# Molecular analysis of the *rfb* (O antigen) gene cluster of *Salmonella enterica* serovar anatum (group E1)

Lei WANG, BSc. (Nankai)
Department of Microbiology
The University of Sydney

A thesis submitted for the degree of
Doctor of Philosophy

December 1992

# Table of Contents

**Table of contents**

**Abstract**

**Statement**

**Acknowledgment**

| | | |
|---|---|---|
| **CHAPTER 1** | **Introduction** | 1 |
| 1.1 | Introduction | 1 |
| 1.2 | Inner membrane | 2 |
| 1.3 | Peptidoglycan | 2 |
| 1.4 | Periplasmic space | 3 |
| 1.5 | Fusion sites between outer and inner membranes | 3 |
| 1.6 | Outer membrane | 4 |
| 1.6.1 | Structure of the outer membrane | 4 |
| 1.6.2 | Role of the outer membrane | 5 |
| 1.7 | Lipopolysaccharide | 6 |
| 1.7.1 | Lipid A | 6 |
| 1.7.2 | Core | 7 |
| 1.7.3 | O antigen | 8 |
| 1.7.4 | The functional significance of O antigen variation | 10 |
| 1.8 | Polysaccharide components other then LPS | 10 |
| 1.8.1 | Vi antigen | 10 |
| 1.8.2 | Capsular antigens | 11 |
| 1.8.2.1 | Group I capsules | 11 |
| 1.8.2.2 | Group II capsules | 12 |
| 1.8.2.3 | Modification of capsules by O-acetylation | 13 |
| 1.8.3 | ECA | 14 |
| 1.9 | LPS biosynthesis | 15 |
| 1.9.1 | LPS biosynthesis - lipid A | 16 |

| | | |
|---|---|---|
| 1.9.2 | LPS biosynthesis - core | 17 |
| 1.9.3 | LPS biosynthesis - O antigen (*rfe*-independent) | 19 |
| 1.9.3.1 | Biosynthesis of the O unit | 19 |
| 1.9.3.2 | Polymerization and translocation | 22 |
| 1.9.3.3 | Localization of the reactions in the O antigen synthesis | 23 |
| 1.9.4 | LPS biosynthesis - O antigen (*rfe*-dependent) | 24 |
| 1.9.4.1 | The initial reaction | 25 |
| 1.9.4.2 | Chain elongation | 26 |
| 1.9.4.3 | Translocation of O antigen to core | 26 |
| 1.10 | Modification of the O antigen | 27 |
| 1.10.1 | O-acetylation | 27 |
| 1.10.2 | Glucosylation | 28 |
| 1.10.3 | Phage conversion | 29 |
| 1.10.4 | Form variation | 30 |
| 1.11 | *rfb* gene cluster | 30 |
| 1.11.1 | *rfb* gene cluster of *S. enterica* serovar typhimurium | 31 |
| 1.11.1.1 | Rhamnose biosynthetic genes | 31 |
| 1.11.1.2 | Abequose biosynthetic genes | 32 |
| 1.11.1.3 | Mannose biosynthetic genes | 32 |
| 1.11.1.4 | The transferase genes | 32 |
| 1.11.1.5 | Structure of the *rfb* gene cluster | 33 |
| 1.11.1.6 | The atypical G+C content | 34 |
| 1.11.2 | *rfb* gene clusters of *S. enterica* serovars typhi and paratyphi | 34 |
| 1.12 | Aim of this project | 35 |
| 1.13 | Taxonomy of *Salmonella* | 36 |

**CHAPTER 2 Materials and Methods**


| | | |
|---|---|---|
| 2.1 | Bacterial strains, bacteriophages and plasmids | 37 |
| 2.2 | Reagents | 37 |
| 2.3 | Growth media | 38 |
| 2.4 | DNA techniques | 39 |
| 2.4.1 | DNA isolation - chromosomal | 39 |
| 2.4.2 | DNA isolation - plasmid | 39 |
| 2.4.3 | DNA isolation - single-stranded DNA | 39 |
| 2.4.4 | Restriction endonuclease digestion | 40 |
| 2.4.5 | Agarose gel electrophoresis | 40 |
| 2.4.6 | Purification of DNA fragments from agarose gels | 40 |
| 2.4.7 | DNA quantitation | 41 |
| 2.4.8 | Dephosphorylation of linearized plasmid DNA | 41 |
| 2.4.9 | Ligation of DNA | 41 |
| 2.4.10 | Transformation | 41 |
| 2.4.11 | Nick translation | 42 |
| 2.4.12 | Colony hybridization | 42 |
| 2.4.13 | Southern hybridization | 42 |
| 2.4.14 | DNA sequencing | 42 |
| 2.5 | Bacteriophage propagation | 43 |
| 2.6 | Bacteriophage sensitivity test | 43 |
| 2.7 | Transduction | 44 |
| 2.8 | Conjugation | 44 |
| 2.9 | *In vivo* packaging of cosmid | 44 |
| 2.10 | Transposon mutagenesis | 45 |
| 2.11 | Membrane preparation | 45 |
| 2.12 | Enzyme assay of galactose transferase | 46 |
| 2.13 | Slide agglutination | 47 |
| 2.14 | Computer analysis of sequence data | 48 |

| | | |
|---|---|---|
| 2.15 | Methods used in the study of *rfbP(T)* mutants | 49 |
| 2.15.1 | Labelling of O antigen with [$^{3}$H]galactose | 49 |
| 2.15.2 | Labelling of O antigen with [$^{14}$C]mannose | 49 |
| 2.15.3 | Purification of the labelled LPS and O antigen | 50 |
| 2.15.4 | Analytical procedures | 50 |
| 2.15.5 | *In vitro* synthesis of oligosaccharide | 51 |

**CHAPTER 3 Cloning and Physical Mapping of the *rfb* Region from Strain M32**

| | | |
|---|---|---|
| 3.1 | Introduction | 53 |
| 3.2 | Preliminary analysis of similarity between *rfb* regions of strains M32 and LT2 | 53 |
| 3.3 | Cloning of the *rfb* region from strain M32 | 54 |
| 3.3.1 | pPR964 | 54 |
| 3.3.2 | pPR966 | 55 |
| 3.3.3 | pPR1030 | 55 |
| 3.3.4 | pPR1029 | 55 |
| 3.3.5 | pPR964, pPR966, pPR1030 and pPR1029 together cover the whole M32 *rfb* region | 56 |
| 3.4 | Restriction mapping | 57 |
| 3.4.1 | Restriction mapping of pPR1029 | 57 |
| 3.4.2 | Restriction mapping of pPR1030, pPR966 and pPR964 | 57 |
| 3.4.3 | Comparison with the *rfb* map of strain LT2 | 58 |
| 3.5 | The region unique to strain M32 | 58 |
| 3.6 | Summary and conclusion | 60 |

**CHAPTER 4 Analysis of the Nucleotide Sequence of M32 *rfb* Region**

| | | |
|---|---|---|
| 4.1 | Introduction | 62 |
| 4.2 | Sequencing strategy | 63 |
| 4.3 | Nucleotide sequence accession number | 65 |
| 4.4 | Sequence analysis | 65 |
| 4.4.1 | Open reading frames | 65 |
| 4.4.2 | Promoters | 66 |
| 4.5 | Proteins encoded by the ORFs | 67 |
| 4.6 | Similarity between *rfb* genes of strain M32 and *rfb* genes from other groups | 70 |
| 4.6.1 | *orf6.1* and *orf7.0* | 70 |
| 4.6.2 | *orf7.9* | 71 |
| 4.6.3 | *orf11.9* | 72 |
| 4.7 | Similarity with other proteins | 73 |
| 4.8 | G+C content and codon usage | 74 |
| 4.8.1 | G+C content | 74 |
| 4.8.2 | Codon usage and amino acid composition | 74 |
| 4.8.3 | P1, P2 and P3 values | 74 |
| 4.9 | Restriction map | 75 |
| 4.10 | Comparison of the *rfb* gene clusters of strains M32 and LT2 | 75 |
| 4.11 | Junctions between the central and flanking regions | 76 |
| 4.12 | Summary and discussion | 77 |

**CHAPTER 5 *rfbP(T)***

| | | |
|---|---|---|
| 5.1 | Introduction | 82 |
| 5.2 | Characterization of *rfbT* strains | 84 |
| 5.3 | *rfbT* mutation is in the *rfbP* gene | 85 |

| | | |
|---|---|---|
| 5.4 | *rfbP(T⁻)* mutants have galactose transferase activity | 86 |
| 5.5 | Characterization of the O-specific products of *rfbP(T⁻)* strains | 87 |
| 5.5.1 | Introduction of a *galE* mutation into various *S. enterica* LT2 strains | 88 |
| 5.5.2 | Introduction of a *pmi* mutation into various *S. enterica* LT2 strains | 89 |
| 5.5.3 | Labelling of the O-specific products | 90 |
| 5.5.4 | Gel filtration of the ethanol soluble radioactive materials | 92 |
| 5.5.5 | Chase of the previously accumulated ethanol soluble radioactive materials | 94 |
| 5.6 | Summary and discussion | 95 |

**CHAPTER 6 Final Discussion and Conclusion**

| | | |
|---|---|---|
| 6.1 | Introduction | 101 |
| 6.2 | Cloning and sequence of the *rfb* gene cluster of strain M32 | 101 |
| 6.3 | Genes common to groups E1 and B | 103 |
| 6.4 | Strain M32 lacks abequose biosynthetic pathway genes | 104 |
| 6.5 | Genes unique to group E1 | 104 |
| 6.5.1 | *orf7.9* (a putative membrane protein) | 104 |
| 6.5.2 | *orf17.4* (O antigen polymerase gene ?) | 105 |
| 6.5.3 | *orf9.6* and *orf10.8* | 106 |
| 6.6 | Structure of the *rfb* gene cluster of strain M32 | 106 |
| 6.7 | Evolution of the *rfb* gene cluster | 107 |
| 6.8 | Evolutionary relationship of *rfb* gene clusters of groups E1 and B | 112 |

| | | |
|---|---|---|
| 6.9 | *rfbP(T)* | 113 |
| 6.9.1 | *rfbT* mutation is in the *rfbP* gene | 114 |
| 6.9.2 | 'T' function is not concerned with the ligation as previously believed | 115 |
| 6.9.3 | The possible 'T' function | 116 |
| 6.9.4 | Dose *rfbP* encode a bifunctional protein ? | 119 |

**Appendix**

**Bibliography**

# ABSTRACT

*Salmonella enterica* is highly polymorphic for the O antigen, which is synthesized under the control of *rfb* genes. To understand the genetic basis of this polymorphism, comparative studies of the O antigen genetics of *S. enterica* groups A, B, C1, C2, D, and E1 have been undertaken in this laboratory (Verma *et al.*, 1988; Verma and Reeves, 1989; Jiang *et al.*, 1991; Liu *et al.*, 1991; Lee *et al.*, 1992; Brown *et al.*, 1992). The work on groups A, B, D and C2 has shown that there is substantial similarity in the *rfb* region, with only limited regions of low similarity and non homology. This thesis describes the cloning and sequence of *rfb* region from *S. enterica* strain M32 (serovar anatum, group E1), and compare it to the *rfb* region of strain LT2 (serovar Typhimurium, group B) as a representative of the related groups B, A and D. Some of the work on other groups was done at the same time as the study of group B.

The *rfb* gene cluster of strain M32 was cloned as a series of overlapping clones, and compared to that of strain LT2 by Southern hybridization and restriction mapping. The data revealed that the M32 *rfb* gene cluster has regions which are essentially identical to that of strain LT2, and which flank a central 6.5kb region of low or non similarity. It was also found that a 1.8kb region in the right end of M32 *rfb* gene cluster is not present in strain LT2. The regions unique to strain M32 were sequenced. Analysis of the sequences revealed six complete open reading frames in the central region and one in the right end region. All the open reading frames are transcribed in the same direction.

*Abstract*

Comparison of the *rfb* gene cluster of strain M32 to that of strain LT2 revealed that genes for biosynthetic pathways common to both strains are conserved and have very similar sequences, and theses genes are TDP-rhamnose and GDP-mannose biosynthetic pathway genes, and the galactose and rhamnose transferase genes. In contrast, the five genes for CDP-abequose synthesis, present in strain LT2, are absent in strain M32. A region containing *orf12.8*, *orf14.1* and the mannose transferase gene of strain LT2 is replaced in strain M32 by the DNA containing *orf7.9*, *orf9.6* and *orf10.8* which do not show any detectable nucleotide sequence similarity with the three LT2 genes. *orf7.9* encodes a protein with 12 predicted transmembrane segments, and the distribution of these 12 transmembrane segments is similar to that of the 12 segments of the protein encoded by *orf12.8* of strain LT2. The function of *orf7.9* is unclear, but it may play a role in O antigen translocation across the inner membrane. *orf10.8* encodes the mannose transferase (Liu *et al.*, 1992), and the function of *orf9.6* is unknown. *orf17.4* which is located in the right end region unique to strain M32 encodes a protein which shows significant similarity to the protein encoded by *rfc* of group B *S. enterica*, and it could be the *rfc* gene of strain M32.

The *rfb* gene clusters of strains M32 and LT2 differ in their central regions but are conserved in flanking regions, with a gradient in similarity level between the central and flanking regions. It was concluded that much of these two gene clusters is of common ancestry, and the substantial differences in the central regions result from one or more recombination events that replaced one set of genes by another unrelated set.

*Abstract*

Like the *rfb* gene cluster of group B, the *rfb* gene cluster of strain M32 is also low in G+C content. It was further concluded that the group E1 *rfb* gene cluster, like that of group B, arose in a low G+C content species and was acquired by *S. enterica* by lateral transfer recently.

The location and functional analysis of *rfbT* is also presented in this thesis.

# STATEMENT

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

[Redaction]

Lei WANG

December 1992

# ACKNOWLEDGMENTS

I would like to thank Professor Peter Reeves for his guidance, valuable comments and informative discussions. His direction and ideas have played a major role in the work presented in this thesis.

Thanks are also due to all the colleagues in the laboratory for many useful and helpful discussions, specially to Dr. Naresh Verma for his helpful discussions at the beginning of this project, and to Brian Neal and Craig Simpson for operating the automated sequencer used for most of the DNA sequencing in this project.

I thank my parents for their advice, encouragement and help throughout of the course of my studies.

Finally, I thank Lu for her patience, encouragement and support throughout.

# CHAPTER ONE: INTRODUCTION

## 1.1 INTRODUCTION

Bacterial cells have a cell wall, which is a barrier separating the inside of the cell from the environment. Bacteria have developed different ways to meet the requirements of living in various ecological niches. These differences are reflected especially in the structures of the cell walls of gram-positive and gram-negative bacteria, which differ fundamentally. Gram-positive bacteria are surrounded by a lipid bilayer with a layer of peptidoglycan, which is exterior to the lipid bilayer. Gram-negative bacteria are surrounded by two lipid bilayer membranes, named inner membrane and outer membrane, and a layer of peptidoglycan between these two membranes.

A characteristic feature of gram-negative bacteria is the presence of lipopolysaccharide (LPS) in the outer membrane. LPS consists of three structural regions: lipid A, an oligosaccharide core and the O antigen. O antigen is a polysaccharide with many repeats of an oligosaccharide unit. In the genus *Salmonella enterica*, O antigen is the major cell surface antigen and shows a high level of structural variation. The genes specifying the O antigen synthesis are in general located in the *rfb* gene cluster.

In this thesis, the molecular analysis of the *rfb* gene cluster from *S. enterica* serovar anatum (group E1) and comparison with those of other groups and in particular that of *S. enterica* serovar typhimurium

(group B) are presented. This chapter reviews the literature on the cell wall structure of gram-negative bacteria, and the structure, biosynthesis and genetics of LPS components and related cell surface antigens in *S. enterica* and other related *Enterobacteriaceae*.

## 1.2 INNER MEMBRANE

The inner membrane of *E. coli* and *S. enterica* contains 75% of the cellular phospholipid and about 6 to 9% of total cellular protein. Three species of phospholipids are found, and they are arranged in a bilayer. This lipid bilayer acts as a hydrophobic barrier allowing the intracellular concentration of small hydrophilic molecules and proteins. At least 100 major protein species are present in the bilayer (Ames and Nikaido, 1976), and they are laterally mobile within it. Many of these proteins are involved in bioenergetic and biosynthetic reactions, and the transport of specific solutes.

With a few exceptions (*e.g.*, water and medium-chain fatty acids (Melchoir and Carruthers, 1983; Maloy *et al.*, 1981)), most metabolites require specific proteins in the inner membrane to mediate their crossing of the membrane. Most of the known transport systems require energy to drive the transport, and the energy is generated from one of the following 6 sources: proton motive force, sodium motive force, ATP, PEP, Coenzyme A + ATP, and Acetyl coenzyme A.

## 1.3 PEPTIDOGLYCAN

The peptidoglycan layer surrounds the inner membrane. The glycan chain of this macromolecule consists of N-acetylglucosamine and N-acetylemuramic acid residues, which are linked by ß(1-4) linkages. The

glycan chains are cross linked by a tetrapeptide composed of L-alanine-D-isoglutamic acid-meso-diaminopimelic acid-D-alanine, while the terminal D-alanine may not present in some cases. This peptidoglycan layer attaches to the outer membrane by linking covalently to some of the lipoprotein in the outer membrane (Braun, 1975). It has been shown that it migrates together with the outer membrane portion when the outer and inner membranes were separated by physical methods such as sucrose gradients (Osborn *et al.*, 1972a). This net like macromolecule functions in maintaining the shape and rigidity of the cell.

## 1.4 PERIPLASMIC SPACE

The space between the outer and inner membranes is called the periplasmic space and appears to play vital roles in cell growth. Many hydrolytic enzymes, such as phosphatases, glycosidases, nucleases, proteases, and lipases, are housed in this space, where they are separated from the essential components inside the cells. Some other important proteins, such as the periplasmic binding proteins, are also found in this space. These proteins are involved in transport of specific amino acids, sugars, vitamins and ions through the inner membrane. Binding proteins interact with their respective inner membrane permease, thereby allowing for the translocation of those small molecules across the inner membrane.

## 1.5 FUSION SITES BETWEEN OUTER MEMBRANE AND INNER MEMBRANE

Sites of fusion between inner and outer membranes have been observed. It has been found that LPS, which is assembled in association

with the inner membrane, is translocated to the outer membrane at these sites (Bayer, 1968). Bayer and Thurow also found that the capsular antigen of *E. coli* is exported from these sites (Bayer and Thurow, 1977). It was also found that a large variety of different species of bacteriophages infect the cells at these sites, and these sites may be the routes by which bacteriophage DNA enters the cell (Bayer, 1979). These sites may also provide the basis for insertion of pili and flagellae and be involved in the export of some proteins.

## 1.6 OUTER MEMBRANE

### 1.6.1 STRUCTURE OF THE OUTER MEMBRANE

The outer membrane contains, in addition to phospholipid and protein, lipopolysaccharide (LPS) as a major constituent and enterobacterial common antigen (ECA) as a minor constituent. In comparison to the inner membrane, the outer membrane contains a small variety of proteins often in rather large quantities. These proteins are entirely different from those of the inner membrane, and with respect to their functions, can be divided into 5 classes: those with structure functions such as *OmpA* protein; those involved in interactions with other cells; those with enzyme activity such as phospholipase A1; and those (porins) which form pores. The functions of some outer membrane proteins are unknown.

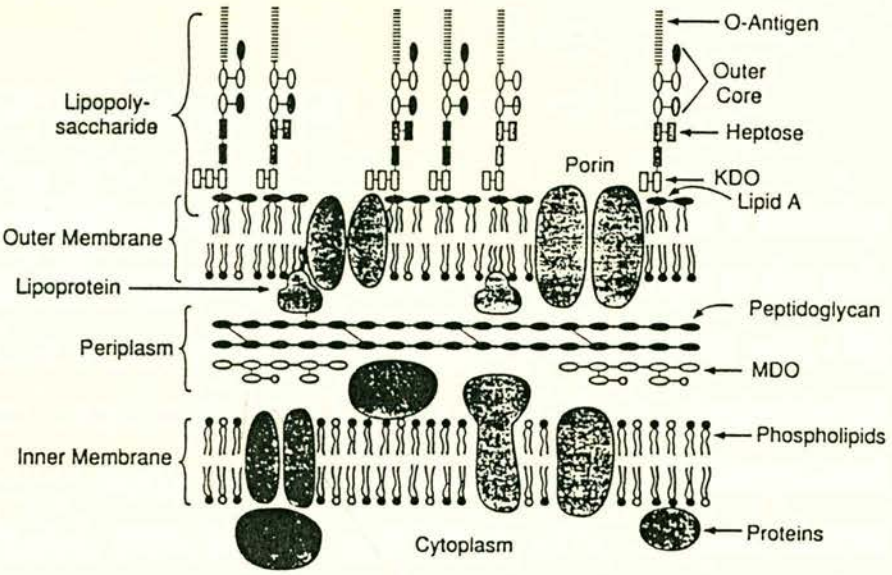
The phospholipid content of the outer membrane is much less than that of the inner membrane. The phospholipid molecules and the lipid A part of LPS molecules form a bilayer which is the basic continuum of the outer membrane, and the polysaccharide part of LPS extends to the outer leaflet of the outer membrane (Figure 1.1). The distribution of

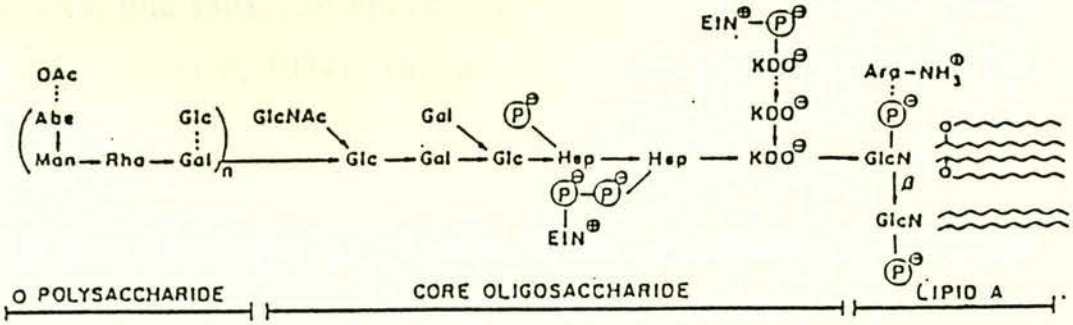
# FIGURE 1.1

## Structure of the outer membrane and LPS

Top. Schematic molecular representation of the *E. coli* envelope. Ovals and rectangles depict sugar residues. Circles represent the polar headgroups of phospholipids. MDO are membrane-derived oligosaccharides (from Raetz, 1990).

Bottom. Structure of the LPS molecule from *S. enterica* serovar typhimurium showing the lipid A, core and O polysaccharide (O antigen region of the molecule (from Nikaido and Vaara, 1987).


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


OAc
Abe
Man
Rha
Glc
Gal
n
GlcNAc
Glc
Gal
Gal
Glc
Hep
Hep
EIN
KDO
KDO
KDO
EIN
Arg-NH3
GlcN
GlcN
O POLYSACCHARIDE
CORE OLIGOSACCHARIDE
LIPID A

lipids in the bilayer is highly asymmetric: the inner leaflet is composed mainly of phospholipid, while the outer leaflet contains protein structures such as pili , flagella and porins, and polysaccharide containing structures such as capsular polysaccharide, ECA and LPS (Nikaido and Vaara, 1987).

### 1.6.2 ROLE OF THE OUTER MEMBRANE

One major role of the outer membrane is to provide a selective barrier to molecules which would otherwise be detrimental to the cell. The key molecule in this exclusion property is LPS. The cells with LPS on their outer membrane showed high resistance to hydrophobic antibiotics such as fusidic acid and erythromycin, and detergents such as bile salts (Leive, 1974; Nikaido and Nakae, 1979). Divalent cations have a strong association with the negatively charged LPS, and tend to form cross bridges between adjacent LPS molecules (Schindler and Osborn, 1979; Nikaido *et al.*, 1977), and this results in the exclusion of hydrophobic and amphiphilic molecules, and detergents from the phospholipid bilayer which is permeable to hydrophobic molecules. Treatment of gram negative cells with EDTA, which chelates divalent cations, results in the cells becoming susceptible to hydrophobic molecules, bile salts, and enzymes such as lysozyme (Nikaido, 1976; Leive, 1965; Leive, 1974). Mutants which lack major portions of the LPS molecule are also susceptible to these molecules.

LPS is also resistant to attack from lipases, phospholipase and sugar hydrolases, and thus shields the cell from these enzymes.

## 1.7 LIPOPOLYSACCHARIDE

Among the various lipopolysaccharides studied, those of *S. enterica* have been most thoroughly investigated, and may serve as the model to which results on lipopolysaccharides from other bacterial families may be compared. Usually, LPS consist of three structural regions: the lipid A moiety which is imbedded in the outer membrane, the core oligosaccharide which links lipid A with the third structural compartment, and the O-specific polysaccharide (Figure 1.1). A term sometimes used as a synonym for LPS is endotoxin. The endotoxic principle of the LPS is expressed by lipid A. The O-specific polysaccharide is also known as O antigen, and the O antigenic specificity is determined by structures of the O-specific polysaccharide.

### 1.7.1 LIPID A

Our knowledge of the lipid A structure in *S. enterica* and *E. coli* is based on studies of only few strains. For those strains, lipid A seems to have a closely related structure, although the fatty acid composition may vary. The lipid A in *S. enterica* contains a backbone of a β-D-(1-6)-linked disaccharide composed of two 2-amino-2-deoxy-D-glucose residues. Both glucosamine residues are substituted by a phosphate group: one is bound in ester linkage to C-4 of the nonreducing glucosmine; the other is bound to C-1 of the reducing glucosmine. Each lipid A molecule has six long-chain fatty acids with the chain length ranging from 12 to 16. Four of these fatty acids substitute the backbone glucosamine directly at positions 2, 3, 2', and 3', two of them link to the hydroxy groups of the two amide-linked (to the backbone) 3-OH-14:0 fatty acids. The lipid A (see Figure 1.4) of *S. enterica* and *E. coli* differs only in that the phosphate group on C-1 of the reducing glucosmine is often substituted with 4-

aminoarabinose in *S. enterica*, while that of *E. coli* carries a pyrophosphate at this position (Raetz, 1990)

Being present in the outer membrane, lipid A plays a important role in the maintenance of the integrity of the outer membrane and the cell's shape. It also interacts with various proteins in the membrane to form different functional complexes.

### 1.7.2 CORE

The core structures are quite uniform in *Enterobacteriaceae*. The best studied structures are those from *S. enterica* serovars typhimurium and minnesota, which appear to be identical (Luderitz *et al.*, 1965; Sutherland *et al.*, 1965; Hellerquist and Lind berg, 1971). The core of *S. enterica* serovar typhimurium contains two regions: a lipid A-distal region called outer core and a inner lipid A-proximal region called inner core. The inner core consists of an oligosaccharide of the core specific sugars, L-glycero-D-manno-heptose (L-D-Hep) and 3-deoxy-mannoctulosonic acid (KDO), each forming a branched trisaccharide (Figure 1.2). The first KDO (KDO I) in the inner core links to lipid A in a acid-labile (ketosidic) linkage. The outer core is a hexose oligosaccharide consisting of D-glucosamine, D-glucose, and D-galactose. The O antigen is attached to the subterminal glucose II unit of the outer core.

Extensive structural studies on other *S. enterica* strains have not been done mainly because the sophisticated chemical methods are very time consuming. Recently, using a murine monoclonal antibody, T6, specific for the D-glucosamine in the outer core, Tsang *et al.* (1991) found that 70 strains belonging to group A to E of *S. enterica* all have D-

glucosamine in the outer core and some of the strains belonging to group F to 67 lack this glycosamine. Extensive studies during the late 1960's and early 1970's showed that bacteriophage Felix O-1 lyses more than 95% of *Salmonella* serovars as a consequence of the presence of the complete *Salmonella* core (Lindberg, 1973)

Five different core types have been observed from a few *E. coli* strains studied so far (Jansson *et al.*, 1981). All five have a structure similar to that of *S. enterica* serovar typhimurium (Figure 1.2).

### 1.7.3 O ANTIGEN

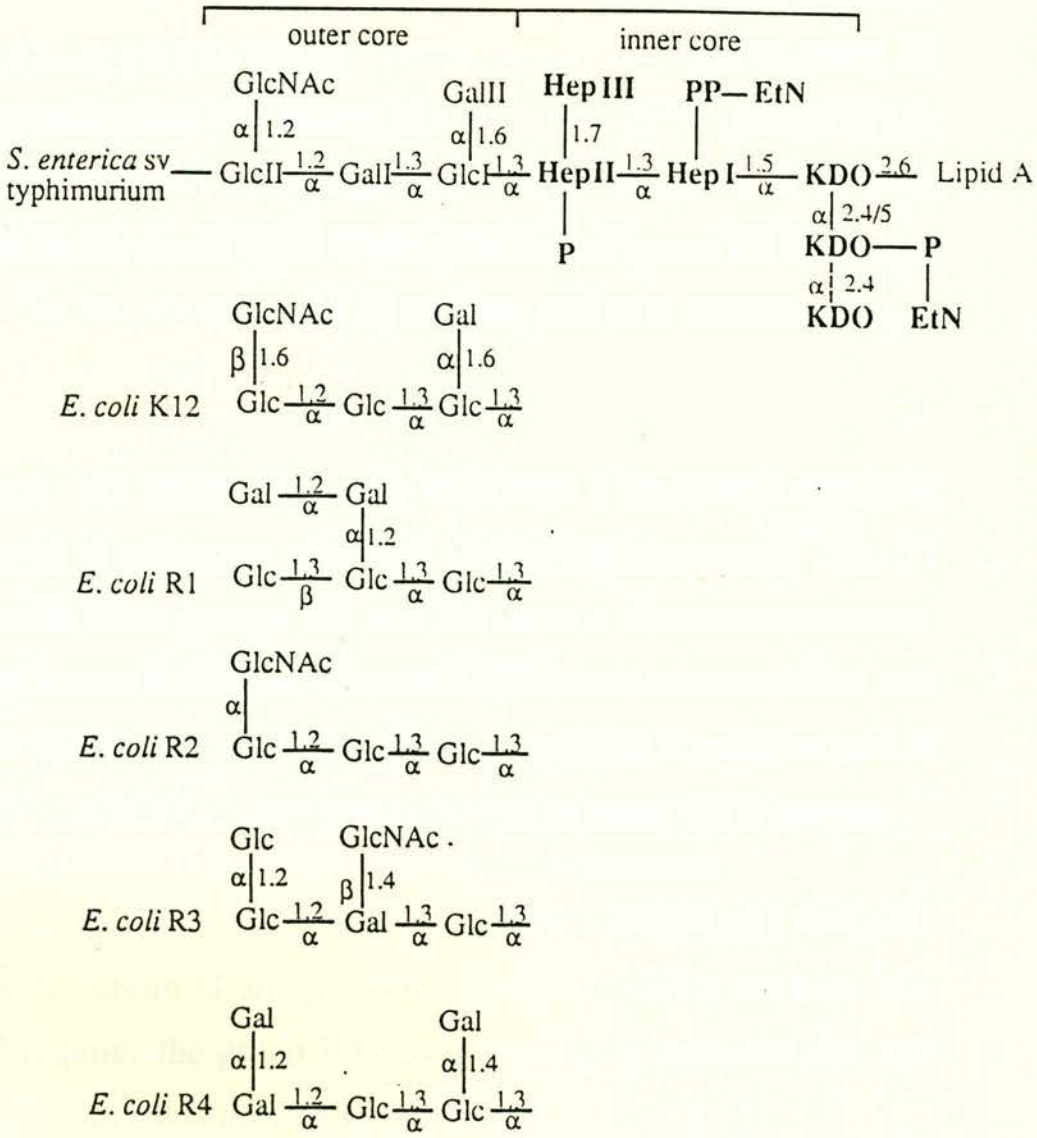
The O antigen is highly polymorphic, and there is a great number of O-specific polysaccharides, differing in composition and/or structure. There are over 60 different O antigen forms in *S. enterica*, and over 100 forms in *E. coli*. They are all made up of repeating oligosaccharide units, and one form differs from others in the composition of each unit and/or the linkages between sugars within the unit, and/or linkages between units. The repeating oligosaccharide unit is also called an O unit, and it typically consists of 3 to 6 sugars. The highly variable nature of the O antigen provides the basis for serotyping *S. enterica* strains into different groups, eg., group A, group B etc., each group then being divided into serovars using other antigens. Extensive chemistry has been done on these polysaccharide and several of the structures are known (Jann and Jann, 1984).

