
pRM118	Tc^R	R. Morona, unpublished
M13mp18	–	Yanisch-Perron *et al.* (1985)
M13mp19	–	Yanisch-Perron *et al.* (1985)

Table 2.4 Oligodeoxynucleotides

name/number[a]	sequence	reference
M13 -20 primer	5' GTAAAACGACGGCCAGT 3'	Yanisch-Perron *et al.* (1985)
#61	5' AACCGCATCCATTTGATA 3'	this study
#128	5' CGCGGCTCAATGGGACTGGAAA 3'	this study
#226	5' CTGAACAACTATTTCGAG 3'	this study
#227	5' CCGAATTTAAAGCAAACC 3'	this study
#274	5' TCCAGTCCCATTGAGCCG 3'	this study
#275	5' CCTAAGGGGCTGGCAACG 3'	this study
#276	5' CACAATTCCACACAAC 3'	Yanisch-Perron *et al.* (1985)
#298	5' CCAACGATGCATTGATAC 3'	this study
#309	5' GAACTGCTGTTGGAGTGG 3'	this study
#316	5' CCAAACAGCTCTTCTACC 3'	this study
#317	5' CTAGTAATCCTAAATGTA 3'	this study
#318	5' TCTAGAAAATACATTGGT 3'	this study
#342	5' CCATTTTATAAGCAGGTC 3'	this study
#343	5' GTGAATATGGTGTTATGC 3'	this study
#344	5' GGCTATAGAGATTTATTT 3'	this study
#345	5' CTGCCAGAGGACTAATCC 3'	this study
#361	5' AAATATTATCTAGAATATGAAA 3'	this study
#362	5' CGGAGCGCAGCAGAGG 3'	de Boer (1984)
#372	5' GCGGTCAATGAGATACTT 3'	this study
#373	5' GCTACTTCCTTAGCCTAA 3'	this study
#374	5' ATTTTGCTAGGTCATGTG 3'	this study
#375	5' GTATCTGAGAGTGATACC 3'	this study
#430	5' ACTGTACTTTCTCGTTAC 3'	this study
#431	5' CTTTCATTGGGTTTGCTT 3'	this study
#565	5' ACGCAGAGCGGCAGTCTG 3'	Hoffman and Wright (1985)

[a] Primer numbers are those designated to a primer in the laboratory catalogue.

constriction without releasing the vacuum. Finally the ampoules were labelled and stored at 4°C.

2.5 Transformation

2.5.1. Transformation of *V. cholerae*

V. cholerae O1 were transformed by electroporation, essentially as described by Stoebner and Payne (1988). Strains were grown overnight on NA and a 20 ml LB culture inoculated from a single colony and grown to an $A_{650} = 0.5$. Cells were pelleted in a bench centrifuge, washed with 20 ml 272 mM sucrose, 7 mM sodium phosphate pH 7.4, 1 mM $MgCl_2$, resuspended in 1 ml of the same buffer and stored on ice. 0.2-1 µg supercoiled plasmid DNA (in a volume of 10 µl) was added to 0.1 ml of cells and incubated on ice for 20 min. The suspension was transferred to a Gene Pulser cuvette (electrode gap of 0.2 cm, purchased from Biorad) and electroporated at 2,000 V, 25 µF capacitance and 200 Ω resistance, producing time constants in the range 2.5-3.5 ms, using a Biorad Gene Pulser. The suspension was immediately diluted in 0.2 ml LB, in order to maximise cell survival (Marcus *et al.*, 1990) and incubated at 37°C for 45 min. The mixture was then plated onto selection plates. Transformation efficiencies were not determined for individual batches of cells.

2.5.2 Transformation of *E. coli*

2.5.2.1 Preparation of competent cells

Competent cells for use in subcloning experiments were made according to a modification of the method of Hanahan (1983) (C.P. Gibbs, personal communication). *E. coli* K-12 strains were grown overnight on fresh ψA at the appropriate temperature. A 5 ml ψB culture was inoculated with a single colony and grown to an $A_{650} = 0.6$. A prewarmed 100 ml ψB was then inoculated with this culture and incubated for a further 2 hours. Cells were chilled on ice for 10 min, pelleted at 4°C in a bench centrifuge, resuspended in 40 ml ice cold 30 mM potassium acetate pH 5.8, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (v/v) glycerol, and incubated on ice for 5 min. Cells were then pelleted in a bench centrifuge at 4°C and resuspended in 4 ml 10 mM MOPS pH 6.5, 10 mM RbCl, 75 mM $CaCl_2$, 15% (v/v) glycerol, incubated on ice for a further 15 min and then snap frozen in 300 µl aliquots and stored at -80°C.

Competent cells for use in transferring plasmids between *E. coli* K-12 strains were made essentially as described by Brown *et al.* (1979). A 100 ml LB culture was inoculated with bacteria isolated from a fresh plate and grown to an $A_{650} = 0.6$ at the appropriate temperature. The cells were chilled on ice for 20 min, pelleted at 4°C in a bench centrifuge, resuspended in 50 ml of ice cold 100 mM $MgCl_2$, centrifuged again and resuspended in 4 ml of ice cold 100 mM $CaCl_2$. This was allowed to stand on ice for 60 min, before the addition of 1 ml ice cold 100 mM $CaCl_2$, 60% (v/v) glycerol. Cells were then snap frozen in 300 μl aliquots and stored at -80°C.

2.5.2.2 Transformation procedure

Competent cells prepared by either procedure were thawed on ice and then mixed (0.1 ml of cells) with DNA (made to a volume of 10 μl with 10 mM Tris-HCl pH 8.0, 1 mM EDTA) and incubated on ice for 30 min. The mixture was then heated at 42°C for 1 1/2 min, returned to ice for 1 - 2 min and then 2 volumes of LB was added, followed by incubation at the appropriate temperature for 45 min. The mixture was then plated directly onto selection plates. Cells prepared for subcloning had transformation efficiencies ranging from 10^7-5 x 10^8 *per* μg supercoiled pUC19 DNA, while cells prepared for transferring plasmids between strains typically had transformation efficiencies within the range 10^5-10^6 *per* μg supercoiled pUC19 DNA (not shown). Frozen cells were found to retain their transformation efficiency for at least 12 months (not shown).

2.6 Bacterial conjugation

Overnight broth cultures grown in NB or LB were diluted 1 in 20 and grown to early exponential phase with slow agitation. Donor and recipient bacteria were mixed at a ratio of 1:10 and the cells pelleted by centrifugation for 10 min in a bench centrifuge. The pellet was gently resuspended in 0.1 ml of LB and spread onto a nitrocellulose membrane filter (0.45mm, type HA, Millipore Corp.) on a NA plate. This plate was incubated for 4-6 hours at the appropriate temperature. The cells were then resuspended in 10 ml NB and samples (neat and serial dilutions) plated onto selective agar and incubated at the appropriate temperature.

2.7 DNA extraction procedures

2.7.1 Small scale plasmid isolation

Small scale plasmid isolation procedures were based on the alkaline lysis method of Birnboim and Doly (1979). Overnight bacterial cultures (1.5 ml) were transferred to a microcentrifuge tube and harvested by centrifugation for 1 min in a microfuge, and resuspended in 0.1 ml 50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA. After 5 min incubation at room temperature, 0.2 ml of 1% (w/v) SDS, 0.2 M NaOH was added and the mixture incubated on ice for 5 min. After the addition of 0.15 ml 3M potassium acetate pH 4.8 and a 15 min incubation on ice, cellular debris was removed by centrifugation (5 min, microfuge). Extracts from *endA*⁺ strains were then extracted once with an equal volume of Tris-saturated phenol (pH 7.5) and once with a chloroform, isoamyl alcohol (24:1) mixture. Extracts from *endA* strains were mixed with a half volume of 7.5 M ammonium acetate, incubated on ice for a further 10 min, centrifuged (3 min, microfuge) and the supernatant collected. In all cases plasmid DNA was then precipitated by the addition of an equal volume of 2-propanol and incubation on ice for 10 min. DNA was collected by centrifugation (15 min, microfuge), washed with 70% (v/v) ethanol, and dried *in vacuo*. The pellet was resuspended in 0.1 ml 10 mM Tris-HCl pH8.0, 1 mM EDTA and stored at 4°C.

2.7.2 Large scale plasmid isolation

Large scale plasmid isolation was performed by the three step alkaline lysis method (Garger *et al.*, 1983). Cells from a litre culture (LB) were harvested (5,000 rpm, 10 min, 4°C, GS-3, Sorvall) and resuspended in 24 ml 50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA. Freshly prepared lysozyme (4 ml of 20 mg/ml in the above solution) was mixed with the cell suspension and incubated at room temperature for 10 min. Next, 56 ml of 1% (w/v) SDS, 0.2 M NaOH was added, followed by 5 min incubation on an ice/water slurry. After the addition of 28 ml 3M potassium acetate pH 4.8 and incubation on an ice/water slurry for a further 15 min, cellular debris was removed by centrifugation (8,000 rpm, 20 min, 4°C, GS-3, Sorvall). The supernatant was then extracted with an equal volume of a Tris-saturated phenol (pH7.5), chloroform, isoamyl alcohol (25:24:1) mixture. Plasmid DNA was precipitated from the aqueous phase by the addition of 0.6 volumes of 2-propanol and incubation at room temperature for 10 min and collected by centrifugation (11,000 rpm, 20 min, 4°C, GS-A,

Sorvall). After washing in 70% (v/v) ethanol, the pellet was dried *in vacuo* and resuspended in 4.8 ml 10 mM Tris-HCl pH 8.0, 1 mM EDTA. CsCl was then added to a density of 1.8 g/ml, followed by 0.8 ml of 10 mg/ml ethidium bromide and the mixture underlaid under 8 ml CsCl in 10 mM Tris-HCl pH 8.0, 1 mM EDTA (density of 1.47 g/ml) in two 5/8 x 3 inch Beckman quickseal polyallomer tubes. The tubes were then filled with the 1.47 g/ml CsCl solution, sealed, and centrifuged (65,000 rpm, 5 hours, 20°C, Ti-80, Beckman L8-80). After centrifugation, supercoiled DNA was removed by side puncture of the tubes with a 19 gauge needle attached to a 2 ml syringe, and ethidium bromide removed by several extractions using isoamyl alcohol. CsCl was then removed by dialysis against three changes of 2 litres 10 mM Tris-HCl pH 8.0, 1 mM EDTA at 4°C. DNA concentration was determined using a LKB Ultrospec Plus spectrophotometer, assuming that 1 A_{260} unit $\equiv$ 50 µg/ml DNA (Maniatis *et al.*, 1982). DNA was then precipitated by the addition of 0.1 volume 3M sodium acetate pH 5.2 and an equal volume of 2-propanol, followed by incubation on ice for 10 min. DNA was collected by centrifugation (15 min, microfuge), washed with 70% (v/v) ethanol, dried *in vacuo*, resuspended at a concentration of 1 µg/µl in 10 mM Tris-HCl pH 8.0, 1 mM EDTA and stored at 4°C.

2.7.3. Preparation of bacterial genomic DNA

Whole genomic DNA was prepared according to Manning *et al.* (1986). Cells from a 20 ml shaken overnight culture were pelleted in a bench centrifuge for 10 min and resuspended in 2 ml 25% sucrose, 50 mM Tris-HCl pH 8.0. 1 ml 10 mg/ml lysozyme in 0.25 M EDTA pH 8.0 was added and the mixture incubated on ice for 20 min. 0.75 ml 10 mM Tris-HCl pH 8.0, 1 mM EDTA and 0.25 ml 5% (w/v) sodium n-lauroyl sarcosine, 50 mM Tris-HCl pH 8.0, 62.5 mM EDTA were added, along with 10 mg solid pronase and the mixture incubated at 65°C for a minimum of 1 hour. This was followed by three extractions with Tris-saturated phenol (pH 7.5) and two with diethyl ether. The genomic DNA was then transferred to dialysis tubing and dialysed against three changes of 2 litres 10 mM Tris-HCl pH 8.0, 1 mM EDTA at 4°C. DNA concentration was determined using a LKB Ultrospec Plus spectrophotometer, assuming that 1 A_{260} U $\equiv$ 50 µg/ml DNA (Maniatis *et al.*, 1982), and DNA was stored at 4°C.

2.7.4 Preparation of genomic DNA for pulsed field gel electrophoresis

Samples for pulsed field gels were prepared by a modification of the procedures of Smith and Cantor (1987) and of Overhauser and Radic (1987) as follows. A 20 ml culture of *V. cholerae* in minimal medium was grown at 30°C to an A_{650} of 0.6 and incubated with chloramphenicol (final concentration of 180 μg/ml) for 60 min. Cells were then pelleted by centrifugation in a bench centrifuge for 10 min, washed in 10 ml 1 M NaCl, 10 mM Tris-HCl pH 8.0, resuspended in 4 ml of the same buffer and warmed to 45°C. Agarose beads were formed by adding 5 ml of 1% LGT agarose in 1 M NaCl, 10 mM Tris-HCl pH 8.0 and 20 ml mineral oil, both at 45°C, to the cells, emulsifying the mixture and pouring into 100 ml ice cold 1 M NaCl, 10 mM Tris-HCl pH 8.0, and stirring the mixture for 5 min. Beads were then pelleted by centrifugation at 4°C and 500 x *g* in 50 ml tubes for 10 min and the mineral oil removed. Tubes were recentrifuged, beads in the pellet combined in a single tube and then centrifuged as before. Beads were then resuspended in 15 ml 1 mg/ml proteinase K, 1% sodium n-lauroyl sarcosine, 0.5 M EDTA pH 9.2, pre-warmed to 65°C and incubated at 65°C for at least 48 hours. A lysozyme treatment step was omitted because of excessive nicking of the DNA during this step (Chapter 4), presumably by secreted DNases (Focareta and Manning, 1991). After proteinase K digestion, the beads were recovered by centrifugation as before and washed in 0.1 mM PMSF in 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA for 2 hours and for a minimum of 4 x 2 hours in 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA at room temperature. Beads were then stored at 4°C.

2.8 Analysis and manipulation of DNA

2.8.1 Restriction endonuclease digestion of DNA in solution

Cleavage reactions with restriction endonucleases were performed on 0.1-10 μg of DNA in a final volume of 20 μl using reaction buffers recommended by the manufacturer. DNA not purified by CsCl centrifugation was digested in the presence of 0.5 μg RNaseA. At least 2 units of restriction enzyme *per* μg DNA was added, with a minimum of 2 units of enzyme being used. Reactions were incubated at 37°C for 1-2 hours, with the exception of digests using *Sma*I, which were incubated at 25°C, and *Bst*BI, which were incubated at 65°C under mineral oil. Reactions for gel electrophoresis were terminated by heating at 65°C for 10 min (except for *Bst*BI reactions, which were heated to 85°C) and prior to loading, were

mixed with a one tenth volume of 25% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol (Maniatis *et al.*, 1982).

Reactions for other procedures were made up to 0.1 ml with Milli Q water, extracted with Tris saturated phenol (pH 7.5), chloroform, isoamyl alcohol (25:24:1) and precipitated by the addition of one tenth volume 3 M sodium acetate pH 5.2 and three volumes 100% ethanol, followed by incubation at -20°C for 30 min. 20 µg glycogen was added as a carrier. DNA was recovered by centrifugation (15 min, microfuge), pellets were washed with 70% (v/v) ethanol and dried *in vacuo*. Pellets were resuspended in an appropriate volume of Milli Q water and stored at -20°C.

Partial digestion with restriction enzymes was performed by serially diluting (2 fold) enzyme in a solution containing 0.1 - 0.5 µg DNA in 1 x restriction buffer, from a starting concentration of 0.5 U *per* µg DNA present, and incubating at 37°C for 60 min. The reaction was then processed as described above and partially digested products recovered by gel purification. Partial digests with *Nde*I were performed with 2U *per* µg (CsCl purified) DNA present in the presence of a one tenth volume of an extract from *E. coli* DH5 prepared by performing a "mock miniprep" (Section 2.7.1, using ammonium acetate precipitation), as this enzyme has been observed to be partially inhibited by miniprep DNA (New England Biolabs catalogue). Work up of these reactions was as above.

2.8.2 Restriction endonuclease digestion of DNA in agarose beads

Agarose embedded DNA was digested essentially as described (Smith and Cantor, 1987; Overhauser and Radic, 1987). Agarose beads were collected by centrifugation at 500 x g, 0.2 ml of beads were resuspended in 0.5 ml of the appropriate digestion buffer (without added BSA) and were washed three times, with 5 min incubation at room temperature between centrifugations. After washing, the beads were resuspended in 0.25 ml of the appropriate digestion buffer containing 1 mg/ml BSA and 25 units of restriction endonuclease, and incubated at 37°C for 16 hours. Beads were then pelleted (2 min, microfuge), resuspended in 0.5 ml 1% sodium n-lauroyl sarcosine, 0.5 M EDTA pH 9.2 and incubated at 55°C for 2 hours. Beads were pelleted and resuspended in the same buffer

containing 1 mg/ml proteinase K, and incubated for a further 2 hours at 55°C. The beads were then loaded on pulsed field gels.

2.8.3 Analytical and preparative separation of restriction fragments

Electrophoresis of digested DNA was carried out at room temperature on horizontal 0.8% or 1% (w/v) agarose gels (Seakem HGT), 14 cm long, 11 cm wide and 0.7 cm thick, in Bethesda Research Laboratories model H5 gel tanks. Gels were run at 100 V for 4-5 hours in 67 mM Tris base, 22 mM boric acid, 1 mM EDTA, final pH 8.8 (TBE). After electrophoresis the gels were stained in distilled water containing 2 μg/ml ethidium bromide for 10-15 min. DNA bands were visualised by trans-illumination with UV light and photographed using either Polaroid 667 positive film or 665 negative film.

Preparative separation of restriction fragments was by one of two methods. Fragments of size ≥ 1 kb were separated on HGT agarose gels as above. After staining, bands were located by brief transillumination with long wave UV light and were excised. The agarose slice was then placed in a length of dialysis tubing containing 0.5 ml TBE and electrophoresed for 20-30 min. The direction of current was reversed for 20 seconds, and the fluid removed from the bag. Fragments of size ≤ 1 kb were separated on 6% polyacrylamide gels (acrylamide: bis acrylamide 30:1) 22 cm long, 16 cm wide and 0.15 cm thick, in TBE. Gels were pre-run at 400 V for 30 min prior to loading and were run at 500 V for 2 hours. After electrophoresis the gels were stained in distilled water containing 2 μg/ml ethidium bromide for 5 min, destained for 10 min in distilled water and the bands were located by brief transillumination with long wave UV light and were excised. The polyacrylamide slice was then placed in a microfuge tube containing 0.5 ml 10 mM Tris-HCl pH 8.0, 1 mM EDTA and incubated at 37°C for 16 hours in the dark, before the fluid was recovered. DNA obtained by either procedure was extracted once with Tris saturated phenol (pH 7.5) and once with chloroform, isoamyl alcohol (24:1). DNA was precipitated with 0.1 volume 3 M sodium acetate pH 5.2, 1 volume of 2-propanol and incubation on ice for 10 min. 20 μg glycogen was added as a carrier. DNA was recovered by centrifugation (15 min, microfuge), pellets were washed with 70% (v/v) ethanol and dried *in vacuo*. DNA was resuspended in an appropriate volume of Milli Q water and stored at -20°C.

2.8.4 Calculation of restriction fragment size

The sizes of restriction endonuclease generated fragments were calculated by comparing their relative mobility with that of *Eco*RI digested *Bacillus subtilis* bacteriophage SPP1 DNA (Ratcliff *et al.*, 1979). The calculated sizes of the SPP1 *Eco*RI fragments used were obtained from Bresatec and were (in kb): 8.51; 7.35; 6.11; 4.84; 3.59; 2.81; 1.95; 1.86; 1.51; 1.39; 1.16; 0.98; 0.72; 0.48; 0.36. For polyacrylamide gels, additional size markers were *Hpa*II digested pUC19 DNA (Yanisch-Perron *et al.*, 1985) and were (in base pairs): 489; 404; 331; 242; 190; 147; 111; 110; 67; 34; 26. Size markers for pulsed field gels were yeast chromosomes or concatamers of λ DNA, purchased from Biorad.

2.8.5 Pulsed field gel electrophoresis

Samples for pulsed field gel electrophoresis were separated on a 1% agarose (pulsed field certified agarose) gel of size 12.7 cm long, 14 cm wide and 0.7 cm thick, using a Biorad CHEF II apparatus. Standard run conditions were: initial A time = 1 second; final A time = 10 seconds; and 200 V for 18 hours at 14°C in 0.5 x TBE.

2.8.6 Dephosphorylation of DNA

Plasmid DNA was dephosphorylated using c.i.p. according to Maniatis *et al.* (1982). ZnCl$_2$ was added to 0.1-2 µg of digested DNA in restriction buffer to a final concentration of 0.1 mM. For restriction digests that generated 5' overhanging ends, 1 unit of c.i.p. was added and the mixture incubated at 37°C for 30 min. For restriction digests that generated 3' overhanging, or blunt ends, 1 unit of c.i.p. was added and the mixture incubated at 37°C for 15 min. Then, a further unit of c.i.p. was added and the mixture incubated at 55°C for a further 45 min. In all cases, 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 volume was made up to 0.1 ml with Milli Q water and the mix was extracted with Tris saturated phenol (pH 7.5), chloroform, isoamyl alcohol (25:24:1). DNA was precipitated with 0.1 volume 3 M sodium acetate pH 5.2, 3 volumes of 100% ethanol and incubation at -20°C for 30 min. 20 µg glycogen was added as a carrier. DNA was recovered by centrifugation (15 min, microfuge), pellets were washed with 70% (v/v) ethanol and dried *in vacuo*. DNA was resuspended in an appropriate volume of Milli Q water and stored at -20°C.

2.8.7 End repair of linear DNA

Protruding ends generated by restriction endonucleases or exonuclease *Bal*31 were repaired using the Klenow fragment of *E. coli* DNA polymerase I (Klenow). DNA in restriction buffer, or resuspended in 20 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$ was mixed with dNTPs to a final concentration of 40 μM and 2 units of Klenow, and incubated at 37°C for 30 min. The volume was made up to 0.1 ml with Milli Q water and the mix was extracted with Tris saturated phenol (pH 7.5), chloroform, isoamyl alcohol (25:24:1). DNA was precipitated with 0.1 volume 3 M sodium acetate pH 5.2, 3 volumes of 100% ethanol and incubation at -20°C for 30 min. 20 μg glycogen was added as a carrier. DNA was recovered by centrifugation (15 min, microfuge), pellets were washed with 70% (v/v) ethanol and dried *in vacuo*. DNA was resuspended in an appropriate volume of Milli Q water and stored at -20°C.

2.8.8 Ligation of DNA fragments

Ligations were performed using approximately 100 ng vector DNA and insert DNA at an approximately 3 fold (Molar) excess in a final volume of 20 μl, in 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5% (w/v) PEG, 40 mM ATP, 0.5 mM DTT, using 1 unit of T4 DNA ligase and incubated overnight at 16°C for "sticky end" ligations or 30°C for "blunt end" ligations. Phosphorylated linkers were ligated to blunt ends generated by Klenow repair, by overnight incubation of 0.2-1 μg plasmid DNA with approximately 1 μg linkers and 2 units T4 DNA ligase in the same buffer and a final volume of 10 μl. The ligation mixture was then digested with the restriction enzyme specific for the linker, excess linkers removed by gel purification and the resulting fragment religated in 0.1 ml of the same buffer.

2.8.9 Exonuclease digestion of linear DNA

2.8.9.1 Generation of small deletions

Small deletions were generated from plasmids linearised with restriction endonucleases using *Bal*31 slow form (Wei *et al.*, 1983). DNA was resuspended to a concentration of 50 pmol DNA ends/ml in 50 μl of 20 mM Tris-HCl pH 8.0, 600 mM NaCl, 12.5 mM $MgCl_2$, 12.5 mM $CaCl_2$, 1 unit of *Bal*31 slow form was added and the mixture incubated at 30°C. Portions were removed at various time points and added to EDTA to a

final concentration of 50 mM and heated at 65°C for 10 min to stop the reaction. Samples were then end repaired with Klenow enzyme and ligated with linkers.

2.8.9.2 Generation of nested deletions

Nested deletions were generated using Exonuclease III (Henikoff, 1984) purchased as the Erase-a-base® system from Promega. Templates were linearised with restriction endonucleases, or alternatively, where unidirectional deletions were required, template was linearised with a restriction endonuclease, protected by end repair using α-phosphothioate dNTPs, and then restricted with a second restriction endonuclease to generate an unprotected end. For digestion, 5 µg DNA was dissolved in 60 µl 66 mM Tris-HCl pH 8.0, 0.66 mM MgCl$_2$, 400 units Exonuclease III was added and the mixture incubated at 30°C. 2.5 µl aliquots were removed at 30 second intervals and added to 7.5 µl 40.5 mM potassium acetate pH 4.6, 338 mM NaCl, 1.35 mM ZnSO$_4$, 6.75% (v/v) glycerol containing 2.25 units of S1 nuclease. The mixture was incubated at room temperature for 30 min before the addition of 1 µl 300 mM Tris base, 50 mM EDTA and incubation at 65°C for 10 min. Samples were then religated.

2.8.10 Labelling of single stranded DNA

Single stranded DNA (oligos) were kinased using γ-[^{32}P]-ATP. 60 ng of primer was incubated with 5 units T4 polynucleotide kinase in the presence of 70 mM Tris-HCl pH 7.6, 10 mM MgCl$_2$, 5 mM DTT and 25-50 µCi γ-[^{32}P]-ATP. The reaction mix was made up to 15 µl with Milli Q water and incubated at 37°C for 30 min. Labelled oligos were separated from unincorporated label by gel electrophoresis as described for small restriction fragments in Section 2.8.3, except that bands for excision were located by brief autoradiography instead of ethidium bromide staining. Percentage incorporation was determined by comparing the recovered material to the starting material by counting in a Beckman model LS6000TA β counter.

2.8.11 Labelling of restriction fragments

Purified restriction fragments were labelled in a random priming reaction (Feinberg and Vogelstein, 1983) incorporating α-[^{32}P]-dCTP, or digoxigenin-11-dUTP (DIG-dUTP).

Reactions incorporating α-[^{32}P]-dCTP were performed by resuspending the fragment to be labelled (60-100 ng) in 6.5 μl Milli Q water, boiling for 5 min and then chilling on ice for 2 min. 12.5 μl of a solution consisting of 40 μM each of dATP, dGTP and dTTP, 100 mM Tris-HCl pH 7.6, 100 mM NaCl, 20 mM MgCl$_2$, 200 μg/ml gelatine and 600 μg/ml random primer (hexamer) was then added, along with 50 μCi α-[^{32}P]-dCTP (5 μl aqueous solution) and 2 units Klenow enzyme, and the mixture incubated at 37°C for 30 min. The reaction was stopped by the addition of EDTA pH 8.0 to a final concentration of 20 mM and unincorporated label was removed by centrifugation through a Sephadex G50 (medium) spun column (Maniatis *et al.*, 1982). Percentage incorporation was determined by comparing recovered material to starting material, by counting in a Beckman model LS6000TA β counter, and was typically about 50% (specific activity in the range 10^7-10^8 cpm/μg DNA: not shown).

Reactions incorporating DIG-dUTP were performed by resuspending the fragment to be labelled (30-50 ng) in 15 μl Milli Q water, boiling for 5 min and then chilling on ice for 2 min. 2μl of a solution containing 1 mM of each dNTP, 0.65 mM DIG-dUTP, 142.5 mM Tris-HCl pH 8.1, 5.7 mM DTT, 14.25 mM MgCl$_2$, 114 mM KCl, 2 μl of 400 μg/ml random primer (hexamer), and 2 units Klenow enzyme were added and the mixture incubated at 37°C overnight. The reaction was stopped by the addition of EDTA pH 8.0 to a concentration of 20 mM. Unincorporated label was removed by precipitating DNA by the addition of a one-tenth volume of 4 M LiCl and three volumes of 100% ethanol followed by incubation at -20°C for 30 min. Glycogen (20 μg) was added as a carrier. DNA was recovered by centrifugation (15 min, microfuge) the pellet washed with 70% (v/v) ethanol, and dried *in vacuo*. The pellet was resuspended in 0.1 ml Milli Q water and stored at -20°C. Percentage incorporation was not determined, but labelling was checked by blotting a tenth of the probe onto nitrocellulose using a dot blot apparatus (see Section 2.8.15) and detecting bound DIG-dUTP (see Section 2.8.13).

2.8.12 Labelling of M13 clones for use as probes

Single strand specific probes were derived from M13 clones by mixing 0.5 μg M13 DNA with 15 ng M13 hybridisation primer (#276; Table 2.4) in 20 mM Tris-HCl pH 8.0, 5

mM MgCl$_2$ (final volume 10 μl), heating at 65°C for 10 min and then cooling slowly to room temperature to anneal. 2μl of a solution containing 1 mM of each dNTP, 0.65 mM DIG-dUTP, 142.5 mM Tris-HCl pH 8.1, 5.7 mM DTT, 14.25 mM MgCl$_2$, 114 mM KCl, 2 units Klenow enzyme were added, the volume adjusted to 20 μl with the addition of Milli Q water and the mixture incubated at 37°C for 30 min. The reaction was stopped with the addition of EDTA pH 8.0 to a final concentration of 20 mM and unincorporated label removed as described for DIG-labelled random primed probes in Section 2.8.11.

2.8.13 DNA gel blot transfer and hybridisation (Southern blot)

Unidirectional transfers of DNA from agarose gels to nitrocellulose membranes (NA45, Schleicher and Schüll) were performed as described by Southern (1975) and modified by Maniatis *et al.* (1982). Prior to hybridisation with isolated DNA fragment or single stranded M13 clone probes, filters were incubated for 4 hours at 42°C in a prehybridisation solution containing 50% (v/v) deionised formamide, 50 mM sodium phosphate pH 6.4, 0.75 M NaCl, 75 mM sodium citrate, 0.1% (w/v) ficoll, 0.1% (w/v) polyvinyl pyrrolidone, 0.1% (w/v) BSA and 100 μg/ml single stranded herring sperm DNA. Prehybridisation solution for oligonucleotide probes was the same except that formamide was omitted and incubation was at 37°C. For hybridisation, probes were heat denatured for 10 min (100°C for fragment probes, 75°C for oligonucleotide probes, no denaturation for M13 probes), added to the filter in prehybridisation solution and incubated for 16 hours at 42°C (37°C for oligonucleotide probes). Filters were washed four times with shaking at room temperature for 5 min in 0.3 M NaCl, 30 mM sodium citrate pH 7.0, 0.1% (w/v) SDS, followed by two washes at 65°C in 30 mM NaCl, 3 mM sodium citrate pH 7.0, 0.1% (w/v) SDS. For oligonucleotide probes, filters were washed four times for 5 min in 0.9 M NaCl, 90 mM sodium citrate pH 7.0, 0.1% (w/v) SDS, followed by rinsing at 37°C in 3 M tetramethyl-ammonium chloride, 50 mM Tris-HCl pH 8.0, 2 mM EDTA, 1 mg/ml SDS and two washes of 20 min at a temperature five degrees below the length dependent temperature of dissociation for the oligonucleotide in the same solution (Wood *et al.*, 1985).

After stringent washes, filters probed with radioactive probes were sealed in plastic bags and placed on X-ray film (Cronex 10T; DuPont) for autoradiography at -80°C with

intensifying screens. Relative intensities of bands obtained from radioactively-probed blots were determined by scanning the autoradiograph with a Molecular Dynamics model 300A computing densitometer. Filters probed with DIG-labelled probes were treated as follows. Filters were washed briefly in 100 mM Tris-HCl pH 7.5, 150 mM NaCl, and then for at least 30 min in 5% (w/v) skim milk in the same buffer. After a second brief wash in the same buffer, the filter was washed in the same buffer containing anti-DIG antibody-horse radish peroxidase conjugate at a dilution of 1/5,000 for a minimum of 30 min. The filter was then washed four times for 5 min in 100 mM Tris-HCl pH 7.5, 150 mM NaCl, followed by two 5 min washes in 10 mM potassium phosphate pH 7.4, 0.82% (w/v) NaCl (PBS), overlaid on 1 ml of ECL detection reagent (Amersham) for 1 min, and then the excess liquid removed with blotting paper and the filter exposed to X-ray film at room temperature. All washes were at room temperature.

2.8.14 Colony hybridisation

Colonies containing DNA hybridising with radioactively labelled DNA fragments or oligonucleotides were detected by the procedure of Grunstein and Hogness (1975). Patched colonies were grown for 5 hours at 37°C. Nitrocellulose disks (8.2 cm diameter, Schleicher and Schüll) were then placed on top of the colonies which were allowed to absorb for 5 min. The filters were then placed on blotting paper saturated with the following solutions for 5 min: i) 10% (w/v) SDS; ii) 1.5 M NaCl, 0.5 M NaOH; iii) 1 M Tris-HCl pH 7.0, 1.5 M NaCl (twice). Filters were then placed colony side up on dry blotting paper to air dry, were baked at 80°C for 1 - 2 hours *in vacuo* and were hybridised with probe as described in Section 2.8.13.

2.8.15 DNA dot blots

DNA dot blots were performed by diluting in 100 µl 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1 M NaCl, adding 6 µl 5 M NaOH and incubating at 65°C for 60 min. The sample was neutralised by the addition of 0.1 ml 3 M NaCl, 0.3 M sodium citrate pH 7.0 and 50 µl 1 M Tris-HCl pH 6.8 and was applied to a nitrocellulose filter using a Biorad dot blot apparatus. Wells were washed once with 0.9 M NaCl, 90 mM sodium citrate pH 7.0 and the

filter dried, baked at 80°C for 1-2 hours *in vacuo*, and prehybridisation and hybridisation steps performed as described in Section 2.8.13.

DIG-labelled probes were evaluated for labelling by dot blotting, except that after baking, filters were treated with anti-DIG antibody-horse radish peroxidase conjugate as described in Section 2.8.13. Probes were deemed sufficient if signal was detectable with 1 min exposure to X-ray film.

2.9 M13 cloning and sequencing procedures

2.9.1 Preparation of M13 replicative form (RF) DNA

Fresh 2 x TY broth (10 ml) was inoculated with 10 µl of an overnight culture of JM101 (in minimal medium containing 50 µg/ml thiamine). A single plaque of M13mp18 or M13mp19 picked from a minimal agar plate with a sterile toothpick was added to this bottle. The culture was grown at 37°C with vigorous shaking for 6 hours. Bacterial cells were removed by centrifugation (bench centrifuge) and the supernatant added to 1 l of 2 x TY broth containing 10 ml of a shaken overnight culture of JM101 (in minimal medium containing thiamine). Following incubation for 16 hours at 37°C with shaking, replicative form (RF) DNA was prepared as described previously for plasmid DNA purification (Section 2.7.2).

2.9.2 Cloning with M13mp18 and M13mp19

The M13 vectors, M13mp18 and M13mp19 (Yanisch-Perron *et al.*, 1985) were used for selective cloning of restriction enzyme generated DNA fragments. Stocks of M13 vectors cleaved with various enzyme combinations and treated with c.i.p. were stored at -20°C. Plasmid DNA was cut with the appropriate enzyme combinations for subcloning into the M13 vectors, with ligation conditions as described in Section 2.8.8.

2.9.3 Transfection of JM101

The *E. coli* strain JM101 was made competent for transformation/transfection by the subcloning efficiency method described in Section 2.5.2.1, except that bacteria were initially propagated on minimal agar containing thiamine instead of ψA. Competent cells (0.1 ml) were added directly to ligation mixes and incubated on ice for 30 min. This was followed by

a 1 1/2 min heat shock at 42°C. Cells were then transferred to sterile test tubes to which were added a mixture of JM101 indicator cells (0.1 ml), 24 mg/ml IPTG (30 µl), 20 mg/ml X-gal in N,N'-dimethyl formamide (30 µl) and 3 ml H-top agar. The mixture was poured as an overlay onto a prewarmed minimal agar plate (with thiamine) and incubated overnight at 37°C.

2.9.4 Screening M13 vectors for inserts

White plaques were picked from X-gal/IPTG containing plates with sterile toothpicks and added to 2 ml 2 x TY broth in 10 ml centrifuge tubes containing a 1:100 dilution of an overnight culture of JM101. These tubes were incubated for 5 hours at 37°C with vigorous shaking. The cells were pelleted by centrifugation (bench centrifuge) and RF DNA, suitable for restriction analysis, was prepared according to the small scale plasmid isolation procedure (Section 2.7.1). After restriction enzyme digestion, DNA was analysed by electrophoresis through 1% (w/v) agarose gels.

2.9.5 Purification of single stranded template DNA

M13 RF DNA containing appropriate inserts was reintroduced into JM101 and single white plaques from this transfection were picked with sterile toothpicks to inoculate 2 ml of 2 x TY broth containing 20 µl of an overnight culture of JM101. After vigorous shaking at 37°C for 6 hours, the culture was transferred to 2 ml Eppendorf tubes and centrifuged for 10 min (microfuge). 1.5 ml of the supernatant was transferred to a clean tube (1.5 ml tube) and recentrifuged for 5 min. A 1 ml aliquot of this supernatant was withdrawn and mixed with 0.27 ml of 20% (w/v) PEG, 2.5 M NaCl, and incubated on ice for 15 min. The phage were pelleted by centrifugation for 5 min (microfuge) and the supernatant discarded. Following another short spin (15 seconds) the remainder of the PEG/NaCl supernatant was removed with a drawn out pasteur pipette. Final traces of supernatant were removed with tissue paper. Pellets were resuspended in 0.3 ml 10 mM Tris-HCl pH 8.0, 1 mM EDTA. Redistilled Tris saturated phenol (pH 7.5) (0.3 ml) was then added to the phage suspension and the tubes were vortexed briefly. After a 5 min period with occasional vortexing, the tubes were centrifuged (3 min, microfuge) and the aqueous phase recovered and extracted with chloroform, isoamyl alcohol (24:1). DNA was precipitated from the aqueous phase by the addition of one tenth volume 5 M NaClO$_4$ and one volume 2-propanol. Tubes were centrifuged for 20 min

(microfuge). DNA pellets were washed once with 70% (v/v) ethanol and dried *in vacuo*. Pellets were resuspended in 50 µl 10 mM Tris-HCl pH 8.0, 1 mM EDTA and stored at -20°C.

2.9.6 Dideoxy sequencing of single stranded DNA with Sequenase™ or SequenaseII™

The dideoxy chain termination procedure of Sanger *et al.* (1977) was modified to encompass the use of Sequenase™ or SequenaseII™ (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 and dITP mixes. The contents of the dGTP mixes are as follows:

Labelling mix (dGTP):   7.5 mM dCTP, dGTP and dTTP

ddA Termination mix (dGTP):   80 mM dNTP, 8 mM ddATP, 50 mM NaCl

ddC Termination mix (dGTP):   80 mM dNTP, 8 mM ddCTP, 50 mM NaCl

ddG Termination mix (dGTP):   80 mM dNTP, 8 mM ddGTP, 50 mM NaCl

ddT Termination mix (dGTP):   80 mM dNTP, 8 mM ddTTP, 50 mM NaCl

The dITP mixes were used to reduce gel artefacts due to secondary structures in DNA synthesised in the sequencing reaction (Barnes *et al.*, 1983; Gough and Murray, 1983). Reactions incorporating dITP mixes were run in lanes adjacent to reactions incorporating dGTP mixes. The dITP mixes were as follows:

Labelling mix (dITP):   15 mM dITP, 7.5 mM dCTP, 7.05 mM dTTP

ddA Termination mix (dITP):   160 mM dITP, 80 mM dATP, dCTP, dTTP,
                                   8 mM ddATP, 50 mM NaCl

ddC Termination mix (dITP):   160 mM dITP, 80 mM dATP, dCTP, dTTP,
                                   8 mM ddCTP, 50 mM NaCl

ddG Termination mix (dITP):   160 mM dITP, 80 mM dATP, dCTP, dTTP,
                                   1.6 mM ddGTP, 50 mM NaCl

ddT Termination mix (dITP):   160 mM dITP, 80 mM dATP, dCTP, dTTP,
                                   8 mM ddTTP, 50 mM NaCl

Normally the labelling mix was diluted 1:5 with Milli Q water to obtain the working concentration. However, to read long sequences in a single reaction, a dilution of 1:2 was used. Primer was annealed to template by incubating 5-10 nM template and 500 nM primer in 40 mM Tris-HCl pH 7.5, 20 mM MgCl$_2$, 50 mM NaCl (in 10 μl) at 65°C for 5 min and cooling slowly to room temperature. To the annealed mixture, 2 μl of the appropriately diluted labelling mix, 1 μl 0.1 M DTT, 5 μCi (0.5 μl) α-[^{35}S]-dATP and 2 μl of diluted Sequenase™ or Sequenase II™ (1:8 dilution in 10 mM Tris-HCl pH 8.0, 1 mM EDTA) was added, spun, mixed and incubated at room temperature for 5 min. 3.5 μl of this mix was then aliquoted into four microfuge tubes, prewarmed to 37°C, each containing 2.5 μl of the appropriate termination mix, then spun briefly to start the termination reaction. After 5 min at 37°C 4 μl 95% (v/v) formamide, 20 mM EDTA pH 8.0, 0.05% (w/v) bromophenol blue, 0.05% xylene cyanol was added to each reaction. Reaction mixes were heated to 100°C for 3 min and 1.2 μl was loaded immediately onto sequencing gels. Samples were stored at -20°C for up to 2 weeks.

2.9.7 Sequencing of supercoiled plasmid template

Supercoiled plasmid DNA was sequenced directly after alkali denaturation (Chen and Seeburg, 1985). 2-4 μg plasmid DNA (if miniprep DNA, DNA was first treated with RNaseA) was diluted to a volume of 18 μl with milli Q water and denatured by the addition of 2 μl 2 M NaOH and incubation at room temperature for 5 min. DNA was precipitated by the addition of 8 μl 5 M ammonium acetate pH 7.5, and 0.1 ml 100% ethanol followed by standing at -20°C for 30 min. DNA was recovered by centrifugation (15 min, microfuge) and the pellet washed with 70% (v/v) ethanol and dried *in vacuo*. The pellet was resuspended in 10 μl of a mixture containing 6-10 ng primer in 40 mM Tris-HCl pH 7.5, 20 mM MgCl$_2$, 50 mM NaCl and incubated at 37°C for 20 min to anneal. Chain termination reactions were performed as described in Section 2.9.6.

2.9.8 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 (sharkstooth) were high density polystyrene (0.4 mm thick) purchased from Bethesda Research Laboratories. The gel mix contained 75 ml

acrylamide stock (5.7% (w/v) acrylamide, 0.3% (w/v) bis acrylamide, 8 M urea, in 1 x sTBE buffer (89 mM Tris base, 89 mM boric acid, 2.5 mM EDTA, final pH 8.3) plus 0.4 ml 25% (w/v) ammonium persulphate and 0.1 ml TEMED. After thorough mixing the gel mix was poured into a clean gel sandwich and the combs inserted upside down. Polymerisation took place for 60 min with the gel in a horizontal position. The gel was mounted on the sequencing apparatus and the combs removed. Gels were pre-electrophoresed at 1,000 V for 30 min. Combs were returned in the correct orientation immediately before sample loading. After samples had been loaded the gel was electrophoresed at a constant voltage of 1,500 V. The bromophenol blue dye front reached the bottom of the gel in about 2 hours. Electrophoresis was for 2, 6, or 10 hours, to allow reading of sequence at different distances from the primer binding site. Plates were separated and tissue paper was lain along the borders of the gel to hold it to the plate during the fixation process, which involved slowly washing the gel with 2 litres of 10% (v/v) acetic acid, 12% methanol, applied with a 60 ml syringe. The gel was then dried at 100°C for 30 min. The gel was placed in direct contact with X-ray film for autoradiography at room temperature for 16-24 hours.

2.9.9 Analysis of DNA sequences

Raw sequence data sets were analysed using the LKB DNA and protein analysis programmes DNASIS and PROSIS. DNA sequences were compared to Genbank and EMBL databases, and predicted protein sequences were compared to PIR and SWISS-PROT databases by Email, using FASTA (Pearson, 1990) or BLAZE (Brutlag et al., in preparation), the latter being temporarily available to the public domain as part of the Genbank contract to Intelligenetics Inc. in late 1992. Multiple sequence alignments were performed using the programme CLUSTAL (Higgins and Sharp, 1989).

2.10 RNA extraction procedures

2.10.1 General RNA extraction method

Total bacterial RNA was isolated by a modification of the procedure of Aiba et al. (1981). 5 ml NB cultures were inoculated with a single colony picked from a fresh plate and grown to an $A_{650} = 0.8$. Cultures were chilled on ice for 10 min and then recovered by centrifugation (10 min, bench centrifuge, 4°C). The pellet was resuspended in 0.5 ml 20 mM

sodium acetate pH 5.5, 1 mM EDTA, 10 mM NaN_3, 0.5% (w/v) SDS and immediately extracted 4 times with hot (65°C) acetate equilibrated phenol (pH 5.5). Each extraction step included a 5 min incubation at 65°C with occasional vortexing. RNA was precipitated by the addition of 2.5 volumes of 100% ethanol and standing at -20°C for 30 min. RNA was recovered by centrifugation (15 min, microfuge), the pellet dried *in vacuo* and resuspended in 0.1 ml 20 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$. Contaminating DNA was removed by the addition of 10 units RNase free DNaseI and incubation at 37°C for 30 min. The mixture was extracted with Tris equilibrated phenol (pH 7.5) and then precipitated twice as above, except that in addition one tenth volume of 3 M sodium acetate pH 5.2 was added. The final dried pellet was resuspended in DEPC treated Milli Q water and stored at -80°C. The concentration of RNA was determined using an LKB Ultrospec plus spectrophotometer, assuming that 1 A_{260} unit ≡ 40 μg/ml RNA (Maniatis *et al.*, 1982). RNA was also analysed by non-denaturing electrophoresis in 1.5% (w/v) agarose mini gels, in a running buffer consisting of 40 mM Tris-acetate pH 8.2, 2 mM EDTA and detection by staining with ethidium bromide as outlined in section 2.8.3.

2.10.2 RNA extraction from cultures stressed for free iron

50 ml NB cultures were inoculated from a single colony picked from a fresh plate and grown to an $A_{650} = 0.4$. At this point cultures were split into two equal portions and the iron chelating agent 2,2' dipyridyl added to one to a final concentration of 0.2 mM and both returned to 37°C incubation. 5 ml samples were removed 0, 30, 60, and 90 min after the addition of 2,2' dipyridyl from both cultures, and RNA was purified as described in Section 2.10.1.

2.11 Analysis and manipulation of RNA

2.11.1 Labelling of RNA by *in vitro* transcription with SP6 or T7 RNA polymerases

RNA probes were derived by run-off transcription of purified, linearised template with bacteriophage SP6 or T7 RNA polymerases. Fragments of interest were subcloned into the vectors pSP64, pSP65 (Melton *et al.*, 1984), pGEM3Zf⁺, or pGEM7Zf⁺ (Promega) and the subclones purified. The resulting plasmids were linearised, extracted with Tris equilibrated phenol (pH 7.5) and precipitated as outlined in Section 2.8.1. Precipitated plasmids were

resuspended in 9 µl DEPC treated Milli Q water, to which 2 µl of a mixture of 5 mM ATP, GTP, CTP, 0.15 mM UTP, 400 mM Tris-HCl pH 7.6, 60 mM MgCl$_2$, 1 mg/ml acetylated BSA, 2 µl 0.1 M DTT, 2 µl of SP6 (5 units/µl) or T7 (2 units/µl) RNA polymerase as appropriate, and 50 µCi (5 µl) α-[^{32}P]-UTP were added. The mixture was incubated at 37°C for 60 min, after which 10 units RNase free DNase was added and the mixture incubated for a further 30 min at 37°C. The sample was then mixed with 10 µl 95% (v/v) formamide, 20 mM EDTA pH 8.0, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol, heated to 70°C for 5 min and loaded on a gel 22 cm long, 16 cm wide and 0.4 mm thick, prepared from 35 ml of the gel mix described in Section 2.9.8 and that had been pre-electrophoresed for 30 min at 400 V. The gel was run at 500 V constant voltage until the bromophenol blue dye front reached the bottom, the plates separated and the full-length run-off transcript located by brief autoradiography (typically 30 seconds) at room temperature. The band was excised, placed in a 1.5 ml reaction tube containing 0.5 ml 15 mM NaCl, 1.5 mM sodium citrate, 200 mM sodium acetate pH 7.0 and incubated at 37°C overnight (Sancar *et al.*, 1982). Probe was then extracted with an equal volume of Tris equilibrated phenol (pH 7.5), and precipitated by the addition of one tenth volume 3 M sodium acetate pH 5.2 and 2.5 volumes of 100% ethanol. Glycogen (20 µg) was added as a carrier. After incubation at -20°C for 30 min, probe was recovered by centrifugation (15 min, microfuge), the pellet washed with 70% (v/v) ethanol, dried *in vacuo* and resuspended in 0.1 ml DEPC treated Milli-Q water. Percentage incorporation was determined by comparing the recovered material to the starting material by counting in a Beckman model LS6000TA β counter.

2.11.2 RNA gel blot transfer and hybridisation (northern blot)

RNA for analysis by northern blotting was denatured by glyoxalation (Thomas, 1980). 10-20 µg RNA was incubated in a mixture of 1 M glyoxal, 10 mM sodium phosphate pH 6.5, 0.1 mM EDTA, 50% (v/v) dimethyl sulphoxide at 50°C for 60 min (final volume of 18 µl). The mixture was then quenched on ice for 2 min, a one tenth volume of 50% (v/v) glycerol containing 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol was added and the mixture loaded on a 1.5% (w/v) agarose gel, poured in an apparatus identical to that described in Section 2.8.3, with a running buffer of 10 mM sodium phosphate pH 6.5, 0.1 mM EDTA. Electrophoresis was at 30 mA constant current with continuous recirculation of the running

buffer. After electrophoresis, gels were set up for unidirectional transfer to nitrocellulose as described by Thomas (1980). Following transfer, filters were air dried and baked for 2 hours at 80°C *in vacuo*. Glyoxal was removed by pouring 500 ml of boiling 10 mM Tris-HCl pH 8.0 onto the filter and allowing it to cool to room temperature. Prehybridisation and hybridisation conditions were as described in Section 2.8.13, except that probes (RNA) were denatured by heating at 70°C for 10 min, instead of 100°C. Filters were washed four times with shaking at room temperature for 5 min in 0.3 M NaCl, 30 mM sodium citrate pH 7.0, 0.1% (w/v) SDS, followed by two washes at 65°C in 30 mM NaCl, 3 mM sodium citrate pH 7.0, 0.1% (w/v) SDS. After washes, filters were sealed in plastic bags and placed on X-ray film (Cronex 10T; DuPont) for autoradiography at -80°C with intensifying screens.

2.11.3 RNA dot blots

RNA for dot blotting was denatured by incubating 5 μg RNA (unless otherwise stated) in a volume of 0.1 ml with three volumes of 18% (w/w) formaldehyde, 1.5 M NaCl, 0.15 M sodium citrate pH 7.0, at 65°C for 10 min. Samples were transferred to nitrocellulose membranes using a Biorad dot blotting apparatus, and the wells were washed once with 1.5 M NaCl, 0.15 M sodium citrate pH 7.0. Filters were air dried and baked for 1-2 hours at 80°C *in vacuo*. Prehybridisation, hybridisation and washes were as described in Section 2.11.2.

2.11.4 5' end mapping by primer extension

Synthetic oligonucleotide primers were end-labelled with γ-[^{32}P]-ATP as described in Section 2.8.10, except that after incubation, the reaction volume was made up to 0.1 ml with Milli Q water and the primer was precipitated with the addition of one tenth volume of 3 M sodium acetate pH 5.2, 20 μg glycogen and 3 volumes of 100% ethanol, followed by incubation at -20°C overnight. Primers were recovered by centrifugation (15 min, microfuge) and the pellets washed with 70% (v/v) ethanol and dried *in vacuo*. Primers were resuspended to 0.1 ng/μl in DEPC treated Milli Q water

1 ng primer was mixed with 20 μg total RNA in a final volume of 0.1 ml 0.2 M NaCl, and precipitated by the addition of three volumes 100% ethanol and incubation at -20°C for 30 min. The precipitate was recovered by centrifugation (15 min, microfuge) and the pellet

washed with 70% (v/v) ethanol and dried *in vacuo*. The pellet was resuspended in 10 μl 10 mM Tris-HCl pH 8.3, 200 mM KCl, heated at 75°C for 3 min and then incubated at 42°C for 60 min to allow hybridisation of the primer to occur. Extension from annealed primers was achieved by the addition of 25 μl 10 mM Tris-HCl pH 8.3, 14 mM MgCl$_2$, 14 mM DTT, 0.7 mM of each dNTP, containing 10 units *per* 25 ml of M-MLV reverse transcriptase and incubating at 42°C for 45 min. Samples were then treated with DNase free RNaseA, extracted with Tris equilibrated phenol (pH 7.5), chloroform, isoamyl alcohol (25:24:1) and precipitated by the addition of one tenth volume 3 M sodium acetate pH 5.2 and three volumes 100% ethanol and standing overnight at -20°C. After centrifugation (15 min, microfuge), pellets were washed with 70% (v/v) ethanol and dried *in vacuo*. Pellets were resuspended in 5 μl Milli Q water, mixed with an equal volume 95% (v/v) formamide, 20 mM EDTA pH 8.0, 0.05% (w/v) bromophenol blue, 0.05% xylene cyanol, heated to 100°C for 3 min and 5 μl loaded onto 6% polyacrylamide-8 M urea gels as described in Section 2.9.8. Electrophoresed and fixed gels were exposed to X-ray film at room temperature. Size markers were sequence reactions generated from appropriate M13 or plasmid templates as described in Sections 2.9.6 and 2.9.7, using the same primer (but not end labelled) as used in the extension reaction.

2.11.5 3' end mapping by RNase protection

RNA probes used were generated as described in Section 2.11.1, and were designed to extend beyond predicted 3' ends. Probes (1 x 10^5 cpm) were mixed with 10 μg total RNA in a final volume of 0.1 ml and precipitated by the addition of one tenth volume 5 M ammonium acetate and three volumes 100% ethanol, followed by incubation at -20°C for 30 min. Two control tubes were set up for each probe, containing probe (1 x 10^5 cpm) and 10 μg yeast RNA. Precipitates were recovered by centrifugation (15 min, microfuge) and pellets dried *in vacuo*. Pellets were resuspended in 20 μl 80% (v/v) deionised formamide, 40 mM PIPES pH 6.4, 400 mM sodium acetate, 1 mM EDTA, heated to 90°C for 3 min and then incubated at 45°C overnight, to allow probe hybridisation to occur. 0.2 ml of a mixture containing 10 mM Tris-HCl pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.1 unit RNaseA and 20 units RNaseT1 was added to all experimental tubes and one control tube, and 0.2 ml of the same buffer but without added RNaseA/T1 was added to the other control tube and the mixtures incubated at

37°C for 30 min. RNase digestion was inhibited by the addition of 20 μl of a mixture containing 10% (w/v) SDS, 0.1 μg/ml proteinase K and incubation at 37°C for a further 15 min. The samples were then extracted with an equal volume of Tris equilibrated phenol (pH 7.5), chloroform, isoamyl alcohol (25:24:1) and RNA precipitated by the addition of 625 μl 100% ethanol and incubation at -20°C for 30 min. RNA was recovered by centrifugation (15 min, microfuge), the pellets washed with 70% (v/v) ethanol and dried *in vacuo*. Pellets were resuspended in 5 μl Milli Q water, mixed with an equal volume 95% (v/v) formamide, 20 mM EDTA pH 8.0, 0.05% (w/v) bromophenol blue, 0.05% xylene cyanol, heated to 70°C for 3 min and 5 μl loaded onto 6% polyacrylamide-8 M urea gels as described in Section 2.9.8. Electrophoresed and fixed gels were exposed to X-ray film at room temperature. Size markers were sequence reactions generated from appropriate M13 or plasmid templates as described in Sections 2.9.6 and 2.9.7, using a primer designed to have the same 5' end as the first protected base of the probe. Size values obtained were approximations, due to the 5 - 10% greater mobility of DNA than RNA under the electrophoresis conditions used (Sambrook *et al.*, 1989).

2.12 Assay for galactokinase (GalK)

Expression of galactokinase (GalK) by reporter gene constructs was determined by introducing plasmids containing the constructs of interest into *E. coli* CB806 and assaying for GalK as described by McKenney *et al.* (1981). 10 ml LB cultures were inoculated with single colonies isolated from fresh MacConkey agar and grown at 37°C overnight. This culture was then diluted 1:10 (10 ml final) in minimal medium containing fructose (0.25%, w/v) as the carbon source and supplemented with 50 μg/ml thiamine. Constructs containing p_{lac} were grown in medium that was also supplemented with 24 μg/ml IPTG. The culture was incubated at 37°C with shaking, until it reached an A_{650} of 0.6. A 1 ml aliquot of the culture was then removed to a fresh tube and cells lysed by adding 40 μl 50 mM Tris-HCl pH 8.0, 100 mM EDTA, 100 mM DTT and one drop of toluene, followed by vortexing for approximately 10 seconds and incubating at 37°C for 60 min. Cell extracts prepared in this manner were used fresh.

D-[1-^{14}C]-galactose was diluted to 4.5 x 10^6 dpm/mmole by mixing 2 μl of labelled galactose (62.5 mCi/mmole) with 0.2 mM non-radioactive D-galactose to a final volume of 1 ml. This mixture was filtered twice through DE81 (Whatman) filters to remove contaminating galactose-1-phosphate before use. For assaying cell extracts, 20 μl of extract (samples with high activity were diluted 1:10 or 1:100 with minimal medium) was mixed with 20 μl 5 mM DTT, 16 mM NaF, 50 μl 200 mM Tris-HCl pH 7.9, 8 mM MgCl$_2$, 3.2 mM ATP, and 10 μl of diluted D-[1-^{14}C]-galactose and incubated at 32°C. 15 μl aliquots were removed at 0, 5, 10, 15, 20 min after the commencement of incubation and spotted onto 23 mm DE81 filter disks. Disks were washed for 2 x 5 min in distilled water, dried, mixed with 3 ml scintillation fluid and radioactivity bound to the filters determined by counting in a Beckman model LS6000TA β counter. Control reactions using minimal medium instead of cell extracts were also performed. Filters were left unwashed to determine total label or washed to determine background cpm. GalK units were determined by plotting the increase in cpm bound to the filters against time. One GalK unit is defined as that amount of GalK required to phosphorylate 1 nmole of galactose.min^{-1}.ml^{-1} for cells at A$_{650}$ of 1.0 (McKenney *et al.*, 1981). The reaction rate in this assay is linear between 0 and 25% conversion of galactose to galactose-1-phosphate (McKenney *et al.*, 1981).

The plasmid copy number of cultures assayed for GalK was determined as described by Projan *et al.* (1983). A second aliquot of 0.5 A$_{650}$ units of cells from the same cultures assayed for GalK was collected by centrifugation (2 min, microfuge), resuspended in 0.1 ml of a mixture of 20 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 mM NaCl, 20% (w/v) sucrose, 25 μg/ml lysozyme and 0.1 mg/ml RNaseA and incubated at 37°C for 30 min. Cells were lysed by the addition of 40 μl of 5% (w/v) SDS and two freeze/thaw cycles. 10 μl of 0.1 mg/ml proteinase K was added and the mixture incubated at 37°C for a further 30 min. The samples were then loaded onto an unsubmerged 1% agarose gel and the wells covered with 0.4% LGT agarose/TBE before submerging, followed by electrophoresis as described in Section 2.8.3. After electrophoresis the gels were stained in distilled water containing 2 μg/ml ethidium bromide for 15 min. DNA bands were visualised by trans-illumination with UV light and photographed using Polaroid 665 negative film. The concentration of DNA in the test samples was determined by scanning the negative with a Molecular Dynamics model

300A computing densitometer and comparing obtained values to those of purified DNAs of known concentration run as standards on the same gel (Projan *et al.*, 1983).

2.13 Protein analysis

2.13.1 Preparation of whole cell samples

Samples of whole cells were prepared for analysis by resuspending 1 ml of an overnight or mid-exponential phase culture in 50 μl PBS, and mixing with an equal volume of 2 x sample buffer (62.5 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol and 0.05% (w/v) bromophenol blue) (Lugtenberg *et al.*, 1975). Samples were then sonicated briefly with a Branson Model B15 Sonifier™ and heated to 100°C for 3 min before storage at -20°C.

2.13.2 T7 RNA polymerase expression system

The plasmid pGP1-2 carries the T7 RNA polymerase gene under the control of the bacteriophage λ p_L promoter, and the λ *cI857* (temperature sensitive) repressor gene (Tabor and Richardson, 1985). When a plasmid containing a T7 RNA polymerase promoter is introduced into cells containing pGP1-2, transcription from this promoter can be induced by shifting the growth temperature from 30°C to 42°C. As the T7 RNA polymerase is resistant to the antibiotic Rif, transcription that is dependent on the host RNA polymerase can be selectively inhibited with this antibiotic, allowing specific labelling or large scale production of proteins encoded by genes cloned behind a T7 RNA polymerase promoter (Tabor and Richardson, 1985).

2.13.2.1 Small scale labelling of proteins with L-[^{35}S]-methionine

A single colony of *E. coli* DH5 containing pGP1-2 and the clone of interest was picked from a fresh plate and 5 ml TB with appropriate antibiotics was inoculated with this colony and incubated at 30°C. When an A_{650} of 0.6 was reached, 0.4 ml of culture was harvested and centrifuged (2 min, microfuge). The pellet was washed three times with minimal medium without added glucose, resuspended in 1 ml methionine assay medium, transferred to 10 ml culture tubes and incubated at 30°C for 60 min with shaking. Cells were then transferred to 42°C. After 15 min incubation, Rif was added to a final concentration of

400 µg/ml and the cultures incubated at 42°C for a further 10 min before shifting incubation to 37°C for 60 min. The samples were pulsed with L-³⁵S-methionine (10 µCi) for 5 min at 37°C, before transfer to 1.5 ml reaction tubes, centrifuged (2 min, microfuge), resuspended in 50 µl PBS, and mixed with an equal volume of 2 x sample buffer (Section 2.13.1). Samples were heated to 100°C for 3 min before SDS-PAGE analysis. Samples were subsequently stored at -20°C for up to 4 weeks. Where desired, globomycin, an inhibitor of type II signal peptidase (Inukai *et al.*, 1978), was added 10 min before the samples were pulsed with L-³⁵S-methionine.

2.13.2.2 Small scale labelling of proteins with [9, 10-³H]-palmitic acid

Samples for labelling with [9,10-³H]-palmitic acid were grown, washed, starved and induced by temperature shift as described in Section 2.13.2.1, except that double quantities of cells were used. After temperature shift, cells were incubated at 37°C for 60 min before the addition of 20 µl 2 mg/ml L-methionine and, where desired, globomycin. Cells were returned to 37°C for 10 min before the addition of [9,10-³H]-palmitic acid (25 µCi) and incubation at 37°C for 1 hour or overnight. Similar results were obtained in either case. After incubation, samples were transferred to 1.5 ml reaction tubes, centrifuged (2 min, microfuge), resuspended in 50 µl PBS, and mixed with an equal volume of 2 x sample buffer (Section 2.13.1). Samples were heated to 100°C for 3 min before SDS-PAGE analysisand then stored at -20°C.

2.13.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on 15% or 20% polyacrylamide gels using a modification of the procedure of Lugtenberg *et al.* (1975) as described previously (Achtman *et al.*, 1978). Gels were 17 cm long, 16 cm wide and 1.5 cm thick. Samples were heated at 100°C for 3 min prior to loading. Gels were generally electrophoresed at 100 V for 5 hours. Proteins were stained with gentle agitation for 15 min in 0.275% (w/v) coomassie brilliant blue R250, 10% (v/v) methanol, 10% (v/v) ethanol, 7.5% (v/v) acetic acid, and destained in the same mixture without coomassie blue. Size markers were generally: phosphorylase B (94 kDa); bovine serum albumin (67 kDa); ovalbumin (43 kDa); carbonic anhydrase (31 kDa); soybean trypsin inhibitor (21.5 kDa); hen egg lysozyme

(14.4 kDa) and were purchased from Biorad. Where necessary, cytochrome C (12.4 kDa) and aprotinin (6.5 kDa) purchased from Sigma, were used in addition to the above. Size markers for gels for western blotting were Biorad prestained markers and were (apparent molecular weights are indicated): myosin (205 kDa); β-galactosidase (116 kDa); bovine serum albumin (80 kDa); ovalbumin (49.5 kDa); carbonic anhydrase (32.5 kDa); soybean trypsin inhibitor (27.5 kDa); hen egg lysozyme (18.5 kDa); and aprotinin (6.5 kDa).

2.13.4 Autoradiography of SDS-PAGE gels

Gels for autoradiography were stained and destained for at least 60 min. Gels containing samples labelled with [9,10-³H]-palmitic acid were in addition treated with Amplify™ (Amersham) for 30 min after destaining. 20% polyacrylamide gels were destained in a solution containing 40% (v/v) methanol, 10% (v/v) acetic acid, 3% (v/v) glycerol for at least 3 hours, in order to protect against cracking during drying (Biorad). Gels containing samples labelled with L-[³⁵S]-methionine were exposed to X-ray film (Cronex 10T, DuPont) at room temperature for 1-7 days. Gels containing samples labelled with [9,10-³H]-palmitic acid were exposed to X-ray film for 1-4 months at -80°C, using intensifying screens. X-ray film was pre-flashed using a Sensitise™ pre-flash unit according to manufacturers instructions (Amersham).

2.13.5 SDS-PAGE gel transfer and protein blotting (western blot)

Western blots were performed by a modification of the procedure of Towbin et al. (1979). Samples were subjected to SDS-PAGE and transferred to nitrocellulose at 0.2 A for 2 hours in a Trans-blot cell (Biorad), using a transfer buffer containing 25 mM Tris base, 192 mM glycine, 5% (v/v) methanol, final pH 8.3. After transfer, the nitrocellulose sheet was incubated for 30 min in 5% skim milk powder in 0.05% (v/v) polyoxyethylenesorbitan monolaurate (Tween 20), 20 mM Tris-HCl pH 7.4, 0.9% (w/v) NaCl (TTBS) to block non-specific protein binding sites. The primary antiserum (rabbit) was diluted 1/5,000 in TTBS containing 0.02% (w/v) skim milk powder and incubated at room temperature with gentle agitation for 2-16 hours. Non-specifically bound antibody was removed by three 10 min washes in TTBS with shaking. Bound antibody was detected using an anti-antibody coupled with horseradish peroxidase and peroxidase substrate. This was accomplished by incubating

the filter for 2 hours (gentle agitation) with goat anti-rabbit IgG-horseradish peroxidase conjugate (Nordic Immunology) at a dilution of 1/5,000 in TTBS. The filter was then washed four times for 5 min with TTBS, followed by two 5 min washes in 20 mM Tris-HCl pH 7.4, 0.9% (w/v) NaCl (TBS). The antigen-antibody complexes were then visualised using peroxidase substrate, which consisted of 9.9 mg 4-chloro-1-naphthol dissolved in 3.3 ml - 20°C methanol added to 16.5 ml TBS containing 15 µl hydrogen peroxide. This mixture was incubated with the filter for 5-15 min with shaking, and the reaction stopped by transferring the filter to 0.1% (w/v) NaN$_3$ (Hawkes *et al.*, 1982).