The O-specific polysaccharides from *S. enterica* groups A, B, D, and E have closely related structures, with a sequence of mannose-rhamnose-galactose as the backbone in the single O unit (Figure 1.3). Although the sugar sequence is the same, the anomeric configuration and/or site of substitution are different. The O-specific polysaccharides of groups A, B, and D differ from each other in the branch sugar residue on mannose: the side-chain sugar is abequose in group B, paratose in

# FIGURE 1.2

## Core structure of *S. enterica* and *E. coli*

The inner (bold) and outer core structure of *S. enterica* serovar typhimurium (top), and the outer core structures of various strains of *E. coli*. The basic inner core structure of *S. enterica* and *E. coli* is conserved, however it should be noted that differences exit both between and within each species, and are mainly through substitutions on the KDO and heptose residues (adopted from P. Brown, PhD thesis)


outer core
inner core
GlcNAc
α|1.2
GalII
α|1.6
Hep III
|1.7
PP—EtN
S. enterica sv typhimurium
GlcII 1.2 α
GalI 1.3 α
GlcI 1.3 α
Hep II 1.3 α
Hep I 1.5 α
KDO 2.6
Lipid A
α|2.4/5
P
KDO—P
α|2.4
KDO
EtN
GlcNAc
β|1.6
Gal
α|1.6
E. coli K12
Glc 1.2 α Glc 1.3 α Glc 1.3 α
Gal 1.2 α Gal
α|1.2
E. coli R1
Glc 1.3 β Glc 1.3 α Glc 1.3 α
GlcNAc
α|
E. coli R2
Glc 1.2 α Glc 1.3 α Glc 1.3 α
Glc
α|1.2
GlcNAc .
β|1.4
E. coli R3
Glc 1.2 α Gal 1.3 α Glc 1.3 α
Gal
α|1.2
Gal
α|1.4
E. coli R4
Gal 1.2 α Glc 1.3 α Glc 1.3 α

group A and tyvelose in group D. Group E1 O antigen differs from those of groups A, B, and D in the linkage between the single O units, in the linkage between mannose and rhamnose within the units, and in the lack of a branch sugar residue on mannose (Figure 1.3).

The O antigens of groups E2, E3 and E4 are closely related to that of group E1 (Figure 1.3). Group E1 strains can be converted to group E2 strains by infection of group E1 bacteria with phage ε15. Bacteriophage ε34 can convert group E2 strains to group E3 strains. The glucosylation on the galactose of group E1 O antigen results the forming of group E4 O antigen, and it is determined by a chromosomal gene *oafC*.

The group C1 O antigen has an O unit with a backbone sequence consisting of 4 mannose and one N-acetylglucosamine residues. The O polysaccharide from groups C2 and C3 have a rhamnose-mannose-mannose-galactose backbone. The main difference between these two forms is the substitution by glucose, being on a mannose unit in group C2 O antigen and on the galactose unit in group C3 O antigen. The glucose residues are then O-acetylated by O-acetyltransferase(s).

O antigens are the receptors for some *S. enterica* phages. ε15 and g341 bind to the E1 group O antigen before they can infect the host cell. Phage P22 requires the group B O antigen as the attachment site. ε34 binds to the group E2 O antigen of *S. enterica*.

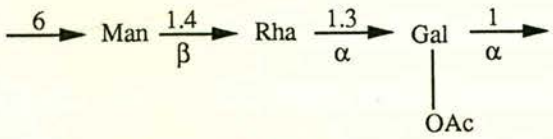
Chain lengths of the O antigen from *E. coli* and *S.enterica* strains are heterogeneous (Goldman and Hunt, 1990; Hitchcock and Brown, 1983; Jann *et al.*, 1975; Palva and Mäkelä, 1980; Peterson and McGroarty, 1985). In *S.enterica* strain LT2, O antigen molecules with 15 to 30 O units are the most predominant while LPS with O antigen chain length

# FIGURE 1.3

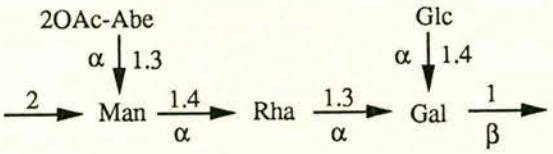
## O antigen structures of *S. enterica*

Structures of O-repeating units from *S. enterica* groups E1, E2, E3, E4 (Robbins and Uchida, 1965; Hellerqvist *et al.*,1971; Mäkelä and Stocker, 1984), and B, A, and D (Jann and Jann, 1984; Mäkelä and Stocker, 1984). Abbreviations: Abe, abequose; Gal, galactose; Glc, glucose; Man, Mannose; Par, paratose; Rha, rhamnose; Tyv, tyvelose; OAc, O-Acetyl.

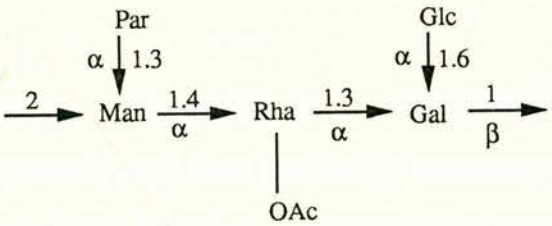
E1



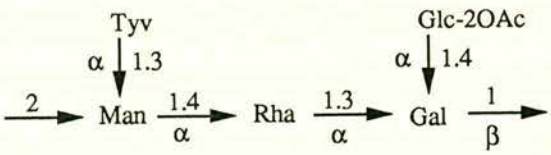

B



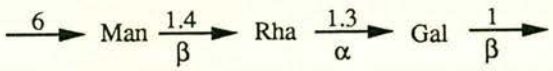

A



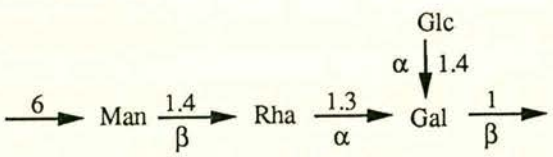

D



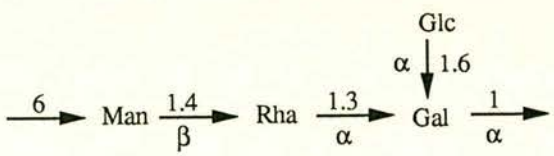

E2




E3




E4



of between 4 and 15 units is produced in such low amounts that it is not detectable with many of the techniques available. This modal distribution of O antigen chain length has also been observed in strains from *Shigella* spp. and *Pseudomonas* (Darveu and Hancock, 1983; Kuzio and Kropinsky, 1983). In some cases, a multi-modal distribution of O antigen is observed whereby a second and third cluster of high molecular weight LPS is apparent (Kusecek *et al.*, 1984; Peterson and McGroarty, 1985; Rivera *et al.*, 1988).

### 1.7.4 THE FUNCTIONAL SIGNIFICANCE OF O ANTIGEN VARIATION

The O antigen is on the surface and appears to be a target of intense selection by the immune system and this probably accounts for the origin and maintenance of the high level of variation. The variation of O antigen changes the sensitivity of the bacteria to bacteriophages, and this will increase the resistance of the bacteria. The variations also allow the bacteria to change and adapt to different environments.

## 1.8 POLYSACCHARIDE COMPONENTS OTHER THAN LPS

Several bacteria produce other extracellular polysaccharides. When these surround the bacterium as a capsule, the cell wall antigens are covered and the immunological properties are determined by the capsular polysaccharide. These polysaccharides are generally, but not always acidic.

### 1.8.1 VI ANTIGEN

It is not type specific, and it is so called because it is a virulence factor which prevents phagocytosis. It has been found in *S. enterica*, *E. coli*, and Citrobacter. Vi antigen is a

found in *S. enterica*, *E. coli*, and Citrobacter. Vi antigen is a homopolysaccharide composed of 2-acetamido-2-deoxy-α-D-galactopyranosyluronic acid residues linked by 1-4 linkage.

### 1.8.2 CAPSULAR ANTIGENS

Serologic studies on *E. coli* indicated that many strains exhibit surface antigens which are distinct from the O antigen. These additional antigens are present in an extracellular envelope or capsule covering the cell wall O antigen, and they were therefore termed capsular antigens (Jann and Jann, 1990). It is generally accepted that some capsules protect pathogenic bacteria against the non-specific host defence. Capsular antigens are acidic polysaccharides with different chemical compositions, and in *E. coli* can be divided into two groups according to their molecular weights and compositions: group I capsules and group II capsules.

#### 1.8.2.1 GROUP I CAPSULES

These polysaccharides in this group consist of repeating units with three to six sugars, and the molecular weights are generally greater than 100 kD. The repeating unit normally contains D-glucose, D-galactose, L-fucose and D-glucuronic. Some contain galacuronic acid instead of the glucuronic in the repeating unit. Some repeating units are substituted ketosidically with pyruvate to give a cyclic 1-carboxyethylidene substitution with an additional negative charge.

There are several indications that group I capsules are bound to core-lipid A, and at least part of the preparation is in fact a lipopolysaccharide (Jann and Jann, 1990). From the K40 antigen of *E.*

*coli* O8:K40, fractions can be obtained with exhibit on SDS-PAGE the ladder-like patterns characteristic of LPS. This material reacts in immunoblots with anti-K40 serum but not with anti-O8 serum (Jann and Jann, 1990). Chemical analysis showed that the material contained KDO (core constituents) and lipid A characteristic fatty acids (Jann and Jann, 1990). The linkage between the capsule and core may be similar to that between O specific polysaccharide and the core because neither of them can be broken by mild acid (Jann and Jann, 1990).

Little is known about the biosynthesis and surface expression of group I capsules. It was shown that the capsular polysaccharide is synthesized by polymerization of oligosaccharide repeating units, linked to undecaprenol pyrophosphate (Jann and Jann, 1990).

Most of the genes involved in the synthesis of group I capsules are located at the *cps* cluster adjacent to *rfb* in *E. coli*, and it is believed that this gene cluster contains at least five genes (*cpsA-E*) (Trisler and Gottesman, 1984). *cpsF* was located outside of the *cps* gene cluster at 89 min (Trisler and Gottesman, 1984). Recently, two *cps* genes from *S. enterica* sv. typhimurium have been identified and sequenced (Stevenson *et al.*, 1991).

### 1.8.2.2 GROUP II CAPSULES

The molecular weights of the capsules in this group are generally less than 50 kD (Jann and Jann, 1990). In contrast to group I capsules, which are relatively homogeneous, the group II capsules differ widely in composition and general structural features.

Some of the group II capsules contain KDO as much as 40% to 60% of the capsular polysaccharide mass, whereas the LPS contains only about 2% to 5% (Unger, 1981). There are five group II capsules known contain glucuronic acid in their repeat units (Jann and Jann, 1990). One of these capsules contains D-fructose as a substituent on the main chain. Two group II capsules contain N-Acetylneuraminic acid as the sole constituent (Jann and Jann, 1990).

Some of the group II capsular antigens are phosphate-containing polymers: they are not polysaccharides in the strict sense, but are comparable to the complex teichoic acids of certain gram-positive bacteria (Wicken and Knox, 1980).

The expression of group II capsular polysaccharides is temperature regulated, being expressed at 37°C but not at 20°C and below (Orskov *et al.*, 1984). Their expression is determined by the chromosomal *kps* genes, located close to the *serA* gene in *E. coli* (Orskov and Nyman, 1974; Orskov *et al.*, 1977). The *kps* gene cluster consists of three regions. Central region 2 directs the polymerization and is characteristic of a given *E. coli* strain and its capsular polysaccharide. Flanking regions 1 and 3 are thought to be responsible for the translocation of the capsular polysaccharide to the cell surface. The size and organization of *kps* gene cluster seem to be very similar, if not identical, in all *E. coli* with group II capsules hitherto studied (Finke *et al.*, 1991).

### 1.8.2.3 MODIFICATION OF CAPSULES BY O-ACETYLATION

Many capsules are O-acetylated, mostly to a low degree of about one acetyl group per four to eight sugar units. Although not every

repeating group in the polysaccharide is substituted, the acetylation is not random, always involving the same hydroxyl group of the same sugar constituent. The acetyl groups are in all instances immunodominant parts of a serologic epitope (Jann and Jann, 1990).

### 1.8.3 ECA

Enterobacterial common antigen (ECA) is found in all members of the *Enterobacteriaceae*. It is an amino sugar-containing heteropolysaccharide comprised of N-acetyl-D-glucosamine, N-acetyl-D-mannosa-minuronic acid (ManNAcA) and 4-acetamido-4,6-dideoxy-D-galactose (4-FucNAc). The component sugars are arranged in a linear chain of trisaccharide repeat units. ECA occurs as two forms: one highly immunogenic form and one poorly immunogenic form. In rough mutants which have a complete core but lack O antigen, the ECA occurs in the highly immunogenic form with the polysaccharide chains being covalently linked to the LPS core region. In normal wild-type strains, the polysaccharide chains O-glycosidically linked to the phospholipid to form the poorly immunogenic form.

In *E coli*, genes from *rfe -rff* gene cluster are involved in the synthesis of ECA. The *rffE* and *rffD* genes encode UDP-GlcNac-2-epimerase and UDP-ManNAc dehydrogenase respectively, and are required for UDP-ManNAc synthesis (Lew *et al.*, 1978). The first two steps in the synthesis of TDP-Fuc4NAc are catalyzed by two enzymes encoded by *rfbA* and *rfbB* from the *rfb* gene cluster (Mäkelä and Mayer, 1974; Mäkelä *et al.*, 1976). The third reaction is catalyzed by a transminase encoded by *rffA* (Meier *et al.*, 1990). *rffM* and *rffT* encode two transferases which are required for the assembly of ECA. A

functional *rfe* gene is also required for synthesis of ECA, and it was suggested that *rfe* may encode UDP-GlcNAc:undercaprenylphosphate GlcNAc-1-phosphate transferase that catalyzes the first reaction in the assembly of the ECA (Meier-Dieter *et al.*, 1992).

## 1.9 LPS BIOSYNTHESIS

The information on this topic is mainly obtained from studies of *S. enterica* and *E. coli*. LPS molecule consists of three structural regions: lipid A, core and O antigen. O antigen is synthesized separately from the synthesis of the lipid A-core complex until their final assembly into the completed LPS molecule.

The sugar components of LPS include both monosaccharides that are common metabolic intermediates or components of other macromolecules of the cells, and others that are found only in LPS. The first type of monosaccharides are biosynthesized through pathways which are determined by housekeeping genes (Sanderson and Roth, 1983). By contrast, in general the genes determining the biosynthesis of LPS-specific sugars are located in the appropriate gene cluster, e.g. the *rfb* and *rfa* clusters. Other components of LPS are the fatty acids and ethanolamine in the lipid A and deep core. Many of them are essential components of other cellular lipids and are built under the direction of the housekeeping genes of lipid metabolism (Raetz 1978, 1982, 1990).

There are two types of O antigen biosynthesis: *rfe*-dependent and *rfe*-independent (Jann and Jann, 1984). In the *rfe*-independent O antigen synthetic system, the single O unit is synthesized under the direction of *rfb* genes before it is polymerized to polysaccharide by the polymerase. In the *rfe*-dependent O antigen synthetic system, a functional *rfe* gene

is needed for O antigen expression. Synthesis of O antigens of groups A, B, D, C2 and E1 is *rfe*-independent mechanism, while synthesis of those of groups C1 and L requires the *rfe* gene, as does that of *E. coli* 08 and 09.

### 1.9.1 LPS BIOSYNTHESIS - LIPID A

The pathway leading to lipid A biosynthesis has not been completely elucidated. Lack of lipid A mutants has delayed the elucidation of both the genetics and the biosynthesis of the lipid A. It is believed that the lipid A-KDO part of the LPS are necessary for the integrity of the outer membrane, and are essential for the life of the bacteria. This may account for the lack of lipid A and KDO mutants.

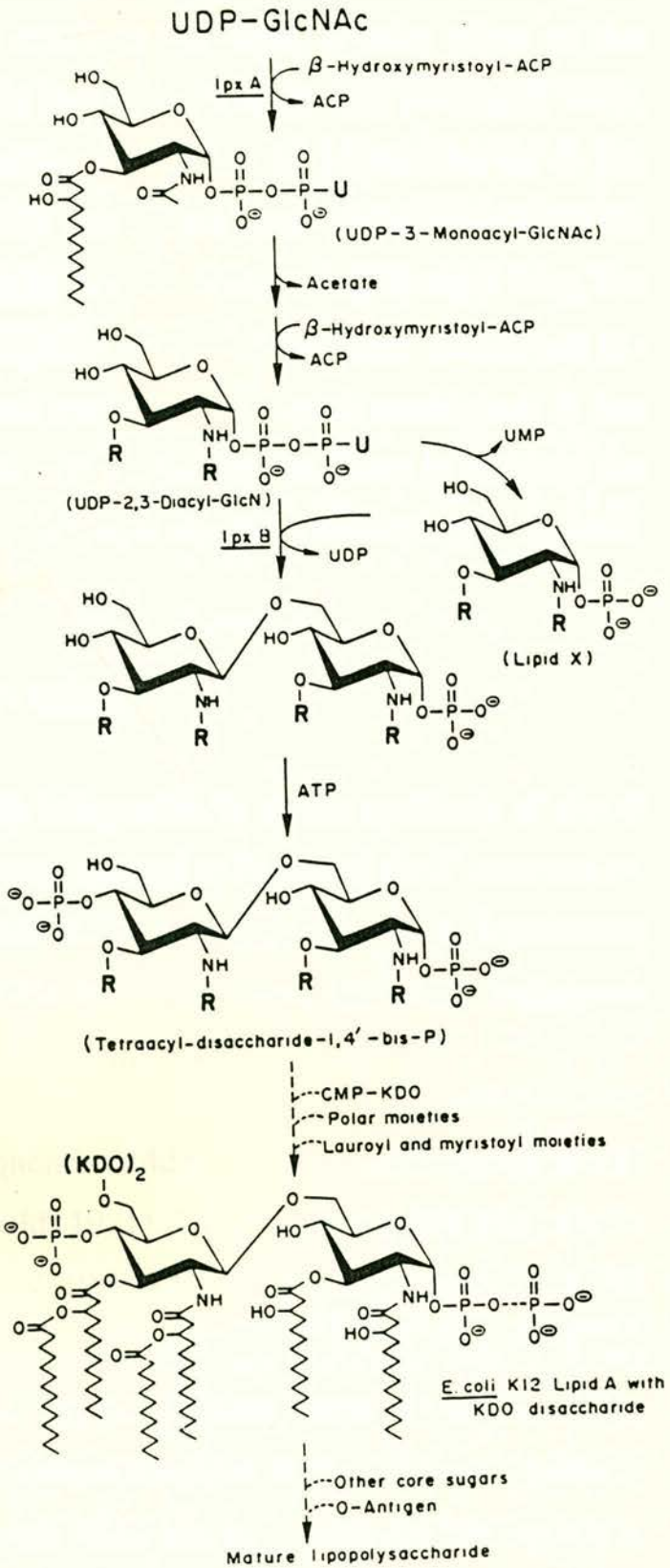
Raetz proposed a model for the biosynthesis of lipid A in *E. coli* shown in Figure 1.4 (Raetz, 1986, 1990). All of the reactions indicated in solid arrows in Figure 1.4 have been characterised in cell extracts, and their products have been verified by spectroscopic methods, as well as by chemical synthesis.

Two genes involved in the biosynthesis of lipid A have been found, and they are *lpxA* and *lpxB*. *LpxA* encodes the enzyme responsible for the first committed step in the lipid A biosynthesis, UDP-N-acetyl-glucosamine(UDP-GLcNAc) acyltransferase. The fourth step in the biosynthesis of lipid A is catalyzed by lipid A disaccharide synthase, encoded by *lpxB*. *LpxA* and *lpxB* are adjacent genes and lie at 4 min. on *E. coli* chromosome, and they have been cloned and sequenced (Crowell *et al.*, 1987; Coleman and Raetz, 1988). It has been found that *lpxB* is located immediately downstream of *lpxA* with the start codon of *lpxB* overlapping the stop codon of *lpxA*, and these two genes are presumably co-transcribed.

# FIGURE 1.4

## Biosynthetic pathway of lipid A

Enzymatic synthesis of lipid A in *E. coli*. *R* designates a β-hydroxymyristoyl moiety, having the 3*R* stereochemistry, and *U* designates uridine. The structure shown at the bottom is the minimal one that appears to be required for cell growth, but some of the molecules are substituted with a pyrophosphate residue, as indicated by the broken line (from Raetz, 1986).


UDP-GlcNAc
Ipx A
β-Hydroxymyristoyl-ACP
ACP
(UDP-3-Monoacyl-GlcNAc)
Acetate
β-Hydroxymyristoyl-ACP
ACP
(UDP-2,3-Diacyl-GlcN)
UMP
Ipx B
UDP
(Lipid X)
ATP
(Tetraacyl-disaccharide-1,4'-bis-P)
CMP-KDO
Polar moieties
Lauroyl and myristoyl moieties
(KDO)2
E. coli K12 Lipid A with KDO disaccharide
Other core sugars
O-Antigen
Mature lipopolysaccharide

Anderson and Raetz (1987) have shown that UDP-GlcNAc acyltransferase of *E. coli* is predominantly cytoplasmic. The subsequent steps in synthesis of the core on the lipid A take place at the cytoplasmic face of the inner membrane (Creeger and Rothfield, 1979; Osborn *et al.*, 1972b). Because the transfer of polysaccharide to the core occurs at the periplasmic face of the inner membrane (McGrath and Osborn, 1991a), the lipid A-core must be transported from the cytoplasmic face to the periplasmic face of the inner membrane for attachment of O antigen and transfer to the outer membrane. It has been shown that maintenance of proton motive force is required for transmembrane transposition of lipid A-core (McGrath and Osborn, 1991b).

## 1.9.2 LPS BIOSYNTHESIS - CORE

The core synthesis proceeds independently of the synthesis of the O antigen. The genes cluster in the *rfa* loci at 79 min of the *S. enterica* sv. typhimurium map are involved in the synthesis of the core (Schnaitman *et al.*, 1991). This gene cluster has been cloned and partly sequenced (MacLachlan *et al.*, 1991; Klena *et al.*, 1992)

Core is synthesized by sequential addition of the constituents to the lipid A (Osborn and Rothfield, 1971). KDO is first added from CMP-KDO as the donor (Heath *et al* ,1966). The only two known CMP-KDO biosynthetic pathway genes in *S. enterica* are *kdsA* and *kdsB*. *KdsA* maps adjacent to the *rfa* cluster, and *kdsB* maps at 16 min. (Clementz and Raetz, 1991). It was found that abnormal forms of lipid A, lacking some of the fatty acids, were accumulated in temperature-sensitive KDO-defective mutants. This accumulation suggests that KDO has to be added before lipid A is completely synthesized. 3 moles of heptose, glucose, galactose and another glucose, are then added sequentially to the incomplete core-

lipid A. Lack of a chain sugar will prevent the addition of all other distal ones.

The donor of the heptose in the core is ADP-L-glycero-D-mannose-heptose (Coleman, 1983). It is believed that *rfaD* encodes the enzyme which catalyzes the last step in the biosynthesis of this donor, and *rfaD* was mapped at the main *rfa* gene cluster. Other steps in this biosynthetic pathway have not been identified.

Mutants in *rfaC*, *rfaD*, *rfaE* and *rfaF* are defective in synthesis or attachment of ADP-L-glycero-D-Mannose-heptose. Of the four genes, only the functions of *rfa*F (transferase of heptose II) and *rfaD* (ADP-heptose epimerase) have been identified (Mäkelä and Stocker, 1984).

Gene *rfaG* is believed to be responsible for addition of the first glucose to the core (Kuo and Stocker, 1972; Austin *et al.*, 1990). *galU* mutants which lack UDP-glucose pyrophosphorylase are unable to synthesize the donor of glucose in the core, UDP-glucose, and therefore have the same phenotype as *rfaE* mutants. Another house-keeping gene, *galE*, determines the epimerization of UDP-glucose to UDP-galactose. Without external supply of galactose, a *galE* mutant will have an incomplete core without galactose, and also for those strains which have galactose in their O antigen they will have an incomplete O antigen.

*rfaB* mapping at *rfa* gene cluster encodes the galactosyl-transferase which adds the branch unit, galactose II, at C-6 of glucose I in the core. *rfaI* encodes another galacosyl transferase for attachment of the galactose I unit, and is also located in *rfa* gene cluster (Kuo and Stocker, 1972).

*rfaJ* is thought to be the structural gene for the glucosyl transferase for addition of the glucose II unit of the core (Kuo and Stocker, 1972). *rfaK* is thought to be the structural gene for N-acetylglucosamine transferase for adding the last sugar, N-acetylglucosamine unit. *rfaJ* and *rfaK* are both located in the main *rfa* gene cluster.

*rfaH* (*sfr*B in *E. coli*) which is located outside of the main *rfa* gene cluster (Kuo and Stocker, 1972) encodes a positive regulator which affects expression of several *rfa* genes (Bailey *et al.*, 1992). *rfaG* which is not regulated by the product of *rfaH* is an exception.

The *rfa* gene cluster of *E. coli* which is located at 81 min has also been cloned and partly sequenced (Schnaitman *et al.*, 1991; Roncero and Casadaban, 1992; Parker *et al.*, 1992; Pradel *et al.*, 1992).

### 1.9.3 LPS BIOSYNTHESIS - O ANTIGEN (*RFE*-INDEPENDENT)

#### 1.9.3.1 BIOSYNTHESIS OF THE O UNIT

The O unit is assembled on undecaprenol by sequential transfer of the monosaccharide residues to their appropriate acceptors (Robbins and Wright, 1971). Biosynthesis of the O antigen has been studied most extensively using *S. enterica* serovar typhimurium strain LT2 (group B) as an example. As mentioned above, the single O unit of *S. enterica* serovar typhimurium is a tetramer containing abequose, galactose,

mannose and rhamnose. The assembly of this single O unit starts by binding the galactose to the undecaprenol via a diphosphate linkage. The other three sugars are then sequentially transferred to the undecaprenol. The sugars used in these reactions are in the high energy dinucleotide sugar forms (CDP-abequose, GDP-mannose, dTDP-rhamnose and UDP-galactose) (Osborn and Tze-Yuen, 1968; Robbins and Wright, 1971; Wright *et al.*, 1967).

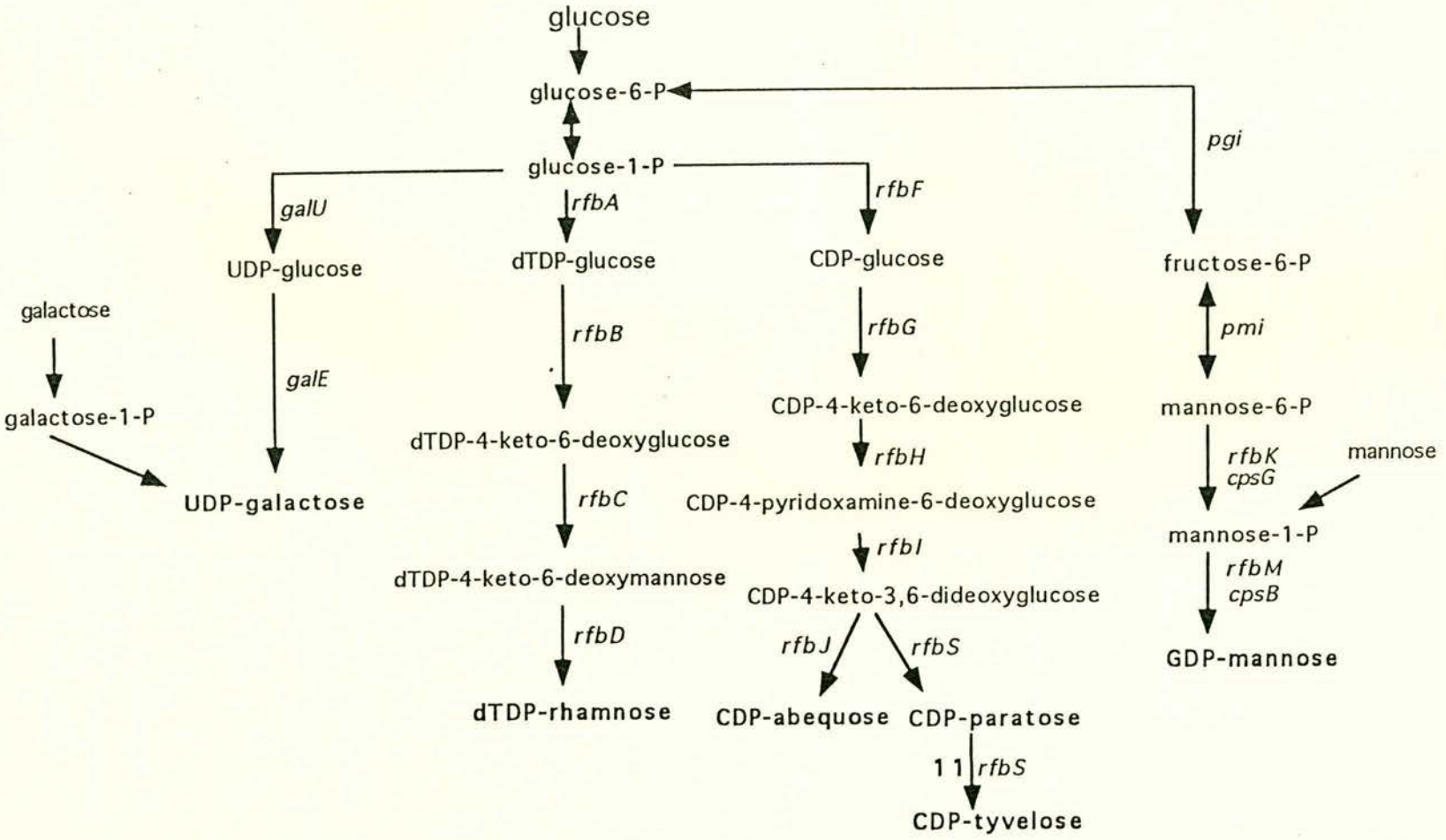
UDP-galactose is synthesized from glucose directed by housekeeping genes in the *gal* operon (Figure 1.5 and Table 1.1). UDP-galactose can also be synthesized from galatose. Strains mutant in one of the *gal* genes, *galE*, have been used in studying the synthesis of LPS. *galE* encodes UDP-glucose-4-epimerase which catalyzes the reversible interconversion of UDP-glucose and UDP-galactose. *galE* mutants are unable to synthsize UDP-galactose except from galactose, thus the synthesis of O antigen and core (galactose is also a constituent of the core) are blocked unless galactose is provided in the medium. Thus, the synthesis of LPS can be controlled by using different media, and the LPS can be specifically labelled by adding radioactive galactose to the medium.

The first step in the GDP-mannose pathway is the isomerization of fructose-6-phosphate to mannose-6-phosphate determined by the *pmi* gene at map position 32 (Figure 1.5 and Table 1.1). Mannose-6-phosphate is then converted to GDP-mannose catalyzed by phosphomannomutase and guanylyltransferase which are encoded by *rfbK* and *rfbM* in the *rfb* gene cluster (Jiang *et al.*, 1991). Mutant blocked in *pmi* can only use extrogenous mannose to synthesize O antigen, and the O antigen can be specifically labelled by feeding the

# FIGURE 1.5

## Synthesis of nucleotide sugar precursors of the O antigen

Biosynthetic pathways for the nucleotide sugar precursors of the O unit of group B strains of *S. enterica*. The genes encoding the enzymes responsible for each step are shown in bold. The enzymes encoded by these genes can be obtained from Table 1.1. The *cps* equivalent is also included for the GDP-mannose pathway (from Mäkelä and Stocker, 1984).


glucose
glucose-6-P
glucose-1-P
pgi
galU
rfbA
rfbF
UDP-glucose
dTDP-glucose
CDP-glucose
fructose-6-P
galactose
rfbB
rfbG
pmi
galE
CDP-4-keto-6-deoxyglucose
mannose-6-P
galactose-1-P
dTDP-4-keto-6-deoxyglucose
rfbH
rfbK
cpsG
mannose
UDP-galactose
rfbC
CDP-4-pyridoxamine-6-deoxyglucose
mannose-1-P
rfbI
dTDP-4-keto-6-deoxymannose
CDP-4-keto-3,6-dideoxyglucose
rfbM
cpsB
rfbD
rfbJ
rfbS
GDP-mannose
dTDP-rhamnose
CDP-abequose
CDP-paratose
11
rfbS
CDP-tyvelose

# TABLE 1.1

## Names of *rfb* genes and enzyme nomenclature

The recommended names are from the Recommendations of the Nomenclature Commit of the International Union of Biochemistry on the Nomenclature and Classification of Enzymes (Academic Press Inc. 1978 and 1984). References for the identification of enzymes are given in the last column (adopted from X-M. Jiang, PhD thesis).

| *rfb* gene | Recommended name for enzyme | Other name | Source or reference |
|---|---|---|---|
| *galU* | Glu-1-P uridylyltransferase (EC 2. 7. 7. 9) | UDP-glucose pyrophosphorylase | Kamogawa & Kurahasi 1965 |
| *galE* | UDP-glu-4-epimerase(EC 5. 1. 3. 2) | UDP-gal-4-epimerase | Maxwell & de Robichon-Szulmajster 1960 |
| *rfbA* | Glu-1-P thymidylyltransferase (EC 2. 7. 7. 24) | dTDP-glucose pyrophosphorylase, Rha-1 | Kornfeld & Glaser 1961 |
| *rfbB* | dTDP-glu-4,6-dehydratase (EC 4. 2. 1. 46) | dTDP-glucose oxidoreductase, Rha-2 | Gilbert *et al* 1965 |
| *rfbC* | dTDP-4-dehydrorhamnose-3,5-epimerase (EC 5. 1. 3. 13) | dTDP-L-rhamnose synthase (EII), Rha3-EII | Melo & Glaser 1968 |
| *rfbD* | dTDP-4-dehydrorhamnose reductase (EC 1. 1. 1. 133) | (dTDP-L-rhamnose synthase EI), Rha3-EI) | Melo & Glaser 1968 |
| *rfbF* | Glu-1-P cytidylyltransferase (EC 2. 7. 7. 33) | CDP-glucose pyrophosphorylase, Abe-1 | Mayer & Ginsburg 1965 |

| | | | |
|---|---|---|---|
| *rfbG* | CDP-glu-4,6-dehydratase (EC 4. 2. 1. 45) | CDP-glucose oxidoreductase, Abe-2 | Hey & Elbein 1970 |
| *rfbH* | | (CDP-4-keto-6-deoxy-D-glucose-3-dehydratse, enzyme E1 | Rubenstein & Strominger 1974 |
| *rfbI* | | (CDP-6-deoxy-3,4-glucoseen reductase), enzyme EIII | Rubenstein & Strominger 1967 |
| *rfbJ* | | (CDP-abequose synthase), enzyme EII | Matsuhashi & Strominger 1967 |
| *rfbS* | | (CDP-paratose synthase) | Hellerqvist *et al* 1969 |
| *rfbE* | | (CDP-2-tyvelose epimerase) | Hellerqvist *et al* 1969 |
| *rfbK* | phosphomannomutase (EC 5. 2. 4. 8) | Man-2 | Luderitz *et al* 1971 |
| *rfbM* | Man-1-P guanylyltransferase (EC 2. 7. 7. 22) | GDP-mannose pyrophosphorylase, guanosine diphosphomannose pyrophosphorylase, Man-3 | Luderitz *et al* 1971 |
| *rfbP* | Undecaprenyl-1-P galactosylphosphotransferase (EC 2. 7. 8. 6) | (galactosyl-1-P transferase; galactosyl-PP-GCL synthetase) | Osborn & Yuan Tze-Yuen 1968 |

| | | | |
|---|---|---|---|
| *rfbN* | | (rhamnosyl transferase) | Weiner *et al* 1965 |
| *rfbU* | | (mannosyl transferase) | Weiner *et al* 1965 |
| *rfbV* | abequosyl transferase (EC 2. 4. 1. 60) | | Osborn & Weiner 1968 |

Abbreviations: P, phosphate; PP, pyrophosphate; gal, galactose (-/syl); rha, rhamnose (-/syl); man, mannose (-/syl); abe, abequose (-/syl) and P-GCL, glycosyl carrier lipid.

mutant with radioactive mannose under the usual conditions of growth (Kent and Osborn, 1968a).

TDP-rhamnose is derived from glucose-1-phosphate by four reactions (Figure 1.5 and Table 1.1), and the four genes specifying these four reactions are all in the *rfb* gene cluster (Jiang *et al.*, 1991). Glucose-1-phosphate is first converted to dTDP-glucose by thymidylytransferase which is encoded by *rfbA*. dTDP-glucose is then sequentially converted to dTDP-4-keto-6-deoxyglucose, dTDP-4-keto-6-deoxymannose and dTDP-mannose by dTDP-D-glucose-4,6-dehydratase, dTDP-6-deoxy-D-glucose-3,5-epimerase and dTDP-6-deoxy-L-mannose dehydrogenase respectively. These three enzymes are encoded by *rfbB*, *rfbC* and *rfbD* respectively (Jiang *et al.*, 1991).

Abequose is one of a rare group of sugars known as 3,6-dideoxyhexoses, which are almost exclusively present in the O antigens of bacteria. CDP-abequose is synthesized from glucose-1-phosphate via five steps, and all the genes specifying these reactions are in the *rfb* gene cluster (Jiang *et al.*, 1991). The five enzymes involved in CDP-abequose synthesis are CDP-glucose-1-phosphate cytidylyltransferase, CDP-glucose-4,6-dehydratase, CDP-4-keto-6-deoxy-D-glucose-3-dehydratase, CDP-6-deoxy-3,4-glucosen reductase and CDP-abequose synthase, and they are encoded by *rfbF*, *rfbG*, *rfbH*, *rfbI* and *rfbJ* respectively (Figure 1.5 and Table 1.1).

The first reaction in the O unit assembly (transferring galactose to undecaprenol from UDP-galactose) is reversible (Nikaido and Nikaido, 1971) in contrast to the subsequent transfers (of another three sugars). The genes responsible for all these four transferring reactions are located in *rfb* gene cluster (Jiang *et al.*, 1991; Liu *et al.*, 1992).

### 1.9.3.2 POLYMERIZATION AND TRANSLOCATION

The O units synthesized on the undecaprenol are polymerized to a long polysaccharide chain while still attached to the undecaprenol. The already partly polymerized chain is transferred from undecaprenyl pyrophosphate to the nonreducing end of a newly formed oligosaccharide-p-p-undecaprenol, and one undecaprenyl pyrophosphate is released. This reaction is repeated until the chain reaches a certain length.

The polymerase which catalyses this reaction is encoded by *rfc* which is located outside the *rfb* gene cluster, between *gal* (at position 18) and *trp* (at position 34) in *S. enterica* groups A, B, and D (Mäkelä and Stocker, 1984; Collins and Hacket, 1991; Naide *et al.*, 1965). In group E1, it has been shown that this gene is probably within the *rfb* gene cluster (Nyman *et al.*, 1979). LT2 strains mutant in *rfc* have single O units in their LPS, and a semirough phenotype having antigenicity but being resistant to O antigen specific phage P22.

As mentioned above, a modal distribution of O antigen chain length has been observed in many strains. Recently Bastin *et al.* in this laboratory have identified and sequenced the gene which determines the chain length of the O antigen, and termed it *cld* (chain length determinant) (Bastin *et al.*, 1992). In *E. coli* O111 and *S. enterica* sv. typhimurium, *cld* lies between *gnd* gene and *his* operon (Bastin *et al.*, 1992). It was proposed that the *cld* protein forms a complex with the polymerase and that the complex has two states, an E state facilitating extension and a T state facilitating transfer to core. The *cld* protein is postulated to enter the E state as O antigen synthesis starts and to shift to the T state after a predetermined time (Bastin *et al.*, 1992).

The polymerized O antigen or the single O unit is then transferred to the lipid A-core from undecaprenol pyrophosphate. The transfer requires the complete core, and *rfa* mutants with incomplete core accumulate undecaprenol linked O antigen (O hapten) (Beckmann *et al.*, 1964; Weiner *et al.*, 1965; Kent and Osborn, 1968b; Lindberg *et al.*, 1972). O haptens were also found in two types of strains, which have mutations in *rfaL* and *rfbT* with complete core, and hence it was proposed that these two genes were responsible for the transferring the O antigen from undecaprenol to core (Mäkelä and Stocker, 1984).

### 1.9.3.3 LOCALIZATION OF THE REACTIONS IN THE O ANTIGEN SYNTHESIS

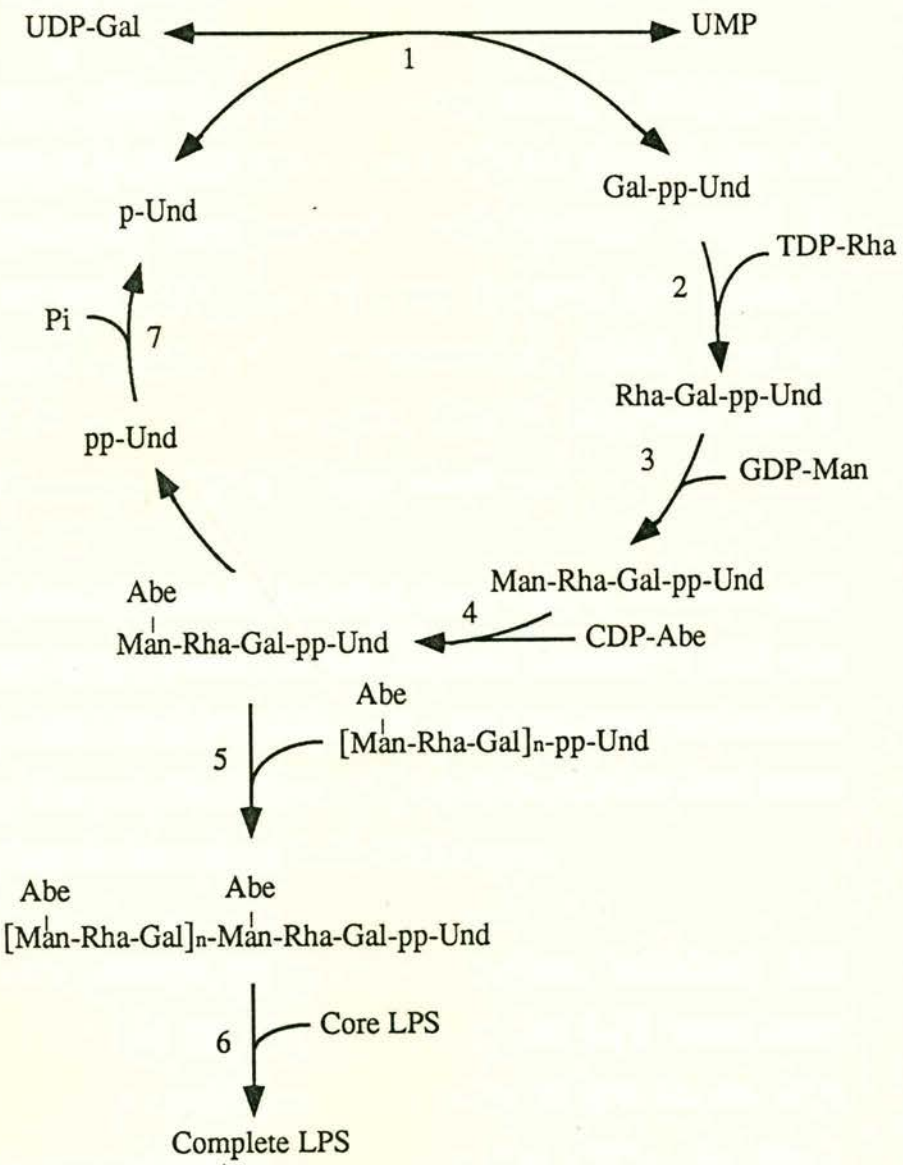
The biosynthetic pathway of LPS is shown in Figure 1.6. All the nucleotide sugars which are precursors of the sugar constituents in the O unit are synthesized by cytosolic enzymes in the cytoplasm. The single O unit is assembled on the inner membrane as shown by a pulse-chase experiment (Osborn *et al.* 1972b). The first enzyme for the assembly of O unit, galactose transferase, is within the isolated inner membrane fraction, and this also suggests that the O unit is synthesized at the inner membrane (Osborn *et al.*, 1972a).

A number of observations led to the conclusion that the attachment of O antigen to core takes place at the periplasmic face of the inner membrane. First immunoelectron microscopy studies have established that newly synthesized O-positive lipopolysaccharide is transiently located at the periplasmic face of the inner membrane as an intermediate in translocation to the outer membrane (Mulford and Osborn, 1983). In the same study it was also shown that undecaprenol linked O polysaccharide accumulated stably at the periplasmic face of inner

# FIGURE 1.6

## Biosynthetic pathway of LPS

Pathway of LPS biosynthesis (McGrath and Osborn, 1991). Steps 1, 2, 3, and 4 are catalyzed by galactose, rhamnose, mannose, and abequose transferases respectively. Step 5 is catalyzed by O antigen polymerase. Abbreviations: Und, undecaprenol; pp, pyrophosphoryl; others as in Figure 1.3.


UDP-Gal
UMP
1
p-Und
Gal-pp-Und
TDP-Rha
2
Pi
7
Rha-Gal-pp-Und
pp-Und
3
GDP-Man
Man-Rha-Gal-pp-Und
Abe
Man-Rha-Gal-pp-Und
4
CDP-Abe
Abe
[Man-Rha-Gal]n-pp-Und
5
Abe
Abe
[Man-Rha-Gal]n-Man-Rha-Gal-pp-Und
6
Core LPS
Complete LPS

membrane when transfer to lipid A-core is blocked by a mutation in core biosynthesis (Mulford and Osborn, 1983). This strongly suggested that the final transfer of O antigen to lipid A-core takes place at the periplasmic face of the inner membrane.

Recently, by doing pulse-chase experiments using a temperature sensitive core mutant McGrath and Osborn (1991a) showed that the transfer of previously accumulated undecaprenol linked O polysaccharide to newly synthesized core is rapid and virtually complete. The quick transfer of O polysaccharide, which is on the periplasmic face of the inner membrane, to core suggested that it occurs at the periplasmic face of the inner membrane.

Using a strain with a *rfc* mutation and a temperature sensitive core mutation, McGrath and Osborn (1991a) also showed that the O unit was chased into lipopolysaccharide rapidly when the core was synthesized. Thus, it is likely the polymerization of O polysaccharide also takes place at the periplasmic face of the inner membrane.

Based on immunoelectron microscopic studies, Muhlradt and coworkers postulated that the translocation of the LPS from inner membrane to the outer membrane takes place at the fusion sites between outer and inner membranes (Muhlradt *et al.* 1973). The details of this translocation remain unknown.

### 1.9.4 LPS BIOSYNTHESIS - O ANTIGEN (*RFE*-DEPENDENT)

It was first found by Mäkelä *et al.* (1970) that in certain bacterial strains such as *S. enterica* strains of groups C1 and L and *E. coli* O8 and O9 strains, the synthesis of O antigen depended on the *rfe* gene

(Jann and Jann, 1984). This gene is not needed in *S. enterica* groups B and E strains, which had been used in the previous biosynthetic studies. The *rfe* gene is required for biosynthesis of the enterobacterial common antigen ECA (Mäkelä and Mayer, 1974). It was observed that semirough (SR) mutants with only one single O unit in the LPS had never been isolated from *E. coli* O8 and O9, and it was concluded that a second mechanism of O antigen synthesis was operating (Jann and Jann, 1984). Much of the work on the *rfe*-dependent O antigen biosynthesis was carried out using *E. coli* O8 and O9 strains, whose O antigens are mannose polymers.

### 1.9.4.1 THE INITIAL REACTION

Analysis of the O8 and O9 polysaccharide, generated *in vitro* and liberated from the membranes with mild acid, revealed in both cases the presence of glucose at the reducing end (Flemming and Jann, 1978a, 1978b). Since the O antigens of O8 and O9 do not contain glucose, it was assumed that there was a intermediary glucose-containing mannose acceptor. This hypothesis was confirmed, and compositional analysis, biochemical evidence and mass spectrometry showed that the intermediary mannose acceptor is α-glucosyl diphosphateundecaprenol.

The biosynthesis of the O8 and O9 polysaccharide is initiated by transfer of α-glucose-1-phosphate to undecaprenyl phosphate to form a-glucosyl-diphosphateundecaprenol, which functions as an acceptor for mannose incorporation. It was also found that β-glucosyl-diphosphateundecaprenol can not function as acceptor, but the α-galactosyl derivative was about as effective an acceptor as the α-glucosyl derivative (Jann and Jann, 1984).

### 1.9.4.2 CHAIN ELONGATION

Direct analysis of the incubation mixtures with SDS-PAGE (Kopmann and Jann, 1975) and column chromatography of the products after liberation from the membranes with mild acid revealed that polymer can be detected within one minute after the addition of GDP-mannose. No intermediate of small molecular size was found. It was therefor concluded that the polysaccharides grow in a single chain mechanism without soluble intermediates.

The finding of glucose at the reducing end of the free mannose (Flemming and Jann, 1979a, 1979b) indicated that the mannose polymer grows at the nonreducing end.

### 1.9.4.3 TRANSLOCATION OF O ANTIGEN TO CORE

Study of the core structure of an *rfe* mutant showed that it lacks a terminal glucose compared to the core from $\overset{\alpha}{\lambda}$*rfe*$^+$ strain. It was concluded that in the translocation reaction, the glucose unit of the mannose acceptor is transferred to a precore to form the complete core before the O antigen is translocated to the core (Weisgerber *et al.*, 1984).

Although much of the present evidence points to a single chain mechanism in *rfe*-dependent O antigen synthesis, SDS-PAGE analysis of LPS from strains of group C1 *S. enterica* and *E. coli* O9 show a typical ladder pattern similar to that observed for *rfe*-independent systems (Kido *et al.*, 1989; Mäkelä and Stocker, 1984). This ladder pattern presumably correspond to the sequential addition of

oligosaccharide repeat units, and is thus a strong argument in favour of polymerization of pre-synthesized O units in these strains.

## 1.10 MODIFICATION OF THE O ANTIGEN

The basic structure of O antigen is synthesized under the direction of *rfb* and *rfc* genes. However the O antigen can be modified, and this modification is not essential for the completion of O antigen. Modification can be a constant part of the LPS structure in one species or one clone; in which case it is responsible for variation between serotypes within one O group. Modification may also cause variation within a single strain, due to rapid on-off variation of expression of a gene; this is termed form variation (Kauffmann, 1941, 1966). Furthermore, the modification may only happen to a portion of the O units in a LPS molecule.

The most common modifications in *S. enterica* are based on O-acetylation and glucosylation of specific residues in the O units, although other substitutions are also found (Kenne and Lindberg, 1983).

### 1.10.1 O-ACETYLATION

In *S. enterica* serovar typhimurium, the abequose residue of the O unit is acetylated, and the gene responsible for this acetylation is *oafA* at position 43 on the chromosome (Mäkelä, 1966; Johnson *et al.*, 1966; Kishi and Iseki, 1973a,b). This O-acetylation on abequose confers O antigen factor 5. The single O unit in the LPS of a *rfc* mutant can also be acetylated, and this may indicate that the acetylation occurs before the polymerazation (Naide *et al.*, 1965). The rhamnose residue of the serovar typhimurium O antigen can also be O-acetylated by lysogenization (Wollin *et al.*, 1987)

The galactose of the O units in group E1 is also modified by the O-acetylation, but the gene(s) responsible for this modification has not been identified. Lysogenization by phage ε15 or g341 prevents this acetylation, and a phage mutant lacks this blocking capacity has been isolated (Uetake *et al.*, 1958; Robbins *et al.*, 1965; Robbins and Uchida, 1965). The O-acetylation on galactose serves as the acceptor of O specific phage g341 but prevent the adsorption of phage ε34 (Robbins and Uchida, 1965).

### 1.10.2 GLUCOSYLATION

Glucosylation of the galactose residue in the O unit of *S. enterica* serovar typhimurium through a 1,4 linkage is determined by chromosomal genes in the *oafR-oafE* cluster close to *purE* at position 11 (Mäkelä and Mäkelä, 1966; Mäkelä, 1973). This glucosylation is responsible for the O antigen factor $12_2$, and for the preventing the adsorption of phages 9NA and KB1 (Mäkelä and Stocker, 1984).

In *S. enterica* group E1 strains, the glucosylation of the galactose residue in the O unit results the new antigenic factor 1. It is determined by chromosomal gene *oafC* (Plosila and Mäkelä, 1972), and the group E strains with antigenic factor 1 are subgrouped into group E4.

The glucosyl donor is glucose, linked via a phosphate to undecaprenol (Nikaido *et al.*, 1971; Nikaido and Nikaido, 1971; Takeshita and Mäkelä, 1971). The undecaprenol-linked O polysaccharide serves as glucosyl acceptor.

## 1.10.3 PHAGE CONVERSION

Many phage genes introduced into bacteria by phage infection can also modify the O antigen. O antigen acts as receptor for many different phages, and a change in the O antigen structure is expected to alter its phage receptor function and prevent the further infection from the same phage.

The galactose in the O unit of *S. enterica* serovar typhimurium can be glucosylated through a 1,6 linkage, and this modification is determined by a gene from phage P22. This modification also uses undecaprenol linked glucose as the donor for glucose (Mäkelä and Stocker, 1984).

The linkage between O units in group E1 strains can be changed by genes from phage ε15. The normal linkage between O units in group E1 is in the α-configuration through C-1 of galatosyl residue to C-6 of mannosyl residue. The phage ε15 changes this linkage to the β-configuration and is responsible for a new antigenic factor 15 (Uetake *et al.*, 1958; Robbins and Uchida, 1965; Robbins *et al.*, 1965). This phage was shown to carry both the structural gene for the new, β-polymerase and the gene specifying a protein inhibitor for the old, α-polymerase (Robbins *et al.*, 1965; Nyman *et al.*, 1979). Strains with O antigenic factor 15 are known as group E2 strains, and are all thought to carry ε15 or equivalent phages.

A glucosylation through a 1,4 linkage on the galactose in the O unit of group E2 strains is effected by genes of the prophage ε34 and recognized as the antigenic factor 34, giving concomitant resistance to adsorption of the phage ε34 itself (Iseki and Hamano, 1958; Uetake *et al.*, 1955). Strains with antigenic factor 34 are put in group E3.

### 1.10.4 FORM VARIATION

Variable expression of two O antigen factors (1 and $12_2$) has been described by Kauffmann in many *S. enterica* strains (Kauffmann, 1941; Kauffmann and Rohde, 1961). In both cases, the variation concerned a glucosylation-type modification. Several other instances of variation have also been described (Mäkelä and Stocker, 1984).

In a culture of a variable strain, some colonies show full expression of the variable factor, as seen by a strong agglutination, while others do not express this factor at all, and few colonies are intermediate. The bacteria can switch from one form to the other, and back again, and this usually occurs at relatively high frequency (once per $10^2$ - $10^4$ cell divisions) (Kauffmann, 1966).

In general, the form variation of the LPS is determined by phage or chromosomal genes outside of *rfb*. Genetics studies with the factor $12_2$ in *S. enterica* serovar typhimurium showed that both its variability and form (+ or -) were determined by *oafR*, which is closely linked to the biosynthetic genes determining the $12_2$-type glucosylation (Mäkelä and Mäkelä, 1966; Mäkelä, 1973).

## 1.11 *RFB* GENE CLUSTER

The specific genes involved in O unit synthesis map together at the *rfb* gene cluster (Mäkelä and Stocker, 1984) at 42 min: on the *S. enterica* chromosome (Sanderson and Roth, 1988). Transconjugation of the *rfb* gene clusters between strains with different O antigen causes a change of O antigen. Hence the *rfb* gene cluster must contain both the genes that determine the synthesis of specific monosaccharides present

only in the O unit and those that determine their correct assembly into the O unit.

### 1.11.1 *RFB* GENE CLUSTER OF *S. ENTERICA* SEROVAR TYPHIMURIUM

Most studies on the *rfb* gene cluster have been done using *S. enterica* serovar typhimurium strain LT2. Only recently, work on the *rfb* gene clusters of other organisms, such as *E. coli* strains and other *S. enterica* strains started. Using a series of deletion mutants that extend into the *rfb* gene cluster from its right end, the *his* operon, a partial gene map of the *rfb* gene cluster from *S. enterica* sv. typhimurium was determined by a combined genetic-biosynthetic approach (Nikaido *et al.*, 1967; Levinthal and Nikaido, 1969). Recently, this gene cluster has been cloned and sequenced, and the positions of most of the genes involved in O antigen synthesis have been determined in this laboratory (Jiang *et al.*, 1991). By making insertions and partial deletions of a cosmid clone which containing the whole *rfb* gene cluster, this *rfb* gene cluster has been identified as a 19kb region with 17 open reading frames (ORFs) (Figure 1.7) (Jiang *et al.*, 1991). The genes of individual biosynthetic pathways are grouped together in the gene cluster (Jiang *et al.*, 1991).

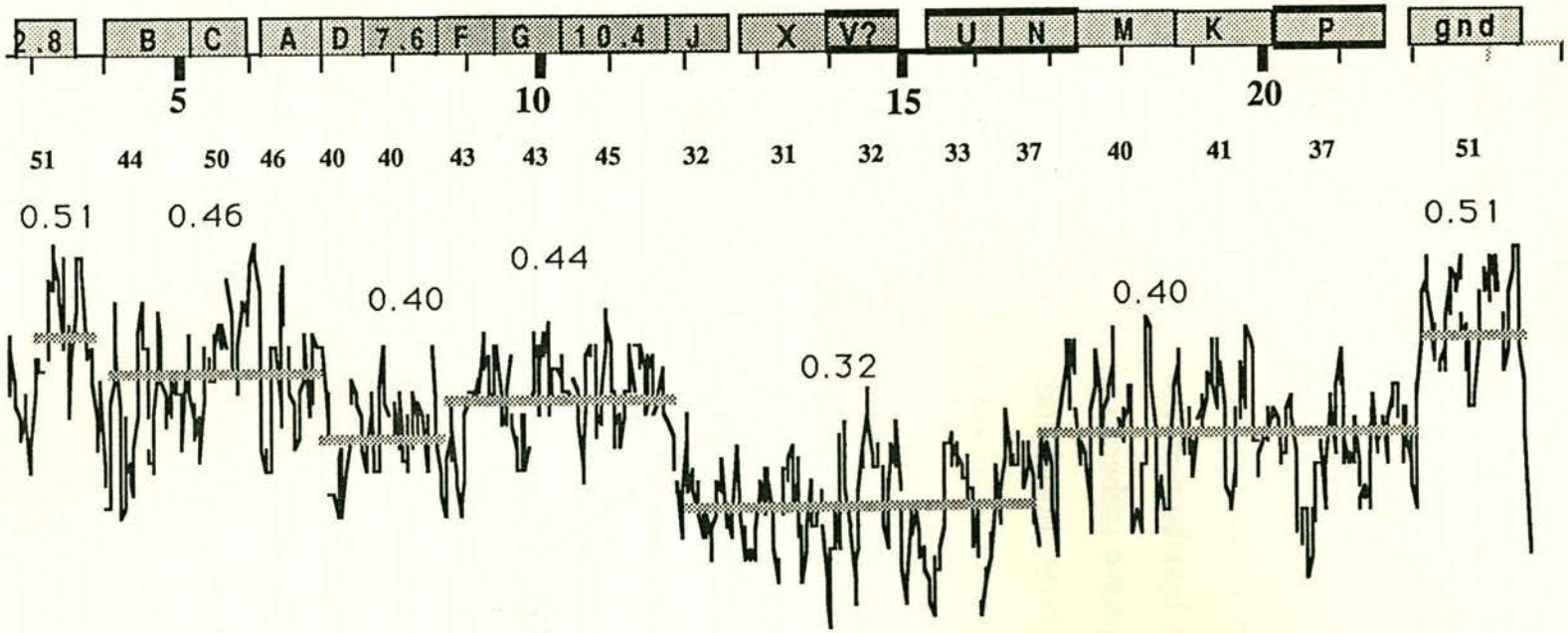
#### 1.11.1.1 RHAMNOSE BIOSYNTHETIC GENES

It takes 4 enzymes to make TDP-rhamnose from glucose-1-phosphate, and they are glucose-1-phosphate thymidylytransferase, dTDP-D-glucose 4,6-dehydratase, dTDP-6-deoxy-D-glucose-3,5 epimerase and dTDP-6-deoxy-l-mannose-dehydrogenase. The genes which encode

# FIGURE 1.7

## The genes and major features of the *rfb* gene cluster of strain LT2

The *rfb* gene cluster of *S. enterica* serovar typhimurium strain LT2 is shown in the upper portion. Genes are indicated by open boxes with the name of that gene within it, Where the function is unknown, the gene has been given a number corresponding to the start position of that ORF. The G+C content is given below each gene and a G+C content plot is given below the map (each point gives an average of over 199 bases).


2.8
B
C
A
D
7.6
F
G
10.4
J
X
V?
U
N
M
K
P
gnd
5
10
15
20
51 44 50 46 40 40 43 43 45 32 31 32 33 37 40 41 37 51
0.51
0.46
0.40
0.44
0.32
0.40
0.51

these four enzymes are all located in the *rfb* gene cluster: they are *rfbA*, *rfbB*, *rfbC* and *rfbD* respectively (Figure 1.7).

#### 1.11.1.2 ABEQUOSE BIOSYNTHETIC GENES

There are 5 steps in the CDP-abequose pathway starting from glucose-1-phosphate. Three of the CDP-abequose pathway genes have been identified (Jiang *et al.*, 1991; Wyk and Reeves, 1989), being *rfbF*, *rfbG* and *rfbJ*. Glucose-1-phosphate cytidylyltransferase, CDP-glucose 4,6-dehydratase and abequose synthase are encoded by these three genes, respectively. The other two genes in this pathway in strain LT2 have not yet been identified directly but comparison of *rfb* gene clusters of groups B and C2 (Brown *et al.*, 1992) suggests that they are *orf7.6* and *orf10.4*. These five genes group in a region adjacent to the rhamnose genes in the *rfb* gene cluster (Figure 1.7).

#### 1.11.1.3 MANNOSE BIOSYNTHETIC GENES

Two genes are responsible for the last two steps in the biosynthesis of GDP-mannose from mannose-6-phosphate. These two genes are *rfbK* and *rfbM*, and they encode phosphomannomutase and mannose-1-phosphate guanylytransferase respectively (Jiang *et al.*, 1991). The *pmi* gene, which encodes the phosphomannoseisomerase, lies at 32 min. on the *S. enterica* chromosome, well away from *rfb* gene cluster.