2.13.6 Colony transfer for blotting with antiserum

Colonies expressing proteins that cross-reacted with specific antisera were detected as follows. Patched colonies were grown for 5 hours at 37°C. Nitrocellulose disks (8.2 cm diameter, Schleicher and Schüll) were then placed on top of the colonies which were allowed to absorb for 5 min. The filters were then placed on blotting paper saturated with 0.5 M HCl, to lyse and fix the colonies, and were left to stand in the dark for 30 min, following which cellular debris was removed with a jet of 0.9% NaCl. Antigen detection was by the method outlined in Section 2.13.5.

2.13.7 Determination of protein concentration

Protein concentration was determined by incubating samples with the BCA™ protein assay reagent (Pierce) in microtitre trays (96 well; Costar) at 37°C for 30 min, then measuring the A$_{570}$ in a Dynatech model MR5000 ELISA tray reader. Concentrations were obtained by interpolation from a standard curve generated from known concentrations of BSA.

2.14 Antisera

Rabbit anti *V. cholerae* O1 Inaba and Ogawa typing antisera were obtained from Wellcome Reagents Ltd. *V. cholerae* O1 Ogawa LPS was detected on a Western blot using mouse anti-live *V. cholerae* O162 antiserum provided by Dr. S. Attridge. Goat anti-bacterial alkaline phosphatase (*E. coli*) was purchased from 5 Prime → 3 Prime Inc®. Goat anti-rabbit-horse radish peroxidase and goat anti-mouse-horse radish peroxidase conjugates were

purchased from Kirkegaard & Perry Laboratories Inc. Sheep anti-DIG-horse radish peroxidase conjugate was purchased from Boehringer Mannheim.

Anti-overexpressed-MrhA and -MrhB antiserum was raised as follows. A 250 ml TB culture of *E. coli* DH5 [pGP1-2, pPM1642] was grown to an A_{650} of 0.6 at 30°C, then shifted to 42°C for 15 min. Rif was added and the cells incubated at 42°C for a further 15 min before being shifted to 37°C overnight. The production of inclusion bodies in cells subjected to temperature shift, but not in cells incubated solely at 30°C, was confirmed by phase contrast microscopy. After incubation, cells were pelleted (5,000 rpm, 10 min, 4°C, GS-3, Sorvall), washed with 50 ml 50 mM Tris-HCl pH 7.5, pelleted as before and resuspended in 10 ml of the same buffer. Cells were disrupted by passing through a French press six times at 8,000 psi and inclusion bodies recovered by low-speed centrifugation (500 x g, 10 min, 4°C, bench centrifuge) then resuspended in 10 ml 50 mM Tris-HCl pH 7.5. MrhA and MrhB were then resolubilised from this mixture by incubation of 2.5 mg (protein) inclusion bodies in a final volume of 5 ml 2 M urea at 37°C for 30 min, followed by centrifugation (35,000 rpm, 60 min, 4°C, Ti-60, Beckman L8-80) to remove insoluble material, and the supernatant was dialysed against four changes of 2 l PBS at 4°C, then centrifuged as above. SDS-PAGE analysis revealed the presence of three species of approximately 10, 27 and 32 kDa in this supernatant (not shown). The presence of the overexpressed proteins in fractions produced was followed throughout by SDS-PAGE analysis (20% acrylamide) on minigels (Biorad Miniprotean) and Coomassie blue staining as described in Section 2.13.3. An 11 week old rabbit from the central animal house of the University of Adelaide was pre-bled and this serum shown not to contain antibodies which cross reacted with *E. coli* DH5, *V. cholerae* Z17561 or O17 whole cells, or the partially purified fraction obtained above, in a western blot (not shown). The rabbit was then immunised with 100 µg of the partially purified fraction emulsified in Freunds complete adjuvant (Commonwealth Serum Laboratories), and boosted four times at two week intervals with 100 µg of the partially purified fraction emulsified in Freunds incomplete adjuvant (Commonwealth Serum Laboratories). The production of anti overexpressed MrhA and MrhB antibodies was confirmed by western blot.

2.15 Porphyrin binding assays

2.15.1 Colony substrate binding assay

Binding of various chromophores to *E. coli* colonies expressing proteins overproduced *via* the T7 induction system was examined by plate induction. Samples to be tested were streaked on minimal medium supplemented with thiamine (50 μg/ml), casamino acids (1% (w/v); Difco), and one of: congo red (0.1 mg/ml); haemin (0.2 mg/ml); haematoporphyrin (0.2 mg/ml); protoporphyrin IX (0.2 mg/ml). Plates were incubated at 30°C for 24 hours for growth and then at 42°C for 24 hours for induction. Binding was determined by a change in the colony pigmentation.

2.15.2 Liquid haemin binding assay

The liquid haemin binding assay used was based loosely on the assay of Hanson and Hansen (1991) incorporating the T7 induction system. 10 ml LB cultures were inoculated with single colonies and grown to an $A_{650} \approx 0.6$ at 30°C. Temperature was shifted to 42°C for 15 min and then to 37°C for 60 min. Cells were then pelleted (10 min, bench centrifuge) and resuspended to a density of 10 A_{650} units/ml in sterile PBS. 90 μl cells were then mixed with 10 μl haemin (10 x final concentration) in 1 mM NaOH and incubated at 37°C for 60 min. The effect of urea on haemin binding was examined by resuspending cells to a density of 20 A_{650} units/ml in sterile PBS and then mixing 45 μl cells with 50 μl 8 M urea and then adding 5 μl haemin (20 x final concentration) in 1 mM NaOH and incubating at 37°C for 60 min. Cells were pelleted by centrifugation (2 min, microfuge) and 90 μl of supernatant transferred to microtitre trays (96 well; Costar) and the amount of haemin remaining in the supernatant determined by measuring the A_{410} in a Dynatech model MR5000 ELISA tray reader and comparing to a standard curve of concentrations of haemin.

2.16 Detection of PhoA activity

PhoA activity of translational fusions to the portion of *phoA* encoding mature PhoA was determined by patching colonies on minimal medium supplemented with thiamine (50 μg/ml) and X-pho (40 μg/ml) and incubating at 37°C overnight. Colonies expressing active PhoA (*i.e.* PhoA located in the periplasmic space) developed a blue colour on incubation (Manoil and Beckwith, 1985).

2.17 Haemagglutination

2.17.1 Preparation of erythrocytes

Erythrocytes were prepared from female BALB/c mice and were collected by retro-orbital bleeding of anaesthetised animals. Blood was harvested in 3.8% (w/v) sodium citrate pH 7.0. Erythrocytes were collected by centrifugation at 500 x g for 5 min and were resuspended in 100 volumes of Krebs Ringer buffer (KR) (Kitchen, 1984) modified by using ADA as a buffer instead of sodium bicarbonate (Freter and Jones, 1976). KR consists of 10 mM ADA pH 7.0, 0.69% (w/v) NaCl, 0.2% (w/v) glucose, 0.016% (w/v) KH_2PO_4, 0.036% (w/v) KCl, 0.029% (w/v) $MgSO_4 \cdot 7H_2O$, 0.039% (w/v) $CaCl_2 \cdot 2H_2O$. D-mannose was also added to a final concentration of 1% (w/v). Erythrocytes were washed three times in KR and resuspended to a final concentration of 1% (v/v). Erythrocytes were used fresh in haemagglutination assays.

2.17.2 Haemagglutination assays

10 ml NB cultures for haemagglutination assays were shaken overnight at 37°C. Bacterial cells were pelleted by centrifugation (10 min, bench centrifuge) and resuspended to a density of 10 A_{650} units in KR. Bacterial suspensions were twofold serially diluted in round-bottomed microtitre trays (no. 1-221-24; Dynatech Laboratories Inc.) in 50 μl KR. A 50 μl aliquot of erythrocytes (Section 2.15.1) was added, the tray was tapped and the erythrocytes allowed to settle for 60 min at 25°C. The haemagglutinating titre was defined as the reciprocal of the highest dilution in which haemagglutination was visible. All assays were performed on cultures grown in triplicate.

2.17.3 Haemagglutination inhibition with antisera

For haemagglutination inhibition assays, the haemagglutination titre for the bacterial suspension was first determined and the suspension adjusted to contain twice the haemagglutination dose (2 haemagglutination units). Antisera were then twofold serially diluted in 50 μl of the bacterial suspension (in KR) in microtitre trays and the mixture incubated at 37°C for 60 min. Erythrocytes were then added and the trays incubated at 25°C for a further 60 min, after which the trays were analysed.

2.18 Adherence to cultured epithelial cell monolayers *in vitro*

2.18.1 Maintenance of the human colon carcinoma cell line Caco-2

The Caco-2 cell line, which originated from a moderately differentiated human colonic adenocarcinoma (Pinto *et al.*, 1983), was obtained from the American Type Culture Collection (ATCC number HTB 37, batch number 10366). The cell line was cultured in minimal Eagle medium (MEM) with Earle salts, non-essential amino acids, and glutamine, (GIBCO Life Technologies) supplemented with 10% (v/v) foetal bovine serum (FBS) (GIBCO) penicillin (12 μg/ml) and Gm (16 μg/ml). For adherence assays, 5 x 10^5 Caco-2 cells (between 25 and 40 passages) were seeded into 12 ml of medium and distributed in 24 well tissue culture trays (Corning Glass Works) containing 13 mm tissue culture cover slips (Mediglass) in each well. After incubation at 37°C for 72-96 hours in a 10% CO_2/90% air atmosphere, the culture medium was changed to MEM with 2% (v/v) FBS. Cell cultures were maintained with biweekly changes of the medium until brush border differentiation was complete at approximately 10-12 days post-inoculation (Panigrahi *et al.*, 1990). Representative samples were examined by scanning electron microscopy to confirm the presence of a brush border layer (not shown).

2.18.2 *In vitro* adherence assay

Bacterial cultures were grown in triplicate at 37°C overnight in NB. Cells were pelleted (10 min, bench centrifuge) and resuspended to 10 A_{650} units in KR. Suspensions were then adjusted to 1 A_{650} unit or 0.1 A_{650} unit in MEM without FBS, penicillin or Gm, but containing 1% (w/v) mannose and 5 μg/ml Cm, and the mixture homogenised. Viable counts were performed on samples by plating appropriate dilutions on NA with appropriate selection. Differentiated Caco-2 cells (Section 2.18.1) were washed three times with PBS containing 1% mannose and overlaid with 0.5 ml of bacterial suspension. After incubation at 37°C for 60 min in a 10% CO_2/90% air atmosphere, monolayers were washed three times with PBS containing 1% mannose, then once with PBS containing 0.1% triton X-100 to lyse the Caco-2 cells, and appropriate dilutions were plated on NA with appropriate selection. Adherence was determined as the percentage of input viable cells remaining associated with the monolayer after washing. Preliminary experiments established that *V. cholerae* 569B is resistant to 0.1% Triton X-100 and that 5 μg/ml Cm is bacteriostatic but does not diminish viable count within

the time course of the experiment (not shown). As an alternative to triton lysis and viable counting, monolayers were fixed in methanol, stained with Giemsa and examined by oil-immersion light microscopy.

2.19 *In vivo* assay for virulence using the infant mouse cholera model

Virulence of *V. cholerae* strains was assessed using the infant mouse model of cholera infection (Attridge and Rowley, 1983a). 3-4 day old infant mice, of body weight 2.4-2.7 g, were removed from their mothers about 5-6 hours before use, to permit emptying of stomach contents. Challenge strains were picked from fresh plates and grown to an $A_{650} \approx 1.0$ at 37°C in NB. Cells were pelleted (10 min, bench centrifuge) and resuspended in 0.1% (w/v) proteose peptone (Difco), 0.85% (w/v) NaCl (PS). Cells were serially diluted 10 fold and 0.1 ml of the suspension fed to groups of eight infant mice using a smooth-tipped hypodermic needle. Viable counts of the administered cultures were also determined by plating dilutions on NA with appropriate selection. Mice were kept on tissue paper in plastic containers for 48 hours at 25°C, after which time the number of survivors was noted and these data used to construct a plot of cumulative percentage mortality against the log dose administered (Reed and Muench, 1938). By interpolation, an estimate of the LD_{50} (the dose of vibrios capable of killing 50% of the mice within 48 hours) was obtained. Strains with an $LD_{50} \geq 1 \times 10^8$ are considered avirulent in this model.

Competition of parent and mutant strains *in vivo* was assessed by orally administering to infant mice a mixture of the two strains to be tested in different ratios. Groups of mice were of sufficient size that triplicate samples could be obtained for each time point. At given times, animals were sacrificed, their entire intestines removed and homogenised in 2 ml 0.9% (w/v) NaCl. *V. cholerae* organisms in these samples were enumerated by viable count. Competition of parent and mutant strains *in vitro* was performed by mixing exponential-phase cultures of the two strains in approximately 1:1 ratio in NB or minimal medium and incubating at 37°C. Samples were taken at 0, 3 and 6 hours for counting of viable organisms in the presence or absence of selection.

2.20 *V. cholerae* motility assay

Motility of *V. cholerae* strains was tested prior to use in animal experiments by overlaying plates with soft agar (Attridge, 1979).   Strains were streaked on NA with appropriate selection and incubated overnight at 37°C. Following incubation, each plate was overlaid with 3 ml H-top agar, allowed to set at room temperature, and incubated at 37°C for 2-3 hours. Colonies of motile bacteria develop a halo on incubation as cells swim into the soft agar overlay.

2.21 Assessment of production of cholera toxin by *V. cholerae*

Prior to use in animal experiments, the cholera toxin production status of strains was determined by a GM1-ELISA assay (Voss and Attridge, 1993).   5 ml AKI cultures were grown and the supernatant collected. Microtitre trays (96 well; Costar) were pre-incubated with 2 μg/ml GM1-ganglioside (Sigma). Trays were washed four times with PBS containing 0.05% (v/v) Tween 20 (PBS-Tween) and non-specific binding sites were blocked by incubating with 20% (w/v) BSA, 0.05% (v/v) Tween 20, 0.0125 M triethanolamine, 0.14 M NaCl (BSA-Tween). This was followed by incubation for 2 hours at 37°C with replicate serial dilutions of a standard cholera toxin preparation (4 μg/ml) or with appropriate dilutions of culture supernatants. The trays were washed with PBS-Tween and incubated for 4 hours at 37°C with the IgG fraction of a polyclonal rabbit anti-cholera toxin serum (diluted 1:70,000). Again, after washing with PBS-Tween, the trays were incubated for 17-19 hours at 4°C with sheep anti-rabbit IgG conjugated with alkaline phosphatase (diluted 1:1,000) in 0.14 M NaCl, 0.25 M triethanolamine, 0.002% (w/v) BSA, 0.5 mM $MgCl_2$, and 1.25 μM $ZnCl_2$. The trays were washed thoroughly with PBS-Tween and incubated for 90 min at 37°C with substrate (1 mg/ml di-sodium p-nitrophenylphosphate (Sigma), 10.5% (w/v) diethanolamine, 1 mM $MgCl_2$, final pH 9.8). The A_{405} was measured using a Titertek Multiscan ELISA tray reader and the toxin content of the test samples estimated by interpolation from the standard curve.

2.22 Analysis of lipopolysaccharide (LPS) expression

Whole cell lysates were prepared by resuspending the pellet from 1.5 ml of an overnight culture, grown in NB, in 50 μl of lysis buffer (2% (w/v) SDS, 4% (v/v) β-mercaptoethanol, 10% (v/v) glycerol, 1 M Tris-HCl pH 6.8, 0.1% (w/v) bromophenol blue),

heating at 100°C for 10 min and then incubating with 2.5 μg proteinase K (in 10 μl lysis buffer) at 65°C for 2-4 hours (Hitchcock and Brown, 1983). Samples were separated on 20% SDS-PAGE gels (Section 2.13.3). Gels were then transferred to nitrocellulose for western blotting as described in Section 2.13.5, or silver stained for LPS as described by Tsai and Frasch (1982).

Chapter 3

Nucleotide sequence of pPM471, a Mannose-Fucose-Resistant Haemagglutinin expressing clone

3.1 Introduction

The genes encoding the MFRHA have been cloned by two methods. Franzon and Manning (1986) probed a plasmid bank derived from the classical strain 569B with an antiserum raised against the haemagglutinin/protease (HA/protease) (Finkelstein and Hanne, 1982). Out of the two thousand colonies screened, a single positive colony was detected. The resulting clone, pPM471, corresponds to a 6.3 kb *Bam*HI fragment of the 569B chromosome encoding the MFRHA: why this strategy produced a MFRHA and not a HA/protease clone is not known (Franzon, 1988). Van Dongen and DeGraaf(1986) isolated a cosmid clone from a bank derived from the El Tor strain C5 by virtue of its MFRHA activity. A cutdown of this cosmid retaining MFRHA activity, pVC hag-11, also contained a 6.3 kb *Bam*HI insert. Both authors assigned MFRHA activity to a 25 kDa protein encoded within this region, but van Dongen *et al.* (1987) altered this assignment to an adjacently encoded 7 kDa protein. The reason for the discrepancy in results between the two groups is not clear.

The nucleotide sequence of a portion of this 6.3 kb fragment has been determined by both groups (van Dongen *et al.*, 1987; Franzon, 1988). Franzon (1988) sequenced the ORF corresponding to the 25 kDa protein, while van Dongen *et al.* (1987) sequenced the upstream ORF encoding the 7 kDa protein as well. Analysis of the pattern of transcription of this region has been undertaken (S. G. Williams, unpublished data; see also Chapter 6) and indicates that the genes for the 7 kDa and 25 kDa proteins are co-transcribed. Some differences in nucleotide sequence exist between the two clones (Chapter 7) although whether these differences lead to the discrepancy in results described above is not clear. The region immediately downstream of the 25 kDa ORF was also sequenced in a separate study and shown to contain an ORF predicted to encode a 16.3 kDa protein (Clark, 1988). The role (if any) that this protein has in MFRHA activity is not known, but it is of interest that a gene replacement mutation that abolishes MFRHA activity in 569B removes most of this ORF as well (Franzon, 1988; Chapter 7). This indicates that expression of MFRHA activity could well involve sequences flanking those already determined. Furthermore, the nature of the MFRHA is not known. A number of bacterial adhesins have been shown to be proteinaceous filamentous appendages called fimbriae, which are often encoded in a single biosynthetic operon (Mooi and DeGraaf, 1985). *V. cholerae* have been shown to bear at least three types of fimbriae (Hall *et al.*, 1988) and although fimbriae have not been detected on MFRHA-

positive cells (Franzon, 1988), the possibility that the MFRHA is fimbrial in nature has not been disproven. Therefore, in order to ascertain the likelihood or otherwise of this possibility, the nucleotide sequence of the region of pPM471 upstream of that previously determined was analysed and is reported, along with some computer analyses, in this chapter.

3.2 Results

3.2.1 Detection of multiple copies of a repeated sequence in pPM471

The genes encoding the MFRHA had been localised previously to a 2.4 kb *Hind*III fragment of *V. cholerae* 569B (Franzon and Manning, 1986) and the original aim of this sequencing project was, in collaboration with others (Franzon, 1988; Clark, 1988), to determine the nucleotide sequence of this fragment. However, van Dongen *et al.* (1987) noted that the MFRHA genes were contained between two copies of a 124 bp direct repeat, and a third copy of this repeat was found at the 3' end of the ORF immediately downstream of the MFRHA genes (Clark, 1988). Therefore, a Southern blot was performed on various digests of pPM471 (Franzon and Manning, 1986) using oligonucleotide #128 (which was designed to bind within the repeat sequence; Chapter 2) as a probe (Fig 3.1A). The availability of sequence covering the entire 2.4 kb *Hind*III fragment indicated that there were three copies of the repeat on this fragment; consequently densitometric analysis of this Southern blot was undertaken. This suggested that there were at least six copies of the repeat contained within pPM471 (Fig. 3.1B).

3.2.2 Nucleotide sequence of pPM471

In order to determine the arrangement and possible function of these repeats and the genes associated with them, it was decided to sequence the entire insert of pPM471. The nucleotide sequence of part of this region has been determined elsewhere (Franzon, 1988; Clark, 1988), and the extent of that part of the region is indicated in Fig. 3.2. In order to be able to determine the sequence of the remainder of pPM471, appropriate subclones in bacteriophage M13mp18 and M13mp19 (represented schematically in Fig. 3.2) were generated as follows. The proximal end of pPM471 was subcloned by ligating *Bam*HI-*Hind*III digested pPM471 with M13mp19 or M13mp18, to generate M13/1601 and M13/1602, respectively. The 0.55 kb *Hind*III fragment was gel purified and ligated with

Figure 3.1: Detection of multiple copies of a repeated sequence in pPM471

Panel A shows a Southern blot of pPM471 DNA probed with oligodcoxynucleotide #128. Panel B shows the integer values of band intensities shown in panel A, as determined from analysis with a Molecular Dynamics model 300A computing densitometer, assuming that the 2.4 kb *Hind*III band contains three copies of sequences that hybridise with the probe. This band is indicated with an arrow in panel A. Lanes are pPM471 DNA digested with: 1, *Bam*III; 2, *Hind*III; and 3, *Alu*I.

A   1   2   3   B   1   2   3

                                           6   1

►   ━━   3
                                               2

                                               1

                                           1   3
                                           1

Figure 3.2: Schematic representation of the subclones used to determine the nucleotide sequence of pPM471.

The numbered scale indicates the position in kb on the sequence, with the positions of some restriction endonuclease recognition sequences indicated. The regions spanned by subclones used to determine the nucleotide sequence are indicated by lines below the map, and are annotated with the subclone name. Subclone construction is detailed in the text. A blank box indicates the portion of pPM471 that was originally sequenced by Franzon (1988) and a shaded box indicates the portion of pPM471 originally sequenced by Clark (1988).

M13/1601

M13/1602

M13/1603

M13/1604

M13/1605

M13/1606

M13/1607

M13/1608

M13/1609

M13/1610

pPM1699

M13/1611

M13/1612

M13/1613

M13/1614

6.302 kb

Bam HI

Hin dIII

Hin dIII

Hin dIII

Bgl II

Hin dIII

Dra I

Acc I

Nru I

Dra I

Eco RI

Eco RI

Bam HI

0 1 2 3 4 5 6
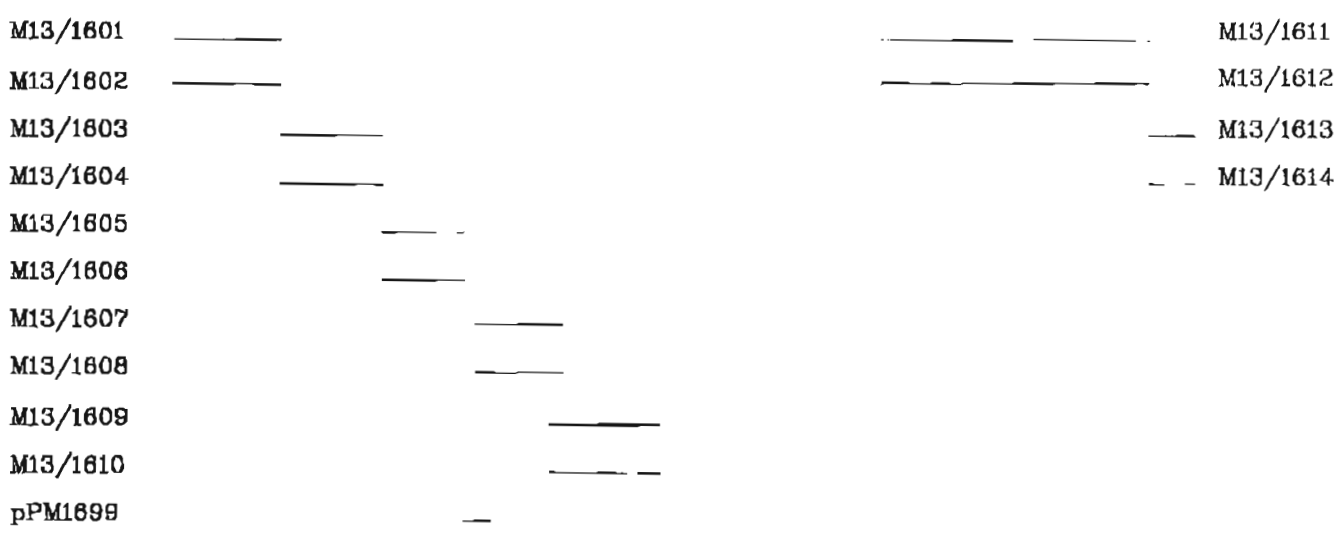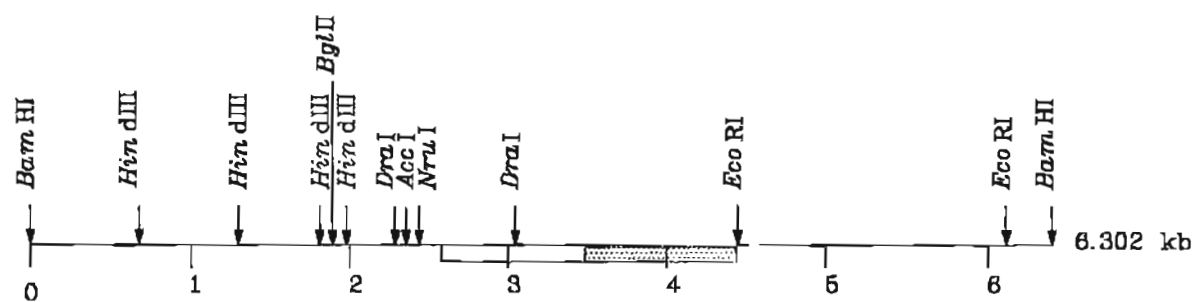
M13mp19 to give M13/1603 and M13/1604. The orientations of insert were determined by restriction mapping with *Cla*I, which cuts close to one end of this *Hind*III fragment. Similarly, the 0.65 kb *Hind*III fragment was gel purified and ligated with M13mp19 to give M13/1605 and M13/1606. The orientations of insert were determined by DNA sequencing. M13/1607 was derived by subcloning the 0.6 kb *Bgl*II-*Nru*I fragment into *Sma*I-*Bam*HI digested M13/mp19. Attempts to clone the same fragment into *Sma*I-*Bam*HI digested M13mp18 were unsuccessful (not shown), consequently M13/1608 was derived by digesting pPM1612 (Chapter 6) with *Acc*I, end repairing, then digesting with *Bgl*II and ligating the mixture with *Sma*I-*Bam*HI digested M13mp19. Sequences generated from M13/1607 and M13/1605 do not overlap, so *Hind*III digested pPM471 was analysed by polyacrylamide gel electrophoresis (PAGE). The presence of an approximately 80 bp *Hind*III fragment was revealed (not shown), which was gel purified, and subcloned into plasmid pK18, to give plasmid pPM1699. The sequence of Franzon (1988) contained a *Dra*I site approximately 400 bp from the proximal end, therefore subcloning of the *Dra*I fragment containing this proximal end would allow the derivation of overlapping sequence. Electrophoresis of *Dra*I digested pPM471 on PAGE gels revealed eight distinct fragments. Five of these, in the size range 0.5-0.7 kb (not shown), were gel purified and the fragment overlapping with the region sequenced by Franzon (1988) identified by dot blot, using oligo #61 (Chapter 6). This fragment was subcloned into *Sma*I digested M13mp19, to give M13/1609 and M13/1610. The region downstream of that sequenced by Clark (1988) was subcloned in two portions. The 1.8 kb *Eco*RI fragment of pPM471 was gel purified and subcloned into M13mp19 to give M13/1611 and M13/1612, respectively. Finally, the distal portion of pPM471 was subcloned by ligating *Eco*RI-*Bam*HI digested pPM471 with *Eco*RI-*Bam*HI digested M13mp19 and M13mp18, to give M13/1613 and M13/1614, respectively. Numerous attempts were made to subclone larger fragments from the portion of pPM471 proximal to the region sequenced by Franzon (1988). However, the derived constructs were extremely unstable, exhibiting a very high rate of deletion (not shown). Subsequent analysis of smaller fragments indicated that the unstable fragments contained multiple copies of a direct repeat (see Fig. 3.3 and Fig. 3.4).

The nucleotide sequence of the subclones was determined by extension from the universal (-20) primer binding site present in M13mp18 and M13mp19 (Yanisch-Perron *et*

Figure 3.3: The complete nucleotide sequence of the insert of pPM471.

ORFs are indicated by the single letter code for amino acids, with the stop codons indicated by an asterisk. For ORFs of the same polarity as the nucleotide sequence, the amino acid sequence is above the nucleotide sequence and putative translation initiation sites are underlined. Underlined sequences between nucleotides 6121 and 6180 annotated with a question mark are a proposed alternative initiation site. For ORFs of the opposite polarity to the nucleotide sequence, the amino acid sequence is below the nucleotide sequence, and putative translation initiation and termination sites are indicated by complementary strand sequence. Nine copies of a 124 bp repeat are indicated by bold type in the nucleotide sequence. Sequence from nucleotides 2688-3380 was originally reported by Franzon (1988), although some corrections are reported here, and sequence from nucleotides 3281-4244 was origianlly reported by Clark (1988), and is again shown with some corrections. The nucleotide sequence of pPM471 is available from the EMBL database under the accession number X64097.

.

```
          D P H F P V Q K L A I A E I I A L R H K C I P K N P R F V Y S K D Q I I N M S L S
     5' GGATCCTCATTTTCCAGTAGCAAAAGCTGGCGCTAGCTGAGATCATTGCACTTAGAAAGTGTATCCTAAGAACCCCAGGTTTGTCCATTCAAAGATCAGATAATTAACATGAGCTTGTC

          T M A K N P C I A Y G I N E R H L S H V R R F V L L E N C I K I I N T *
   121 CACAATGGCTAAAAATCCCGACATTGCTTATGGAATCAATGAGGAGCATCTTTCACACGTGCGTAGAGAGGTTTTACTTGAAATGGGTATCAAAATAATAAACACATAACAAACGCTTCA
   241 AGAGGGACAGCCAACGCGTGGCATTTTTATTATGCGTTGGTTTTTGTGATTATGCGTTAAGTTAGTAGTAGCGTTGGCTGCCCCTTAAGCGGGCGTTATAAACCTAATGGGAA

          M N L E H P Q E S D F P D L L I K W I D S D E L N Y L W G C P A Y V P P L T Y E
   361 ATCATGAACCTGCAAGAATTCGGATTTCCAGAATCGGATTTTGACCTGCTTATAAAATGGATTGACTCAGATGAGTTGAACTACTTATGGGTTGCCCTGCGTATGTTTTCCCGCTGACTTATGAA

          Q I N S H C S K A E V F P Y L K V V K G R H A G R V E L Y K V T D H C Y R I C R
   481 CAGATCCATTCCCACTGTAGCAAAGCGGAAGTATTTCCCTATCTGCTAAAAGTGAAAGGTAGACATGCCGGGTTTGTTGAGTTGTACAAAGTCACCGATCAGCAATATCGGATTGCCGT

          V F I S N A Y R G C Q G L S K S H M L H L L I D K A R L D F S A T K L S L G V F R Q
   601 GTCTTTATTTCAAATGCTTACCGTGGGCAAGGGCTATCAAAATCGGATGTGTGATGTTATTAATCGATAAAGCCCGATTAGATGTTTCTGCAAACCAAGCTTAGCTTAGGCGTTTTTGAACAG

          N T V A R K C Y R G S L C F R V V S F H I C T R A F N G K L W D L V R M E K R L *
   721 AACACAGTTGCTAGAAAATGCTATGGCGTCATTAGGTTTTGAAGTGGTTTCGACAGAAATTGGTACAAGAGCTTTTAATGGTAAGTTATGGGATCTCGGTGCGAATGGAAAAAACGCTTAATAA
   841 CAAACGCCTCAAGAGGGAACTGTCAACGCGTGGCGTTTCCAGTCCCATTGAGCCGCGGTGGTGCTGTTGTTGTGTTTGAGTTTAGTGGTAGTGCGTTGTCAGCCCCTTAGGCGGGCGTTA

          M R A I P L I L C S V L L N C C L G M P E S V K P V S D F P L H N Y L
   961 GGTATTTGGAGATATCATGGAGCTATCTTTTTGATTCTTTGCTCTGTTTTAAATGGCTGCTTGCCGGAATCATGAAACCAGTGTCGGATTTTGAACTGAACAACTATT

          G K W Y E V A R L D H S F E R G L S Q V T A E Y R V R N D Q G I S V L N R G Y S
  1081 AGGCAAATGGTACGAGGTTGCTCGACTCGATCACTCCTTTGAAAGAGGTTTAAGTCAGGTTACTGCGGAATACCGTGTTCGAAATGATGGTGGTATTTCGGTTCTTAATCGTGGTTATCC

          M E K G F N K E A R G K A Y F V N G B T D G Y L K V S P F G Y G S Y V V F H
  1201 TGAAGAGAAAGGTGAGTGGAACGAAGCGGCAGGTAAAGCTTACTTTGTGAATGGCTCAACAGATGGCTATCTGAAAGTTTCATTTTTTGGCCCGTTTTATGGCTCCTACGTAGTGTTTGA

          L D R E N Y S Y A F V S C F P N T E Y L W L L S R T P P T V E R G I L D K F I P M S
  1321 GTTAGACCGTGAAAACTACAGTTATGCTTTTGTGTCAGGTTTCCCGAATACAGAATATCTGTGGTTACTTTCAAGAACGCCGACCGTAGAACGAGGCATTCTGGACAAGTTCATAGAAATGTC

          K R K G P D T N R L I Y V D L Q *
  1441 GAAAGAAGCGTGGTCTTTGATACAAATCGGCTTATTTACGTTCAGCTGCAATAAATACCTAACAAACGCCTCAAGAGGGACTGTCAACGCGTGGCGTTTCCAGTCCCATTGAGCCGCGGTGG
  1561 TTTCCGGTTGTTGTGTTTGAGTTTAGTGGTAATACGTTGCCAGCCCCTTAGGCGGGCGTTAAATAC
                                                              3' AGTAGC 5'
                                                              COOH-     * D D L R Q L A K T Y V E S R R

  1681 TACCGACAGCTAAAACAGTAACTACGATGATGTCATTTTCAACTTGATAGACTAAGCGATAAGCAGACTGCCGAAGTTTTATTTTATAAATGTCTCAGCGCCAGACAGTTTGGCTGATG
          K G V A L V T V V S I D N E V Q V V L R Y G S G Q R L K I K Y I N E A G S K A S
  1801 GAACATGTGGATTTTCAAGGCGTTCCTATTAGTTTCTTCTTGAACTGCCGTTGGAGTGGAACGGCTAACTTTTTCCATTCCTTCAAAGCACTCTTTTGAACCTCAAGCCTTATAGGTCATCG
          P V H P N E L R M I L K K K F Q Q Q L P V A L K K W E K I A S K K F E L K Y T M-NH2
                                                              3' AAT   GTA 5
                                                              COOH *   L D D

  1921 ATATTTACCGATACGCTGTCTTCACTTAAGCGTTCTTTGGCAATAGCCAGTAGCTCAAGATCTTCAAGCTTATCCATCATGATCTCGTAGGTCGTTGCAGGGACGCAGTAGAAAGCCGGC
          I N V S V S D E S I R E K A I A L L E I D E L K D H N I H Y T S A P V C Y P A P
  2041 TCATTTCGGTTGAGTACAAGCAAGAGGAGCACCGAAGGCAGTAGTTGCAACTTTCATTCGGGTTGCTTTAAATTCGGTAATGCTTGCTGCAACATCAGCTAGGATTCTAGCGGTCATTTAA
          E N R N L V A V P A G P A E T A V K M P N A K P E T I S A A V D A L I K T T M-NH2
                                                              3 GTA
  2161 TCGGTGCCTTTATCTGGTCTTTAATTGGGTCATTATAGTCTTGAAAATGCAGTTTAACAAACGCCTCAAGAGGGACTGTCAACGCGCGGCGTTTCCAGTCCCAATGAGCCGCGGTGGTGC
                                                              AGCA 5'
  2281 AGTTGTTGTGTTTGGGTTTAGTGTTATGCGTTGCCAGCGCCCTGACGGGGGCGGTAGGCATAAAATTCATAAGCAATTGCTTGATATTGTCCATATTGCCCATATTTAAATTATGGCATT

                          M F D L T K A R C P C R G C S K V A N N A E R V E R L Z G L R N M G
  2401 AAAATAAAAGGAGTAGACCATTAATGCCAGATTTAACGAAAGCTCGCTGTCCGTGTCGTGGATCGAATAAAGTTGCGAACAATGCAGAAGAGGTAGAAGAGCTGTTTGGCCTCAGAAATATGG

          D G T I R V Q S Y C R E C R G L H C E A S S P K C N H H A *
  2521 GTGATGGTACGAATCGGCGTACAATCTATATTGTCGGCGAATGTCGGAGGCTTGCATTGTGAAGCTAGTAATCCTAAATGTAATCATCATCGTAGGTATTTTGATGATTTGTTAGATAAGAAA

  2641 ATGAATATAGGTTAATGAACAGATACGATTTTAACGTATCTGTCATTTCATGTTAGGTTATGCGTTTCCCTCGATATTAGGCAATTTAATGTCAAAATTTATCAAATGGATGCGGTTGATT
                                                  M S K I Y Q M D A V C W

          L K T L E N C S V D L F I T D P P Y E S L R K Y R Q I V T C T R L K M S K R S S
  2761 GGCTTAAAAACACTTGAAAATCGTAGTGTTGATCTGTTCATCACTGATCCACCATATGAATCGCTAGAAAATATAGACAAATAGTTACGACTACACGGTTAAAAGAGAGTAAATCATCGA

          N Q W P S V F P N T P E E F R P V V R V L K K G E H P Y L F C D Q E T F A I
  2881 GCAATCAATGGTTAGTGTTTTTCCTAACACTAGGTTTGAAGAGTTGTTTCGTGAAGTTTATAGAGTGCTAAAAAAAGGTTCTCATTTCGTATTTATTTTGCGACCAGGAAACTATGTTTT

          A K F I A H S V C M R F W K P I V W D K C C R I C H G Y E Y R A R Y K F I L P P A
  3001 TGGCGAAAACCAATTGCGCATAGCGTAAAGTGTAGGCTTTAAATTTTGGAAGCCTACAGTTTGGGATAAGTGTGTAGATATGCGGATACCATTATCGTGCTAGATAGTGAATTTAATTCTATTTTCGG

          K G K R K L N D L G U P D V L E Y K R V W K G Y P T M K P V E L L E V L I R G S
  3121 AGAAAGGAAAGAGAAAGCTTAAATGATTTAAGTCGTTCCTGATGTGTTGGAATATAGAGAGTTTGGAAAAGGCTACCCAACAGAAGAAGCCAGTTGAACTTCTGGAGGTTTGATTAGGCGAA

          S S R N E I V A D S F G G A T L I A A N N S R K Y I G C D I S M R A H T Y
  3241 GGTCTTCAGAAAATGAAATTGTAGCGGATTCATTTTTTGGTTCAGGCGCAACTTTAATTGCAGCTAATAATCTGTCTAGAAAATACAGTGGTTGTGATATATCAATGTGTGCAACATGAGT

          F K N R A *
  3361 ATTTTTAAGAATGAAAGCTTGAAACTTATTTTATGCCTAACAAACGCCTCAAGAGGGACTGTCAACGCGTGGCGTTTCCAGTCCCATTGAGCCGCGGTGGTGCAGTTGTTGTGTTTGAGTTTA
  3481 GTGGTAATGCGTTGCCAGCCCCTTAGGCGGGCGTTAGCAATTCACAGTTTGGCTTTATTTTTGTTGATGCATTCTTCCCTGATTATGTATAAGTCGTTTTGAAAATAACCTTTAACACCATG

          M F K S N F E V Q L L V N A I R F R K P Y F A N D F
  3601 ATTACCATTAGTTTCTTGCTAAGCTTAAATAACATGGAGAATAAAGGTAAGAAAGGTACAGTTGCTAGTAAATGCTATTCGGTTTCGTAAACCCTACAGGGTTTAATGATT

          Q V T A S E L K V K K E S V H L S K I K S I E K K L L G F K K D N V V K V I T Y A
  3721 TTCAAGTAACAGCATCTGAGCTGAAAGTGAAAAAAGAGAGTTATCACTTAAGTAAAATCAAGAGTATTGAACTGAAGCTGCTTGGCTTCAAAGATAACGTGGTTAAAGTGACAACATACG

          I L B A A T W A F V F E F G E F P A L L I A L I A L V S M R K Y E K V E F K
  3841 CCATTTTAGTCAGCAGCGAGCTGGGCTTTGTTACCGAGTTTGGTTTTTTTTGCGGCTCCTAATTGCTCTGATTTTGGCTTTAGCTTGTCGATGAGAAAGTATGAGTTGAAAGTAGAGTTTA

          G T D E T G D Y W V S I A R C H T E D E F H I L D N I Y T R L K S R I *
  3961 AAGGTACAAGATGAAACGGGTGATTACTGGGTATCAATAGCCCGTTGTTGCAGCGGAAGATGAATTTCAATTTATACTGAAGTCGAAAAGCAAGAATCTAGCACAGAGGCT
  4081 CGTCAATTTTGGCGAGCTTTGAGTTTATGCTCCGGACCTAACAAACGCCTCAAGAGGGACTGTCAACGCGTGGCGTTTCCAGTCCCATTGAGCCGCGGTGGTGCATTAGGTGAATGTGGTTT

                          M L V N E M F F S P L A R F R Q
  4201 GAGTTTAGTGGTGGTGCGTTTGTCATCCCCTTAGGCGGGCGTTAGGTGAATGTAAGTGGATTGAACAAGGACAGGTTTAGTAAATGGTTTTTTAGCTTTTGGCTCGTTTTCGTTCA

          L F W G R F H A H D A V L C Q A Y G E Q Q S E L Q V Q G Y G Q V V K V L P D E R
  4321 GCTTTTTTCGGTCCGGTTTCATGCCAATGACGCCGTTTTACAACAAGCTTATCAATCGCAACAAAGTCGATTTACAGGTTCAAGCATTCGGACAGGTAGTGAAAGTGTTACCTGACGACAA

          D G S K H Q K F I L K L N S G Q T L L V A H N I D L A P R I P N L R V C D S V R
  4441 TGATGGTTCAAAGCATCAAAAGTTCATCTTAAAGCTCAATTAGCGGACAAAACATTGCTGGTTGCCCATAACATCGACTTAGCACCGGAGAATTCCGAACTTGAAAGTTGGCGATAGTGTTGA

          F M V N M N G T K K K V G F F V G L I K I P I Q N R H A H G W L K H N G Q V V Y R
  4561 GTTTTATGGTGAATATGAATGGAACAAAAAGGTGGGATTCTTTGTTGGTCTCATCAAAATTCCTATTCAGAATCGTCATGCTCATGGTTGGTTAAAGCACAATGGGCAGGTGTCAGAGTAAAT
  4681 TCACCTAACAAACGCCTCAAGCGGGGACTGTCAACGCGCGGCGTTTCCAGTCCCAATGAGCCGCGGTGGTGCGTTGAGTTTAGTGGTAATGCGTTGCCAGCCCCTTAGGC

                      M E K R R R Q K Y S L L E F I V V Q S G N Q Y D N A A S V L A
  4801 GGGGCGTTAGGTTTACTGAGTGAAGTCTTGATGGAAAAGAAGAGAAGACAAAAGTACTCGTTGTTAGAACCAATAGTTGTTCAGTCTGGGAAACCAATATCTAAAATGCGATGGTTGC

          S T D G N S W V A G I N A S L A I K I Y L F S P I I T E H V Q G Y S R F M G I
  4921 AAGCACGGATGGAAAATCGTGGGTAGCCGTGTATCAATGTATCGTTGGCTATAGAGATTTATTCAAAATCATTTTTGGATTACAGGACATCAAGTGCAAGGGTACTCAAGATTTATTGGGAAC

          V H K S E D L E K L P K K L H P Q K Q S V L I R A V K N V A P G I D L V A E L T
  5041 CGTTCATAAATCACGAGGACTAGCAAAATTATTTAAGAAACTACAGCCGCAAAAACAACTCGGTACTTTCAGAGGCAGGCTCAAACATCATTGTCTCCAGGCATCGGATCTAGTTTGCTGAAACTGAC

          K Y R N V F T N A R Y T F E P N N V P S M G S Q L I N L A S V L R R A V I H I I
  5161 TAAATACCGGAAATGTCTTTACGAATGCTAGGGTATACTTTGAACCAAACAATGTTCCATCTATGGGTTCTGACCTAATAAATTTAGCTAGTGTTTTAAGACGTGCGGTTATCCATATCAT

          N V H Y P V K F I G L E K A V N E I L S E R O T *
  5281 AAATGTTCATTATCCAGTGATTAAACCCATAGGTCTTGAAAAGGCGGTCAATGAGATACTTTCGACACGGCAAACATAACAAACGCCTCAAGAGGGACTGTCAACGTCGTAGCCGTTCCA

                                              M K R I K Y I L
  5401 GTCCCAATGAGCCGCGGTGGTGGTTGGTTGTTGTGTTTGGGTTTAGTGGTAATGCGTTGCCAGCCCCTTAGGCGGGCGTTAGGCTAAAGGAAGTAGCAAAAATATTAAAATACATTTTT

          F I G V V A Y M P F V S A N L S K L D K G I S L L N T L N Y P A I P K N T L L K
  5521 ATTTATTTGCGTTGTAGCGTACAGCCCATTTGTAAGTGCTAATCTATCGCATTTGGATAAAAGCATAAGTGTCCTTAATACATTAAAATATCCTGCAATATTTAAAATACATTACTTAA

          R E S V S D I R K G G L P C P A S L K I C K Q S H T L S G E H I V R I A A K R P I S V S E
  5641 AGAGTGTGATAAGGGCGTCCCACAGTTCGCTCACTAAAAAGTTTGTAAACAAATCGAACGAAAGAACGGATTCTGCTAAGAGTTCGTTTCGTATTCATATGTATCGA

          S D T D R L I A F W G G S D G V R I S N G L V V L W Q Q I G R L D V L T D T D L Y
  5761 GAGTGGATAAGGAGTAGTGCTTTTTGGGGTGGAAGCGACGGTGTGCGTATTTCTAACGGGCTTGTTGTTTTATGGCAGCAGATCGGACGTTTAGATGTTCTGACCGATACTGATTTGTA

          I L D K F N Q T A E H R K A S M A F A D A V N S V G I L L E I E R I L S S Q *
  5881 TATTTTAGACAAATTTAACCAAACAGCAGAAAATAAAGCATCTATGGCATTTGCCGACAATGTTAATGTAAGTGTGGCAAATAATCTTAGAAATCGAAGAAATTCTAAGCAGTCAATAGCC
  6001 TAACAAAGCGTTAAAGCGGGACTTGGCACGCGTGGTGGTTGGTTAAGGTTAATGCGGGTTCACTTTTGTATTACGTGCCGTCGCCCCTTAACGCGGCGTT

          ?         ? (K) M L T F K N V S R M L M G L V L V G A S F G P Q V V A K L T L
  6120 AGGTTGCTCAACGAAAGGTGAATTCTGTGTAAATGTTAACCTTTAAAAATGTTTCACGAAATGCCTATGAGTTTGGTTTTGGTGGGTCGCGTCACCGCAAGTTTATGCAAAGTTAACACTG

          D E F C E H R Y D L A F A S I G D R M C
  6241 GATGAATTTTGCGAAAATGCGGATAAGACTCTAGCCTTTGCGGCCTATAAGCGATAGAATGGATCC 3'
                                                              6302
```

Figure 3.4: Schematic representation of features of the nucleotide sequence shown in Figure 3.3.

The numbered scale indicates the position in kb on the sequence, with the positions of some restriction endonuclease recognition sequences indicated. ORFs are indicated by blank boxes with the size in kiloDaltons and the assigned name below. A hatched portion at the proximal end of ORF2 indicates the presence of a predicted signal peptide. Filled boxes represent the location of copies of a 124 bp direct repeat. The predicted direction of transcription is indicated by arrows.

BamHI MluI HindIII MluI EcoRV HindIII MluI HindIII HindIII BglII NruI EcoRV XbaI MluI XbaI MluI HindIII EcoRI MluI EcoRI BamHI

0   1   2   3   4   5   6   6.302 kb

18k   19k   11k9k 7k   25k   16k   15k   20k   18.6k

mrhA   mrhB

ORF1   ORF2 ORF3.2   ORF6   ORF7   ORF8   ORF9

ORF3.1
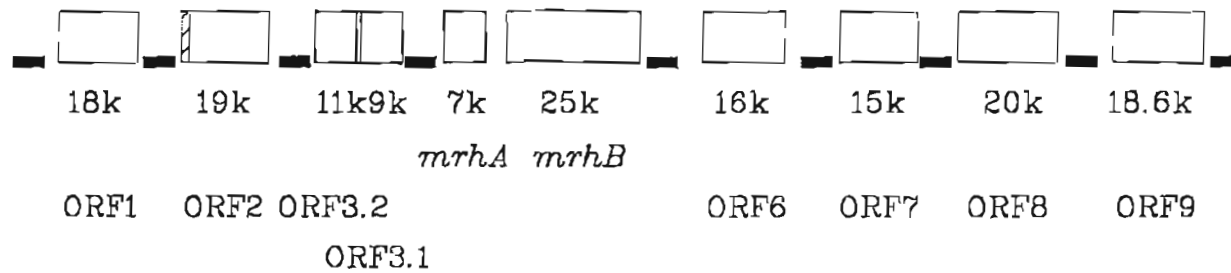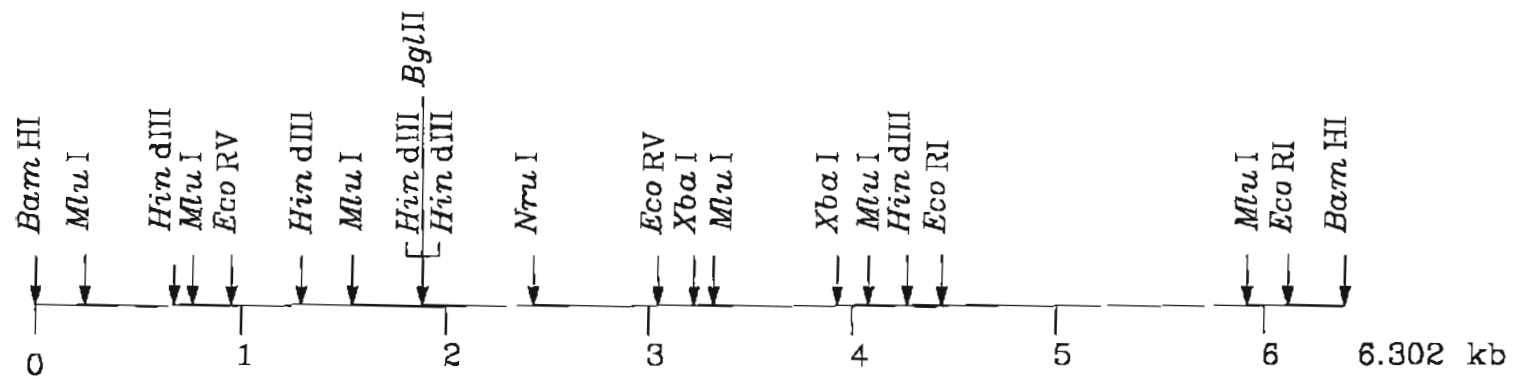
direction of transcription

■ 124 bp direct repeats

al., 1985), or from the forward and reverse primer binding sites present in pK18 (Pridmore, 1987). Where sequence data sets did not extend over the entire cloned region, oligo primers were designed to bind approximately 100 nucleotides from the end of the data set, to obtain overlapping data sets (primer binding sites are not shown, but the sequence of the primers used is presented in Chapter 2). Where possible, M13 subclones were constructed so that sequence data sets overlapped adjacent subclones. Where such an arrangement was not possible, the junction sites of subclones were sequenced directly from pPM471, using primers designed to bind approximately 100 nucleotides from the subclone junction (not shown). The nucleotide sequence of the entire insert of pPM471 is reported in Fig. 3.3, and shown schematically in Fig 3.4.

Ten ORFs are apparent within the sequenced region, eight of which are of the same predicted transcriptional polarity. In addition, two partial ORFs occur at either end of the cloned DNA, and both of these are of the same predicted polarity as the majority of ORFs on the clone (Fig. 3.3). Two of the ORFs are the same as those sequenced by van Dongen *et al.* (1987) and encode the 7 kDa and 25 kDa proteins that constitute the MFRHA. It is proposed to name these genes *mrh*, for D-mannose, L-fucose resistant haemagglutinin; *mrhA* for the 7 kDa ORF and *mrhB* for the 25 kDa ORF, respectively. Further analysis of these sequences is shown in Chapter 7. The other ORFs have been designated numbers as shown in Fig. 3.4. For example, the 16.3 kDa ORF sequenced by Clark (1988) is here designated ORF6. ORF2 is predicted to have a type II signal sequence, i.e. to be a lipoprotein (Pugsley, 1989). Further analysis of this ORF is presented in Chapter 5.

Surprisingly, more copies of the repeat sequence were found than were predicted from Fig 3.1; nine as opposed to the six predicted (Fig. 3.3). This discrepancy can be explained by the imperfect match between the oligo probe used in Fig. 3.1 and all copies of the repeat, as the stringent hybridisation conditions used preclude binding of the probe to three of the copies sequenced (a single mismatch would lead to a decrease of 5-10°C in T_m; Wood *et al.*, 1985). In accordance with the terminology used elsewhere (Fujitani *et al.*, 1991), it is proposed to name these repeats VCR, for *V. cholerae* repetitive DNA sequence. Further analysis of these repeats is undertaken in Chapter 4.

An analysis of the size and amino acid content of the complete ORFs is shown in Table 3.1, along with predicted isoelectric point (pI) values. In all cases, the utilisation of amino acids is not striking. The isoelectric point values, however, show a range of distributions, from the acidic (pI=4.19 for ORF3.1; Table 3.1) to the basic (pI=10.68 for ORF3.2). The pattern of codon usage, as well as the G+C% of the ORFs is shown in Table 3.2. Statistically significant comparisons of the codon usage of most of the ORFs is difficult, due to their small size. For example, the 3 proline residues of ORF3.1 (Table 3.1) are all encoded by the triplet pair CCA, the most common codon for proline in *V. cholerae* sequences available (Table 3.2) but the one phenylalanine residue of ORF3.1 (Table 3.1) is encoded by the "rare" UUC codon (Table 3.2). However, by considering the average codon usage for ORFs encoded on pPM471, some differences in codon usage between these ORFs and *V. cholerae* sequences available can be seen for arginine codon usage (Table 3.2).

3.2.3 Computer analysis of open reading frames

The predicted protein sequences obtained were used to search the PIR (Release 33) and SWISS-PROT (Release 22) databases for similar entries, using the BLAZE algorithm with access provided by Intelligenetics Inc. as part of the Genbank service (Brutlag *et al.*, in preparation). Significant matches to entries in the databases were obtained for the following: ORF2, ORF3.1, ORF 3.2 and *mrhB*. The results for ORF2 and *mrhB* will be presented in Chapters 5 and 7 respectively, while those for ORF3.1 and ORF3.2 are shown in Fig. 3.5.

3.2.3.1 ORF3.1

ORF 3.1 shows 32% identity to the amino-terminal 72 amino-acids of the 179 amino-acid DnaT protein (Primosomal protein I, and here designated $DnaT_{(1-72)}$) of *E. coli* (Masai *et al.*, 1986; Masai and Arai, 1988). This result is intriguing in that DnaT is a component of the primosome complex involved in normal and stable cellular DNA replication (Zavitz and Marians, 1991). No such role has been postulated for any protein encoded within pPM471 and must be considered unlikely for ORF3.1 as: it shows 'low' similarity to $DnaT_{(1-72)}$ (32% identity); it is significantly truncated with respect to DnaT (Fig 3.5A); and *dnaT* in *E. coli* exists in an operon containing at least one other primosome complex component, *dnaC* (Masai and Arai, 1988). Basic computer predictions of secondary structure and hydropathy of

Table 3.1 The amino acid content and size predicted for the 10 complete ORFs encoded within pPM471. The predicted isoelectric point values are also shown. All data are derived from the sequence shown in Fig. 3.3, as detailed in Chapter 2.

ORF			ORF1	ORF2	ORF3.2	ORF3.1	*mrhA*	*mrhB*	ORF6	ORF7	ORF8	ORF9
Size (aa)			158	171	95	82	63	217	141	134	176	166
Residue	Phe	Mol%	6.33	6.43	2.08	3.66	1.59	7.37	6.38	7.46	2.84	4.82
		No.	10	11	2	3	1	16	9	10	5	8
	Leu	Mol%	12.66	10.53	12.5	9.76	6.35	8.29	12.06	10.45	10.8	12.05
		No.	20	18	12	8	4	18	17	14	19	20
	Ile	Mol%	4.43	3.51	5.21	6.1	1.59	5.53	7.09	2.24	7.39	9.64
		No.	7	6	5	5	1	12	10	3	13	16
	Met	Mol%	2.53	1.75	1.04	4.88	3.17	2.3	1.42	1.49	1.7	2.41
		No.	4	3	1	4	2	5	2	2	3	4
	Val	Mol%	6.96	8.19	9.38	8.54	4.76	6.45	7.8	8.21	9.66	6.63
		No.	11	14	9	7	3	14	11	11	17	11
	Ser	Mol%	6.96	8.19	5.21	7.32	3.17	8.76	6.38	4.48	7.39	9.64
		No.	11	14	5	6	2	19	9	6	13	16
	Pro	Mol%	1.9	2.92	3.12	4.88	4.76	3.69	1.42	2.99	3.98	1.81
		No.	3	5	3	4	3	8	2	4	7	3
	Thr	Mol%	3.8	3.51	3.12	6.1	3.17	4.15	6.38	1.49	4.55	4.82
		No.	6	6	3	5	2	9	9	2	8	8
	Ala	Mol%	5.06	3.51	6.25	14.63	7.94	4.61	7.8	5.22	7.95	6.63
		No.	8	6	6	12	5	10	11	7	4	11
	Tyr	Mol%	5.06	7.02	5.21	2.44	1.59	5.53	4.26	2.99	3.98	3.01
		No.	8	12	5	2	1	12	6	4	7	5
	His	Mol%	1.9	0.58	1.04	0	4.76	1.38	1.42	5.97	3.98	1.2
		No.	3	1	1	0	3	3	2	8	7	2
	Gln	Mol%	3.16	1.75	6.25	0	1.59	2.3	1.42	9.7	5.11	4.22
		No.	5	3	6	0	1	5	2	13	9	7
	Asn	Mol%	3.16	5.26	3.12	4.88	9.52	3.69	3.55	6.72	6.25	6.02
		No.	5	9	3	4	6	8	5	9	11	10
	Lys	Mol%	6.96	4.68	14.58	4.88	4.76	8.76	10.64	6.72	6.82	7.23
		No.	11	8	14	4	3	19	15	9	12	12
	Asp	Mol%	5.06	4.09	3.12	7.32	3.17	4.61	4.26	5.97	2.27	6.63
		No.	8	7	3	6	2	10	6	8	4	11
	Glu	Mol%	8.23	9.36	8.33	8.54	9.52	8.29	8.51	3.73	5.11	4.82
		No.	13	16	8	7	6	18	12	5	9	8
	Cys	Mol%	2.53	1.17	0	1.22	11.11	1.84	0.71	0	0.57	1.81
		No.	4	2	0	1	7	4	1	0	1	3
	Trp	Mol%	1.9	1.75	1.04	0	0	2.3	2.13	2.24	0.57	1.2
		No.	3	3	1	0	0	5	3	3	1	2
	Arg	Mol%	5.7	6.43	6.25	3.66	9.52	5.99	3.55	3.73	5.11	1.81
		No.	9	11	6	3	6	13	5	5	9	3

		Gly	Mol%	5.7	9.36	3.12	1.22	7.94	4.15	2.84	8.21	3.98	3.61
			No.	9	16	3	1	5	9	4	11	7	6
Mol. wt.				18411	19670	11247	8937	7034	25583	16333	15506	19834	18616
Isoelectric point				6.75	4.99	10.68	4.19	7.59	8.92	9.09	9.11	10.5	5

Table 3.2: Codon usage of the ORFs presented in Fig. 3.2 compared to sequenced genes of *V. cholerae* O1 and to genes expressed at high or low level from *E. coli*[a]. The G+C content (Mol%) of each ORF is also indicated.

Codon/aa		ORF1	*vlp*	ORF 3.2	ORF 3.1	*mrhA*	*mrhB*	ORF6	ORF7	ORF8	ORF9	pPM471[b]	*V. cholerae*	*E. coli* (high)	*E. coli* (low)
UUU	Phe	90	90.9	0	33.3	100	81.2	88.8	70	100	75	80	70.6	18.3	64.9
UUC	Phe	10	9.1	100	66.7	0	18.8	11.1	30	0	25	20	29.4	81.7	35.1
UUA	Leu	45	33.3	25	12.5	25	27.7	11.8	42.9	31.6	40	31.3	28.2	1.4	16.4
UUG	Leu	15	11.1	8.3	0	25	22.2	23.5	28.6	15.8	15	16.7	18.9	1.5	15.3
CUU	Leu	10	22.2	25	25	0	16.7	5.92	14.3	21.1	10	15.3	17.9	3.5	12
CUC	Leu	0	5.5	8.3	12.5	25	5.5	5.9	7.1	0	15	6.7	9.7	4.8	10.8
CUA	Leu	10	0	16.7	37.5	0	16.7	17.6	0	21.1	20	14	12.6	0.6	3.5
CUG	Leu	20	27.8	16.7	12.5	25	11.1	35.3	7.1	10.5	0	16	12.6	88.2	41.9
AUU	Ile	57.1	66.7	20	40	100	50	70	33.3	23.1	62.5	50	50.7	18.6	49.5
AUC	Ile	28.6	16.7	40	60	0	41.7	20	66.6	30.8	12.5	24.4	28.4	81.2	39.6
AUA	Ile	14.3	16.7	40	0	0	8.3	10	0	46.1	25	25.6	20.9	0.2	10.9
AUG	Met	100	100	100	100	100	100	100	100	100	100	100	100	100	100
GUU	Val	45.4	50	66.7	42.9	33.3	71.4	16.7	50	52.9	18.2	46.4	34.5	50.6	26
GUC	Val	18.2	0	22.2	42.9	0	0	0	0	17.6	27.3	10	15.7	6.8	22.8
GUA	Val	9.1	21.4	11.1	14.3	66.6	14.3	50	16.7	11.8	45.4	24.5	25.1	26.2	15.7
GUG	Val	27.3	28.6	0	0	0	14.3	33.3	33.3	17.6	9.1	19.1	24.7	16.4	35.5
UCU	Ser	9.1	14.3	40	0	0	21.1	11.1	0	23.1	12.5	14.9	20.9	38.8	14
UCC	Ser	9.1	14.3	0	0	0	0	0	0	0	0	3	7.7	34.4	12.8
UCA	Ser	36.4	35.7	20	0	50	31.6	22.2	16.7	23.1	18.8	25.7	21.6	2.1	13.4
UCG	Ser	27.3	21.4	0	16.7	0	10.5	11.1	16.7	30.8	18.8	17.8	11.9	3.6	16.2
AGU	Ser	0	14.3	20	33.3	50	26.3	44.4	33.3	7.7	37.5	23.8	22.1	1.6	16.1
AGC	Ser	18.2	0	20	50	0	10.5	11.1	33.3	15.4	12.5	14.9	15.9	19.4	27.4
CCU	Pro	33.3	0	0	50	33.3	37.5	50	50	14.3	33.3	28.6	30.5	6.1	16
CCC	Pro	33.3	20	0	0	0	0	50	0	14.3	0	9.5	13.1	1.0	15.5
CCA	Pro	0	20	100	25	33.3	62.5	0	0	71.4	66.6	42.9	36	11.9	23.5
CCG	Pro	33.3	60	0	25	33.3	0	0	50	0	0	19	20.5	81	45
ACU	Thr	16.7	33.3	66.7	40	0	55.5	22.2	50	25	25	32.8	26.6	40.3	37.8
ACC	Thr	33.3	0	33.3	60	0	0	0	0	25	12.5	15.5	25.7	52.2	21
ACA	Thr	50	50	0	0	0	33.3	44.4	50	37.5	62.5	37.9	27.3	3	37
ACG	Thr	0	16.7	0	0	100	11.1	33.3	0	12.5	0	13.8	20.4	4.5	26.2

Codon	AA														
GCU	Ala	37.5	66.7	50	41.7	40	30	45.4	57.1	42.9	45.4	44.4	31.7	38.8	15.3
GCC	Ala	25	0	33.3	16.7	20	0	9.1	28.6	0	9.1	12.2	16.2	8.8	28.8
GCA	Ala	12.5	0	0	41.7	20	30	18.2	14.3	35.7	36.4	24.4	29.8	26.6	22.3
GCG	Ala	25	33.3	16.7	0	20	40	27.3	0	21.4	9.1	18.9	22.4	25.8	34
UAU	Tyr	87.5	50	80	0	100	83.3	50	75	71.4	60	67.7	66.4	24.5	65.2
UAC	Tyr	12.5	50	20	100	0	16.7	50	25	28.6	40	32.3	33.6	75.5	34.8
CAU	His	66.7	0	100	0	100	100	50	75	100	50	80	65.7	18.4	60
CAC	His	33.3	100	0	0	0	0	50	25	0	50	20	34.3	81.6	40
CAA	Gln	60	33.3	50	0	100	80	50	69.2	88.8	71.4	68.6	65.7	10.7	38.9
CAG	Gln	40	66.7	50	0	0	20	50	30.8	11.1	28.6	31.4	34.3	89.3	61.1
AAU	Asn	40	66.7	66.7	50	83.3	87.5	60	66.7	90.9	90	74.3	62.3	5.5	53.6
AAC	Asn	60	33.3	33.3	50	16.7	12.5	40	33.3	9.1	10	25.7	37.7	94.5	46.4
AAA	Lys	81.8	62.5	57.1	75	100	63.2	80	77.7	83.3	91.7	71.9	71.6	79	72.5
AAG	Lys	18.2	37.5	42.9	25	0	36.8	20	22.2	16.7	8.3	28.1	28.4	21	27.5
GAU	Asp	75	71.4	0	66.7	100	90	100	50	50	63.6	69.2	77	34.2	68.3
GAC	Asp	25	28.6	100	33.3	0	10	0	50	50	36.4	30.8	23	65.8	31.7
GAA	Glu	61.5	68.8	75	57.1	66.7	72.2	50	60	66.7	75	65.7	64.5	79.7	65.7
GAG	Glu	38.5	31.2	25	42.9	33.3	27.8	50	40	33.3	25	34.3	35.5	21.3	34.3
UGU	Cys	25	0	0	0	85.7	75	100	0	100	100	65.2	61.2	39.6	45.3
UGC	Cys	75	100	0	100	14.3	25	0	0	0	0	34.8	38.8	60.4	54.7
UGG	Trp	100	100	100	0	0	100	100	100	100	100	100	100	100	100
CGU	Arg	22.2	36.4	0	0	16.7	23.1	40	60	11.1	0	22.9	28.1	74.3	31.8
CGC	Arg	0	0	50	38.3	50	0	0	20	0	0	11.4	20.5	25.0	40.3
CGA	Arg	22.2	27.3	16.7	33.3	16.7	0	0	0	11.1	33.3	14.3	15.6	0	9.6
CGG	Arg	22.2	9.1	33.3	0	0	7.7	20	0	11.1	0	11.4	6.7	0.2	10.8
AGA	Arg	33.3	27.3	0	33.3	16.7	53.8	40	20	55.5	33.3	34.3	21.4	0	4.6
AGG	Arg	0	0	0	0	0	15.40	0	11.1	33.3	5.7	7.6	0	2.9
GGU	Gly	55.5	43.8	66.7	100	40	44.4	75	45.4	28.6	33.3	46.5	42.5	58.4	35.5
GGC	Gly	11.1	50	33.3	0	40	33.3	25	9.1	28.6	16.7	28.2	25.1	40	33.2
GGA	Gly	0	0	0	0	20	22.2	0	27.3	28.6	33.3	14.1	17.9	0.5	14.6
GGG	Gly	33.3	6.2	0	0	0	0	0	18.2	14.3	16.7	11.3	14.6	1.1	16.7
UAA	Stop	100	100	0	100	0	0	0	100	100	0	50	45.6	95.8	51.1
UAG	Stop	0	0	0	0	100	0	100	0	0	100	30	27.2	4.2	4.2
UGA	Stop	0	0	100	0	0	100	0	0	0	0	20	27.2	0	44.7

G+C Mol% 39.7 41.9 41.3 45.5 44.4 34.3 37.6 40.2 36.9 33.1 [c]   47–49[d]

[a] This codon usage table was compiled from 72 sequenced *V. cholerae* O1 genes (P. A. Manning, manuscript in preparation). The 10 ORFs presented constitute 10 of these genes. The codon usage for genes expressed at high and low levels in *E. coli* (Andersson and Kurland, 1990) is included by way of comparison.