#### 1.11.1.4 THE TRANSFERASE GENES

The genes encoding the transferases for O antigen assembly all lie in the *rfb* gene cluster. *orf16.5* encodes the rhamnose transferase, and *orf15.4* encodes the mannose one (Liu and Reeves, unpublished data).

*orf14.1* is probably the abequose transferase gene (Liu and Reeves, unpublished data).

Because UDP-galactose is generally available in *S. enterica*, there are no UDP-galactose biosynthetic pathway genes in *rfb* gene cluster but only a UDP-galactose transferase gene, *rfbP*. *rfbP* maps at the right end of the *rfb* gene cluster (Jiang *et al.*, 1991), and encodes the transferase catalyzing the first reaction in the O antigen biosynthesis.

### 1.11.1.5 STRUCTURE OF THE *RFB* GENE CLUSTER

The expression of the *rfb* genes in deletion mutants that have the deletion extending from the *his* operon into the *rfb* gene cluster suggests that the direction of transcription in the *rfb* gene cluster is towards the *his* operon (Levinthal and Nikaido, 1969). This early work covered only part of the gene cluster. By analysis of the DNA sequence, it has been found all 17 ORFs in this cluster are read in one direction (towards *his* operon) with no ORFs found in the other direction (Jiang *et al.*, 1991).

Using the available promoter searching programs, two potential promoters have been found with the mRNA startpoint at positions 2803 and 3927 (Jiang *et al.*, 1991). It was also found that those two promoters are likely to be under positive control because they are not strong enough to function without modification (Jiang *et al.*, 1991). These two promoters are at the upstream end of the *rfb* gene cluster, suggesting that the whole *rfb* gene cluster is in one operon.

### 1.11.1.6 THE ATYPICAL G+C CONTENT

All of the DNA in the LT2 *rfb* gene cluster has a much lower G+C content than the usual G+C content for *S. enterica*. Most *S. enterica* genes have a G+C content of about 0.51, but the *rfb* genes range from 0.32 to 0.46 (Jiang *et al.*, 1991). To account for this atypical G+C content, it has been proposed that the LT2 *rfb* gene cluster was only recently transferred to *S. enterica* from a species with low G+C content (Jiang *et al.*, 1991). The presence in the *rfb* gene cluster of regions with different G+C content (Figure 1.7) suggests different histories for these regions.

## 1.11.2 *RFB* GENE CLUSTERS OF *S. ENTERICA* SEROVARS TYPHI AND PARATYPHI

The O antigens of *S. enterica* serovars typhi and paratyphi have the same mannosyl-rhamnosyl-galactosyl 'backbone' as serovar typhimurium but differ in the 3,6-dideoxyhexose linked to the mannose, being paratose in serovar paratyphi, tyvelose in serovar typhi and abequose in serovar typhimurium (Figure 1.3). The pathways for the synthesis of these group specific sugars differ only in their final steps: in group B strains, CDP-4-keto-3,6-dideoxyhexose is converted to CDP-abequose by abequose synthase (encoded by *rfbJ*), whereas paratose synthase (encoded by *rfbS*) of group A and D strains converts it to CDP-paratose (Matsuhashi *et al.*, 1966; Matsuhashi and Strominger, 1967). CDP-paratose is further converted into CDP-tyvelose by paratose epimerase (encoded by *rfbE*) in group D strains.

The *rfb* gene clusters of groups A and D have been cloned and sequenced in this lab (Verma *et al.*, 1988; Verma and Reeves, 1989;

Liu *et al.*, 1991). The major difference between the *rfb* gene cluster of group B and these two lies in a region which contains three genes (*rfbJ*, *orf12.8* and *orf14.1*) in group B, and in groups A and D three homologous genes plus one additional gene. *rfbS* encoding the paratose synthase, which carries out a step parallel to that of abequose synthase (encoded by *rfbJ*), is one of the three homologous genes in groups A and D. Comparison of *rfbS* and *rfbJ* shows that although they are related, there has been substantial divergence; thus they may have a common ancestor but diverged a very long time ago (Verma and Reeves, 1989). *rfbE* is the additional gene in groups A and D, and encodes paratose epimerase which converts CDP-paratose to CDP-tyvelose: this gene is inactive in group A being in a mutant form (Verma and Reeves, 1989). The remaining genes in the *rfb* gene cluster of these three strains appear to be essentially identical. There is also a triplication of the sequence from position 14.4 to position 17.2 in some group A strains, but this probably has no effect on function.

## 1.12 AIMS OF THIS THESIS

As mentioned above, the O antigen of *S. enterica* is highly polymorphic. It is on the surface and appears to be a target of intense selection by the immune system and this probably accounts for the origin and maintenance of the high level of variation. The aims of the work in this thesis were to determine the basis of the variation between O antigens of group E1 and groups A, B and D at DNA level, in an attempt to better understand how bacteria adjust to change in their environment.

Group E1 is interesting because it represents the other group E strains. As mentioned above, the variations among groups E1, E2, E3 and E4 are due to phage conversion or the expression of the *rafC* gene. Thus, all group E strains probably are the same in the *rfb* region. Also, the study of group E1 would complete the study of the four groups (A, B, D, and E) which all have a trisaccharide (mannosyl-rhamnosyl-galactose) as the backbone in O units.

## 1.13 TAXONOMY OF *SALMONELLA*

O antigen variation was one of the criteria used to divide *Salmonella* into different species. Now, *Salmonella* isolates are considered to belong to a single species, *Salmonella enterica*, because of the high degree of nucleic acid relatedness among them as shown by DNA/DNA hybridization (Le Minor and Popoff, 1987; Brenner, 1984; Crosa *et al.*, 1973). Within the species, isolates are grouped into group A, group B etc. on the basis of their O antigen specificity, each group then being divided into serovars using other antigens. In this thesis the old 'species' names such as "typhimurium" or "anatum" are used as serovar names as recommended by Le Minor and Popoff (1987).

*S. enterica* serovars typhimurium (strain LT2), typhi (strain Ty21a), paratyphi (strain A IMVS1316) and anatum (strain M32) are used as representatives of groups B, D, A and E1 respectively. For the sake of brevity, we may use only the group name instead of the whole name in this thesis such as group B instead of *S. enterica* serovar typhimurium (strain LT2).

# CHAPTER TWO: MATERIALS AND METHODS

## 2.1 BACTERIAL STRAINS, BACTERIOPHAGES AND PLASMIDS

The bacterial strains, bacteriophages and plasmids used are listed in Tables 2.1 and 2.2.

## 2.2 REAGENTS

Restriction enzymes, DNA polymerase I, calf intestinal alkaline phosphatase, RNase and T4 DNA ligase were purchased from Boehringer Mannheim Biochemicals, Indianapolis, Ind.; the incubation buffer set for restriction enzymes was also from Boehringer Mannheim; exonuclease III (Exo III), S1 nuclease and Klenow DNA polymerase were from Promega. Dye-labelled M13 standard primer and dye-labelled dideoxy terminator were from Promega; specific primers for dye dideoxy termination sequencing were synthesized on an Applied Biosystems DNA synthesizer; deoxy-nucleotides, dideoxy-nucleotides and Taq DNA polymerase were purchased from Pharmacia LKB Biotechnology; an Applied Biosystems 370A DNA sequencer was from Applied Biosystems. Pro-A-Gene matrix was from BIO-RAD. UDP-galactose, GDP-mannose and decaprenol phosphate were purchased from Sigma. $\alpha$-[$P^{32}$] dCTP was from BRESA; deoxyadenosine 5'-$\alpha$-[$^{35}S$]thiotriphosphate, triethylammonium salt, UDP-[$^{14}C$]galactose, [$^{3}H$]galactose and GDP-

## Table 2.1 Bacterial strains

| Strain | Lab. stock no. | Characteristics | Reference |
|---|---|---|---|
| *Salmonella* | | | |
| M32 | | Serovar anatum, group E1 | IMVS[a] |
| *his-388* | P9029 | Serovar typhimurium LT2. Δ(*hisD-metG*) | Nikaido *et al.* (1967) |
| *his-642* | P9073 | Serovar typhimurium LT2. Δ(*hisI-rfbK*) | Nikaido *et al.* (1967) |
| SL1196 | P9285 | Serovar typhimurium LT2. *hisD27, trpA8 (ColE1-30), H1-a, H2-e,n, xgal-437, xyl-412, rfbT* | Wilkinson *et al.* (1972) |
| SL1197 | P9286 | Serovar typhimurium LT2. *hisD27, trpA8 (ColE1-30), H1-a, H2-e,n, xgal-437, xyl-412, rfbT* | Wilkinson *et al.* (1972) |
| SL1198 | P9287 | Serovar typhimurium LT2. *hisD27, trpA8 (ColE1-30), H1-a, H2-e,n, xgal-437, xyl-412, rfbT* | Wilkinson *et al.* (1972) |

[a], Institute of Medical and Veterinary Science, Adelaide, Australia

| | | | |
|---|---|---|---|
| SL1654 | P9003 | Serovar typhimurium LT2. Group B. *hsdL, trpB2, nml, flaA66, rpsl, xylT404, ilv452, metE554, metA22, hsdA* | Ornellas and Stocker (1974) |
| SL3848 | P9300 | Serovar typhimurium LT2. *rfaL447, gal-851* | Roantree *et al.* (1977) |
| SGSC647 | P9138 | Serovar typhimurium LT2. *Zbi 812::Tn10, galE* | Sanderson and Roth (1983) |
| TT10251 | P9306 | Serovar typhimurium LT2. *pmi::Mud1-8* | Hughes and Roth (1984) |
| *his-515* | P9031 | Serovar typhimurium LT2. Δ(*flaA-rfbH*) | Nikaido *et al.* (1967) |
| | P9317 | As P9286, *pmi::Mud1-8* | This study |
| | P9318 | As P9073, *Zbi 812::Tn10, galE* | This study |
| | P9320 | As P9285, *Zbi 812::Tn10, galE* | This study |
| | P9321 | As P9286, *Zbi 812::Tn10, galE* | This study |
| | P9322 | As P9300, *Zbi 812::Tn10, galE* | This study |
| | P9324 | P9285 carrying pPR1381 | This study |
| | P9325 | P9286 carrying pPR1381 | This study |

*E. coli* K12

| | | | |
|---|---|---|---|
| JM101 | P2398 | *recA1, endA1, thi, gyrA96, hsdR17, supE44, relA1, Δ(lac-proAB), (F' episome traD36, proAB, lacI$^q$, lacΔM15)* | Yanisch-Perron *et al.* (1985) |
| x2819 | P3851 | *F$^-$, lacY1, glnV44, galK2, galT22, l$^-$, ΔthyA57, metB1, hsdR2 lysogenized with lcI1857, b2, redb3, S7* | Jacobs *et al.* (1986) |

**Table 2.2 Plasmids and phages**

| Plasmids or phages | Characteristics | Reference or source |
|---|---|---|
| Cloning vectors | | |
| pGB2 | Replicon: SC101<br>Genetic marks: $str^r$, $spe^r$ | Churchward *et al.* (1984) |
| pPR637 | Replicon: SC101<br>Genetic marks: $str^r$, $spe^r$ | Brahmbhatt (1987) |
| pUC8 and pUC9 | Replicon: pMB1<br>Genetic mark: $amp^r$ | Viera and Messing (1982) |
| pUC18 and pUC19 | Replicon: pMB1<br>Genetic mark: $amp^r$ | Norrander *et al.* (1983) |
| pGEM-7ZF(+) | Replicon: pMB1<br>Genetic mark: $amp^r$ | Promega Corporation |
| pT7T3 19U | Replicon: pMB1<br>Genetic mark: $amp^r$ | Pharmacia LKB |
| Plasmids | | |
| pADE206 | $amp^r$; vector, pUC18; insert, *rfc* gene | Collins and Hackett (1991) |
| pSUP::Tn5-B50 | $cml^r$, $tet^r$; vector, pSUP102-Gm; the donor of Tn5-B50 | Simon *et al.* (1989) |
| pPR333 | $amp^r$, $cml^r$; vector, pUC7; insert, 3.27kb carring $cml^r$ gene from Tn1725 | Jiang *et al.* (1991) |

| | | |
|---|---|---|
| pPR1009 | kan$^r$, str$^r$, spe$^r$; vector, pPR691; insert, LT2 *rfb* DNA from positions -10 to 24 | Jiang *et al.* (1991) |
| pPR300 | amp$^r$; vector, pJRD158; insert, LT2 *rfb* DNA from positons 8.4 to 11.4 | Brahmbhatt *et al.* (1988) |
| pPR301 | amp$^r$; vector, pEMBL8; insert, LT2 *rfb* DNA from positons 11.4 to 21.3 | Brahmbhatt *et al.* (1988) |
| pPR341 | cml$^r$; vector, pPR328; insert, LT2 *rfb* DNA from positons 5.3 to 11.4 | Brahmbhatt *et al.* (1988) |
| pPR465 | amp$^r$; vector, pUC9; insert, LT2 *rfb* DNA from positons -1.8 to 5.3 | Brahmbhatt *et al.* (1988) |
| pPR472 | amp$^r$; vector, pUC9; insert, LT2 *rfb* DNA from positons 21.3 to 24.7 | Brahmbhatt *et al.* (1988) |
| pPR686 | str$^r$, spe$^r$; vector, pPR637; insert, LT2 *rfb* DNA from positions 17.7 to 23.0 | Jiang *et al.* (1991) |
| pPR964 | amp$^r$; vector, pUC18; insert, M32 *rfb* DNA from positions 15.7 to position 31.2 | This study |
| pPR966 | amp$^r$; vector, pUC19; insert, M32 *rfb* DNA from positions 9.7 to 16.8 | This study |
| pPR1029 | amp$^r$; vector, pUC8; insert, M32 *rfb* DNA from positions 0.0 to 9.7 | This study |

| | | |
|---|---|---|
| pPR1030 | amp$^r$; vector, pUC8; insert, M32 *rfb* DNA from positions 7.2 to 11.1 | This study |
| pPR1031 | amp$^r$; vector, pUC8; insert, M32 *rfb* DNA from positions -1.7 to 5.3 | This study |
| pPR1033 | amp$^r$; vector, pUC18; insert, M32 *rfb* DNA from positions 7.2 to 11.1, in right orientation for expression | This study |
| pPR1034 | amp$^r$; vector, pUC19; insert, M32 *rfb* DNA from positions 7.2 to 11.1, in right orientation for expression | This study |
| pPR1135 | amp$^r$; vector, pT7T3 19U; insert, M32 *rfb* DNA from positions 11.2 to 13.6, in right orientation for expression | This study |
| pPR1136 | amp$^r$; vector, pT7T3 19U; insert, M32 *rfb* DNA from positions 11.2 to 13.6, in opposite orientation for expression | This study |
| pPR1153 | amp$^r$; vector, pT7T3 19U; insert, M32 *rfb* DNA from positions 5.3 to 7.2, in right orientation for expression | This study |
| pPR1154 | amp$^r$; vector, pT7T3 19U; insert, M32 *rfb* DNA from positions 11.2 to 13.6, in opposite orientation for expression | This study |

| | | |
|---|---|---|
| pPR1169 | amp$^r$; vector, pGEM-7ZF(+); insert, M32 *rfb* DNA from positions 15.7 to 19.1 | This study |
| pPR1173 | amp$^r$; vector, pGEM-7ZF(+); insert, M32 *rfb* DNA from positions 15.7 to 19.1, in right orientation for expression | This study |
| pPR1381 | amp$^r$; vector, pGEM-7ZF(+); insert, LT2 *rfb* DNA from positions 19.7 to 21.7, in wrong orientation for expression | This study |

Bacteriophages

| | |
|---|---|
| M13Mp18 and M13Mp19 | Norrander *et al.* (1983) |
| M13KO7 | Vieira and Messing (1987) |
| P22 | Zinder (1953) |
| Felix anti-O phage | Felix and Callow (1943) |
| λb221cI857Pam80 ::Tn5-B50 | Simon *et al.* (1989) |

[$^{14}$C]mannose were from Amersham; [$^{14}$C]mannose was purchased from Sigma; TDP-rhamnose and TDP-[$^{14}$C]rhamnose were kindly provided by Dr. A.A. Lindberg; other radiochemicals were from BRESATEC LIMITED. LGT and HGT agarose from Seakem; nitrocellulose from Schleicher; X-ray film from Kodak. *S. enterica* antisera 04 and 05 were from Wellcome Diagnostics. All other chemicals were of analytical grade and were obtained from Sigma Chemical Co. and Ajax Chemicals Pty. Ltd, Sydney.

## 2.3 GROWTH MEDIA

Nutrient broth (NB) (Difco 0003, 16g/L; NaCl, 5g/L) was used for growth of bacteria in broth cultures. Nutrient agar was obtained by addition of agar (Difco 0045, 16g/L) to nutrient broth. Soft agar contained equal volumes of nutrient broth and nutrient agar.

Minimal medium was prepared as described by Davis and Mingoli (1950), and was solidified with the addition of agar (Difco 0140-01, 16g/L). Carbon sources for minimal medium were glucose (5mg/ml) or maltose (10mg/ml). Broth 2xYT (Sanger *et al.*, 1980) was used in propagating phage M13, phagemids pT7T3 and pGEM, and their derivatives.

Antibiotics were added as needed to media at the following concentrations: ampicillin, 25μg/ml; chloramphenicol, 25μg/ml; kanamycin, 50μg/ml; spectinomycin, 10μg/ml; streptomycin, 20μg/ml; tetracycline, 16μg/ml.

## 2.4 DNA TECHNIQUES

### 2.4.1 DNA ISOLATION - CHROMOSOMAL DNA

Chromosomal DNA was prepared by the modified method of Ardeshir *et al.* (1981): the bacterial cells are lysed by treated with lysozyme and followed EDTA/proteinase K (50 μg/ml), the lysed mixture is sequentially extracted by phenol and either and then dialysed against 1xTE.

### 2.4.2 DNA ISOLATION - PLASMID

Small-scale preparation of plasmid DNA was carried out using the alkaline lysis method which is a modification of the methods of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981). Cells were harvested from 1.5 ml of a overnight culture and lysed by alkali, extracted with phenol:chloroform and ether, and the double-stranded DNA is precipitated with 2.5 volumes of cold ethanol.

Large-scale preparation of plasmid DNA was carried out by lysing the cells with alkali and purifying the DNA using the two-step CsCl gradient method of Garger *et al.* (1983).

### 2.4.3 DNA ISOLATION - SINGLE-STRANDED DNA

Single-stranded DNA was prepared using the method of Sanger *et al.* (1980). M13 and its derivatives were propagated on *E. coli* K12 strain JM101. Phagmids and their derivatives were packaged as single-stranded DNA (ssDNA) on *E. coli* K12 NM522 during super-infection with helper phage M13K07 (Vieira and Messing, 1987).

## 2.4.4 RESTRICTION ENDONULEASE DIGESTION

Restriction digestions of DNA were carried out in the appropriate buffer as suggested by the manufacturer supplying the enzyme. Usually 0.2-1 μg of plasmid DNA was digested with 2 units of enzyme in a total volume of 20 μl at 37°C for 1.5 hours. 1-2 μg chromosomal DNA was digested with 20 units of enzyme at 37°C for 2 hours in a total volume of 20 μl. When necessary, digestions were terminated either by heating the digestion mixture at 65 °C for 15 minutes or by extracting the mixture with phenol, ether, and ethanol precipitation if heat stable enzymes were used.

## 2.4.5 AGAROSE GEL ELECTROPHORESIS

0.7-1.0% agarose gels in 0.5xTBE buffer were used to analyse restriction digested or undigested DNA samples. Gels were stained with ethidium bromide (0.5-1.0 mg/ml) and the bands were visualized on a UV transilluminator. DNA fragments of phage SPP-1 digested with *Eco*RI was used as DNA size markers.

## 2.4.6 PURIFICATION OF DNA FRAGMENTS FROM AGAROSE GELS

The DNA band of interest was located using a long-wavelength UV lamp, and the slice of agarose containing the band of interest was cut out. Prep-A-Gene DNA Purification Kit (from BIO-RAD) was used to recover DNA from the slice of agarose gel using the method provided by BIO-RAD.

### 2.4.7 DNA QUANTITATION

DNA was quantitated by measuring the enhancement of fluorescence of bis-benzimida-zole (Hoechst 33258) on binding to DNA using the TKO 100 - Dedicated Mini Fluorometer and the protocol provided by Hoefer Scientific Instruments.

### 2.4.8 DEPHOSPHORYLATION OF LINEARIZED PLASMID DNA

5'-phosphate groups of linear DNAs were removed using calf intestinal alkaline phosphatase (CIP) and the method described by Maniatis *et al.* (1982).

### 2.4.9 LIGATION OF DNA

Ligation reaction of DNA with cohesive ends was carried out in a total volume of 20 µl containing 0.3-1 µg DNA (with a ratio of approximately 1:3 of vector to insert), 20mM Tris-Cl(pH 7.6), 5 mM MgCl2, 5 mM dithiothreitol, 50 µg/ml bovine serum albumin, and 1 unit T4 DNA ligase. This reaction mixture was incubated at 4 $^{o}$C for overnight, and used directly for the transformation. Ligation of blunt-ended DNA was performed under similar conditions as ligation of cohesive ended DNA except that PEG 8000 (at a concentration of 5%) and more ligase (10 units) were included in the ligation mixture.

### 2.4.10 TRANSFORMATION

Electrotransformation method was used to introduce DNA into bacterial cells. The bacterial cells for electroporation were prepared using the method provided by BIO-RAD: 100 ml bacteria were grown

to mid-log phase, chilled, centrifuged, washed twice with 0.9% NaCl and once with 10% glycerol, and then resuspended in 1 ml of 10% glycerol. The electroporation was carried out with a mixture of 1 µl ligation mixture and 100 µl cell suspension using the Pulse Controller (BIO-RAD) and the method provided by BIO-RAD.

### 2.4.11 NICK TRANSLATION

Nick translation of plasmid DNA or DNA fragments extracted from agarose gels was performed by the method described by Maniatis *et al.* (1982).

### 2.4.12 COLONY HYBRIDIZATION

Colony hybridization was carried out by the method described by Maniatis *et al.* (1982).

### 2.4.13 SOUTHERN HYBRIDIZATION

Southern transfer and hybridization of DNA immobilized on filters to radioactive probes was performed by the methods described by Maniatis *et al.* (1982). The high and low stringency hybridization experiments were as described by Howely *et al.* (1979).

### 2.4.14 DNA SEQUENCING

All sequencing was by the chain termination technique of Sanger *et al.* (1977): most was done using nested sets of deletions. Custom made oligonucleotides complementary to sequenced *rfb* DNA were used as primers for filling in a few gaps and for the sequencing of one small

region.

Nested sets of deletions were made using exonuclease III and the procedure described by Henikoff (1984), and pT7T3-19U, pGEM-7zf(+), pUC18 or pUC19 as the vector. pT7T3, pGEM and their derivatives were packaged as single-stranded DNA (ssDNA) during super-infection with helper phage M13K07 (Vieira and Messing, 1987). The inserts in pUC18 and pUC19 were transferred into M13mp18 or M13mp19, and ssDNA was made from these clones. In both cases the ssDNA was sequenced by the dideoxy method, in general using fluorescent dye-labelled M13 standard primer and running the reaction mixture on an ABI 370A sequencer. The small amount of sequencing using custom made primers was done with $^{35}S$ labelled dATP, and the gels were read after autoradiography.

## 2.5 BACTERIOPHAGE PROPAGATION

Bacteriophages were propagated by mixing $5x10^3$ bacteriophages and 100 µl of a fresh log phase bacterial culture, and this mixture was left standing for 30 min at 37°C and then poured on the top of a NA plate after adding 4 ml of soft agar (45 - 49°C). Plates were incubated upright at 37°C overnight to obtain the complete lysis, and left standing at 4°C overnight after 6 ml SM buffer was poured onto the agar. Bacteria were removed from lysates by filtration (0.2 mm). Propagation of λbz21cI857Pam80 was carried out in an $Su^+$ host, such as *E. coli* C600.

## 2.6 BACTERIOPHAGE SENSITIVITY TEST

The cross streak method was used. Bacteriophage lysate was first

streaked across a NA plate, and a loop of bacterial culture was then streaked across after the phage streak dried completely. The plate was then incubated at 37°C. After 8-12 hours, there should be a clearing on the cross if the bacteria are sensitive to the phage.

## 2.7 TRANSDUCTION

Target bacterial cells were prepared by centrifugation and resuspension in 1/10 volume of NB. 0.1 ml of cells was mixed with phage lysate of the same volume, and incubated standing at 37°C for 30 min. Cells were resuspended in 0.1 ml NB after being washed twice with 0.9% NaCl, and then spreaded on the selection plates.

## 2.8 CONJUGATION

Log phase recipient and donor cultures were mixed together (1 ml of each) and incubated standing for 1 hour at 37°C. Cells were pelleted and resuspended in 1 ml of NB before being spreaded on the selection plates.

## 2.9 IN VIVO PACKAGING OF COSMID

Cosmids were packaged into bacteriophage λ either by infecting the cells containing cosmids with λ phage or by using a lysogen of *E. coli* strain (P3851) which contains an integrated prophage that is defective in normal excision function and the procedure described by Jacobs *et al.* (1986).

## 2.10 TRANSPOSON MUTAGENESIS

Insertion mutagenesis of cosmid was performed using Tn5 and Tn1725, and the method as follows: the plasmid carrying the transposon was moved into the *in vivo* packaging strain X2819 (P3851) (Jacobs *et al.*, 1986) containing the target cosmid either by conjugation or transformation. A few hundred colonies which contained both the donor plasmid and the target cosmid were incubated in a 10 ml NB for overnight at 30°C. A 1:20 dilution into 20 ml NB was made, and the cosmid was packaged in vivo by heat induction using the procedure described by Jacobs *et al.* (1986). Packaged cosmid was transduced into an *E. coli* strain, and restriction mapping was done with each cosmid to locate the insertion.

Insertion mutagenesis of plasmid was performed in a Su° *E. coli* strain, P3127, using λbz21cI857Pam80::Tn5-B60: 10 ml late log phase culture of P3127 containing target plasmid was pelleted and resuspended in 2 ml SM buffer before 2 ml λ::Tn5-B50 lysate was added. The mixture was incubated standing at 37°C for 30 min, and then 10 ml NB was added before being spread on plates containing kanamycin to select for strains carrying Tn5.

## 2.11 MEMBRANE PREPARATION

Whole membrane samples for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) were prepared from 10 ml cultures which were pelleted and washed with 10 ml of 30 mM Tris-HCl, pH 8.0. Following the addition of 20 µl of lysozyme (2 mg/ml) in 0.1 M EDTA, the mixture was incubated for 30 min on ice before being frozen at -70°C. After 30 min, the mixture was thawed

and 3 ml of 3mM EDTA was added before sonication four times for 15 seconds on ice. Cellular debris was removed by centrifugation at 2.5 x $10^3$ g for 15 min and the supernatant centrifuged at 2.6 x $10^4$ g for 1 hour. The resulting pellet was resuspended in 50 µl of sample buffer (Lugtenberg *et al.*, 1975); 10 µl of this preparation was boiled for 5 min and 1 µl of proteinase K (200 µg/ml) added before loading onto the gel.

Membrane fraction for enzyme assay and *in vitro* synthesis of oligosaccharide was prepared from 100 ml of culture which was grown for 3 to 4 h (or to OD A530=0.4) after 1:10 inoculation. The cell pellet was washed once with 0.9% NaCl solution, suspended in 3ml of buffer A (50mM Tris-acetate, pH 8.5 and 1mM EDTA), and kept at -20 °C for 30 min, then sonicated 3 times for 15 sec periods with 1-2 min cooling between bursts. The sonicated sample was centrifuged at $3.0x10^3g$ for 10 min to remove unbroken cells, and the turbid supernatant fraction was centrifuged at $4.8x10^4g$ for 1 h. To wash the membrane, the pellet was resuspended in the same volume of buffer A, again centrifuged at $4.8x10^4g$ for 1 h, and resuspended in 0.3 to 0.5ml of buffer A.

## 2.12 ASSAY OF GALACTOSE TRANSFERASE

Galactosyl transferase (encoded by *rfbP*) catalyses the first reaction of the O antigen biosynthesis (Figure 1.5). When UDP-[$^{14}$C]galactose is used, radioactivity will be transferred to the hydrophobic undecaprenyl phosphate and becomes extractable by appropriate solvents.

The transfer of galactose from UDP-[$^{14}$C]galactose to undecaprenyl

phosphate was measured by the modified method of Osborn *et al.* (1972b). 4µl freshly diluted undecaprenyl phosphate (4x diluted in a mixture of chloroform and methanol (C: M=3:1) of original stock, final concentration is 6 nmol.) were added to each reaction tube, then dried under vacuum. Undecaprenyl phosphate was then resuspended in 10µl methanol and 15µl 0.15% Triton X100, and then 12.5µl of 1M Tris-acetate (pH 8.5), 5µl of 0.5M MgCl2, 5µl of 50mM EDTA, 20µl of envelope fraction (1 to 10 mg protein ), 0.33 µM (6µl) of UDP-[$^{14}$C]galactose (20x diluted in $H_2O$, 15,000 cpm) were added. MQ water was then added to make the total volume to 250 µl. After incubation at 37°C for 15min, the reaction was stopped by addition of $CHCl_3$-methanol (C:M, 2:1). This mixture was transferred into a scintillation vial, vortexed briefly to obtain a single phase, and kept at room temperature for 20 min. 0.8ml of PSUP (15 ml chloroform, 240 ml methanol, 1.83g KCl and 235 ml water) was added to the vial, vortexed for 20 sec, centrifuged for 2 min, and the top aqueous phase was carefully sucked out and discarded. The C: M phase was washed 3x with PSUP, dried down (under 150W light), resuspended in 4 ml of scintillant LIPO LUMA, and counted in a liquid scintillation counter

## 2.13 SLIDE AGGLUTINATION

O antigen epitopes on cells was detected by slide agglutination as described by Leinonen (1985). A small amount of bacterial culture was mixed with a drop of a suitable dilution of antiserum on a glass slide. The slide was tilted up and down for 2 minutes and examined against a dark background. As a control, 4% NaCl instead of antiserum was used to detect auto-agglutination.

## 2.14 COMPUTER ANALYSIS OF SEQUENCE DATA

Sequence data were analysed using the MBIS System (Bucholz and Reisner, 1986) at the CSIRO Biotechnology Laboratory, Sydney and the Australian National Genomic Information Service (ANGIS) at Sydney University: both incorporate several sets of programs. Sequence data was assembled using the SAP program of Staden (Staden, 1982, 1986); molecular weights and hydrophobicity indices were calculated by the program NIP (Staden, 1982); the programs PepPlot of GCG (Devereux *et al.*, 1984) and CHOU (Chou and Fasman, 1974) were used to do the secondary structure prediction on a amino acid sequence; the program ALOM based on the method of Klein *et al.* (1985) was used to identify potential transmembrane segments, which were then further checked using the algorithm described by Eisenberg *et al.* (1984). The program BESTFIT in the GCG package, which uses the "local homology" algorithm of Smith and Waterman (1981), and the program MTX (Reisner and Bucholtz, 1986), were used to do sequence comparisons. The percentages of identity and similarity given below are derived using BESTFIT and the parameters: gap weight, 3.0; gap length weight, 0.10; similarity levels of up to 40% are found for unrelated proteins: levels over this were taken to be significant. The program SHUFFLE (GCG package) which randomizes the order of the symbols in a sequence was used to generate shuffle sequence. The programs MATCH and STATSEARCH based on the method of Wilbur and Lipman (1983) were used to search databanks for sequences with homology to the open reading frames (ORFs) found in our sequence. Promoter search programs were generously provided by Dr O'Neill: AUTOST22 used the criteria published by O'Neill and Chiafari (1989), ALIGNIC2

searched for promoters on the basis of information content as described by O'Neill (1989a), and MOLBvM13 used the information content Berg-von Hippel function described by O'Neill (1989b). The program VONHIPIC was used to calculate the index for each presumptive promoter, as described by O'Neill (1989b). A training back-propagation neural networks (O'Neill, 1991) to define the promoters was run for us by Dr. O'Neill on his computer.

## 2.15 METHODS USED IN THE STUDY OF *RFBP(T)* MUTANTS

### 2.15.1 LABELLING OF O ANTIGEN WITH [$^3$H]GALACTOSE

1 ml over night culture (P9320, P9321, P9322, P9318 and P9138) was inoculated in 10 ml NB, and grown for 3 hours at 37°C to a density of approximately 3 x $10^7$/ml. A mix of 1 nmol [3H]galactose (about 23,265,840 cpm) and 4 nmol cold galactose was then added, and the incubation was continued for 1 min. Cells were collected by centrifugation after the culture was cooled on ice.

For label and chase, cells were labelled for 1 min as above before a 10000-fold excess of cold galactose was added to the culture, and incubation was continued for another 10 min at 37°C. Cells were then cooled on ice and collected by centrifugation.

### 2.15.2 LABELLING OF O ANTIGEN WITH [$^{14}$C]MANNOSE

1 ml over night culture (P9306 or P9317) was inoculated in 10 ml NB with 1% mannose, and grown for 3.5 hr at 30°C. Cells were collected by centrifugation, and washed twice with NB before they were

resuspended in 10 ml NB and grown for 20 min at 37°C. 0.01µmol $^{14}$C-mannose was then added, and the incubation was continued for another 10 min. Labelled cells were collected by centrifugation after the culture was cooled on ice.

For label and chase, cells were labelled as above, and washed twice with NB before being resuspended in 10 ml NB and shaken for 20 min at 37°C. 1% mannose was added and the incubation continued for 12 min before the culture was placed on ice.

## 2.15.3 PURIFICATION OF THE LABELLED LPS AND O ANTIGEN

LPS and O antigen precursor were extracted by the method of McGrath and Osborn (1991a): Labelled cells were first washed twice with 2 ml of cold 50% acetone. Acetone-washed cells were resuspended in 1 ml water and extracted twice with 45% phenol at 70°C for 15 min each. LPS was precipitated from the aqueous phase with cold 65% ethanol-10mM sodium acetate. Free O antigen which had been cleaved from undecaprenyl pyrophosphate by hot phenol remained in the ethanol supernatant. The ethanol supernatant was freeze dried before resuspended in water.

## 2.15.4 ANALYTICAL PROCEDURES

Gel filtration was done with a Sephadex G-15 column (1.5 x 40 cm) using 50 mM ammonium acetate (pH 7.0) as packing and elution buffer. Samples were applied in 100 µl 10% CsCl, and elution rate was 60µl per min. Fractions of 0.5 ml were collected and radioactivity was counted in 5 ml of ASC II (Amersham) in a MINAXI liquid

scintillation counter.

SDS-polyacrylamide gel electrophoresis and silver staining of LPS for visualizing LPS was carried out as described by Brown *et al.* (1992). SDS-polyacrylamide gel with radioactive LPS was dried, and the distribution of radioactivity in the gel was quantitated using a Molecular Dynamics Phospho-Imager 400B. Subsequent data analysis employed the Molecular Dynamics Image Quant Computer Software package.

### 2.15.5 *IN VITRO* SYNTHESIS OF OLIGOSACCHARIDE

[$^{14}$C-Rha]-Gal-lipid, [$^{14}$C-Man]-Rha-Gal-lipid and [$^{14}$C-Rha]-Man-Man-Gal-lipid intermediates were synthesized *in vitro* using the whole cell envelope fractions of strains P9298, P9299 and P4706 as the enzymetic sources,respectively. Strains P9298 and P9299 contain the plasmids carrying strain LT2 *rfb* genes for galactose and rhamnose transferases, galactose, rhamnose and mannose transferases, respectively (D. Liu and P.R. Reeves, unpublished data). Strain P4706 contains a plasmid carrying all the transferase genes for synthesis of the single O unit (Rha-Man-Man-Gal) of strain M67 (D. Liu and P.R. Reeves, unpublished data). The preparation of the cell envelope fraction was described above (see Membrane Preparation, P45), and the synthesis was carried out by the method of Osborn *et al.* (1972b). The incubation mixture contained 6 nmole decaprenol phosphate (as the lipid-carrier), 50 mM Tris-acetate (pH 8.5), 10 mM $MgCl_2$, 1 mMEDTA, 20 µl membrane fraction and 0.33 mM [14C] nucleotide sugar in a final volume of 0.25 ml. Other cold nucleotide sugars were added to the reaction mixture as required to a

final concentration of 0.1 mM. The reaction was carried out at $37^0$C for 15 min, and the synthesized decaprenol linked oligosaccharide was extracted by 4 ml of $CHCl_3$:methanol (C:M, 2:1). In each case, the radioactive nucleotide sugar used was the last sugar added to the oligosaccharide by the transferase, and the radioactivity in the C:M phase must be derived from the decaprenol linked full length oligosaccharide. The oligosaccharides were cleaved from lipid-carrier by extraction with hot phenol (45% phenol, $70^o$C, 15 min.) three times.

# CHAPTER 3: CLONING AND PHYSICAL MAPPING OF THE *RFB* REGION FROM STRAIN M32

## 3.1 INTRODUCTION

The aim of the work discussed in this chapter was to clone the DNA covering the whole M32 *rfb* region, and compare it with the *rfb* region of strain LT2 (group B). When I started this work, the *rfb* region of strain LT2 had been cloned in this laboratory (Brahmbhatt *et al.*, 1988). A series of clones which together cover the whole LT2 *rfb* region were used as the probe for Southern hybridization against the M32 chromosome, and one of them was used as the probe to start the chromosomal walking procedure to clone the whole M32 *rfb* region. Restriction enzyme mapping of the M32 *rfb* region was then carried out. Comparison of the *rfb* regions of strains M32 and LT2 was carried out using the cloned M32 and LT2 DNA.

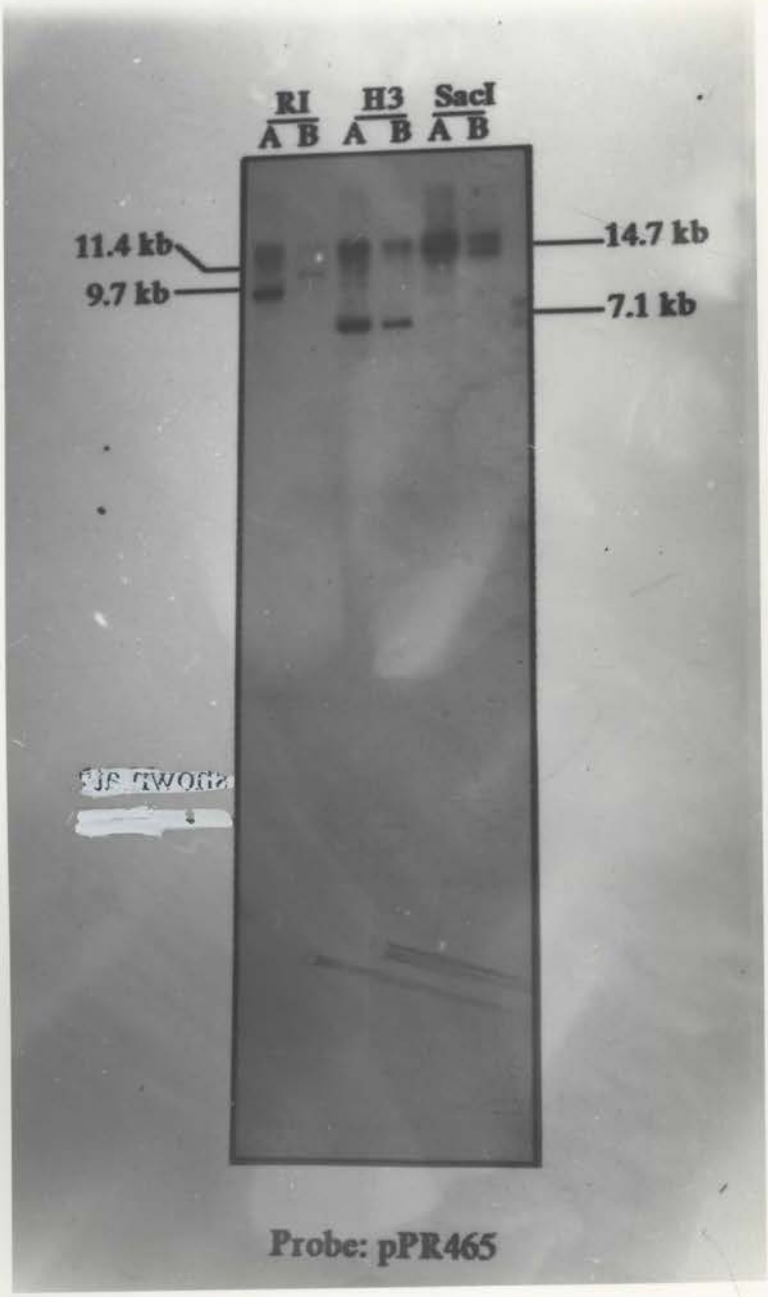
## 3.2 PRELIMINARY ANALYSIS OF SIMILARITY BETWEEN *RFB* REGIONS OF STRAINS M32 AND LT2

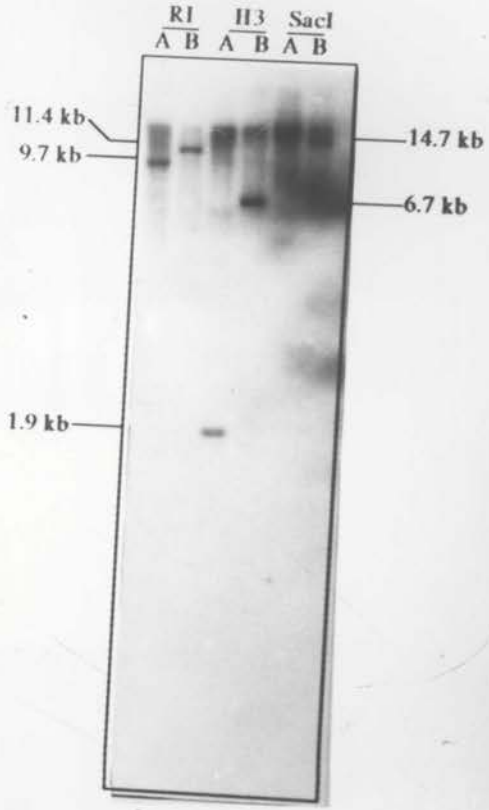
The O antigen of strain M32 has a similar structure to that of strain LT2, and we expected part of the LT2 *rfb* region would be homologous to strain M32 *rfb* DNA, allowing us to use the strain LT2 DNA as a probe to start cloning of the M32 *rfb* region. The inserts of the LT2 plasmids pPR300, pPR301 and pPR465 (see Figure 3.7) were used as radioactive probes against restriction enzyme digested

# FIGURE 3.1

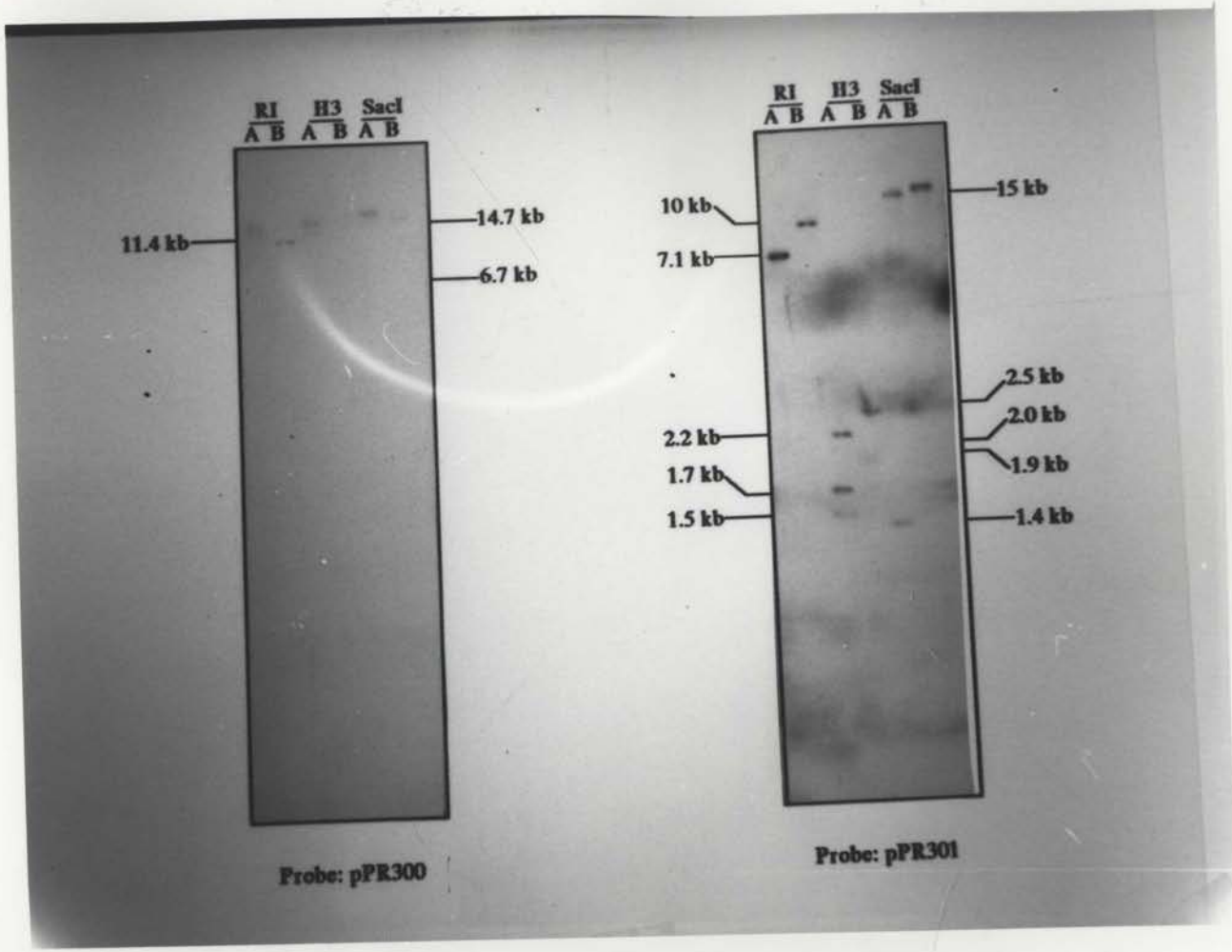
## Similarity between the *rfb* regions of strains M32 and LT2

Southern hybridization analysis of restriction enzyme digests of the M32 chromosomal DNA using strain LT2 plasmids pPR300, pPR301, pPR341 and pPR465 as probes. LT2 chromosomal DNA digested with same restriction enzymes were run in adjacent lanes. The probe used is indicated, and the lanes A and B are M32 chromosome and LT2 chromosome respectively. The restriction enzymes used are *Eco*RI (RI), *Hin*dIII (H3) and *Sac*I, and they are shown at top.


RI
A B
H3
A B
SacI
A B
11.4 kb
9.7 kb
14.7 kb
7.1 kb
Probe: pPR465




Probe: pPR341


RI
A B
H3
A B
SacI
A B
11.4 kb
14.7 kb
6.7 kb
Probe: pPR300
RI
A B
H3
A B
SacI
A B
10 kb
7.1 kb
2.2 kb
1.7 kb
1.5 kb
15 kb
2.5 kb
2.0 kb
1.9 kb
1.4 kb
Probe: pPR301

chromosomal DNA of strain M32. The inserts of pPR301 and pPR465 which cover the two ends of the LT2 *rfb* region hybridized strongly to M32 chromosomal DNA whereas the insert of pPR300 which covers the middle part of LT2 *rfb* region did not hybridize (Figure 3.1).

## 3.3 CLONING OF THE *RFB* REGION FROM STRAIN M32

The insert of pPR301 (Jiang *et al.*, 1991) hybridized to a 15kb *Sac*I fragment of M32 chromosomal DNA, and the strategy used consisted of cloning this 15kb *Sac*I piece from the chromosome of strain M32 using [$^{32}$P]-labelled pPR301 as probe in colony hybridization. This 15kb *Sac*I fragment was then used to start "chromosomal walking" to clone the whole *rfb* region of strain M32.

### 3.3.1 pPR964

M32 chromosomal DNA was cleaved completely with *Sac*I, and ligated with *Sac*I digested pUC18 DNA. The ligation mixture was transformed into *E.coli* K12 strain P3584. A total of 3000 clones were screened by colony hybridization using [$^{32}$P]-labelled pPR301 as probe. One clone, designated P4250, was obtained, and the plasmid contained in this clone was designated pPR964 (Figure 3.2).

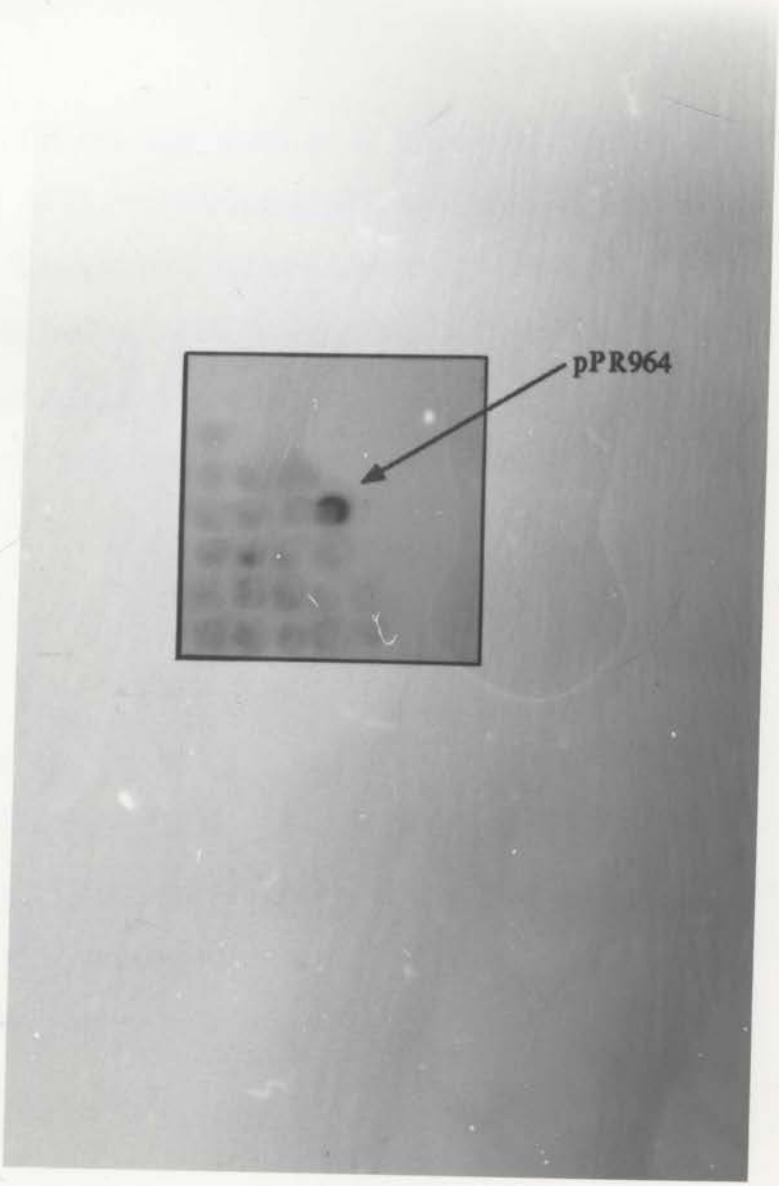
Restriction enzyme analysis of pPR964 showed that it contained 15kb DNA, and some of the important restriction sites are shown in Figure 3.3.

A Southern hybridization analysis was conducted to confirm that pPR964 contains *rfb* DNA. The inserts of pPR301 and pPR472 which cover the right end of strain LT2 *rfb* region and the *gnd-his* region

# FIGURE 3.2

## Identification of plasmid clone pPR964 by colony hybridization

Autoradiogram showing the identification of pPR964 from a *Sac*I bank of M32 chromosomal DNA in pUC18. The insert of pPR301 (Jiang *et al.*, 1991) from strain LT2 was used as probe.


pPR964

respectively (see Figure 3.7) (Jiang *et al.*, 1991) were used separately as probes against restriction enzyme digested pPR964. As shown in Figures 3.3, the left end of pPR964 hybridizes to pPR301 and the right end to pPR472. Thus pPR964 covers the region from the right end of *rfb* region through to the *gnd-his* region.

### 3.3.2 pPR966

The insert of pPR964 was used as a radioactive probe to screen a *EcoR*I bank of M32 chromosomal DNA in pUC19. Several colonies were obtained. Restriction enzyme analysis showed that one of these clones, pPR966, had a 7.1kb insert, and that the 1.1 kb *Sac*I/*Eco*RI fragment on the right end of this insert overlapped with the left end of pPR964 (Figure 3.4), with pPR964 extending further into the *rfb* region.

### 3.3.3 pPR1030

pPR1030 was obtained by screening a *Hin*dIII bank of M32 chromosomal DNA in pUC8 using the insert of pPR966 as radioactive probe. Restriction enzyme analysis showed that pPR1030 contained 3.9kb DNA, and it overlapped with pPR966 by a 1.4kb *Eco*RI/*Hin*dIII fragment which locates on the left end of pPR966 (Figure 3.4).

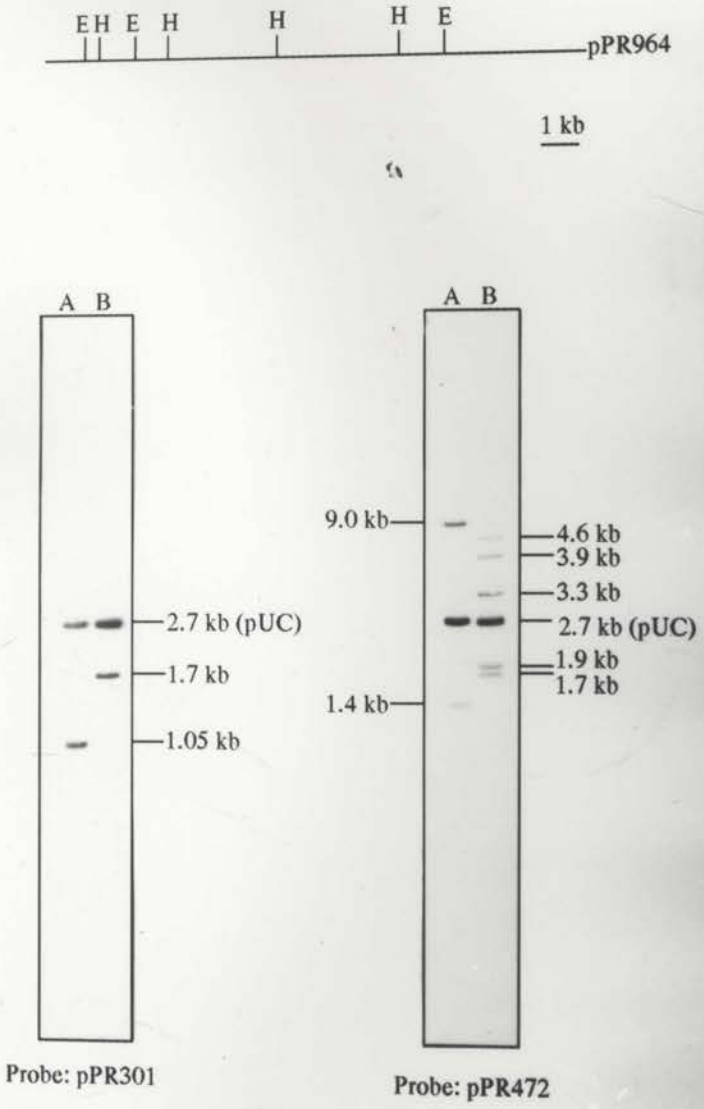
### 3.3.4 pPR1029

The chromosomal walking procedure was repeated to select pPR1029 from a *Eco*RI bank of M32 chromosomal DNA in pUC8 using the insert of pPR1030 as probe. Restriction analysis showed that pPR1029

# FIGURE 3.3

## Confirmation of plasmid pPR964

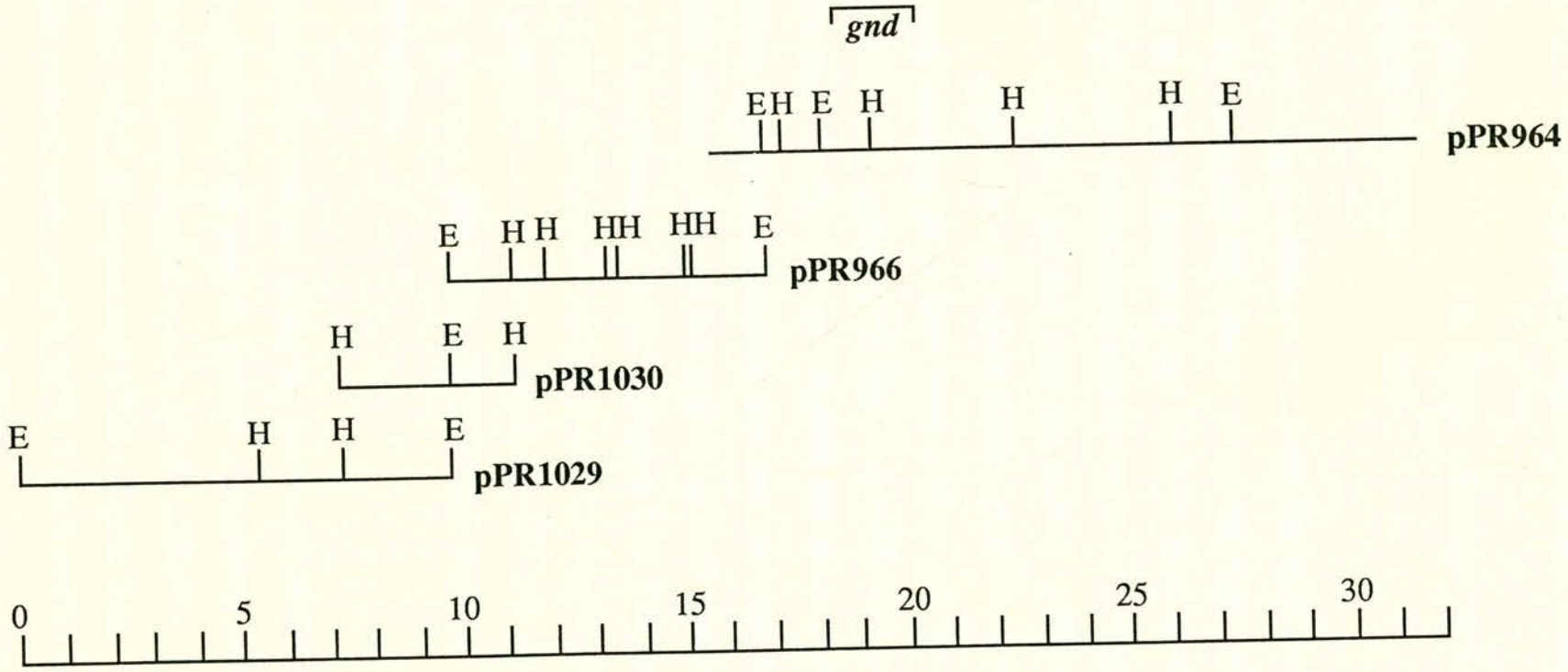
The *Eco*RI (E) and *Hin*dIII (H) restriction endonuclease map of pPR964 is shown at top. Below the map is the autoradiogram showing the Southern hybridization results of blotted *Eco*RI-*Sac*I (lane A) and *Hin*dIII-*Sac*I (lane B) restriction fragments of pPR964 probed with LT2 plasmids pPR301 and pPR472. Note that the 2.7 kb fragment present in all tracks represents vector:vector (pUC) hybridization.


E H E H H H E
pPR964
1 kb
A B
2.7 kb (pUC)
1.7 kb
1.05 kb
Probe: pPR301
A B
9.0 kb
1.4 kb
4.6 kb
3.9 kb
3.3 kb
2.7 kb (pUC)
1.9 kb
1.7 kb
Probe: pPR472

# FIGURE 3.4

## *Eco*RI and *Hin*dIII restriction maps of M32 plasmid clones obtained by chromosomal walking

*Eco*RI (E) and *Hin*dIII (H) restriction maps of pPR964, pPR966, pPR1030 and pPR1029. The approximate position of *gnd* are indicated at top. Map positions (kilobase pairs) are indicated on the lower part of the diagram.


gnd
E H H E H H H E
pPR964
E H H HH HH E
pPR966
H E H
pPR1030
E H H E
pPR1029
0 5 10 15 20 25 30

contained a 9.7kb *Eco*RI insert, and the 2.5kb *Hin*dIII/*Eco*RI fragment on its right end overlapped with the left end of pPR1030 (Figure 3.4).

### 3.3.5 pPR964, pPR966, pPR1030 AND pPR1029 TOGETHER COVER THE WHOLE M32 *RFB* REGION

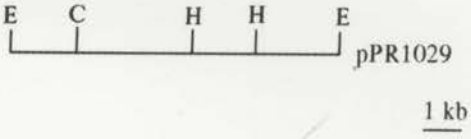
The chromosomal walking procedure starts from the right end of M32 *rfb* gene cluster by using pPR301 which covers the right end of LT2 *rfb* gene cluster as probe to screen a *Sac*I bank of M32 chromosomal DNA in pUC18. pPR964 was selected and shown to cover the right end region of *rfb* gene cluster and the *gnd-his* region. pPR966 which extends to the left of pPR964 was selected using pPR964 as probe, and this chromosomal walking procedure went further left by obtaining pPR1030 and pPR1029 using pPR966 and pPR1030 as probes respectively. To show if pPR1029 had reached the left end of *rfb* gene cluster, a Southern hybridization experiment was performed using the insert of pPR465 which covers the left end of LT2 *rfb* region as probe against the restriction digested pPR1029 (Figure 3.5). The fact that the insert of pPR1029 hybridized strongly to the insert of pPR465 indicates that pPR1029 covers the left end of M32 *rfb* gene cluster. Thus, pPR964, pPR966, pPR1030 and pPR1029 together cover the whole M32 *rfb* region.

Southern hybridization analysis using the inserts of pPR964, pPR966, pPR1030 and pPR1029 as probes against themselves and the M32 chromosomal DNA was carried out to check if the inserts of these clones were from M32 chromosomal DNA (Figure 3.6). The results showed that the general patterns of *Eco*RI, *Hin*dIII and *Sac*I sites were the same on both plasmids and the chromosome, and hence the inserts of these clones were from M32 chromosomal DNA and there had been

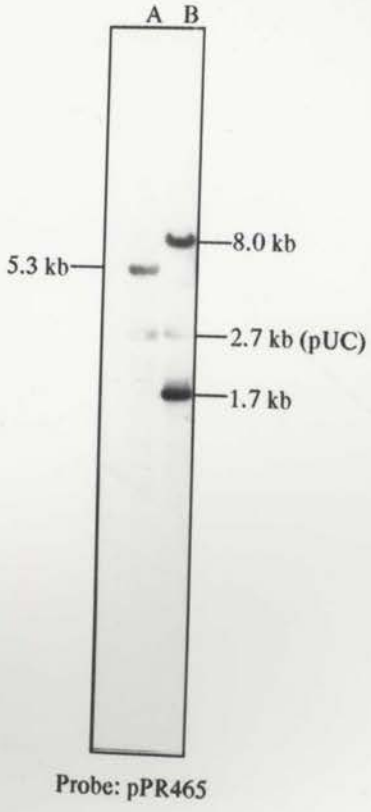
# FIGURE 3.5

## Location of the left end of the M32 *rfb* region

Southern hybridization analysis of restriction enzyme digests of pPR1029 using LT2 plasmid pPR465 as probe. The *Eco*RI (E), *Hin*dIII (H) and *Cla*I (C) restriction map of pPR1029 is shown at top. Lanes designation are as follows: A, pPR1029 doubly digested with *Eco*RI and *Hin*dIII; B, pPR1029 doubly digested with *Eco*RI and *Cla*I.