[b]The codon usage of ORFs present on pPM471 is expressed as the sum of the occurrence of each codon as a percentage of the total usage for each amino acid, not as an average of the codon usage for each ORF.

[c] The region as a whole has a G+C content of 40.6%., which drops to 38% when the 9 copies of VCR, which have a G+C content of 52.8% (collectively) are excluded from the analysis.

[d] The G+C content of *V. cholerae* quoted is from Baumann *et al.* (1984)

Figure 3.5: Pairwise alignments of ORF3.1 and ORF3.2 with similar sequences.

Panel A shows an alignment between ORF3.1 and the amino-terminal 72 amino acids of DnaT of *E. coli* (Masai *et al.*, 1986; Masai and Arai, 1988). Full-length DnaT is 179 amino acids in length. Panel B shows an alignment between ORF3.2 and the hypothetical RelE protein of *E. coli* (Bech *et al.*, 1985). Identical residues in the alignments are joined by a solid line (|), similar residues are joined by a dotted line (:). The boundaries of the region of similarity are indicated by an X above and below the alignment. Numbers annotating the sequence indicate the position from the amino terminus of each residue. The alignments were performed using the algorithm BLAZE (Brutlag *et al.*, in preparation).

A

```
            X          10         20         30         40         50
ORF3.1  MTTRILADVAASITEFKANPMKVATSAFGAPVAVLNRNEPAFYCVPASTYEIMM
        |::|:|:     :|  :    :    |  : | |: |||:  |  |||| |  ::
DnaT    MSSRVLTPDVVGIDALVHDHQTVLAKAEGGVVAVFANNAPAFYAVTPA------
            X          10         20         30         40

             60         70         X
ORF3.1  DKLEDLELLAIAKERLSEDSVSVNIDDL
         :|   ||||:  |:|:  :   | :||
DnaT    -RLA--ELLALE-EKLARPGSDVALDDQ
          50         60        70X
```

B

```
            X          10         20         30         40         50
ORF3.2  MTYKLEFKKSALKEWKKLAVPLQQQFKKKLIERLENPHVPSAKLSGAENIYKIK
        |:|  |:|    |||||:||:   :::|:||||:| ||:|:: :  || |   : ||||
RelE    MAYFLDFDERALKEWRKLGSTVREQLKKKLVEVLESPRIEANKLRGMPDCYKIK
            X          10         20         30         40         50

             60         70         80         90   X
ORF3.2  LRQSGYRLVYQVENDIIVVTVLAVGKRERSEVYTKALQRLD
        ||  ||||||||||  :: :||  |::||||||||||: |: |:
RelE    LRSSGYRLVYQVIDEKVVVFVISVGKRERSEVYSEAVKRIL
             60         70         80         90   X
```

ORF3.1 protein and DnaT$_{(1-72)}$ show that both are slightly hydrophobic in nature (mean values of 0.20 for ORF3.1 and 0.28 for DnaT$_{(1-72)}$), and show alternating stretches of hydrophobic and hydrophilic residues (Fig. 3.6A). Secondary structure predictions do not show marked similarity beyond predicting α-helical character for the similar hydrophilic domains that lie within residues 57-70 of ORF3.1 and 49-60 of DnaT (Fig. 3.6B).

3.2.3.2 ORF3.2

ORF3.2 protein shows 56% identity with the hypothetical RelE protein of *E. coli* (Bech *et al.*, 1985). Information on this protein is scant, but it is contained within the *relB* operon and has been hypothesised to be the translational inhibitor that accumulates during amino-acid starvation of *relB* mutants (O. H. Karlström, personal communication). Again, this similarity is surprising, given that pPM471 has not been suspected of encoding proteins of any such activity (see also Section 3.2.4). Comparison of the predicted secondary structure and hydropathy plots of ORF3.2 protein and RelE were again performed (Fig 3.7). In this case the two proteins show distinct similarity, with overall hydrophilic character (mean values of -0.53 for ORF3.2 and -0.36 for RelE) and only one markedly hydrophobic domain which occurs between residues 60-75 for both proteins (Fig. 3.7A). The predicted secondary structures are more revealing in this case, predicting a predominantly α-helical amino terminal domain, followed by a β-sheet domain that overlaps with the hydrophobic portion of the proteins, followed by a turn and the tail of the protein (Fig. 3.7B). All these predictions together strongly suggest a similar function for the two proteins.

3.2.3.3 Other ORFs

The predicted products of ORFs 1, 6, 7, 8 and 9 did not give any significant positive scores when compared with PIR or SWISS-PROT databases, using either the BLAZE (Brutlag *et al.*, in preparation) or FASTA (Pearson, 1990) algorithms (not shown). The hydropathy plots of these proteins were determined and are shown in Fig. 3.8. Most of these proteins are hydrophilic, with only ORF9 showing significant hydrophobicity. None of these ORFs have amino-terminal signal sequences typical of exported proteins (Pugsley, 1989). Therefore, determination of the cellular location of the products of these ORFs requires further analysis.

Figure 3.6: Analysis of the sequences of ORF3.1 and the amino-terminal 72 amino acids of DnaT of *E. coli*.

Panel A shows hydropathy plots for ORF3.1 (top) and $DnaT_{(1-72)}$ (bottom) determined according to Kyte and Doolittle (1982), using a window of six amino acids. Panel B shows secondary structure predictions for ORF3.1 (top) and $DnaT_{(1-72)}$ (bottom) determined using the algorithm of Chou and Fasman (1978).
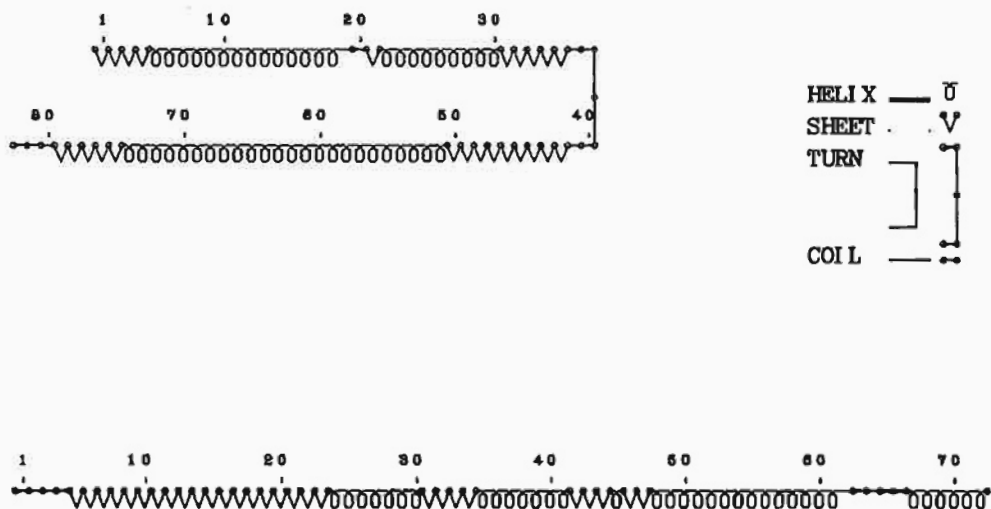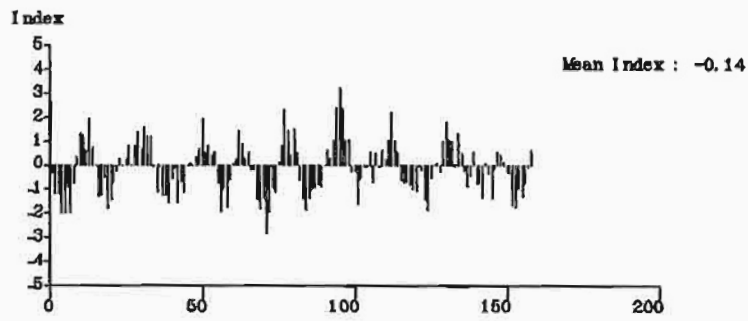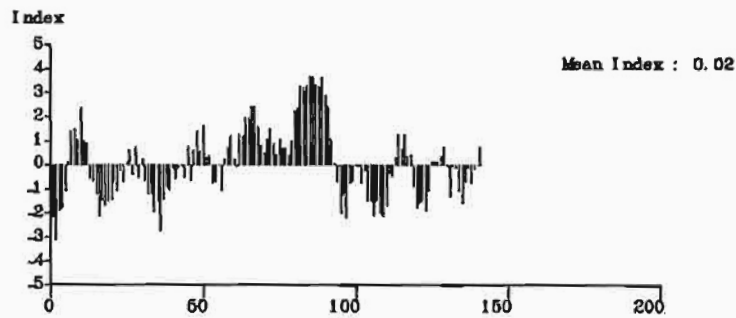
A

Index

5
4
3 Mean Index : 0.20
2
1
0
-1
-2
-3
-4
-5

0   50   100

Index

5
4
3 Mean Index : 0.28
2
1
0
-1
-2
-3
-4
-5

0   50   100

B

HELIX ———— Ū
SHEET .   V
TURN
COIL ————
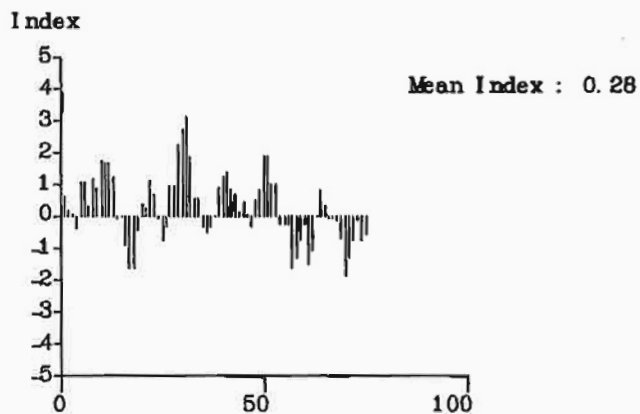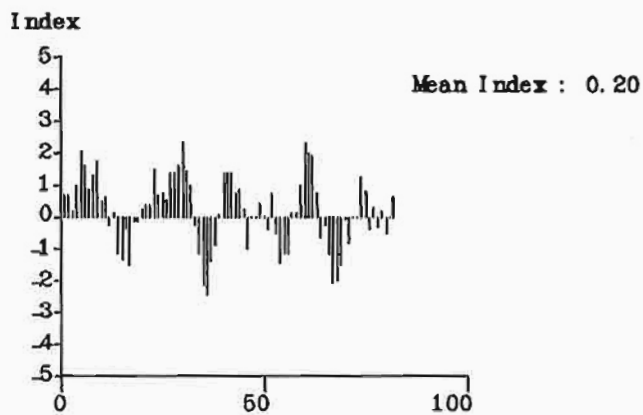
Figure 3.7: Analysis of the sequences of ORF3.2 and the hypothetical RelE protein of *E. coli.*

Panel A shows hydropathy plots for ORF3.2 (top) and RelE (bottom) determined according to Kyte and Doolittle (1982), using a window of six amino acids. Panel B shows secondary structure predictions for ORF3.2 (top) and RelE (bottom) determined using the algorithm of Chou and Fasman (1978).

A

Index

Mean Index : −0.53

Index

Mean Index : −0.36

B

HELIX
SHEET
TURN
COIL
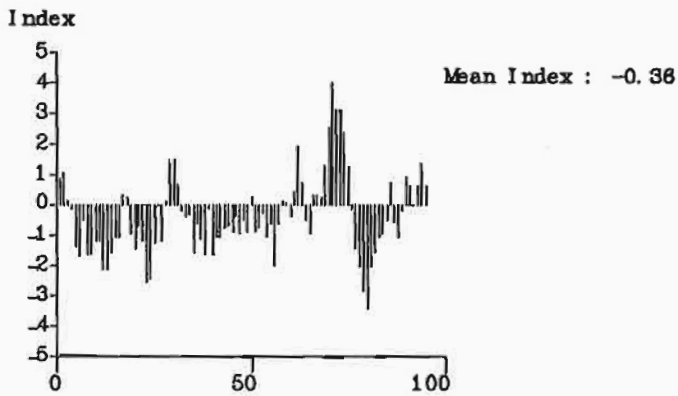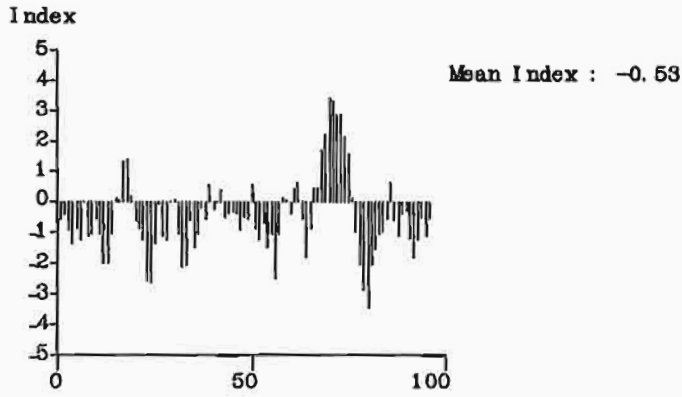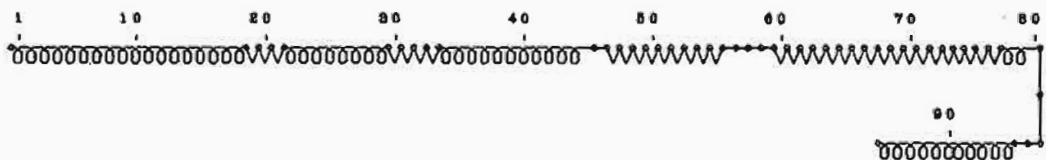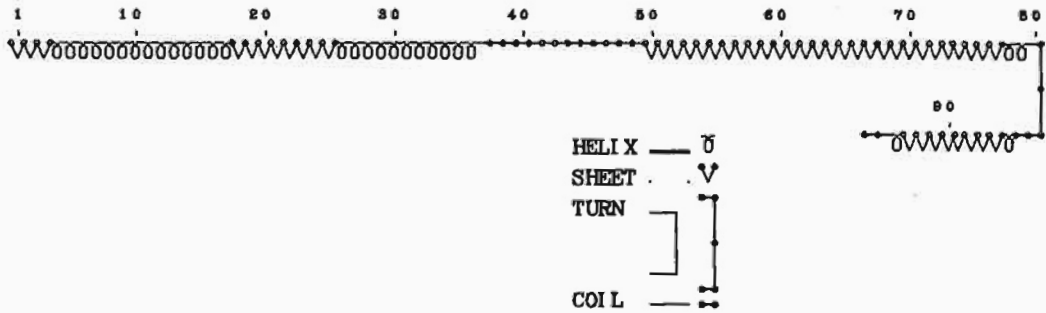
Figure 3.8: Hydropathy plots of other proteins encoded within pPM471.

Panel A shows a hydropathy plot for ORF1, Panel B shows a hydropathy plot for ORF6, Panel C shows a hydropathy plot for ORF7, Panel D shows a hydropathy plot for ORF8, and Panel E shows a hydropathy plot for ORF9, determined according to Kyte and Doolittle (1982), using a window of six amino acids.

A Index

Mean Index : -0.14

B Index

Mean Index : 0.02

C Index

Mean Index : -0.56

D Index

Mean Index : -0.09

E Index

Mean Index : 0.15
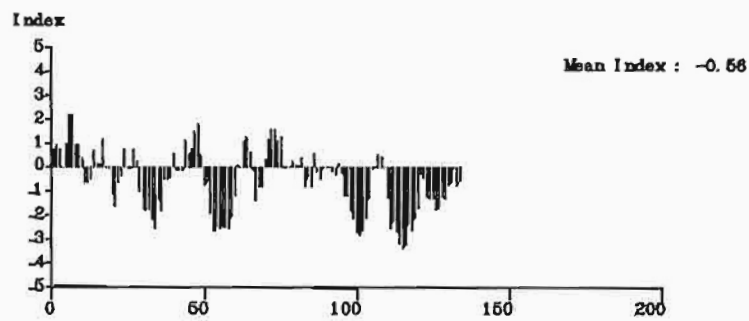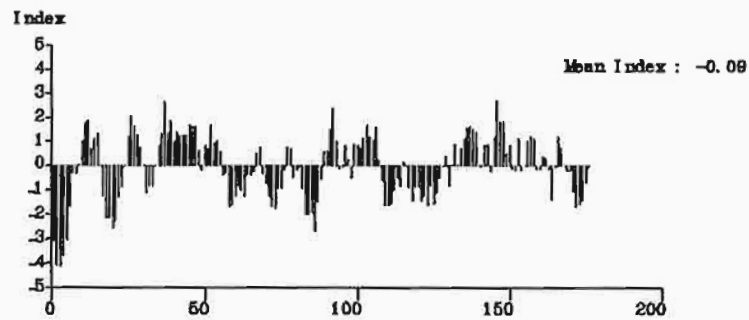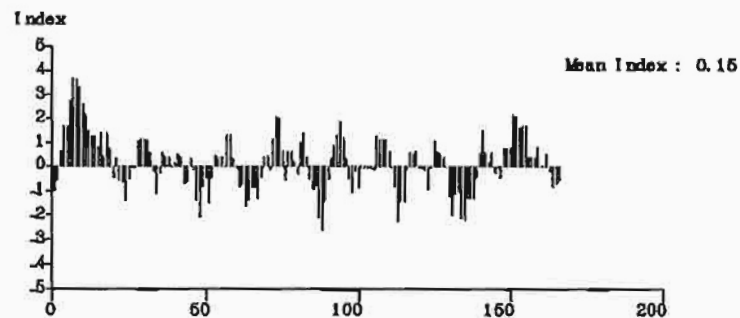
3.2.3.4 Partial ORFs

Both ends of pPM471 contain potential partial open reading frames, which overlap the *Bam*HI sites at the end of the clone (Fig 3.3). Sequence is available only for the carboxy terminal portion of the proximal ORF, and the amino terminal portion of the distal ORF (for which two possible start sites have been identified; Fig. 3.3). These partial sequences were compared with PIR and SWISS-PROT databases, as above, but did not yield any significant positive scores (not shown) and so were not considered further.

3.2.4 Attempts to engineer mutations in ORF3.1 and ORF3.2

In order to analyse further the functions of ORF3.1 and ORF3.2, attempts were made to engineer mutations in the two ORFs. Two strategies were used: i) insertion of the Km^R gene cartridge from pUC4-K (Vieira and Messing, 1982) in the unique *Bg*III site of pPM471 (between codon 58 and 59 of ORF3 1), which should disrupt ORF3.1 but not ORF3.2; and ii) linearising pPM471 with *Bg*III and digesting with Exonuclease III under conditions that should lead to small deletions only (Chapter 2). Despite several attempts, neither strategy was successful. The only Km^R Ap^R transformants obtained from the first strategy all had large deletions in the surrounding region, and all transformants screened from the second strategy had deletions of larger than 1 kb (not shown). This is despite both strategies working for other loci within pPM471 (Chapters 5 and 7, respectively). No further analyses were attempted.

3.3 Discussion

The nucleotide sequence of the entire insert of the MFRHA clone pPM471 has been determined and is presented here. This extends the sequence reported by van Dongen *et al.* (1987), to include eight more ORFs linked to the genes encoding the MFRHA. A BLAZE search (Brutlag *et al.*, in preparation) of the PIR and SWISS-PROT databases revealed homologous entries for: ORF2, ORF3.1, ORF3.2 and *mrhB*. Further analysis of ORF2 is presented in Chapter 5, while further analysis of *mrhA* and *mrhB* is presented in Chapter 7, so these are not discussed further here. The other ORFs (ORFs: 1, 6, 7, 8, and 9) do not show extensive similarity to known protein sequences. Analyses of the primary structure of these ORFs (Table 3.1; Fig. 3.8) have not revealed any clues as to their possible functions, or

cellular location, and will not be considered further in this study. The similarities shown by these ORFs make it seem unlikely that the MFRHA is fimbrial in nature, as the genes required for fimbrial biosynthesis so far characterised fall into one of a few well defined families (Hultgren *et al.*, 1993), none of which includes any ORF found on pPM471.

Comparison of the G+C% of these ORFs shows that they differ significantly from the G+C% of *V. cholerae* as a whole (Table 3.2). One interpretation of this observation is that it indicates a foreign origin for the genes (Reeves, 1993). However, such regions of G+C heterogeneity are known to occur in both prokaryotic and eukaryotic chromosomes, and other explanations for their occurrence are also possible (Sueoka, 1988). Comparison of the codon usage of ORFs encoded on pPM471 (Table 3.2) with those of available *V. cholerae* sequences (P.A. Manning, in preparation) has been performed here, but is difficult to evaluate for a number of reasons. The dataset that *V. cholerae* codon usage has been taken from may not accurately represent codon usage in *V. cholerae* as it almost exclusively contains virulence gene sequences: few sequences of housekeeping genes are available (P.A. Manning, in preparation). As cited above, most ORFs are too small for the numbers obtained to be statistically reliable and it is only by comparing the average codon usage for all ORFs together that any trends can be seen. In this case, only arginine codon usage seems to differ from that seen for *V. cholerae* as a whole, with less use of the common codons CGU and CGC, and more abundance of AGA (Table 3.2). So, analysis of codon usage does not preclude a foreign origin of these genes, but does not provide strong supporting evidence either.

ORF3.1 shows similarity to the amino terminal 72 amino acids of DnaT (primosomal protein i) of *E. coli* (Fig. 3.5A). Information regarding the function of domains of DnaT is not available, so it is not possible to make confident predictions as to what function, if any, the product of ORF3.1 might have. The overall hydrophobicity of the protein (Fig. 3.6) and its marked acidity (pI=4.19; Table 3.1) indicate that it is unlikely to interact directly with nucleic acid. However, as DnaT interacts with other members of the primosome, most notably DnaC, with which it is cotranscribed (Masai and Arai, 1988), the possibility of

ORF3.1 interacting with ORF3.2 should be considered, given their predicted translational coupling (Fig.3.3), and widely differing isoelectric points (Table 3.1).

ORF3.2 is similar to the hypothetical RelE protein of *E. coli* (Fig. 3.5B). This enables a confident prediction of the function of ORF3.2, namely that of a translational inhibitor. Attempts to obtain deletion mutants in ORF3.1 and ORF3.2 were unsuccessful, which was interpreted as indicating that the products of the desired constructs are lethal to the cell. The physiological role of these proteins remains unknown.

The other striking feature of the sequence as presented is the occurrence of the nine copies of the direct repeat VCR (Figures 3.3; 3.4), all of which are found in extragenic locations. Repetitive extragenic sequence elements are described in a number of bacterial genera, although no general physiological role has been assigned to them (Lupski and Weinstock, 1992). Nevertheless, further analysis of these elements has been undertaken, and is presented in the next Chapter.

Chapter 4

Mapping of VCR, a repeated sequence present in the *Vibrio cholerae* chromosome

4.1 Introduction

The occurrence of repetitive sequences in eubacterial chromosomes has been recognised since the characterisation of the REP (for Repetitive Extragenic Palindrome; Stern *et al.*, 1984) or PU (for Palindromic Unit; Gilson *et al.*, 1984) sequence in *E. coli* and *Salmonella typhimurium* (Higgins *et al.*, 1982). This sequence element is about 40 bp long and exhibits imperfect dyad symmetry, consisting of a highly conserved stem, and a variable loop region, along with a 4 bp tail that gives the element an orientation (Gilson *et al.*, 1984; Stern *et al.*, 1984; Gilson *et al.*, 1987a). REP sequences often occur in clusters called REP elements, in which up to four REP sequences can exist, but always in alternating orientation (Gilson *et al.*, 1987a; Higgins *et al.*, 1988). Physical mapping has shown that at least 90 such REP elements exist, randomly distributed throughout the *E. coli* chromosome, and that REP sequences constitute around 0.5 to 1% of the total genome (Dimri *et al.*, 1992). In addition, REP sequences are organised into larger units that also include conserved non-palindromic sequences that are called bacterial interspersed mosaic elements or BIME (Gilson *et al.*, 1991a; Gilson *et al.*, 1991b). The occurrence of REP sequences has now been identified in other members of the family Enterobacteriaceae (Versalovic *et al.*, 1991; Dimri *et al.*, 1992) and REP-related sequences have been detected in taxa as distant from the Enterobacteriaceae as *Deinococcus radiophilus*, *Herpetosiphon giganteus* and the archaebacterium *Halobacterium halobium* (Versalovic *et al.*, 1991). A number of other interspersed repetitive sequences have also been identified although only one of these, the IRU or ERIC sequence (Sharples and Lloyd, 1990; Hulton *et al.*, 1991) has been found in more than one family (Sharples and Lloyd, 1990; Hulton *et al.*, 1991; Versalovic *et al.*, 1991). Other repeat sequence elements for which nucleotide sequence is available include: a 192 bp repeat from *D. radiodurans* (Lennon *et al.*, 1991); an element with a 545 bp central domain and variable flanking domains from *Mycobacterium leprae* (Woods and Cole, 1990); a 129 bp element from *M. tuberculosis* (Eisenach *et al.*, 1990); an 87 bp repeat from *Myxococcus xanthus* (Fujitani *et al.*, 1991); and a 26 bp repeat from *Neisseria gonorrhoeae* and *N. meningitidis* (Correia *et al.*, 1986). Several repeats of around 400 bp in size have been identified in *Mycoplasma pneumoniae* (Su *et al.*, 1988; Wenzel and Herrmann, 1988; Colman *et al.*, 1990) but these appear to be remnants of genes found in the same region as the repeats (Colman *et al.*, 1990) so are not considered further here. A number of other microorganisms have also

been shown to possess interspersed repetitive sequences (reviewed in Lupski and Weinstock, 1992), but none of these sequences have been well characterised.

A number of possible functions have been suggested for repetitive sequences, most notably for REP sequences. Gilson *et al.*, (1987b) have suggested a role for REP in bacterial speciation, and subsequent work by Hofnung and coworkers has suggested that REP is restricted to a few members of the Enterobacteriaceae (Gilson *et al.*, 1990a). However, the validity of this assertion has been questioned by Versalovic *et al.*, (1991), who have detected REP-related sequences across the eubacterial kingdom. Subsets of REP sequences have been shown to have a variety of functions. These include: transcriptional terminators (Gilson *et al.*, 1986b); protectors of upstream RNA from digestion by 3'-5' exonuclease (Newbury *et al.*, 1987; Stern *et al.*, 1988); and inhibitors of translation of downstream cistrons (Stern *et al.*, 1988). However, none of these functions can explain the maintenance of sequence homology observed (Higgins *et al.*, 1988; Yang and Ames, 1988; Gilson *et al.*, 1990b), which must therefore be explained by some other function. To this end, REP has been shown to interact specifically with a number of proteins *in vitro* including: DNA gyrase (Yang and Ames, 1988) in a complex with histone-like protein HU (Yang and Ames, 1990); DNA polymerase I, and a protein complex (Gilson *et al.*, 1990b); and an unidentified nucleoid-associated protein (Gilson *et al.*, 1986a). REP also interacts with DNA gyrase *in vivo* (Yang, Chen and Ames, unpublished manuscript cited in Shyamala *et al.*, 1990). This has led to the suggestion of involvement of REP in nucleoid formation by providing a site on DNA where scaffold proteins might assemble (Yang and Ames, 1988; Yang and Ames, 1990; Gilson *et al.*, 1990b). For comparison, it is interesting to note that eukaryotic "matrix association regions" contain repetitive sequence elements that contain topoisomerase II (the eukaryotic homologue of bacterial DNA gyrase; Yang and Ames, 1988) binding sites (Blasquez *et al.*, 1989; Sperry *et al.*, 1989), although other distinct repetitive sequences are also involved in scaffold attachment that is independent of topoisomerase II (Neuer-Nitsche *et al.*, 1988). REP sequences have also been shown to be the end points of *recA* independent illegitimate recombination events in *S. typhimurium* (Shyamala *et al.*, 1990). These events were proposed to be mediated by DNA gyrase because of its demonstrated involvement in *recA*-independent illegitimate recombination events (Naito *et al.*, 1984) and because illegitimate recombination

events have been documented in eukaryotes that were shown to be associated with "matrix attachment regions" and that required topoisomerase II (Sperry *et al.*, 1989).

All of the above strongly suggests a role for REP sequences in nucleoid attachment. However a different explanation for the sequence conservation observed has been proposed, namely that REP sequences are selfish DNA (Higgins *et al.*, 1988), a proposal that seems to be based largely on the authors' inability to purify proteins that bind to REP. Higgins *et al.* (1988) cite gene conversion as a mechanism to explain sequence conservation. In the light of the discovery of reverse transcriptase in *E. coli* (Inouye and Inouye, 1991) and because many REP sequences are transcribed (Higgins *et al.*, 1988) this possibility needs to be considered seriously, regardless of the functions of REP, as reverse transcription has recently been shown to play a role in gene conversion in the budding yeast *Saccharomyces cerevisiae* (Derr and Strathern, 1993). Functions for the other repetitive sequences mentioned above have not been proposed, and in the absence of any recognisable similarities between them, the possibility that these sequences are selfish DNA remains attractive (Lupski and Weinstock, 1992).

Aspects of the arguments presented above may be relevant when virulence genes are considered. Repeat sequences associated with *pilE* and *pilS* loci in *N. gonorrhoeae* are responsible for the chromosomal rearrangements in antigenic variation (Haas and Meyer, 1986), albeit in a *recA*-dependent manner (Koomey *et al.*, 1987). Chromosomally-encoded virulence determinants in pathogenic *E. coli* are located in "pathogenicity islands" rather than being randomly distributed throughout the chromosome and these "islands" can be deleted, although no mechanism has been proposed to mediate this phenomenon (Hacker *et al.*, 1990). Also, the degree of DNA supercoiling can affect the expression of virulence genes, acting in response to environmental signals such as changes in osmolarity or temperature, but the mechanism by which this occurs is far from clear (Dorman, 1991), and there has been no suggestion of a link between supercoiling-dependent gene expression and REP sequences.

In Chapter 3, a previously described direct repeat (van Dongen *et al.*, 1987), designated VCR, was shown to exist in multiple copies linked to the *mrh* locus in *V. cholerae* O1. Further analysis of this repeat has been undertaken, including: copy number and linkage

in *V. cholerae* O1; distribution in related species and genera; possible interactions between VCR and DNA gyrase; and possible involvement of VCR in previously described deletion events, and is described in this Chapter.

4.2 Results

4.2.1. Sequence comparisons of VCR

The most striking feature of the nucleotide sequence presented in Chapter 3 is the occurrence of nine 124 bp repeats, all of which are in the same orientation relative to one another, and all of which are found only in an extragenic location (Chapter 3). The sequences of these repeats have been aligned using the CLUSTAL algorithm (Higgins and Sharp, 1989) and this alignment, along with a consensus sequence based on the alignment, is shown in Fig. 4.1. The individual copies of VCR show an overall similarity of about 92%. Several stretches of dyad symmetry can be identified within the repeats, and the whole repeat itself shows imperfect dyad symmetry which, if transcribed, would give a stable secondary structure. In the case of the consensus sequence, a loop with a free energy of -58.9 kCal.mol^{-1} is predicted (Fig. 4.2).

4.2.2. *V. cholerae* O1 contains multiple copies of VCR

The number of copies of VCR found within pPM471 raises the possibility of the occurrence of further copies of VCR in the chromosome of *V. cholerae*. Fig. 4.3 shows a Southern blot of *Bam*HI-digested chromosomal DNA from 10 different *V. cholerae* O1 strains of both biotypes probed with oligo #128. Given that the probe used will bind to six of the nine VCR sequences of pPM471 under the conditions used (Chapter 3), an estimate of the number of VCRs in the chromosome can be obtained by quantitating the density of the bands in Fig. 4.3A, as shown in Fig. 4.3B. This gives an estimate of 90 to 100 copies of VCR in strains of the classical biotype and 60 to 80 copies of VCR in strains of the El Tor biotype (but see below). To demonstrate that the bands observed are not due to partial digestion, the filter was stripped and re-probed using the 0.7 kb *Cla*I-*Xba*I fragment of pPM1127, which corresponds to *mrhB* (Franzon and Manning, 1986), and the result is shown in Fig. 4.4. All strains tested show a single hybridising band, of 6.3 kb in size, with the exception of C31, in which this gene is deleted (van Dongen and DeGraaf, 1986; see below)

Figure 4.1: Alignment of thirteen sequenced copies of VCR.

The sequences VCR1 to VCR9 are the nine copies of VCR whose sequence is presented in Chapter 3. VCR21 and VCR22 are from van Dongen *et al.* (1987), and VCR 31 and VCR32 are from Takeda *et al.* (1991b). A consensus sequence based on the alignment is also shown. Nucleotides identical to the corresponding base in the consensus sequence are in bold type and nucleotides identical in all copies presented in the alignment are marked with an asterisk. The alignment was performed using the multiple alignment program CLUSTAL (Higgins and Sharp, 1989).

```
VCR1      TAACAAACGCTTCAAGAGGGACAGCCAACG-CGTGGCATTTTTATTAT-GCGTTGGTTTTTGTG
VCR2      TAACAAACGCCTCAAGAGGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
VCR3      TAACAAACGCCTCAAGAGGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
VCR4      TAACAAACGCCTCAAGAGGGACTGTCAACG-CGCGGCGTTTCCAGTCC-CAATGAGCCGCGGTG
VCR5      TAACAAACGCCTCAAGAGGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
VCR6      TAACAAACGCCTCAAGAAGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
VCR7      TAACAAACGCCTCAAGCGGGACTGTCAACG-CGCGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
VCR8      TAACAAACGCCTCAAGAGGGACTGTCAACGTCGTAGCGTTTCCAGTCC-CAATGAGCCGCGGTG
VCR9      TAACAAAGCGTTAAAGCGGGACTTGGCACG-CGTGGCATTTTCAGTTTGCGGTAAGTT-TAGTG
VCR21     TAACAAACGCCTCAAGAGGGACTGTCAACG-CGCGGC-TTTCCAGTCC-CAATGAGCCGCGGTG
VCR22     TAACAAACGCCTCAAGAGGGACTGTCAACG-CGCGGC-TTTCCAGTCC-CAATGAGCAACGGTG
VCR31     TAACAAACGCCTCAAGAGGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
VCR32     TAACAAACGCCTCAAGAGGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CAATGAGCCGCGGTG
          *******    *  *** ****    *** **  ** ***  * *      *   *    ***

Consensus TAACAAACGCCTCAAGAGGGACTGTCAACG-CGTGGCGTTTCCAGTCC-CATTGAGCCGCGGTG
               10        20        30        40        50        60


VCR1      ATTATGG--TGTTATGCGTTAAGTTAGTAGTA--GCGTTGGCT-GCCCCTTAAGCGGGCGTTAT
VCR2      GTTGCTGT-TGTTGTGTTTGAGTTTAGTGGTAGTGCGTT-GTCAGCCCCTTAGGCGGGCGTTAG
VCR3      GTTTCGGT-TGTTGTGTTTGAGTTTAGTGGTAATACGTT-GCCAGCCCCTTAGGCGGGCGTTAA
VCR4      GTTGCAGT-TGTTGTGTTTGGGTTTAGTG-TTATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR5      GTTGCAGT-TGTTGTGTTTGAGTTTAGTGTTAATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR6      GTTTCGGT-TGTTGTGTTTGAGTTTAGTGGT-GTGCGTTTGTCATCCCCTTAGGCGGGCGTTAG
VCR7      GTTTCGGT-TGTTGTGTTTGAGTTTAGTGGTAATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR8      GTTTTGGT-TGTTGTGTTTGGGTT-AGTGGTAATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR9      -TTAAGGTGTAATGCGGTC-ACTTTTGTA-TTACGTGCCTGC--GCCCCTTAACGCGGCGTTAG
VCR21     GTTGCAGT-TGTTGTGTTTGGGTTTAGTG-TTATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR22     GTTACAGT-TGTTGTGTTTGAGTTTAGTG-TAATACGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR31     GTTACAGT-TGTTGTGTTTGAGTTTGGTGGTAATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
VCR32     GTTATGGC-TGTTGTGTTTGAGTTTAGTG-TTATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
          **   *   *   *   *        *  **  *     *   *   ******* *******

Consensus GTTTCGGT-TGTTGTGTTTGAGTTTAGTGGTAATGCGTT-GCCAGCCCCTTAGGCGGGCGTTAG
               70        80        90       100       110       120
```

Figure 4.2: Predicted secondary structure for the transcribed consensus sequence of VCR.

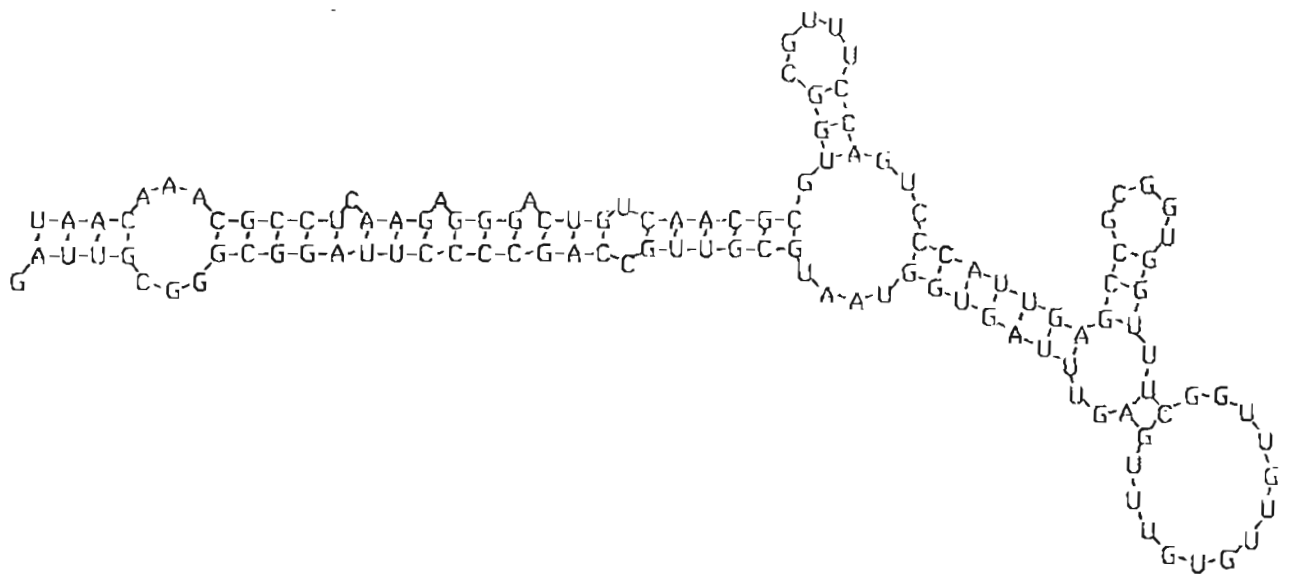This structure has a free energy value of -58.9 kCal.Mol^{-1}.

Figure 4.3: Occurrence of VCR in the *V. cholerae* O1 chromosome.

Panel A shows a Southern blot of chromosomal DNAs from ten *V. cholerae* O1 strains digested with *Bam*HI and probed with oligo #128. The band corresponding to the 6.3 kb *Bam*HI fragment of pPM471 is indicated with an arrow. Panel B shows the integer values of band intensities shown in panel A, as determined from analysis with a Molecular Dynamics model 300A computing densitometer, assuming that the 6.3 kb *Bam*HI band contains six copies of sequences that hybridise with the probe, which is underlined to indicate this assumption. Lanes are DNA from the strains: 1, 569B; 2, CA401; 3, CA411; 4, AA14041; 5, Z17561; 6, O17; 7, C5; 8, C31; 9, AA14073; 10, BM69.

[a] Values were not obtained for O17 or C5 due to excessive smeariness of the sample DNAs.

[b] Values were not obtained for C31 due to the absence of a band of known sequence and hence VCR copy number status.

-

B	1	2	3	4	5	6	7	8	9	10
	56	53	51	54	44				22	21
		22	20	28	27				32	49
	13					ND[a]	ND[a]	ND[b]		
	12	13	14	13	13					
	6	6	6	6	6				6	6
	6	5	5	5	5					
	1	1	1	1	1				1	1
	1	1	1	1	1					

Figure 4.4: Southern blot of *V. cholerae* O1 chromosomal DNAs digested with *Bam*HI and probed with a *mrhB* specific probe.

The filter is the same filter shown in Figure 4.3A, after stripping and re-probing with a probe derived from the 0.7 kb *Cla*I-*Xha*I fragment of pPM1127 (Franzon and Manning, 1986). Lanes are DNA from the strains: 1, 569B; 2, CA401; 3, CA411; 4, AA14041; 5, Z17561; 6, O17; 7, C5; 8, C31; 9, AA14073; 10, BM69.

1　2　3　4　5　6　7　8　9　10

Fig. 4.5A shows the same chromosomal DNAs digested with *Hind*III and probed with oligo #128. In all cases, a large number of bands are visible. Again, in order to demonstrate that the bands observed are not due to partial digestion, the filter was stripped and re-probed using the 0.7 kb *Cla*I-*Xba*I fragment of pPM1127 (Franzon and Manning, 1986) which is shown in Fig. 4.5B. A single hybridising band of 2.4 kb in size is seen in all cases, with the exception of C31 (as above).

As the oligo probe used in Figures 4.3 and 4.5A only binds to 6 out of the nine copies of VCR present on the 6.3 kb *Bam*HI fragment (Chapter 3), the possibility that significant numbers of VCR had been missed in the analysis shown above was investigated by stripping the filters shown in Figures 4.4 and 4.5B and re-probing them using the 6.3 kb *Bam*HI fragment of pPM471 (Franzon and Manning, 1986), and by probing a filter similar to that used in Fig. 4.3A with the 0.25 kb *Xba*I-*Nsi*I fragment of pPM1631 (Chapter 6); all of these hybridisations are shown in Fig. 4.6. The 6.3 kb probe hybridises to the same number of bands as oligo #128, as does the 0.25 kb *Xba*I-*Nsi*I probe, albeit with differing intensities, which indicates that the estimate obtained above for the total number of copies of VCR is not wildly inaccurate due to "missed" copies of VCR in Fig. 4.3.

4.2.3 Distribution of VCR on the *V. cholerae* chromosome

The pattern of hybridisation observed shows remarkable similarity to the pattern observed when a number of MFRHA positive cosmid clones were probed with the 2.4 kb *Hind*III fragment indicated in Fig.4.5 (Franzon, 1988). As this fragment contains three copies of VCR (Chapter 3), the possibility that the copies of VCR observed in the chromosome of *V. cholerae* are linked was investigated by analysis of large restriction fragments separated on a pulsed field gel.

4.2.3.1. Extracting DNA from *V. cholerae* for large-scale mapping

The initial approaches used to extract DNA for large scale mapping centred on those published by Smith and Cantor (1987), Smith *et al.* (1987) and Overhauser and Radic (1987), as detailed in Chapter 2. However, when DNA isolated by these methods was digested with *Not*I and run on a pulsed-field gel, excessive smearing was observed (Fig. 4.7A), which was

Figure 4.5: Southern blot of *Hind*III digested *V. cholerae* chromosomal DNAs.

Panel A shows a filter probed with oligo #128. Panel B shows the same filter after stripping and re-probing with a probe derived from the 0.7 kb *Cla*I-*Xba*I fragment of pPM1127 (Franzon and Manning, 1986). The corresponding bands between the two panels are indicated with arrows. Lanes are DNA from the strains: 1, 569B; 2, CA401; 3, CA411; 4, AA14041; 5, Z17561; 6, O17; 7, C5; 8, C31; 9, AA14073; 10, BM69.
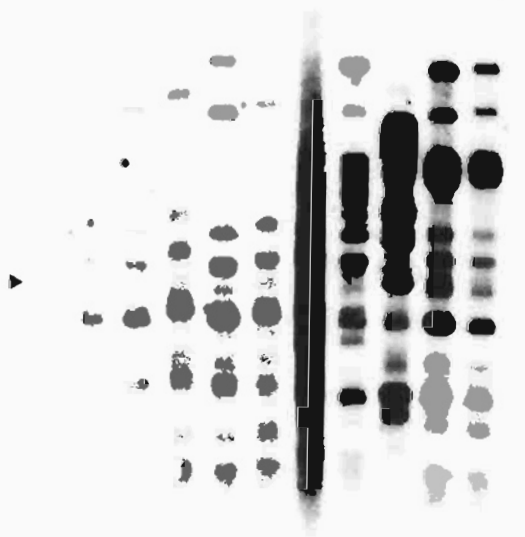
Figure 4.6: Southern blot of *V. cholerae* chromosomal DNAs.

Panel A shows *Bam*HI digested DNAs hybridised with a probe derived from the 6.3 kb *Bam*HI fragment of pPM471. Panel B shows *Hind*III digested DNAs hybridised with the same probe as that used in panel A. Panel C shows *Bam*HI digested DNAs hybridised with a probe derived from the 0.2 kb *Xba*I-*Nsi*I fragment of pPM1631 (Chapter 6). The 6.3 kb *Bam*HI fragment and the 2.4 kb *Hind*III fragment (as appropriate) that hybridises with the *mrhB*-specific probe used in Fig. 4.4 and Fig. 4.5B are indicated with arrows. Lanes are DNA from the strains: 1, 569B; 2, CA401; 3, CA411; 4, AA14041; 5, Z17561; 6, O17; 7, C5; 8, C31; 9, AA14073; 10, BM69.
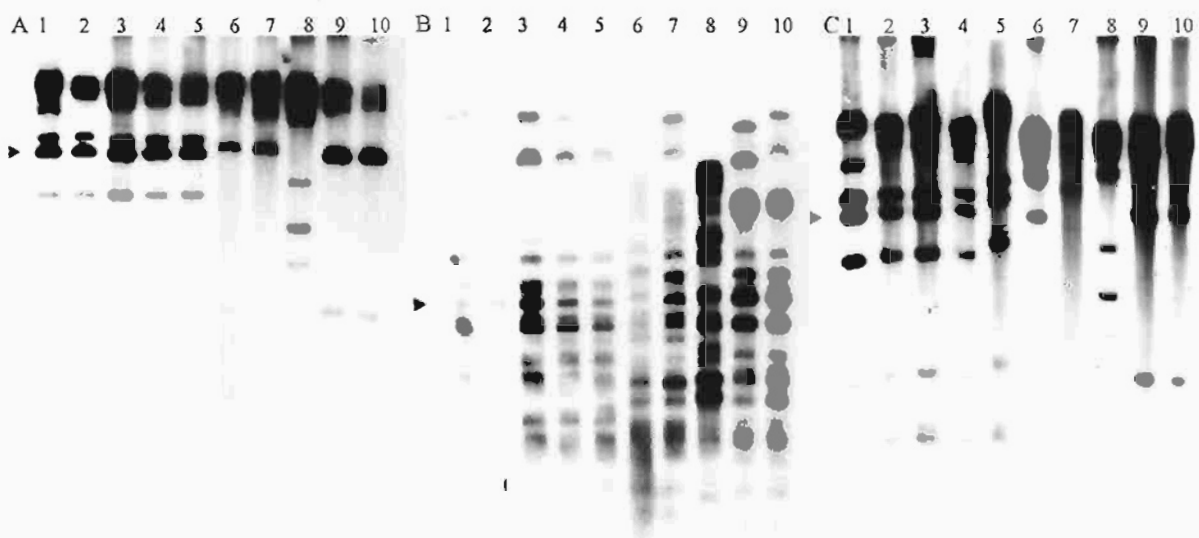
Figure 4.7: Determination of suitable conditions for preparation of samples for pulsed field gel analysis from *V. cholerae*.

Panel A shows a 1.5% agarose gel subjected to the run conditions: initial A time = 50"; final A time = 100"; and 200 V for 21 hours. Panel B shows a 1% agarose gel subjected to the run conditions: initial A time = 1"; final A time = 10"; and 200 V for 18 hours. Lanes are: 1, yeast chromosomes; 2, DNA prepared from 569B essentially according to Smith and Cantor (1987) and Smith *et al.* (1987) and digested with *Not*I; 3, DNA prepared from 569B as for lane 2 except that samples were prepared from agarose beads (Overhauser and Radic, 1987); 4 and 5, DNA prepared from C31 as described in Chapter 2 and digested with *Not*I, or *Sfi*I, respectively.
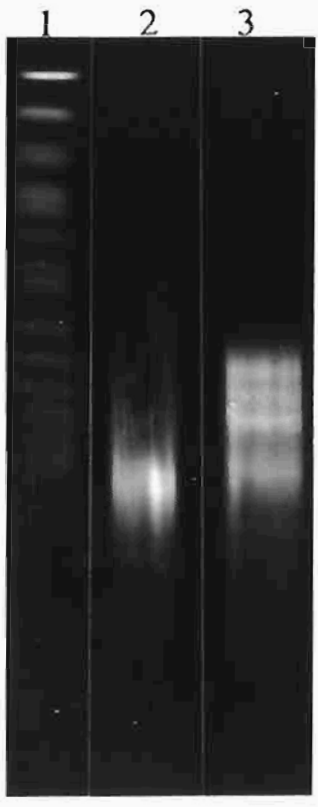
presumably caused by the *V. cholerae* secreted DNases (Focareta and Manning, 1987; Focareta and Manning, 1991), as encountered previously with conventional methods of chromosomal DNA preparation (Manning *et al.*, 1986). The smearing seen in Fig. 4.7A is worse for cells set in agarose blocks, rather than beads, presumably due to longer times for diffusion of protease to the middle of the block as opposed to the middle of the beads (*c.f.* Overhauser and Radic, 1987), and so setting cells in beads was chosen as the standard condition. Omission of lysozyme treatment of beads and Proteinase K digestion of the cells in 0.5 M EDTA pH 9.2 was found to rectify the problem of smearing (Fig. 4.7B) and so these were included in the final method. The preparation conditions used subsequently are detailed in Chapter 2.

4.2.3.2. Mapping copies of VCR

A pulsed field gel of *V. cholerae* chromosomal DNAs cut with *NotI* or *SfiI* was probed with oligo #128 (Fig. 4.8A). If VCRs were randomly distributed in the chromosome, then a multitude of hybridising bands would be observed. However, this is not the case. Two hybridising bands are observed for 569B with sizes of 285 and 195 kb for *NotI* digested DNA, and of 265 and 205 kb for *SfiI* digested DNA. For O17, two hybridising species are observed with *NotI* digested DNA, of 265 and 125 kb, and a single band of 300 kb is observed with *SfiI* digested DNA. Finally, C31 shows single bands of 245 and 235 kb with *NotI* and *SfiI* digested DNA, respectively (Fig. 4.8A). This filter was stripped and reprobed with the 0.7 kb *ClaI-XbaI* fragment of pPM1127, which hybridises with: two fragments of 285 and 195 kb in 569B *NotI* digested DNA; two fragments of 265 and 205 kb in 569B *SfiI* digested DNA; and single fragments of 265 and 300 kb in O17 *NotI* and *SfiI* digested DNA, respectively. As expected, no hybridising fragments were detected with C31 (Fig. 4.8B). There is ample evidence to suggest that *mrhB* exists as a single copy gene in 569B (Franzon, 1988; Figures 4.4, 4.5B) so the larger of the two bands detected in 569B *NotI* and *SfiI* cut, is most probably due to partial digestion. To ensure that most copies of VCR are being detected in this analysis (c.f. Fig. 4.6) the filter was again stripped and reprobed with the 0.25 kb *XbaI-NsiI* fragment of pPM1631. A pattern of hybridisation indistinguishable to that obtained with oligo #128 was obtained (Fig. 4.8C), indicating that this is indeed the case.

Figure 4.8: Southern blots of pulsed field gels.

Panel A shows a Southern blot of a pulsed field gel probed with oligo #128. Panel B shows the same filter after stripping and re-probing with a probe derived from the 0.7 kb *ClaI-XbaI* fragment of pPM1127 (Franzon and Manning, 1986). Panel C shows the same filter again, after stripping and re-probing with a probe derived from the 0.2 kb *XbaI-NsiI* fragment of pPM1631 (Chapter 6). Gel run conditions were the same as those described for Figure 4.7B. Lanes are DNA from: 1, 569B, digested with *NotI*; 2, 569B, digested with *SfiI*; 3, O17, digested with *NotI*; 4, O17, digested with *SfiI*; 5, C31, digested with *NotI*; 6, C31, digested with *SfiI*.

A　1　2　3　4　5　6　　B　1　2　3　4　5　6　　C　1　2　3　4　5　6

4.2.4 Specificity of VCR to *V. cholerae*

To see whether VCR-related sequences could be detected in other members of the family Vibrionaceae, a number of different species from the genera *Vibrio* and *Aeromonas* were probed with oligo #128 (Fig. 4.9). Hybridisation was observed in a number of non-O1 *V. cholerae* (Fig. 4.9A), but no hybridisation was observed for the other *Vibrio* sp. or *Aeromonas* sp. tested (Fig. 4.9B). In order to test whether sequences related to *mrhB* were present in any of these other strains tested, the filters were stripped and reprobed with the 0.7 kb *ClaI-XbaI* fragment of pPM1127 (Fig. 4.10). However, a positive signal was obtained only for *V. cholerae* O1 strains 569B and C21 (Fig. 4.10A), all other strains tested were negative.

4.2.5 Similarities between VCR and other repeats

REP sequences have been shown to be the endpoints of *recA*-independent illegitimate recombination events in *S. typhimurium*, a property that is attributed to their interacting with DNA gyrase (Shyamala *et al.*, 1990). While a direct involvement of VCR in similar events has not been shown, plasmids and M13 phage that with hindsight contain multiple copies of VCR show extreme instability in a variety of *rec*$^+$, *gyr*$^+$, *E. coli* strains (Chapter 3; Franzon, 1988; F.K. DeGraaf, personal communication). Propagation of plasmids in *recA* strains has led to a reduction of this problem (not shown). Homologous recombination in *E. coli* requires a minimum of 40 to 50 bp of terminal homology (Smith, 1988), making *recA*-dependent recombination a possible mediator of this instability. However, DNA gyrase is capable of mediating illegitimate recombination events in the absence of more than 4 bp of terminal homology (Naito *et al.*, 1984; Reece and Maxwell, 1991), and is possibly involved in homologous recombination (Reece and Maxwell, 1991), so either or both genes could be involved in the observed phenotype. RecA shows no sequence specificity in its interactions with DNA (Smith, 1988), but DNA gyrase binds with high affinity to a recognisable sequence (Yang and Ames, 1988; Condemine and Smith, 1990). This allows the two possibilities to be distinguished by comparing the sequences in question. Therefore, REP, VCR and other repetitive sequences discussed above, were compared (pairwise), using the FASTA algorithm (Pearson, 1990). No strong, overall similarity could be detected between any two repeats, but small blocks of similarity exist between VCR and REP, and this alignment is shown in Fig.

Figure 4.9: Southern blot of other bacterial chromosomal DNAs probed with a VCR-specific probe.

Panel A shows a Southern blot of DNAs isolated from the listed O1 and non-O1 *V. cholerae* digested with *Bam*HI. Lanes are: 1, 569B; 2, C21; 3, NCV165; 4, BV7; 5, BV22; 6, BV41; 7, N34; 8, N41; 9, N125; 10, N50. Panel B shows a Southern blot of DNA isolated from the listed *Aeromonas*, *V. parahaemolyticus*, *V. mimicus* and *V. fluvialis* strains digested with *Bam*HI. *V. cholerae* 569B is included by way of comparison. Lanes are: 1, *V. cholerae* 569B; 2, *A. hydrophila* A006; 3, *Aeromonas* sp. AB1; 4, *A. sobria* A187; 5, *A. sobria* A191; 6, *A. caviae* V14; 7, *V. parahaemolyticus* NCTC10884; 8, *V. parahaemolyticus* NCTC10885; 9, *V. mimicus*, V800; 10, *V. fluvialis* V564. In both cases the probe is oligo #128. The 6.3 kb *Bam*HI band of *V. cholerae* O1 that hybridises with a *mrhB*-specific probe is indicated with an arrow.

Figure 4.10: Southern blot of other bacterial chromosomal DNAs probed with a *mrhB*-specific probe.

Panel A shows a Southern blot of DNAs isolated from the listed O1 and non-O1 *V. cholerae* digested with *Bam*HI. Lanes are: 1, 569B; 2, C21; 3, NCV165; 4, BV7; 5, BV22; 6, BV41; 7, N34; 8, N41; 9, N125; 10, N50. Panel B shows a Southern blot of DNA isolated from the listed *Aeromonas*, *V. parahaemolyticus*, *V. mimicus* and *V. fluvialis* strains digested with *Bam*HI. *V. cholerae* 569B is included by way of comparison. Lanes are: 1, *V. cholerae* 569B; 2, *A. hydrophila* A006; 3, *Aeromonas* sp. AB1; 4, *A. sobria* A187; 5, *A. sobria* A191; 6, *A. caviae* V14; 7, *V. parahaemolyticus* NCTC10884; 8, *V. parahaemolyticus* NCTC10885; 9, *V. mimicus*, V800; 10, *V. fluvialis* V564. In both cases the probe is a probe derived from the 0.7 kb *Cla*I-*Xba*I fragment of pPM1127 (Franzon and Manning, 1986). The 6.3 kb *Bam*HI band of *V. cholerae* O1 that hybridises with a *mrhB*-specific probe is indicated with an arrow.

A   1 2 3 4 5 6 7 8 9 10   B   1 2 3 4 5 6 7 8 9 10

4.11A. Furthermore, the region of REP identified in this analysis overlaps with a region that shows 70% identity with a strong binding site for DNA gyrase found in pBR322 (included in Fig. 4.11A by way of illustration; Yang and Ames, 1988) and a multiple alignment between REP, VCR and the DNA gyrase binding site is shown in Fig. 4.11B. The similarity between these sequences is not conclusive, but raises the possibility that DNA gyrase reacts with VCR, as has been demonstrated for REP (Yang and Ames, 1988; Yang and Ames, 1990). DNA polymerase I has also been shown to bind specifically to REP. The recognition sequence identified is not highly conserved, beyond being G+C rich (Gilson *et al.*, 1990b), and although G+C rich runs can be identified in other repeats examined (not shown) this cannot be taken as evidence for any interaction with DNA polymerase I: direct experimental evidence is required before any assertion can be made.

4.2.6 Does VCR interact with DNA gyrase?

In order to test the prediction that DNA gyrase might interact with VCR, attempts were made to isolate *gyr* mutants in *V. cholerae* by virtue of their resistance to gyrase inhibitors. In *E. coli*, *gyrA* mutants have been isolated by virtue of their resistance to the quinolone group of drugs such as nalidixic acid (Gellert *et al.*, 1977), while *gyrB* mutants have been isolated due to their resistance to the coumarins, such as coumermycin A (Contreras and Maxwell, 1992). Therefore, conditions similar to those used for *E. coli* were used in numerous attempts to obtain Nal^R mutants of *V. cholerae* Z17561 and O17. No mutants were obtained, even after treatment with the mutagen EMS (Lin *et al.*, 1962). However, Nal^R mutants were readily obtained from C31. The reason for this strain difference is not known (see also Chapter 8).

Two Nal^R variants of strain C31 were isolated and were indistinguishable from the parental strain in slide agglutination, haemolytic activity, phage sensitivity and biochemical tests (not shown). However, both exhibited drastically reduced growth rates *in vitro*, and one had lost the ability to grow on minimal medium (not shown). Two experimental approaches were followed to confirm that these Nal^R mutants were *gyr* mutants. Numerous attempts to introduce reporter plasmids into these strains, to enable analysis of plasmid superhelicity on chloroquine gels, was unsuccessful, due either to an inability to transform these strains (with

Figure 4.11: Sequence similarities between VCR, REP and a DNA gyrase binding site from plasmid pBR322.

Panel A shows pairwise alignments in all possible combinations between the three sequences. Numbers annotating the sequence are: for VCR, from the left end of the consensus (Figure 4.1); for REP and pBR322, according to Yang and Ames (1988). Identical nucleotides are indicated by a vertical stroke. The alignments were performed using the FASTA algorithm (Pearson, 1990). Panel B shows a multiple alignment of the three sequences generated with CLUSTAL (Higgins and Sharp, 1989). Nucleotides identical in all sequences are indicated with an asterisk.

A

```
                          990          1000
Gyrase (pBR322)   CTGGAT--GGCCTTCCCCATTAT
                  ||| ||  ||| | | || ||||
REP consensus     CGT-ATGCGGCGTGCGCC-TTAT
                          10           20


                          100          110
RVC consensus     GTAATGCGTTGCCAGCCCCTTAG
                  | ||||| || || |||||
REP consensus     CTGATGCG--GCGTGCGCCTTAT
                          10           20


                          100          110
RVC consensus     GTAATGCGTTGCCAGCCCCTTAG
                  |        |    | ||| |||
Gyrase (pBR322)   CTGGATGGCCTTC--CCCATTAT
                          990          1000
```

B

```
Gyrase      CTGGATGGCCTTC--CCCA-TTAT
REP con.    CTGATGCGGCGTGCGCC---TTAT
VCR con.    GTAATGCGTTGCCAGCCCC-TTAG
ERIC con.   CTG-CGTTCGCTCACCCCAGTCAC
            *              **    * *
```

pACYC184 based replicons) or extreme instability of the introduced plasmids (with pGEM3Zf$^+$ based replicons) (not shown). The second approach was to obtain crude extracts from C31 and the Nal^R variants, and to assay these for DNA gyrase activity in an *in vitro* gyrase assay. Despite several different approaches, a vigorous contaminating DNase activity could not be separated from fractions predicted to contain DNA gyrase. Therefore, it was not possible to confirm the nature of the Nal^R mutation in these strains.

4.2.7 VCR as deletion endpoints

Van Dongen and DeGraaf (1986) identified a strain, C31, in which the entire 6.3 kb *Bam*HI region associated with MFRHA activity seemed to be deleted. However, weak cross-hybridisation seen on their blot and numerous cross-hybridising bands detected with various probes used above raised the possibility that the deletion they observed was in fact contained within the MFRHA$^!$ 6.3 kb *Bam*III fragment. So, in an attempt to define this deletion, *Bam*HI and *Hind*III digested chromosomal DNAs from 569B and C31 were probed with M13 clones M13/1601 and M13/1613, which correspond to the proximal and distal ends ends of plasmid pPM471, respectively (Fig. 3.2; Fig. 4.12). The distal probe hybridises to a single fragment for 569B (a 6.3 kb band for *Bam*HI digested DNA and a 6.5 kb band for *Hind*III digested DNA) and no fragment for C31 (Fig. 4.12B), but the proximal fragment hybridises to two fragments for 569B, of 11 and 6.3 kb for *Bam*HI digested DNA and 2.8 and 1.2 kb for *Hind*III digested DNA, and one fragment for C31, of 10 kb and 5.4 kb for *Bam*HI and *Hind*III digested DNA, respectively (Fig 4.12A). A similar result was obtained with other *V. cholerae* strains (*c.f.* Figures 4.3, 4.4, and 4.6; not shown). Cross-hybridising fragments were also observed when 569B and C31 *Bam*HI digested DNAs were probed with the 0.35 kb *Eco*RV-*Not*I fragment of pPM1648 (Chapter 5), as shown in Fig. 4.12C. In this case, fragments of 6.3 kb and 4.5 kb are observed with *Bam*HI digested 569B DNA, and a fragment of 11.5 kb is observed with *Bam*HI digested C31 DNA. Bands were not apparent with *Hind*III digested DNA in this case. This raises the possibility that some sequences contained within pPM471 have been duplicated, but also that the deletion in C31 is larger than the cloned DNA. Therefore, mapping the deletion end-points was not pursued further.

Figure 4.12: Southern blotting to determine the end points of the deletion in strain C31.

Panel A shows a filter hybridised with a probe derived from M13/1601, representing the proximal end of the insert in pPM471 (Chapter 3). Panel B shows a filter hybridised with a probe derived from M13/1613, representing the distal end of the insert in pPM471 (Chapter 3). Panel C shows a filter probed with the 0.3 kb *EcoRV-NotI* fragment of pPM1648, specific for ORF2 (Chapter 5). The 6.3 kb *Bam*HI band of *V. cholerae* O1 that hybridises with a *mrhB*-specific probe is indicated with an arrow. Lanes are: 1, 569B, *Bam*HI digested; 2, 569B, *Hind*III digested; 3, C31, *Bam*HI digested; 4, C31, *Hind*III digested.
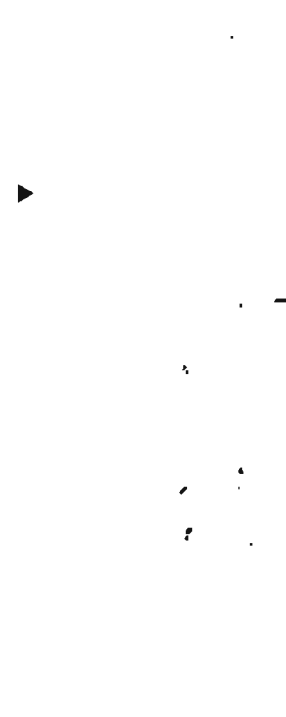
4.3 Discussion

Van Dongen *et al.* (1987) identified two copies of a repeat sequence that in this study has been shown to be present in multiple copies in the chromosome of *V. cholerae* O1. An estimate of 60 to 100 copies of VCR in *V. cholerae* O1 was obtained, but this is an underestimation as the estimate assumes: that all fragments of a chromosomal digest will transfer with equal efficiency, regardless of size, under the conditions used; that the oligo probe used will bind to all copies of VCR; and that the X-ray film quantitated has not reached saturation for the most intense bands. It is known that the oligonucleotide binds to only six of the nine copies of VCR present on pPM471 at the degree of stringency used (Chapter 3). However, the VCR fragment probe (random primed) that is capable of binding to all copies of VCR, does not reveal any major new bands when used to probe a filter identical to that probed with the oligo probe.

VCR shows several regions of dyad symmetry, and the whole repeat shows imperfect symmetry, that if transcribed would form a stable secondary structure. This raises the possibility that VCR might play a role in transcription termination in *V. cholerae*. Transcription has been shown to terminate within at least one of the repeats but it seems unlikely that all copies of VCR act as transcriptional terminators, as RNase protection experiments using VCR specific probes of either orientation reveal multiple bands consistent with the presence of full-length copies of VCR within *V. cholerae* total RNA (Chapter 6). Furthermore, the repeat upstream of *mrhA* is unlikely to act as a terminator, due to the polarity of the flanking ORFs (Chapter 3). Similarly, only a subset of REP sequences have been shown to act as transcriptional terminators (Gilson *et al.*, 1986b). Therefore, some other role must lead to their maintenance.

Repetitive extragenic sequences have been characterised from a number of bacterial genera, but do not show any recognisable sequence similarities (Lupski and Weinstock, 1992). *E. coli* REP sequences have been shown to bind DNA polymerase I (Gilson *et al.*, 1990b), and DNA gyrase (Yang and Ames, 1988) in a complex interaction that involves other proteins including the histone-like protein HU (Yang and Ames, 1990). In *S. typhimurium*, REP sequences have been shown to be the end points of recombination events that occur in a

recA independent manner (Shyamala *et al.*, 1990), and in *E. coli*, DNA gyrase (topoisomerase II) has been shown to catalyse illegitimate recombination (Naito *et al.*, 1984). Sequence similarities can be predicted between VCR, and the DNA gyrase binding sites of REP and pBR322 (Yang and Ames, 1988; Yang and Ames, 1990), but attempts to obtain a *V. cholerae gyr* mutant in order to determine whether there are similar interactions between VCR and DNA gyrase have been unsuccessful. Nevertheless, one possible role for VCR is analogous to that proposed for REP sequences in Enterobacteriaceae, namely to provide anchoring sites in DNA for the assembly of nucleoid proteins (Yang and Ames, 1988; Gilson *et al.*, 1990b; Yang and Ames, 1990). However, as REP sequences are distributed more or less randomly throughout the *E. coli* chromosome (Dimri *et al.*, 1992) whereas VCRs are restricted to a region of approximately 10% of the *V. cholerae* chromosome, this does not seem to be a likely role for VCR. It is also worth noting that although REP-related sequences have been detected in diverse bacterial species, at least one member of Vibrionaceae, *V. vulnificus*, does not contain REP-related sequences (Versalovic *et al.*, 1991).

The most striking feature of VCR presented here is this localisation to a limited region of the chromosome. An upper limit on the region occupied by VCR has been determined by pulsed field gel electrophoresis, but the question remains as to what the lower limit is. Assuming a density of VCR repeats equivalent to that seen on pPM471 (*i.e.* nine in 6 kb) a lower limit of around 42 to 70 kb (depending on the strain) can be predicted, assuming that the number of repeats determined in Fig. 4.3B is valid. The orientation of the nine copies of VCR within pPM471 is such that recombination events between them will lead to deletion of the intervening DNA, and the presence of multiple copies of VCR on multicopy plasmids and M13 vectors (determined in hindsight!) leads to a very high rate of deletion events in *E. coli* *recA*[+] strains (Chapter 3; Franzon, 1988; F.K. DeGraaf, personal communication).

Such a "deletogenic" activity would make VCR an effector sequence in *V. cholerae* for deletion events associated with pathogenicity islands, in which virulence determinants that become undesirable, such as adhesins in a host mounting a successful immune response, can be jettisoned (albeit irreversibly) (Hacker *et al.*, 1990). It remains to be determined if genes other than *mrh* associated with VCR are virulence genes, and what the frequency of

recombination events between VCR sequences are. It is possible that all copies of VCR are of the same orientation in the *V. cholerae* chromosome, although we have not determined whether this is the case or not. Despite this, the VCR region of *V. cholerae* O1 seems to be stable within biotypes, with strains isolated over a forty year period showing a similar banding pattern on Southern hybridisation. The one exception to this, the El Tor strain C31 has a deletion covering the *mrh* region (van Dongen and DeGraaf, 1986), and although we have been unable to determine whether the deletion end points are associated with VCR sequences or not, it is possible that VCR was involved in the rearrangement, by analogy with deletions observed with multicopy plasmids.

One consequence of this model is that VCR is predicted to gradually disappear from the genome. Analysis of the occurrence of VCR in non-O1 strains indicated that some strains contained only a few copies, while others contained numerous copies (assuming that the differences observed are not the result of sequence divergence leading to loss of binding of the oligonucleotide probe used). As the non-O1 strains examined were isolated from diverse sources (Chapter 2), it is feasible that differing selection pressures being exerted on the bacteria have led to differing rates of deletion events. However, analysis of the deletion in strain C31 has indicated that the proximal portion of the region represented by pPM471 has been duplicated, suggesting that deletions are not the only possible recombination events between copies of VCR.

An alternative possible role for VCR is to provide a region where foreign genes can integrate into the chromosome. Out of 500 *V. cholerae* O1 isolates screened by Takeda *et al.* (1991a), one was identified that was ST⁻ (heat stable toxin). DNA sequencing revealed that the gene is flanked by copies of VCR suggesting that these sequences are somehow involved in the acquisition of the ST gene (Takeda *et al.*, 1990a). Elements for gene integration, termed integrons, have been identified in Gram negative bacteria and exhibit conserved sequences (Collis and Hall, 1992) but these do not show similarity to VCR, so if VCR does fulfill a similar function, then it represents a different class of integron from those described previously. However, as sequencing has revealed that ST genes in virulent non-O1 *V. cholerae* are also flanked by copies of VCR (Takeda *et al.*, 1991b) it is also possible to

explain the acquisition of the ST gene by this strain as occurring by a conjugation or transduction event from a non-O1 strain, without invoking a novel integration mechanism. Another alternative is that the extreme clustering of VCR could be due to the entire region including VCR sequences originating in a foreign host, but the presence of genes such as *mrh* exclusively in O1 *V. cholerae* (for strains thus far examined) and ST toxin genes predominantly in non-O1 *V. cholerae* (Takeda *et al.*, 1991b) makes this possibility seem less likely. Similarly, the suggestion that a single copy of VCR has integrated and is slowly spreading throughout the chromosome requires an explanation as to why the spread is restricted to one region of the chromosome and is not occurring randomly. Nevertheless, there is not evidence to discount any of these possibilities.

Although the biological function of VCR is not known, the extreme clustering and maintenance of these elements suggests that they fulfill the same, or a similar role. An earlier proposal that REP sequences may be involved in bacterial speciation within Enterobacteriaccae (Gilson *et al.*, 1987b) has been questioned, as REP-related sequences have recently been identified in diverse bacterial taxons (Versalovic *et al.*, 1991). VCR-related sequences have not been detected in bacteria other than V. cholerae in the limited survey reported here, but as Takeda *et al.* (1991a) have identified cross-reacting sequences in *V. mimicus* by colony blot, it seems possible that VCR exists in other species of *Vibrio*, in addition to *V. cholerae*. Identification of proteins (if any) that interact specifically with VCR should provide some clues as to their function, and hopefully help explain how they have persisted in what should be a deletogenic arrangement without being lost from the *V. cholerae* genome.

Chapter 5

Characterisation of a haemin-binding protein gene, linked to genes encoding the Mannose-Fucose-Resistant Haemagglutinin of *Vibrio cholerae* O1

5.1 Introduction

Free iron is a limiting nutrient in mammalian tissues and secretions as it is tightly bound by high-affinity iron-binding proteins, such as lactoferrin and transferrin. Hence, iron acquisition is of extreme nutritional importance for pathogenic bacteria, being required as a cofactor for enzymes involved in electron transport and intermediary metabolism (Williams and Roberts, 1989; Crosa, 1989). Several strategies have evolved for fulfilling this requirement with many bacteria, amongst them *V. cholerae*, secreting iron binding compounds, or siderophores, that are taken up by specific, energy (TonB) dependent transport systems when complexed with iron (Williams and Roberts, 1989; Crosa, 1989). However, siderophore production is not essential for virulence in *V. cholerae* O1 (Sigel *et al.*, 1985), indicating the presence of other iron-acquisition systems. A number of pathogenic bacteria are also capable of binding host factor-iron complexes. Well studied examples include haemin binding in *Shigella flexneri* and enteroinvasive *E. coli* (EIEC), which correlates with binding of the dye Congo red and with virulence (Stugard *et al.*, 1989), and transferrin and lactoferrin binding by *Neisseria gonorrhoeae* (Lee and Schryvers, 1988). *V. cholerae* may well utilise both strategies of iron acquisition, as it possesses haem-iron utilisation genes that transport the entire haemin moiety (Henderson and Payne, 1993), and some O1 strains exhibit specific lactoferrin binding properties (Ascencio *et al.*, 1992).

Proteins involved in iron uptake in Gram negative bacteria show a number of features in common. They are usually in the size range of 60-100 kDa and the transport systems are dependent on the TonB protein for energy (Crosa, 1989; Postle, 1990). For example, haemin-binding outer membrane proteins that have been identified include: a 101 kDa protein from *S. flexneri* and EIEC (Stugard *et al.*, 1989); a 78 kDa protein from *Yersinia enterocolitica* (Stojiljkovic and Hantke, 1992); and a 77 kDa protein from *V. cholerae* O1 (Henderson and Payne, 1993). A few proteins involved in iron transport have been shown to be lipoproteins. A 51 kDa haemin-binding lipoprotein has been identified in *Haemophilus influenzae* type b (Hanson and Hansen, 1991), and iron-ferrichrome uptake in the Gram positive bacterium *Bacillus subtilis* is dependent on a 33 kDa lipoprotein (Schneider and Hantke, 1993). A 14 kDa lipoprotein from *Y. enterocolitica* has been shown to facilitate ferrioxamine uptake in an *E. coli fhuE* mutant, but this appears to be due to overexpression of the lipoprotein causing

changes in outer membrane permeability rather than high-affinity uptake (Bäumler and Hantke, 1992).

The α_2-microglobulin (or 'lipocalin') protein superfamily consists of a number of eukaryotic proteins that are small hydrophobic molecule transporters. Superfamily members have diverse biological functions, which include roles in: sperm maturation (Brooks et al., 1986); vision (Schubert et al., 1986; Berman et al., 1987); the acute phase response (Pervaiz and Brew, 1987); the complement cascade (Haefliger et al., 1987; Ng et al., 1987); insect camouflage (Holden et al., 1987; Huber et al., 1987); senescence (Richardson et al., 1987); olfaction (Snyder et al., 1988); and retinol transport (Godovac-Zimmermann, 1988). Mammalian apolipoprotein D (ApoD) is also a member of the superfamily (Drayna et al., 1986; Drayna et al., 1987; Spreyer et al., 1990) and although originally presumed to be involved in cholesterol esterification or transport, has recently been shown to bind bilirubin (Peitsch and Boguski, 1990) and to be involved in the acute phase response, by virtue of induction of its expression in regenerating sciatic nerve (Spreyer et al., 1990). The level of primary structure similarity between superfamily members is low (pairwise alignments typically show around 30% sequence identity and multiple alignments show few common residues), but crystallisation of a number of superfamily members has revealed a common theme of a globular protein consisting of eight strands of anti-parallel β-sheet forming a hydrophobic cleft or pocket that acts as the ligand binding site. The size of superfamily members is also similar, varying from 18 to 20 kDa (unmodified). The structure seems well suited to the binding of labile and/or insoluble molecules to facilitate transport (Sawyer, 1987; Godovac-Zimmermann, 1988).

As noted in Chapter 3, the 19.6 kDa ORF2 shows sequence similarity to members of the α_2-microglobulin superfamily. Examination of the predicted protein sequence also revealed similarity to the amino-terminus of bacterial lipoproteins. Lipoproteins have been associated with the toxin-coregulated-pilus (TCP) and a possible adhesin gene cluster acf in V. cholerae (Parsot et al., 1991). In the former case the TcpC lipoprotein has been proposed to act as a membrane anchor for the fimbrial structure (Ogierman and Manning, 1992b). A similar proposal was made for ORF2, namely that it acted as a membrane anchor for the

MFRHA (Manning, 1991). Therefore, further analysis of ORF2 was undertaken, and is reported below.

5.2 Results

5.2.1 ORF2 encodes a lipoprotein

The predicted amino acid sequence of ORF2 (Chapter 3) is shown in Fig. 5.1, along with a consensus for processing sites recognised by lipoprotein signal peptidase (Pugsley, 1989). A good match with the consensus sequence is evident, leading to the prediction that the product of ORF2 is a lipoprotein. Furthermore, presuming that the actual cleavage site is as proposed, the protein is predicted to be located in the outer membrane, as position 2 in the predicted mature protein is occupied by a non-negatively charged amino acid, leucine (Yamaguchi *et al.*, 1988).

In order to confirm the prediction that ORF2 encodes a lipoprotein, *Hinc*II digested pPM471 (Franzon and Manning, 1986) was ligated with pGEM5Zf$^+$ and transformed into *E. coli* DH5α. After selection on NA plates supplemented with IPTG and X-Gal, 100 white colonies were picked and screened by colony blot, using the insert of plasmid pPM1648 (Chapter 4) as a probe (not shown). Six positive colonies were selected, purified and plasmid DNA analysed by restriction enzyme digestion (not shown). All contained the same 0.65 kb insert in an orientation such that transcription from the vector T7 promoter would lead to expression of the predicted ORF (not shown). One of these was picked and designated pPM1651 (Fig. 5.2).

Plasmid pPM1651 and the vector pGEM5Zf$^+$ were then introduced into *E. coli* DH5[pGP1-2] for analysis of the predicted ORF. Overexpression from the vector T7 promoter was performed in the presence of L-[^{35}S]-methionine and the presence or absence of globomycin, an inhibitor of the lipoprotein signal peptidase (Inukai *et al.*, 1978), and the cells analysed by SDS-PAGE (Fig. 5.3A). Two products were evident from pPM1651 in the absence of globomycin, an 18 kDa band and a 19.5 kDa band. In the presence of globomycin, only a 19.5 kDa band remained detectable. This result indicates that globomycin inhibits the processing of a protein encoded on pPM1651, which is therefore likely to be a lipoprotein

Figure 5.1: Nucleotide sequence and amino acid sequence of ORF2

The sequence shown here is based on that presented in Figure 3.3, with numbers below the nucleotide sequence representing the position in the sequence presented in Figure 3.3. Nucleotides in bold represent the distal and proximal ends of the direct repeats VCR2 and VCR3, respectively. The consensus amino acid sequence for the processing site of bacterial lipoproteins is from Pugsley (1989) with the cysteine residue that forms the fatty acylation site and amino terminus of mature lipoproteins represented in bold. The protein targeting signal at position number 2 of the mature protein is from Yamaguchi *et al.* (1988).

```
                                                           ⎧        V
                                           Consensus-  ⎨  L    L    S
                                                           ⎩        A
                       M   R   A   I   F   L   I   L   C   S   V   L   L   N
        TAGGTATTTGGAGATATCATGAGAGCTATCTTTTTGATTCTTTGCTCTGTTTTATTAAAT
                                                 1000


        A       D/E   –cytoplasmic membrane
        G    C   other–outer membrane
        G   C   L   G   M   P   E   S   V   K   P   V   S   D   F   E   L   N   N   Y
        GGCTGCTTGGGCATGCCCGAATCAGTAAAACCAGTGTCGGATTTTGAACTGAACAACTAT


        L   G   K   W   Y   E   V   A   R   L   D   H   S   F   E   R   G   L   S   Q
        TTAGGCAAATGGTACGAGGTTGCTCGACTCGATCACTCCTTTGAAAGAGGTTTAAGTCAG
                       1100

        V   T   A   E   Y   R   V   R   N   D   G   G   I   S   V   L   N   R   G   Y
        GTTACTGCGGAATACCGTGTTCGAAATGATGGTGGTATTTCGGTTCTTAATCGTGGTTAT


        S   E   E   K   G   E   W   K   E   A   E   G   K   A   Y   F   V   N   G   S
        TCTGAAGAGAAAGGTGAGTGGAAGGAAGCGGAAGGTAAAGCTTACTTTGTGAATGGCTCA
1200

        T   D   G   Y   L   K   V   S   F   F   G   P   F   Y   G   S   Y   V   V   F
        ACAGATGGCTATCTGAAGGTTTCATTTTTTGGCCCGTTTTATGGCTCCTACGTAGTGTTT
                                                 1300

        E   L   D   R   E   N   Y   S   Y   A   F   V   S   G   P   N   T   E   Y   L
        GAGTTAGACCGTGAAAACTACAGTTATGCTTTTGTGTCAGGGCCGAATACAGAATATCTG


        W   L   L   S   R   T   P   T   V   E   R   G   I   L   D   K   F   I   E   M
        TGGTTACTTTCAAGAACGCCGACTGTAGAACGAGGCATTCTGGACAAGTTCATAGAAATG
                       1400

        S   K   E   R   G   F   D   T   N   R   L   I   Y   V   Q   L   Q   *
        TCGAAAGAGCGTGGTTTTGATACAAATCGGCTTATTTACGTTCAGCTGCAATAAATACCT
```

Figure 5.2: Construction of subclones used in this chapter.

Numbers refer to the distance from the proximal end of the clone in kb and some restriction sites in the region are indicated. The expanded region shows ORF2 (*vlpA*) in greater detail, with the ORF represented by an open box. The flanking copies of the direct repeat VCR are represented by hatched boxes. The portions subcloned are represented by solid black lines, annotated with plasmid number. Details of plasmid construction are in the text.
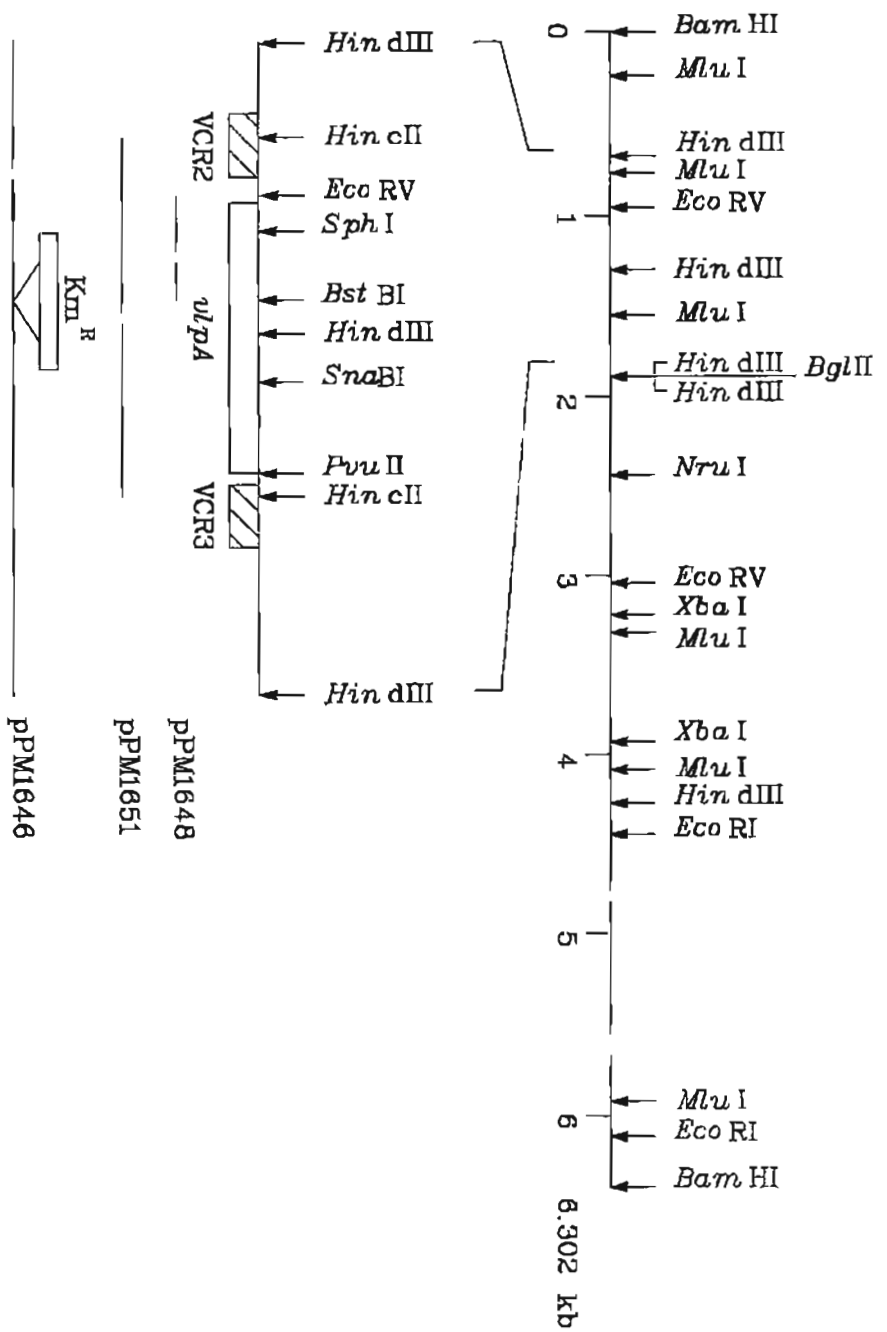
Figure 5.3: Demonstration that ORF2 (*vlpA*) encodes a lipoprotein.

Panel A shows L-[^{35}S]-methionine labelled proteins, and Panel B shows [9, 10-^{3}H]-palmitic acid proteins separated on 15% SDS-PAGE gels. Lanes are *E. coli* DH5[pGP1-2] also containing: 1 and 2, pGEM5Zf$^+$; 3 and 4, pPM1651. Lanes 2 and 4 were labelled in the presence of globomycin. The positions of standard molecular weight markers (Chapter 2) are indicated by lines to the left of lane 1 in each panel.
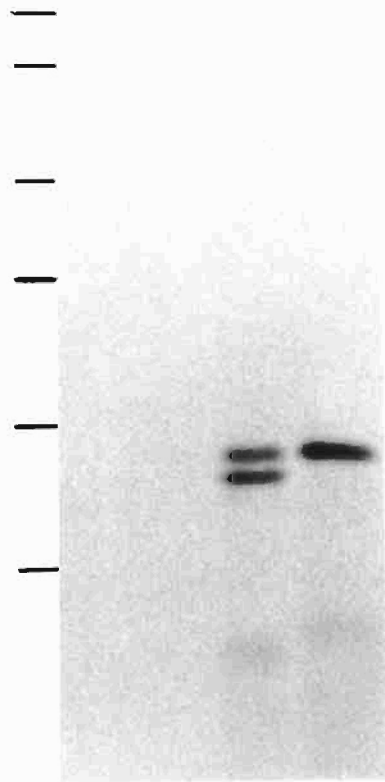
(Inukai *et al.*, 1978). In order to confirm this result, T7 overexpression was then performed on the same strains in the presence of [9, 10-^{3}H] palmitic acid, and the presence or absence of globomycin, and the resulting mixture analysed by SDS-PAGE (Fig. 5.3B). In this case a product was detectable only for pPM1651 in the absence of globomycin, an 18 kDa band. This indicates that in *E. coli* K-12, the product of ORF2 is a lipoprotein, and is therefore extremely likely to be a lipoprotein in *V. cholerae* as well. ORF2 was therefore renamed *vlpA*, for *Vibrio cholerae* lipoprotein A.

5.2.2 VlpA is a member of the α$_2$-microglobulin superfamily

The deduced amino acid sequence of VlpA was compared to entries in PIR (Release 33) and SWISS-PROT (Release 22) databases, using the BLAZE algorithm (Brutlag *et al.*, in preparation). Significant matches were obtained with a number of members of the α$_2$-microglobulin protein superfamily, with the best three alignments shown in Fig. 5.4. As has been noted for pairwise alignments of other members of the superfamily (Sawyer, 1987; Godovac-Zimmermann, 1988), the level of primary structure similarity is not high, with VlpA precursor giving scores of: 29% identity with rat ApoD precursor (Spreyer *et al.*, 1990); 26% identity with human ApoD precursor (Drayna *et al.*, 1986; Drayna *et al.*, 1987); and 25% identity with the bilin-binding protein (Bbp) from the white cabbage butterfly *Pieris brassicae* (Suter *et al.*, 1988). Matches were also obtained with a number of other members of the superfamily (not shown), including: 25% identity to the chicken purpurin precursor (Schubert *et al.*, 1986; Berman *et al.*, 1987); 24% identity with donkey β-lactoglobulin (Godovac-Zimmermann *et al.*, 1988); 23% identity with a rat androgen-dependent epididymal secretory protein precursor (Brooks *et al.*, 1986) and with a chicken quiescence-specific protein precursor (Bedard *et al.*, 1989); 22% identity with insecticyanin A (IcyA), a bilin binding protein from the tobacco hornworm *Manduca sexta* (Riley *et al.*, 1984) and with the crustacyanin C1 subunit of the carotenoid pigment complex of the European lobster *Homarus gammarus* (Keen *et al.*, 1991); and 21% identity with the crustacyanin A2 subunit of *H. gammarus* (Keen *et al.*, 1991), human (Colantuoni *et al.*, 1983) and bovine (Berni *et al.*, 1990) retinol binding proteins, and the minor major urinary protein 15 precursor involved in pheromone transport in the urine of male mice (Clark *et al.*, 1985).

Figure 5.4: Pairwise alignment of pre-VlpA with similar sequences lodged in the PIR and SWISS-PROT databases.

Panel A shows an alignment with Rat ApoD (mature) (Spreyer *et al.*, 1990), panel B shows an alignment with human ApoD (pre-protein) (Drayna *et al.*, 1986; Drayna *et al.*, 1987), and panel C shows an alignment with Bbp from *P. brassicae* (mature) (Suter *et al.*, 1988). Identical residues in the alignments are joined by a solid line (|), similar residues are joined by a dotted line (:). The boundaries of the region of similarity are indicated by an X above and below the alignment. Numbers annotating the sequence indicate the position from the amino terminus of each residue. The alignments were performed using the algorithm BLAZE (Brutlag *et al.*, in preparation).

A

```
            10        20      X      30         40        50
 MRAIFLILCSVLLNGCLGMPESVKPVSD-FELNNYLGKWYEVARLDHSFERGLSQVTAEY
            ||  :  |::::|||:|||: ::  |||:|    :  |:|
 LATLAGLFTTTEGQSFHLGKCPSPPVQENFDVKKYLGRWYEIEKIPVSFEKGNC-IQANY
    10        20        30 X       40        50        60

    60        70        80        90        100       110
 RVRNDGGISVLNRGYSEEKGEWKEAEGKAYFVNGSTDGYLKVSFFG---PFYGSYVVFEL
    : ::|  |  |||:    :  |  :: |||  |  |  :  |  ||:   |   ::|  ::
 SLMENGNIKVLNKELRPD-GTLNQVEGEAKQSNMSEPAKLEVQFFSLMPP--APYWILAT
       70        80        90        100       110       120

    120         130       140       150       X160
 DRENYSYAFVSGP-----NTEYLWLLSRTPTVERGILD--KFIEMSKERGFDTNRLIYVQ
 |  |:|:  :       : :|:|:|:||  |  :     :   |:|   |::
 DYESYALVYSCTTFFWFFHVDYVWILGRNPYLPPETITYLKYILTSNDIDIAKITTKDQA
       130       140       150       160       170       180

 170
   L

   N
```

B

```
 X          10        20                    30        40        50
 MRAIFLILCSVLLNGCLGMPE---------SVKPVSD-FELNNYLGKWYEVARLDHSFER
 |  ::|:|  :   |  |  :|  :|          :   || :  |::|:|||:|||: ::  :||
 MVMLLLLLSA--LAGLFGAAEGQAFHLGKCPNPPVQENFDVNKYLGRWYEIEKIPTTFEN
 X          10        20        30        40        50

            60        70        80        90        100
 GLSQVTAEYRVRNDGGISVLNRGYSEEKGEWKEAEGKAYFVNGSTDGYLKVSF--FGPFY
 |   :  |:|  :  ::|  |||:    :  |   ::  ||  |  |  |  |  |
 GRC-IQANYSLMENGKIKVLNQELRAD-GTVNQIEGEATPVNLTEPAKLEVKFSWFMP-S
    60        70        80        90        100       110

            120         130       140       150       160
 GSYVVFELDRENYSYAFVSGP-----NTEYLWLLSRTPTVERGILDKFIEMSKERGFDTN
 ::|  ::  |  |||:  :        : ::  |:|:  |  :     :|  :  ::        |  :
 APYWILATDYENYALVYSCTCIIQLFHVDFAWILARNPLNPPETVDSLKNILTSNNIDVK
    120       130       140       150       160       170

 X   170
 RLIYVQL
 ::
 KMTVTDQ
  X180
```

C

```
            X         30         40         50                  60
LLNGCLGMPESVKPVSDFELNNYLGKWYEVARLDHSFER----GLSQVTAEYRVRNDGGI
    | |   | | | | :| : : | |   | | |   | | | :   :|   |:     | :: |:|              |
NVYHDGACPE-VKPVDNFDWSNYHGKWWEVAKYPNSVEKYGKCGWAEYTPE-------GK
            10         20         30         40         50


    70         80         90                   100        110
SVLNRGYSEEKGEWKEAEGKAYFVNGST----DG-----YLKVSFFGPFYGS------YV
    | |     |           | |   | | :: | ::       |:        |   |:             :
SVKVSNY-------HVIHGKEYFIEGTAYFVGDSKIGKIYHKLT-----YGGVTKENVFN
                 60         70         80              90        100


    120               130        140        150        160
VFELDRENYSYAFVS------GPNTEYLWLLSRTPTV---ERGILDKFIEMSKERGFDTN
    |:    |    | |    ::              : :::|:|||:   :       :    ::::   :            |:
VLSTDNKNYIIGYYCKYDEDKKGHQDFVWVLSRSKVLTGEAKTAVENY--LIGSPVVDSQ
                 110        120        130        140        150


    170
RLIYVQL
    :|:|  ::
KLVYSDF
160     X
```

Despite the divergent biological roles and ligands of the proteins listed above, the alignment hinted at a possible function for VlpA. The best match, ApoD, has recently been shown to bind bilirubin IX, in an approximately one-to-one molar ratio (Peitsch and Boguski, 1990). The third-best match, Bbp, also binds a porphyrin, biliverdin IX (Huber *et al.*, 1987). Detailed crystallographic structures are available for Bbp (Huber *et al.*, 1987) and another insect biliverdin IX binding protein, insecticyanin (IcyA) from the tobacco hornworm *M. sexta* (Holden *et al.*, 1987). Peitsch and Boguski (1990) have built a detailed structural model of Human ApoD based on its similarity to these proteins, then used this model to predict porphyrin binding by ApoD. To see if VlpA might also fit into this porphyrin-binding group of 'lipocalins,' a multiple alignment between the proteins was performed (Fig. 5.5). As multiple alignments between members of the superfamily typically yield only a few residues in common (Sawyer, 1987; Godovac-Zimmermann, 1988) the presence of residues in VlpA that are similar to residues that interact with the ligand in Bbp and IcyA raises the possibility that VlpA also binds a porphyrin.

Preliminary structural comparison of the proteins aligned in Fig. 5.5 was also undertaken (Fig. 5.6). Hydropathy plots do not show marked similarity to one another, but all proteins have an overall hydrophilic nature, with some hydrophobic domains predominantly in the carboxy-terminal half of the proteins (Fig. 5.6A). The most striking feature is the hydrophobic amino-terminus of the mature portion of VlpA, a structural feature that is predicted to promote association of the lipid moiety of the processed protein with the cell membrane. Comparison of predicted secondary structures of the proteins (Fig. 5.6B), does not reveal striking similarities. A more rigorous analysis, consisting of building a model for VlpA based on its similarities to the crystallised Bbp and IcyA, is currently being undertaken elsewhere.

5.2.3 Examining activities of VlpA

A mutant in *vlpA* was constructed by linearising pPM471 (Franzon and Manning, 1986) with *Bst*BI, religating with the Km^R resistance cartridge from pUC4-K (Viera and Messing, 1982) and transforming the mixture into *E. coli* HB101. Ap^R, Km^R colonies were obtained, but exhibited a drastically reduced growth rate (colonies took 2-3 days to appear on

Figure 5.5: Multiple alignment of VlpA with porphyrin binding members of the α_2-microglobulin superfamily.

Identical sequences are marked with an asterisk, similar residues with a dot. Residues in Bbp (Huber *et al.*, 1987) and IcyA (Holden *et al.*, 1987) that have been shown, by X-ray crystallography, to interact with the ligand are in bold type and underlined, while residues in IcyA in which the side chain lies close to the ligand are underlined (Holden *et al.*, 1987). The alignment was performed using the multiple alignment program CLUSTAL (Higgins and Sharp, 1989). The alignment covers the mature portions of the respective proteins.

```
VlpA        ----CLGMPESVKPVSDFELNNYLGKWYEVARLDHSFERGLSQVTAEYRV
ApoD Rat    -QSFHLGKCPSPPVQENFDVKKYLGRWYEIEKIPVSFEK-GNCIQANYSL
ApoD Human  -QAFHLGKCPNPPVQENFDVNKYLGRWYEIEKIPTTFEN-GRCIQANYSL
Bbp         -NVYHDGACPEVKPVDNFDWSNYHGKWWEVAKYPNSVEKYGKCGWAEYT-
IcyA        GDIFYPGYCPDVKPVNDFDLSAFAGAWHEIAKLPLENENQGKCTIAEYK-
                    *             .*.    . * * *. .        *          *.*

VlpA        RNDGGISVLNRGYSEEKGEWKEAEGKAYFVNGSTDG-----YLKVSFFGP
ApoD Rat    MENGNIKVLNKELRPD-GTLNQVEGEAKQSNMSEPAK-----LEVQFFSL
ApoD Human  MENGKIKVLNQELRAD-GTVNQIEGEATPVNLTEPAK-----LEVKFSWF
Bbp         -PEGKSVKVSNYHVIH-GKEYFIEGTAYPVGDSKI---GKIYHKLTYGGV
IcyA        -YDGKKASVYNSFVSN-GVKEYMEGDLEIAPDAKYTKQGKYVMTFKFGQR
                  .*          .          .  *        ** .                .

VlpA        F-YGSYVVFELDRENYSYAFVSG------PNTEYLWLLSRTPTVERGILD
ApoD Rat    MPPAPYWILATDYESYALVYSCTTFF-WFFHVDYVWILGRNPYLPPETIT
ApoD Human  MPSAPYWILATDYENYALVYSCTCII-QLFHVDFAWILARNPNLPPETVD
Bbp         TKENVFNVLSTDNKNYIIGYYCKYDEDKKGHQDFVWVLSRSKVLTGEAKT
IcyA        VVNLVPWVLATDYKNYAINYNCDYHPDKKAHSIHAWILSKSKVLEGNTKE
                  ..    *      .*       .                 .         *.*..       .

VlpA        --KFIEMSKERGFDTNRLIYVQL------------------
ApoD Rat    YLKYILT--SNDIDIAKITTKD--QANC-----------PDFL
ApoD Human  SLKNILT--SNNIDVKKMTVTD--QVNC-----------PKLS
Bbp         AVENYLIG-SPVVDSQKLVYSDFSEAACKV----------N
IcyA        VVDNVLKTFSHLIDASKFISNDFSEAACQYSTTYSLTGPDRH
                        *        .          .
```

Figure 5.6: Comparison of VlpA with porphyrin binding members of the α_2-microglobulin superfamily.

Panel A shows hydropathy plots of (from top to bottom) VlpA, ApoD (rat), ApoD (human), Bbp (*P. brassicae*), and IcyA (*M. sexta*), determined according to Kyte and Doolittle (1982), using a window of six amino acids. Panel B shows secondary structure predictions for the same proteins (from left to right) determined by the use of the algorithm of Chou and Fasman (1978).

A
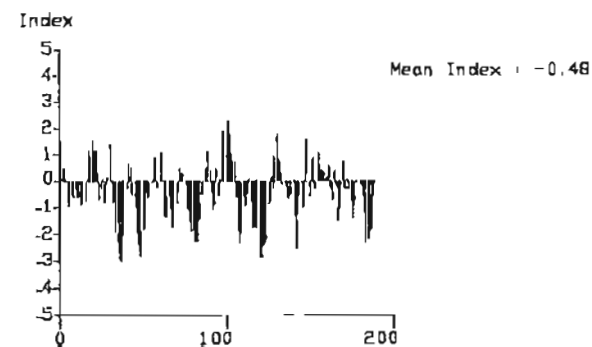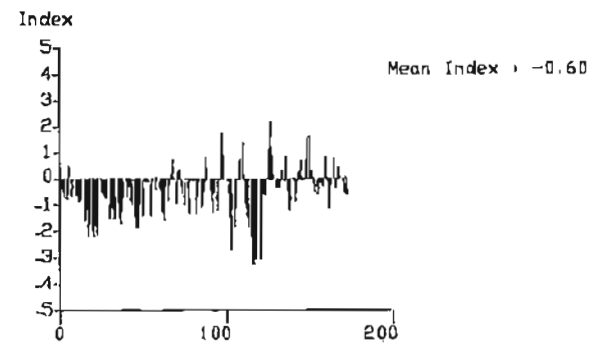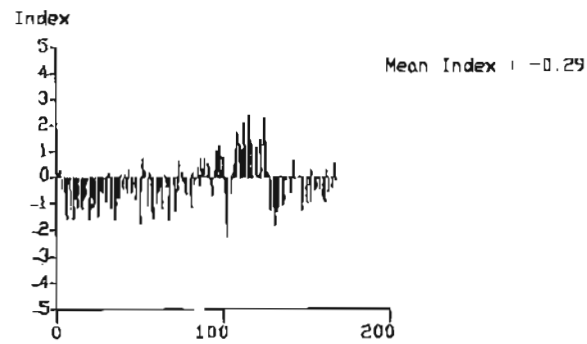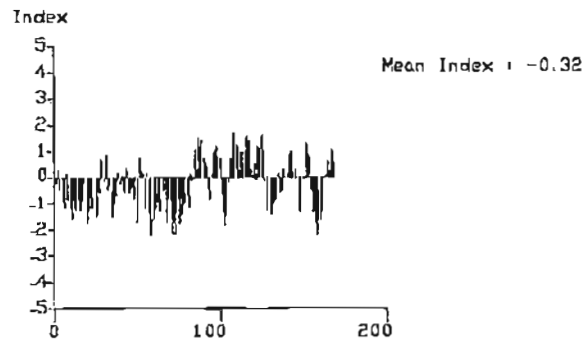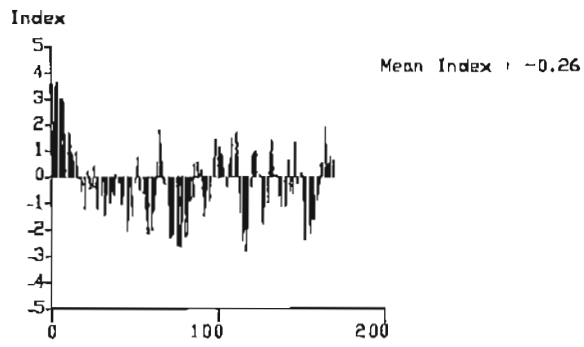
B
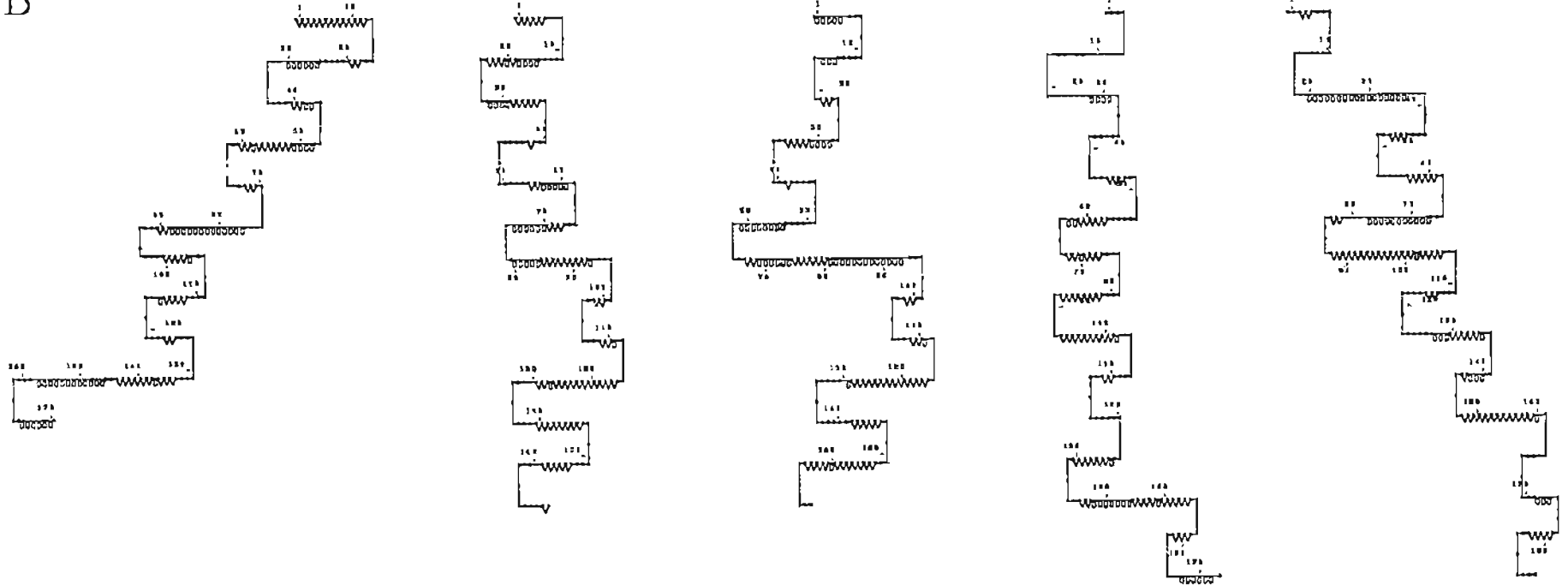
NA plates with selection). Restriction analysis of plasmids isolated from 6 such colonies showed they contained the Km^R cartridge inserted in either orientation in the unique *Bst*BI site (not shown): one in which the cartridge was of the same orientation as *vlpA* was designated pPM1646 (Fig. 5.2).

The role played by VlpA in the MFRHA phenotype exhibited by pPM471 was examined by comparing the haemagglutination titre of *E. coli* hosts into which pPM471 or pPM1646 had been introduced (Table 5.1). Despite the reduced growth rate of cells containing pPM1646, no difference in titre was observed in the host strain examined, indicating that at least in *E. coli*, VlpA is not directly involved in MFRHA activity.

In order to test for a role in porphyrin binding, *E. coli* HB101 containing pBR322, pPM471 or pPM1646 were streaked on media containing the dye congo red, binding of which has been correlated with haem binding (Stugard *et al.*, 1989), or haemin. Despite varying growth temperature and media composition (including stressing for free iron by adding the iron chelator 2, 2' dipyridyl), conditions could not be identified that led to pPM471 mediating binding of either compound by *E. coli* HB101 (not shown). Similarly, attempts to detect *vlpA* specific transcripts in *V. cholerae* 569B and O17 grown in NB or NB containing 2,2' dipyridyl (and in which expression of the iron-starvation induced *hlyA* was activated; not shown) were unsuccessful. Therefore, *E. coli* DH5[pGP1-2, pPM1651] were streaked on minimal medium containing casamino acids and congo red, incubated at 30°C for 24 hours to allow visible colony formation, and then at 42°C for 24 hours, to induce expression of *vlpA* from the subclone. In this case, congo red binding was readily apparent, whereas congo red binding by *E. coli* DH5[pGP1-2, pGEM5Zf$^+$] was not (Fig. 5.7A), indicating that the induced protein may bind a porphyrin. In an attempt to define a specific substrate for VlpA, the same strains were streaked on minimal medium containing casamino acids and haemin, haematoporphyrin or protoporphyrin IX, followed by a similar incubation regime (Fig. 5.7B-D). Significant binding was observed for both haemin and haematoporphyrin, but only limited binding for protoporphyrin IX.

Table 5.1 The influence of a *vlpA*::Km^R mutation on MFRHA activity in an *E. coli* clone

plasmid	relevant genotype[a]	haemagglutination titre in *E. coli* AAEC189[b]
pBR322	vector	0
pPM471	*vlpA$^+$*, *mrhA$^+$B$^+$*	16
pPM1646	*vlpA*::Km^R	16

[a] described in detail in the text.

[b] Haemagglutination titre is the result for three independently isolated transformants.

Figure 5.7: Plate binding of chromophores by *E. coli* K-12 expressing VlpA.

Plates contain: Congo red (A); haemin (B); haematoporphyrin (C); or protoporphyrin IX (D). Bacteria streaked on the top half of all plates are *E. coli* DH5[pGP1-2, pGEM5Zf$^+$], while bacteria streaked on the bottom half of all plates are *E. coli* DH5[pGP1-2, pPM1651].

Examination of the haemin binding exhibited by *E. coli* DH5[pGP1-2, pPM1651] was also undertaken using a liquid haemin binding assay. In this system, cells expressing VlpA show an approximately linear increase in the amount of haemin bound that is directly proportional to the amount added, whereas cells containing vector alone show limited binding that becomes saturated within the limit of the assay (Fig. 5.8). The addition of urea to a final concentration of 4 M led to an inhibition of haemin binding (Fig. 5.8), suggesting that the haemin binding observed is due to a hydrophobic interaction (Peitsch and Boguski, 1990). No alteration of the binding observed with cells containing vector alone was observed on the addition of urea in the presence of haemin (not shown).

Ideally, examination of the expression and function of *vlpA* should be undertaken in the homologous host, *V. cholerae*. Therefore, an attempt was made to transfer the *vlpA*::Km^R mutation into *V. cholerae* O1 strains using pCACTUS, a derivative of pIB307 (Bloomfield *et al.*, 1991b) containing the *sacB* gene of *B. subtilis* (C.A. Clark, unpublished). This vector has been designed to allow the transfer of non-selectable alleles by virtue of a temperature-sensitive replicon and a positive (Cm^R) and negative (*sacB*) selection marker (Bloomfield *et al.*, 1991b; C.A. Clark, unpublished). Plasmid pPM1646 was digested with *Hinc*II and ligated with *Sma*I digested pCACTUS, the resulting mixture transformed into *E. coli* DH5α and Km^R, Cm^R colonies selected at 30°C. Resulting plasmids were characterised by restriction analysis, one being selected and designated pPM1647 (not shown). This plasmid was then introduced into *V. cholerae* strains Z17561 and O17 and growth conditions followed to promote exchange of plasmid and chromosomal *vlpA* alleles (C.A. Clark, unpublished). Six resulting colonies from each strain were then screened by Southern blot, but in all cases integration of *vlpA*::Km^R occurred at a second site (not shown). Therefore, further attempts to isolate *vlpA* mutants of *V. cholerae* were not made.

5.2.4 Carboxy-terminal truncation mutants of VlpA

In order to identify regions of VlpA involved in ligand binding, a series of carboxy-terminal translational fusions between *vlpA* and *phoA* of *E. coli* was constructed. First, the truncated *phoA* gene (a gene fragment encoding only the mature portion of the protein was used) from pCH39 (Hoffman and Wright, 1985) was subcloned as a *Pst*I-*Xho*I fragment into

Figure 5.8: Liquid binding of haemin by *E. coli* K-12 expressing VlpA.

The strains examined are *E. coli* DH5[pGP1-2] containing an additional plasmid as indicated by annotation of the curves.
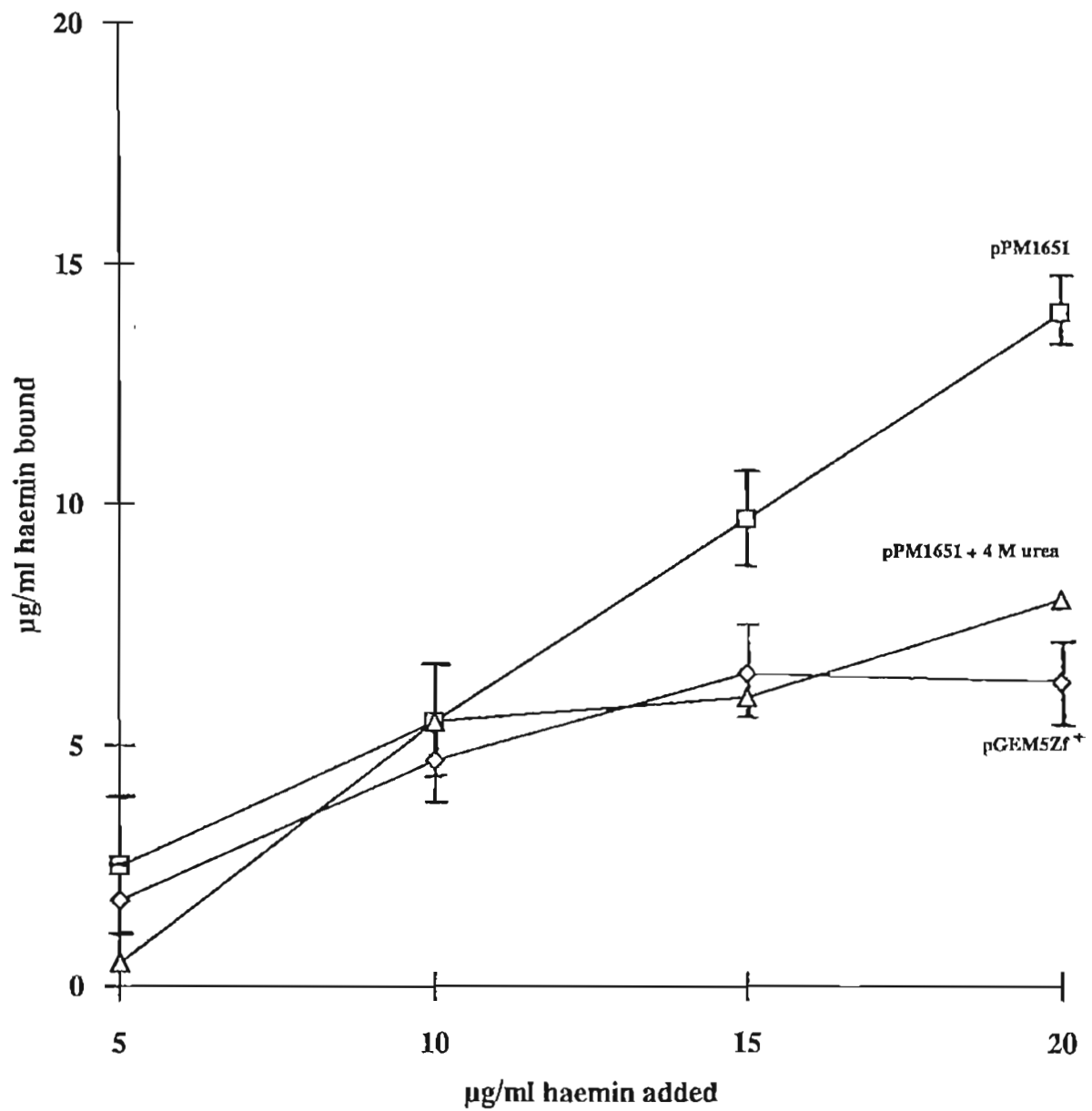
*Pst*I, *Sal*I digested pPM1651, to generate plasmid pPM1652. Next, pPM1652 was digested with *Not*I, protected by end repair, digested with *Spe*I, and digested unidirectionally with exonuclease III. Electrophoretic analysis of the exonuclease III digestion indicated that *Spe*I digestion was incomplete (not shown). Nevertheless, ligation mixes containing digestion products were transformed into *E. coli* CC118. One thousand two hundred transformants were screened by colony blot with anti-PhoA antiserum, to detect in-frame fusions between *vlpA* and *phoA*. Sixty four positively reacting colonies were obtained (not shown), and were patched onto minimal medium containing X-pho as well as being re-screened by colony blot. Colour phenotypes ranging from white to dark blue were obtained on X-pho containing media (not shown). Sixty two of these colonies still reacted with the antiserum (not shown), and plasmid DNA was isolated from these and subjected to restriction analysis (not shown), indicating a spread of fusions that covered the entire *vlpA* coding region. The fusion junctions of eighteen plasmids, that represented the spread of fusions obtained as well as exhibiting the full range of colour phenotypes on X-pho containing medium, were sequenced, the results indicating that eleven independent fusions had been isolated (Fig 5.9). These fusion-bearing plasmids were designated pPM1664 to pPM1674, inclusively. It was anticipated that some information on the topology of VlpA could also be obtained from analysis of these fusion proteins. However, multiple isolates had been isolated for some of the fusion points, and some of these exhibited more than one colour phenotype on X-pho containing media. Therefore, attempts to use the fusions to study the cellular location of VlpA were not made (see also Section 5.3). Finally, in order to check the stability of the fusion proteins to proteolytic degradation, a western blot was performed on whole cell samples of *E. coli* CC118 containing the plasmids, using an anti-PhoA antiserum (Fig. 5.10). In all cases, the predominant reactive band observed corresponded in size to that predicted for full length fusion protein as predicted from the sequence, indicating that significant degradation of the fusion proteins had not occurred.

The plasmids described in Fig 5.9, along with the "parental" plasmid pPM1652, were then transferred into *E. coli* DH5[pGP1-2] in order to overexpress the fusion proteins and to examine their haemin binding activity in a liquid haemin binding assay (Fig. 5.11). In this experiment, VlpA expressed from both pPM1651 and pPM1652 show haemin binding

Figure 5.9: Construction of *vlpA-phoA* fusion proteins.

Panel A shows the nucleotide and predicted amino acid sequence of the insert of pPM1652. Restriction sites at the 5' and 3' ends of *vlpA* and the start of *phoA* are indicated, as are intervening vector sequences. The predicted VlpA amino acid sequence is also shown, with the end points of the eleven fusion proteins indicated by the appropriate plasmid number. Sequences of VCR2 and VCR3 contained within pPM1652 are indicated by bold type. Panel B shows the nucleotide and predicted amino acid sequence across the junction of the eleven fusion proteins indicated in Panel A. The junction point is indicated by a space in the nucleotide sequence.

A

pGEM5Zf⁺GAT **ACGCGTGGCGTTTCCAGTCCCATTGAGCCGCGGT**
  *Eco*RV/*Hinc*II

GGTTGCTGTTGTTGTGTTTGAGTTTAGTGGTAGTGCGTTGTCAGCCCCTTAGGCGGGCGT
900

```
                 M   R   A   I   F   L   I   L   C   S   V   L   L   N
TAGGTATTTGGAGATATCATGAGAGCTATCTTTTTGATTCTTTGCTCTGTTTTATTAAAT
                                        1000

     G   C   L   G   M   P   E   S   V   K   P   V   S   D   F   E   L   N   N   Y
GGCTGCTTGGGCATGCCCGAATCAGTAAAACCAGTGTCGGATTTTGAACTGAACAACTAT
              |               |                         |
         pPM1674         pPM1673                    pPM1672

     L   G   K   W   Y   E   V   A   R   L   D   H   S   F   E   R   G   L   S   Q
TTAGGCAAATGGTACGAGGTTGCTCGACTCGATCACTCCTTTGAAAGAGGTTTAAGTCAG
              1100

     V   T   A   E   Y   R   V   R   N   D   G   G   I   S   V   L   N   R   G   Y
GTTACTGCGGAATACCGTGTTCGAAATGATGGTGGTATTTCGGTTCTTAATCGTGGTTAT
        |   |                                     |
   pPM1671  |                                 pPM1669
        pPM1670

     S   E   E   K   G   E   W   K   E   A   E   G   K   A   Y   F   V   N   G   S
TCTGAAGAGAAAGGTGAGTGGAAGGAAGCGGAAGGTAAAGCTTACTTTGTGAATGGCTCA
1200                      |
                     pPM1668

     T   D   G   Y   L   K   V   S   F   F   G   P   F   Y   G   S   Y   V   V   F
ACAGATGGCTATCTGAAGGTTTCATTTTTTGGCCCGTTTTATGGCTCCTACGTAGTGTTT
                        |                    1300              |
                   pPM1667                                pPM1666

     E   L   D   R   E   N   Y   S   Y   A   F   V   S   G   P   N   T   E   Y   L
GAGTTAGACCGTGAAAACTACAGTTATGCTTTTGTGTCAGGGCCGAATACAGAATATCTG
                            |
                       pPM1665

     W   L   L   S   R   T   P   T   V   E   R   G   I   L   D   K   F   I   E   M
TGGTTACTTTCAAGAACGCCGACTGTAGAACGAGGCATTCTGGACAAGTTCATAGAAATG
                        1400

     S   K   E   R   G   F   D   T   N   R   L   I   Y   V   Q   L   Q   *
TCGAAAGAGCGTGGTTTTGATACAAATCGGCTTATTTACGTTCAGCTGCAATAAATACCT
                            |
                       pPM1664
```

AACAAACGCCTCAAGAGGGACTGTC ATCACTAGTGCGGCCGC <u>CTGCAGCCCTGTTCTGGA</u>
         *Hinc*II/*Eco*RV   *Pst*I
                        pGEM5Zf⁺   *phoA*

B

pPM1674

```
          N   G   C   L   A   A   C   S   P   V   L   E   -PhoA
          GGCTATCAGAAG GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1673

```
          L   G   M   P   A   A   C   S   P   V   L   E   -PhoA
          TTGGGCATGCCC GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1672

```
          K   P   V   S   A   A   C   S   P   V   L   E   -PhoA
          AAACCAGTGTCG GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1671

```
          S   Q   V   T   A   A   C   S   P   V   L   E   -PhoA
          AGTCAGGTTACT GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1670

```
          Q   V   T   A   A   A   C   S   P   V   L   E   -PhoA
          CAGGTTACTGCG GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1669

```
          G   G   I   S   A   A   C   S   P   V   L   E   -PhoA
          GGTGGTATTTCG GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1668

```
          G   E   W   K   A   A   C   S   P   V   L   E   -PhoA
          GGTGAGTGGAAG GCCGCCTGCAGCCCTGTTCTGAA-phoA
                            PstI
```

pPM1667

```
          G   Y   L   K   A   A   C   S   P   V   L   E   -PhoA
          GGCTATCTGAAG GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1666

```
          G   S   Y   V   A   A   C   S   P   V   L   E   -PhoA
          GGCTCCTACGTA GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1665

```
          N   Y   S   Y   A   A   C   S   P   V   L   E   -PhoA
          AACTACAGTTAT GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

pPM1664