1 kb



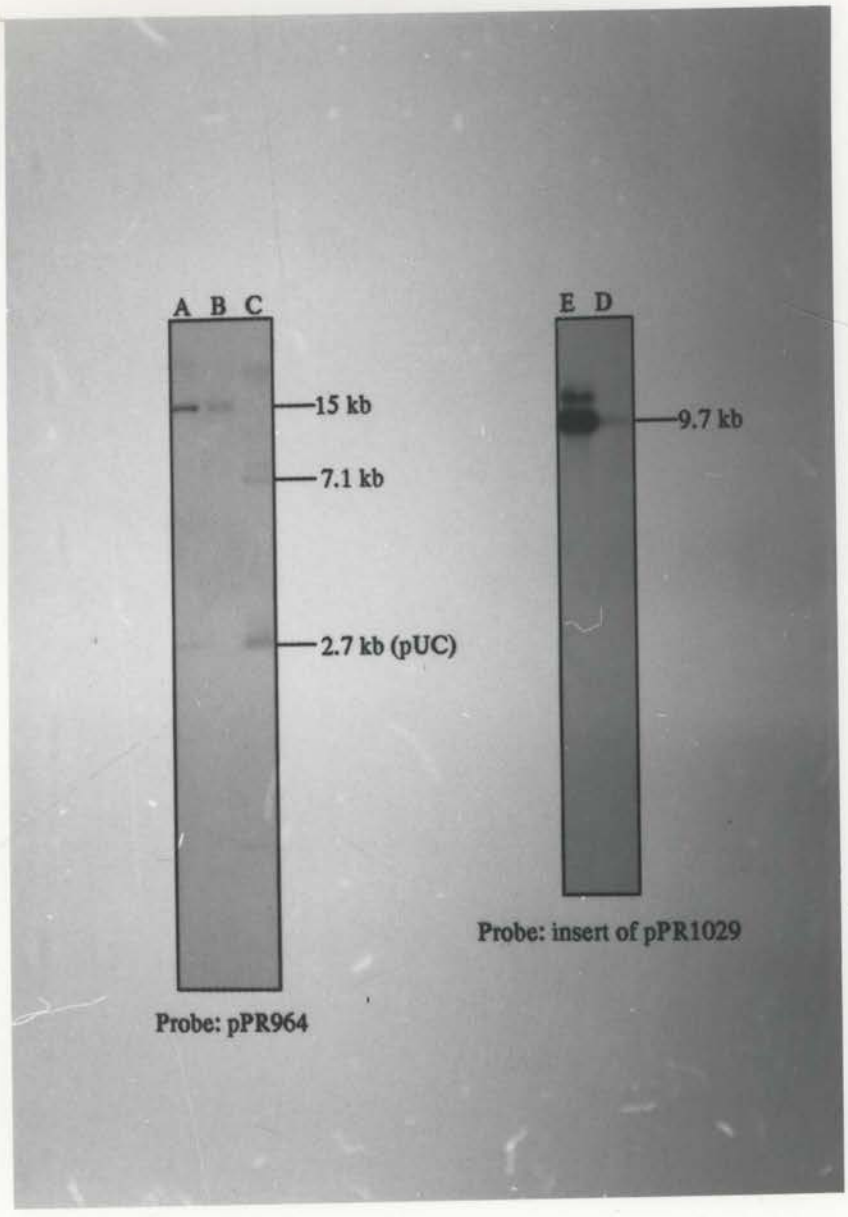

Probe: pPR465

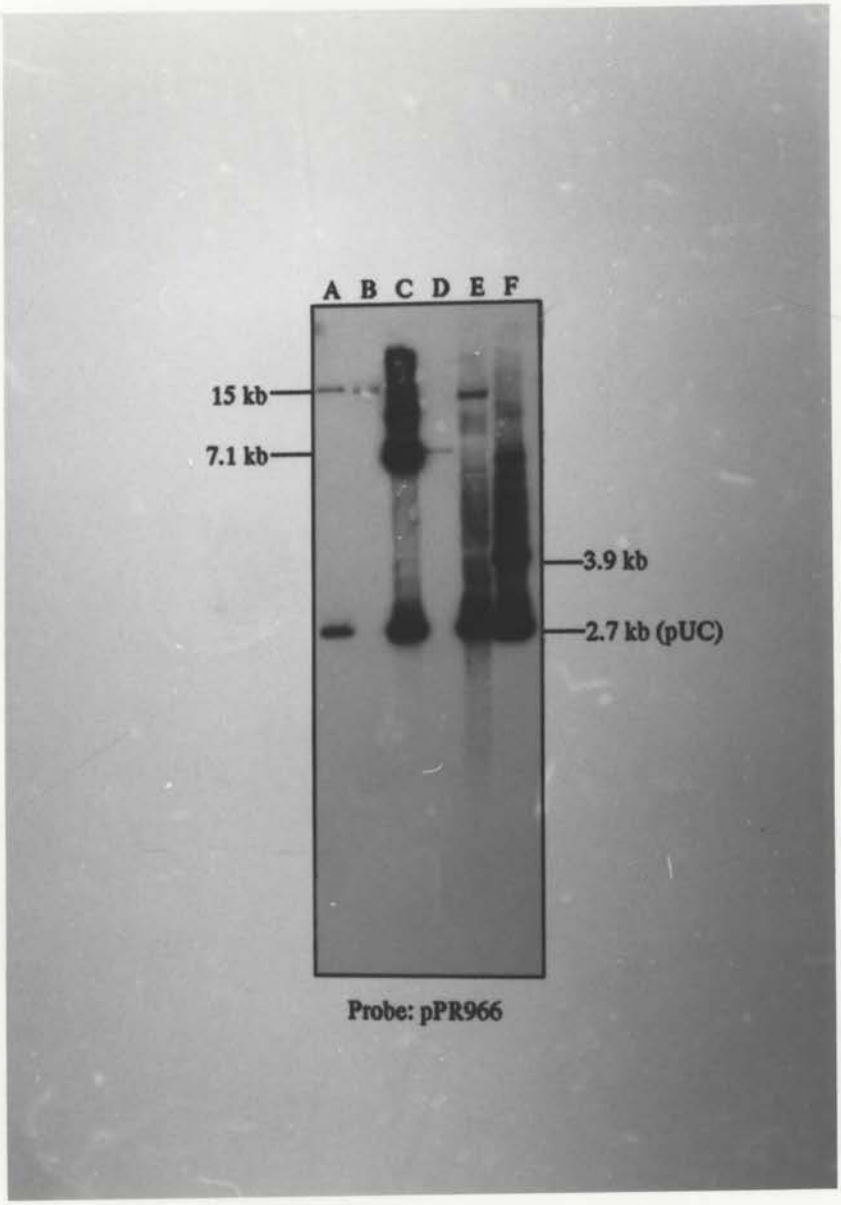
# FIGURE 3.6

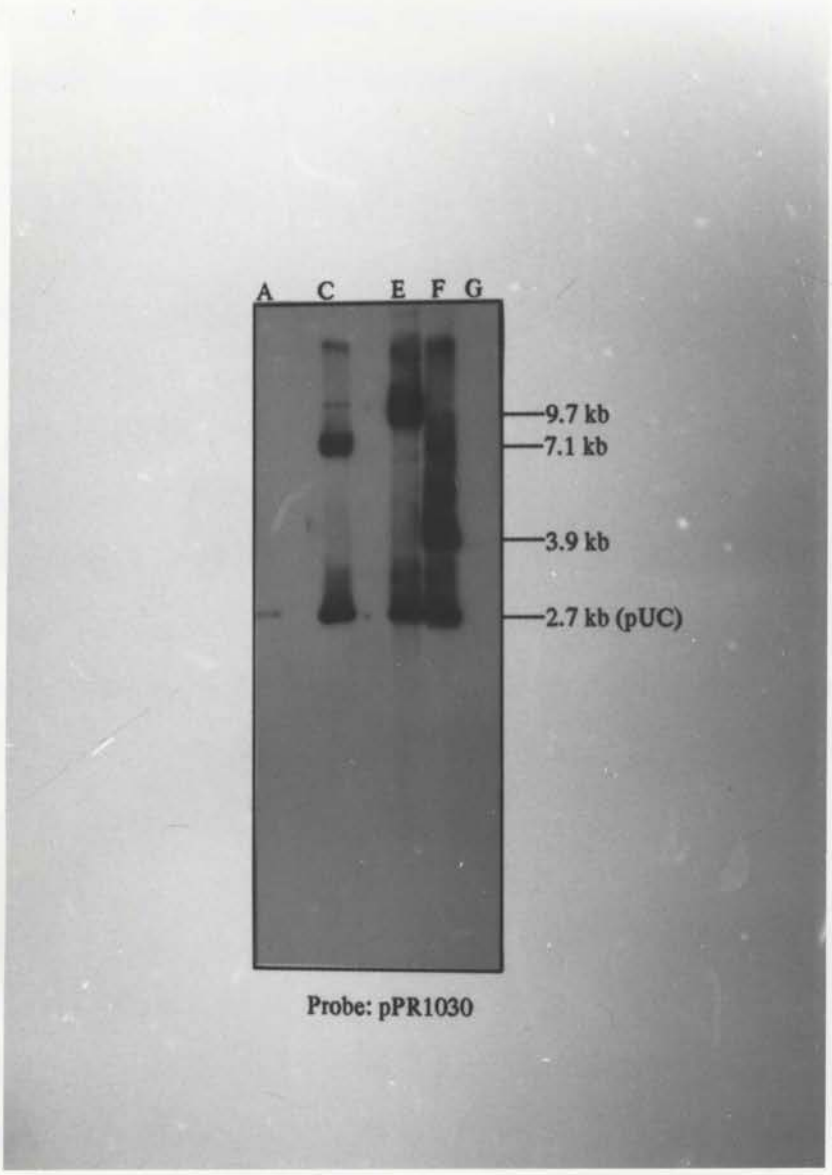
## Confirmation of plasmids pPR964, pPR966, pPR1030 and pPR1029

Southern hybridization analysis of restriction enzyme digests of pPR964, pPR966, pPR1030, pPR1029 and the M32 chromosomal DNA to determine if these four plasmids represent a contiguous region of the chromosome of strain M32. DNA of plasmids pPR964, pPR966, pPR1030 and pPR1029 were used as probes. The lanes are lablled as follows: A, pPR964 digested with *Sac*I; B, M32 chromosome digested with *Sac*I; C: pPR966 digested with *Eco*RI; D: M32 chromosome digested with *Eco*RI; E: pPR1029 digested with *Eco*RI; F: pPR1030 digested with *Hin*dIII; G: M32 chromosome digested with *Hin*dIII.

Note that the 2.7kb fragment present in some tracks represents vector:vector (pUC) hybridization. Also note that the 12.4kb band in track E of some autoradiograms resulted from a partial DNA digest.


A B C
15 kb
7.1 kb
2.7 kb (pUC)
Probe: pPR964
E D
9.7 kb
Probe: insert of pPR1029


A B C D E F
15 kb
7.1 kb
3.9 kb
2.7 kb (pUC)
Probe: pPR966


A
C
E
F
G
9.7 kb
7.1 kb
3.9 kb
2.7 kb (pUC)
Probe: pPR1030

no detectable structural rearrangement during cloning. The results also showed that the inserts of pPR964 and pPR966 hybridized with each other; and pPR966 and pPR1029 both hybridized with pPR1030. These data confirmed that these clones represented continuous overlapping DNA segments from strain M32 chromosome.

## 3.4 RESTRICTION MAPPING

### 3.4.1 RESTRICTION MAPPING OF pPR1029

Plasmid pPR1029 was mapped with the restriction enzymes *Bam*HI, *Bgl*II, *Cla*I, *Hin*dIII, *Kpn*I, *Pst*I and *Sac*I, and the map was compared to those of pPR465 and pPR341 which contain LT2 *rfb* DNA from position 0.0 to position 5.3 and from position 5.3 to position 11.4 respectively (Figure 3.7). Most of the restriction sites on the region from position 0.0 to position 5.6 of strain LT2 appeared to be conserved in pPR1029 (Figure 3.7).

The restriction sites of pPR465 and pPR1029 are identical except for an additional *Cla*I site at position 3.2 in pPR465 (Figure 3.7). The insert of pPR465 includes the 0.00 position used in the map of strain LT2 (Jiang *et al.*, 1991), and as the two strains are identical in this region, the map of strain M32 was given the same start position (Figure 3.7).

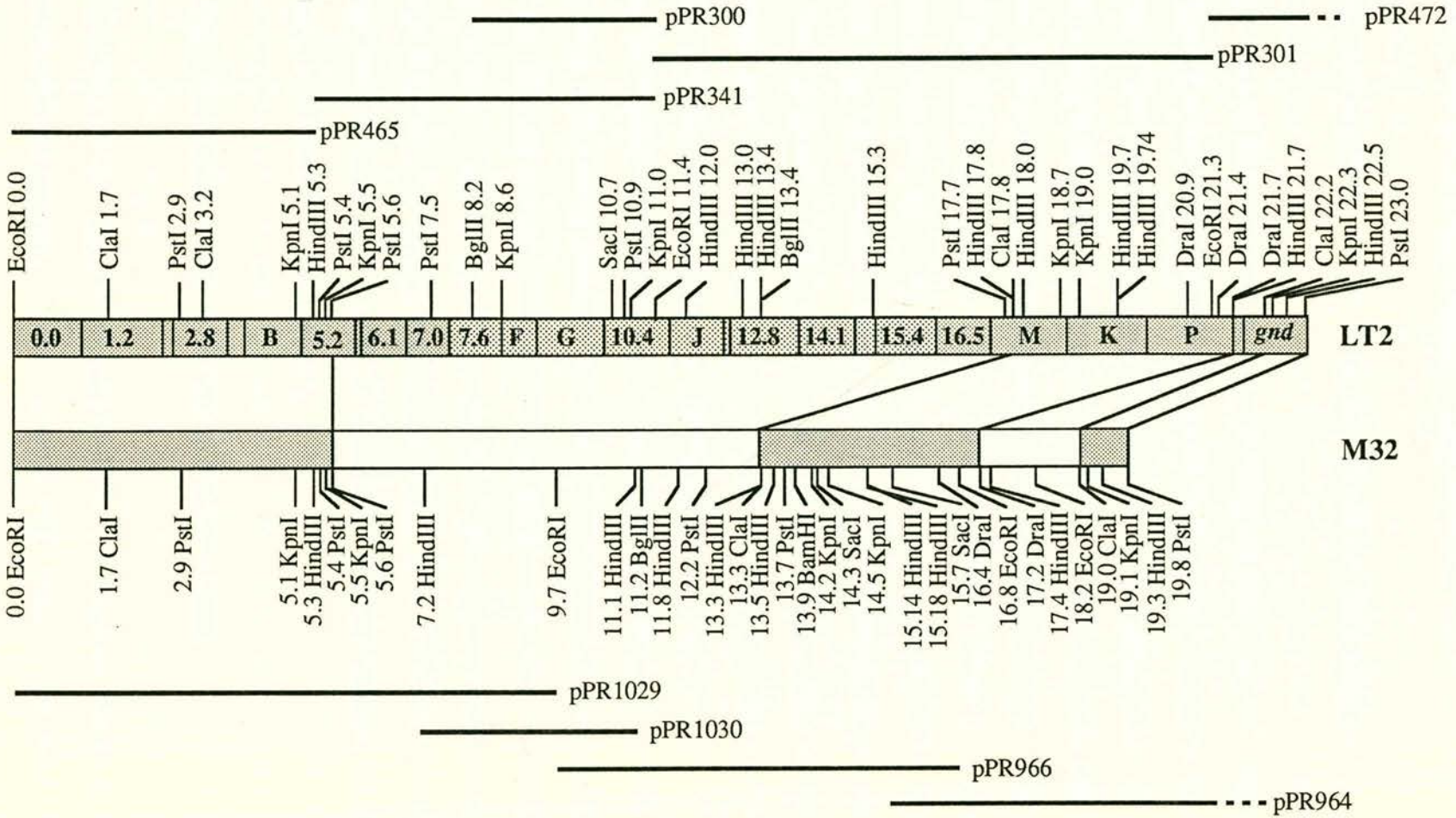
### 3.4.2 RESTRICTION MAPPING OF pPR1030, pPR966 AND pPR964

pPR1030, pPR966 and pPR964 were also mapped with *Bam*HI, *Bgl*II, *Cla*I, *Hin*dIII, *Kpn*I, *Pst*I and *Sac*I (Figure 3.7). pPR1030 has 2.5kb

# FIGURE 3.7

## Restriction map of the *rfb* gene cluster of strain M32

Detailed restriction map of the *rfb* gene cluster and adjacent DNA of strain M32. Restriction enzymes used were *Bam*HI, *Bgl*II, *Cla*I, *Hin*dIII, *Kpn*I, *Pst*I and *Sac*I. Restriction enzyme *Dra*I was only used in the region from position 16.4 to position 19.1 on M32 map. The corresponding region of strain LT2 is also shown . The region of strain LT2 is shown in black, and the region of strain M32 which exhibits identity to that of strain LT2 is shown likewise. The LT2 plasmids (Jiang *et al.*, 1991) used as probes are shown in the upper portion of the diagram, and the M32 plasmids in the lower portion. pPR964 extends through the *gnd* gene to a *Sac*I site which is not shown on the map.


pPR300
pPR472
pPR301
pPR341
pPR465
EcoRI 0.0
ClaI 1.7
PstI 2.9
ClaI 3.2
KpnI 5.1
HindIII 5.3
PstI 5.4
KpnI 5.5
PstI 5.6
PstI 7.5
BglII 8.2
KpnI 8.6
SacI 10.7
PstI 10.9
KpnI 11.0
EcoRI 11.4
HindIII 12.0
HindIII 13.0
HindIII 13.4
BglII 13.4
HindIII 15.3
PstI 17.7
HindIII 17.8
ClaI 17.8
HindIII 18.0
KpnI 18.7
KpnI 19.0
HindIII 19.7
HindIII 19.74
DraI 20.9
EcoRI 21.3
DraI 21.4
DraI 21.7
HindIII 21.7
ClaI 22.2
KpnI 22.3
HindIII 22.5
PstI 23.0
0.0
1.2
2.8
B
5.2
6.1
7.0
7.6
F
G
10.4
J
12.8
14.1
15.4
16.5
M
K
P
gnd
LT2
M32
0.0 EcoRI
1.7 ClaI
2.9 PstI
5.1 KpnI
5.3 HindIII
5.4 PstI
5.5 KpnI
5.6 PstI
7.2 HindIII
9.7 EcoRI
11.1 HindIII
11.2 BglII
11.8 HindIII
12.2 PstI
13.3 HindIII
13.3 ClaI
13.5 HindIII
13.7 PstI
13.9 BamHI
14.2 KpnI
14.3 SacI
14.5 KpnI
15.14 HindIII
15.18 HindIII
15.7 SacI
16.4 DraI
16.8 EcoRI
17.2 DraI
17.4 HindIII
18.2 EcoRI
19.0 ClaI
19.1 KpnI
19.3 HindIII
19.8 PstI
pPR1029
pPR1030
pPR966
pPR964

DNA overlapping with pPR1029 and 1.4kb overlapping with pPR966. pPR964 overlaps with pPR966 by 1.1kb (Figure 3.7).

### 3.4.3 COMPARISON WITH THE *RFB* MAP OF STRAIN LT2

Comparison of the restriction enzyme map of the M32 *rfb* region with that of strain LT2 showed that most of the restriction sites from position 0.0 to position 5.6 and from position 17.8 to position 21.3 on the LT2 *rfb* map were conserved in M32, with the *Eco*RI site at position 0.0 in strain LT2 corresponding to the *Eco*RI site at position 0.0 of the M32 map and the *Cla*I site at position 17.8 of LT2 map corresponding to the *Cla*I site at position 13.3 of the M32 map (Figure 3.7). Four sites (*Pst*I at position 13.7, *Bam*HI at position 13.9, *Sac*I at position 14.3 and *Sac*I at position 15.7) of the thirteen restriction sites (all 6 base pair sites) within the region from position 13.3 to 16.8 on the M32 map were absent in strain LT2 (Figure 3.7): it was estimated that there was approximately 5% (4x100/13x6) nucleotide differences between strains M32 and LT2 in this region. One site (*Cla*I at position 3.2) of the nine restriction sites within the region from position 0.0 to 5.6 on the LT2 *rfb* map was not present in strain M32 giving an approximate estimate of 1 to 2% nucleotide differences between strains M32 and LT2 in this region, or an average of 3% for the two regions.

## 3.5 THE REGION UNIQUE TO STRAIN M32

The first two of the regions of high similarity discussed above flank different central regions in each strain with neither similarity in restriction map nor detectable homology by Southern blotting. To define the junctions between these regions, we probed *Eco*RI and

*Hin*dIII digested M32 clones (pPR1029, pPR1030, pPR966 and pPR964) with the DNA of LT2 clones, pPR465, pPR341, pPR301 and pPR472 (Figure 3.8).

The 5.3kb *Eco*RI-*Hin*dIII fragment of pPR1029, which contains M32 DNA between map positions 0.0 and 5.3, hybridized strongly with the insert of pPR465. The 1.9kb *Hin*dIII fragment of pPR1029 from position 5.3 to position 7.2 hybridized strongly with the insert of pPR341. The 2.5kb *Eco*RI/*Hin*dIII fragment of pPR1029 (position 7.2 to position 9.7) did not exhibit any detectable similarity with any of these probes.

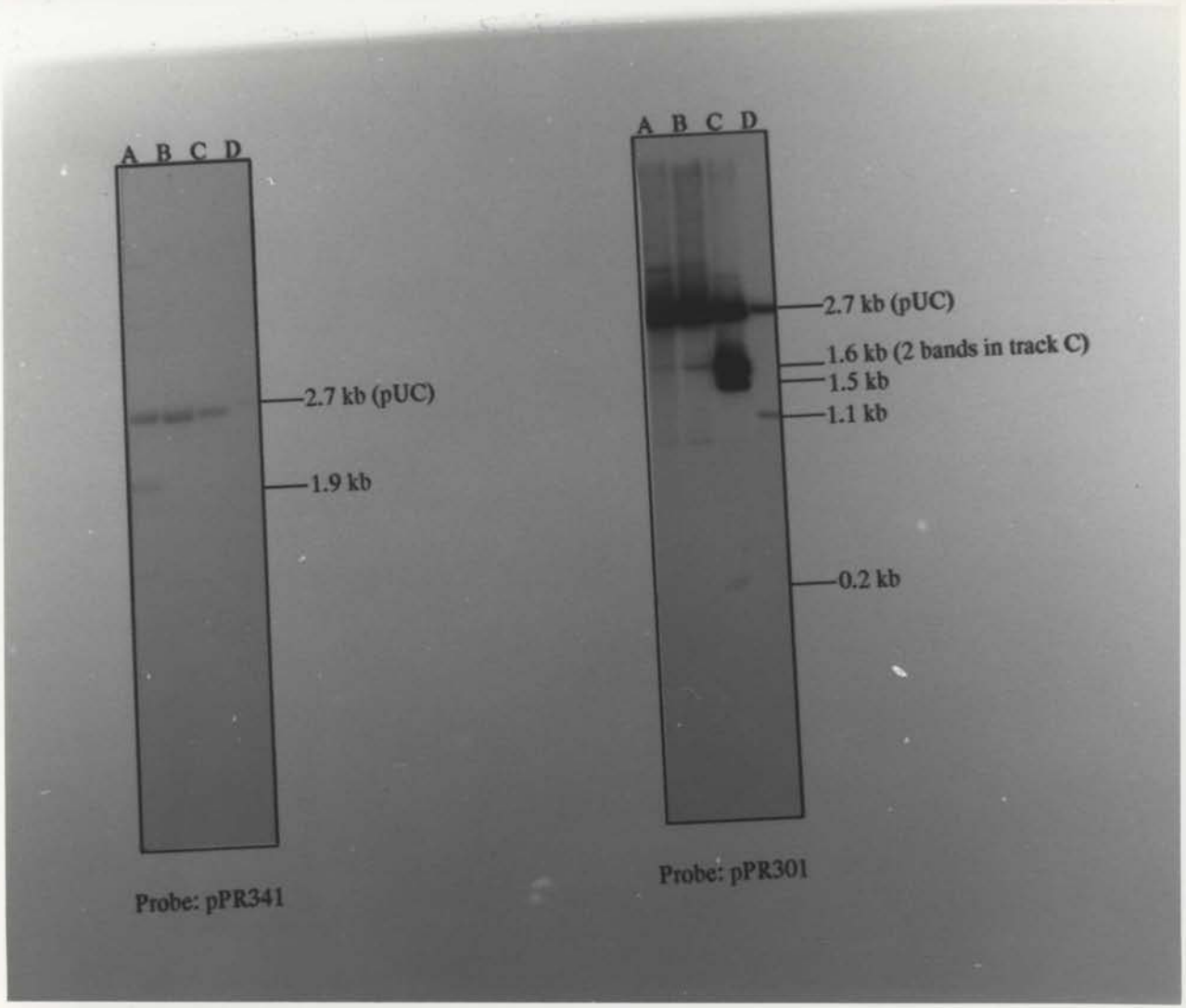
The 1.6kb *Hin*dIII fragment (position 13.5 to position 15.1) and 1.6kb *Hin*dIII-*Eco*RI fragment (position 15.2 to position 16.8) of pPR966 both hybridized strongly with the insert of pPR301. The 1.5kb *Hin*dIII piece (position 11.8 to position 13.3) hybridized with the insert of pPR301. The 1.4kb *Eco*RI/*Hin*dIII piece (position 9.7 to position 11.1) and 0.7kb *Hin*dIII piece (position 11.1 to position 11.8) did not hybridize with any of the probes.

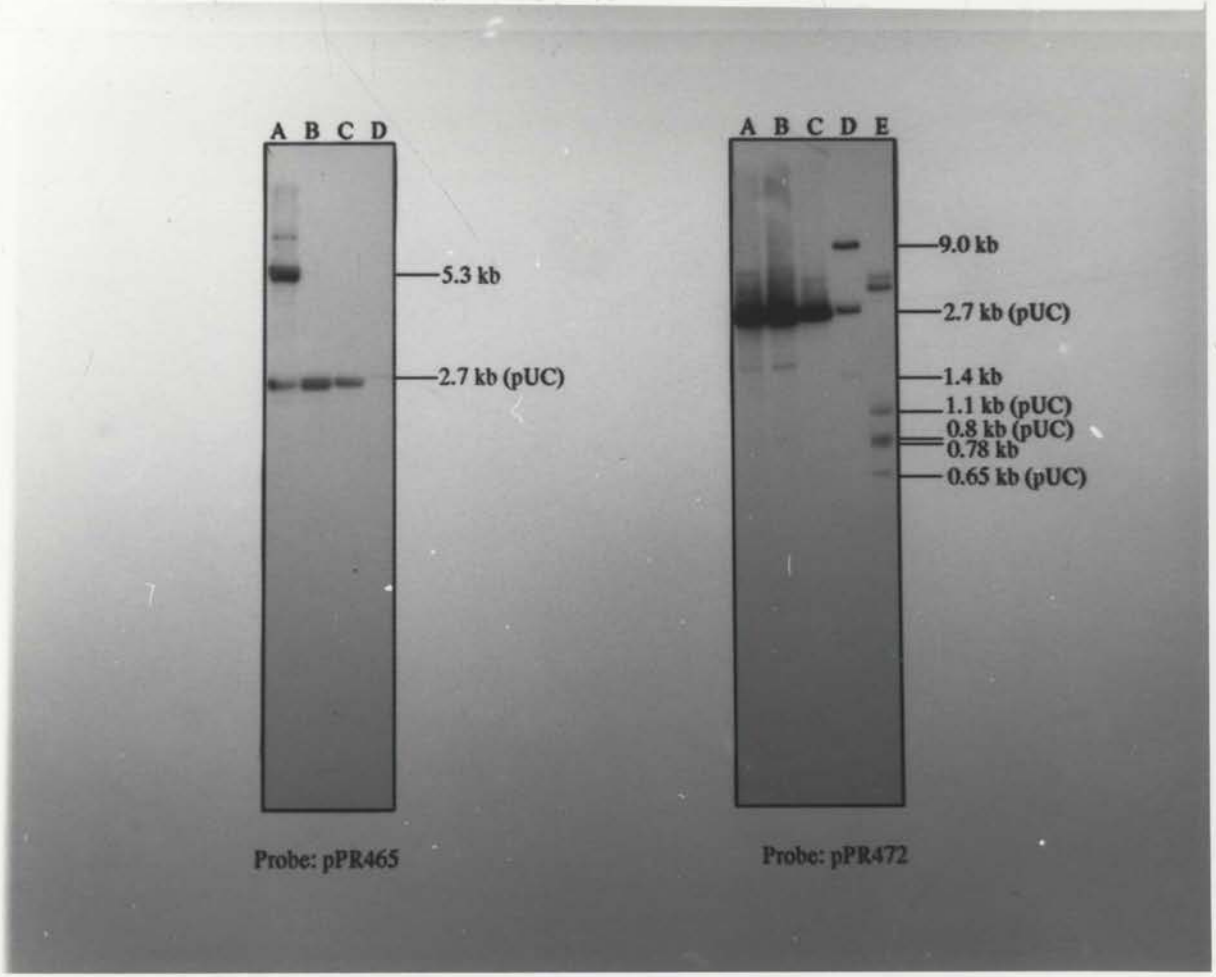
The 1.4kb *Eco*RI fragment (position 16.8 to position 18.2) of pPR964 hybridized to the insert of pPR472, which covers the region from position 21.3 through the *gnd* gene of strain LT2. To further define the junction between similar and non-hybridizing regions in this region, the insert of pPR472 was used as a radioactive probe against a *Dra*I digest of the insert of pPR964. A 0.78kb *Dra*I fragment from position 16.4 to position 17.2 on the M32 map hybridized strongly. Thus, the junction in this region must be between position 16.8 and position 17.2 of strain M32.

# FIGURE 3.8

## Comparison of the *rfb* regions of strains M32 and LT2 by Southern hybridization

The M32 plasmids pPR964, pPR966, pPR1030 and pPR1029 were doubly digested and probed with the LT2 plsmids pPR465, pPR341, pPR301 and pPR472. Probe used is indicated below each autoradiogram. Lanes are designated as follows: A, pPR1029 doubly digested with *Eco*RI and *Hin*dIII; B, pPR1030 doubly digested with *Eco*RI and *Hin*dIII; C, pPR966 doubly digested with *Eco*RI and *Hin*dIII; D, pPR964 doubly digested with *Eco*RI and *Sac*I; E, pPR964 doubly digested with *Dra*I and *Sac*I. The bands representing the vector:vector hybridization are indicated.


A B C D
2.7 kb (pUC)
1.9 kb
Probe: pPR341
A B C D
2.7 kb (pUC)
1.6 kb (2 bands in track C)
1.5 kb
1.1 kb
0.2 kb
Probe: pPR301


A B C D
5.3 kb
2.7 kb (pUC)
Probe: pPR465
A B C D E
9.0 kb
2.7 kb (pUC)
1.4 kb
1.1 kb (pUC)
0.8 kb (pUC)
0.78 kb
0.65 kb (pUC)
Probe: pPR472

# TABLE 3.1

## Comparison of the *rfb* gene regions of strains M32 and LT2 by Southern hybridization

Southern hybridization data using inserts of strain LT2 *rfb* clones as probes against *Eco*RI and *Hin*dIII digests of strain M32 clones. The map position of each LT2 insert is given, together with the map position of each *Eco*RI and *Hin*dIII digested segment of M32 inserts. For M32 clone pPR964 only the segment located in the *rfb* region is listed, and the segment from position 16.4 to 17.2 is a *Dra*I digest.

Symbols: +++, hybridizing strongly at high stringency; -, not hybridizing at high stringency.

## Table 3.1

**Southern Hybridization Data**

| The inserts of M32 clones | | The inserts of LT2 clones as probes | | | |
|---|---|---|---|---|---|
| | | pPR465 0.0-5.3 | pPR341 5.3-11.4 | pPR301 11.4-21.3 | pPR472 21.3-24.6 |
| pPR1029 | 0.0-5.3 | +++ | - | - | - |
| | 5.3-7.2 | - | +++ | - | - |
| pPR1030 | 7.2-9.7 | - | - | - | - |
| | 9.7-11.1 | - | - | - | - |
| | 11.1-11.8 | - | - | - | - |
| pPR966 | 11.8-13.3 | - | - | +++ | - |
| | 13.3-13.5 | - | - | +++ | - |
| | 13.5-15.1 | - | - | +++ | - |
| | 15.2-16.8 | - | - | +++ | - |
| pPR964 | 16.8-18.2 | - | - | - | +++ |
| | 16.4-17.2 | - | - | - | +++ |

The above results, summarized in Table 3.1, show that the first boundary is between positions 5.3 to 7.2 of strain M32, and the second boundary is between positions 11.8 to 13.3. The restriction maps also show differences in the region from positions 17.2 to 19.0 of strain M32. The 0.5kb region from positions 21.7 to 22.2 of strain LT2 is replaced in strain M32 by 1.8 kb DNA.

## 3.6 SUMMARY AND CONCLUSION

The *rfb* gene cluster of strain M32 has been cloned as a series of overlapping clones. Comparison of the *rfb* gene cluster of strain M32 with that of strain LT2 by Southern hybridization and restriction mapping revealed that these two gene clusters have a central unique region flanked by 2 regions which are essentially identical (Figure 3.7), the boundaries between the central region and the flanking regions being from position 5.3 to position 7.2 and from position 11.8 to position 13.3 on the M32 map. The comparison also showed differences in the region from position 17.2 to position 19.0 of strain M32 (Figure 3.7): a 0.5kb DNA from position 21.7 to position 22.2 of strain LT2 is replaced in strain M32 by a 1.8kb DNA.

The DNA from position 0.0 to position 5.3 of strain LT2 is conserved in strain M32. In strain LT2, the DNA of this region contains *rfb* gene *rfbB* (Jiang *et al.*, 1991), and the same must apply to strain M32. *rfbB* encodes dTDP-D-glucose-4,6-dehydratase which catalyzes one step in the four step dTDP-rhamnose synthesis (Nikaido *et al.*, 1967; Gabriel and Lenten 1978). Another three ORFs presented in this region of strain LT2 are *orf0.0*, *orf1.2* and *orf2.8*: which are not *rfb* genes although *orf2.8* has not been completely excluded as yet (Jiang *et al.*,

1991). These three genes are also clearly present in the same positions in strain M32.

The central region of strain M32 from position 7.2 to position 11.8 is unique and replaces the LT2 *rfb* DNA from position 7.2 to position 16.5. This central region of strain LT2 contains not only all the CDP-abequose pathway genes but also *orf12.8*, *orf14.1*, *orf15.4* and part of *orf7.0* . *orf12.8* encodes protein of unknown function with 12 transmembrane segments (Jiang *et al.*, 1991). *orf14.1* and *orf15.4* are the candidates for abequose and mannose transferase (Liu *et al.*, 1991). The linkage between mannose and rhamnose in M32 O antigen is different from that in LT2 O antigen (Figure 1.3), and the mannose transferase genes of these two strains must be different. *orf7.0* is one of the dTDP-rhamnose pathway genes (Jiang *et al.*, 1991).

The right hand flanking region contains *rfbK*, *rfbM* and *rfbP* in strain LT2, and the same genes must be present in strain M32. *rfbK* and *rfbM* are the only *rfb* genes necessary for CDP-mannose synthesis. *rfbP* encodes galactosyl-p-p-undecaprenol synthetase (galactose transferase).

A 0.5kb DNA between *rfbP* and *gnd* in strain LT2 is replaced by a 1.8kb DNA in strain M32.

# CHAPTER 4: ANALYSIS OF THE NUCLEOTIDE SEQUENCE OF M32 *RFB* REGION

## 4.1 INTRODUCTION

As shown in the previous chapter, the M32 *rfb* DNA from position 0.0 to position 5.3 and from position 13.3 to position 16.8 is nearly identical in strain LT2, and the DNA from position 7.2 to position 11.8 and from position 17.2 to position 19.0 of strain M32 shows no significant homology with LT2 *rfb* DNA based on restriction map comparison and Southern Blotting. To enable a detailed comparison with the *rfb* regions of strain LT2 and those of other groups, the regions unique to strain M32 and the junction regions were sequenced.

Group E1 O antigen, like that of groups B, A, D, C1 and C2, contains mannose and rhamnose, and the genes involved in the synthesis of these two sugars are expected to be in the identical regions in strain M32. The central region of LT2 *rfb* region carrying the genes for the synthesis of abequose which is not present in group E1 O antigen is not expected to be found in M32 *rfb* region.

The polymerase which catalyzes the linkage between subunits of O antigen is encoded by *rfc* which is located far away from the *rfb* region in group A, B and D (Mäkelä and Stocker, 1984). In group B, *rfc* gene has been located between *gal* and *trp*, and been sequenced recently (Mäkelä and Stocker, 1984; Collins and Hacket, 1991; Naide *et al.*,

1965). In group E1, Nyman *et al.* (1979) found that 49 of 50 *his*$^+$ recombinants between a wild type strain and a *rfc*$^-$/*his*$^-$ strain were *rfc*$^+$ and therefore proposed that *rfc* gene was within the *rfb* region in this group.

## 4.2 SEQUENCING STRATEGY

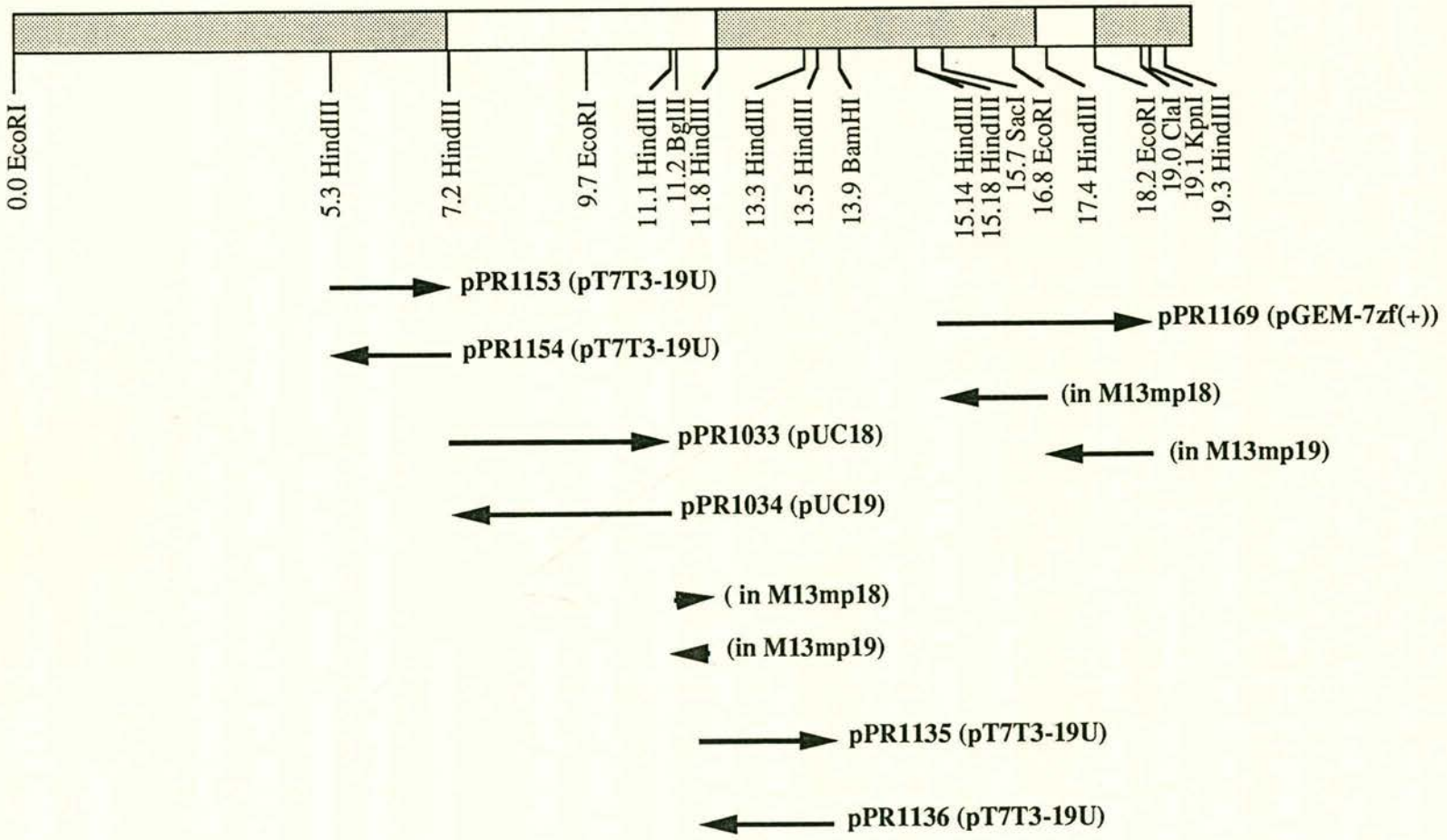
The regions from position 5337 to position 13000 and from position 16808 to position 19094 were sequenced. In both cases, the new sequence extended into the regions nearly identical in strains M32 and LT2.

To enable sequencing of the unique regions, a few subclones which cover those regions were constructed (Figure.4.1). Nested deletion families were then constructed from these subclones to facilitate sequencing. Nested sets of deletions were made using exonuclease III (Henikoff, 1984), and pT7T3-19U (from Pharmacia), pGEM-7zf(+) (from Promega), pUC18 or pUC19 (Norrander *et al.*, 1983) as the vector. Exonuclease III digests from a 5' protruding or blunt end, while leaving a 3' protruding end intact. Thus, the vector can be protected from digestion by creating a 3' overhang closest to the vector, and a 5' overhang closest to the insert will allow digestion of the insert DNA, the extent being dependent on the time allowed for digestion. pT7T3 and pGEM carry the IG region from M13 which enables derivatives to be packaged as single-stranded DNA (ssDNA) during super-infection with helper phage M13K07 (Vieira and Messing, 1987). The inserts in pUC18 and pUC19 were transferred into M13mp18 or M13mp19 (Norrander *et al.*, 1983) and ssDNA was made from these clones. In both cases the ssDNA was sequenced by the dideoxy method, in general using fluorescent dye-labelled M13

# FIGURE 4.1

## Construction of clones for sequencing the *rfb* region of strain M32

The clones are shown underneath a partial restriction map of M32 *rfb* region, in which the regions homologous to strain LT2 are shown in grey and the regions showing nohomology to that of strain LT2 are represented as open box. The vectors into which each DNA fragment was cloned are listed in brackets, and the orientation of each insert is indicated by an arrow.


0.0 EcoRI
5.3 HindIII
7.2 HindIII
9.7 EcoRI
11.1 HindIII
11.2 BglII
11.8 HindIII
13.3 HindIII
13.5 HindIII
13.9 BamHI
15.14 HindIII
15.18 HindIII
15.7 SacI
16.8 EcoRI
17.4 HindIII
18.2 EcoRI
19.0 ClaI
19.1 KpnI
19.3 HindIII
pPR1153 (pT7T3-19U)
pPR1154 (pT7T3-19U)
pPR1033 (pUC18)
pPR1034 (pUC19)
( in M13mp18)
(in M13mp19)
pPR1135 (pT7T3-19U)
pPR1136 (pT7T3-19U)
pPR1169 (pGEM-7zf(+))
(in M13mp18)
(in M13mp19)

standard primer and running the reaction mixture on an ABI 370A sequencer. The small amount of sequencing using custom made primers was done with [$^{35}$S]-labelled dATP, and the gels were read after autoradiography.

The region from position 5.3 (*Hin*dIII) to 7.2 (*Hin*dIII) was cloned into the *Hin*dIII site of pT7T3 in both orientations to give pPR1153 (forward orientation) and pPR1154 (reverse orientation). Nested deletions were made from these two clones using *Bam*HI and *Sac*I as 5' and 3' overhang enzymes.

The region from position 7.2 (*Hin*dIII) to 11.1 (*Hin*dIII) was cloned into pUC18 and pUC19 to give pPR1033 (forward orientation) and pPR1034 (reverse orientation): nested sets of deletions were made using *Sal*I and *Sac*I as the 5' and 3' overhang enzymes and then were transferred to M13mp19.

The fragment from position 11.1 (*Hin*dIII) to 11.2 (*Bgl*II) was sequenced by cloning it into M13mp18 and M13mp19 and then sequencing in both orientations.

The fragment from position 11.2 (*Bgl*II) to 13.9 (*Bam*HI) was cloned into pT7T3 in both orientations to give pPR1135 (forward orientation) and pPR1136 (reverse orientation): nested sets of deletions were made using *Sac*I and *Sma*I as 3' overhang and blunt end enzymes. The sequence from position 11.1 to position 13.0 was obtained from these two nested sets of deletions. A few gaps in this area were filled in using custom made primers.

The 3.4kb *Sac*I-*Kpn*I fragment (position 15.7 to position 19.1) was

cloned into pGEM to give pPR1169. Nested deletions were made using *Kpn*I and *Xba*I as 3' and 5' overhang enzymes from pPR1169, and one direction of the sequence from position 16,808 to position 19,094 was obtained using these deletions. The other direction was sequenced by cloning the *Sac*I-*Hin*dIII fragment (position 15.7 to position 17.4) and the *Hin*dIII-*Kpn*I fragment (position 17.4 to position 19.1) into M13mp18 and M13mp19 respectively, and using custom made primers.

The sequences, orientations and positions of the primers used in sequencing are listed in Table 4.1. The members of the deletion families used for sequencing are shown in Figure 4.2.

The sequence data was thoroughly checked and no errors was detected. The sequences obtained are shown in Figures 4.3 and 4.4.

## 4.3 NUCLEOTIDE SEQUENCE ACCESSION NUMBER

The sequences from position 5337 to position 13000 and from position 16808 to position 19094 presented here have been assigned the EMBL accession numbers X60665 and X60666, respectively.

## 4.4 SEQUENCE ANALYSIS

### 4.4.1 OPEN READING FRAMES

Shepherd's RNY preference method based on the finding of the preference of many genes for the use of codons in form RNY where R represents purine, Y represents pyrimidine, and N represents any base (Shephered, 1981) was used to predict genes in the two sequenced regions.

# FIGURE 4.2a

## Clones used in obtaining the sequence from positions 5.3 to 13.0

The direction of the sequence obtained from each clone is indicated by arrow. The length of the sequence is indicated by the length of the line, with the dashed lines representing the sequences derived using synthetic oligonucleotides as primers and the plain lines representing the sequences derived from deletion clones constructed using exonuclease III.

The lines to the right of **A** represent the deletion clones of plasmids pPR1153, pPR1154, pPR1135 and pPR1136. These clones were constructed in pT7T3, and were *in vivo* packaged as single stranded DNA using helper phage M13KO7.

The lines to the right of **B** represent the deletion clones of pPR1033 and pPR1034. These were constructed in pUC18 and pUC19, and subsequently cloned into M13mp18 and M13mp19 for sequencing.

B
A
HindIII 5.3
HindIII 7.2
HindIII 11.1
BglII 11.2
BamHI 13.9

# FIGURE 4.2b

## Clones used in obtaining the sequence from positions 16.8 to 19.1

The direction of the sequence obtained from each clone is indicated by arrow. The length of the sequence is indicated by the length of the line, with the dashed lines representing the sequences derived using synthetic oligonucleotides as primers and the plain lines representing the sequences derived from deletion clones constructed using exonuclease III.

Those lines to the right of **A** represent sequence derived from deletion clones made from pPR1166 (a pGEM clone carrying the *Sac*I-*Kpn*I fragment from positions 15.7 to 19.1) using *Kpn*I and *Xho*I as 3' and 5' overhang enzymes. The line to the right of **B** represents sequence derived from a M13mp18 clone with carrying the *Sac*I-*Hin*dIII fragment from positions 15.7 to 17.4. The lines to the right of C represent sequence derived from a M13mp19 clone carrying the *Hin*dIII-*Kpn*I fragment from positions 17.4 to 19.1.

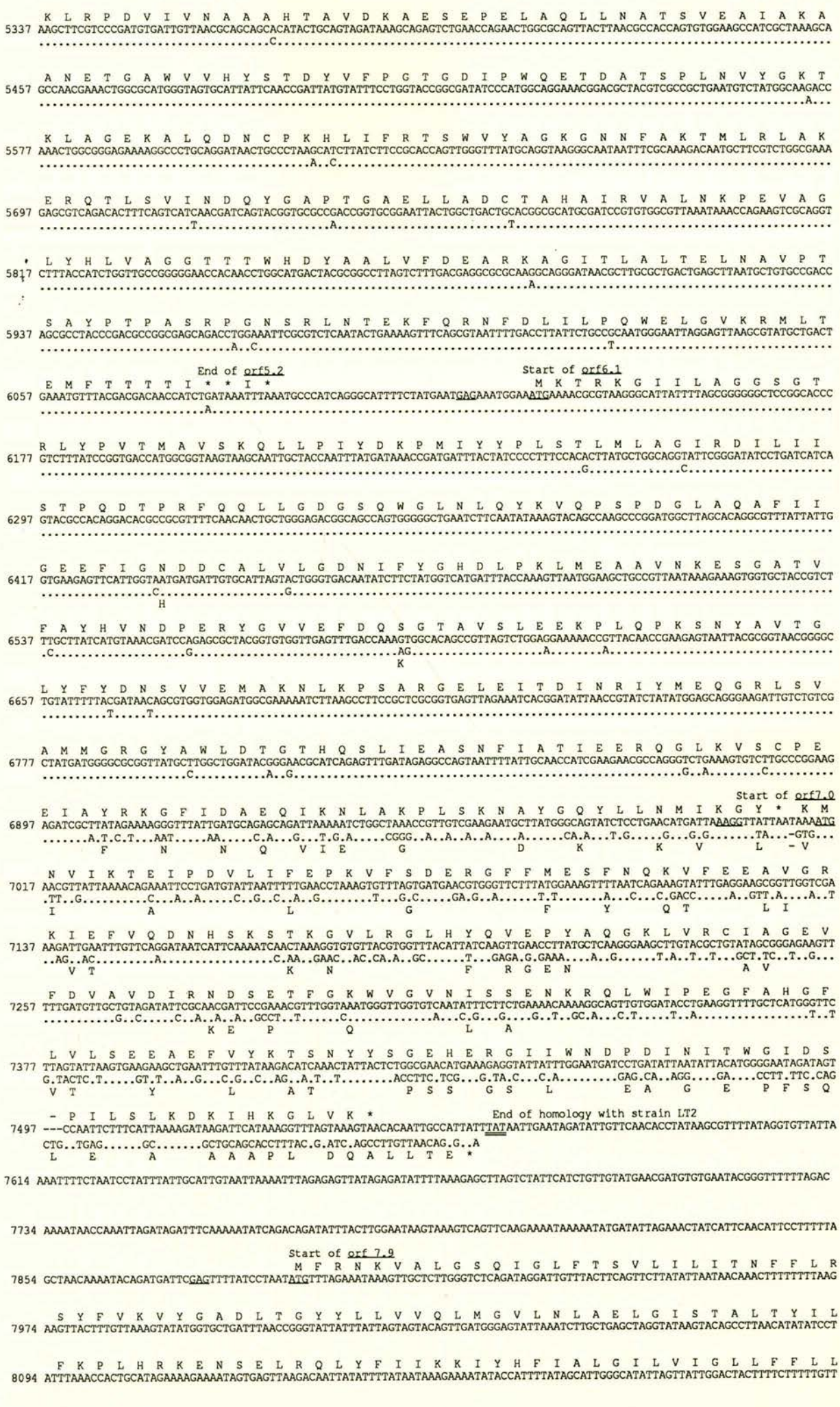
A
B
C
SacI 15.7
EcoRI 16.8
HindIII 17.4
EcoRI 18.2
KpnI 19.1

# FIGURE 4.3

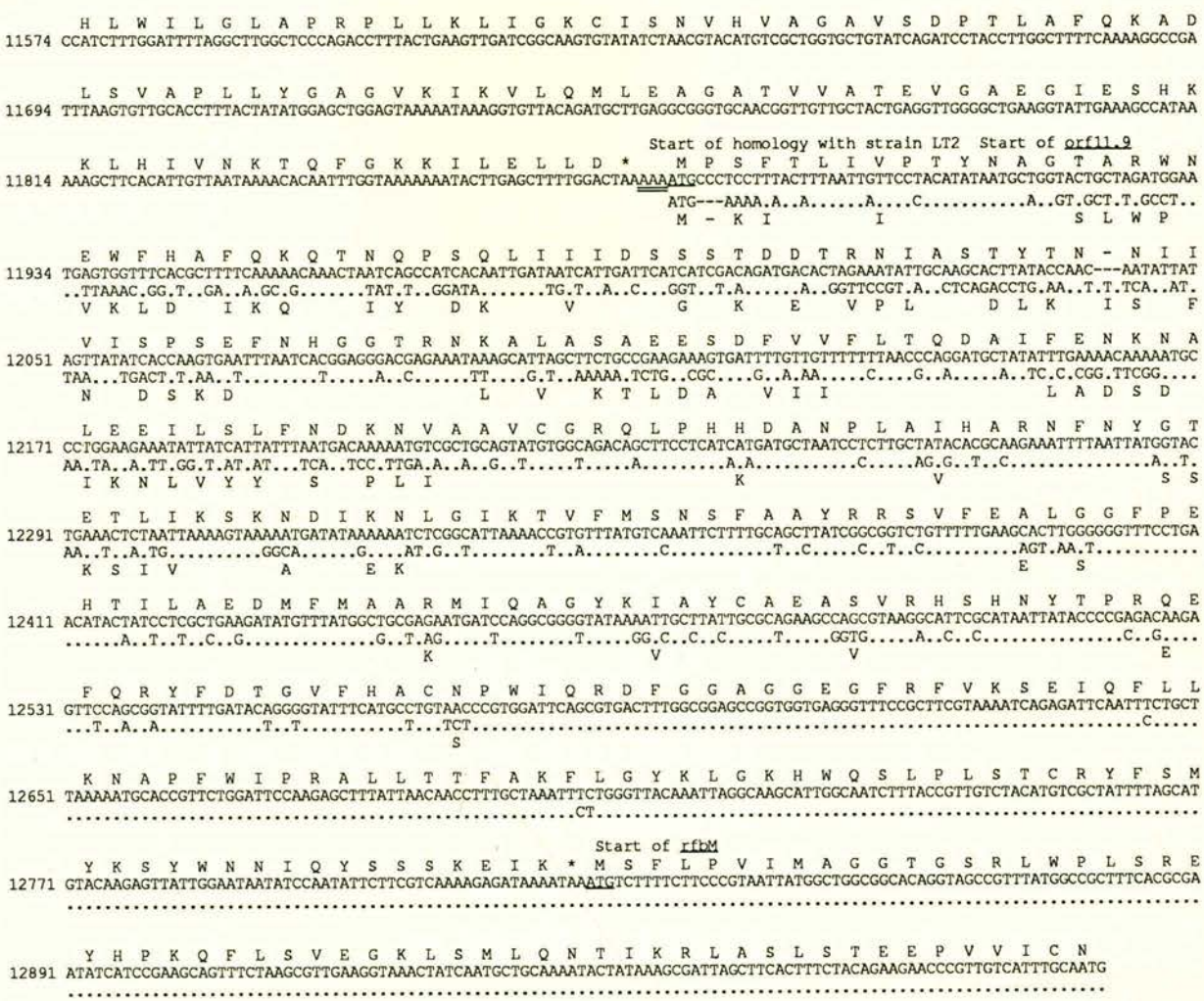
## Sequence of the central region of M32 *rfb*

The DNA sequence from position 5,337 to position 13,000 of the *rfb* region of strain M32, with the amino acid sequences of open reading frames shown above the DNA sequence.

The start position and Shine-Dalgarno sites, where present, are underlined. The sequences of the corresponding segments of strain LT2 are given underneath, with only the nucleotide and amino acid differences presented


End of orf5.2
Start of orf6.1
Start of orf7.9
End of homology with strain LT2

```
      L  N  S  I  V  N  A  S  I  S  P  E  N  L  Y  I  T  W  G  V  F  V  I  S  T  S  L  S  Y  L  Y  S  A  Q  S  V  I  L  T  A
8214  ATTAAATTCTATAGTAAATGCAAGTATATCCCCTGAAAATCTATATATAACATGGGGGGTGTTTGTTATAAGTACATCATTATCATATTTATACTCTGCTCAATCTGTCATTTTAACAGC

      D  Q  N  V  Y  L  V  K  L  I  T  G  L  T  R  S  L  A  Y  I  L  Q  I  F  L  M  I  C  G  V  S  F  W  I  V  C  A  I  E  L
8334  AGACCAGAATGTCTATTTGGTTAAATTAATTACTGGGTTAACTAGAAGTTTAGCGTATATATTGCAAATATTTCTAATGATTTGCGGCGTCTCCTTTTGGATTGTGTGCGCTATTGAATT

      L  S  N  V  I  Q  L  I  L  F  N  K  L  T  L  K  K  Y  P  Q  L  V  K  L  D  I  T  D  T  I  N  K  E  N  I  I  S  V  K  S
8454  ACTTTCTAATGTAATACAACTTATTCTTTTTAACAAACTAACATTAAAAAAATATCCGCAATTAGTAAAGTTGGATATTACTGATACTATCAATAAAGAAAATATAATAAGTGTAAAATC

      K  I  K  R  E  I  K  Y  T  F  V  H  K  I  A  G  V  V  V  F  N  T  D  Y  L  I  I  S  V  F  L  G  I  T  V  I  T  S  F  S
8574  TAAAATAAAAAGAGAGATAAAATATACGTTTGTACATAAAATTGCCGGTGTTGTTGTATTTAATACAGATTATTTAATTATATCTGTCTTTCTAGGTATAACGGTTATCACATCTTTTTC

      S  Y  M  M  L  I  Q  A  L  A  F  L  I  S  A  I  A  S  P  L  G  A  Y  I  G  A  K  L  Y  S  D  S  K  E  K  V  I  R  I  I
8694  TTCCTATATGATGTTAATTCAAGCATTGGCTTTTCTGATATCAGCAATAGCTTCCCCACTGGGTGCCTATATAGGAGCAAAACTATATTCTGATAGTAAGGAAAAAGTTATCAGGATTAT

      S  L  Y  N  S  L  F  F  M  L  G  I  F  C  A  Y  M  F  Y  I  S  S  S  T  F  V  S  L  W  M  G  K  E  I  V  L  S  Q  Q  L
8814  ATCACTTTATAATTCACTTTTTTTTATGCTAGGTATTTTTTGCGCGTATATGTTTTATATATCTTCAAGTACATTTGTATCCTTATGGATGGGAAAAGAAATTGTACTTAGCCAACAATT

      V  L  L  L  S  V  N  C  F  V  L  I  A  R  I  S  F  D  V  A  K  V  G  L  G  Y  M  S  D  I  E  L  P  L  L  E  A  I  I  N
8934  AGTTTTGTTACTTTCTGTTAATTGCTTCGTATTGATTGCAAGGATCTCTTTTGATGTTGCAAAAGTAGGCCTTGGTTATATGTCAGATATTGAATTACCCCTACTGGAAGCCATAATTAA

      I  I  V  S  L  V  L  V  H  Y  L  G  L  N  G  I  V  I  G  T  I  V  S  N  I  I  V  V  M  I  L  K  P  V  F  L  Y  K  N  A
9054  TATTATCGTTTCATTAGTTTTAGTACATTATTTAGGTTTAAATGGGATAGTTATTGGAACTATCGTTAGTAATATCATCGTTGTCATGATCTTGAAACCAGTATTTTTATACAAAAATGC

      L  K  S  D  N  A  F  I  I  H  E  L  I  R  S  W  F  F  I  S  I  F  L  L  S  I  F  F  I  T  R  A  K  V  I  V  S  N  E  W
9174  CCTTAAATCGGATAATGCATTTATTATTCATGAGCTTATAAGGTCATGGTTCTTTATAAGCATATTCTTGCTTTCCATCTTTTTTATTACAAGGGCAAAAGTTATAGTCAGCAACGAATG

      L  D  F  F  I  Q  V  C  M  S  S  I  T  P  L  L  L  L  I  M  I  Y  S  V  F  D  S  N  V  R  K  F  M  F  V  M  F  K  K  K
9294  GCTAGATTTCTTTATACAAGTGTGTATGAGTAGCATTACTCCTTTATTGTTATTGATAATGATATACAGCGTCTTTGATAGCAACGTTCGGAAATTCATGTTTGTTATGTTTAAAAAAAA

      Y  R  E  N  *
9414  GTATAGGGAAAATTGATTCTCTAAAATTGGAGAAAAAGATCGTTGCCTTGATAGTTTGATAATGAAACCGATATATATTAAGCGTAGTGGTGTACGCAAAAGCGACTATTTGGCTATTGA

                                                                                  Start of orf 9.6
                                                                                  M  V  F  L  A  W  F  I  A  I  G  I
9534  AAATAAGGTATAAATCTAAACAAGATAAATATATGTAATAAAGTATTATCTTATTCCTTATCGCTTAATAATATGAGATGTGTTATGGTATTTTTAGCTTGGTTTATTGCTATTGGGATA

      Y  I  I  C  G  I  L  Y  L  Q  F  P  Q  L  S  P  L  G  A  L  A  Y  T  D  T  V  G  D  V  I  M  V  V  S  T  G  V  V  G  V
9654  TATATAATATGTGGAATTCTTTATCTTCAATTTCCTCAATTATCACCACTAGGGGCCCTTGCATATACTGATACAGTGGGCGATGTAATCATGGTCGTATCCACGGGGGTAGTGGGAGTT

      F  L  G  Y  F  A  F  S  K  I  K  L  A  G  I  K  D  F  N  N  S  C  F  I  N  F  C  S  V  C  F  I  I  S  A  F  L  I  F  A
9774  TTTCTCGGCTATTTTGCATTTTCTAAAATCAAGTTAGCAGGAATAAAAGATTTTAATAATTCGTGTTTTATAAATTTTTGTTCAGTATGCTTTATTATTTCAGCGTTTCTAATTTTCGCC

      L  G  V  A  A  Y  G  G  Y  I  A  F  L  H  T  P  Y  I  A  I  Y  E  G  S  A  E  N  Q  T  R  D  I  L  I  S  S  S  G  L  L
9894  TTAGGTGTCGCTGCATATGGTGGTTACATAGCATTTCTACATACGCCTTATATAGCAATCTATGAAGGTAGTGCAGAAAATCAAACCAGAGATATTCTCATATCTTCATCGGGATTGCTT

      A  N  Y  A  L  L  T  S  I  R  V  N  F  K  E  I  S  I  T  N  K  V  V  I  A  V  S  L  L  I  L  V  S  I  F  I  Q  G  R  R
10014 GCTAACTATGCACTATTAACTAGCATTCGTGTCAACTTTAAAGAAATATCAATAACTAATAAAGTTGTCATAGCAGTATCGCTATTAATACTTGTTTCTATATTTATTCAGGGACGAAGA

      E  N  L  I  L  L  I  L  C  F  V  S  F  Y  M  L  N  Y  K  I  N  F  K  R  I  I  K  V  S  I  I  C  I  I  M  F  F  I  A  G
10134 GAAAATTTAATTTTGCTAATATTATGTTTCGTGTCATTCTATATGCTAAATTATAAAATTAATTTTAAACGTATTATCAAAGTAAGCATTATTTGTATAATAATGTTTTTCATTGCTGGA

      L  G  L  Y  L  R  E  S  T  S  T  S  G  G  S  V  F  T  A  I  P  F  A  V  M  Y  E  T  H  F  S  L  A  T  L  A  N  E  V  R
10254 TTAGGTTTATATTTGCGTGAAAGCACTTCAACTTCAGGTGGAAGTGTTTTTACAGCTATACCTTTTGCCGTGATGTATGAAACACATTTTTCTTTAGCGACTTTGGCTAATGAAGTGAGA

      T  H  L  Y  N  N  L  P  Y  G  G  V  L  D  L  F  S  P  I  L  F  I  I  P  A  F  I  Y  G  I  F  G  L  D  K  Q  S  L  F  E
10374 ACGCATTTATATAATAATTTACCTTATGGGGGAGTATTAGATTTATTTTCTCCCATATTATTTATTATCCCTGCTTTTATATATGGCATTTTTGGTTTAGATAAACAATCTTTGTTCGAG

      N  N  E  M  R  F  Y  E  D  K  G  G  Q  F  I  F  T  E  A  F  H  S  L  G  Y  V  G  V  F  L  H  G  L  I  L  G  V  M  L  I
10494 AATAATGAGATGCGTTTTTATGAGGATAAAGGTGGGCAGTTTATATTCACAGAGGCTTTTCATTCACTTGGTTATGTCGGAGTATTCCTGCATGGTTTAATCTTGGGTGTAATGCTAATA

      V  F  Y  R  V  A  K  K  S  R  L  I  I  Y  H  F  P  I  I  S  L  I  F  V  A  M  R  K  D  L  T  Y  G  V  K  Y  I  S  L  L
10614 GTATTTTATCGTGTAGCCAAGAAAAGCAGGCTCATAATTTATCATTTTCCAATTATTTCCCTCATATTTGTTGCAATGAGAAAAGACCTTACCTATGGTGTTAAATATATTTCTCTTTTA

                                                                  Start of orf10.8
                                                                  M  V  E  L  K  M  H  K  V  L  F  I  S  A  F  H  P  G  A  R  G  S
      F  I  F  M  M  I  F  Y  F  I  Y  K  L  L  P  L  K  N  N  G  *
10734 TTCATATTTATGATGATTTTTTATTTCATCTATAAGTTACTCCCACTTAAAAATAATGGTTGAATTAAAAATGCATAAAGTTCTTTTTATTAGTGCCTTTCATCCTGGGGCGAGAGGTTC

      I  G  A  G  E  A  I  C  G  D  N  L  K  K  H  V  N  N  G  S  Q  V  D  V  I  V  I  S  P  A  T  Q  Y  E  N  P  E  I  K  A
10854 GATCGGTGCTGGTGAGGCCATTTGTGGTGATAATTTAAAAAAACATGTAAATAATGGCAGTCAAGTCGATGTTATTGTTATATCTCCTGCAACACAATATGAAAATCCAGAAATTAAAGC

      L  C  H  S  Y  E  V  I  N  V  S  K  K  D  Y  L  K  G  I  L  F  N  A  K  Y  K  S  F  I  A  P  W  F  F  T  R  V  N  K  S
10974 ACTATGCCATTCTTATGAAGTAATCAACGTTAGTAAGAAAGATTATCTTAAAGGGATACTTTTCAATGCTAAATATAAATCTTTTATTGCGCCATGGTTTTTTACTAGAGTAAATAAAAG

      L  I  K  R  I  N  N  K  I  S  H  T  N  Y  D  L  V  W  L  D  F  P  S  V  L  G  V  V  K  H  I  L  H  P  N  I  H  Y  F  A
11094 CTTAATTAAACGGATAAATAATAAGATCTCTCACACGAATTACGACTTAGTATGGCTTGATTTTCCATCAGTTTTAGGAGTGGTAAAACATATTTTACATCCAAATATTCATTATTTTGC

      H  D  V  V  S  Q  R  T  S  R  S  L  T  K  K  P  I  N  K  E  I  A  K  V  E  A  N  L  F  C  K  L  K  K  I  V  A  M  S  H
11214 CCATGATGTTGTATCACAGCGCACAAGTCGTTCTCTTACAAAGAAACCTATCAATAAAGAGATTGCAAAAGTTGAGGCAAACTTATTCTGTAAGTTGAAAAAAATTGTGGCAATGTCTCA

      K  D  K  L  L  L  E  Q  M  N  Y  K  G  V  I  E  V  A  D  L  G  V  Q  K  V  G  E  V  L  N  C  I  P  I  E  E  V  V  D  K
11334 TAAAGATAAATTGTTACTTGAACAAATGAATTATAAAGGAGTTATAGAGGTTGCTGATCTTGGAGTACAAAAAGTTGGTGAGGTATTAAATTGTATTCCGATTGAAGAGGTTGTAGATAA

      F  K  D  R  K  N  L  I  F  F  G  Y  M  K  R  A  E  N  H  W  S  I  I  W  F  I  F  F  V  F  L  K  I  R  K  Q  N  P  H  I
11454 ATTTAAGGATAGAAAAAATTTAATTTTTTTTGGTTACATGAAGCGTGCGGAAAACCACTGGTCTATAATATGGTTTATATTTTTTGTTTTTCTTAAAATAAGAAAACAAAATCCACATAT
```