```
          F   D   T   N   A   A   C   S   P   V   L   E   -PhoA
          TTTGATACAAAT GCCGCCTGCAGCCCTGTTCTGGAA-phoA
                            PstI
```

Figure 5.10: Expression of *vlpA-phoA* fusion proteins.

Shown is a western blot of total cellular protein probed with anti-PhoA antiserum. Lanes are *E. coli*: 1, C75a; 2, G206; 3, CC118[pGEM5Zf⁺]; 4, CC118[pPM1652]; 5, CC118[pPM1664]; 6, CC118[pPM1665]; 7, CC118[pPM1666]; 8, CC118[pPM1667]; 9, CC118[pPM1668]; 10, CC118[pPM1669]; 11, CC118[pPM1670]; 12, CC118[pPM1671]; 13, CC118[pPM1672]; 14, CC118[pPM1673]; 15, CC118[pPM1674].
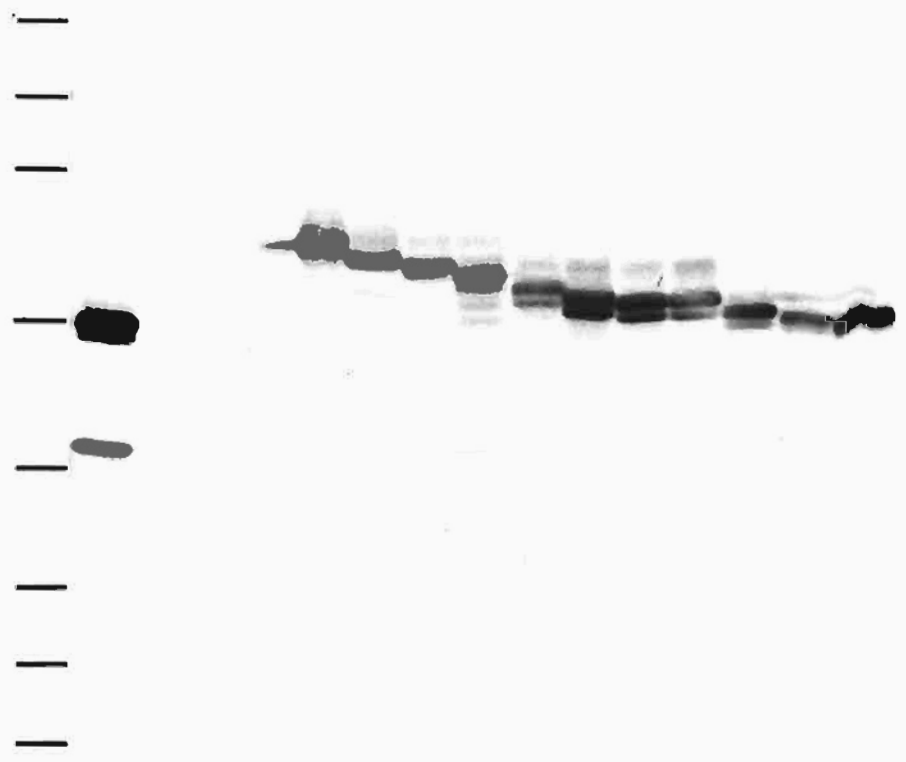
Figure 5.11: Liquid binding of haemin by *E. coli* K-12 expressing VlpA-PhoA fusion proteins.

Binding by *E. coli* DH5[pGP1-2] also containing the indicated plasmids is represented with the negative and positive controls, *E. coli* DH5[pGP1-2, pGEM5Zf$^+$] and *E. coli* DH5[pGP1-2, pPM1651], respectively, for comparison. The individual curves are annotated with appropriate plasmid numbers.

Row 1, left graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1652 / pCRM521⁺

Row 1, middle graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1664 / pCRM521⁺

Row 1, right graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1665 / pCRM521⁺

Row 2, left graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1666 / pCRM521⁺

Row 2, middle graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1667 / pCRM521⁺

Row 2, right graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1668 / pCRM521⁺

Row 3, left graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1669 / pCRM521⁺

Row 3, middle graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1670 / pCRM521⁺

Row 3, right graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1671 / pCRM521⁺

Row 4, left graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1672 / pCRM521⁺

Row 4, middle graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1673 / pCRM521⁺

Row 4, right graph:
µg/ml haemin bound — µg/ml haemin added
pPM1651 / pPM1674 / pCRM521⁺

comparable to that observed for pPM1651 in Fig. 5.8, while pGEM5Zf⁺ also shows binding comparable to that observed previously. Of the deletion constructs: pPM1664, pPM1665, pPM1666 and pPM1667 mediate binding comparable to the wild-type VlpA expressing clone; pPM1668, pPM1669 and pPM1671 show reduced binding; while pPM1670, pPM1672, pPM1673 and pPM1674 show binding comparable to the vector. This indicates that it is possible to destroy haemin binding by deleting a sufficient portion of *vlpA*, and thus that the haemin binding observed above is due to VlpA.

5.2.5 Detection of *vlpA* related sequences in *V. cholerae* O1

The possibility that sequences closely related to *vlpA* exist in the *V. cholerae* O1 chromosome has been raised by: the identification of a second band cross-reacting with a pPM1648-derived probe in *V. cholerae* 569B; a cross-reacting band in *V. cholerae* C31 (Chapter 4); and realisation that in all cases screened, the pCACTUS introduced *vlpA*::Km^R allele had integrated at a second site (Section 5.2.3). Therefore a Southern blot was performed on *Bam*HI digested chromosomal DNA purified from a number of *V. cholerae* O1 strains, using a probe derived from pPM1651 (Fig. 5.12). As noted previously (for a smaller *vlpA*-specific probe; Chapter 4) *V. cholerae* C31, in which the entire region encoded by pPM471 has been deleted (van Dongen and DeGraaf, 1986; Chapter 4), cross-hybridises with this probe, exhibiting a single band of 10 kb. All other strains examined show multiple bands, with the largest number, four, seen with *V. cholerae* O17. This indicates that *vlpA* is present in multiple copies in most strains of *V. cholerae* O1.

5.3 Discussion

Analysis of nucleotide sequence data (see also Chapter 3) leads to the prediction that one of the ORFs (ORF2) linked to the *mrhAB* operon encodes a lipoprotein. This gene, designated *vlpA*, encodes a 19.6 kDa pre-protein that is processed to form an 18 kDa protein on cleavage of the signal peptide. Processing is completely inhibited by the cyclic peptide antibiotic globomycin, an inhibitor of signal peptidase II (Inukai *et al.*, 1978) and labelling of the protein with tritiated palmitic acid is only observed in the absence of globomycin. These data confirm that the protein is expressed as a lipoprotein in *E. coli* K-12, and thus is also extremely likely to be a lipoprotein in *V. cholerae*.

Figure 5.12: Southern blot of *V. cholerae* O1 strains probed with a *vlpA*-specific probe.

The filter shown was hybridised with a probe derived from the 0.5 kb *Eco*RV-*Pvu*II fragment of pPM1651 (Figure 5.1). Lanes are DNA from the strains: 1, 569B; 2, CA401; 3, CA411; 4, C21; 5, AA14041; 6, Z17561; 7, O17; 8, C5; 9, C31; 10, H1; 11, AA14073; 12, BM69.

The presence of a lipoprotein gene in close proximity to an adhesin operon is of interest as lipoproteins have been identified in the gene clusters encoding two other *V. cholerae* colonisation factors: TCP, where the *tcpC* gene encodes a lipoprotein (Parsot *et al.*, 1991; Ogierman and Manning, 1992b); and ACF, a gene cluster also proposed to be involved in colonisation, where *acfD* encodes a lipoprotein (Parsot *et al.*, 1991). Given that the second residue of the mature portion of a lipoprotein determines the ultimate cellular location of the protein, with a negatively charged residue leading to location in the cytoplasmic membrane, while other residues lead to location in the outer membrane (Yamaguchi *et al.*, 1988), all three of these lipoproteins are predicted to be localised in the outer membrane, as VlpA has a leucine residue, TcpC has a serine residue (Ogierman and Manning, 1992b) and AcfD has a lysine residue (Parsot *et al.*, 1991). This possibility led to the proposal that VlpA might play a role in anchoring the MFRHA to the bacterial cell surface (Manning, 1991). Similarly TcpC has been suggested to provide a platform for TCP (Taylor *et al.*, 1988; Ogierman and Manning, 1992b), more specifically, by the interaction between the lipid portion of the protein and the outer membrane, rather than acting as a porin as identified in other characterised fimbrial systems (Ogierman and Manning, 1992b). Tn*phoA* insertion mutants in *tcpC* are defective in TCP biogenesis (Taylor *et al.*, 1988), supporting this assertion, although probable polar effects of Tn*phoA* insertions on downstream genes also required for TCP biogenesis could also explain this result. However, a *vlpA*::Km^R mutation has no effect on MFRHA activity, suggesting that at least in *E. coli* clones, VlpA is not required for anchoring of the MFRHA.

Comparison of the predicted VlpA sequence with sequences in the PIR and SWISS-PROT databases revealed that VlpA shows similarity to the α_2-microglobulin gene family. The biological roles of these proteins are diverse, but all of those characterised to date are eukaryotic soluble transporters of hydrophobic molecules (Sawyer, 1987; Godovac-Zimmermann, 1988). VlpA shows highest similarity with two superfamily members, rat (Spreyer *et al.*, 1990) and human (Drayna *et al.*, 1986; Drayna *et al.*, 1987) ApoD, and Bbp from the white cabbage butterfly *P. brassicae* (Suter *et al.*, 1988), which are porphyrin binding proteins (Huber *et al.*, 1987; Peitsch and Boguski, 1990) and which, with IcyA from the tobacco horn worm *M. sexta* (Riley *et al.*, 1984), form a subgroup of members of the

superfamily that are more closely related to each other than to other superfamily members (Peitsch and Boguski, 1990). However, some important differences between VlpA and these proteins are apparent. The first of these is the probable membrane association of VlpA. Most bacterial lipoproteins (from Gram negative bacteria) seem to associate with the inner face of the outer membrane, although some extracellular enzymes, such as the *Klebsiella pneumoniae* pullulanase are surface exposed (Pugsley, 1989). Most sequenced outer membrane lipoproteins are devoid of potential membrane anchoring sequences other than the amino terminal lipid moieties (Pugsley, 1989) and it is possible that VlpA, being related to globular, soluble proteins, falls into this category. The second difference is the lack of cysteine residues (other than the amino-terminal cysteine that is the presumed site of lipidation) in the mature portion of VlpA. The presence of two disulphide bridges is an important feature for the maintenance of the β-barrel structure of ApoD (Peitsch and Boguski, 1990), Bbp (Huber *et al.*, 1987) and IcyA (Holden *et al.*, 1987), the described porphyrin binding proteins examined here. Both Bbp and IcyA have been shown to exist as tetramers (Huber *et al.*, 1987; Holden *et al.*, 1987) and ApoD forms a dimer (Peitsch and Boguski, 1990). Residues involved in subunit association of both Bbp and IcyA are located in the amino terminal portion of the proteins (Huber *et al.*, 1987; Holden *et al.*, 1987) and it seems possible that the lipidation of VlpA may prevent a similar interaction from occurring. The structure of VlpA therefore presents an interesting problem, and is currently being examined by others.

The results of the sequence comparison suggested a possible function for VlpA, as the best matches were to that subset of the α_2-microglobulin superfamily involved in porphyrin binding (Peitsch and Boguski, 1990). Attempts to detect binding of haemin and the acid diazo dye Congo red, binding of which correlates with haemin binding (Stugard *et al.*, 1989) to *E. coli* cells containing cloned *vlpA* were unsuccessful, and it was only by overproducing the protein after cloning the gene under the control of a bacteriophage T7 promoter that binding of congo red, haemin, haematoporphyrin and, to a lesser degree, protoporphyrin IX, could be detected. The construction of a series of progressive carboxy-terminal deletions fused to the mature portion of PhoA demonstrated that the deletion of a sufficient portion of the protein led to the loss of haemin binding, although approximately one third of the protein was sufficient for binding in this experiment. Removal of the carboxy-terminal portion of the

protein had no effect on haemin binding (plasmids pPM1664 to pPM1667), although the end point in pPM1667 is within a region of residues important for porphyrin binding for both Bbp and IcyA. An intermediate phenotype is displayed by pPM1668 and pPM1669, in which the remainder of the central region of VlpA that aligns with residues important for porphyrin binding in Bbp and IcyA is progressively removed. The first deletion in which haemin binding comparable to the negative control is observed has lost most residues that align with residues involved in porphyrin binding by Bbp and IcyA (pPM1670). Thus the progressive removal of residues in VlpA that align with residues that are important for porphyrin binding in Bbp and IcyA leads to a gradual loss of haemin binding by VlpA-PhoA fusion proteins, suggesting that the similar residues in VlpA might also be involved in ligand interactions. PhoA fusions to a series of carboxy-terminal deletions have been used to study the membrane topology of a number of proteins, for example the mannitol permease of *E. coli* (Sugiyama *et al.*, 1991), but seem to be limited in their application to study of the topology of outer membrane proteins, possibly due to hindrance of translocation of the hydrophilic PhoA portion of the fusion through the outer membrane (Murphy and Klebba, 1989). The isolation of *vlpA-phoA* fusions at the same nucleotide that gave different colony colour on X-pho containing media, despite the apparent stability of the fusion proteins, precluded the possibility of using these fusions to study the topology and location of VlpA. Such studies await the availability of a suitable antiserum.

The activity of VlpA as a haemin (and other porphyrin) binding protein suggested that the protein may be involved in iron uptake. A precedent for this proposal has been set by the identification of a haemin binding lipoprotein in *H. influenzae* (Hanson and Hansen, 1991), and lipoproteins involved in iron uptake in *B. subtilis* (Schneider and Hantke, 1993). However, several pathogenic bacteria have been shown to bind haem without necessarily taking it up. These include: *Aeromonas salmonicida*, where the surface array virulence protein binds congo red, haemin and protoporphyrin IX (Kay *et al.*, 1985); *S. flexneri*, where haemin, protoporphyrin IX and congo red binding are associated with virulence (Daskaleros and Payne, 1987; Stugard *et al.*, 1989); *Y. pestis*, which exhibits haemin and congo red binding, known as the pigmentation (Pgm) phenotype, loss of which leads to avirulence in mice (Perry *et al.*, 1990); and *H. influenzae*, in which a haemin binding lipoprotein has been

identified (Hanson and Hansen, 1991). The role played by porphyrin binding in the pathogenesis of these bacteria is not known, and may well be different in different pathogens. Unlike other pathogens, *H. influenzae* shows an absolute requirement for the porphyrin provided by haemin, although it also utilises iron acquired from haemin (Hanson and Hansen, 1991). Precoating of *S. flexneri* with haemin or congo red enhances the invasion of HeLa cells *in vitro*, prompting the suggestion that haemin binding may facilitate target cell entry by the bacterium (Daskaleros and Payne, 1987; Stugard *et al.*, 1989). *V. cholerae* has been shown to bind Congo red, with mutants unable to do so showing reduced virulence in a chicken embryo model that can be restored by the addition of exogenous iron (Payne and Finkelstein, 1977). Genes that confer the ability to utilise iron from haemin and haemoglobin have been isolated recently from *V. cholerae* (Henderson and Payne, 1993) and encode a 26 kDa cytoplasmic membrane protein and a 77 kDa outer membrane protein - clearly distinct from VlpA. However, it is a possibility that an accessory haemin binding protein may enhance the effectiveness of a haemin utilisation system *in vivo*, where any such system must compete with host haemin binding proteins. It seems unlikely that a role similar to that proposed for haemin binding in *H. influenzae* or *S. flexneri* is fulfilled by VlpA, as *V. cholerae* O1 neither exhibits a requirement for exogenous haemin for growth, nor is invasive (Baumann *et al.*, 1984). Conditions under which *vlpA* is expressed are yet to be determined, and a survey of ToxR-regulated lipoproteins did not identify proteins related to VlpA, so an understanding of the expression of *vlpA* remains a major stumbling block to an understanding of its function.

Probing a sample of *V. cholerae* O1 strains with a *vlpA* specific probe revealed that most strains have at least two copies of *vlpA*-related sequences. The one exception to this, C31, has a chromosomal deletion that encompasses the entire region cloned on plasmid pPM471 (Chapter 4), so that the single sequence detected is not *vlpA*. The presence of multiple copies of *vlpA*-related sequences, when *vlpA* is located in a region of the chromosome known to contain direct repeat sequences implicated in recombination events (Chapter 4), is significant and will be discussed in Chapter 8. Further analysis should reveal whether these related sequences are copies of *vlpA*, whether they are also associated with copies of VCR, and whether any of these related sequences encode active or silent genes.

Chapter 6

Transcriptional organisation of the *mrh* locus encoding the Mannose-Fucose-Resistant Haemagglutinin of *Vibrio cholerae* O1

6.1 Introduction

The pathogenesis of *V. cholerae* O1 requires the coordinated expression of a large number of genes in order to bring about the observed changes as free-living *Vibrios* are ingested, colonise the intestinal epithelium, and secrete toxins (Rabbani, 1986; Manning, 1987; Taylor, 1989). The best characterised of these genes include genes encoding cholera toxin (Mekalanos *et al.*, 1983) and TCP fimbriae (Taylor *et al.*, 1987) which, along with a number of other virulence genes, form a regulon under the control of the ToxRST regulatory system (reviewed in DiRita, 1992). The existence of this regulon, which encompasses genes necessary for both colonisation and toxin secretion, has led to the view that the ToxRST system is the major global regulatory system of virulence in *V. cholerae* (Miller *et al.*, 1989; Parsot and Mekalanos, 1990; DiRita, 1992), but recent results have shown this to be an over-simplification. Expression of the toxin HlyA is regulated both by Fur (Stoebner and Payne, 1988) and HlyU, and appears to be independent of ToxR (Williams and Manning, 1991). Mutation in at least one other Fur regulated gene, *irgA*, attenuates *V. cholerae* (Goldberg *et al.*, 1990b), but it is not yet known whether *fur* mutations are attenuating in a similar manner to a *toxR* mutant (Taylor *et al.*, 1987; Herrington *et al.*, 1988).

Colonisation has been shown to be essential in order for an infection to be established (Schrank and Verwey, 1976), so consequently we are interested in the control of expression of genes that mediate adhesion. A number of haemagglutinins have been identified which are expressed at different phases of *in vitro* growth (Hanne and Finkelstein, 1982), and the *tcp* genes are expressed under different *in vitro* growth conditions in strains of classical and El Tor biotype (Jonson *et al.*, 1990). This differential expression may reflect differential requirements for the adhesins during the course of an infection, as Teppema *et al.* (1987) have suggested that strains differing in haemagglutinating phenotype *in vitro* show some preference for adherence to different sites on the intestinal epithelium, at least in a rabbit tied-loop model.

Chapters 3 and 4 have detailed the genetic organisation of the region surrounding the genes encoding the MFRHA in *V. cholerae*. There is some controversy as to which of two genes encodes the MFRHA: van Dongen *et al.* (1987) have suggested a 7 kDa ORF, *mrhA*, while Franzon and Manning, (1986) have suggested a 25 kDa ORF, *mrhB*. The

transcriptional organisation of these two genes has been determined, as described in this Chapter, and it has been demonstrated that the two genes are cotranscribed, suggesting that both may be important for MFRHA activity.

6.2 Results

6.2.1 The MFRHA locus consists of at least two genes

MFRHA activity has been assigned to two adjacent ORFs located within a 6.3 kb *Bam*HI fragment of *V. cholerae* O1 (Franzon and Manning, 1986; van Dongen and DeGraaf, 1986; van Dongen *et al.*, 1987). By examining the haemagglutinating activity mediated by minimal fragments in *E. coli* K-12, van Dongen *et al.* (1987) assigned MFRHA activity to a 7 kDa protein, here called MrhA, while Franzon and Manning (1986) assigned MFRHA activity to a 15 kDa protein, here called MrhB. The genetic organisation of this region is shown in Fig. 6.1, along with a representation of the subclones used to probe the transcriptional organisation of the region. Subclones were constructed as follows.

A *Bg*/II-*Nru*I fragment of pPM471 (Franzon and Manning, 1986), spanning nucleotides 1978-2554, was sub-cloned into *Sma*I-*Bam*HI digested pSP64 and pSP65 (Melton *et al.*, 1984) to give plasmids pPM1611 and pPM1612, respectively. Digestion of these plasmids with *Acc*I, followed by end repair and religation, leading to deletion of the 0.5 kb *Acc*I fragment, gave pPM1613 and pPM1614, respectively. The 0.7 kb *Cla*I-*Xba*I fragment from plasmid pPM1127 (Franzon and Manning, 1986) was sub-cloned into *Acc*I-*Xba*I digested pSP64 and pSP65 to give pPM1615 and pPM1616, respectively. Plasmid pPM1618 was constructed by subcloning a *Dra*I fragment identified by dot blot hybridisation with oligo #61 (Chapter 3), spanning nucleotides 2387-3032 in the 6.3 kb *Bam*HI fragment of pPM471, into pK18 (Pridmore, 1987). An *Ahu*I-*Fok*I fragment internal to this *Dra*I fragment was subcloned into: the *Hinc*II site of pGEM3Zf$^+$ (Promega) to give pPM1630; the *Sma*I site of pPM3024 (Williams and Manning, 1991) to give pPM1632 and pPM1633 (both orientations of insert); and the *Sma*I site of pKL600 (K. McKenny, unpublished) to give pPM1634 and pPM1635 (both orientations of insert). Subclones were obtained by ligating *Ahu*I-*Fok*I digested pPM1618 DNA with vector, and then screening progeny colonies by colony blot using oligo #61 to identify subclones containing the appropriate insert. Subclones were

Figure 6.1: Schematic representation of the plasmid pPM471, with the *mrhAB* region and derived subclones expanded.

Numbers refer to the distance from the proximal end of the clone in kb and some restriction sites in the region are indicated. The genes *mrhA* and *mrhB* are represented by open boxes, and two flanking copies of the direct repeat VCR are represented by hatched boxes. The portions subcloned are represented by solid black lines or arrows, annotated with plasmid number. In the case of subclones in the vectors pSP64 and pSP65, arrows represent the polarity of the subcloned region with respect to the SP6 promoter of the vectors. In the case of pPM1630, the arrow represents polarity with respect to the T7 promoter of the vector. In all pPM3024 and pKL600 subclones, arrows represent the orientation of insert with respect to the vector *galK* gene. Details of plasmid construction are in the text. The binding site and orientation of oligos are indicated by small arrows.
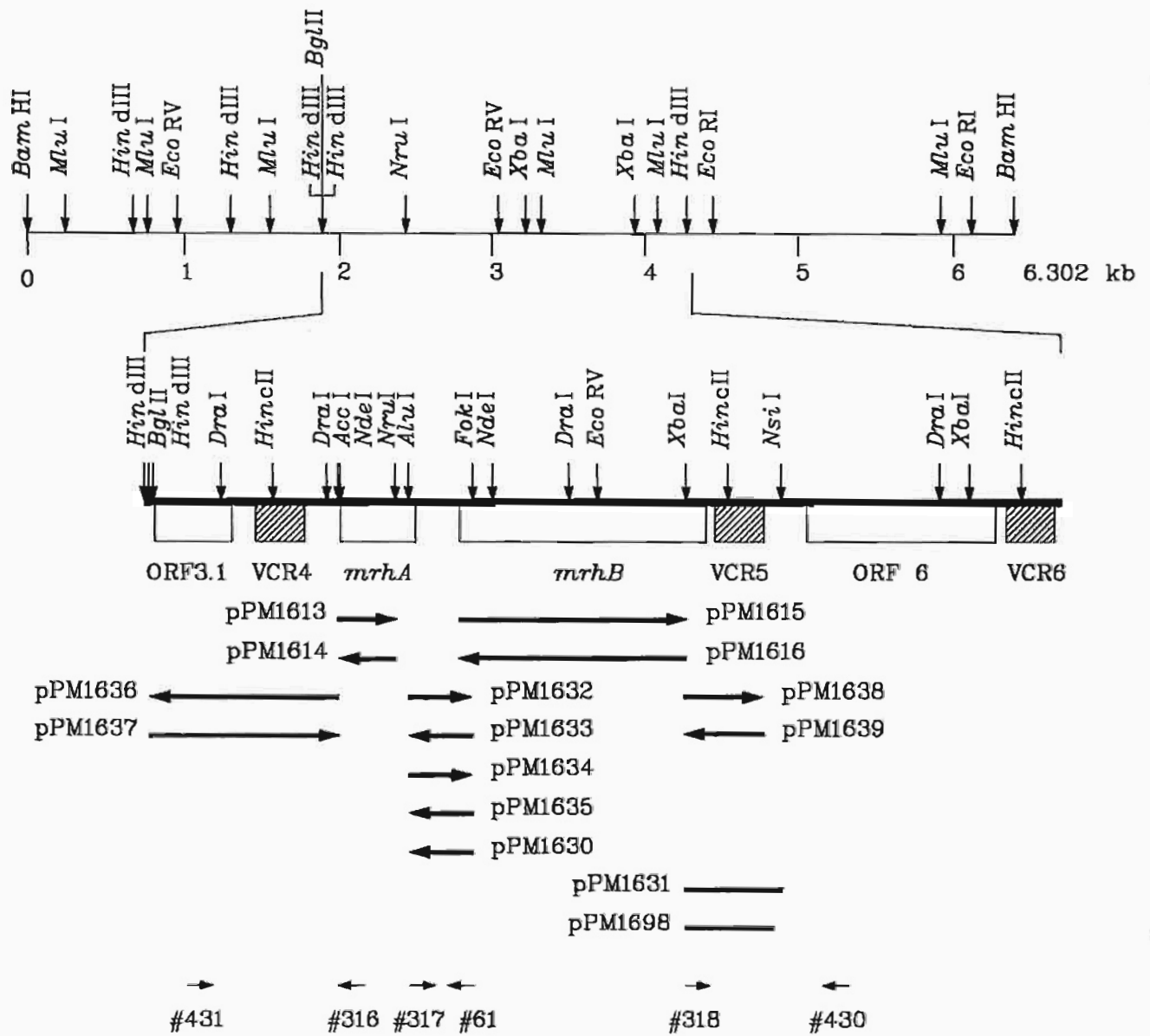
confirmed by DNA sequencing. Plasmids pPM1636 and pPM1637 were generated by subcloning the (end filled) *Acc*I fragment of pPM1611 into the *Sma*I site of pPM3024, and represent both orientations of this insert (as confirmed by restriction analysis). Plasmid pPM1631 was generated by subcloning a *Xba*I-*Nsi*I fragment (nucleotides 3316-3552 in pPM471 insert) into plasmid pGEM7Zf⁺ (Promega). Plasmid pPM1698 contains an *Xba*I fragment internal to this insert, derived as follows. Plasmid pPM1631 was linearised with *Nsi*I, and digested with *Bal*31 slow form (Wei *et al.*, 1983). Digested samples were end repaired and religated with *Xba*I linkers. The mixture was then digested with *Xba*I and the digested fragment representing the 0.2 kb *Xba*I-*Nsi*I fragment of pPM1631 purified from a PAGE gel and ligated with pK18 (Pridmore, 1987). Plasmid DNA was isolated from twelve resulting colonies and analysed by DNA sequencing (not shown). One plasmid, in which nucleotides from 3524-3550 had been replaced with an *Xba*I linker, was picked and named pPM1698. The 0.25 kb *Xba*I fragment from pPM1698 was subcloned into pKL600 to give pPM1638 and pPM1639 (both orientations of insert), which were confirmed by DNA sequencing (not shown). Plasmids pPM1640 and pPM1641 were generated by subcloning the 300 bp *Pvu*II fragment of pUC19 bearing p_{lacUV5} (Yanisch-Perron *et al.*, 1985), into pPM3024. Plasmids were recovered from red, chloramphenicol-sensitive colonies on McConkey plates (pPM1640) or white, chloramphenicol-resistant colonies (pPM1641) and were confirmed by restriction mapping (not shown).

6.2.2 Analysis of transcription in the *mrhAB* region

Northern blots were performed on total RNA extracted from *E. coli* HB101 harbouring a MFRHA positive clone, pPM471 (Franzon and Manning, 1986) and are shown in Fig. 6.2. When a *mrhA* specific probe was used, two hybridising species were observed with HB101[pPM471]: an abundant one of approximately 400 nucleotides, and a much weaker one of approximately 1200 nucleotides (Fig. 6.2A). When a *mrhB* specific probe was used, a weak signal of approximately 1200 nucleotides was observed (Fig 6.2B). No signals were observed with HB101[pBR322] (Fig. 6.2). This suggests that the two genes are co-transcribed. To see if this was also the case in *V. cholerae*, northern blots were performed on total RNA extracted from four different strains of both biotypes (Fig. 6.3). Again, a *mrhA* specific probe hybridised to two RNA species in three of the four strains tested, an abundant

Figure 6.2: Northern blot of RNA extracted from *E. coli* HB101 harbouring an MFRHA expressing clone.

Panel A is probed with a *mrhA* specific probe and panel B is probed with a *mrhB* specific probe. Probes are derived from plasmids pPM1614 and pPM1616, respectively. Lanes are total RNA prepared from: 1, HB101[pBR322]; 2, HB101[pPM471].
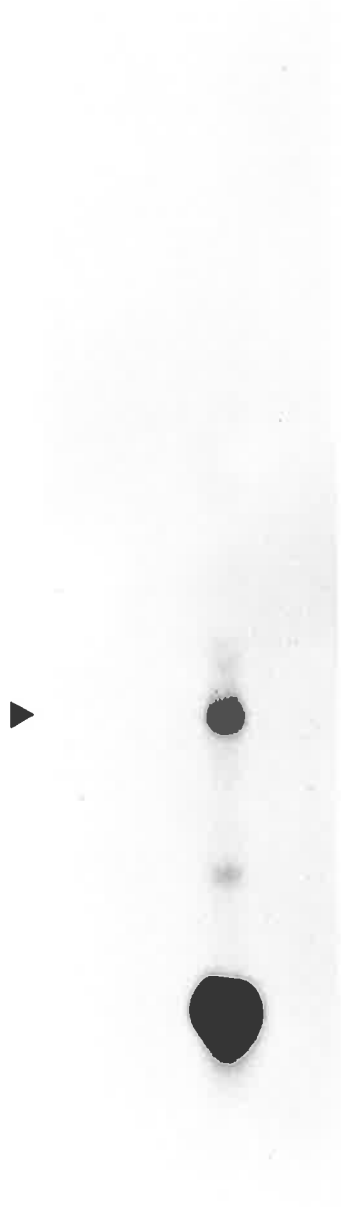
Figure 6.3: Northern blot of RNA extracted from four strains of *V. cholerae* O1.

Panel A is probed with a *mrhA* specific probe and panel B is probed with a *mrhB* specific probe. Probes are derived from plasmids pPM1614 and pPM1616, respectively. Lanes are total RNA prepared from *V. cholerae*: 1, 569B; 2, Z17561; 3, O17; 4, C31.

A 1 2 3 4   B   1 2 3 4

►   ►

one of approximately 400 nucleotides and a weak signal of about 1200 nucleotides (Fig. 6.3A), and a *mrhB* specific probe hybridised to a single weak signal of about 1200 nucleotides (Fig. 6.3B). No signal was observed for strain C31, in which these genes have been deleted (van Dongen and DeGraaf, 1986; Chapter 4). The signals observed are much stronger for the *E. coli* clone than for *V. cholerae*, which is presumably a gene dosage effect due to differences in copy number between the clone and chromosomal loci. In all cases, no hybridisation is observed with probes complementary to those used in Figures 6.2 and 6.3 (data not shown).

6.2.3 Identification of a unique promoter for *mrhA* and *mrhB*

In order to detect the 5' ends of any transcripts in this region, primer extension assays were performed. Extension products consistent with a promoter lying immediately 5' to *mrhA* were obtained with two primers: oligo #316 gave an extension product of 113 nucleotides in length both with *E. coli* HB101[pPM471] and with *V. cholerae* 569B (Fig. 6.4A), and oligo #61 gave extension products for both *E. coli* LE392[pPM471] and *V. cholerae* 569B of approximately 360 nucleotides (S.G. Williams, personal communication). Both of these are consistent with a unique 5' end at nucleotide 2393 for transcripts covering both *mrhA* and *mrhB*. Sequences showing a close match to the consensus sequence for *E. coli* σ^{70} promoters (O'Neill, 1989) can be identified upstream of this site (Fig. 6.4B) although the spacing between the putative -35 and -10 boxes (14 bp) is significantly shorter than the consensus. To ascertain whether this mapped 5' end corresponded to a promoter, plasmids pPM1637, pPM1636, pPM1632 and pPM1633, which are based on the promoter detection vector pPM3024 (Section 6.2.1), were introduced into *E. coli* CB806. Positive and negative controls for this experiment were provided by plasmids pPM1640 and pPM1641, which contain the 300 b.p. *Pvu*II fragment of pUC19 containing P_{lacUV5} (Yanisch-Perron *et al.*, 1985) in pPM3024 (Section 6.2.1). Appropriate orientation of a fragment bearing a promoter in pPM3024 leads to galactokinase expression, when introduced into a *galK* background (Schneider and Beck, 1986). Thus, galactokinase activity was determined for each construct after introduction into *E. coli* CB806 and the values are shown in Table 6.1. Plasmid pPM1640 gave high galactokinase activity and pPM1641 gave negligible activity (Table 6.1) consistent with the orientation of p_{lac} in these subclones (Yanisch-Perron *et al.*, 1985).

Figure 6.4: 5' end mapping of *mrhAB* transcripts.

Panel A shows a primer extension assay using oligo #316. Lanes are RNA purified from: 1, *E. coli* HB101[pBR322]; 2, *E. coli* HB101[pPM471]; 3, *V. cholerae* C31; 4, *V. cholerae* 569B. DNA size markers are a sequence ladder generated from M13/1609 using oligo #316 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix. Panel B shows the sequence of the region around the mapped 5' end, with +1 and predicted -35 and -10 boxes, as well as the start codon of *mrhA* in bold. The *E. coli* σ^{70} -35 and -10 consensus sequences (O'Neill, 1989) are shown for comparison.

A

B
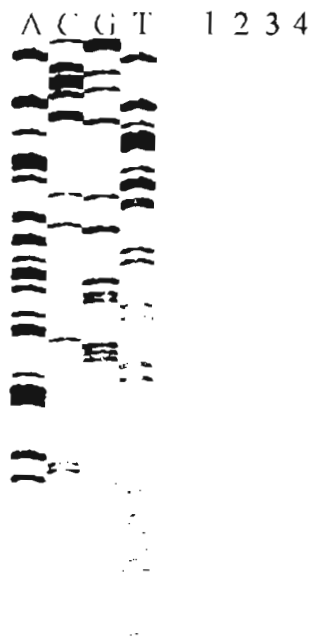
```
                2360            2370            2380            2390
                 |   -35          |               |   -10          |
AGCAATTGCTTGATATTCTCCATATTGCCCATATTTAAAT
Consensus        TTGACA              17          TATAAT
```

```
                2400            2410            2420
       +1         |               |               |
TATGGTCATTAAATAAAAGGAGTAGACCATATG...mrh4...
```

Table 6.1

Galactokinase activity of subclones in the promoter detection vector pPM3024 in *E. coli* CB806.

Plasmid	colony colour on McConkey agar	GalK activity[a] (Units)	Plasmid yield[a,b]
pPM3024	white	1.5±1.1	1.0±0.08
pPM1632	white[c]	25±6.1	0.9±0.08
pPM1633	white[c]	35±8.2	0.94±0.03
pPM1636	red	101±28	1.0±0.01
pPM1637	red	150±4.1	1.0±0.09
pPM1640	red	556±114	0.64±0.07
pPM1641	white	<1	0.78±0.03

[a] The results presented are the mean and standard deviation for assays performed on three independent transformants of each plasmid.

[b] Plasmid yield is expressed as μg/ml of culture of A_{650} 1 and is the mean and standard deviation for assays performed on three independent transformants of each plasmid.

[c] White single colonies turn red on extended incubation.

Plasmid pPM1637 gave high galactokinase activity (Table 6.1), indicating the presence of a promoter on this insert, which was called p_{mrh}. The opposite orientation, pPM1636, also gives elevated galactokinase activity, but the presence of a divergent ORF immediately upstream of *mrhA* (Chapter 3) is consistent with activity from this construct (see also Section 6.2.7). Both pPM1632 and pPM1633 also give galactokinase activity, albeit significantly lower than for pPM1637 (Table 6.1), suggesting that there may be a promoter in the *mrhA-B* intergenic region. Examination of the sequence of the constructs reveals that in both cases promoters may have been constructed by the subcloning. In an attempt to locate these predicted promoters, primer extensions were performed using total RNA extracted from *E. coli* CB806 containing the plasmids pPM3024, pPM1632, pPM1633, pPM1636 and pPM1637, and oligo #362, a primer that binds within *galK* (de Boer, 1984). An extension product of 73 nucleotides was observed for plasmid pPM1637 (Fig. 6.5), which is in agreement with the 5' end mapped for p_{mrh} (Fig. 6.4). Extension products were not apparent with RNA extracted from cells containing the other plasmids (Fig. 6.5), so that 5' ends could not be mapped by this method for these constructs. The plasmid copy-number was determined from the same cultures as galactokinase activity. All plasmids show approximately the same copy number in the cultures assayed for galactokinase (Table 6.1).

6.2.4 Determination of transcription termination sites

6.2.4.1 Subcloning into the terminator detection vector pKL600

The marked difference in abundance of *mrhA* and *mrhB* transcripts observed in Figures 6.2 and 6.3 cannot be explained by differential activity of two promoters (Section 6.2.3). Therefore some other mechanism must lead to the observed difference. Two possibilities are: that there is a terminator that is "read through" lying between the two ORFs; or that the *mrhA* transcript is more stable than the *mrhAB* transcript. To distinguish between these possibilities, plasmids pPM1634, pPM1635, pPM1638 and pPM1639, which are based on the terminator detection vector pKL600 (Section 6.2.1), were introduced into *E. coli* CB806. This vector contains p_{lac} driving transcription of *galK*, with the two separated only by a multiple cloning site. Introduction of a fragment containing a terminator leads to a reduction in galactokinase activity when the construct is introduced into a *galK* background (K. McKenny, unpublished results). Galactokinase activity was determined for each

Figure 6.5: Primer extension analysis of *galK* reporter gene fusion constructs.

The primer extension assay was performed using oligo #362, specific for *galK* (de Boer, 1984). Lanes are RNA purified from *E. coli*: 1, CB806[pPM3024]; 2, CB806[pPM1632]; 3, CB806[pPM1633]; 4, CB806[pPM1636]; 5, CB806[pPM1637]. DNA size markers are a sequence ladder generated from pPM3024, using oligo #362 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix.

A C G T 1 2 3 4 5

construct after introduction into *E. coli* CB806 and the values obtained are shown in Table 6.2. Plasmid pPM1634 gave no detectable galactokinase activity, indicating the presence of a terminator on this insert. A reduction in galactokinase activity was also observed with pPM1635, where the insert is cloned in the opposite orientation (75% reduction of galactokinase activity by comparison with pKL600) (Table 6.2). Plasmid pPM1638 also shows a significant reduction in galactokinase activity (63% reduction), but the opposite orientation, pPM1639, shows only a slight reduction in galactokinase activity (32% reduction). Determination of plasmid copy number in the cultures used for galactokinase assays indicated that the copy number was substantially reduced in all cases where significant galactokinase expression occurred (Table 6.2). This reduction in copy number explains the low galactokinase activity as measured for pKL600 in comparison to pPM1640 (Table 6.1), both of which contain p$_{lac}$ driving *galK*. The reason for this observation is not known.

6.2.4.2 3'-end mapping by RNase protection assay

In order to map the predicted 3' ends, a RNase protection assay was used. Riboprobes generated from plasmid pPM1630, which should anneal to the *mrhA-B* intergenic region, gave several major protected products of 97-111 nucleotides, given that RNA migrates 5-10% slower than DNA under the electrophoretic conditions used (Sambrook *et al.*, 1989) (Fig. 6.6A). These protected products were observed using RNA from both *V. cholerae* 569B and *E. coli* HB101[pPM471], and are consistent with a major 3' end localised to nucleotides 2684-2692. Immediately 5' to this mapped end is a predicted secondary structure, shown in Fig. 6.6B, with a free energy of -19.8 kCal.Mol.$^{-1}$. This structure is predicted to act as a "leaky" Rho-independent terminator, as the stem-loop is followed by a U-rich tail (Fig. 6.6B), and because of the absence of recognisable features characteristic of Rho-dependent terminators (Alifano *et al.*, 1991). A protected product of 192-202 nucleotides and of much weaker intensity, which corresponds to "full-length" *mrhAB* transcript, is also apparent (Fig. 6.6A).

In an attempt to map the 3' end of the *mrhAB* transcript, riboprobes generated from plasmid pPM1631 were used in a RNase protection assay in which multiple bands were observed with both polarities of probe (Fig. 6.7). As these probes hybridise with VCR, a reiterated sequence present nine times in the *E. coli* clone and at least 60 times in *V. cholerae*

Table 6.2

Galactokinase activity of subclones in the terminator detection vector pKL600 in *E. coli* CB806.

Plasmid	colony colour on McConkey agar	GalK activity[a] (Units)	Plasmid yield[ab]
pKL600	red	158±23	<0.1[c]
pPM1634	white	<1	0.97±0.06
pPM1635	red	40±7.3	<0.1[c]
pPM1638	red	59±4.5	<0.1[c]
pPM1639	red	108±9.6	<0.1[c]

[a] The results presented are the mean and standard deviation for assays performed on three independent transformants of each plasmid.

[b] Plasmid yield is expressed as μg/ml of culture of $A_{650} = 1$ and is the mean and standard deviation for assays performed on three independent transformants of each plasmid.

[c] Plasmid bands were detectable on ethidium bromide stained agarose gels (not shown) but were of insufficient intensity for quantitation.

Figure 6.6: 3' end mapping of *mrhA* transcripts.

Panel A shows a RNase protection assay using probes generated from plasmid pPM1630: complementary to *mrhAB* mRNA (lanes 1 to 6); and of the same sense as *mrhAB* mRNA (lanes 7 to 12). Lanes contain target RNAs: 1 and 7, undigested probe; 2 and 8, yeast RNA; 3 and 9, *V. cholerae* 569B; 4 and 10, *V. cholerae* C31; 5 and 11, *E. coli* HB101[pBR322]; 6 and 12, *E. coli* HB101[pPM471]. DNA size markers are a sequence ladder generated from pPM1630, using oligo #317 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix. Panel B shows a predicted RNA secondary structure of free energy = -19.8 kCal.Mol^{-1}, located within -2 to 12 nucleotides of the mapped 3' end.

A

A C G T   1 2 3 4 5 6 7 8 9 10 11 12

B

5'
U
A
U
A
G U
G
U A-A-U-G-A-C-A-G-A-U-A-C-G   A-U
                                      U U
U-U-A-C-U-G-U-C-U-A-U-G-C   A-A-U
C
U
A
U
G
G
U
U
3'
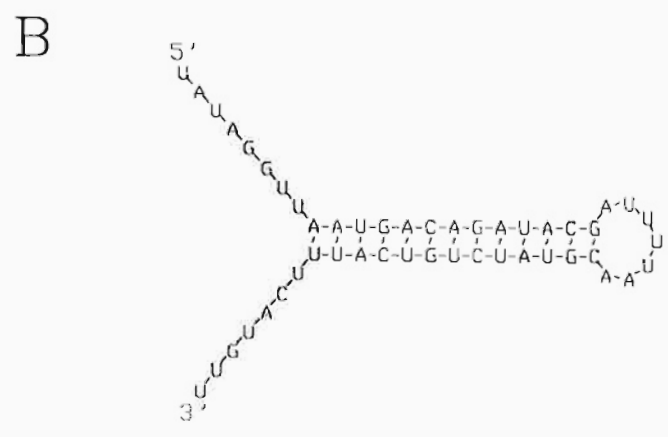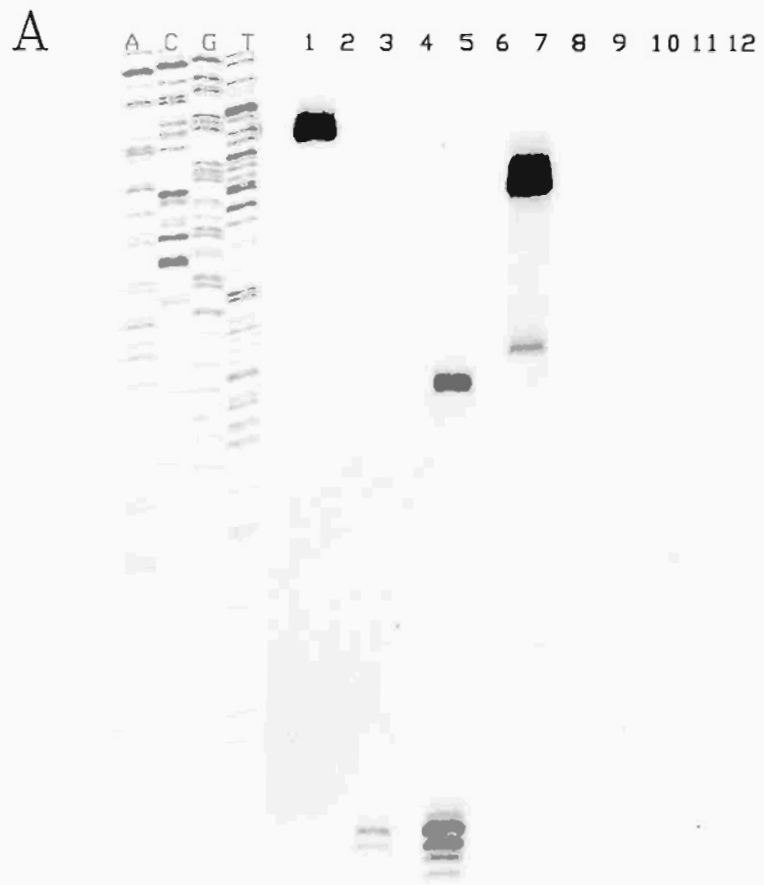
Figure 6.7: RNase protection assay using a VCR containing probe.

Probes used were generated from plasmid pPM1631: complementary to *mrhAB* mRNA (lanes 1 to 6); and of the same sense as *mrhAB* mRNA (lanes 7 to 12). Lanes contain target RNAs: 1 and 7, yeast RNA; 2 and 8, undigested probe; 3 and 9, *E. coli* HB101[pBR322]; 4 and 10, *E. coli* HB101[pPM471]; 5 and 11, *V. cholerae* C31; 6 and 12, *V. cholerae* 569B. DNA size markers are a sequence ladder generated from pPM1631, using oligo #318 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix.

A C G T   1 2 3 4 5 6 7 8 9 10 11 12
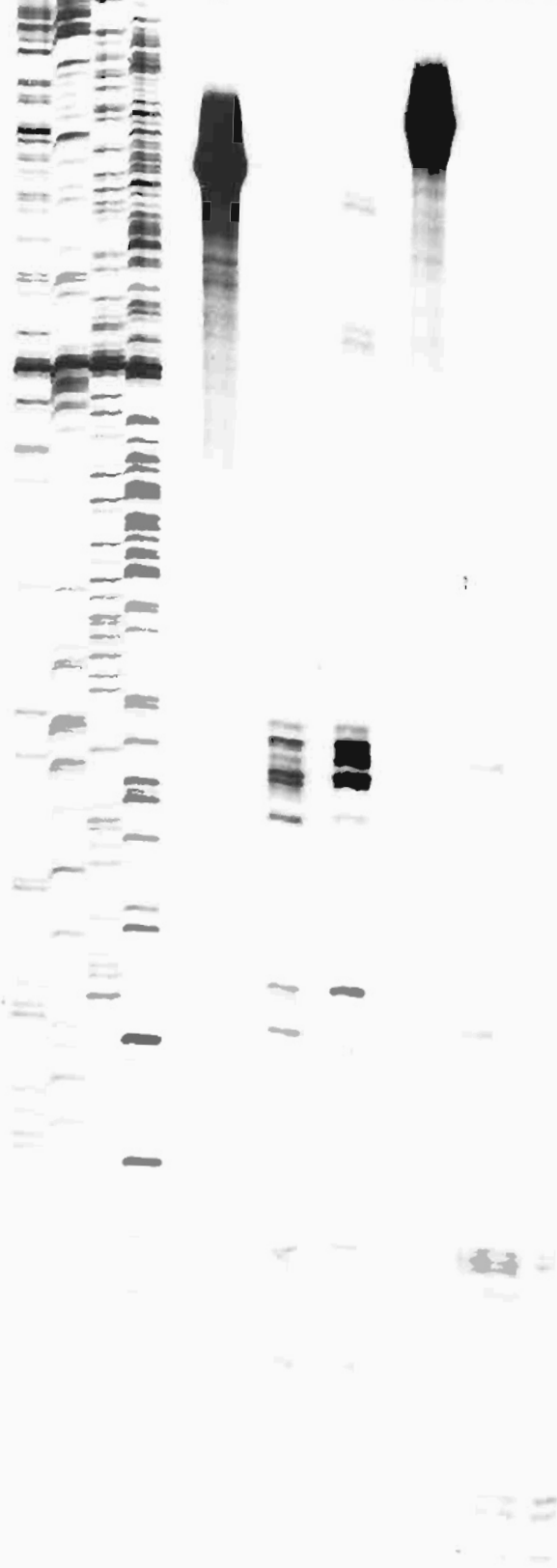
(Chapters 3 and 4), this result could be due either to multiple ends originating from multiple copies of VCR, or to partial protection of the probes by VCR, which contains an imperfect dyad symmetry (Chapter 4). The lack of any signal in the yeast RNA lane (Fig. 6.7) indicates that the probe is not self protecting. In order to successfully map 3' ends associated with the *mrhAB* transcript, total RNA was extracted from cells containing the terminator-probes pPM1638 and pPM1639 and was used in a RNase protection assay. Two major protection products were observed for sense strand probe and pPM1638 derived RNA: a 199-210 nucleotide product, corresponding to full length or "read-through" transcript; and four products of similar length, the major one being of 117-124 nucleotides, which corresponds to a 3' end mapping to nucleotides 3432-3439 (Fig. 6.8A). Analysis of the sequence 5' to this mapped end reveals that a weak secondary structure is capable of forming. This structure has a predicted free energy change of -15.2 kCal.Mol^{-1} (Fig. 6.8B) and is predicted to act as a weak Rho-independent terminator because of the lack of a C-rich G-poor region downstream of the stop codon, proposed as the signal for Rho interaction with mRNA (Alifano *et al.*, 1991), although the tail of this structure is not markedly U-rich. Comparison of the result obtained in Fig. 6.8A with that of Fig. 6.7 indicates that the most abundant bands observed in Fig. 6.7, are also of 117-124 nucleotides. These bands could either correspond to the 3' ends of *mrhB* transcripts, or to other full-length copies of VCR. Distinguishing between these explanations would be possible only if 3' end labelled probes were used, and this was not attempted. A weak protected fragment of 145-153 nucleotides is also observed with RNA from *E. coli* HB101[pPM471] (Fig. 6.9), but not for *V. cholerae* 569B or for *E. coli* CB806[pPM1638] (Fig. 6.10). A stable secondary structure ($\Delta G = -39.1$ kCal.Mol.$^{-1}$) with a U-rich tail is predicted to lie immediately upstream of this 3' end (Fig. 6.8C) which is also predicted to act as a Rho-independent terminator.

6.2.5 Is *mrhAB* transcription controlled by iron concentration?

The patterns of transcription of *mrhA* and of the iron-starvation-inducible haemolysin gene *hlyA* (Stoebner and Payne, 1988) in a variety of laboratory media have been analysed and very similar results were obtained for *mrhA* and *hlyA* (S. G. Williams, personal communication). Analysis of the sequence around the identified p$_{mrh}$ revealed a 17/19 bp match to a predicted Fur box associated with the *hlyA* promoter (Williams and Manning,

Figure 6.8: 3' end mapping of subclones in the terminator detection vector pKL600.

Panel A shows a RNase protection assay using probes generated from plasmid pPM1631: complementary to *mrhAB* mRNA (lanes 1 to 5); and of the same sense as *mrhAB* mRNA (lanes 6 to 10). Lanes contain target RNAs: 1 and 6, undigested probe; 2 and 7, yeast RNA; 3 and 8, *E. coli* CB806[pKL600]; 4 and 9, *E. coli* CB806[pPM1638]; 5 and 10, *E. coli* CB806[pPM1639]. DNA size markers are a sequence ladder generated from pPM1631, using oligo #318 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix. Panel B shows a predicted RNA secondary structure of free energy = -15.2 kCal.Mol^{-1}, located within 2 to 9 nucleotides of the mapped 3' end. Panel C shows a predicted RNA secondary structure of free energy -39.1 kCal.Mol^{-1}, located within 0 to 9 nucleotides of a mapped 3' end observed only in *E. coli* HB101[pPM471] (Figure 6.7).

A

A C G T   1 2 3 4 5   6 7 8 9 10

B   ENERGY −15.2

C   ENERGY −39.1

Figure 6.9: Examination of the effect of iron-starvation on *hlyA* and *mrhA* transcription.

Panel A shows a match between the predicted Fur box of *hlyA*, a sequence in the promoter region of *mrh*, and the consensus sequence for the *E. coli* Fur box (de Lorenzo *et al.*, 1987). The sequences shown are from nucleotides 2385 - 2403 of the *mrh* sequence and from nucleotides 545 - 563 of the *hlyA* sequence (Alm *et al.*, 1988). Panel B shows total RNA isolated from a *V. cholerae* 569B culture grown in the absence or presence of 2,2'dipyridyl for the times indicated, probed with a *hlyA*-specific probe (Williams and Manning, 1991). Control samples are the indicated amount of the 90 minute sample. Panel C shows samples identical to those of Panel B probed with a *mrhA*-specific probe derived from plasmid pPM1614.

A

mrh	T T A A A T T A T G G T C A T T A A A
hlyA	T T A A A T T A G G C T C A T T A A A
E. coli	G A T A A T G A T A A T C A T T A T C
	* * * * * * * * * *

B

	0' 30' 60' 90'	5 0.5 0.05 µg
no added 2,2' dipyridyl		
+ 0.25mM 2,2' dipyridyl		

C

	0' 30' 60' 90'	5 0.5 0.05 µg
no added 2,2' dipyridyl		
+ 0.25mM 2,2' dipyridyl		
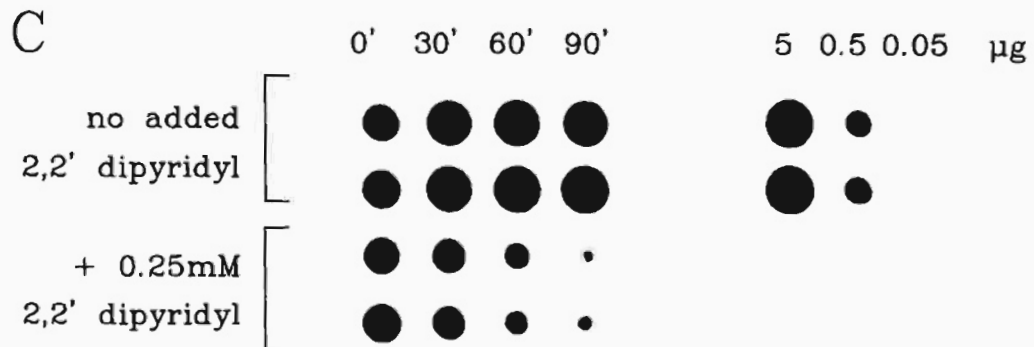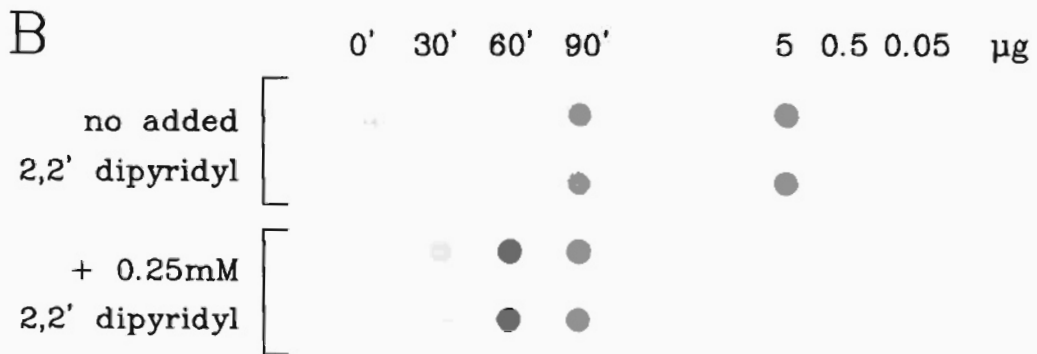
Figure 6.10: 5' end mapping of transcripts covering ORF6.

Panel A shows a primer extension assay using oligo #430. Lanes are RNA purified from: 1, *E. coli* HB101[pBR322]; 2, *E. coli* HB101[pPM471]; 3, *V. cholerae* C31; 4, *V. cholerae* 569B. DNA size markers are a sequence ladder generated from pPM471 using oligo #430 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix. Panel B shows the sequence of the region around the mapped 5' end, with +1 and predicted -35 and -10 boxes, as well as the start codon of the ORF in bold. The *E. coli* σ^{70} -35 and -10 consensus sequences (O'Neill, 1989) are shown for comparison.

A

A C G T   1 2 3 4

B
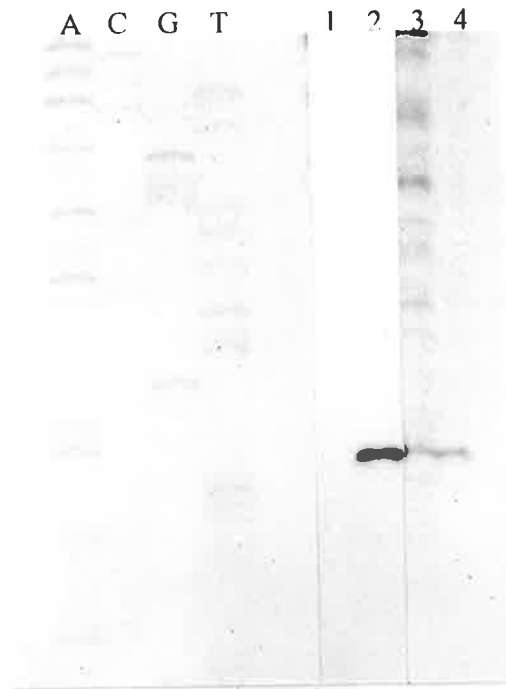
```
          3540              3550              3560              3570
           |               -35   |                 |        -10    |
GCTTTATTTTTGTTGATGCATTCTCCCTGATTATGTATAAG
Consensus       TTGACA            17        TATAAT


          3580                        3640
  +1       |                          |
TCGTTTTGA- 50 -CATGGAGAATAAAGGTG...ORF6...
```

1991), as shown in Fig. 6.9A. Therefore, the effect of iron-starvation on transcription from P_{mrh} was examined. Total RNA was isolated from *V. cholerae* O17 at various time points after treatment with the iron-chelator 2,2' dipyridyl and a dot-blot performed, using an *hlyA*- or *mrhA*-specific probe (Fig. 6.9B and C, respectively) on the same samples. Fur regulation of *hlyA* transcription is at the level of transcription (S. G. Williams, personal communication) and good induction of *hlyA* specific transcription was seen under these conditions (Fig. 6.9B). However, *mrhA* transcription seems to be repressed under conditions of iron starvation (Fig. 6.9C). Although Fur has recently been shown to activate expression of some genes in *S. typhimurium* (Foster and Hall, 1992), this decrease in *mrhA* transcription is more likely to be the result of an indirect effect of iron starvation than loss of Fur activation (see Section 6.3).

6.2.6 MFRHA expression is also independent of other known regulators of virulence in *V. cholerae*

In addition to the Fur regulon, two other regulators of virulence gene expression have been identified in *V. cholerae* - the well characterised ToxRST regulon (for example, DiRita, 1992) and HlyU (Williams *et al.*, 1993). In order to test whether either of these regulate expression of MFRHA activity, the haemagglutination titres of *toxR* mutants and a *hlyU* mutant were determined and compared with the parental strains (Table 6.3). No difference in titre is apparent for any of the strains tested, indicating that MFRHA expression is independent of both of these regulators, under the conditions used here.

6.2.7 Mapping of two putative promoters adjacent to the *mrhAB* operon

As a preliminary to analysis of a possible link in expression between *mrh* and adjacent genes, the putative promoters for ORF6 and ORF3.1 were mapped using a primer extension assay. Oligo #430, which is complementary to nucleotides 3655-3672, gives an extension product of 96 nucleotides with RNA from both *E. coli* HB101[pPM471] and *V. cholerae* 569B, although the band intensity for the *E. coli* clone is much stronger (Fig. 6.10A). This 5' end maps to nucleotide 3576 (Fig. 6.10B). An extension product is not apparent for *V. cholerae* C31, consistent with this strain having a deletion within this region (van Dongen and DeGraaf, 1986; Chapter 4). Oligo #431, which is the same as nucleotides 2090-2107, gives an extension product of 95 nucleotides with *E. coli* HB101[pPM471] (Fig. 6.11A), suggesting

Table 6.3

Haemagglutination titres of *toxR* and *hlyU* mutant *V. cholerae* strains.

Strain	Relevant genotype	Haemagglutination titre[a]
O17	wild type	128
V795	*toxR*::Km[R]	128
V876	*hlyU*::Km[R]	128
Z17561	wild type	64 (32)[b]
V885	*toxR*::Km[R]	64

[a] Haemagglutinations were performed on three independent cultures and are presented as the reciprocal of the highest dilution still giving visible agglutination.

[b] A number in brackets indicates a single sample that gave a different result to the majority.

Figure 6.11:   5' end mapping of transcripts covering ORF3.1.

Panel A shows a primer extension assay using oligo #431. Lanes are RNA purified from: 1, *E. coli* HB101[pBR322]; 2, *E. coli* HB101[pPM471]. DNA size markers are a sequence ladder generated from pPM471 using oligo #431 as a primer. Lanes are labelled A, C, G, and T to indicate the appropriate 2',3'-dideoxynucleotide in the reaction mix. Panel B shows the sequence of the region around the mapped 5' end, with +1 and predicted -35 and -10 boxes, as well as the start codon of the ORF in bold. The *E. coli* σ^{70} -35 and -10 consensus sequences (O'Neill, 1989) are shown for comparison.

A

| A | C | G | T | 1 | 2 |

B

2220   2210   2200   2190
| -35 | | -10 | +1

TTTGTTAAACTGCATTTTCAAGACTATAATGACCCAATTA

Consensus   TTGACA   17   TATAAT

2180   2170   2160
| | |

AAGACCACATAAAGGACCGATTAAATG...ORF3.1...

a 5' end that maps to nucleotide 2185 (Fig. 6.11B) but a ladder with both *V. cholerae* 569B and C31 (not shown). The reason for this discrepancy is not clear. In both cases sequences can be found upstream of the mapped 5' ends that show similarity to the consensus for *E. coli* σ^{70} promoters (O'Neill, 1989), although the spacing between the predicted -35 and -10 boxes is shorter (14 bp) than the consensus for the promoter suggested upstream of ORF6 (Fig. 6.10B). The significance of this spacing will be discussed in Section 6.3.

6.3 Discussion

In this Chapter, the pattern of transcription of the *mrhAB* region has been determined, and is summarised as follows. Transcription initiates at a promoter, p_{mrh}, proximal to *mrhA*, which is abundantly transcribed. Transcription from p_{mrh} terminates close to a putative stem loop structure between *mrhA* and *mrhB*, with some "readthrough" occurring to give transcription of *mrhB*. This full-length transcript terminates within a conserved repeat sequence (named VCR) downstream of *mrhB*, albeit not with great efficiency. An explanation for the "weakness" of this terminator as measured may be found in the low level of transcription of *mrhB*, as terminator efficiency can be directly linked to promoter efficiency (Jaquet and Reiss, 1992). This is an interesting possibility because if the conserved repeat VCR acts generally as a transcriptional terminator, then we can predict that genes whose expression rely on this terminator activity would be expressed at low levels. Not inconsistent with this is the observation that most proteins encoded on pPM471 are poorly expressed in *E. coli* minicells (Franzon and Manning, 1986) and all except for *mrhA* have VCR sequences at their 3' ends (Chapter 3), although we have not determined whether or not any of these other copies of VCR show terminator activity.

The transcriptional organisation described here contrasts with that seen in fimbrial operons in *E. coli*, where differential decay of a primary transcript ensures unequal abundance of the mRNAs of cotranscribed genes whose products are required in unequal amounts (Båga *et al.*, 1988; Nilsson and Uhlin, 1991; Bilge *et al.*, 1993). Processing of a primary transcript has also been proposed to occur in the *tcp* fimbrial gene cluster of *V. cholerae* O1 (Higgins *et al.*, 1992). Preliminary experiments indicate that the half life of *mrhA* and *mrhAB* mRNA in *V. cholerae* 569B is short (not shown), suggesting that the intercistronic stem loop described

is not a significant barrier to 3'-5' exonuclease degradation, unlike some stem loop structures in *E. coli* (Newbury *et al.*, 1987). The possibility of this "attenuation" was suggested by van Dongen *et al.* (1987), and its occurrence explains, in the absence of data on the translational efficiencies of the two genes, the relative abundance of the two proteins MrhA and MrhB (Franzon and Manning, 1986; van Dongen and DeGraaf, 1986; van Dongen *et al.*, 1987).

The expression of a number of *V. cholerae* virulence determinants has been shown to be regulated by one of a few regulatory factors, including, the ToxRST regulon (DiRita, 1992), iron stress (Stoebner and Payne, 1988; Goldberg *et al.*, 1990b), and HlyU (Williams *et al.*, 1993). The presence of a 17/19 bp match to a predicted "Fur box" upstream of the Fur-regulated *hlyA* (Stoebner and Payne, 1988) around the +1 of p_{mrh} led to the suggestion that *mrhAB* may be part of the Fur regulon. However, RNA dot blots performed with samples from cultures stressed for free iron show that while the Fur regulated gene *hlyA* is activated, *mrhAB* is actually repressed under these conditions. Fur has recently been shown to activate the expression of some genes in *S. typhimurium* (Foster and Hall, 1992), offering a possible explanation for this result, but this is considered unlikely as the proposed Fur-binding site overlaps the -10 and +1 region of p_{mrh} (Fig. 6.9A) and Fur binding would therefore be predicted to sterically hinder RNA polymerase association with the promoter. ToxR and HlyU do not appear to regulate transcription from p_{mrh} either, as mutations in both *toxR* and *hlyU* do not affect the haemagglutination titre of strains O17 (*toxR* and *hlyU*) and Z17561 (*toxR*). This is consistent with the observation that the introduction of a *toxR* clone does not alter the haemagglutination titre of cloned *mrh* in *E. coli* (van Dongen and DeGraaf, 1986) although that experiment did not account for a possible involvement of ToxS or ToxT in ToxR activation (DiRita, 1992). Thus, the study of MFRHA expression may uncover other mechanisms of virulence gene control in *V. cholerae* as it does not appear to be controlled by known regulators of virulence gene expression.

Another feature of p_{mrh} is the atypical spacing between predicted -35 and -10 boxes: 14 nucleotides (Fig. 6.4B) as opposed to 17 nucleotides for the *E. coli* σ^{70} promoter consensus (O'Neill, 1989). This may reflect either the need for an activator protein for expression from p_{mrh}, or that it belongs to a group of promoters with non-consensus spacing. Promoters with

non-consensus spacing have been proposed to be sensitive to DNA twist, and thus to changes in environmental conditions that affect the state of DNA supercoiling (Wang and Syvanen, 1992). In the case of p_{mrh} a decrease in negative supercoiling would be predicted to increase relative spacing between -35 and -10 boxes and thus activate expression (Wang and Syvanen, 1992), such as occurs when *E. coli* encounters a change in its nutrient source or in its transition from exponential to stationary phase (Balke and Gralla, 1987). MFRHA activity appears during late exponential phase and is maximal in stationary phase in *V. cholerae* growing *in vitro* (Hanne and Finkelstein, 1982) suggesting that DNA supercoiling may play a role in expression from p_{mrh}. The expression of a number of environmentally regulated genes is controlled by changes in the level of DNA supercoiling in a variety of pathogenic bacteria, including adhesin operons whose expression is temperature sensitive (Dorman, 1991). However, mutagenesis in the p_{mrh} promoter region is necessary to determine whether the predicted -35 box is involved in promoter activity or not, and to confirm the apparent atypical spacing between -35 and -10 boxes. Similarly, determination of a role for DNA supercoiling in *mrh* expression requires the construction of gyrase mutants in *V. cholerae*.

Mapping of predicted promoters adjacent to *mrhAB* has revealed that one, p_{ORF6}, also has atypical spacing between predicted -35 and -10 boxes, again of 14 nucleotides. It is not known whether expression from this promoter is coordinated with expression from p_{mrh}, but comparison of the sequences in the vicinity of p_{mrh} and p_{ORF6} did not reveal similarities between the two, other than the non-consensus spacing of the predicted -35 and -10 boxes. Evidence presented in Chapter 7 suggests that the product of ORF6 influences expression from p_{mrh}, but whether this is a direct or indirect effect remains to be determined. The promoter upstream from p_{mrh} shows consensus spacing between predicted -35 and -10 boxes. A 5' end was detected only in *E. coli* HB101[pPM471], so it remains to be determined if this predicted promoter is active in *V. cholerae* as well. One of the two ORFs that are most probably transcribed from this promoter, ORF3.2, is predicted to encode a general inhibitor of translation (Chapter 3), a gene product that, because of its presumed toxicity, is not expected to be expressed under normal growth conditions. Therefore, the signals governing expression from this promoter are expected to be different from those governing expression from p_{mrh}, transcription from which is readily detectable during *in vitro* growth.

The biogenesis and exact nature of the MFRHA on the cell surface remains unclear. There is also some controversy as to whether *mrhA* or *mrhB* is the MFRHA structural gene: data supporting the assertion that *mrhA* (van Dongen *et al.*, 1987) and *mrhB* (Franzon and Manning, 1986) encode the adhesin are available. Demonstration that the two genes form an operon supports the possibility that both may be involved in the MFRHA phenotype, and suggest that previous experiments eliminating one or the other gene (Franzon and Manning, 1986; van Dongen *et al.*, 1987) may have been complicated by polar effects of these mutations. Attempts to resolve this anomaly are presented in Chapter 7.

Chapter 7

Examination of the role of the Mannose-Fucose-Resistant Haemagglutinin of *Vibrio cholerae* O1 in bacterial adherence and virulence

7.1 Introduction

As outlined in earlier Chapters, many adhesins possessed by pathogenic bacteria are proteinaceous filaments called fimbriae. The biogenesis of these structures is complex and requires the presence of a number of proteins in addition to the major structural subunit (Mooi and DeGraaf, 1985; Hultgren *et al.*, 1993). Different fimbriae have different adhesin locations, possessing either a specialised "tip" adhesin (a minor subunit located at the tip of a fimbria) or the major subunit itself being the adhesin (and thus having adhesin along the length of a fimbria). However, not all adhesins form fimbriae (Hultgren *et al.*, 1993).

The physical nature of the MFRHA is of considerable interest, given the discrepancy between results obtained in this laboratory (Franzon and Manning, 1986; Franzon *et al.*, 1993) and elsewhere (van Dongen *et al.*, 1987) as to which gene encoded on the 6.3 kb *Bam*H1 fragment (Chapter 3), encodes MFRHA activity. Franzon and Manning (1986) originally identified the 25 kDa ORF (*mrhB*) as the MFRHA by analysing haemagglutinating activities associated with different plasmid subclones, as did van Dongen and DeGraaf (1986). However, further analysis of their original and additional subclones led van Dongen *et al.* (1987) to reassign MFRHA activity to the adjacent 7 kDa ORF (*mrhA*) (these results are summarised in Fig. 7.1). Data presented in Chapter 6 has shown that *mrhA* and *mrhB* form an operon, and a possible explanation for the discrepancies between results obtained by the two groups is that "normal" expression of the operon is disrupted, as subclones that delete different portions of the two genes have been analysed by the two groups (Franzon and Manning, 1986; van Dongen and DeGraaf, 1986; van Dongen *et al.*, 1987).

Franzon (1988) also engineered a gene replacement mutation in the classical strain 569B, in which the distal portion of *mrhB* and most of the adjacent ORF6 (Clark, 1988; Chapter 3) was replaced by the Km^R resistance cartridge of pUC4-K (Vieira and Messing, 1982). This mutation dramatically attenuated 569B, although the strain still carried a plasmid, pPH1JI (Beringer *et al.*, 1978), used during mutant construction, and which has subsequently been shown to attenuate *V. cholerae* O1 strains of both biotype (Focareta, 1989). Since the mutation was not limited to one gene, the discrepancy as to which gene encodes the MFRHA has not been resolved. Therefore, further analysis of ORFs in this region is necessary in order to assign function to them. Preliminary analysis, with the view to

Figure 7.1: A summary of subclones in the MFRHA region used to assign MFRHA activity to a particular gene.

Panel A shows a schematic representation of pPM471 and some subclones derived from it by Franzon and Manning (1986). Numbers refer to the distance from the proximal end of pPM471 in kb and some restriction sites in the region are indicated. Panel B shows a schematic representation of pVC hag-11 and some subclones derived from it by van Dongen and DeGraaf (1986). Numbers refer to the distance from the proximal end of pVC hag-11 in kb and some restriction sites in the region are indicated. Sequences originating from outside the 6.3 kb *Bam*HI fragment in some plasmids are indicated with a dashed line and are not drawn to scale. Panel C shows a schematic representation of pVC hag-232 (van Dongen and DeGraaf, 1986) and some subclones derived from it by van Dongen *et al.* (1987). Some restriction sites in pVC hag-232 are indicated. The portions subcloned are represented by solid black lines annotated with plasmid number, and the ability or otherwise to mediate MFRHA activity on *E. coli* K-12 containing the plasmid is indicated. Details of plasmid construction are in Franzon and Manning (1986), van Dongen and DeGraaf (1986) or van Dongen *et al.* (1987).
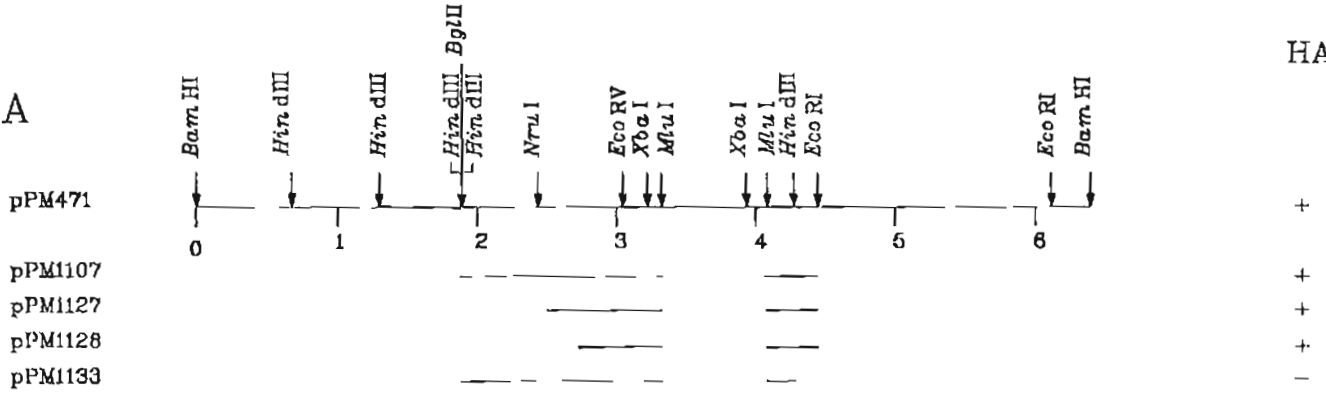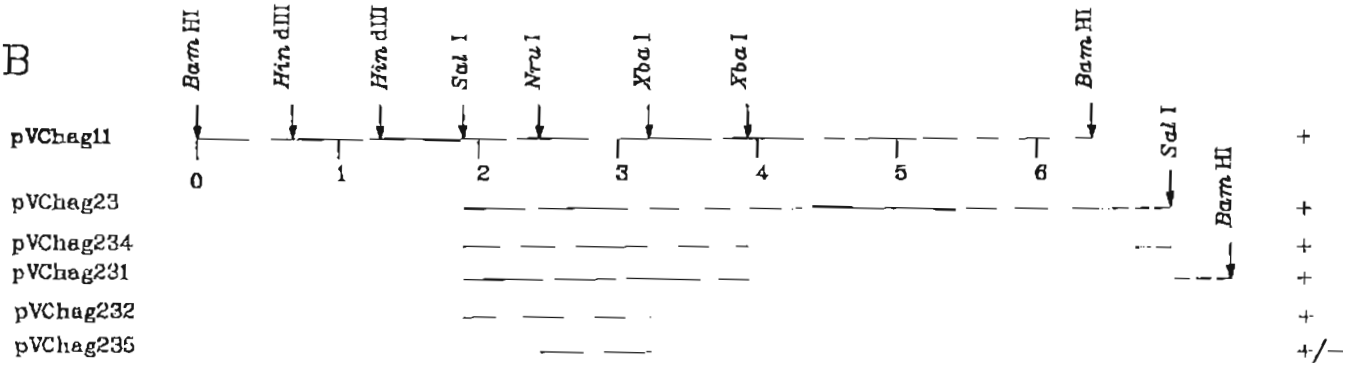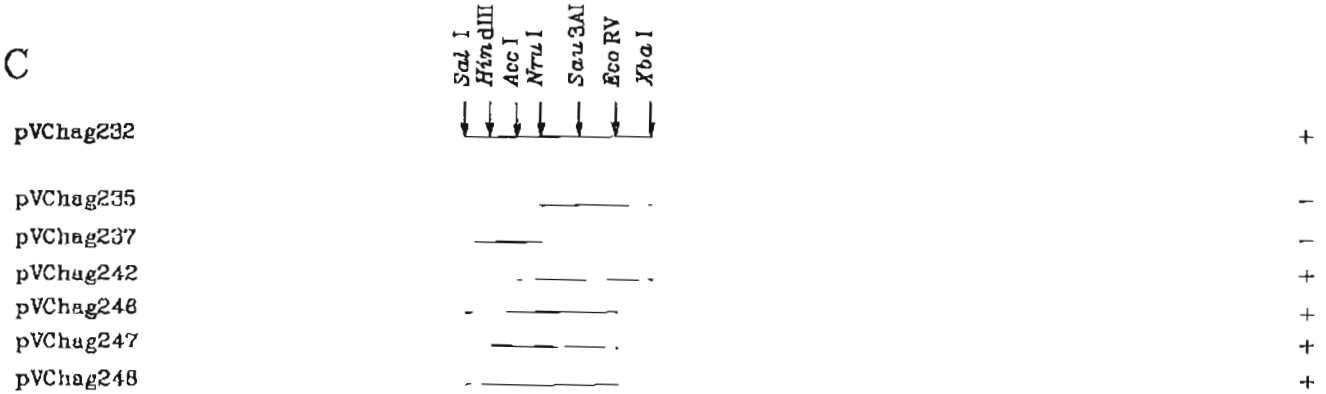
resolving the conflict as to which ORF encodes the MFRHA, has been undertaken, and is presented in this Chapter.

7.2 Results

7.2.1 Analysis of the MFRHA::Km^R mutant V761

As discussed above, Franzon (1988) constructed a gene replacement mutant, by replacing the region between the two *Xba*I sites of pPM471 with the Km^R cartridge of pUC4-K (Vieira and Messing, 1982). The resulting plasmid, pPM1146 (Fig. 7.1), was then used as the starting point to introduce this mutation into *V. cholerae* 569B, to produce strain V761 (Franzon, 1988).

Although V761 was attenuated, the effect of the mutation on haemagglutination was not reported. To confirm an effect, plasmids were introduced into *E. coli* LE392 and HB101, the host strains used by Franzon (1988) and van Dongen *et al.* (1987), respectively, and their ability to mediate haemagglutination examined (Table 7.1). In both strains, pPM471 and the minimal clone from this laboratory, pPM1127 (Franzon and Manning, 1986) mediate haemagglutination, whereas the gene replacement mutant, pPM1146, has lost the ability to haemagglutinate. This indicates that the mutation has affected the MFRHA phenotype in the heterologous *E. coli* host.

As V761 contains the plasmid pPH1JI (Franzon, 1988), which has been shown to attenuate *V. cholerae* O1 strains (Focareta, 1989), this plasmid was cured by mating with *E. coli* DH5[pME305], a temperature-sensitive replicon of the same incompatibility group (IncP) as pPH1JI (Table 2.3), followed by selection for Ap^R, Tc^R at 30°C. Colonies were checked for curing of pPH1JI on the basis of sensitivity to each of Gm, Sm and Sp (separately), and pME305 was then cured by growth in the absence of selection at 42°C. One motile Km^R, Ap^S, Tc^S colony was picked and designated V762.

The ability of this strain to mediate haemagglutination was then compared to the parent strain (Table 7.2). The observed haemagglutination of the parental strain 569B is abolished by the mutation introduced into V762. Attempts to complement this loss of

Table 7.1 Haemagglutination titres of MFRHA clones in *E. coli* K-12

plasmid	Haemagglutination titre in *E. coli*	
	LE392	HB101
pBR322	1	1
pPM471[a]	16	16
pPM1127[a]	16	16
pPM1146[a]	1	1

[a] These plasmids were constructed by Franzon (1988).

Table 7.2 Virulence of *V. cholerae* 569B and MFRHA⁻ derivatives

strain	Haemagglutination titre[a]	cholera toxin production[b]	LD_{50}[c]
569B	16	+++	7×10^5
V762	1	+ ·+	$>1 \times 10^8$
V762[pBR322]	1	+++	$>1 \times 10^8$
V762[pPM471]	1	+·+	$>1 \times 10^8$
V762[pPM1127]	1	+++	$>1 \times 10^8$
V762[pPM1654]	1	+++	$>1 \times 10^8$

[a] Haemagglutination titre is the result for three independently isolated colonies.

[b] Cholera toxin production was similar in all cases, being approximately 40 µg/ml of culture supernatant.

[c] Groups of eight mice were fed 10-fold serial dilutions of 569B; all other strains were administered to a single group of mice at the maximal dose.

MFRHA activity by introduction of the cloned genes were unsuccessful (Table 7.2). Initially, introduction of the original clone, pPM471, was attempted but the possibility of rearrangement of the clone as observed in *recA⁺* *E. coli* K-12 (Chapter 3) forced the use of smaller fragments. Both the minimal fragment (*mrhB⁻*) encoding haemagglutinating activity that is carried on plasmid pPM1127 (Franzon and Manning, 1986; Franzon, 1988) and pPM1654, a *mrhA⁺B⁻* construct (see Section 7.2.5 for construction) fail to restore haemagglutination (Table 7.2).

The effect of the mutation on the virulence of *V. cholerae* in the absence of pPHIJI was assessed (Table 7.2). The mutation abolishes virulence in the infant mouse model, but again virulence cannot be restored by introduction of the cloned genes. A defect in the mutant strain is seen in competition experiments (Fig. 7.2), in which the parent strain (569B) rapidly outgrows the mutant (V762). However, numbers of the mutant strain recovered stabilised at approximately 500-fold fewer than the parent, indicating that V762 retains some ability to colonise This inability to compete is not observed *in vitro*, where ratios of 569B to V762 do not alter significantly over the course of the experiments in either rich (NB) or minimal media (not shown).

Due to the low haemagglutination titres observed in the above experiments, the feasibility of *in vitro* adherence to the brush border-forming cultured epithelial cell line Caco-2 (Pinto *et al.*, 1983) as an alternative assay for MFRHA activity was examined. At the two input concentrations of bacterial cells examined, an approximately two- to three-fold difference between the adherence of 569B and V762 was discernible (Table 7.3). Provision of *mrhB⁻* did not lead to complementation with this assay either, with V762[pBR322] giving comparable results to V762[pPM1127]. Thus, at least with strains derived from 569B, it seems that adherence to Caco-2 cells does not present an attractive alternative to haemagglutination as a means for assessing adherence *in vitro*.

Figure 7.2: Colonisation of infant mice by *I' cholerae* 569B and its MFRHA-negative variant V762.

Infant mice were fed 5×10^5 cells (low dose) or 4×10^7 cells (high dose) of either 569B or its MFRHA-negative variant V762. At 21 and 44 hours, the mice were sacrificed, their entire intestines were removed, and the numbers of *I' cholerae* organisms were determined by viable count.
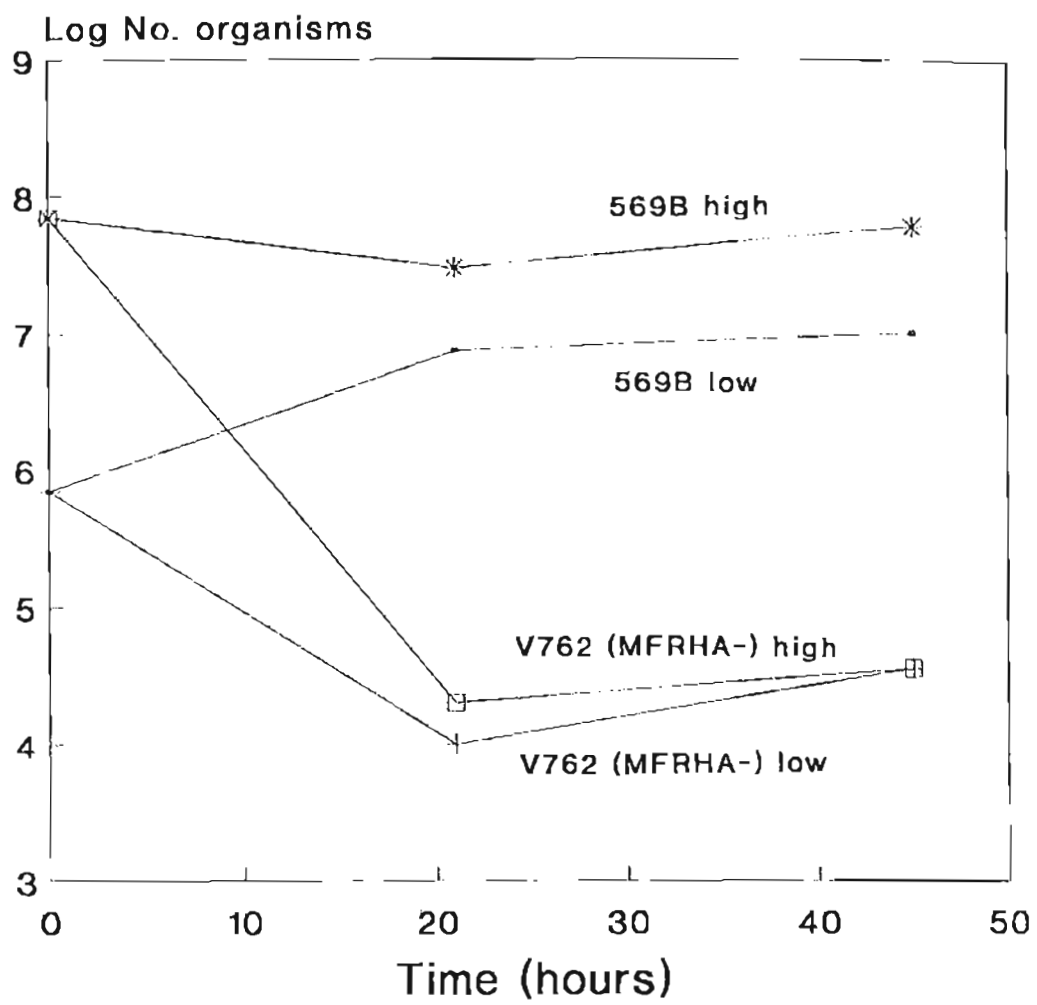
Log No. organisms

569B high

569B low

V762 (MFRHA-) high

V762 (MFRHA-) low

Time (hours)

Table 7.3 Adherence *in vitro* of *V. cholerae* 569B and MFRHA⁻ derivatives to Caco-2 cells

strain	input concentration[a]	percent adherent[b]
569B	10^8	0.45 ± 0.13
V762	10^8	0.14 ± 0.036
569B	10^7	0.66 ± 0.23
V762	10^7	0.19 ± 0.11
V762[pBR322]	10^8	0.14 ± 0.016
V762[pPM1127]	10^8	0.14 ± 0.046
V762[pBR322]	10^7	0.2 ± 0.005
V762[pPM1127]	10^7	0.15 ± 0.055

[a] in cfu/ml. Approximate values were obtained from A_{650} values. The exact concentration was determined by viable count.

[b] expressed as the percentage of input viable count recovered after washing of the monolayer. Results are presented as mean ± standard deviation for three independent cultures.

7.2.2 Effect of the presence of lipopolysaccharide on the haemagglutination titre of MFRHA$^+$ E. coli K-12

One aspect of the isolation of MFRHA$^+$ clones has been the finding that the haemagglutination titre of clones in *E. coli* K-12 is much lower (with the exception of 569B) than that of the *V. cholerae* strains from which the clones were isolated (Franzon and Manning, 1986; van Dongen and DeGraaf, 1986; Table 7.4). One possible explanation for this effect is that the MFRHA requires the presence of *V. cholerae* lipopolysaccharide for correct presentation on the cell surface. Therefore, pRM242, a clone of Ogawa serotype *rfb* genes derived from pPM1002 (Manning *et al.*, 1986) in a modified pSC101 replicon, pRM118, (R. Morona, unpublished), and pRM118 were introduced into *E. coli* HB101 containing pBR322 or pPM471. Cells were judged to be expressing *V. cholerae* O1 Ogawa-specific LPS by a positive slide agglutination reaction with Ogawa- but not Inaba-specific antisera and by silver staining and western blot analysis of crude LPS preparations from the cells separated by SDS-PAGE (not shown). Examination of the haemagglutination titres of these cells revealed no difference in the titre when LPS producing clones were introduced (Table 7.4) indicating that the absence of a homologous LPS environment was not responsible for the differences in titre between MFRHA$^+$ *V. cholerae* and *E. coli* K-12.

7.2.3 Construction of new *mrhA* and *mrhB* mutants

As the *mrh* mutant constructed by Franzon (1988) only removes the twenty carboxy-terminal amino acids of MrhB and in addition removes the bulk of the downstream ORF6 (Chapter 3), a series of mutants, specific for *mrhA* or *mrhB*, or both, was constructed (Fig. 7.3). First, the insert of pPM471 was subcloned into pUC19 (Yanisch-Perron *et al.*, 1985), creating pPM1619. Linearisation of this plasmid with *Nru*I, followed by bidirectional digestion with Exonuclease III and religation of the products, gave a series of plasmids. Sequencing of six of these (sized by restriction analysis; not shown) revealed that two contained deletions that were entirely within *mrhA*. The *Bam*HI fragments of *V. cholerae* DNA in these plasmids were then re-subcloned into pBR322 (Bolivar *et al.*, 1977) in the same orientation as pPM471 and the resulting plasmids named pPM1620 and pPM1621 (Fig. 7.3). The deletions span nucleotides 2491-2609 and 2444-2706 in pPM1620 and pPM1621, respectively (not shown). Second, pPM1619 DNA was linearised with *Nru*I and partially

Table 7.4 Haemagglutination titres of:

A some *V. cholerae* O1 strains used in this study

strain	biotype	Haemagglutination titre[a]
569B[b]	classical	16 (8)
Z17561	classical	32 (16)
O17[c]	El Tor	256 (64)
C5[d]	El Tor	64
C31	El Tor	4 (2)

[a] Haemagglutination results are combined from several experiments performed in triplicate. Numbers presented are the maximal titres obtained, with numbers in parentheses representing minimal titres obtained.

[b] 569B is the strain from which the MFRHA clone pPM471 was derived (Franzon and Manning, 1986).

[c] A series of MFRHA expressing cosmids was isolated from O17, which conferred MFRHA activity comparable to that conferred by pPM471, on *E. coli* LE392 (Franzon, 1988).

[d] C5 is the strain from which the MFRHA clone pVC hag-11 was derived (van Dongen and DeGraaf, 1986).

**B *E. coli* MFRHA clones expressing *V. cholerae* LPS**

E. coli HB101 with:	Haemagglutination titre[a] in the presence of:	
	pRM118[b]	pRM242[b]
pBR322	0	0
pPM471	16	16

[a] Haemagglutination titre is the result for three independently isolated transformants.

[b] Plasmids pRM118 and pRM242 are described in the text.

Figure 7.3: Construction of subclones for examination of the roles of *mrhA* and *mrhB* in the MFRHA phenotype.

The 6.3 kb *Bam*HI fragment contained within pPM471 and pVC hag-11 is indicated, with subclones derived from the respective plasmids indicated below. Numbers refer to the distance from the proximal end of the clone in kb and some restriction sites in the region are indicated. The genes *mrhA*, *mrhB* and the predicted ORF6 are represented by open boxes, and three copies of the direct repeat VCR are represented by hatched boxes. The portions subcloned are represented by solid black lines annotated with plasmid number. Spaces in the lines indicate the regions absent in deletion mutants, and the position of a *Sal*I linker-insertion mutation in pPM1626 is indicated with a triangle. Details of plasmid construction are in Sections 7.2.3 and 7.2.4.
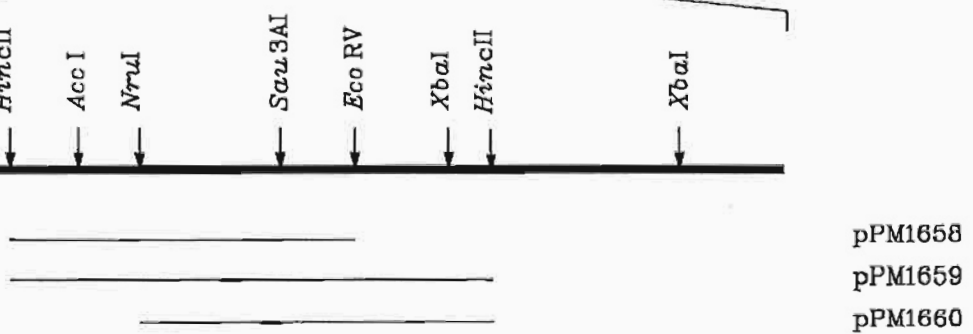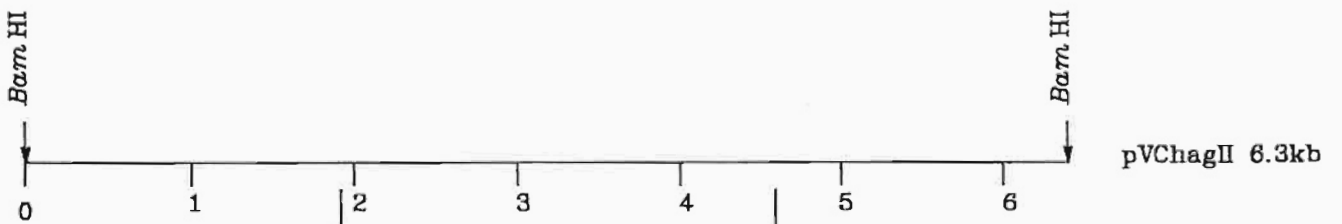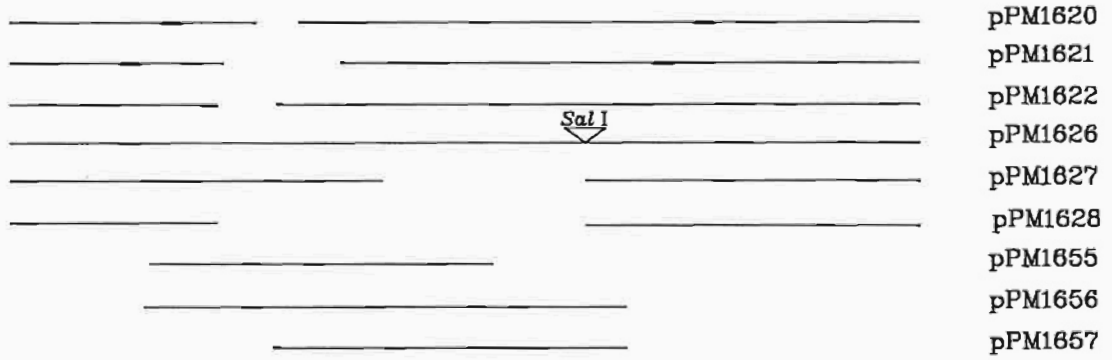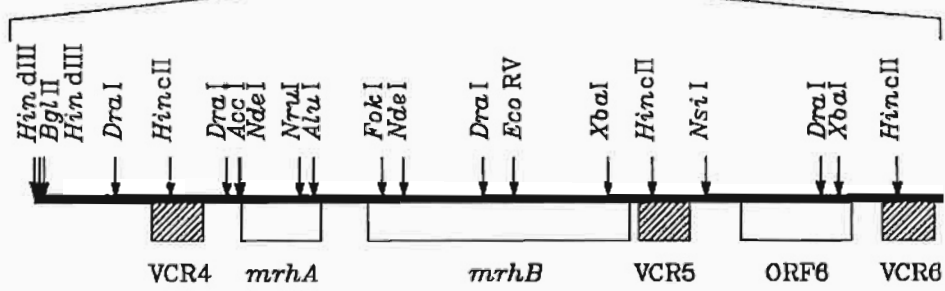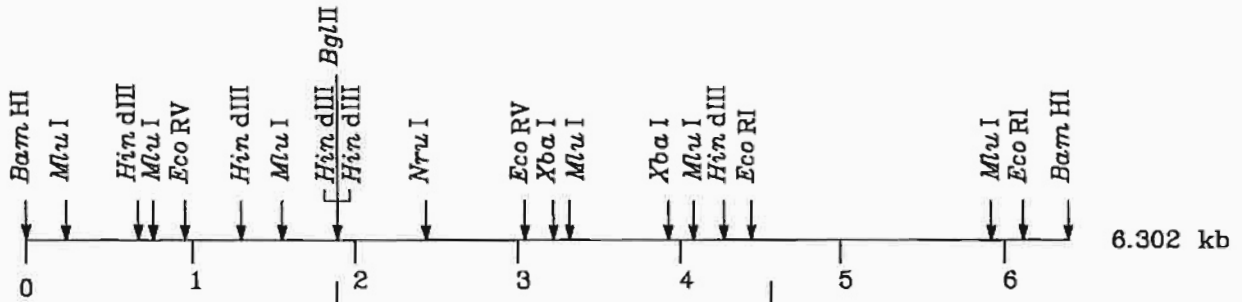
digested with *Nde*I. The largest partially digested product was isolated, end repaired and religated. After confirmation of the *mrhA* deletion by restriction mapping, the *Bam*HI insert of this plasmid was reintroduced into pBR322 in the same orientation as pPM471 and the resulting plasmid named pPM1622. Third, plasmid pPM1623 was created by subcloning the entire insert in pPM471 into the *Bam*HI site of plasmid pACYC177 (Chang and Cohen, 1978). Partial digestion with *Xba*I, followed by end repair and religation with *Sal*I linkers gave plasmids pPM1624, in which the *Xba*I site in *mrhB* had been replaced with a *Sal*I site, and pPM1625, in which the *Xba*I site in ORF6 had been replaced with a *Sal*I site (confirmed by restriction mapping; not shown). Plasmid pPM1626 was created by reintroducing the insert of pPM1624 into pBR322 in the same orientation as pPM471 (Fig. 7.3). Linearisation of pPM1624 with *Sal*I, followed by partial digestion with *Nde*I and purification and recircularisation of the two largest partially digested products gave plasmids in which *mrhB* or *mrhAB* had been deleted (not shown). The inserts from these plasmids were reintroduced into pBR322, again in the same orientation as the insert of pPM471, and were named pPM1627 and pPM1628, respectively (Fig. 7.3). Numerous attempts to engineer an ORF6-specific deletion by partial digestion with *Nsi*I, of *Sal*I linearised pPM1625, were unsuccessful (not shown). *E. coli* K-12 strains harbouring plasmids pPM1624, pPM1626 and pPM1627 showed a reduced growth rate in comparison to strains bearing other pPM471 derived plasmids (not shown). Similarly, *E. coli* strains containing pPM1146 show a reduced growth rate in comparison to strains bearing pPM471, although a similar retardation is not observed on introduction of the lesion into *V. cholerae* 569B (not shown).

The effect of these *mrhA* and *mrhB* mutations on the haemagglutination titre of *E. coli* AAEC189 was examined. Unlike *E. coli* HB101, which produces Type 1 fimbriae under some conditions, AAEC189 is bald, due to a Δ*fim* lesion (Bloomfield *et al.*, 1991a). In addition, plasmid instability due to recombination between multiple copies of VCR (Chapter 4) is reduced in AAEC189, a *recA* strain, in comparison to the *recA*[+] LE392 (not shown). The only deletion mutant to abolish haemagglutination in this strain was pPM1628, a Δ*mrhAB* mutation (Table 7.5). Single mutants in both *mrhA* and *mrhB* have no effect on the haemagglutination titre of *E. coli* AAEC189 (Table 7.5). Therefore, the effect of the plasmids examined in Table 7.1 on the haemagglutination titre of *E. coli* AAEC189 was also examined.

Table 7.5   Haemagglutination titres of _E. coli_ AAEC189 containing pPM471 and derivatives

plasmid	_mrh_ genotype[a]	Haemagglutination titre[b]
pBR322	vector	1
pPM471	_mrhA_$^+$_B_$^+$	16
pPM1620	_mrhA1_	16 (32)[c]
pPM1621	_mrhA2_	16
pPM1622	_mrhA3_	16
pPM1626	_mrhB1_	16
pPM1627	_mrhB2_	16
pPM1628	Δ_mrhAB_	1
pBR322	vector	10
pPM471	_mrhA_$^+$_B_$^+$	16
pPM1127	_mrhB_$^+$	16 (32)[c]
pPM1146	_mrhB_::Km^R, ΔORF6	16

[a] described in detail in the text.

[b] Haemagglutination titre is the result for three independently isolated transformants.

[c] Some agglutination was evident in the fifth well.

In this strain, all plasmids mediate haemagglutination, including pPM1146, the Franzon (1988) gene replacement mutant (Table 7.5). The reason for the discrepancy between this result and Table 7.1 is not known (see Section 7.3).

The transcriptional organisation of *mrhAB* was established in Chapter 6, in which "read through" of a transcriptional terminator between *mrhA* and *mrhB* led to expression of the latter. As two of the *mrhA* lesions are predicted to disrupt this terminator, the effect of these and other lesions on the expression of the genes was examined. Total RNA was extracted from *E. coli* AAEC189 containing the indicated plasmids and analysed by northern blot. A *mrhA*-specific probe hybridised with an abundant 400 nucleotide band with plasmids pPM471, pPM1626, pPM1627, a weak 400 nucleotide band with pPM1146 and a weak 250 nucleotide band with pPM1620 (Fig. 7.4). Bands were not detected with strains bearing the other plasmids. In addition a weak 650 nucleotide band was detected with pPM471, pPM1626 and pPM1627 and of 500 nucleotides with pPM1620, which is probably analogous to the 650 nucleotide band detected for pPM471 with the same probe in Chapter 6. A band corresponding to the full-length *mrhAB* was not detected with this probe. Similarly, bands were not detected with a *mrhB* probe (not shown). Thus, *mrhB* mutations do not appear to affect transcription from p_{mrh}, but a mutation in the downstream ORF6 does.

7.2.4. Comparison of identical subclones derived from *V. cholerae* strains 569B and C5

The nucleotide sequence of the entire insert of the original MFRHA$^+$ clone pPM471, isolated from the classical strain 569B (Franzon and Manning, 1986) has been presented in Chapter 3. MFRHA-associated genes have also been cloned from an El Tor strain, C5 (van Dongen and DeGraaf, 1986) and *mrhAB* sequenced (van Dongen *et al.*, 1987). In the light of the differences in results between the two groups and compounded by results presented in Section 7.2.3, comparison of the two protein sequences was undertaken, and is shown in Fig. 7.5. The sequence of *mrhA* (the 7 kDa ORF) is identical to that of van Dongen *et al.* (1987) (Fig. 7.4A), but *mrhB* (the 25 kDa ORF) shows six changes at the nucleotide level. Four of these lead to changes in the predicted amino acid sequences (based in part on the choice of different starts to the ORFs: van Dongen *et al.* (1987) chose the first possible ATG codon for *mrhB*, whereas the start site preceded by the best match to a ribosome binding site is used

Figure 7.4: Analysis of transcription of *mrhA* and *mrhB* mutant alleles in *E. coli* AAEC189.

The filter shown is probed with a *mrhA* specific probe derived from plasmid pPM1614. Lanes are total RNA prepared from *E. coli* AAEC189 containing: 1, pBR322; 2, pPM471; 3, pPM1127; 4, pPM1146; 5, pPM1620; 6, pPM1621; 7, pPM1622; 8, pPM1626; 9, pPM1627; 10, pPM1628. Plasmids are described in Figures 7.1 and 7.3.
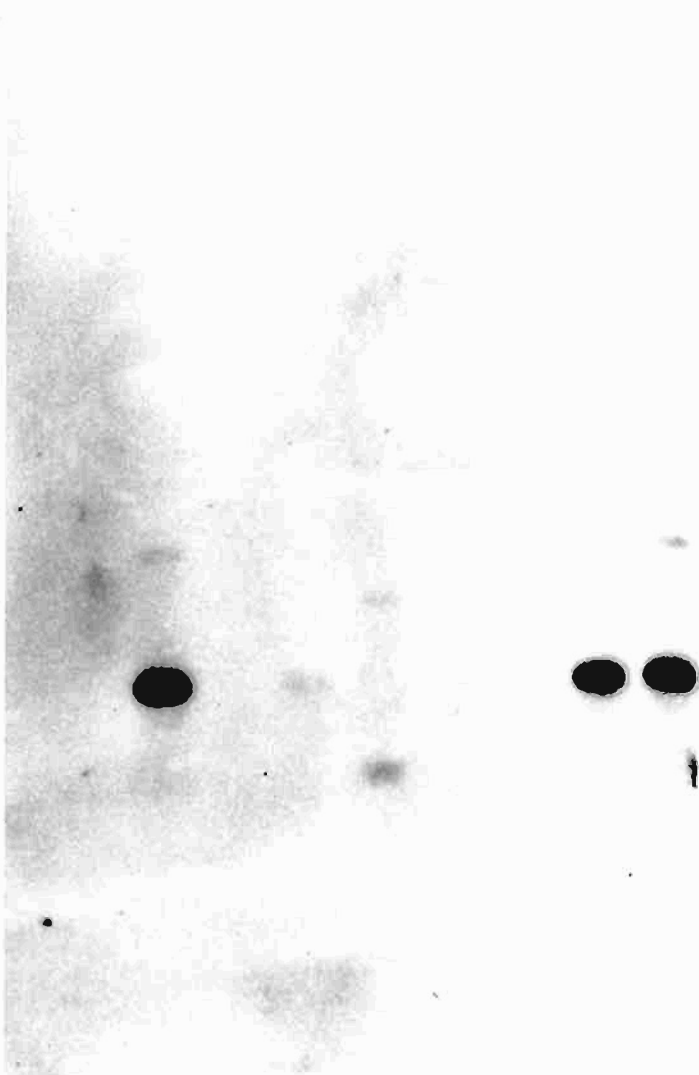
1　2　3　4　5　6　7　8　9　10

Figure 7.5: Comparison of MrhA and MrhB sequences derived from *V. cholerae* strains 569B (classical) and C5 (El Tor).

Panel A shows an alignment between MrhA from 569B (Chapter 3) and C5 (van Dongen *et al.*, 1987). Panel B shows an alignment between MrhB from 569B (Chapter 3) and C5 (van Dongen *et al.*, 1987). The plasmid from which each allele originated is indicated. Identical residues in the alignments are joined by a solid line (|), similar residues are joined by a dotted line (:). The boundaries of the region of similarity are indicated by an X above and below the alignment. Numbers annotating the sequence indicate the position from the amino terminus of each residue. The sequence of MrhB from pVC hag-11 is numbered according to van Dongen *et al.* (1987). The alignments were performed using the algorithm BLAZE (Brutlag *et al.*, in preparation).

A

```
                   X          10          20          30          40
MrhA (pPM471)      MPDLTKARCPCRGCNKVANNAEEVEELFGLRNMGDGTIRVQSYC
                   |||||||||||||||||||||||||||||||||||||||||||
MrhA (pVC hag-11)  MPDLTKARCPCRGCNKVANNAEEVEELFGLRNMGDGTIRVQSYC
                   X          10          20          30          40


                              50          60 X
MrhA (pPM471)      ERECRGLHCEASNPKCNHHA
                   ||||||||||||||||||||
MrhA (pVC hag-11)  ERECRGLHCEASNPKCNHHA
                              50          60 X
```

B

```
                   X          10          20          30          40
MrhB (pPM471)      MSKIYQMDAVDWLKTLENCSVDLFITDPPYESLEKYRQIVTTTR
                   |||||||||||||||||||||||||||||||||||||| ||||
MrhB (pVC hag-11)  MSKIYQMDAVDWLKTLENCSVDLFITDPPYESLEKYRQIGTTTR
                   X          20          30          40          50


                              50          60          70          80
MrhB (pPM471)      LKESKSSSNQWFSVFPNTRFEELFREVYRVLKKGSHFYLFCDQE
                   |||||||||||||||||||||||||||||||||||||||||||
MrhB (pVC hag-11)  LKESKSSSNQWFSVFPNTRFEELFREVYRVLKKGSHFYLFCDQE
                              60          70          80          90          100


                   90          100         110         120         130
MrhB (pPM471)      TMFLAKPIAESVGFKFWKPIVWDKCRIGMGYHYRARYEFILFFE
                   |||||||||||||||||||||||||| |||||||||||||||||
MrhB (pVC hag-11)  TMFLAKPIAESVGFKFWKPIVWDKCAIGMGYHYRARYEFILFFE
                              110         120         130         140


                   140         150         160         170
MrhB (pPM471)      KGKRKLNDLSVPDVLEYKRVWKGYPTEKPVELLEVLIRQSSSEN
                   |||||||||||||||||||||||||||||||||||||||||||
MrhB (pVC hag-11)  KGKRKLNDLSVPDVLEYKRVWKGYPTEKPVELLEVLIRQSSSEN
                              150         160         170         180


                   180         190         200         210         X
MrhB (pPM471)      EIVADSFFGSGATLIAANNLSRKYIGCDISMSAHEYFKNRA
                   |||||||||||:||||||||||||||||||| |||||||||
MrhB (pVC hag-11)  EIVADSFFGSGSTLIAANNLSRKYIGCDISSSAHEYFKNRA
                              190         200         210         220         X
```

here, as discussed in Chapter 3). Three of these amino acid substitutions are non-conservative (Fig. 7.5B). However, these do not lead to significant changes in the hydropathy plot (Fig. 7.6A) and only minor changes in the predicted secondary structure (Fig. 7.6B). These probably reflect differences in the *V. cholerae* strains that the two sequences are derived from. Both sequences show an overall hydrophilic character. A hydropathy plot for MrhA shows that the protein should be quite hydrophilic, without any markedly hydrophobic domains (Fig. 7.7).

One possible explanation for the difference in the results obtained by this group (Franzon and Manning, 1986; Franzon, 1988) and DeGraaf and co-workers (van Dongen *et al.*, 1987) is that the difference in sequence between the *V. cholerae* 569B and C5 *mrhB* alleles has somehow led to different MFRHA activities in the *E. coli* clones examined (Fig. 7.1). It is worth noting that van Dongen and DeGraaf (1986) originally designated pVC hag-235 as their minimal MFRHA$^+$ clone, in which the insert spanned from the unique *NruI* site to the proximal *XbaI* site (Fig. 7.1). This clone was subsequently described as MFRHA$^-$, and all subsequent analyses were conducted with subclones that extended further upstream of the *NruI* site, but not downstream from this proximal *XbaI* site (van Dongen *et al.*, 1987). It is also worth noting that the Franzon (1988) Km^R replacement mutation analysed above (Section 7.2.1) involves a deletion of sequences distal to this *XbaI* site. In the light of the above information, an attempt was made to re-examine the MFRHA activities associated with similarly constructed subclones of the two MFRHA clones.

First, the *mrhAB* operons were subcloned by ligating *HincII* digested pPM471 (Franzon and Manning, 1986) or pVC hag-11 (van Dongen and DeGraaf, 1986) DNA with *HincII* digested pGEM3Zf$^+$. One hundred transformants from each ligation were screened for hybridisation with oligo #61, which binds within *mrhB*, and plasmids isolated from six positive colonies from each transformation subjected to restriction analysis (not shown). Both orientations of the 1.2 kb insert were isolated from both reactions, and were designated pPM1642 and pPM1643, for inserts derived from pPM471, with *mrhAB* in the same and opposite orientation as the vector T7 promoter, respectively, and pPM1644 and pPM1645, for inserts derived from pVC hag-11, with *mrhAB* in the same and opposite orientation as the

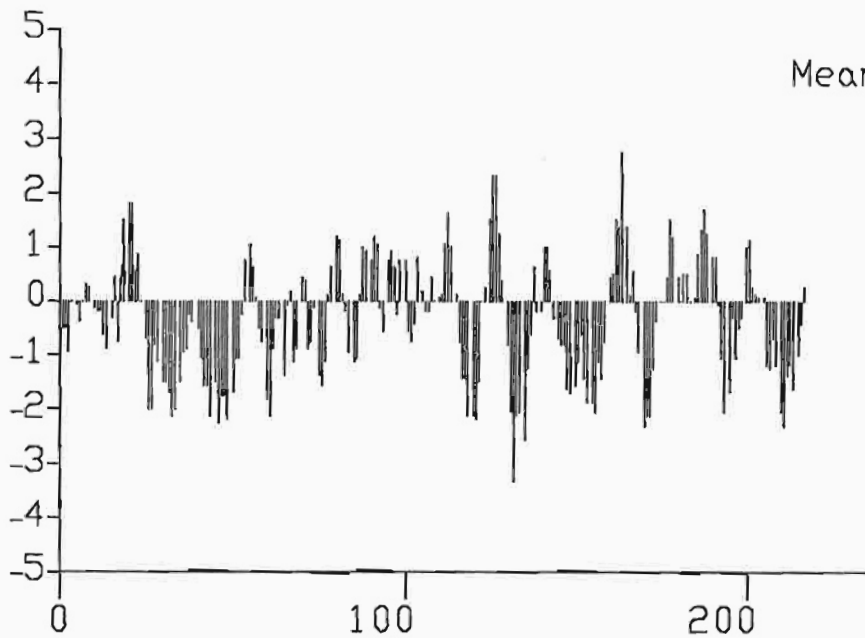**Figure 7.6:** Prediction of structural features of MrhB.

Panel A shows hydropathy plots for MrhB derived from 569B (top) and C5 (bottom) determined according to Kyte and Doolittle (1982), using a window of six amino acids. Panel B shows secondary structure predictions for MrhB derived from 569B (top) and C5 (bottom) determined using the algorithm of Chou and Fasman (1978).
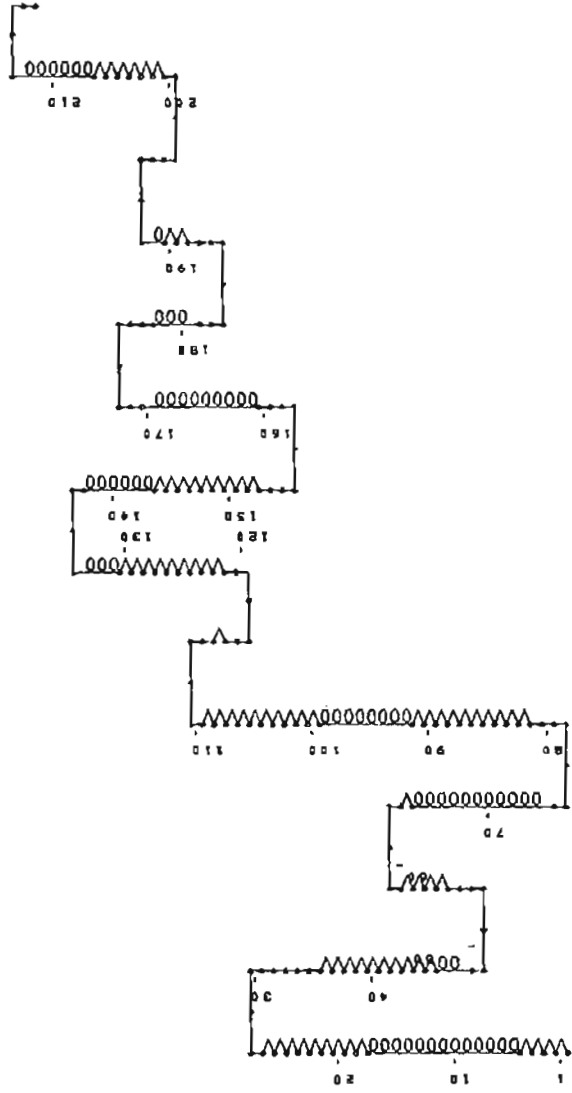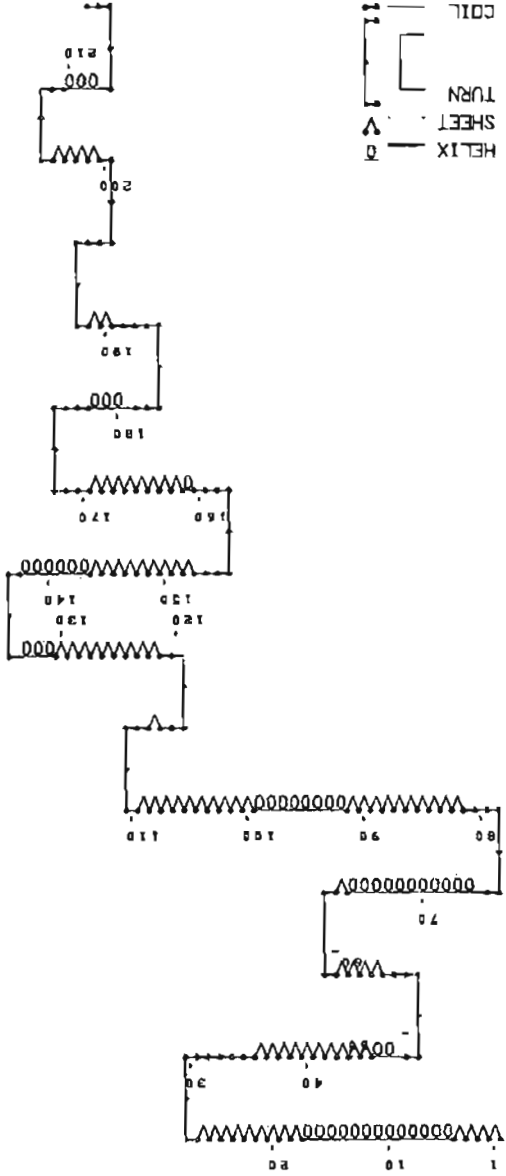
A

Index

Mean Index : −0.37

Index

Mean Index : −0.39
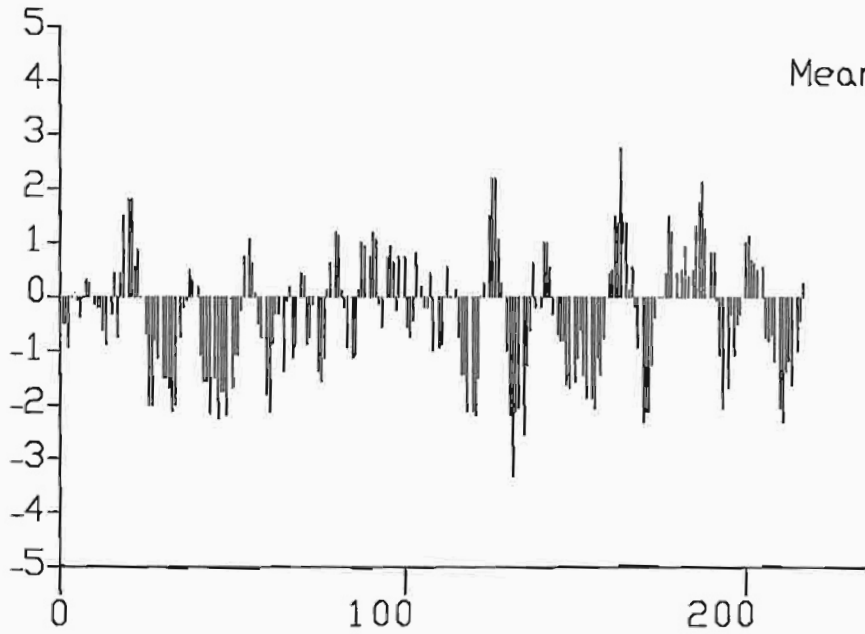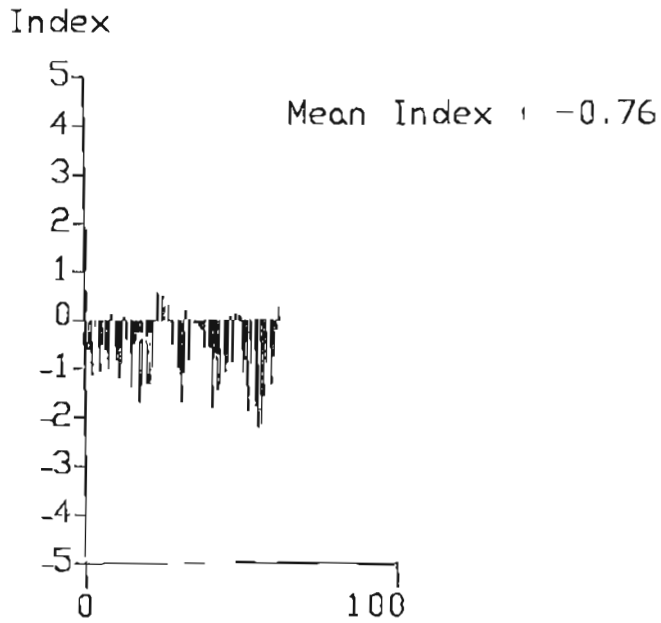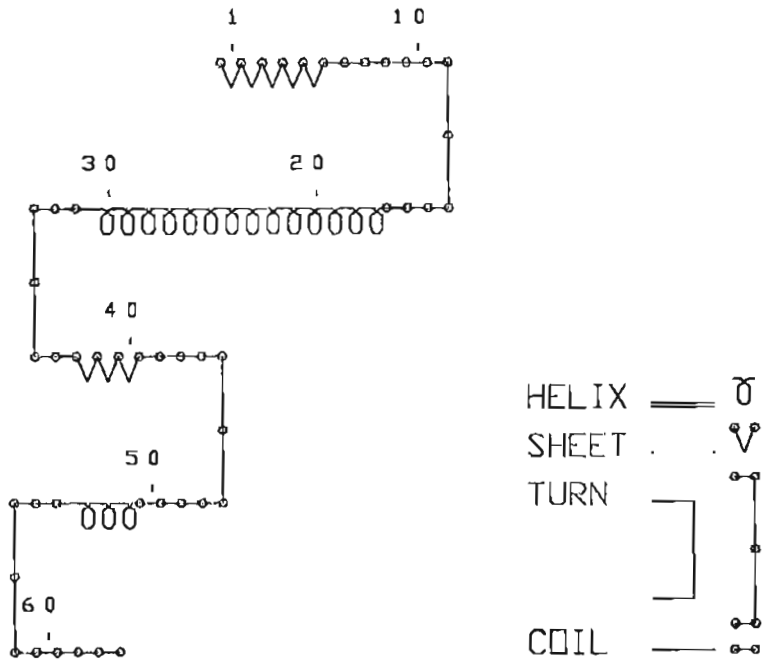
HELIX

SHEET

TURN

COIL

B

Figure 7.7: Prediction of structural features of MrhA.

Panel A shows a hydropathy plot for MrhA determined according to Kyte and Doolittle (1982), using a window of six amino acids. Panel B shows a secondary structure prediction for MrhA determined using the algorithm of Chou and Fasman (1978).

A

Index

5
4   Mean Index : -0.76
3
2
1
0
-1
-2
-3
-4
-5
   0   100

B

1   10

30   20

40

50

60

HELIX ═══ ♈
SHEET . . ♈
TURN
COIL ───

vector T7 promoter, respectively (not shown). Plasmid pPM1654, examined in Section 7.2.1, consists of the insert of pPM1643 subcloned into the *Eco*RV site of pBR322, such that *mrh* is in the same orientation as the vector TcR gene (not shown). Plasmids pPM1655, pPM1656 and pPM1657 consist of the 0.8 kb *Hinc*II-*Eco*RV fragment, the 1.2 kb *Hinc*II fragment, and the 0.85 kb *Nru*I-*Hinc*II fragment, respectively, of pPM1643 subcloned into the *Eco*RV site of pACYC184 such that *mrhAB* is in the same orientation as the vector TcR gene (Fig. 7.3). Plasmids pPM1658, pPM1659 and pPM1660 consist of the 0.8 kb *Hinc*II-*Eco*RV fragment, the 1.2 kb *Hinc*II fragment, and the 0.85 kb *Nru*I-*Hinc*II fragment, respectively, of pPM1645 subcloned into the *Eco*RV site of pACYC184 such that *mrhAB* is in the same orientation as the vector TcR gene (Fig. 7.3).

The MFRHA activity associated with these constructs was then examined in *E. coli* AAEC189. All constructs mediate haemagglutination (Table 7.6), suggesting that at least in AAEC189, either *mrhA* or *mrhB* is sufficient for MFRHA activity. The significance of these and of results presented earlier will be discussed in Section 7.3. Strains harbouring plasmids pPM1655 and pPM1658 show a reduced growth rate in both liquid and on solid media in comparison to strains containing the other plasmids (not shown).

7.2.5 Other attempts to identify the true identity of the MFRHA

As the aim of establishing the true identity of the MFRHA had still not been fulfilled, attempts were made to introduce the Δ*mrhAB* lesion of pPM1628 into *V. cholerae* and to then complement the (predicted) MFRHA⁻ phenotype with the different plasmids described in Section 7.2.5. To this end, the 1.15 kb *Bgl*II-*Xba*I fragment of plasmid pPM1628 was subcloned into pCACTUS, and the resulting plasmid (confirmed by restriction analysis) designated pPM1629 (not shown). This plasmid was then transformed into *V. cholerae* strains Z17561, SA33 (a *tcpA*::KmR derivative of Z17561 constructed by Dr. S. Attridge), and O17, and recombination events in which the plasmid allele replaced the chromosomal allele were selected for and the plasmid cured by virtue of its temperature sensitivity (not shown). The selection protocol should result in 50% of the progeny cells bearing the mutant allele instead of the parental allele (C. Clark, unpublished), and has been shown to work in the construction of *tcpA* mutants in *V. cholerae* Z17561 (S. Attridge, personal communication).

Table 7.6 Haemagglutination titres of *E. coli* AAEC189 containing similarly constructed derivatives of pPM471 or pVC hag-11

plasmid	*mrh* genotype[a]	Haemagglutination titre[b]
pACYC184	vector	1
pVC hag-11	$mrhA^- B^+$	16 (32)[c]
pPM1655	$mrhA^-$	16
pPM1656	$mrhA^- B^-$	16
pPM1657	$mrhB^-$	16 (32)[c]
pPM1658	$mrhA^-$	16
pPM1659	$mrhA^+ B^-$	16
pPM1660	$mrhB^-$	16

[a] described in detail in the text.

[b] Haemagglutination titre is the result for three independently isolated transformants.

[c] Some agglutination was evident in the fifth well.

However, when 24 Z17561[pPM1629] and 12 of each of SA33[pPM1629] and O17[pPM1629] colonies subjected to the selection protocol were screened by Southern hybridisation, all had retained the parental *mrhAB* locus (not shown). Allele-specific difficulties with isolating altered strains have been reported for similar protocols using vectors from which pCACTUS was derived (Bloomfield *et al.*, 1991b) and it is possible that the difficulty reported here is a similar problem.

A final attempt to identify the MFRHA was made by over-producing MrhA and MrhB proteins and raising a rabbit antiserum against a semipurified fraction containing both proteins. Therefore, plasmids pPM1642, pPM1643, pPM1644 and pPM1645 were introduced into *E. coli* DH5[pGP1-2]. Metabolic labelling of plasmid-encoded proteins in a small scale T7 over-expression revealed elevated production of proteins of 12 kDa and 25 kDa from pPM1642 but not from pPM1643 (Fig. 7.8). In contrast, pPM1644 was unstable in DH5[pGP1-2]: colonies showed a reduced growth rate and plasmid encoded proteins were not apparent with metabolic labelling, while pPM1645 gave a profile indistinguishable from pPM1643 (not shown). In addition to the 12 kDa and 25 kDa proteins, a protein of about 40 kDa was apparent with pPM1642, pPM1643 and pPM1645. Only proteins cloned downstream and in the correct orientation to be expressed from the T7 promoter should be labelled with this protocol (Tabor and Richardson, 1985) and the vector-encoded *bla* is in the wrong orientation (Promega). Therefore, the identity of this band is not known.

MrhA and MrhB were overproduced on a large scale from *E. coli* DH5[pGP1-2, pPM1642] and the resulting proteins semi-purified as described in Chapter 2. Soluble material consisting of a mixture of both proteins was used to immunise a rabbit. The resulting antiserum reacted strongly with overproduced protein in a western blot but did not give a detectable reaction with MrhA nor MrhB in either *E. coli* AAEC189[pPM471], *V. cholerae* Z17561 or O17 whole cells in western blots (not shown). Similarly the antiserum did not inhibit haemagglutination of the above mentioned strains to a greater degree than pre-immune serum (not shown).

Figure 7.8: Over-expression of MrhA and MrhB proteins.

An autoradiograph of a 20% polyacrylamide gel is shown, on which *E. coli* cells overproducing protein labelled with L-[^{35}S]-Met have been fractionated. Lanes are *E. coli* DH5[pGP1-2] also containing: 1, pGEM3Zf$^+$; 2, pPM1642; 3, pPM1643. The positions of standard molecular weight markers (Chapter 2) are indicated by lines to the left of lane 1. The positions of additional molecular weight markers, Cytochrome C (12.4 kDa) and Aprotinin (6.5 kDa), are indicated by arrows. Plasmid construction is detailed in the text.
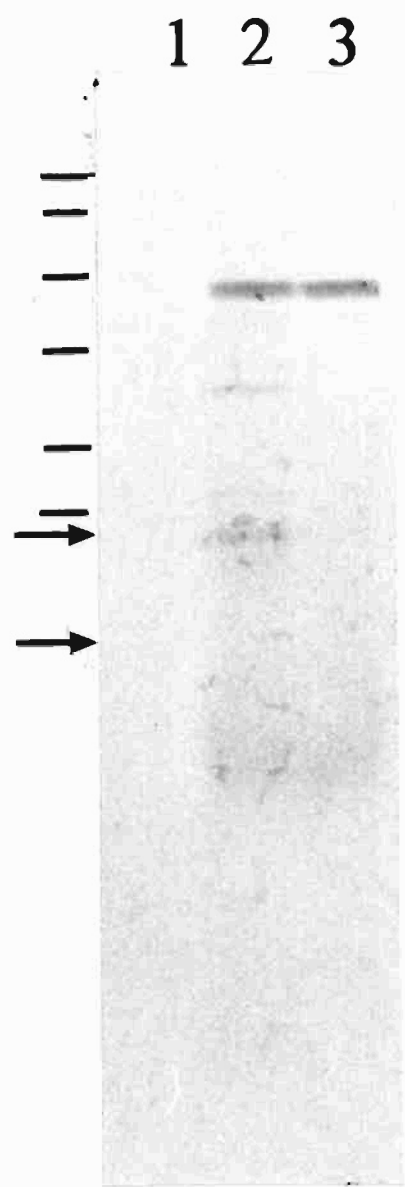
7.2.6 Similarity of MrhB to a family of DNA methyl transferases

The amino acid sequence of both MrhA and MrhB was compared to entries in PIR (Release 38) and SWISS-PROT (Release 27) databases, using the FASTA algorithm (Pearson, 1990). No significant matches were obtained for MrhA (not shown), but MrhB shows similarity with a number of DNA methyl-transferases associated with restriction systems. The most significant pairwise alignments are with: the *Moraxella bovis MboII* methylase, which N6 methylates the 3' adenine in the sequence 5' GAAGA 3' (Bocklage *et al.*, 1991); the *Haemophilus influenzae* Rf *HinfI* methylase, which N6 methylates the adenine in the sequence 5' GANTC 3' (Chandrasegaran *et al.*, 1988a); the *Bacillus amyloliquefaciens* H cellular *BamHI* methylase, which N4 methylates the 5' cytosine in the sequence 5' GGATCC 3' (Vanek *et al.*, 1990); the *Haemophilus haemolyticus HhaI* methylase, of the same specificity as the *HinfI* methylase (Schoner *et al.*, 1983; Chandrasegaran *et al.*, 1988b); the *Streptococcus pneumoniae DpnA* methylase, which N6 methylates the adenine in the sequence 5' GATC 3' (Lacks *et al.*, 1986; de la Campa *et al.*, 1987); the *B. amyloliquefaciens* proviral H2 *BamHI* methylase, of the same specificity as the cellular *BamHI* methylase (Connaughton *et al.*, 1990); the *Enterobacter cloacae Eca*I methylase, which probably N6 methylates the adenine in the sequence 5' GGTNACC 3' (Brenner *et al.*, 1990); and the *Rhodobacter sphaeroides Rsr*I methylase, which N6 methylates the 3' adenine in the sequence 5' GAATTC 3' (Kaszubska *et al.*, 1989) (not shown). As sequence similarities between the methylases listed here have been noted by several authors (Kaszubska *et al.*, 1989; Brenner *et al.*, 1990; Bocklage *et al.*, 1991) a multiple alignment was performed with the sequences listed above (Fig. 7.9). The overall level of similarity between the nine sequences aligned is low, with significant matches between all sequences being limited to two regions. The first centres on residues 27 - 30 of MrhB and is the DPPY motif, thought to be the active site of the methylases, while the second, from residues 183 to 187 is the FXGXG motif (in both MrhB sequences, FFGSG) proposed as the S-adenosyl methionine binding site (Klimasauskas *et al.*, 1989). Other limited similarities can be seen between all sequences in the region of these two blocks (Fig. 7.9). This alignment suggests that MrhB may be a DNA methylase, the significance of which is discussed in Section 7.3.

Figure 7.9: Multiple alignment of MrhB with DNA methyl-transferases associated with restriction enzymes.

Identical sequences are marked with an asterisk, similar residues with a dot. The alignment was performed using the multiple alignment program CLUSTAL (Higgins and Sharp, 1989). The MrhB sequence used is that presented in Chapter 3.

```
MrhB       M-------------------------------------------------------
MboIIM     ML------------------------------------------------------
HinfIM     M-------------------------------------------------------
cBamHIM    MYAVLKNYNVNKKDLLAYKENGVIPTDKVLNGILSYLGMTKVELELKLGR
HhaIIM     MFSHQD-------------------------YLS----------------------
DpnA       M-------------------------------------------------------
vBamHIM    MKE-----------------------------------------------------
EcaIM      MAVGLNK----------KEIDGPASQRAVA----------CDLEPALPP
RsrIM      MANR-SHHNAGHRAMNALRKSG----------------QKHSSESQLG-
                *

MrhB       ------------------SKIYQ----------------------------M--
MboIIM     -------------EINKIHQ-------------------------------M--
HinfIM     -----------------------------MKENINDFLNTILK
cBamHIM    IPAGLEDVFLNNTKEIAKILENKNSVKLNEFNSIQEIKPYFYTDLGKLYN
HhaIIM     --------FVNKKNKMNGI-----------------------------------
DpnA       ---------------------------------TKPYYNKNKMILVH
vBamHIM    ------------------------------------------------------
EcaIM      IGARVTLNYPGKMDE-SIILQKKDTKYLRVGSDFAKESSLILPNSFIWSD
RsrIM      -----------SSEIGTTRH----------------------------VYD

MrhB       --DAVDWLKTLENCSVDLFITDPPYESLEKYRQIVTTTRLKESKSSSNQW
MboIIM     --NCFDFLDQVENKSVQLAVIDPPYN--------LSKADWDSFDSHNE--
HinfIM     G-DCIEKLKTIPNESIDLIFADPPY-FMQTEGKLLRTNGDEFSGVD-DEW
cBamHIM    G-DCLELFKQVPDENVDTIFADPPF-NLDKE------------YDEGVT
HhaIIM     -----DLFKLIPDKAVKIAFFDPQYRGVLD--KMSYGNEGKGRGKERAAL
DpnA       S-DTFKFLSKMKPESMDMIFADPPY-FLSNGG--ISNSGGQVVSVDKGDW
vBamHIM    ---------NIGDCTIDLTVTSPPYDDLRN-------------------
EcaIM      NSLALNRL-MVEGKKAKLIYLDPPYATGMGFSS----------RSNEHAY
RsrIM      VCDCLDTLAKLPDDSVQLIICDPPYNIM--------LADWDDHMD----
                    .              *  .

MrhB       --FSVFP--NTRFEELFREVYRVLKKGSHF-YLF------CDQETMFLAK
MboIIM     --FLPFT--YRWIDKVLDK----LDKDGSL-YIFNT-PFNCAFICQYL--
HinfIM     DKFNDFVEYDSFCELWLKECKRILKSTGSI-WVIGSFQ-N--------IY
cBamHIM    DK-NSFSGYLDWYYKWIDECIRVLKPGGSL-FIYNIPKWN--------TY
HhaIIM     PQMTD-----EIIQQFINEFERVLLPNGYLFLWVDKFH---------LVE
DpnA       DKISSFEEKHEFNRKWIRLAKEVLKPNGTV-WISGSLH-N--------IY
vBamHIM    -----YNGYSFNFEETAQELYRVTKEGGVVVWVVGDKTHKGSETGS--SF
EcaIM      DDCLTEAAYLEFMRRRLILMREILDDDGTI-YV---------HIGHQMLG
RsrIM      --YIGWA--KRW----LAEAERVLSPTGSI-AIFGGLQYQGEAGSGDLIS
                                                        .

MrhB       PIA---ESVGFKFWKP--IVWDKCRIGMGYH---YRARYEFILFFEKGKR
MboIIM     -VS---KGMIFQNW----ITWDK-RDGMGSAKRGFSTGQETILFFSKSKN
HinfIM     RIGYIMQNLDF--WILNDVIWNK-----TNPVPNFGGTR----------
cBamHIM    LSEYLNRKLNFRNWITVDMKFGL-----PIQNRLYPANYSLLYYVKGDKP
HhaIIM     GVKPWLENTP-SLSVVDMLTWDKQKIGMGYRTR---RRSEYLVVIQKEPK
DpnA       SVGMALEQEGF--KILNNITWQK-----TNPAPNLSCRY----------
vBamHIM    RQALYFKELGFNL--HDTMIYEKDSISFPDKNRYYQI-FEYMFIFSKGKP
EcaIM      ELKCLLDEIFGRERFINLITRRKCS-SKNSTKNNFANLNDYILCYSKGKK
RsrIM      IISHMRQNSKMLLANL--IIWN-YPNGMS-AQRFFANRHEEIAWFAKTKK
                                                        .

MrhB       ---KL--NDLSVP----------------------DVLEYKRVWKGY
MboIIM     HTFNY--DEVRVPY-ESTDRIKHASEK----------GILKNGKRWFPN
HinfIM     --FCN--AHETMLW----CSKC-KKNK-----------FTFNYKTM-KHL
cBamHIM    KTFNV--QRIPLQT----CPHCGREIK----------DYGGYKNK-MN-
HhaIIM     K----------------------------------------------
DpnA       --FTH--STETILW----ARKNDKKAR----------HYYNYDLM-KEL
vBamHIM    KTINL-LADRKNKW-YNGKKHIKGHYR----------KMDGEKVRHHKQ
EcaIM      YIWNRPLKKPDAEWLAKEYPKTDSKGQFKLVPIHAPGVRHGATGGEWKGM
RsrIM      YFFDL--DAVREPY-DEETKAAYMKDK----------RLNPESV-----
```

```
MrhB       --P-----------------------------------------------------
MboIIM     --PNGRL---CG---EV--------WHFSSQRHKE----------------
HinfIM     --NQEKQ---ER---SV--------WSLSLCTGKE----------------
cBamHIM    --PKGVT---LS---DV--------WSDIY--------------------
HhaIIM     -------------------AKITWTLHNIP-------------DVW
DpnA       --NDGKQ---MK---DV--------WTGSL---TK----------------
vBamHIM    --NLLKE---FGVRFNI--------WRIPNGHQKS----------------
EcaIM      LPPPGKHWQYTPEKLDILDASGDIHWSKTGNPRRKVYLTDDKSIGYTDYW
RsrIM      --EKGRN---PT---NV--------WRMS---------------------


MrhB       ------------------TEKPVELLEVLIRQSSSENEIVADSFFGSGATL
MboIIM     ----KVNGKTV-KLTHITPKPRDLIERIIRASSNPNDLVLDCFMGSGTTA
HinfIM     ----RIK-DEEGKKAHSTQKPESLLYKVILSSSKPNDVVLDPFFGTGTTG
cBamHIM    ----PVR-HSSSKNRKFNELSVKLLDRIITMSTNEGDVVLDPFGGSGTTF
HhaIIM     AE-------KLQSKPHTHSKPIEMQKQLILATTQEGDLILDPASGGYSVF
DpnA       ----KVE-KWAGK--HPTQKPEYLLERIILASTKEGDYILDPFVGSGTTG
vBamHIM    ----TLDKIAF---QHPAIFPEKLAEDHILSWSNEGDIVFDPFMGSGTTA
EcaIM      EEFRDAHHQSILVTGYPTEKNFNMMKLIVGASSNPGDLVIDPFCGSGSTL
RsrIM      ----RLNGNSLERVGHPTQKPAAVIERLVRALSHPGSTVLDFFAGSGVTA
                          .      .    .         . *     *

MrhB       IAANNLSRKYIGCDIS---MSAHE-------------------------Y
MboIIM     IVAKKLGRNFIGCDMNAEYVNQAN-------------------------F
HinfIM     AVAKALGRNYIGIEREQKYIDVAEKRLREIKPNPNDIELLS--LEIKPPK
cBamHIM    AVSEMLGRKWIGFELGN--CEIIKERL---KNKDKDKKLLGKVYEEKNKL
HhaIIM     ECCKQTNRNFIGCDL----------------------------------
DpnA       VVAKRLGRRFIGIDAEKEYLKIARKRLEA---------------------
vBamHIM    KMAALNNRKYIGTEISKEYCDIANERL-------KNYIILHKRMEGK---
EcaIM      HAASLLQRKWIGIDESLFAAKTVMKRFAIGRAPMGDY-------------
RsrIM      RVAIQEGRNSICTDAAPVF---KE-------------------------Y
              *    *    .

MrhB       FKNR---------------------------------------------
MboIIM     VLNQLEI------------------------------------------
HinfIM     VP--MKTLIEADFLRVGQTLFDKNENAICIVTQDGNVKDNEETLSIHKMS
cBamHIM    FPNRVKELRKKNGLWIDDD-FRQD--------HEGNSKGDKKNENNDQIS
HhaIIM     -------------------------------------------------
DpnA       -------------------------------------------------
vBamHIM    -------------------------------------------------
EcaIM      --------------VNTSLNKQTELPLSL-------NETARHEYVSNDF
RsrIM      YQKQLTFLQ-------DDGLIDK--------------------------


MrhB       ----------------------------------------A
MboIIM     ---------------------------------------N
HinfIM     AKYLNKTNNNGWDYFYLFRNNNFITLDSLRYEYTNQ
cBamHIM    ----------------------LSLE---------
HhaIIM     -IFGDD-----------------------ENEQD
DpnA       ------------------------------ENETN
vBamHIM    ------------GY--------------RLPQVHS
EcaIM      NIYVDELTASVSKNELAEIQKAYRDLKA-----NQQ
RsrIM      ----------ARSYEIVEGAANFGA--ALQRGDVAS
```

7.3 Discussion

As noted in Section 7.1, analysis of the gene replacement mutant engineered by Franzon (1988) implicated the MFRHA in virulence, but the results were not clear in that strain V761 still contained a plasmid, pPH1JI, that was subsequently shown to cause attenuation of *V. cholerae* O1 strains (Focareta, 1989). Strain V762, constructed by curing this plasmid, was also attenuated in the infant mouse model. This result is in agreement with an earlier observation that the expression of MFRHA activity by input bacteria correlates with increased virulence of *V. cholerae* in a rabbit model (Guinée *et al.*, 1985). In addition, the effect of the gene replacement mutant on MFRHA activity *in vitro* had not been determined by Franzon (1988). Examination of this activity in *E. coli* K-12 strains LE392 and HB101 carrying the cloned wild type or mutant genes, and *V. cholerae* 569B or the gene replacement derivative V762, indicated that introduction of the mutation led to a loss of MFRHA activity in both cases. However, attempts to complement the mutation in V762 by reintroduction of the cloned genes was unsuccessful.

The inability to complement this mutation with the cloned genes presents the possibility that a mutation other than that introduced by Franzon (1988) is responsible for the observed attenuation of 569B. However, neither cholera toxin production, motility, nor other potentially attenuating factors analysed (namely LPS structure, ability to chemotact towards glucose and L-methionine, and auxotrophy) were affected in the mutant, although this is by no means an exhaustive survey of such factors. As the mutation introduces a Km^R gene, it is possible that the cost of kanamycin resistance is responsible for the (non-complementable) attenuation of V762, but this possibility seems remote as introduction of the same resistance gene at two other loci, *dns* (Focareta and Manning, 1991), and *rfbT* (Strocher *et al.*, 1992) does not affect virulence. Another possible explanation lies in the choice of cloning vehicle for reintroducing cloned genes into *V. cholerae* V762. All subclones examined were based on the vector pBR322, and this vector is unstable in the absence of continuous selection in *V. cholerae* O1 (Marcus *et al.*, 1990). However, the stability of plasmids examined here was variable in the presence or absence of selection *in vitro*, with no reliable trend discernible (not shown), indicating that plasmid instability is not an adequate explanation for the inability to complement this mutation.

The identification of a suitable vector for complementation studies still presents a problem for the genetic analyses of cloned genes in *V. cholerae*. A pSC101 based vector, pPM2182 (U. Stroeher, personal communication), consisting of the RP4 *mob* region (Simon *et al.*, 1983) cloned into pGB2 (Churchward *et al.*, 1984), has been used successfully as the vehicle for cloned genes in the complementation of introduced mutations in this laboratory (Williams *et al.*, 1993). This vector carries a Sp^R marker, and the possibility exists that spontaneous chromosomally encoded resistance due to *rpsE* mutations (Cerretti *et al.*, 1983) could lead to false positives in viable counts with cells that had lost the plasmid. However, the use of a sufficiently high spectinomycin concentration (100 μg/ml) circumvents this problem, and the vector appears to be stable in the absence of selection (S. Attridge, personal communication). Two other vectors have been examined in this laboratory, pACYC184 (U. Stroeher, personal communication), a derivative of the p15A cryptic plasmid (Chang and Cohen, 1978) and pWSK29 (S. Attridge, personal communication), an Ap^R derivative of pSC101 (Wang and Kushner, 1991), and neither are sufficiently stable for use in complementation analysis. Thus, pGB2, or its mobilisable derivative pPM2182 seem to be the best available vectors for future complementation analysis.

569B expresses a lower level of MFRHA activity *in vitro* than other *V. cholerae* strains examined, so the feasibility of using *in vitro* adherence to a differentiated human colonic adenocarcinoma cell line, Caco-2 (Pinto *et al.*, 1983), was examined. The difference in adherence observed between 569B and its MFRHA$^-$ derivative V762 was approximately three-fold, as opposed to at least sixteen-fold for comparison of the same pair of strains in a haemagglutination assay. Although significant, this low sensitivity may well result from non-specific adherence of bacteria to exposed plastic in the wells of the trays used in the assays (for example, the sides), as *V. cholerae* have been observed to readily adhere to inert surfaces by hydrophobic means (Kabir and Ali, 1983). Alternatively, adherence to the intestinal epithelium *in vivo* has been observed with strains that have lost the ability to haemagglutinate (Teppema *et al.*, 1987), indicating that the residual adherence observed could be due to another MFRHA, such as TCP (Taylor *et al.*, 1987). Therefore, if adherence to Caco-2 cells is to become a useful assay for examining adherence of *V. cholerae in vitro*, further optimisation of assay conditions is necessary. The lower level of MFRHA activity observed

with *E. coli* clones than the homologous *V. cholerae* hosts (with the exception of 569B) is also of interest, especially as transcription of *mrhAB* occurs at a higher level with *E. coli* clones than from the *V. cholerae* chromosomal locus (Chapter 6). A possible explanation of this result was that absence of an appropriate LPS background in the heterologous host led to the MFRHA being unable to form a fully active conformation on the cell surface, but introduction of an Ogawa-expressing LPS clone into *E. coli* HB101[pPM471] did not affect the haemagglutination titre of this strain one way or the other. Thus, the explanation for the difference in MFRHA titre between *E. coli* clones and *V. cholerae* O1 needs to be found elsewhere. This result also suggests that whatever the nature of the MFRHA, it is likely to be extended at least some distance from the cell surface, as a shielding effect (*i.e.* a reduction in titre due to steric hindrance by the LPS chains, of a membrane-located adhesin), similar to that observed with other 'adhesins' such as the *Yersinia pseudotuberculosis* invasin protein (Voorhis *et al.*, 1991), was not observed.

As noted above, the mutation of Franzon (1988) removes most of ORF6 in addition to the distal end of *mrhB*, and comparison of transcription of *mrhA* from the mutant plasmids pPM1146 (a deletion of the 0.7 kb *XbaI* fragment; Franzon, 1988) and pPM1626 (a linker insertion in the proximal *XbaI* site) in *E. coli* AAEC189 indicates that in the absence of ORF6, transcription from p$_{mrh}$ is dramatically reduced. This suggests an involvement of the product of ORF6 in *mrh* expression. This protein does not show any features reminiscent of other proteins (Chapter 3), so further work is required to elucidate its role. This down-regulation raises the possibility that the attenuation observed on introduction of the mutation into 569B is due to the absence of the product of ORF6, a hypothesis supported by the lack of complementation observed following the reintroduction of the minimal clones pPM1127 ($mrhB^+$) and pPM1654 ($mrhA^+B^+$). However, if this were the case, then the mutation should be complementable by the original clone, pPM471, which is not the case. Attempts to construct an ORF6-specific mutation were unsuccessful.

The resolution of the issue of whether *mrhA* or *mrhB* encodes the MFRHA was the major aim of this study. To this end, a new series of deletion mutants was constructed, deficient solely in *mrhA*, or *mrhB*, or both. Examination of the MFRHA activity conferred on

E. coli AAEC189 by these constructs indicated that the presence of either was sufficient. The decision to change the host strain being used was taken because LE392, the strain initially used in this laboratory, is *recA+*, and problems with the stability of plasmids containing the VCR repeat have been observed in *recA+* backgrounds (Chapter 4) and because HB101, the host favoured by DeGraaf and co-workers (van Dongen and DeGraaf, 1986; van Dongen *et al.*, 1987) has been shown to express type I fimbriae under certain conditions (Bloomfield *et al*, 1991a). Therefore, it was surprising that pPM1146, which is not capable of conferring MFRHA activity on LE392 or HB101, confers MFRHA activity on AAEC189. Tracing the pedigrees of these respective strains has revealed that they are derived from different backgrounds: LE392 (L. Endquist, unpublished, cited in Maniatis *et al.*, 1982) is derived from ED8654 (Murray *et al.*, 1977) and shares a common ancestor with the lineage from which AAEC189 is derived (Bloomfield *et al.*, 1991a), but the two strains are separated by at least twelve intermediates, and crosses with other *E. coli* K-12 derivatives are present in both lineages (Bachmann, 1989); HB101 is derived from an *E. coli* K-12/*E. coli* B cross (Boyer and Roulland-Dussoix, 1969). However, there are no obvious known genetic differences that might lead to the observed phenotypic differences (see also below). Transcription of *mrhA* was also analysed in these mutants and in the heterologous host, the loss of *mrhB* did not affect *mrhA* transcription, while the decrease of *mrhA*-specific message observed with the *mrhA* mutants probably reflects the deletion of sequences capable of cross-hybridising with the probe.

The construction of identical sub-clones from the two original sources, pPM471 (Franzon and Manning, 1986) and pVC hag-11 (van Dongen and DeGraaf, 1986) was undertaken to see if comparable subclones would exhibit comparable phenotypes on introduction into the same host strain. The results obtained from these experiments were the same as those obtained in earlier experiments, namely that the presence of either *mrhA* or *mrhB* is sufficient for MFRHA activity in the heterologous host. This is in contrast to the results of Franzon and Manning (1986) and of van Dongen *et al.* (1987). The former concluded that *mrhB* was the structural gene on the basis of analysis of successive subclones generated by *Bal*31 deletion from an upstream *Cla*I site in a derivative of pPM471 called pPM1107. The latter analysed the MFRHA activities associated with restriction fragments

derived from pVC hag-232 (van Dongen and DeGraaf, 1986), a subclone spanning the region from the *Sal*I site present upstream of *mrhA* in C5 (van Dongen and DeGraaf, 1986; van Dongen *et al.*, 1987), but not 569B (Chapter 3), to the *Xba*I site at the distal end of *mrhB*. As the Franzon (1988) gene replacement mutant commences at this *Xba*I site, these results seem entirely contradictory, but the realisation that either *mrhA* or *mrhB* may be sufficient for MFRHA activity in *E. coli* reconciles these results, as Franzon and Manning (1986) would first have removed *mrhA* with no obvious effect as *mrhB* was still present, and van Dongen *et al.* (1987) commenced with a plasmid that was effectively *mrhB⁻*. In fact it is easy to see that if each group had followed different strategies for determining the minimal sequence necessary for MFRHA activity, they might well have arrived at the opposite conclusions from those that they actually did.

Comparison of the sequences obtained for the 569B (Chapter 3) and C5 (van Dongen *et al.*, 1987) *mrhAB* operons has revealed that the two *mrhA* sequences are identical, but that there are four changes between the *mrhB* sequences, all of which lead to substitutions in the predicted protein sequences. These changes are not of great consequence, however, as they do not lead to significant changes in the predicted character of the protein. There are three possible start sites to the *mrhB* open reading frame: van Dongen *et al.* (1987) favoured the first of these in their analysis (at position 2688 (Chapter 3), presumably to cover all options), while the second of these has been favoured here (at position 2727; Chapter 3) as it is the only ATG codon of the three possibilities to be preceded by a possible match to the Shine-Dalgarno consensus sequence. Furthermore, as the first ATG codon is 7 nucleotides downstream of the predicted stem-loop involved in termination of transcription downstream of *mrhA* (Chapter 6), access to a ribosome-binding site in this region is likely to be limited.

Of far greater interest is the result of comparison of the predicted sequences of MrhA and MrhB to entries in the PIR and SWISS-PROT databases. No significant matches were found for MrhA, but MrhB shows similarity to a number of DNA methyl transferases, listed in Section 7.2.4. A multiple alignment of these sequences revealed two blocks of similarity between all sequences, that centred on the proposed methylase active site and the S-adenosyl methionine binding site of these proteins (Klimasauskas *et al.*, 1989). Removal of MrhB

from the sequences compared in this multiple alignment does not lead to a significant change in the alignment (not shown). Although predictions of protein secondary structure by methods such as that of Chou and Fasman (1978) are not overly reliable, similarities between the two blocks of similarity of the various DNA methyl-transferases have been noted (Guschlbauer, 1988). Comparison of the secondary structure predictions for the two MrhB sequences has been undertaken here, and similarities with the predicted secondary structures of other DNA methyl-transferases are apparent (not shown). Thus, it seems likely that MrhB is a DNA methyl-transferase.

How does this revelation affect our understanding of the nature of the MFRHA? Phase variation of the major subunit gene (*papA*) of pyelonephritis associated pili in uropathogenic *E. coli* is controlled by differential inhibition of DNA methylation at one of two GATC (Dam) sites upstream of the *papBA* promoter: one of these sites is unmethylated only in transcriptionally active (phase on) cells, the other only in transcriptionally silent (phase off) cells (Blyn *et al.*, 1990). This effect is mediated by a locus, initially called *mbf* for methylation blocking factor, which is required for *pap* transcription (Braaten *et al.*, 1991) and which has subsequently been shown to be synonomous with the leucine-responsive regulatory protein, or Lrp (Platko *et al.*, 1990; Braaten *et al.*, 1992). Lrp has been identified independently several times as the regulator of various members of its regulon, many of which are involved in amino acid metabolism (Newman *et al.*, 1992), in a manner similar to another regulator of virulence genes in *E. coli* that also regulates other metabolic functions, the histone-like protein H-NS (Dorman, 1991). Although the environmental signals associated with the regulon involving H-NS are clear, namely medium osmolarity and temperature (Dorman, 1991), no such physiological trigger is apparent for Lrp. Newman *et al.* (1992) have suggested that the Lrp regulon allows *E. coli* to distinguish between "feast and famine", in other words whether the cell is in an intestinal (rich) or external (lean) environment. Complicating this regulatory issue is that not all genes whose expression is affected by Lrp are sensitive to the level of leucine in the growth medium (Newman *et al.*, 1992). For example, expression of the *fan* operon, encoding K99 fimbriae, is significantly reduced by exogenous leucine, while expression of the *pap* operon is unresponsive to leucine (Braaten *et al.*, 1992). Expression of the regulatory gene *lrp* is itself unaffected by exogenous

leucine, although it is subject to the control of Lrp (Newman *et al.*, 1992). Lrp seems to be involved in methylation at sites other than the *pap* promoter, as *metK* mutant strains, in which most S-adenosyl methionine synthetase activity is lost, also carry a supressor mutation which is often an *lrp* allele (Lin *et al.*, 1990). In the case of *pap* phase variation, Lrp seems to act by binding near and sterically hindering access to one or other Dam site involved in phase variation (Braaten *et al.*, 1992). It is not clear how this state of unmethylation affects promoter activity, but by analogy with the inhibition of gene expression by cytosine methylation at CpG sites in eukaryotes (Razin and Cedar, 1991), the phase on methylation pattern should derepress the *papBA* promoter (Blyn *et al.*, 1990).

DNA methylation states controlling *pap* phase variation provide a precedent for suggesting a similar activity being involved somehow in MFRHA activity associated with *mrhAB*. Experiments reported to date have not precluded the possibility that the cloned genes are regulators and not effectors of adherence: resolution of this awaits the availability of suitable antisera. Similarly, attempts to examine the effect of Dam methylation on MFRHA activity in *E. coli* were unsuccessful, due to the inability to transform a *dam*, *dcm* mutant strain, *E. coli* JM110 (Yanisch-Perron *et al.*, 1985), with constructs bearing an intact *mrhA* allele, regardless of the presence or absence of *mrhB* (not shown). Assuming however, that the *mrhAB* operon does indeed encode an adhesin, the available data best fits a model in which MrhA is the adhesin and MrhB performs a putative regulatory role that includes activation of an as yet uncharacterised mannose resistant haemagglutinin in the heterologous (*E. coli*) host. The regulation of expression of both these proteins may also require expression of the neighbouring ORF6. Such functions are of course still extremely speculative, and further analysis is needed in order to unravel the Gordian knot that the attempt to identify the effector of MFRHA activity reported here, has created.

Chapter 8

Perspectives

8.1 Genetic organisation of the region encoding the MFRHA

The initial stages of this study involved the characterisation of the original MFRHA clone, pPM471 (Franzon and Manning, 1986). The nucleotide sequence of portions of this clone was determined by Franzon (1988) and Clark (1988). Additional sequencing revealed the presence of at least three copies of a direct repeat originally identified flanking the MFRHA structural genes (van Dongen *et al.*, 1987) and probing pPM471 DNA with an oligodeoxynucleotide probe specific for this repeat revealed the presence of multiple copies. In order to undertake a more extensive analysis of these repeats, the nucleotide sequence of the entire insert of pPM471 was determined.

8.1.1 Open reading frames linked to the genes encoding the MFRHA

There are ten complete open reading frames within the insert of pPM471 (Chapter 3). Two of these have been suggested to be the structural genes for the MFRHA (Franzon and Manning, 1986; van Dongen and DeGraaf, 1986; van Dongen *et al.*, 1987; Franzon, 1988), and were designated *mrhA* and *mrhB*, to reflect their location relative to one another. These two genes lie between two copies of the direct repeat sequence, assigned the acronym VCR. The other complete ORFs are all flanked by copies of VCR, with the exception of another pair of ORFs that are predicted to form an operon and to be translationally coupled. Comparison of the predicted amino acid sequences of these other ORFs to entries in databases has revealed similarity in only three cases: the functions of the proteins as suggested by these similarities precludes any obvious role for them in the haemagglutinating phenotype. On the contrary, it seems unlikely that these proteins function in concert as one, ORF2, is related to a superfamily of soluble proteins involved in the transport of hydrophobic molecules in eukaryotes, the second shows similarity to the amino-terminal half of an *E. coli* protein involved in the priming step of DNA replication and the third shows similarity to a hypothesised translational inhibitor involved in the stringent response in *E. coli*. Speculation on the functions of the other five ORFs is not possible, as they do not show similarity to sequences in the databases, and are not predicted to be exported from the cytoplasm. Two partial ORFs are also apparent at the ends of the insert in pPM471. However, neither of these partial sequences shows similarity to known sequences, so again, the roles (if any) of these sequences cannot yet be determined. The genetic organisation observed within the insert of pPM471 does not show similarity with characterised fimbrial gene clusters, suggesting, along

with the lack of sequence similarity between MrhA, MrhB and known components of fimbrial systems, that the MFRHA is not fimbrial in nature.

8.1.2 A direct repeat associated with the genes encoding the MFRHA

As mentioned above, the presence of nine copies of a repeat sequence named VCR throughout the insert of pPM471, is of considerable interest. The repeats are 124 nucleotides in length, and exhibit numerous short stretches of internal inverted repeat structure (van Dongen *et al.*, 1987; Chapter 4). Furthermore, the entire repeat shows imperfect dyad symmetry, with enough consistent mis-matches between the two halves to be able to assign an orientation to the sequence: all copies of VCR present on pPM471 have the same orientation. Additional copies of VCR were shown to exist on the chromosome of *V. cholerae* O1, and although almost certainly an underestimate, quantitation of the intensities of bands hybridising with a VCR-specific probe revealed that there are around 100 copies of VCR in classical strains, and 60 to 80 copies of VCR in El Tor strains of *V. cholerae* O1. Large scale mapping, using pulsed field gel electrophoresis, raised the possibility of most if not all copies of VCR being confined to less than 10% of the *V. cholerae* O1 chromosome. Such a distribution is in marked contrast to the distribution of repeat sequences in the chromosomes of other eubacteria. The best characterised of these repeats, REP sequences, are distributed more or less randomly throughout the *E. coli* chromosome (Dimri *et al.*, 1992), a distribution that is compatible with the proposed role of REP sequences in providing sites on the chromosome (DNA) upon which assembly of scaffold proteins, such as DNA gyrase, the histone like protein HU and DNA polymerase I, can occur (Yang and Ames, 1988; Yang and Ames, 1990; Gilson *et al.*, 1990b). The distribution of REP sequences also roughly matches the distribution of sites of DNA gyrase binding (Condemine and Smith, 1990). Other functions attributed to REP sequences, such as acting as transcriptional terminators (Gilson *et al.*, 1986b), stabilisers of mRNA (Newbury *et al.*, 1987), or the end points in illegitimate recombination events (Shyamala *et al.*, 1990), seem to be secondary to this role. However, an alternative hypothesis that REP sequences are "selfish" DNA, which would also explain the sequence conservation between REP sequences (Higgins *et al.*, 1988), has not been discounted.

As the distribution of VCR sequences is limited to approximately 10% of the *V. cholerae* chromosome, a function similar to that proposed for REP sequences, seems unlikely. At least one copy of VCR within pPM471 acts as a transcriptional terminator (Chapter 6), although it seems unlikely that this is a general property attributable to VCR (*cf.* the subset of REP sequences that act as transcription terminators; Gilson *et al.*, 1986b). RNase protection experiments, in which a VCR-specific probe was hybridised with total *V. cholerae* RNA, indicated that at least part of many copies of VCR are transcribed (and in both orientations), and that some of these transcripts contain full length copies of VCR. Comparison of the nucleotide sequence of the consensus sequences for REP, VCR and a known DNA gyrase binding site led to the suggestion that VCR might also bind DNA gyrase (Chapter 4), as has been shown for REP (Yang and Ames, 1988; Yang and Ames, 1990). Attempts to isolate *gyr* mutants in *V. cholerae* strains Z17561 and O17, by virtue of nalidixic acid resistance were unsuccessful, even after chemical mutagenesis. The isolation of Nal^R mutants has been reported, but required a "two step" process (which was not defined) (Parker *et al.*, 1979). However, Nal^R mutants were obtained readily from strain C31, in which the entire chromosomal region represented by the clone pPM471 has been deleted. The significance of this observation will be returned to in Section 8.2. Attempts to verify that these Nal^R strains were *gyr* mutants, and to use them to probe possible interactions between DNA gyrase and VCR, were unsuccessful. Therefore, suggestions as to a function for VCR remain speculative.

The presence of multiple copies of VCR in a direct repeat orientation leads to extreme instability when present in multicopy plasmids in *E. coli* K-12 *recA*$^+$ strains. A *V. cholerae* O1 strain (C31) has been identified, which harbours a deletion encompassing the entire region represented in the clone pPM471. Although mapping of the deletion end points was not attempted, it has been proposed that this deletion may have arisen by homologous recombination between two copies of VCR leading to deletion of the intervening sequence. If all chromosomal copies of VCR are of the same orientation as those found within the sequenced region, then this 'deletogenic' orientation might be expected, with time, to lead to the gradual loss of VCR from the *V. cholerae* chromosome. With the exception of strain C31, this does not appear to be the case, as strains isolated over a forty year period show

remarkable consistency in hybridisation pattern within biotypes. Furthermore, at least part of the sequence contained within the region represented by pPM471 appears to have been duplicated, in that two hybridising bands are detected in strain 569B, when fragment probes corresponding to the proximal end of pPM471, and ORF2 are used. Single bands, probably corresponding to the second copy detected in 569B, are detected in strain C31. Furthermore, DNA gel blot analysis of a number of *V. cholerae* O1 strains with a fragment probe corresponding to ORF2 revealed the presence of up to four copies of ORF2-related sequences (Chapter 5). Although these related sequences have not been demonstrated to be associated with copies of VCR, this result raises the possibility of extensive rearrangements having occurred within the VCR-linked region of the chromosome. In addition a single, rare, ST$^+$ isolate of *V. cholerae* O1 has been identified, in which the ST toxin gene is flanked by copies of VCR (Takeda *et al.*, 1991a). Thus, VCR does not seem to be 'on the way out', but instead could somehow be involved in the acquisition of genes from foreign sources. VCR-related sequences have been detected in non-O1 *V. cholerae* and also in *V. mimicus* by Takeda *et al.* (1991a). The numbers in non-O1 *V. cholerae* seem to be much more variable, ranging from a few copies to numbers comparable to O1 strains (Chapter 4), although quantitation of the hybridising bands was not possible, due to the absence of a band containing known numbers of VCR. This variability in copy number could also be explained by sequence variation between VCRs in non-O1 strains. However, the survey reported here is hardly exhaustive, and further work is needed to resolve these issues.

Alternatively, a single copy of VCR, or the entire region, could have been imported from a foreign source. However, if a single copy were imported, then any mechanism proposed for its spread should explain how progeny copies have remained relatively localised, rather than being dispersed at random. Similarly, it seems unlikely that the entire VCR region was imported into *V. cholerae*, because if this were the case, then the variability in VCR copy number between O1 and non-O1 strains would need to be explained by differential stability of the region between strains.

8.1.3 A haemin-binding protein gene linked to the genes encoding the MFRHA

As mentioned in Section 8.1.1, one of the genes encoded within pPM471, ORF2, shows significant sequence similarity with the α_2-microglobulin (lipocalin) superfamily. This similarity was of sufficient interest for further analysis of ORF2 and its product to be undertaken. The predicted amino acid sequence of the product of ORF2 has features characteristic of the signal peptide and cleavage site of bacterial lipoproteins at its amino terminus. Analysis has shown that the protein is processed in *E. coli* K-12 in the absence, but not the presence of the antibiotic globomycin, an inhibitor of type II signal peptidase that is responsible for the processing of lipoproteins. In addition, the protein can only be labelled with ^{3}H-palmitate in the absence of globomycin, showing that it is a lipoprotein in *E. coli*, and is thus extremely likely to be a lipoprotein in *V. cholerae* as well. On this basis, ORF2 was named *vlpA* (Chapter 5).

A function for VlpA was suggested by its similarity to several members of the α_2-microglobulin superfamily, as the best matches are to a subset of lipocalins that are involved in porphyrin binding (Peitsch and Boguski, 1990). Plate binding experiments using a temperature-inducible clone showed that VlpA conferred on *E. coli* the ability to bind the porphyrins haemin, haematoporphyrin and to a lesser extent protoporphyrin IX, as well as the dye Congo red. Liquid binding assays confirmed this result for haemin, and suggested that the interaction was hydrophobic in nature. Furthermore, the construction of a series of gene fusions between the mature portion of *phoA* and progressive carboxy-terminal truncations of *vlpA* showed that this binding activity is gradually lost, and is absent when approximately half of *vlpA* is removed. Comparison of VlpA with similar proteins for which detailed molecular structures are available (Holden *et al.*, 1987; Huber *et al.*, 1987) suggests that the loss of residues responsible for ligand binding is responsible for this loss of activity. Thus, VlpA seems to act as a porphyrin binding protein. A haemin uptake system has been identified in *V. cholerae* O1, which bears no resemblance to VlpA, and which is sufficient to confer the ability to utilise haemin as a sole iron source on *E. coli* (Henderson and Payne, 1993). Hence, if VlpA is involved in iron scavenging, it most likely plays a secondary role to this alternative system, perhaps more analogous to the haemin binding properties associated with the virulence plasmid of *Shigella flexneri* (Stugard *et al.*, 1989) or the surface array virulence

protein of *Aeromonas salmonicida* (Kay *et al.*, 1985). It should be noted that the above experiments were performed with constructs allowing expression of *vlpA* from a bacteriophage T7 promoter, as conditions under which *vlpA* is expressed were not identified.

Despite having functional similarities to porphyrin-binding lipocalins, in addition to relatedness in primary structure, the structure of VlpA represents an interesting diversion from other members characterised to date. First, the protein is a lipoprotein and is predicted to be at least membrane-associated, unlike other members characterised to date, which are soluble proteins (Sawyer, 1987; Godovac-Zimmermann, 1988). Second, an important structural feature of the haemin binding lipocalins is a pair of disulphide bridges, which holds the β-barrel globular structure of the proteins together (Holden *et al.*, 1987; Huber *et al.*, 1987; Peitsch and Boguski, 1990). This feature is entirely absent from VlpA, as the only cysteine residue in the mature portion of the protein is the amino terminal residue that presumably is the site of fatty acylation. Thus, solving a structure for VlpA could provide insights into the maintenance of structure of these proteins, a task currently being undertaken elsewhere.

8.2 Expression of the genes encoding the MFRHA

The genes encoding the MFRHA were cloned from the classical strain 569B (Franzon and Manning, 1986) and the El Tor strain C5 (van Dongen and DeGraaf, 1986). Analysis of subclones derived from these original clones led to conflicting conclusions regarding which open reading frame represented the structural gene for the MFRHA, either a 7 kDa ORF (van Dongen *et al.*, 1987), or a 25 kDa ORF, here designated *mrhA* and *mrhB*, respectively (Franzon and Manning, 1986). The two genes in question are adjacent, and lie between two copies of VCR, an organisation that led van Dongen *et al.* (1987) to propose that they may be co-transcribed. Extensive analysis of *mrhA* and *mrhB* mRNA in both *V. cholerae* and *E. coli* K-12 clones, and of the activities of reporter gene constructs derived from segments of the two genes in *E. coli* K-12, has demonstrated that this is the case. The two genes are transcribed from a single promoter upstream of *mrhA*, with the bulk of the transcripts terminating immediately 3' to a predicted secondary structure lying between the two genes.

Transcription of *mrhB* occurs by read-through of this structure, and terminates within the downstream copy of VCR.

Given the postulated role in virulence for the MFRHA, the factors influencing expression of *mrhAB* are of interest. Examination of MFRHA activity in defined mutants suggests that expression is independent of the regulators ToxR or HlyU. Similarly, analysis of *mrhA* transcription has shown that while transcription of the toxin *hlyA*, whose expression is affected by the availability of free iron (Stoebner and Payne, 1988), is activated by the addition of the iron chelator 2,2'-dipyridyl to exponential phase cultures, transcription of *mrhA* is actually repressed by addition of the iron chelator. Fur, the regulator presumably responsible for the iron-dependence of transcription of *hlyA*, also has been shown to act as a transcriptional activator in *Salmonella typhimurium* (Foster and Hall, 1992), but it has been suggested (Chapter 6) that this apparent repression by the sequestering of iron may be due to a secondary effect, such as reduced growth rate, in the stressed cultures. Analysis of *mrhA* transcription in a series of *mrhA* and *mrhB* mutants has suggested that the downstream ORF6 may also be involved in transcription from p_{mrh}, as transcription is dramatically reduced when this ORF is removed along with the distal portion of *mrhB*, but not when the distal portion of *mrhB* alone is removed (Chapter 7). This result seemingly contradicts the activity associated with a promoter-reporter gene construct, which gave significant activity from a fragment containing p_{mrh}, as confirmed by primer extension analysis (Chapter 6). However, activity as measured from p_{mrh} is markedly lower than that measured from p_{lac}, so the possibility that the provision of the product of ORF6 would lead to increased expression from the reporter construct, should not be dismissed.

An interesting feature of p_{mrh} is the spacing between the predicted -35 and -10 boxes. The consensus spacing between these boxes for *E. coli* σ^{70} promoters is 17 nucleotides (O'Neill, 1989); the observed spacing for p_{mrh} is 14 nucleotides (Chapter 6). Similarly, the promoter upstream of ORF6 was mapped, and the predicted -35 and -10 boxes were also separated by 14 nucleotides. Promoter element spacing provides a mechanism by which the state of DNA supercoiling might affect transcription, with an increase in the state of negative supercoiling of DNA shifting expression toward promoters with greater than consensus

spacing, while relaxation of negative supercoiling would shift expression toward promoters of less than consensus spacing (Wang and Syvanen, 1992). Relaxation of DNA supercoiling is known to occur in *E. coli* during the transition from exponential growth to stationary phase (Balke and Gralla, 1987), and the expression of a number of environmentally regulated genes is controlled by changes in the level of DNA supercoiling in a variety of pathogenic bacteria (Dorman, 1991). As MFRHA activity appears during late exponential phase and is maximal in stationary phase (Hanne and Finkelstein, 1982), this suggests a role for DNA supercoiling in MFRHA expression. Both p_{mrh} and p_{ORF6} are close to VCR sequences, predicted in Chapter 4 to be capable of interacting with DNA gyrase. If this were the case then conditions that led to gyrase interaction with VCR would be expected to repress expression from these adjacent promoters, in the absence of compensating regulatory factors. However, such an assertion is extremely speculative, and mutagenesis in the p_{mrh} and p_{ORF6} promoter regions, as well as the availability of reliable *gyr* mutants is required for its testing.

On the other hand, the promoter upstream of ORF3.1 and ORF3.2, which presumably directs expression of both ORFs, has also been successfully mapped in pPM471 (but not in *V. cholerae*), and the predicted -35 and -10 boxes of this promoter show consensus spacing. This promoter is divergent from p_{mrh} and analysis of promoter-reporter gene fusion constructs in *E. coli* suggests that it is weaker than p_{mrh}. Similarities of the predicted protein products of ORF3.1 and ORF3.2 with known sequences (*dnaT* and *relE* of *E. coli*, respectively) has suggested functions for them (Section 8.1.1): in particular ORF3.2 is predicted to encode a translational inhibitor. Therefore, expression of these ORFs is not expected under normal *in vitro* growth conditions. As promoter mapping was not successful in *V. cholerae* it seems possible that the ORF3.2 present in pPM471 has mutated, so that it is no longer functional. A *relE* allele has been described in which RelE has lost activity because of an $R_{45} \rightarrow T$ mutation (Bech *et al.*, 1985) - the corresponding amino acid in ORF3.2 is S_{45}. On the other hand, attempts to obtain mutants in ORF3.1 and ORF3.2 were unsuccessful, which was attributed to the lethality of the desired products (Chapter 3). Nevertheless, the expression or otherwise of these ORFs needs to be considered in any model invoking DNA supercoiling, as the promoter as mapped is also adjacent to a copy of VCR.

Attempts to elucidate whether *mrhA* or *mrhB* encodes the MFRHA have shown that either gene is sufficient to confer MFRHA activity on *E. coli* cells (Chapter 7). Experimental difficulties have led to the situation in *V. cholerae* being less clear, but a mutation that removes the distal portion of *mrhB* and most of the adjacent ORF6 abolishes both MFRHA activity and virulence of 569B. As discussed above, the ORF6 mutation probably leads to down regulation of *mrhAB* transcription. MrhA does not show sequence similarity to entries in the databases, but MrhB is similar to a number of DNA methyl transferases associated with type II restriction enzymes. As inhibition of *dam*-mediated DNA methylation by a global regulator has been shown recently to control phase variation of the *pap* fimbrial adhesin genes in *E. coli* (Braaten *et al.*, 1992), it is possible that the putative DNA methylating activity of MrhB fulfills some regulatory function. MrhB is not essential for activity of p_{mrh} (Chapter 7), suggesting that if it is a regulator, then it is not autoregulatory. The transcriptional organisation of the *mrhAB* operon, with read-through of a terminator leading to *mrhB* transcription is reminiscent of the *senS* regulatory gene in *Bacillus subtilis*, in which the structural gene is preceded by a transcriptional terminator (Wang and Doi, 1990). This mechanism seems to be in place to avoid lethal over-production of the *senS* gene product, a situation that does not appear to be so crucial for *mrhB*, at least in *E. coli* K-12 subclones, as removal of the predicted stem-loop structure between *mrhA* and *mrhB*, or expression from vector promoters, has no apparent effect on the host (Chapter 7). Similarly, up-regulation of the *kdpDE* operon, encoding the regulators of the *kdpABC* operon responsible for potassium uptake in *E. coli*, occurs *via* read-through of an upstream terminator during induction of *kdpABC* expression (Polarek *et al.*, 1992). Thus, although the biochemical nature of the products of the *mrhAB* operon remain to be satisfactorily explained, the transcriptional organisation could be explained as a structural gene cotranscribed with a regulator, and a second, adjacent, regulatory gene being transcribed from a separate promoter.

8.3 Future directions

This project has examined questions on the nature of the MFRHA and the control of its expression. The genes associated with MFRHA activity occupy an interesting region of the *V. cholerae* chromosome, and may be linked to other genes involved in virulence. Three genes have now been implicated in the phenotype, at least in *E. coli* clones, but the exact role

played by each is still largely speculative. In particular, the possibility remains that a regulator rather than an effector of haemagglutination has been cloned, that is inducing expression of one MFRHA in *V. cholerae* and another in *E. coli*. Experiments along two lines are necessary to resolve the problems established here. One is to introduce mutations specific for *mrhA*, *mrhB*, or ORF6, and combinations of these, into *V. cholerae* so that their effects on haemagglutination and virulence can be determined. The second is the development of effective antisera against these proteins, which would allow an examination of the cellular location of the proteins, and, *via* inhibition of haemagglutination, determination of whether *mrhAB* encodes a regulator, or effector, of adherence.

Bibliography

Abdulkadir, S.A. 1991. Cholera and cholera-like diarrhoeal diseases: why the differences in severity? *Med Hypotheses* **34**: 278-281.

Achtman, M., Schwukow, S., Helmuth, R., Morelli, G. and Manning, P.A. 1978. Cell-cell interactions in conjugating *Escherichia coli*: Con⁻ mutants and stabilisation of mating aggregates. *Molec Gen Genet* **164**: 171-183.

Adamowicz, M., Kelley, P.M. and Nickerson, K.W. 1991. Detergent (sodium dodecyl sulphate) shock proteins in *Escherichia coli*. *J Bacteriol* **173**: 229-233.

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.

Albert, M.J., Ansaruzzaman, M., Bardhan, P.K., Faruque, A.S.G., Faruque, S.M., Islam, M.S., Mahalanabis, D., Sack, R.B., Salam, M.A., Siddique, A.K. and Yunus, M. 1993a. Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. *Lancet* **342**: 387-390.

Albert, M.J., Siddique, A.K., Islam, M.S., Faruque, A.S.G., Ansaruzzan, M., Faruque, S.M. and Sack, R.B. 1993b. Large outbreak of clinical cholera due to *Vibrio cholerae* non-O1 in Bangladesh. *Lancet* **341**: 704.

Alifano, P., Rivellini, F., Limauro, D., Bruni, C.B. and Carlomagno, M.S. 1991. A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* **64**: 553-363.

Alm, R.A. 1989. *Molecular characterisation of the haemolysin determinant of* Vibrio cholerae *O1*. PhD thesis. University of Adelaide, Adelaide, S.A. Australia.

Alm, R.A. and Manning, P.A. 1990a. Characterisation of the *hlyB* gene and its role in the production of the El Tor haemolysin of *Vibrio cholerae* O1. *Mol 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., Mayrhofer, G., Kotlarski, I. and Manning, P.A. 1991. The amino-terminal domain of the El Tor haemolysin of *Vibrio cholerae* O1 is expressed in classical strains and is cytotoxic. *Vaccine* **9**: 588-594.

Alm, R.A., Stroeher, U.H. and Manning, P.A. 1988. Extracellular proteins of *Vibrio cholerae*: nucleotide sequence of the structural gene (*hlyA*) for the haemolysin of the haemolytic El Tor strain O17 and characterisation of the *hlyA* mutation in the non-haemolytic classical strain 569B. *Mol Microbiol* **2**: 481-488.

Almeida, R., Chen, F., Brenner, F., Evins, G., Cook, W.L., Blake, P.A. and Wachsmuth, I.K. 1989. United States isolates of nontoxigenic *Vibrio cholerae* O1. In *Proceedings of the 25th joint conference on cholera and related diarrhoeal diseases.* National Institutes of Health. Bethesda, Md. USA. p 61.

Ames, P. and Parkinson, J.S. 1988. Transmembrane signalling by bacterial chemoreceptors: *E. coli* transducers with locked signal output. *Cell* **55**: 817-826.

Andersson, S.G.E. and Kurland, C.G. 1990. Codon preferences in free-living microorganisms. *Microbiol Rev* **54**: 198-210.

Asakura, H., Morita, A., Morishita, T., Tsuchiya, M., Fukumi, H., Ohashi, M., Ulyangco, C. and Castro, A. 1973. Pathologic findings from intestinal biopsy specimens in human cholera. *Am J Dig Dis* **18**: 271-278.

Ascencio, F., Ljungh, A. and Wadstrom, T. 1992. Lactoferrin binding properties of *Vibrio cholerae*. *Microbios* **70**: 103-117.

Attridge, S.A. 1979. *Infection by, and immunity to*, Vibrio cholerae. PhD thesis. University of Adelaide, Adelaide, S.A. Australia.

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.

Attridge, S.R. and Rowley, D. 1983c. Prophylactic significance of the nonlipopolysaccharide antigens of *Vibrio cholerae*. *J Infect Dis* **148**: 931-939.

Attridge, S.R., Voss, E. and Manning, P.A. 1994. The role of toxin-coregulated pili in the pathogenesis of *Vibrio cholerae* O1 El Tor. *Microb Pathog* In press.

Bachmann, B.J. 1989. Derivatives and genotypes of some mutant derivatives of *Escherichia coli* K-12. In Escherichia coli *and* Salmonella typhimurium: *cellular and*

molecular biology Vol II. Neidhardt, F.C. (ed). American Society for Microbiology. Washington, DC. USA. pp1190-1219.

Båga, M., Görannson, M., Normark, S. and Uhlin, B.E. 1988. Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. *Cell* **52:** 197-206.

Bagg, A. and Neilands, J.B. 1987. Molecular mechanism of regulation of siderophore-mediated iron assimilation. *Microbiol Rev* **51:** 509-518.

Baine, W.B., Vasil, M.L. and Holmes, R.K. 1978. Genetic mapping of mutations in independently isolated nontoxigenic mutants of *Vibrio cholerae. Infect Immun* **21:** 194-200.

Bales, G.L. and Lankford, C.E. 1961. An adherence factor of *Vibrio cholerae.* In *Bacteriol Proc Abstr* **M64:** 118.

Balke, V.L. and Gralla, J.D. 1987. Changes in the linking number of supercoiled DNA accompanying growth transitions in *Escherichia coli. J Bacteriol* **169:** 4499-4506.

Banerjee, K.K., Ghosh, A.N., Dutta-Roy, K., Pal, S.K. and Ghose, A.C. 1990. Purification and characterisation of a novel haemagglutinin from *Vibrio cholerae. Infect Immun* **58:** 3698-3705.

Bardwell, J.C.A., McGovern, K. and Beckwith, J. 1991. Identification of a protein required for disulphide bond formation *in vivo. Cell* **67:** 581-589.

Barkla, D.H., Whitehead, R.H. and Hayward, I.P. 1992. Effects of cholera toxin on human colon carcinoma cell lines. *Path* **24:** 296-301.

Barnes, R.D. 1987. *Invertebrate Zoology*, 5th edn. Saunders College/Harcourt Brace Jovanovich. Orlando, Fla. USA.

Barnes, W.M., Bevan, M. and Son, P.H. 1983. Kilo sequencing: creation of an ordered nest of asymmetric deletions across a large target sequence carried on phage M13. *Meth Enzymol* **101:** 98-122.

Bartowsky, E.J., Attridge, S.R., Thomas, J.C., Mayrhofer, G. and Manning, P.A. 1990. Role of the P plasmid in attenuation of *Vibrio cholerae* O1. *Infect Immun* **58:** 3129-3134.

Bartowsky, E.J., Morelli, G., Kamke, M. and Manning, P.A. 1987. Characterisation and restriction analysis of the P sex factor and the cryptic plasmid of *Vibrio cholerae* strain V58. *Plasmid* **18**: 1-7.

Barua, D. 1988. Cholera during the last hundred years (1884-1983). In Vibrio cholerae *and cholera*. Takeda, Y. and Kuwahara, S. (eds). KTK Scientific Publishers. Tokyo. Japan. pp 9-32.

Barua, D. and Burrows, W. 1974. *Cholera.* W.B. Saunders. Philadelphia, Pa. USA.

Barua, D. and Paguio, A.S. 1977. ABO blood groups and cholera. *Am Hum Biol* **4**: 489-492.

Baudry, B., Fasano, A., Ketley, J. and Kaper, J.B. 1992. Cloning of a gene (*zot*) encoding a new toxin produced by *Vibrio cholerae*. *Infect Immun* **60**: 428-434.

Baudry, B., Fasano, A., Ketley, J.M. and Kaper, J.B. 1991. Cloning and characterisation of the *Vibrio cholerae zot* gene encoding a toxin which affects intestinal tight junctions. In *Proceedings of the 27th joint conference on cholera and related diarrhoeal diseases.* National Institutes of health. Bethesda, Md. USA. pp 46-47.

Baumann, P., Furniss, A.L. and Lee, J.V. 1984. Genus I. *Vibrio* Pacini 1854, 411^{AL}. In *Bergeys Manual of systematic bacteriology Vol I.* Krieg, N.R. and Holt, J.G. (eds). Williams and Wilkins. Baltimore, Md. USA. pp 518-538.

Bäumler, A. J. and Hantke, K. 1992. A lipoprotein of *Yersinia enterocolitica* facilitates ferrioxamine uptake in *Escherichia coli. J Bacteriol* **174**: 1029-1035.

Bech, F.W., Jørgensen, S.T., Diderichsen, B. and Karlström, O.H. 1985. Sequence of the *relB* transcription unit from *Escherichia coli* and identification of the *relB* gene. *EMBO J* **4**: 1059-1066.

Bedard, P.-A., Yannoni, Y., Simmons, D.L. and Erikson, R.L. 1989. Rapid repression of quiescence-specific gene expression by epidermal growth factor, insulin, and pp60^{v-src}. *Mol Cell Biol* **9**: 1371-1375.

Bellamy, J.E.C., Knop, J., Steele, E.J., Chaicumpa, W. and Rowley, D. 1975. Antibody cross-linking as a factor in immunity to cholera in infant mice. *J Infect Dis* **132**: 181-188.

Bennun, F.R., Roth, G.A., Monferran, C.G. and Cumar, F.A. 1989. Binding of cholera toxin to pig intestinal mucosa glycosphingolipids: relationship with the ABO blood group systems. *Infect Immun* **57:** 969-974.

Beringer, J.E., Benyon, J.L., Buchanan-Wollaston, A.V. and Johnston, A.W.B. 1978. Transfer of drug resistance transposon Tn*5* to *Rhizobium*. *Nature* **276:** 633-634.

Berman, P., Gray, P., Chen, E., Keyser, K., Ehrlich, D., Karten, H., LaCorbiere, M., Esch, F. and Schubert, D. 1987. Sequence analysis, cellular localisation and expression of a neuroretina adhesion and cell survival molecule. *Cell* **51:** 135-142.

Berni, R., Stoppini, M., Zapponi, M.C., Meloni, M.L., Monaco, H.L. and Zanotti, G. 1990. The bovine plasma retinol-binding protein: amino acid sequence, interaction with transthyretin, crystallisation and preliminary X-ray data. *Eur J Biochem* **192:** 507-513.

Bhaskaran, K. and Sinha, V.B. 1971. Transmissible plasmid factors and fertility inhibition in *Vibrio cholerae*. *J Gen Microbiol* **69:** 89-97.

Bhattacharjee, J.W. and Srivastava, B.S. 1978. Mannose sensitive haemagglutinins in adherence of *Vibrio cholerae* El Tor to intestine. *J Gen Microbiol* **107:** 407-410.

Bilge, S.S., Apostol, J.M., Jr., Aldape, M.A. and Moseley, S.L. 1993. mRNA processing independent of RNaseIII and RnaseE in the expression of the F1845 fimbrial adhesin of *Escherichia coli*. *Proc Natl Acad Sci USA* **90:** 1455-1459.

Birnboim, H.C. and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7:** 1513-1522.

Biswas, S.K., Chowdhury, R. and Das, J. 1992. A 14 kiloDalton inner membrane protein of *Vibrio cholerae* biotype El Tor confers resistance to group IV cholera phage infection to classical vibrios. *J Bacteriol* **174:** 6221-6229.

Blasquez, V.C., Sperry, A.O., Cockerill, P.N, and Garrard, W.T. 1989. Protein:DNA interactions at chromosomal loop attachment sites. *Genome* **31:** 503-509.

Blenkinsopp, S.A., Khoury, A.E. and Costerton, J.W. 1992. Electrical enhancement of biocide efficacy against *Pseudomonas aeruginosa* biofilms. *Appl Environ Microbiol* **58:** 3770-3773.

Bloomfield, I.C., McClain, M.S. and Eisenstein, B.I. 1991a. Type I fimbriae mutants of *Escherichia coli* K-12: characterisation of recognised afimbriate strains and construction of new *fim* deletion mutants. *Mol Microbiol* **5**: 1439-1445.

Bloomfield, I.C., Vaughn, V., Fest, R.F. and Eisenstein, B.I. 1991b. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. *Mol Microbiol* **5**: 1447-1457.

Blyn, L.B., Braaten, B.A. and Low, D.A. 1990. Regulation of *pap* pilin phase variation by a mechanism involving differential Dam methylation states. *EMBO J* **9**: 4045-4054.

Bocklage, H., Heeger, K. and Müller-Hill, B. 1991. Cloning and characterisation of the *Mbo*II restriction-modification system. *Nucleic Acids Res* **19**: 1007-1013.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H and Falkow, S. 1977. Construction and characterisation of new cloning vehicles II: a multipurpose cloning system. *Gene* **2**: 95-113.

Bollinger, J., Park, C., Harayama, S. and Hazelbauer, G.L. 1984. Structure of the Trg protein: homologies with, and differences from, other sensory transducers of *Escherichia coli*. *Proc Natl Acad Sci USA* **81**: 3287-3291.

Booth, B.A., Boesman-Finkelstein, M. and Finkelstein, R.A. 1983. *Vibrio cholerae* soluble haemagglutinin/protease is a metalloenzyme. *Infect Immun* **42**: 639-644.

Booth, B.A., Boesman-Kinkelstein, M. and Finkelstein, R.A. 1984. *Vibrio cholerae* soluble haemagglutinin/protease nicks cholera-enterotoxin. *Infect Immun* **45**: 558-560.

Boyer, H.W. and Roulland-Dussoix, D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* **41**: 459-472.

Braaten, B.A., Blyn, L.B., Skinner, B.S. and Low, D.A. 1991. Evidence for a methylation-blocking factor (*mbf*) locus involved in *pap* pilus expression and phase variation in *Escherichia coli*. *J Bacteriol* **173**: 1789-1800.

Braaten, B.A., Platko, J.V., van der Woude, M.K., Simons, B.H., DeGraaf, F.K., Calvo, J.M. and Low, D.A. 1992. Leucine-responsive regulatory protein controls the expression of both the *pap* and *fan* pili operons in *Escherichia coli*. *Proc Natl Acad Sci USA* **89**: 4250-4254.

Brenner, V., Venetianer, P. and Kiss, A. 1990. Cloning and nucleotide sequence of the gene encoding the *Eca*I DNA methyl transferase. *Nucleic Acids Res* **18:** 355-359.

Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P. and Tiver, K.K. 1986. Molecular cloning of the cDNA for two major androgen-dependent secretory proteins of 18.5 kiloDaltons synthesised by the rat epididymis. *J Biol Chem* **261:** 4956-4961.

Brown, M.C.M., Weston, A., Saunders, J.R. and Humphreys, G.O. 1979. Transformation of *E. coli* C600 by plasmid DNA at different phases of growth. *FEMS Microbiol Lett* **5:** 219-222.

Brown, M.H. and Manning, P.A. 1985. Haemolysin genes of *Vibrio cholerae*: presence of homologous DNA in non-haemolytic and non-O1 strains. *FEMS Microbiol Lett* **30:** 197-201.

Brutlag, D.L., Dautricourt, J.-P., Diaz, R., Fier, J. and Stamm, R. BLAZE(TM): an implementation of the Smith-Waterman sequence comparison algorithm on a massively parallel computer. In preparation.

Butterton, J.R., Stoebner, J.A., Payne, S.M. and Calderwood, S.B. 1992. Cloning, sequencing and transcriptional regulation of *viuA*, the gene encoding the ferric vibriobactin receptor of *Vibrio cholerae*. *J Bacteriol* **174:**3729-3738

Carpenter, C.C.J. 1992. The treatment of cholera: clinical science at the bedside. *J Infect Dis* **166:** 2-14.

Cash, R.A., Music, S.I., Libonati, J.P., Craig, I.P., Pierce, N.F. and Hornick, R.B. 1974. Response of man to infection with *Vibrio cholerae* II: protection from illness afforded by previous disease and vaccine. *J Infect Dis* **130:** 325-333.

Cerretti, D.P., Dean, D., Davis, G.R., Bedwell, D.A. and Nomura, M. 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. *Nucleic Acids Res* **11:** 2599-2616.

Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. 1988a. Cloning and sequencing the *Hin*fI restriction and modification genes. *Gene* **70:** 387-392.

Chandrasegaran, S., Wu, L.P., Valda, E. and Smith, H.O. 1988b. Overproduction and purification of the M.*Hha*II methyltransferase from *Haemophilus haemolyticus*. *Gene* **74:** 15-21.

Chang, A.C.Y. and Cohen, S.N. 1978. Construction and characterisation of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. *J Bacteriol* **134**: 1141-1156.

Chao, W.L., Chen, R.S. and Tai, C.L. 1988. Factors affecting the survival of pathogenic bacteria in subtropical river water. *J Chinese Natl Microbiol Immunol* **21**: 85-92.

Chaudhuri, A. and De, S. 1977. Cholera and blood groups. *Lancet* **II**: 404.

Chen, E.Y. and Seeburg, P.H. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4**: 165-170.

Cheney, C.P. Schad, P.A. and Boedeker, E.C. 1981. Appearance with age of receptors on host intestinal epithelial cells for enteropathogenic *Escherichia coli* (EPEC). *Clin Res* **29**: 305A.

Chitnis, D.S., Sharma, K.D. and Koynat, R.S. 1982. Role of somatic antigen of *Vibrio cholerae* in adhesion to intestinal mucosa. *J Med Microbiol* **5**: 53-61.

Chongsa-nguan, M., Chaicumpa, W., Moolasart, P., Kandhasingha, P., Shimada, T., Kurazouo, H. and Takeda, Y. 1993. *Vibrio cholerae* O139 Bengal in Bangkok. *Lancet* **342**: 431-432.

Chou, P.Y. and Fasman, G.D. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol* **47**: 145-148.

Churchward, G., Belin, D. and Nagamine, Y. 1984. A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. *Gene* **31**: 165-171.

Clark, A.J., Chave-Cox, A., Ma, X. and Bishop, J.O. 1985. Analysis of mouse major urinary protein genes: variation between the exonic sequences of group I genes and a comparison with an active gene outwith (*sic*) group I both suggest that gene conversion has occurred between MUP genes. *EMBO J* **4**: 3167-3171.

Clark, C.A. 1988. *Localisation of the MFRHA in* Vibrio cholerae *and the nucleotide sequence of the neighbouring 16.3 kDa encoding gene.* Honours thesis. Department of Microbiology and Immunology, University of Adelaide. Adelaide, S.A. Australia.

Clemens, J.D., van Loon, F., Sack, D.A., Rao, M.R., Ahmed, F., Chakraborty, J., Kay, B.A., Khan, M.R., Yunus, M., Harris, J.R., Svennerholm, A.-M., and Holmgren,

J. 1991. Biotype as determinant of natural immunising effect of cholera. *Lancet* **337**: 883-884.

Colantuoni, V., Romano, V., Bensi, G., Santoro, C., Costanzo, F., Rangei, G. and Cortese, R. 1983. Cloning and sequencing of a full-length cDNA coding for human retinol-binding protein. *Nucleic Acids Res* **11**: 7769-7776.

Collis, C.M. and Hall, R.M. 1992. Site-specific deletion and rearrangement of integron insert genes catalysed by the integron DNA integrase. *J Bacteriol* **174**: 1574-1585.

Colman, S.D., Hu, P.-C. and Bott, K.F. 1990. Prevalence of novel repeat sequences in and around the P1 operon in the genome of *Mycoplasma pneumoniae*. *Gene* **87**: 91-96.

Colwell, R.R., Brayton, P.R., Grimes, D.J., Roszak, D.B., Huq, S.A. and Palmer, L.M. 1985. Viable but non culturable *Vibrio cholerae* and related pathogens in the environment: implications for release of genetically engineered microorganisms. *Biotechnology* **3**: 817-820.

Condemine, G. and Smith, C.L. 1990. Transcription regulates oxolinic acid-induced DNA gyrase cleavage at specific sites on the *E. coli* chromosome. *Nucleic Acids Res* **18**: 7389-7396.

Connaughton, J.F., Kaloss, W.D., Vanek, P.G., Nardone, G.A. and Chirikjian, J.G. 1990. The complete sequence of the *Bacillus amyloliquefaciens* proviral H2 *Bam*HI methylase gene. *Nucleic Acids Res* **18**: 4002.

Contreras, A. and Maxwell, A. 1992. *gyrB* mutations which confer coumarin resistance also affect DNA supercoiling and ATP hydrolysis by *Escherichia coli* DNA gyrase. *Mol Microbiol* **6**: 1617-1624.

Correia, F.F., Inouye, S. and Inouye, M. 1986. A 26 bp repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. *J Bacteriol* **167**: 1009-1015.

Crosa, J.H. 1989. Genetics and molecular biology of siderophore-related iron transport in bacteria. *Microbiol Rev* **53**: 517-530.

Crowther, R.S., Roomi, N.W., Fahim, R.E.F. and Forstner, J.F. 1987. *Vibrio cholerae* metalloprotease degrades intestinal mucin and facilitates enterotoxin-induced secretion from rat intestine. *Biochim Biophys Acta* **924**: 393-402.

Darfeuille-Michaud, A., Aubel, D., Chauviere, G., Rich, C., Bourges, M., Servin, A. and Joly, B. 1990. Adhesion of enterotoxigenic *Escherichia coli* to the human colon carcinoma cell line Caco-2 in culture. *Infect Immun* **58**: 893-902.

Daskaleros, P.A. and Payne, S.M. 1987. Congo red binding phenotype is associated with haemin binding and increased infectivity of *Shigella flexneri* in the HeLa cell model. *Infect Immun* **55**: 1393-1398.

De, S.N. 1959. Enterotoxigenicity of bacteria-free culture-filtrate of *Vibrio cholerae*. *Nature* **183**: 1533-1534.

de Boer, H.A. 1984. A versatile plasmid system for the study of prokaryotic transcription signals in *Escherichia coli*. *Gene* **30**: 251-255.

de la Campa, A., Kale, P., Springhorn, S.S. and Lacks, S.A. 1987. Proteins encoded by the *Dpn*II restriction gene cassette: two methylases and an endonuclease. *J Mol Biol* **196**: 457-469.

de Lorenzo, V., Wee, S., Herrero, M. and Neilands, J.B. 1987. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. *J Bacteriol* **169**: 2624-2630.

De Moor, C.E. 1963. A non-haemolytic vibrio. *Trop Geogr Med* **15**: 97-107.

Derr, L.K. and Strathern, J.N. 1993. A role for reverse transcripts in gene conversion. *Nature* **361**: 170-173.

Difco Manual. 1984. 10th edn. Difco Laboratories Inc. Detroit, Mi. USA.

Dimri, G.P., Rudd, K.E., Morgan, M.K., Bayout, H. and Ames, G.F.-L. 1992. Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. *J Bacteriol* **174**: 4583-4593.

DiRita, V. 1992. Co-ordinate expression of virulence genes by ToxR in *Vibrio cholerae*. *Mol Microbiol* **6**: 451-458.

DiRita, V.J. and Mekalanos, J.J. 1991. Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxD, results in signal transduction and transcriptional activation. *Cell* **64**: 29-37.

DiRita, V.J., Parsot, C. , Jander, G. and Mekalanos, J.J. 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. *Proc Natl Acad Sci USA* **88**: 5403-5407.

Dorman, C.J. 1991. DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. *Infect Immun* **59**: 745-749.

Drasar, B.S. and Barrow, P.A. 1985. *Intestinal Microbiology*. Van Nostrand Reinhold (UK). Wolkingham, Berks. UK.

Drayna, D., Fielding, C., McLean, J., Baer, B., Castro, G., Chen, E., Comstock, L., Henzel, W., Kohr, W., Rhee, L., Wion, K. and Lawn, R. 1986. Cloning and expression of human apolipoproteinD cDNA. *J Biol Chem* **261**: 16535-16539.

Drayna, D.T., McLean, J.W., Wion, K.L., Trent, J.M., Drabkin, H.A. and Lawn, R.M. 1987. Human apolipoproteinD gene: gene sequence, chromosomal localisation and homology to the α_{3_u}-globulin superfamily. *DNA* **6**: 199-204.

Dutta, N.K., Panse, M.W. and Kulkarni, D.R. 1959. Role of cholera toxin in experimental cholera. *J Bacteriol* **78**: 594-595.

Ehara, M., Iwami, M., Ichinose, Y., Shimotori, S., Kangethe, S.K. and Nakamura, S. 1991. Purification and characterisation of fimbriae from fimbriate *Vibrio cholerae* O1 strain Bgd 17. *Trop Med* **33**: 109-125.

Eisenach, K.D., Cave, M.D., Bates, J.H. and Crawford, J.T. 1990. Polymerase chain reaction amplification of a repetitive DNA sequence specific for *Mycobacterium tuberculosis*. *J Infect Dis* **161**: 977-981.

Elliott, H.L., Carpenter, C.C.J., Sack, R.B. and Yardley, J.H. 1970. Small bowel morphology in experimental canine cholera. *Lab Invest* **22**: 112-120.

Etzler, M.E. and Branstrator, M.L. 1979. Cell surface components of intestinal epithelial cells and their relationship to cellular differentiation. *Ciba Found Symp* **70 (New Series)**: 51-68.

Ewald, P.W. 1993. The evolution of virulence. *Sci Amer* **268(4)**: 56-62.

Faast, R., Ogierman, M.A., Stroeher, U.H and Manning, P.A. 1989. Nucleotide sequence of the structural gene, *tcpA*, for a major pilin subunit of *Vibrio cholerae*. *Gene* **85**: 227-231.

Fasano, A., Baudry, B., Pumplin, D.W., Wasserman, S.S., Tall, B.D., Ketley, J.M. and Kaper, J.B. 1991. *Vibrio cholerae* produces a second enterotoxin, which affects the intestinal tight junctions. *Proc Natl Acad Sci USA* **88:** 5242-5246.

Feinberg, A.P. and Vogelstein, B. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132:** 6-13.

Ferencz, A., Ørskov, I., Ørskov, F. and Klemm, P. 1980. Deoxyribonuclear acid is a significant component of the small intestine mucous. *Acta Path Microbiol Scand* **80:** 347-348.

Fetherstone, J.D., Schuetze, P. and Perry, R.D. 1992. Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. *Mol Microbiol* **6:** 2693-2704.

Filkova, S.L., Ilina, T.S., Gintsburg, A.L., Iamishevskii, N.V. and Smirnov, G.B. 1988. The role of the RS1 sequence of *Vibrio cholerae* in the amplification of a segment of plasmid DNA carrying the tetracycline resistance gene and cholera toxin genes. *Genetika* **24:** 821-828.

Filkova, S.L., Ilina, T.S. and Smirnov, G.B. 1989. RecA-independence of the process of amplification caused by the RS1 sequence of *Vibrio cholerae*. *Mol Gen Mikrobiol Virusol* **1989:** 3-6.

Finkelstein, R.A. 1973. Cholera. *CRC Crit Rev Microbiol* **2:** 553-623.

Finkelstein, R.A. 1988. Cholera-the great teacher: 100 years since Koch's postulates. In Vibrio cholerae *and cholera*. Takeda, Y. and Kuwahara, S. (eds). KTK Scientific Publishers. Tokyo. Japan. pp 57-77.

Finkelstein, R.A., Arita, M., Clements, J.D. and Nelson, E.T. 1978. Isolation and purification of an adhesive factor ("Cholera lectin") from *Vibrio cholerae*. In *Proceedings of the 13th joint conference on cholera*. US-Japan Cooperative Medical Science Programmes. National Institutes of Health. Bethesda, Md. USA. pp 137-151.

Finkelstein, R.A., Boesman-Finkelstein, M., Chang, Y, and Häse, C.C. 1992. *Vibrio cholerae* haemagglutinin/protease, colonial variation, virulence and detachment. *Infect Immun* **60:** 472-478.

Finkelstein, R.A., Boesman-Finkelstein, M. and Holt, P. 1983. *Vibrio cholerae* haemagglutinin/lectin/protease hydrolyses fibronectin and ovomucin: F.M. Burnet revisited. *Proc Natl Acad Sci USA* **80:** 1092-1095.

Finkelstein, R.A. and Hanne, L.F. 1982. Purification and characterisation of the soluble haemagglutinin (cholera lectin) produced by *Vibrio cholerae. Infect Immun* **36:** 1199-1208.

Finkelstein, R.A. and LoSpalluto, J.J. 1969. Pathogenesis of experimental cholera: a preparation and isolation of choleragen and choleragenoid. *J Exp Med* **130:** 185-202.

Finkelstein, R.A. and Mukerjee, S. 1963. Haemagglutination: a rapid method for differentiating *Vibrio cholerae* and El Tor vibrios. *Proc Soc Exp Biol Med* **112:** 355-359.

Finn, T.M., Reisner, J., Germanier, R. and Cryz, S.J. 1987. Cell-associated haemagglutinin-deficient mutant of *Vibrio cholerae. Infect Immun* **55:** 942-946.

Focareta, T. 1989. *The extracellular DNase(s) of* Vibrio cholerae. PhD thesis. University of Adelaide. Adelaide, S.A. Australia.

Focareta, T. and Manning, P.A. 1987. Extracellular proteins of *Vibrio cholerae*: molecular cloning, nucleotide sequence and characterisation of the deoxyribonuclease (DNase) together with its periplasmic localisation in *Escherichia coli* K-12. *Gene* **53:** 31-40.

Focareta, T, and Manning, P.A. 1991. Distinguishing between the extracellular DNases of *Vibrio cholerae* and development of a transformation system. *Mol Microbiol* **5:** 2547-2555.

Foster, J.W. 1993. The acid tolerance response of *Salmonella typhimurium* involves transient synthesis of key acid shock proteins. *J Bacteriol* **175:** 1981-1987.

Foster, J.W. and Hall, H.K. 1990. Adaptive acidification tolerance response of *Salmonella typhimurium. J. Bacteriol* **172:** 771-778.

Foster, J.W. and Hall, H.K. 1991. Inducible pH homeostasis and the acid tolerance response of *Salmonella typhimurium. J Bacteriol* **173:** 5129-5135.

Foster, J.W. and Hall, H.K. 1992. Effect of *Salmonella typhimurium* ferric uptake regulator (*fur*) mutations on iron- and pH-regulated protein synthesis. *J Bacteriol* **174:** 4317-4323.

Franzon, V.L. 1988. *Haemagglutinins of* Vibrio cholerae: *molecular characterisation of the mannose-fucose-resistant haemagglutinin (MFRHA)*. PhD thesis. University of Adelaide. Adelaide, S.A. Australia.

Franzon, V.L. and Manning, P.A. 1986. Molecular cloning and expression in *Escherichia coli* K-12 of the gene for a haemagglutinin from *Vibrio cholerae*. *Infect Immun* **52:** 279-284.

Freter, R. 1980. Association of enterotoxigenic bacteria with the mucosa of the small intestine: mechanisms and pathogenic implications. *Nobel Symp* **43:** 155-170.

Freter, R., Allweiss, B., O'Brien, P.C.M., Halstead, S.A. and Macsai, M.S. 1981a. Role of chemotaxis in the association of motile bacteria with intestinal mucosa: *in vitro* studies. *Infect Immun* **34:** 241-249.

Freter, R. and Jones, G.W. 1976. Adhesive properties of *Vibrio cholerae*: nature of the interaction with intact mucosal surfaces. *Infect Immun* **14:** 246-256.

Freter, R., O'Brien, P.C.M. and Macsai, M.S. 1981b. Role of chemotaxis in the association of motile bacteria with intestinal mucosa: *in vivo* studies. *Infect Immun* **34:** 234-240.

Fujitani, S., Komano, T. and Inouye, S. 1991. A unique repetitive DNA sequence in the *Myxococcus xanthus* genome. *J Bacteriol* **173:** 2125-2127.

Galen, J.E., Ketley, J.M., Fasano, A., Richardson, S.H., Wasserman, S.S. and Kaper, J.B. 1992. Role of the *Vibrio cholerae* neuraminidase in the function of cholera toxin. *Infect Immun* **60:** 406-415.

Gangarosa, E.J., Sanati, A., Saghari, H. and Feeley, J.C. 1967. Multiple serotypes of *Vibrio cholerae* isolated from a case of cholera. *Lancet* **I:** 646-648.

Garger, S.J., Griffith, O.M. and Gill, 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.

Gellert, M., Mizuuchi, K., O'Dea, M.H., Itoh, T. and Tomizawa, J.-I. 1977. Nalidixic acid resistance: a second genetic character involved in DNA gyrase activity. *Proc Natl Acad Sci USA* **74:** 4772-4776.

Gilson, E., Bachellier, S., Perrin, S., Perrin, D., Grimont, P.A.D., Grimont, F. and Hofnung, M. 1990a. Palindromic unit highly repetitive DNA sequences exhibit species specificity within Enterobacteriaceae. *Res Microbiol* **141:** 1103-1116.

Gilson, E., Clément, J.M., Brutlag, D. and Hofnung, M. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. *EMBO J* **3:** 1417-1421.

Gilson, E., Clément, J.-M., Perrin, D. and Hofnung, M. 1987a. Palindromic units: a case of highly repetitive DNA sequences in bacteria. *Trends Genet* **3:** 226-230.

Gilson, E., Perrin, D., Clément, J.-M., Szmelcman, S., Dassa, E. and Hofnung, M. 1986a. Palindromic units from *E. coli* as binding sites for a chromoid-associated protein. *FEBS Lett* **206:** 323-328.

Gilson, E., Perrin, D. and Hofnung, M. 1990b. DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organisation. *Nucleic Acids Res* **18:** 3941-3952.

Gilson, E., Perrin, D. Saurin, W. and Hofnung, M. 1987b. Species specificity of bacterial palindromic units. *J Mol Evol* **25:** 371-373.

Gilson, E., Rousset, J.P., Clément, J.M. and Hofnung, M. 1986b. A subfamily of *E. coli* palindromic units implicated in transcription termination? *Ann Inst Pasteur/Microbiol* **137B:** 259-270.

Gilson, E., Saurin, W., Perrin, D., Bachelier, S. and Hofnung, M. 1991a. The BIME family of bacterial highly repetitive sequences. *Res Microbiol* **142:** 217-222.

Gilson, E., Saurin, W., Perrin, D., Bachelier, S., and Hofnung, M. 1991b. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). *Nucleic Acids Res* **19:** 1375-1383.

Glass, R.I., Holmgren, J., Haley, C.E., Khan, M.R., Svennerholm, A.-M., Sfoll, B.J., Belajet, Hossain, K.M., Black, R.E., Yunus, M. and Barua, D. 1985. Predisposition for cholera of individuals with O blood group. *Am J Epidemiol* **121:** 791-796.

Godovac-Zimmermann, J. 1988. The structural motif of β-lactoglobulin and retinol binding protein: a basic framework for binding and transport of small hydrophobic molecules? *Trends Biochem Sci* **13:** 64-66.

Godovac-Zimmermann, J., Conti, A., James, L. and Napolitano, L. 1988. Microanalysis of the amino-acid sequence of monomeric β-lactoglobulin I from donkey (*Equus asinus*) milk: the primary structure and its homology with a superfamily of hydrophobic molecule transporters. *Biol Chem Hoppe Seyler* **369**: 171-179.

Goldberg, I. and Mekalanos, J.J. 1986. Effect of a *recA* mutation on cholera toxin gene amplification and deletion events. *J Bacteriol* **165**: 723-731.

Goldberg, M.B., Boyko, S.A., Butterton, J.R., Stoebner, J.A., Payne, S.M. and Calderwood, S.B. 1992. Characterisation of a *Vibrio cholerae* virulence factor homologous to the family of TonB-dependent proteins. *Mol Microbiol* **6**: 2407-2418.

Goldberg, M.B., Boyko, S.A. and Calderwood, S.B. 1990a. Transcriptional regulation by iron of a *Vibrio cholerae* virulence gene and homology of the gene to the *Escherichia coli* Fur system. *J Bacteriol* **172**: 6863-6870.

Goldberg, M.B., Boyko, S.A. and Calderwood, S.B. 1991. Positive transcriptional regulation of an iron-regulated virulence gene in *Vibrio cholerae*. *Proc Natl Acad Sci USA* **88**: 1125-1129.

Goldberg, M.B., DiRita, V.J. and Calderwood, S.B. 1990b. Identification of an iron-regulated virulence determinant in *Vibrio cholerae*, using Tn*phoA* mutagenesis. *Infect Immun* **58**: 55-60.

Goldberg, S.L. and Murphy, J.R. 1984. Molecular cloning of the haemolysin determinant from *Vibrio cholerae* El Tor. *J Bacteriol* **160**: 239-244.

Goldberg, S.A. and Murphy, J.R. 1985. Cloning and characterisation of the haemolysin determinants from *Vibrio cholerae* RV79 (Hly$^+$), RV79 (Hly$^-$) and 569B. *J Bacteriol* **162**: 35-41.

Goodson, M. and Rowbury, R.T. 1989. Habituation to normal lethal acidity by prior growth of *Escherichia coli* at a sublethal acid pH value. *Lett Appl Microbiol* **8**: 77-79.

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.

Greenough, W.B., III, and Molla, A.M. 1988. Oral rehydration therapy (ORT): present and future. In Vibrio cholerae *and cholera*. Takeda, Y. and Kuwahara, S. (eds). KTK Scientific Publishers. Tokyo. Japan. pp117-127.

Grimes, D.J., Atwell, R.W., Brayton, P.R., Rollins, D.A., Roszak, D.B., Singleton, F.L., Tamplin, M.L. and Colwell, R.R. 1986. The fate of enteric pathogenic bacteria in estuarine and marine environments. *Microbiol Sci* **3**: 324-329.

Grunstein, M. and Hogness, D. 1975. Colony hybridisation: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci USA* **72**: 3961-3965.

Guinée, P.A.M., Jansen, W.H. and Peters, P.W. 1985. *Vibrio cholerae* infection and immunity in an adult rabbit model. *Zbl Bakt Hyg* A **259**: 118-131.

Guschlbauer, W. 1988. The DNA and *S*-adenosylmethionine-binding regions of *Eco*Dam and related methyltransferases. *Gene* **74**: 211-214.

Gustafsson, B. and Holme, T. 1985. Immunological characterisation of *Vibrio cholerae* O1 lipopolysaccharide, O-sidechain, and the core with monoclonal antibodies. *Infect Immun* **49**: 275-280.

Haas, R. and Meyer, T.F. 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. *Cell* **44**: 107-115.

Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R. and Goebel, W. 1990. Deletions of chromosomal regions coding for fimbriae and haemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. *Microb Pathog* **8**: 213-225.

Haefliger, J.A., Jenne, D., Stanley, K.K. and Tschopp, J. 1987. Structural homology of human complement component C8$_7$ and plasma protein HC: identity of the cysteine bond pattern. *Biochem Biophys Res Commun* **149**: 750-754.

Hall, R.H. and Drasar, B.S. 1990. *Vibrio cholerae* HlyA haemolysin is processed by proteolysis. *Infect Immun* **58**: 3375-3379.

Hall, R.H., Khambaty, F.M., Kothary, M. and Keasler, S.P. 1993. Non-O1 *Vibrio cholerae*. *Lancet* **342**: 430.

Hall, R.H., Vial, P.A., Kaper, J.B., Mekalanos, J.J. and Levine, M.M. 1988. Morphological studies on fimbriae expressed by *Vibrio cholerae* O1. *Microb Pathog* **4**: 257-265.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557-580.

Hanne, L.F. and Finkelstein, R.A. 1982. Characterisation and distribution of the haemagglutinins produced by *Vibrio cholerae*. *Infect Immun* **36:** 209-214.

Hanson, M.S. and Hansen, E.J. 1991. Molecular cloning, partial purification and characterisation of a haem-binding lipoprotein from *Haemophilus influenzae* type B. *Mol Microbiol* **5:** 267-278.

Häse, C.C. and Finkelstein, R.A. 1990. Comparison of the *Vibrio cholerae* haemagglutinin/protease and the *Pseudomonas aeruginosa* elastase. *Infect Immun* **58:** 4011-4015.

Häse, C.C. and Finkelstein, R.A. 1991. Cloning and nucleotide sequence of the *Vibrio cholerae* haemagglutinin/protease (HA/protease) gene and construction of an HA/protease-negative strain. *J Bacteriol* **173:** 3311-3317.

Hawkes, R., Niday, E. and Gordon, J. 1982. A dot-immuno binding assay for monoclonal and other antibodies. *Anal Biochem* **119:** 142-147.

Henderson, D.P. and Payne, S.M. 1993. Cloning and characterisation of the *Vibrio cholerae* genes encoding the utilisation of iron from haemin and haemoglobin. *Mol Microbiol* **7:** 461-469.

Henikoff, S. 1984. Unidirectional digestion with Exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28:** 351-359.

Herrington, D.A., Hall, R.H., Losonsky, G., Mekalanos, J.J., Taylor, R.K. and Levine, M.M. 1988. Toxin, toxin coregulated pili and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J Exp Med* **168:** 1487-1492.

Higgins, C.F., Ames, G.F.-L., Barnes, W.M., Clément, J.M. and Hofnung, M. 1982. A novel intercistronic regulatory element of prokaryotic operons. *Nature* **298:** 760-762.

Higgins, C.F., McLaren, R.S. and Newbury, S.F. 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? - a review. *Gene* **72:** 3-14.

Higgins, D.E., Nazareno, E. and DiRita, V.J. 1992. The virulence gene activator ToxT from *Vibrio cholerae* is a member of the AraC family of transcriptional activators. *J Bacteriol* **174:** 6974-6980.

Higgins, D.G. and Sharp, P.M. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Cabios* **5**: 151-153.

Hirst, T.R. and Holmgren, J. 1987. Transient entry of enterotoxin subunits into the periplasm during their secretion from *Vibrio cholerae*. *J Bacteriol* **169**: 1037-1045.

Hitchcock, P.J. and Brown, T.M. 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver stained polyacrylamide gels. *J Bacteriol* **154**: 269-277.

Ho, A.S.Y., Sohel, I. and Schoolnik, G.K. 1992. Cloning and characterisation of *fxp*, the flexible pilin gene of *Aeromonas hydrophila*. *Mol Microbiol* **6**: 2725-2732.

Hoepelman, A.I.M and Tuomanen, E.I. 1992. Consequences of microbial attachment directing host cell functions with adhesins. *Infect Immun* **60**: 1729-1733.

Hoffman, C.S. and Wright, A. 1985. Fusions of secreted proteins to alkaline phosphatase: an approach to studying protein secretion. *Proc Natl Acad Sci USA* **82**: 5107-5111.

Holden, H.M., Rypniewski, W.R., Law, J.H and Rayment, I. 1987. The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* at 2.6 Å resolution. *EMBO J* **6**: 1565-1570.

Holmes, R.K., Vasil, M.L and Finkelstein, R.A. 1975. Studies on toxigenesis in *Vibrio cholerae* III: characterisation of non-toxigenic mutants *in vitro* and in experimental animals. *J Clin Invest* **55**: 551-560.

Holmgren, J. 1981. Actions of cholera toxin and the prevention and treatment of cholera. *Nature* **292**: 413-417.

Holmgren, J. 1991. Mucosal immunity and vaccination. *FEMS Microbiol Immunol* **89**: 1-10.

Holmgren, J., Svennerholm, A.-M., Lönnroth, I., Fall-Persson, M., Markman, B. and Lundbeck, H. 1977. Development of improved cholera vaccine based on subunit toxoid. *Nature* **269**: 602-604.

Honda, T. and Finkelstein, R.A. 1979. Purification and characterisation of a haemolysin produced by *Vibrio cholerae* biotype El Tor: another toxic substance produced by cholera vibrios. *Infect Immun* **26**: 1020-1027.

Hornick, R.B., Music, S.I., Wenzel, R., Cash, R., Libonati, J.P., Snyder, M.J. and Woodward, T.E. 1971. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. *Bull NY Acad Med* **47**: 1181-1191.

Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. and Kayser, H. 1987. Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J Mol Biol* **198**: 499-513.

Hultgren, S.J., Abraham, S., Caparon, M., Falk, P., St. Geme, J.W., III, and Normark, S. 1993. Pilus and non-pilus bacterial adhesins: assembly and function in cell recognition. *Cell* **73**: 887-901.

Hulton, C.S.J., Higgins, C.F. and Sharp, P.M. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium*, and other enterobacteria. *Mol Microbiol* **5**: 825-834.

Huq, A., Small, E.B., West, P.A., Huq, M.I., Rahman, R. and Colwell, R.R. 1983. Ecological relationships between *Vibrio cholerae* and planktonic copepods. *Appl Environ Microbiol* **45**: 275-283.

Huq, A., West, P.A., Small, E.B., Huq, M.I., and Colwell, R.R. 1984. Influence of water temperature, salinity and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Appl Environ Microbiol* **48**: 420-424.

Imbesi, F. and Manning, P.A. 1982. Biotype-specific restriction and modification of DNA in *Vibrio cholerae*. *J Clin Microbiol* **16**: 552-554.

Inouye, M. and Inouye, S. 1991. msDNA and bacterial reverse transcriptase. *Ann Rev Microbiol* **45**: 163-186.

Inukai, M., Takeuchi, M., Schimizu, K. and Arai, M. 1978. Mechanism of action of globomycin. *J Antibiotics* **31**: 1203-1205.

Islam, M.S. 1990. Increased toxin production by *Vibrio cholerae* O1 during survival with a green alga, *Rhizoclonium fontanum*, in an artificial aquatic environment. *Microbiol Immunol* **34**: 557-563.

Islam, M.S., Drasar, B.S. and Bradley, D.J. 1989. Attachment of toxigenic *Vibrio cholerae* O1 to various freshwater plants and survival with a filamentous green alga *Rhizoclonium fontanum*. *J Trop Med Hyg* **92**: 396-401.

Islam, M.S., Drasar, B.S. and Bradley, D.J. 1990a. Survival of toxigenic *Vibrio cholerae* O1 with a common duckweed, *Lemna minor*, in artificial aquatic ecosystems. *Trans R Soc Trop Med Hyg* **84**: 422-424.

Islam, M.S., Drasar, B.S. and Bradley, D.J. 1990b. Long-term persistence of toxigenic *Vibrio cholerae* O1 in the mucilaginous sheath of a blue-green alga, *Anabaena variabilis*. *J Trop Med Hyg* **93**: 133-139.

Iwanaga, M., Nakasone, N and Ehara, M. 1989. Pili of *Vibrio cholerae* O1 biotype El Tor: a comparitive study on adhesive and non-adhesive strains. *Microbiol Immunol* **33**: 1-9.

Iwanaga, M., Yamamoto, K., Higa, N., Ichinose, Y., Nakasone, N. and Tanabe, M. 1986. Culture conditions for stimulating cholera toxin production by *Vibrio cholerae* O1 El Tor. *Microbiol Immunol* **30**: 1075-1083.

Jaquet, M.A. and Reiss, C. 1992. *In vivo* control of promoter and terminator efficiencies at a distance. *Mol Microbiol* **6**: 1681-1691.

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.

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.

Jonson, G. 1991. *Expression of virulence factors by classical and El Tor* Vibrio cholerae. PhD thesis. University of Göteborg. Göteborg. Sweden.

Jonson, G., Holmgren, J. and Svennerholm, A.-M. 1991. Identification of a mannose-binding pilus on *Vibrio cholerae* El Tor. *Microb Pathog* **11**: 433-441.

Jonson, G., Holmgren, J. and Svennerholm, A.-M. 1992. Analysis of expression of toxin-coregulated pili in classical and El Tor *Vibrio cholerae* O1 *in vitro* and *in vivo*. *Infect Immun* **60**: 4278-4284.

Jonson, G., Sanchez, J. and Svennerholm, A.-M. 1989a. Expression and detection of different biotype-associated cell-bound haemagglutinins of *Vibrio cholerae* O1. *J Gen Microbiol* **135**: 111-120.

Jonson, G., Svennerholm, A.-M. and Holmgren, J. 1989b. *Vibrio cholerae* express cell surface antigens during intestinal infection which are not expressed during *in vitro* culture. *Infect Immun* **57**: 1809-1815.

Jonson, G., Svennerholm, A.-M. and Holmgren, J. 1990. Expression of virulence factors by classical and El Tor *Vibrio cholerae in vivo* and *in vitro*. *FEMS Microbiol Ecol* **74**: 221-228.

Kabir, S., Ahmad, N. and Ali, S. 1984. Neuraminidase production by *Vibrio cholerae* O1 and other diarrhoeagenic bacteria. *Infect Immun* **44**: 747-749.

Kabir, S. and Ali, S. 1983. Characterisation of surface properties of *Vibrio cholerae*. *Infect Immun* **39**: 1048-1058.

Kaper, J.B. and Baldini, M.M. 1992. Genetics. In *Cholera*. Barua, D. and Greenough, W.B., III. (eds). Plenum Medical Book Company, New York, NY. USA. pp 69-94.

Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumport, R.I. 1989. Purification, cloning and sequence analysis of *Rsr*I DNA methyl transferase: lack of homology between two enzymes, *Rsr*I and *Eco*RI that methylate the same nucleotide in identical recognition sequences. *Nucleic Acids Res* **17**: 10403-10425.

Kaufmann, M.R., Seyer, J.M. and Taylor, R.K. 1991. Processing of TCP pilin by TcpJ typifies a common step intrinsic to a newly recognised pathway of extracellular protein secretion by Gram-negative bacteria. *Genes Develop* **5**: 1834-1846.

Kaufmann, M.R., Shaw, C.E., Jones, I.D. and Taylor, R.K. 1993. Biogenesis and regulation of the *Vibrio cholerae* toxin-coregulated pilus: analogies to other virulence factor secretory systems. *Gene* **126**: 43-49.

Kay, W.W., Phipps, B.M., Ishiguro, E.E. and Trust, T.J. 1985. Porphyrin binding by the surface array virulence protein of *Aeromonas salmonicida*. *J Bacteriol* **164**: 1332-1336.

Keen, J.N., Caceres, I., Eliopoulos, E.E., Zagalsky, P.F. and Findlay, J.B.C. 1991. Complete sequence and model for the A_2 subunit of the carotenoid pigment complex, crustacyanin. *Eur J Biochem* **197**: 407-417.

Kenne, L., Lindberg, B., Unger, P. Gustafsson, B. and Holme, T. 1982. Structural studies of the *Vibrio cholerae* O-antigen. *Carbohydr Res* **100**: 341-349.

Kenne, L., Lindberg, B., Unger, P., Holme, T. and Holmgren, J. 1979. Structure studies of the *Vibrio cholerae* O-antigen. *Carbohydr Res* **68**: C14-C16.

Kitchen, I. *Textbook of* in vitro *practical Pharmacology.* Blackwell Scientific Publications. Oxford. UK.

Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Janulaitis, A. 1989. Sequence motifs characteristic of DNA [cytosine-N4] methyl transferases: similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res* **17**: 9823-9832.

Koch, R. 1883. Der seitens des Gch. Reg.-Rath Dr. R. Koch an den Staatssekretär des innem, Herrn Staatsminister v. Betticher Excellenz erstattete Bericht. *Dtsche Med Wschr* **9**: 615-617.

Koch, R. 1884a. Fünfter Bericht des Leiters der deutschen wissenschaftlichen Commission zur Erforschung der Cholera. *Dtsche Med Wschr* **10**: 111-112.

Koch, R. 1884b. Sechster Bericht des leiters der deutschen wissenschaftlichen Commission zur Erforschung der Cholera. *Dtsche Med Wschr* **10**: 191-192.

Koch, R. 1886. Further researches on cholera. *B Med J* **I**: 62-66.

Koomey, M., Gotschlich, E.C., Robbins, K., Bergström, S. and Swanson, J. 1987. Effects of *recA* mutations on pilus antigenic variation and phase transitions in *Neisseria gonorrhoeae.* *Genetics* **117**: 391-398.

Kothary, M.A. and Kreger, A.S. 1985. Production and partial characterisation of an elastolytic protease of *Vibrio vulnificus.* *Infect Immun* **50**: 534-540.

Kreutzer, K., Pratt, C. and Torriani, A. 1975. Genetic analysis of regulatory mutants of alkaline phosphatase of *Escherichia coli.* *Genetics* **81**: 459-468.

Kugimiya, W., Otani, Y., Hashimoto, Y. and Tagaki, Y. 1986. Molecular cloning and nucleotide sequence of the lipase gene from *Pseudomonas fragi.* *Biochem Biophys Res Commun* **141**: 185-190.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**: 105-132.

Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. and Greenberg, W. 1986. Genetic basis of the complimentary *DpnI* and *DpnII* restriction systems of *S. pneumoniae*: an intercellular cassette mechanism. *Cell* **46**: 993-1000.

Lawrence, J.B. 1990. A fluorescence *in situ* hybridisation approach for gene mapping and the study of nuclear organisation. In *Genome analysis VolI: genetic and physical mapping*. Davies, K.E. and Tilghman, S.M. (eds). Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY. USA.

Leblond, P., Demuyter, P., Simonet, J.M. and Decaris, B. 1990. Genetic instability and hypervariability in *Streptomyces ambofaciens*: towards an understanding of a mechanism of genome plasticity. *Mol Microbiol* **4**: 707-714.

Lee, B.C. and Schryvers, A.B. 1988. Specificity of the lactoferrin and transferrin receptors in *Neisseria gonorrhoeae*. *Mol Microbiol* **2**: 827-829.

Lencer, W.I., Reinhart, F.D. and Neutra, M.R. 1990. Interaction of cholera toxin with cloned human goblet cells in monolayer culture. *Am J Physiol* **258**: 96-102.

Lennon, E., Gutman, P.D., Yao, H. and Minton, K.W. 1991. A highly conserved repeated chromosomal sequence in the radio resistant bacterium *Deinococcus radiophilus* SARK. *J Bacteriol* **173**: 2137-2140.

Levine, M.M., Kaper, J.B., Black, R.E. and Clements, M.L. 1983. New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development. *Microbiol Rev* **47**: 510-550.

Levine, M.M., Kaper, J.B., Herrington, D., Losonsky, G., Morris, J.G., Clements, M.L., Black, R.E., Tall, B. and Hall, R. 1988. Volunteer studies of deletion mutants of *Vibrio cholerae* O1 prepared by recombinant techniques. *Infect Immun* **56**: 161-167.

Levine, M.M. and Pierce, N.F. 1992. Immunity and vaccine development. In *Cholera*. Barua, D. and Greenough, W.B., III. (eds). Plenum Medical Book Company. New York, NY. USA. pp 285-327.

Lin, E.C.C., Lerner, S.A. and Jorgensen, S.E. 1962. A method for isolating constitutive mutants for carbohydrate-catabolising enzymes. *Biochem Biophys Acta* **60**: 422-424.

Lin, R.T., D'Ari, R. and Newmann, E.B. 1990. The leucine regulon of *Escherichia coli* K-12: a mutation in *rblA* alters expression of L-leucine-dependent metabolic operons. *J Bacteriol* **172**: 4529-4535.

Lin, Z., Kumagai, K., Baba, K., Mekalanos, J.J. and Nishibuchi, M. 1993. *Vibrio parahaemolyticus* has a homologue of the *Vibrio cholerae toxRS* operon that mediates environmentally induced regulation of the thermostable direct haemolysin gene. *J Bacteriol* **175**: 3844-3855.

Litwin, C.M., Boyko, S.A. and Calderwood, S.B. 1992. Cloning, sequencing and transcriptional regulation of the *Vibrio cholerae fur* gene. *J Bacteriol* **174**: 1897-1903.

Lobley, R.W. 1991. The enterocyte and its brush border. In *Molecular pathogenesis of gastrointestinal infections*. Wädstrom, T., Mäkelä, P.H., Svennerholm, A.-M. and Wolf-Watz, H. (eds). Plenum Press. New York, NY. USA. pp 1-8.

Lohia, A., Chatterjee, A.N. and Das, J. 1984. Lysis of *Vibrio cholerae* cells: direct isolation of the outer membrane from whole cells by treatment with urea. *J Gen Microbiol* **130**: 2027-2033.

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.

Lund, B., Marklund, B.-I., Strömberg, N., Lindberg, F., Karlsson, K.-A. and Normark, S. 1988. Uropathogenic *Escherichia coli* can express serologically identical pili of different receptor binding specificities. *Mol Microbiol* **2**: 255-263.

Lupski, J.R. and Weinstock, G.M. 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J Bacteriol* **174**: 4525-4529.

McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. 1981. A system to study promoter and terminator signals recognised by *Escherichia coli* RNA polymerase. In *Gene amplification and analysis VolII: analysis of nucleic acids by enzymatic methods*. Chirikjian, I.C. and Papas, T.S. (eds). pp384-415.

McNeill, W.H. 1976. *Plagues and People*. Anchor Press/Doubleday. Garden City, NY. USA.

Mahan, M.J., Slauch, J.M. and Mekalanos, J.J. 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**: 686-688.

Maniatis, T., Fritsch, E.F. and Sambrook, J. 1982. *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratories. Cold Spring Harbour, NY. USA.

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. 1991. Molecular analysis of potential adhesions of *Vibrio cholerae* O1. In *Molecular pathogenesis of gastrointestinal infections.* Wädstrom, T., Mäkelä, P.H., Svennerholm, A.-M. and Wolf-Watz, H. (eds). Plenum Press. New York, NY. USA. pp 139-146.

Manning, P.A. 1992. Molecular design of cholera vaccines. *Vaccine* **10**: 1015-1021.

Manning, P.A. 1994. Surface associated and soluble components of *Vibrio cholerae* are involved in bacteria-host interactions. *Curr Top Microbiol Immunol* In press.

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 Inaba and Ogawa serotypes of the *Vibrio cholerae* O1 lipopolysaccharides and their potential for vaccine development. *Infect Immun* **53**: 272-277.

Manning, P.A., Stroeher, U.H. and Morona, R. 1993. Molecular basis for O-antigen biosynthesis in *Vibrio cholerae* O1: Ogawa-Inaba switching . In Vibrio cholerae *and cholera.* Wachsmuth, I., Blake, P. and Olsvik, Ø. (eds). American Society for Microbiology. Washington DC. USA. In press.

Manoil, C. and Beckwith, J. 1985. Tn*phoA*: a transposable probe for protein export signals. *Proc Natl Acad Sci USA* **82**: 8129-8133.

Marcus, H., Ketley, J.M., Kaper, J.B. and Holmes, R.K. 1990. Effects of DNase production, plasmid size and restriction barriers on transformation of *Vibrio cholerae* by electroporation and osmotic shock. *FEMS Microbiol Lett* **68**: 149-154.

Masai, H. and Arai, K.-I. 1988. Operon structure of *dnaT* and *dnaC* essential for normal and stable DNA replication of *Escherichia coli* chromosome. *J Biol Chem* **263**: 15083-15093.

Masai, H., Bond, M.W. and Arai, K.-I. 1986. Cloning of the *Escherichia coli* gene for primosomal protein i: the relationship to *dnaT*, essential for chromosomal replication. *Proc Natl Acad Sci USA* **83:** 1256-1260.

Maurelli, A.T. 1989. Temperature regulation of virulence genes in pathogenic bacteria: a general strategy for human pathogens? *Microb Pathog* **7:** 1-10.

Mekalanos, J.J. 1983. Duplication and amplification of toxin genes in *Vibrio cholerae*. *Cell* **35:** 253-263.

Mekalanos, J.J. 1985. Cholera toxin: genetic analysis, regulation and role in pathogenesis. *Curr Top Microbiol Immunol* **118:** 97-118.

Mekalanos, J.J. 1992. Environmental signals controlling expression of virulence determinants in bacteria. *J Bacteriol* **174:** 1-7.

Mekalanos, J.J. and Murphy, J.R. 1980. Regulation of cholera toxin production in *Vibrio cholerae*: genetic analysis of phenotypic instability in hypertoxigenic mutants. *J Bacteriol* **141:** 570-576.

Mekalanos, J.J., Sublett, R.D. and Romig, W.R. 1979. Genetic mapping of toxin regulatory mutations in *Vibrio cholerae*. *J Bacteriol* **139:** 859-865.

Mekalanos, J.J., Swart, 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.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridisation probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* **12:** 7035-7056.

Mercurio, A. and Manning, P.A. 1985. Cellular localisation and export in *Escherichia coli* K-12 of the soluble haemolysin of *Vibrio cholerae* El Tor. *Mol Gen Genet* **200:** 472-475.

Miller, C.J., Feachem, R.G. and Drasar, B.S. 1985. Cholera epidemiology in developed and developing countries: new thoughts on transmission, seasonality and control. *Lancet* **I:** 261-263.

Miller, J.H. 1972. *Experiments in Molecular Genetics.* Cold Spring Harbour Laboratories. Cold Spring Harbour, NY. USA.

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. 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation in *Vibrio cholerae* of outer membrane proteins and virulence determinants require *toxR*. *J Bacteriol* **170:** 2575-2583.

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.

Mishra, L. and Holmes, R.K. 1987. Transcription of cholera toxin operons in wild-type and mutant strains of *Vibrio cholerae*. *Infect Immun* **55:** 1529-1532.

Mitra, S.N. 1989. Mutation induced by vibriophage PS166 infection changes biotype and phage type of *Vibrio cholerae*. *J Med Microbiol* **30:** 137-141.

Monferran, C.G., Roth, G.A. and Cumar, F.A. 1990. Inhibition of cholera toxin binding to membrane receptors by pig gastric mucin-derived glycopeptides: different effect depending on the ABO blood group antigenic determinants. *Infect Immun* **58:** 3966-3972.

Mooi, F.R. and DeGraaf, F.K. 1985. Molecular biology of fimbriae of enterotoxigenic *Escherichia coli*. *Curr Top Microbiol Immunol* **118:** 119-138.

Morona, R., Matthews, M.S., 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. *Mol Gen Genet* **224:** 405-412.

Murphy, C.K. and Klebba, P.E. 1989. Export of FepA::PhoA fusion proteins to the outer membrane of *Escherichia coli* K-12. *J Bacteriol* **171:** 5894-5900.

Murray, N.E., Brammer, W.J. and Murray, K. 1977. Lamboid phages that simplify the recovery of *in vitro* recombinants. *Mol Gen Genet* **150:** 53-61.

Naito, A., Naito, S. and Ikeda, H. 1984. Homology is not required for recombination mediated by DNA gyrase of *Escherichia coli*. *Mol Gen Genet* **193:** 238-243.

Nelson, E.T., Clements, J.D. and Finkelstein, R.A. 1976. *Vibrio cholerae* adherence and colonisation in experimental cholera: electron microscopic studies. *Infect Immun* **14:** 527-547.

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.

Neuer-Nitsche, B., Lu, X, and Werner, D. 1988. Functional role of a highly repetitive DNA sequence in anchorage of the mouse genome. *Nucleic Acids Res* **16:** 8351-8360.

Newbury, S.F., Smith, N.H., Robinson, E.C., Hiles, I.D. and Higgins, C.F. 1987. Stabilisation of translationally active mRNA by prokaryotic REP sequences. *Cell* **48:** 297-310.

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.

Newman, E.B., D'Ari, R. and Lin, R.T. 1992. The leucine-Lrp regulon in *E. coli*: a global response in search of a raison d'être. *Cell* **68:** 617-619.

Ng, S.C., Rao, A.G., Howard, O.M. and Sodetz, J.M. 1987. The eighth component of human complement: evidence that it is an oligomeric serum protein assembled from products of three different genes. *Biochemistry* **26:** 5229-5233.

Nickerson, K.W. and Aspedon, A. 1992. Detergent-shock response in enteric bacteria. *Mol Microbiol* **6:** 957-961.

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

Nilsson, P. and Uhlin, B.E. 1991. Differential decay of a polycistronic *Escherichia coli* transcript is initiated by RNaseE-dependent endonucleolytic processing. *Mol Microbiol* **5:** 1791-1799.

Nishima, Y. Miyoshi, S.-I., Nagase, A. and Shinoda, S. 1992. Significant role of an exocellular protease in utilisation of haem by *Vibrio vulnificus*. *Infect Immun* **60:** 2128-2132.

Nordström, C., Dahlqvist, A. and Jofesson, L. 1967. Quantitative determination of enzymes in different parts of the villi and crypts of rat small intestine: comparisons of alkaline phosphatase, disaccharidases and dipeptidases. *J Histochem Cytochem* **15**: 713-721.

O'Brien, A.D., Chen, M.E., Holmes, R.K., Kaper, J.B. and Levine, M.M. 1984. Environmental and human isolates of *Vibrio cholerae* and *Vibrio parahaemolyticus* produce a *Shigella dysenteriae* 1 (Shiga-like) cytotoxin. *Lancet* **I**: 77-78.

Ogawa, A., Kato, J.-I., Watanabe, H., Nair, B.G. and Takeda, T. 1990. Cloning and nucleotide sequence of a heat-stable enterotoxin gene from *Vibrio cholerae* non-O1 isolated from a patient with travellers diarrhoea. *Infect Immun* **58**: 3325-3329.

Ogg, J.E., Ryder, R.A. and Smith, H.L., Jr. 1989. Isolation of *Vibrio cholerae* from aquatic birds in Colorado and Utah. *Appl Environ Microbiol* **55**: 95-99.

Ogierman, M.A. and Manning, P.A. 1992a. Homology of TcpN, a putative regulatory protein of *Vibrio cholerae*, to the AraC family of transcriptional activators. *Gene* **116**: 93-97.

Ogierman, M.A. and Manning, P.A. 1992b. TCP pilus biosynthesis in *Vibrio cholerae* O1: gene sequence of *tcpC* encoding an outer membrane lipoprotein. *FEMS Microbiol Lett* **97**: 179-184.

Ogierman, M.A., Zabihi, S., Mourtzios, L. and Manning, P.A. 1993. Genetic organisation and sequence of the promoter-distal region of the *tcp* gene cluster of *Vibrio cholerae*. *Gene* **126**: 51-60.

Olsvik, Ø., Wahlberg, J., Petterson, B., Uhlén, M., Popovic, T., Wachsmuth, I.K. and Fields, P.I. 1993. Use of automated sequencing of polymerase chain reaction – generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *J Clin Microbiol* **31**: 22-25.

O'Neill, M.C. 1989. *Escherichia coli* promoters I: consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organisation. *J Biol Chem* **264**: 5522-5530.

Osek, J., Svennerholm, A.-M. and Holmgren, J. 1992. Protection against *Vibrio cholerae* El Tor infection by specific antibodies against mannose-binding haemagglutinin pili. *Infect Immun* **60**: 4961-4964.

Otteman, K.M., DiRita, V.J. and Mekalanos, J.J. 1992. ToxR proteins with substitutions in residues conserved with ompR fail to activate transcription from the cholera toxin promoter. *J Bacteriol* **174**: 6807-6814.

Overbye, L.J., Sandkvist, M. and Bagdasarian, M. 1993. Genes required for extracellular secretion of enterotoxin are clustered in *Vibrio cholerae*. *Gene* **132**: 101-106.

Overhauser, J. and Radic, M.Z. 1987. Encapsulation of cells in agarose beads for use with pulsed-field gel electrophoresis. *Focus* **9(3)**: 8-9.

Owen, R.L., Pierce, N.F., Apple, R.T. and Cray, W.C., Jr. 1986. M cell transport of *Vibrio cholerae* from the intestinal lumen into Peyer's-patches: a mechanism for antigen sampling and for microbial transepithelial migration. *J Infect Dis* **153**: 1108-1118.

Panigrahi, P., Tall, B.D., Russel, R.G., Detolla, L.J. and Morris, J.G., Jr. 1990. Development of an *in vitro* model for study of non-O1 *Vibrio cholerae* virulence using Caco-2 cells. *Infect Immun* **58**: 3415-3424.

Parker, C., Gauthier, D., Tate, A., Richardson, K., and Romig, W.R. 1979. Expanded linkage map of *Vibrio cholerae*. *Genetics* **91**: 191-214.

Parsot, C. and Mekalanos, J.J. 1990. Expression of ToxR, the transcriptional activator of the virulence factors in *Vibrio cholerae*, is moderated by the heat shock response. *Proc Natl Acad Sci USA* **87**: 9898-9902.

Parsot, C. and Mekalanos, J.J. 1991. Expression of the *Vibrio cholerae* gene encoding aldehyde dehydrogenase is under the control of ToxR, the cholera toxin transcriptional activator. *J Bacteriol* **173**: 2842-2851.

Parsot, C. and Mekalanos, J.J. 1992. Structural analysis of the *acfA* and *acfD* genes of *Vibrio cholerae*: effects of DNA topology and transcriptional activators on expression. *J Bacteriol* **174**: 5211-5218.

Parsot, C., Taxman, E. and Mekalanos, J.J. 1991. ToxR regulates the production of lipoproteins and the expression of serum resistance in *Vibrio cholerae*. *Proc Natl Acad Sci USA* **88**: 1641-1645.

Paruchuri, D.K., Seifert, H.S., Ajioka, R.S., Karlsson, K.-A. and So, M. 1990. Identification and characterisation of a *Neisseria gonorrhoeae* gene encoding a glycolipid-binding adhesin. *Proc Natl Acad Sci USA*. **87**: 333-337.

Patel, P., Marrs, C.F., Mattick, J.S., Russel, W.W., Taylor, R.K. and Koomey, M. 1991. Shared antigenicity and immunogenicity of type 4 pilins expressed by *Pseudomonas aeruginosa, Moraxella bovis, Neisseria gonorrhoeae, Dichelobacter nodosus* and *Vibrio cholerae. Infect Immun* **59**: 4674-4676.

Paul, S., Chaudhuri, K., Chatterjee, A.N. and Das, J. 1992. Presence of exposed phospholipids in the outer membrane of *Vibrio cholerae. J Gen Microbiol* **138**: 755-761.

Paul, S., Sen, A.K., Banerjee, N., Chatterjee, A.N. and Das, J. 1990. Lipid A mutants of *Vibrio cholerae*: isolation and partial characterisation. *Biochem Biophys Res Commun* **169**: 116-122.

Payne, S.M. and Finkelstein, R.A. 1977. Detection and differentiation of iron-responsive avirulent mutants on Congo red agar. *Infect Immun* **18**: 94-98.

Pearson, G.D.N. and Mekalanos, J.J. 1989. Recombination and replication activities of RS1, a repetitive sequence associated with the cholera toxin genetic element. In *Proceedings of the 25ᵗʰ joint conference on cholera and related diarrhoeal diseases.* National Institutes of Health. Bethesda, Md. USA pp 23-24.

Pearson, G.D.N., Woods, A., Chiang, S.L. and Mekalanos, J.J. 1993. CTX genetic element encodes a site-specific recombination system and an intestinal colonisation factor. *Proc Natl Acad Sci USA* **90**: 3750-3754.

Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* **183**: 63-98.

Peek, J.A. and Taylor, R.K. 1992. Characterisation of a periplasmic thiol: disulphide interchange protein required for the functional maturation of secreted virulence factors of *Vibrio cholerae. Proc Natl Acad Sci USA* **89**: 6210-6214.

Peitsch, M.C. and Boguski, M.S. 1990. Is apolipoprotein D a mammalian bilin-binding protein? *New Biol* **2**: 197-206.

Perry, R.D., Pendrak, M.L. and Schuetze, P. 1990. Identification and cloning of a haemin storage locus involved in the pigmentation phenotype of *Yersinia pestis. J Bacteriol* **172**: 5929-5937.

Pervaiz, S. and Brew, K. 1987. Homology and structure-function correlations between α_1-acid glycoprotein and serum retinol-binding proteins and its relatives. *FASEB J* **1**: 209-214.

Peterson, K.M. and Mekalanos, J.J. 1988. Characterisation of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonisation. *Infect Immun* **56**: 2822-2829.

Peterson, K.M., Taylor, R.K., Miller, V.L., DiRita, V.J. and Mekalanos, J.J. 1988. Coordinate regulation of virulence determinants in *Vibrio cholerae*. In *Molecular biology and infectious diseases*. Schwartz, M. (ed). Elsevier. Paris. France. 165-172.

Pierce, N.F., Cray, W.C., Jr., Kaper, J.B. and Mekalanos, J.J. 1988. Determinants of immunogenicity and mechanisms of protection by virulent and mutant *Vibrio cholerae* O1 in rabbits. *Infect Immun* **56**: 142-148.

Pinto, M., Robine-Leon, S., Appay, M.D., Kedinger, M., Triadon, N., Dussaulx, E., Lacroix, B., Simon-Assmann, P., Haffen, K., Fogh, J. and Zweibaum, A. 1983. Enterocyte-like differentiation and polarisation of the human colon carcinoma cell line Caco-2 in culture. *Biol Cell* **47**: 323-330.

Platko, J.V., Willins, D.A. and Calvo, J.M. 1990. The *ilvIH* operon of *Escherichia coli* is positively regulated. *J Bacteriol* **172**: 4563-4570.

Polarek, J.W., Williams, G. and Epstein, W. 1992. The products of the *kdpDE* operon are required for the expression of the Kdp ATPase of *Escherichia coli*. *J Bacteriol* **174**: 2145-2151.

Pollitzer, R. 1959. *Cholera*. World Health Organisation. Geneva. Switzerland.

Postle, K. 1990. TonB and the Gram-negative dilemma. *Mol Microbiol* **4**: 2019-2025.

Potten, C.S. and Allen, T.D. 1977. Ultrastructure of cell loss in intestinal mucosa. *J Ultrastruct Res* **60**: 272-277.

Pridmore, R.D. 1987. New and versatile cloning vectors with kanamycin-resistance marker. *Gene* **56**: 309-312.

Projan, S.J., Carleton, S. and Novick, R.P. 1983. Determination of plasmid copy number by fluorescence densitometry. *Plasmid* **9**: 182-190.

Pugsley, A.P. 1989. *Protein targeting.* Academic Press. San Diego, Ca. USA.

Rabbani, G.H. 1986. Cholera. *Clin Gastroenterol* **15:** 507-528.

Rader, A.E. and Murphy, J.R. 1988. Nucleotide sequences and comparison of the haemolysin determinants of *Vibrio cholerae* El Tor RV79 (Hly⁺), RV79 (Hly⁻) and classical strain 569B (Hly⁻). *Infect Immun* **56:** 1414-1419.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S.K., Nair, G.B., Shimada, T., Takeda, T., Karasawa, T., Kurazano, H., Pal, A. and Takeda, Y. 1993. Emergence of novel strain of *Vibrio cholerae* with epidemic potential in southern and eastern India. *Lancet* **341:** 703-704.

Ratcliffe, S.W., Luh, J., Ganesan, 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. *Mol Gen Genet* **168:** 165-172.

Razin, A. and Cedar, H. 1991. DNA methylation and gene expression. *Microbiol Rev* **55:** 451-458.

Redmond, J.W. 1979. The structure of the O-antigen side chain of the lipopolysaccharide of *Vibrio cholerae* 569B (Inaba). *Biochim Biophys Acta* **542:** 378-384.

Redmond, J.W., Korsch, M.J. and Jackson, G.D.F. 1973. Immunochemical studies of the O-antigens of *Vibrio cholerae*: partial characterisation of an acid labile antigenic determinant. *Aust J Exp Biol Med Sci* **51:** 229-235.

Reece, R.J. and Maxwell, A. 1991. DNA gyrase: structure and function. *CRC Crit Rev Biochem Mol Biol* **26:** 335-375.

Reed, L.J. and Muench, H. 1938. A simple method for estimating fifty per cent end points. *Am J Hyg* **27:** 493-497.

Reeves, P. 1993. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet* **9:** 17-22.

Rella, M., Mercenier, A. and Haas, D. 1985. Transposon insertion mutagenesis of *Pseudomonas aeruginosa* with Tn5 derivative: application to physical mapping of the *arc* gene cluster. *Gene* **33:** 293-303.

Richardson, A., Butler, J.A., Rutherford, M.S., Sensei, I., Gu, M.Z., Fernandez, G. and Chiang, W.H. 1987. Effect of age and dietary restriction on the expression of α_{2u}-globulin. *J Biol Chem* **262**: 12821-12825.

Richardson, K. 1991. Roles of motility and flagellar structure in pathogenicity of *Vibrio cholerae*: analysis of motility mutants in three animal models. *Infect Immun* **59**: 2727-2736.

Richardson, K., Kaper, J.B. and Levine, M.M. 1989. Human immune response to *Vibrio cholerae* O1 whole cells and isolated outer membrane antigens. *Infect Immun* **57**: 495-501.

Richardson, K., Michalski, J. and Kaper, J.B. 1986. Haemolysin production and cloning of two haemolysin determinants from classical *Vibrio cholerae*. *Infect Immun* **54**: 415-420.

Richardson, S.H. 1969. Factors influencing *in vitro* skin permeability factor production by *Vibrio cholerae*. *J Bacteriol* **100**: 27-34.

Riley, C.T., Barbeau, B.K., Kein, P.S., Kédzy, F.J., Heinrikson, R.L. and Law, J.H. 1984. The covalent protein structure of insecticyanin, a blue biliprotein from the haemolymph of the tobacco hornworm *Manduca sexta* L.. *J Biol Chem* **259**: 13159-13165.

Rosenberg, C.E. 1962. *The cholera years*. The University of Chicago. Chicago, Il. USA.

Rutter, J.M., Jones, G.W., Brown, G.T.H., Burrows, M.R. and Luther, P.D. 1976. Antibacterial activity in colostrum and milk associated with protection of piglets against enteric disease caused by K88-positive *Escherichia coli*. *Infect Immun* **13**: 667-676.

Sakazaki, R. and Tamura, K. 1971. Somatic antigen variation in *Vibrio cholerae*. *Jpn J Med Sci Biol* **24**: 93-100.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular cloning: a laboratory manual*. 2nd edn. Cold Spring Harbour Laboratories. Cold Spring Harbour, NY. USA.

Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. 1982. The *uvrB* gene of *Escherichia coli* has both *lexA* repressed and *lexA* independent promoters. *Cell* **28**: 523-530.

Sandkvist, M., Morales, V. and Bagdasarian, M. 1993. A protein required for secretion of cholera toxin through the outer membrane of *Vibrio cholerae*. *Gene* **123:** 81-86.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463-5467.

Sawyer, L. 1987. One fold among many. *Nature* **327:** 659.

Schlager, T.A., Wanke, C.A. and Guerrant, R.L. 1990. Net fluid secretion and impaired villous function induced by colonisation of the small intestine by nontoxigenic colonising *Escherichia coli*. *Infect Immun* **58:** 1337-1343.

Schneider, D.R. and Parker, C.D. 1978. Isolation and characterisation of protease-deficient mutants of *Vibrio cholerae*. *J Infect Dis* **138:** 143-151.

Schneider, K. and Beck, C.F. 1986. Promoter-probe vectors for the analysis of divergently arranged promoters. *Gene* **42:** 37-48.

Schneider, R. and Hantke, K. 1993. Iron-hydroxamate uptake systems in *Bacillus subtilis*: identification of a lipoprotein as part of a binding protein-dependent transport system. *Mol Microbiol* **8:** 111-121.

Schoner, B., Kelly, S. and Smith, H.O. 1983. The nucleotide sequence of the *Hha*II restriction and modification genes from *Haemophilus haemolyticus*. *Gene* **24:** 227-236.

Schrank, G.D. and Verwey, W.F. 1976. Distribution of cholera organisms in experimental *Vibrio cholerae* infections: proposed mechanisms of pathogenesis and antibacterial immunity. *Infect Immun* **13:** 195-203.

Schubert, D., LaCorbiere, M. and Esch, F. 1986. A chick neural retina adhesion and survival molecule is a retinol-binding protein. *J Cell Biol* **102:** 2295-2301.

Sciortino, C.V. and Finkelstein, R.A. 1983. *Vibrio cholerae* expresses iron-regulated outer membrane proteins *in vivo*. *Infect Immun* **42:** 990-996.

Sharma, D.P., Attridge, S., Hackett, J. and Rowley, D. 1987. Nonlipopolysaccharide protective antigens shared by classical and El Tor biotypes of *Vibrio cholerae*. *J Infect Dis* **155:** 716-723.

Sharma, D.P., Stroeher, U.H., Thomas, J.C., Manning, P.A. and Attridge, S.R. 1989a. The toxin-coregulated pilus (TCP) of *Vibrio cholerae*: molecular cloning of genes involved in pilus biosynthesis and evaluation of TCP as a protective antigen in the infant mouse model. *Microb Pathog* **7**: 437-448.

Sharma, D.P., THomas, C.J., Hall, R.H., Levine, M.M. and Attridge, S.R. 1989b. Significance of toxin-coregulated pili as protective antigens of *Vibrio cholerae* in the infant mouse model. *Vaccine* **7**: 451-456.

Sharples, G.J. and Lloyd, R.G. 1990. A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res* **18**: 6503-6508.

Shaw, C.E., Peterson, K.M., Sun, D., Mekalanos, J.J. and Taylor, R.K. 1989. TCP pilus expression and biogenesis by classical and El Tor biotypes of *Vibrio cholerae* O1. In *Molecular mechanisms of microbial adhesion*. Switalski, L., Höök, M. and Beachey, E. (eds). Springer Verlag. New York, NY. USA. pp 23-35.

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* **56**: 2822-2829.

Sheehy, T.W., Sprintz, H., Augerson, W.S. and Formal, S.B. 1966. Laboratory *Vibrio cholerae* infection in the United States. *J Am Med Assoc* **197**: 321-325.

Shimada, T., Nair, G.B., Deb, B.C., Albert, M.J., Sack, R.B. and Takeda, Y. 1993. Epidemic of *Vibrio cholerae* non-O1 in India and Bangladesh. *Lancet* **341**: 1347.

Shyamala, V., Schneider, E. and Ames, G.F.-L. 1990. Tandem chromosomal duplications: role of the REP sequences in the recombination event at the join-point. *EMBO J* **9**: 939-946.

Siddique, A.K., Baqui, A.H., Eusol, A., Haider, K., Hossain, M.A., Bashir, I. and Zaman, K. 1991. Survival of classic cholera in Bangladesh. *Lancet* **337**: 1125-1127.

Sigel, S.P. and Payne, S.M. 1982. Effect of iron limitation on growth, siderophore production, and expression of outer membrane proteins of *Vibrio cholerae*. *J Bacteriol* **150**: 148-155.

Sigel, S.P., Stoebner, J.A. and Payne, S.M. 1985. Iron-vibriobactin transport system is not required for virulence of *Vibrio cholerae*. *Infect Immun* **47**: 360-362.

Simon, R., Priefer, U. and Pühler, A. 1983. A broad host range mobilisation system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Biotechnology* **1:** 784-791.

Sixma, T.K., Pronk, S.E., Kalk, K.H., Wartna, E.S., van Zanten, B.A.M., Witholt, B. and Hol, W.G.J. 1991. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli. Nature* **351:** 371-377.

Smith, C.L. and Cantor, C.R. 1987. Purification, specific fragmentation and separation of large DNA molecules. *Methods Enzymol* **155:** 449-467.

Smith, C.L., Econome, J.G., Schutt, A., Kleo, S. and Cantor, C.R. 1987. A physical map of the *Escherichia coli* K-12 genome. *Science* **236:** 1448-1453.

Smith, G.R. 1988. Homologous recombination in prokaryotes. *Microbiol Rev* **52:** 1-28.

Smith, H. 1990. Pathogenicity and the microbe *in vivo. J Gen Microbiol* **136:** 377-383.

Smith, M.W. 1985. Selective expression of brush border hydrolases by mouse Peyer's-patches and jejunal villous enterocytes. *J Cell Physiol* **124:** 219-225.

Snow, J. 1855. *On the mode of communication of cholera.* 2nd edn. London. UK.

Snyder, S.H., Sklar, P.B. and Pevsner, J. 1988. Molecular mechanisms of olfaction. *J Biol Chem* **263:** 13971-13974.

Southern, E.M. 1975. Detection of specific sequence among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98:** 503-517.

Spangler, B.D. 1992. Structure and function of cholera toxin and the related *Escherichia coli* heat labile enterotoxin. *Microbiol Rev* **56:** 622-647.

Sperry, A.O., Blasquez, V.C. and Garrard, W.T. 1989. Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. *Proc Natl Acad Sci USA* **86:** 5497-5501.

Sporecke, I., Castro, D. and Mekalanos, J.J. 1984. Genetic mapping of *Vibrio cholerae* enterotoxin structural genes. *J Bacteriol* **157:** 253-261

Spreyer, P., Schaal, H., Kuhn, G., Rothe, T., Unterbeck, A., Olek, K. and Müller, H.W. 1990. Regeneration-associated high level expression of apolipoprotein D mRNA in endoneural fibroblasts of peripheral nerve. *EMBO J* **9:** 2479-2484.

Srivastava, R., Sinha, V.B. and Srivastava, B.S. 1980. Events in the pathogenesis of experimental cholera: role of bacterial adherence and multiplication. *J Med Microbiol* **13:** 1-9.

Steele, E.J., Chaicumpa, W. and Rowley, D. 1975. Further evidence for cross-linking as a protective factor in experimental cholera: properties of antibody fragments. *J Infect Dis* **132:** 175-180.

Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37:** 1015-1026.

Stern, M.J., Prossnitz, E. and Ames, G.F.-L. 1988. Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *S typhimurium*: involvement of REP sequences. *Mol Microbiol* **2:** 141-152.

Stock, J.B., Ninfa, A.J. and Stock, A.M. 1989. Protein phosphorylation and regulation of the adaptive responses in bacteria. *Microbiol Rev* **53:** 450-490.

Stock, J.B., Stock, A.M. and Mottonen, J.M. 1990. Signal transduction in bacteria. *Nature* **344:** 395-400.

Stoebner, J.A., Butterton, J.R., Calderwood, S.B. and Payne, S.M. 1992. Identification of the vibriobactin receptor of *Vibrio cholerae*. *J Bacteriol* **174:** 3270-3274.

Stoebner, J.A. and Payne, S.M. 1988. Iron regulated haemolysin production and utilisation of haem and haemolysin by *Vibrio cholerae*. *Infect Immun* **56:** 2891-2895.

Stojiljkovic, I. and Hantke, K. 1992. Haemin uptake system of *Yersinia enterocolitica*: similarities with other TonB-dependent systems in Gram negative bacteria. *EMBO J* **11:** 4359-4367.

Stroeher, U.H., Karageorgos, L.E., Morona, R. and Manning, P.A. 1992. Serotype conversion in *Vibrio cholerae* O1. *Proc Natl Acad Sci USA* **89:** 2566-2570.

Strömberg, N., Deal, C., Nyberg, G., Normark, S., So, M. and Karlsson, K.-A. 1988. Identification of carbohydrate structures which are possible receptors for *Neisseria gonorrhoeae*. *Proc Natl Acad Sci USA* **85**: 4902-4906.

Stugard, C.E., Daskaleros, P.A. and Payne, S.M. 1989. A 101-kilodalton haem-binding protein associated with Congo red binding and virulence in *Shigella flexneri* and enteroinvasive *Escherichia coli* strains. *Infect Immun* **57**: 3534-3539.

Su, C.-J., Chavoya, A. and Baseman, J.B. 1988. Regions of *Mycoplasma pneumoniae* cytadhesin P1 structural gene exist as multiple copies. *Infect Immun* **56**: 3157-3161.

Sueoka, N. 1988. Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci USA* **85**: 2653-2657.

Sugiyama, J.E., Mahmoodian, S. and Jacobson, G.R. 1991. Membrane topology analysis of *Escherichia coli* mannitol permease by using a nested-deletion method to create *mtlA-phoA* fusions. *Proc Natl Acad Sci USA* **88**: 9603-9607.

Sukupolvi, S. 1989. *Outer membrane permeability mutants of* Salmonella typhimurium *and* Escherichia coli. PhD thesis. University of Helsinki. Helsinki. Finland.

Sukupolvi, S. and Vaara, M. 1989. *Salmonella typhimurium* and *Escherichia coli* mutants with increased outer membrane permeability to hydrophobic compounds. *Biochim Biophys Acta* **988**: 377-387.

Sun, D., Mekalanos, J.J. and Taylor, R.K. 1990a. Antibodies directed against the toxin-coregulated pilus isolated from *Vibrio cholerae* provide protection in the infant mouse experimental cholera model. *J Infect Dis* **161**: 1231-1236.

Sun, D., Seyer, J.M., Kovari, I., Sumrada, R.A. and Taylor, R.K. 1991. Localisation of protective epitopes within the pilin subunit of the *Vibrio cholerae* toxin-coregulated pilus. *Infect Immun* **59**: 114-118.

Sun, D., Tillman, D.A., marion, T.N. and Taylor, R.K. 1990b. Production and characterisation of monoclonal antibodies to the toxin-coregulated pilus (TCP) of *Vibrio cholerae* that protect against experimental cholera in infant mice. *Serodiagn Immunother Infect Dis* **4**: 73-81.

Suter, F., Kayser, H. and Zuber, H. 1988. The complete amino acid sequence of the bilin-binding protein from *Pieris brassicae* and its similarity to a family of serum transport proteins like the retinol binding proteins. *Biol Chem Hoppe Seyler* **369**: 497-505.

Swerdlow, D.L. and Ries, A.A. 1993. *Vibrio cholerae* non-O1 – the eighth pandemic? *Lancet* **342:** 382-383.

Tabor, S. and Richardson, C.C. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* **82:** 1074-1078.

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.

Takeda, T., Ogawa, A., Abe, H., Kasuga, H., Nair, B.G., Pal, S.C. and Watanabe, H. 1991a. A gene encoding a heat-stable enterotoxin (O1-ST) of *Vibrio cholerae* O1 El Tor Inaba is flanked by direct repeats. In *Proceedings of the 27th joint conference on cholera and related diarrhoeal diseases.* National Institutes of Health. Bethesda, Md. USA. pp 82-87.

Takeda, T., Peina, Y., Ogawa, A., Dohi, S., Abe, H., Nair, B.G. and Pal, S.C. 1991b. Detection of heat-stable enterotoxin in a cholera toxin gene-positive strain of *Vibrio cholerae* O1. *FEMS Microbiol Lett* **80:** 23-28.

Tamplin, M.L., Ganzens, A.L., Huq, A., Sack, D.A. and Colwell, R.R. 1990. Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. *Appl Environ Microbiol* **56:** 1977-1980.

Tartof, K.D. and Hobbs, C.A. 1987. Improved media for growing plasmid and cosmid clones. *Focus* **9(3):** 12.

Taylor, R.K. 1989. Genetic studies of enterotoxin and other potential virulence factors of *Vibrio cholerae*. In *Genetics of bacterial diversity.* Hopwood, D.A. and Chater, K.F. (eds). Academic Press. London. UK. pp 309-329.

Taylor, R.K., Miller, V.L., Furlong, D.B. and Mekalanos, J.J. 1987. Use of *phoA* gene fusions to identify a pilus colonisation factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci USA* **84:** 2833-2837.

Taylor, R., Shaw, C., Peterson, K., Spears, P. and Mekalanos, J. 1988. Safe, live *Vibrio cholerae* vaccines? *Vaccine* **6:** 151-154.

Teppema, J.S., Guinée, P.A.M., Ibrahim, A.A., Pâques, M. and Ruitenberg, E.J. 1987. *In vivo* adherence and colonisation of *Vibrio cholerae* strains that differ in haemagglutinating activity and motility. *Infect Immun* **55:** 2093-2102.

Thom, S., Warhurst, D. and Drasar, B.S. 1992. Association of *Vibrio cholerae* with freshwater amoebae. *J Med Microbiol* **36:** 303-306.

Thomas, P.S. 1980. Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc Natl Acad Sci USA* **77:** 5201-5205.

Towbin, H., Staehlin, T and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci USA* **76:** 4350-4354.

Trucksis, M., Galen, J.E., Michalski, J., Fasano, A. and Kaper, J.B. 1993. Accessory cholera enterotoxin (Ace), the third toxin of a *Vibrio cholerae* virulence cassette. *Proc Natl Acad Sci USA* **90:** 5267-5271.

Tsai, C.M. and Frasch, C.E. 1982. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Anal Biochem* **119:** 115-119.

Tweedy, J.M. and Park, R.W.A. 1968. Evidence for the presence of fimbriae (pili) on vibrio species. *J Gen Microbiol* **51:** 235-244.

van Dongen, W.M.A.M. and DeGraaf, F.K. 1986. Molecular cloning of a gene coding for a *Vibrio cholerae* haemagglutinin. *J Gen Microbiol* **132:** 2225-2234.

van Dongen, W.M.A.M., van Vlerken, M.A.M. and DeGraaf, F.K. 1987. Nucleotide sequence of a DNA fragment encoding a *Vibrio cholerae* haemagglutinin. *Mol Gen (Life Sci Adv)* **6:** 85-91.

Vanek, P.G., Connaughton, J.F., Kaloss, W.D. and Chirikjian, J.G. 1990. The complete sequence of the *Bacillus amyloliquefaciens* strain H cellular *Bam*HI methylase gene. *Nucleic Acids Res* **18:** 6145.

van Heyningen, S. 1991. The ring on a finger. *Nature* **351:** 351.

Varcoe, L.T. 1992. *The characterisation of a 28 kDa protein coregulated with the haemolysin of* Vibrio cholerae. Honours thesis. Department of Microbiology and Immunology, University of Adelaide. Adelaide, S.A. Australia.

Vasil, M.L., Holmes, R.K. and Finkelstein. R.A. 1975. Conjugal transfer of a chromosomal gene determining production of enterotoxin in *Vibrio cholerae*. *Science* **187:** 849-850.

Versalovic, J., Koeuth, T. and Lupkski, J.R. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**: 6823-6831.

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

Vimr, E.R., Lawrisuk, L., Galen, J. and Kaper, J.B. 1988. Cloning and expression of the *Vibrio cholerae* neuraminidase gene *nanH* in *Escherichia coli*. *J Bacteriol* **170**: 1495-1504.

von Mechow, S., Vaidhya, A.B. and Bramucci, M.G. 1985. Mapping of a gene that regulates haemolysin production in *Vibrio cholerae*. *J Bacteriol* **163**: 799-802.

Voorhis, D.L., Dillon, S., Formal, S.B. and Isberg, R.R. 1991. An O antigen can interfere with the function of the *Yersinia pseudotuberculosis* invasin protein. *Mol Microbiol* **5**: 317-325.

Voss, E. and Attridge, S.R. 1993. *In vitro* production of toxin-coregulated pili by *Vibrio cholerae* El Tor. *Microb Pathog* In press.

Wang, J.-Y. and Syvanen, M. 1992. DNA twist as a transcriptional sensor for environmental changes. *Mol Microbiol* **6**: 1861-1866.

Wang, L.-F. and Doi, R.H. 1990. Complex character of *senS*, a novel gene for regulating expression of extracellular-protein genes of *Bacillus subtilis*. *J Bacteriol* **172**: 1939-1947.

Wang, R.F. and Kushner, S.R. 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* **100**: 195-199.

Wanke, C.A. and Guerrant, R.L. 1987. Small-bowel colonisation alone is a cause of diarrhoea. *Infect Immun* **55**: 1924-1926.

Wei, C.-F., Alianelli, G.A., Bencen, G.H. and Gray, H.B., Jr. 1983. Isolation and comparison of two molecular species of the *Bal*31 nuclease from *Alteromonas espejiana* with distinct kinetic properties. *J Biol Chem* **258**: 13506-13512.

Wenzel, R. and Herrmann, R. 1988. Repetitive DNA sequences in *Mycoplasma pneumoniae*. *Nucleic Acids Res* **16**: 8337-8350.

Wick, M.J., Madara, J.L., Fields, B.N. and Normark, S.J. 1991. Molecular cross talk between epithelial cells and pathogenic microorganisms. *Cell* **67:** 651-659.

Wikstrom, M., Jonson, G. and Svennerholm, A.-M. 1991. Production and characterisation of monoclonal antibodies to *Vibrio cholerae* soluble haemagglutinin. *APMIS* **99:** 249-256.

Williams, P.H. and Roberts, M. 1989. Iron scavenging in the pathogenesis of *Escherichia coli*. In *Genetics of bacterial diversity*. Hopwood, D.A. and Chater, K.F. (eds). Academic Press. London. UK. pp 331-350.

Williams, S.G., Attridge, S.R. and Manning, P.A. 1993. The transcriptional activator HlyU of *Vibrio cholerae*: nucleotide sequence and role in virulence gene expression. *Mol Microbiol* **9:** 751-760.

Williams, S.G. and Manning, P.A. 1991. Transcription of the *Vibrio cholerae* haemolysin gene, *hlyA*, and cloning of a positive regulator, *hlyU*. *Mol Microbiol* **5:** 2031-2038.

Winner, L., III, Mack, J., Weltzin, R., Mekalanos, J.J., Krachenbuhl, J.-P. and Neutra, M.R. 1991. New model for analysis of mucosal immunity: intestinal secretion of specific monoclonal immunoglobulin A from hybridoma tumours protects against *Vibrio cholerae* infection. *Infect Immun* **59:** 977-982.

Woese, C.R. 1987. Bacterial evolution. *Microbiol Rev* **51:** 221-271.

Wood, W.I., Gitschier, J., Lasky, L.A. and Lawn, R. 1985. Base composition independent hybridisation in tetramethyl ammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc Natl Acad Sci USA* **82:** 1585-1588.

Woods, S.A. and Cole, S.T. 1990. A family of dispersed repeats in *Mycobacterium leprae*. *Mol Microbiol* **4:** 1745-1751.

Yamaguchi, K., Yu, F. and Inouye, M. 1988. A single amino acid determinant of the membrane localisation of lipoproteins in *E. coli*. *Cell* **53:** 423-432.

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/haemolysin of *Vibrio cholerae* O1 biotype El Tor: nucleotide sequence of the structural gene (*hlyA*) and characterisation of the processed products. *Infect Immun* **58:** 4106-4116.

Yamamoto, T., Kamano, T., Uchimura, M., Iwanaga, M. and Yokota, T. 1988. *Vibrio cholerae* O1 adherence to villi and lymphoid follicle epithelium: *in vitro* model using formalin-treated human small intestine and correlation between adherence and cell-associated haemagglutinin activity. *Infect Immun* **56:** 3241-3250.

Yamamoto, T. and Yokota, T. 1988a. Electron microscopic study study of *Vibrio cholerae* O1 adherence to the mucous coat and villous surface in the small intestine. *Infect Immun* **56:** 2753-2759.

Yamamoto, T. and Yokota, Y. 1988b. *Vibrio cholerae* non-O1: production of cell associated haemagglutinins and *in vivo* adherence to mucous coat and epithelial surfaces of the villi and lymphoid follicles of human small intestines treated with formalin. *J Clin Microbiol* **26:** 2018-2024.

Yamamoto, T. and Yokota, Y. 1989a. *Vibrio cholerae* O1 adherence to human small intestinal M cells *in vitro*. *J Infect Dis* **160:** 168-169.

Yamamoto, T. and Yokota, Y. 1989b. Adherence targets of *Vibrio parahaemolyticus* in human small intestines. *Infect Immun* **57:** 2410-2419.

Yang, Y. and Ames, G.F.-L. 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc Natl Acad Sci USA* **85:** 8850 - 8854.

Yang, Y. and Ames, G.F.-L. 1990. The family of repetitive extragenic palindromic sequences: interaction with DNA gyrase and histone-like protein HU. In *The bacterial chromosome*. Riley, M. and Drlica, K (eds). American Society for Microbiology. Washington, DC. USA. pp 211-225.

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

Young, D.B. and Broadbent, D.A. 1982. Biochemical characterisation of extracellular proteases from *Vibrio cholerae*. *Infect Immun* **37:** 875-883.

Yu, J., Webb, H. and Hirst, T.R. 1992. A homologue of the *Escherichia coli* DsbA protein involved in disulphide bond formation is required for enterotoxin biogenesis in *Vibrio cholerae*. *Mol Microbiol* **6:** 1949-1958.

Zavitz, K.H. and Marians, K.S. 1991. Dissecting the functional role of PriA protein-catalysed primosome assembly in *Escherichia coli* DNA replication. *Mol Microbiol* **5:** 2869-2873.