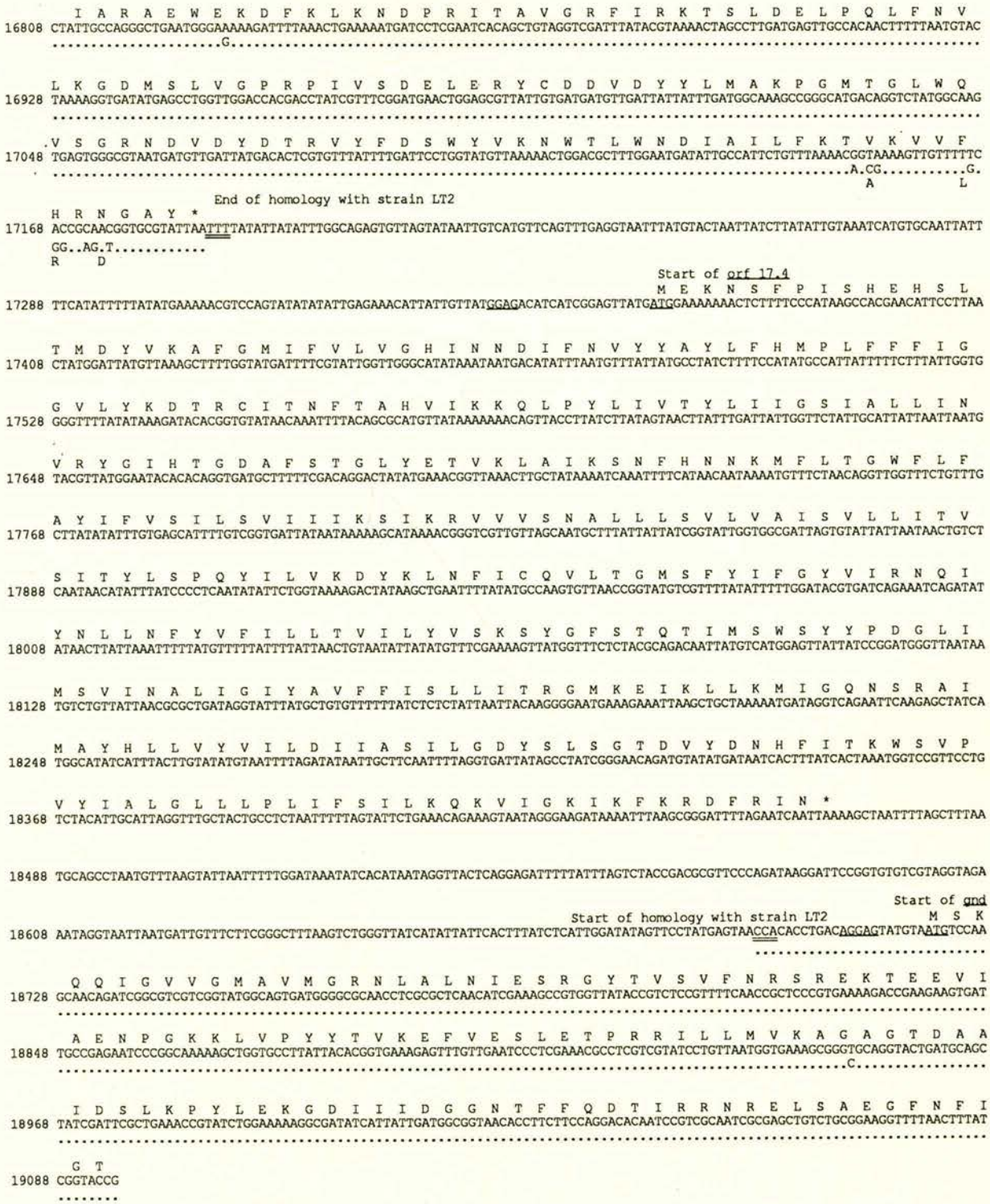

Start of homology with strain LT2
Start of orf11.9
Start of rfbM

# FIGURE 4.4

## Sequence from position 16,808 to 19,094 of M32 *rfb* region

The DNA sequence from position 16,808 to position 19,095 of the *rfb* region of strain M32, with the amino acid sequences of open reading frames shown above the DNA sequence.

The start position and Shine-Dalgarno sites, where present, are underlined. The sequences of the corresponding segments of strain LT2 are given underneath, with only the nucleotide and amino acid differences presented


End of homology with strain LT2
Start of orf 17.4
Start of homology with strain LT2
Start of gnd

# TABLE 4.1

## Synthetic oligonucleotides used in sequence

The oligonucleotides used in sequencing are listed in this table. The orientation of each oligonucleotide is indicated as whether it is the reverse complement of the sequence (rcm) or not. The positions of the sequence that the oligonucleotides bind to, or represent is also shown.

| NAME | SEQUENCE | ORIENTATION | POSITION |
|---|---|---|---|
| Oligo#182 | GTTCAAGTAACAATTTAT | RCM | 11360-11343 |
| Oligo#176 | TGTTTCTCAATATATATA | RCM | 17337-17320 |
| Oligo#177 | TAATCAAATAAGTTACTA | RCM | 17624-17607 |
| Oligo#178 | AGTTATATATCTGATTTC | RCM | 18017-18000 |
| Oligo#179 | TGAGTAACCTATTATGTG | RCM | 18548-18531 |
| Oligo#180 | GAAATCAGATATATAACT | | 18000-18017 |

7 complete open reading frames (ORFs) were found in these two sequences (Figures 4.5 & 4.6), all on the same strand. No ORF longer than 200 base pairs was found on the opposite strand (Figures 4.5 & 4.6). The ORFs were named according to the start position on the map (e.g., *orf17.4* starts at approximately base 17400 on M32 *rfb* map). *orf6.1*, *orf7.0*, *orf7.9*, *orf9.6* and *orf10.8* are located in the central region, and are 876, 528, 1,536, 1,175 and 1,083 base pairs in length respectively. *orf17.4* is 1,104 base pairs in length, and is located in the right end of M32 *rfb* gene cluster.

The sequences were checked for presence of the Shine-Dalgarno sequence which is the sequence showing complementation with the 16S ribosomal RNA sequence and presumed to be the ribosome binding site. The sequence of the initiation region of each ORF is given in Table 4.2. Of the seven ORFs, two (*orf10.8* and *orf11.9*) do not have Shine-Dalgarno sequences. *orf10.8* starts with an ATG codon two bases upstream of the stop codon of *orf9.6*, and the putative start codon of *orf11.9* lies three bases after the stop codon of *orf10.8*: these two ORFs are presumably initiated by translational coupling. *orf9.6* , *orf7.9* and *orf6.1* have only three base Shine-Dalgarno sequences and no possibility of translational readthrough, and are perhaps translated at low efficiency. *orf7.0* and *orf17.4* have good Shine-Dalgarno sequences.

### 4.4.2 PROMOTERS

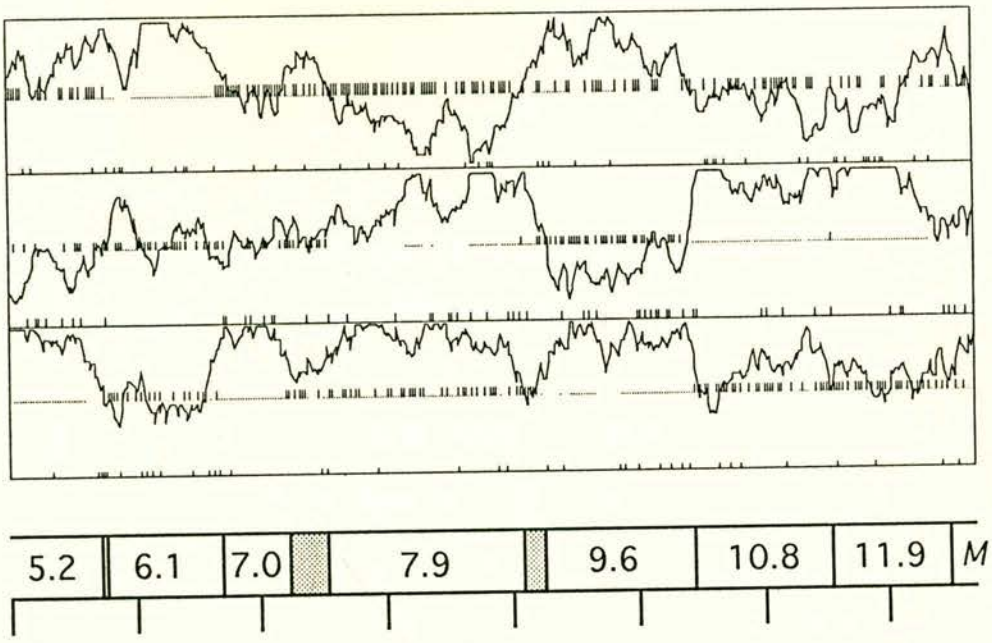
Five programs from Dr. O'Neill were used to search for promoters in the sequenced M32 *rfb* region. A large number of potential promoters were found, but only two locate in the intergenic gap. These two with the mRNA startpoints at positions 9502 and 9563 locate in the

# FIGURE 4.5

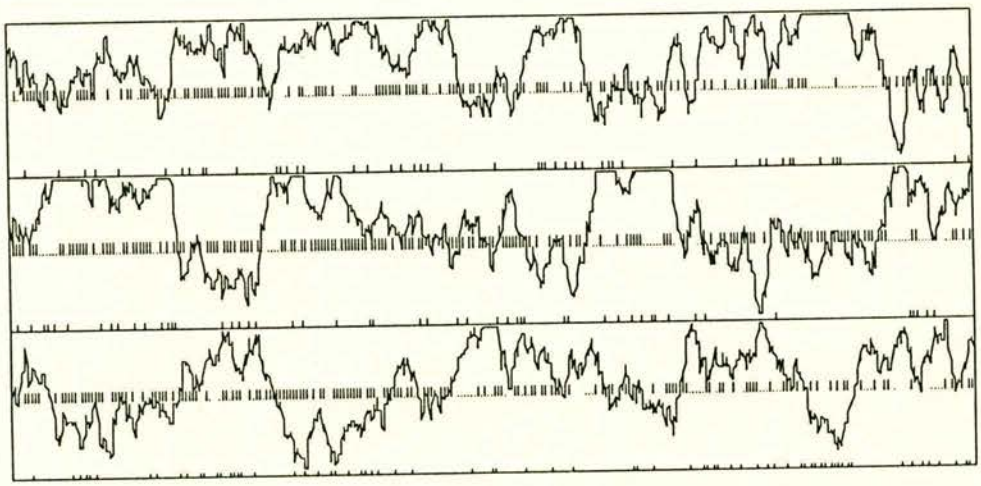
## Determination of coding regions in sequence from position 5,337 to 13,000

The Shepard RNY preference method was used to predict open reading frames in (**A**) the forward direction and (**B**) the reverse direction. The three panels represent the three possible reading frames. The short vertical lines along the central and bottom lines of each panel represent termination and start codons respectively. The curve indicates the extent to which the three bases of the codon are of the form: purine; either; pyrimidine, and thus have a high likelihood of being coding DNA. The predicted open reading frames are shown as the open boxes below the plot labelled A.

A




B

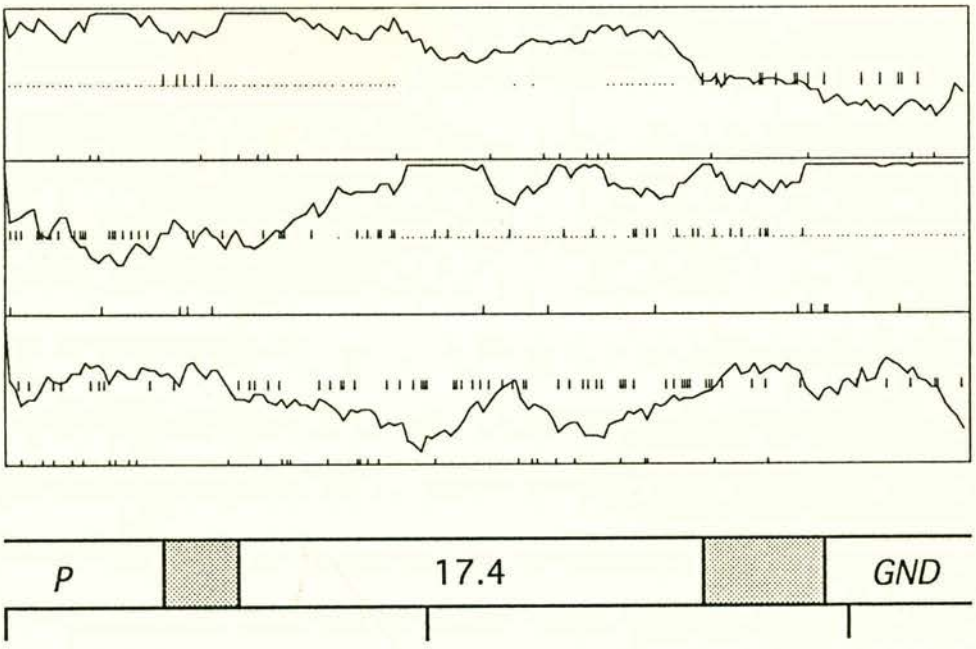


# FIGURE 4.6

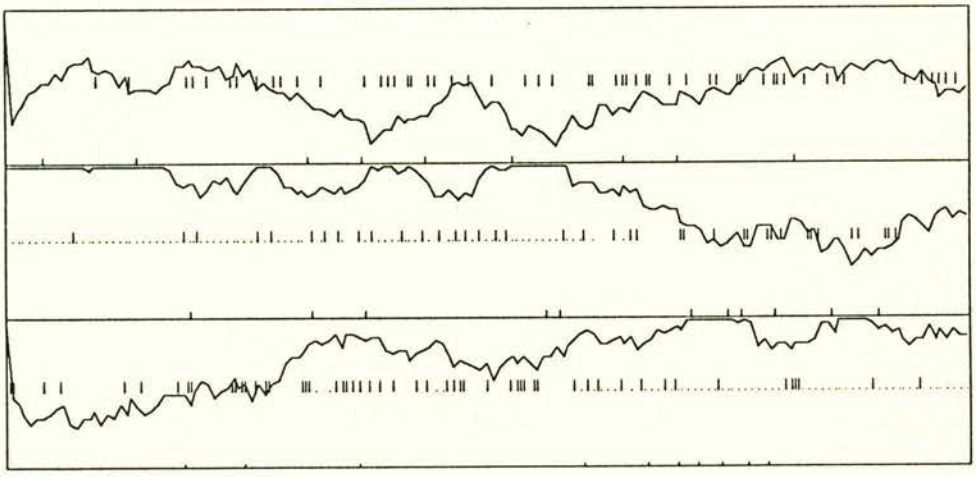
## Determination of coding regions in sequence from position 16,808 to 19,094

The Shepard RNY preference method was used to predict open reading frames in (**A**) the forward direction and (**B**) the reverse direction. The three panels represent the three possible reading frames. The short vertical lines along the central and bottom lines of each panel represent termination and start codons respectively. The curve indicates the extent to which the three bases of the codon are of the form: purine; either; pyrimidine, and thus have a high likelihood of being coding DNA. The predicted open reading frames are shown as the open boxes below the plot labelled A.

A




B



# TABLE 4.2

## Features of the initiation regions of the ORFs in the sequenced *rfb* region of strain M32

Capital letters for bases indicate potential Shine-Dalgarno sites and the presumed start codon. The termination codon of the preceding gene, if it is in the region shown, is indicated by underlining

**Regions of translational initiation in the *rfb* region of strain M32**

| ORFs | | Space between SD and first codon |
|---|---|---|
| *orf6.1* (M32) | tctatgaatGAGaaatggaaATGaaaacgcgta | 8 |
| *orf7.0* (M32) | gattaaAGGttattaataaaATGaacgttatta | 11 |
| *orf7.9* (M32) | gattcGAGttttatcctaatATGtttagaaata | 12 |
| *orf9.6* (M32) | ttaataatatGAGatgtgttATGgtatttttag | 7 |
| *orf10.8* (M32) | gttactcccacttaaaaataATGgttgaattaa | |
| *orf11.9* (M32) | ttgagcttttggactaaaaaATGccctccttta | |
| *orf17.4* (M32) | gagacatcatcGGAGttatgATGgaaaaaaact | 5 |

intergenic gap between *orf7.9* and *orf9.6*, and have Berg-von Hippel indices of 35.3 and 4.5 respectively. The Berg-von Hipple index indicates the *in vitro* performance of a promoter, and the lower the value the stronger the promoter ranks (O'Neill, 1989). In strain LT2 there are two potential promoters in front of the first *rfb* gene (*rfbB*) with the mRNA startpoints at position 2803 and 3927 (Jiang *et al.*, 1991). These two potential promoters have Berg-von Hippel indices of 7.3 and 4.8 respectively, and locate in the region which is essentially identical in strain M32. We suggest that the *rfb* region of strain M32 is under the same promoter as strain LT2, and may have an additional internal promoter in the intergenic gap between *orf7.9* and *orf9.6*. These promoters are not strong enough to function without positive controls, since the Berg-von Hipple indices for all the four potential promoters are too high.

## 4.5 PROTEINS ENCODED BY THE ORFS

The nucleotide sequence of each ORF was translated into an amino acid sequence using the program NIP (Staden, 1982), and the molecular weight, hydrophobicity and charge distribution of the protein were also determined by the same program. The programs PEPPLOT of GCG (Devereux *et al.*, 1984), and CHOU (Chou and Fasman, 1974) were used to predict secondary structure, potential transmembrane segments and charge distribution. Program ALCOM (Klein *et al.*, 1985) was also used to identify potential transmembrane segments, which were then further checked using the algorithm described by Eisenberg *et al.* (1984). The algorithm of Eisenberg *et al.* was used to determine the hydrophobic moment of each transmenbrane segment. Hydrophobic moment is a measure of the amphilicity of a protein segment, and in

general, the segments fall into 4 classes: (1) monomeric transmembrane anchors, which have the highest hydrophobicity and low hydrophobic moment, (2) paired segments or segments bundled together in membranes, which have a high hydrophobic moment and low hydrophobicity, (3) segments of the surface-seeking proteins, which have a highest hydrophobic moment, and (4) segments from globular proteins, which have the lowest hydrophobicity and hydrophobic moment (Eisenberg *et al.*, 1984).

The deduced protein product of *orf6.1* (represented as *orf6.1p*) contains 292 amino acids and has a molecular weight of 32,441 daltons. This protein has a low hydrophobicity of 1.4, and secondary structure predictions indicate it is composed of both α-helices and β-sheets (Figure 4.7).

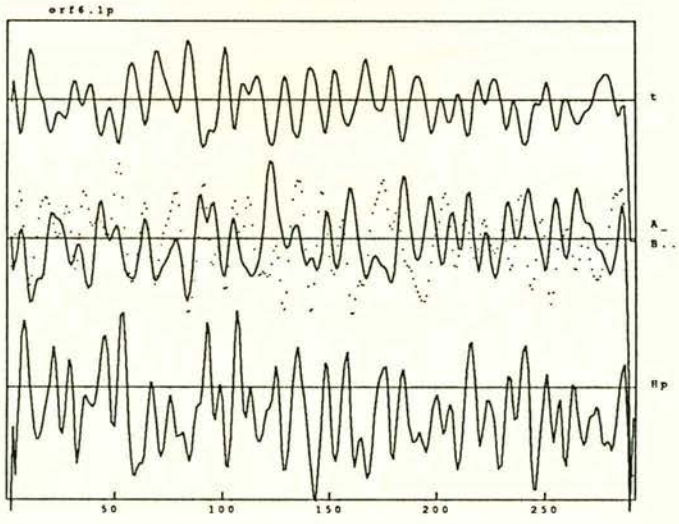
*orf7.0p* has 176 amino acids with a molecular weight of 20,170 daltons. This protein has a hydrophobicity of -4.1, indicating that it is a hydrophilic protein. Secondary structure predictions indicate that it contains both α-helices and β-sheets (Figure 4.7).

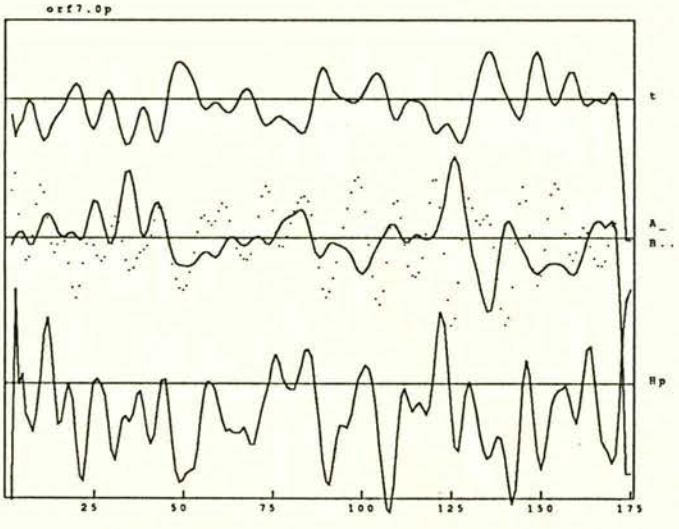
*orf7.9p* is 512 amino acid residues in length, and has a molecular weight of 58,187 daltons. It is a hydrophobic protein with a hydrophobicity of 207.9. It contains 12 transmembrane segments as predicted by programs ALOM (Table 4.3), PEPLOT and the algorithm of Eisenberg *et al.* (1984). The hydrophobic moment of each transmembrane segment was calculated based on the algorithm of Eisenberg *et al.* (1984). The secondary structure of the transmembrane segment can not be predicted by currently available programs because these programs consider that β-sheet conformation is highest for Val, Ile, Tyr, Phe, Trp, Leu which are the most hydrophobic residues

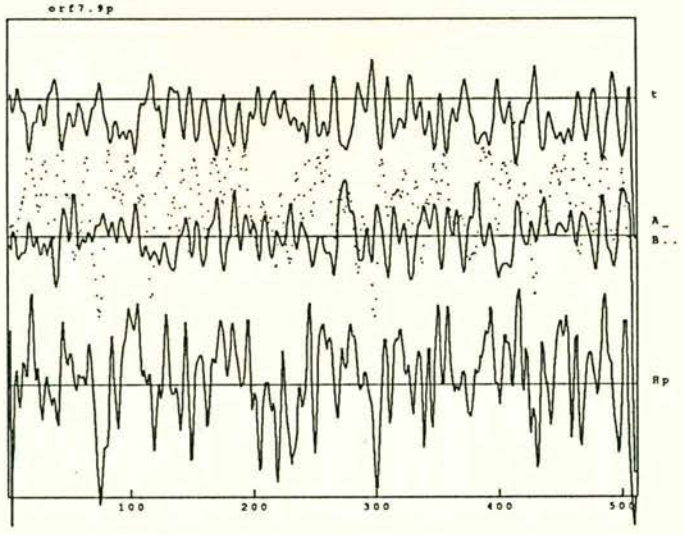
# FIGURE 4.7

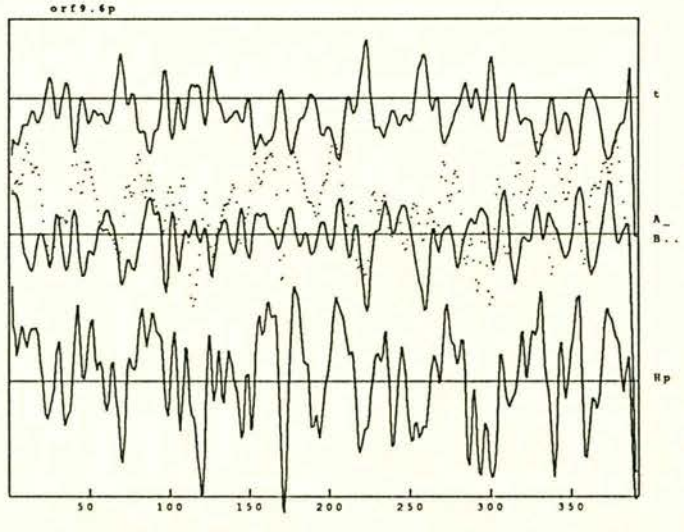
## Secondary structure predictions for the proteins encoded by the sequenced *rfb* region of strain M32

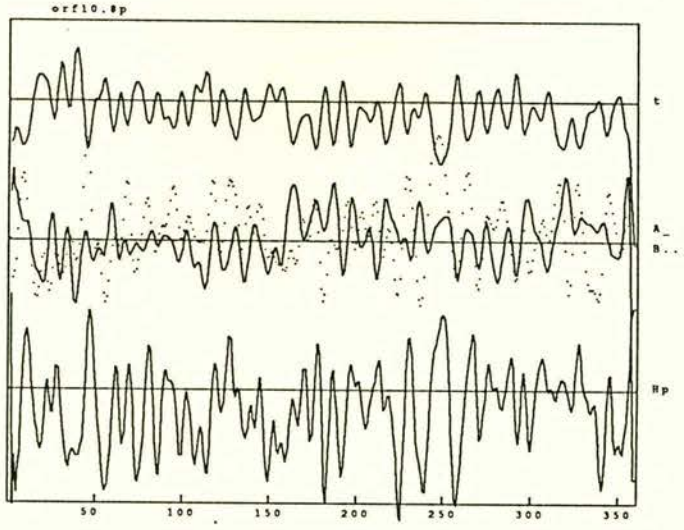
Hydrophobicity and secondary structure predictions of the deduced amino acid sequence of the sequenced M32 *rfb* genes. The uppermost panel represents predicted turns using the algorithm of Chou and Fasman (1978): the curve rises above the central line when a turn promoter is encountered, and drops below the line at a turn inhibitor. The middle panel represents a prediction of α-helix and β-pleat structure using the algorithm of Chou and Fasman (1978): the dashed and solid lines are β-pleat and α-helix respectively. The bottom panel represents the hydrophobicity of the protein calculated using the algorithm of Kyte and Doolittle (1982): the curve rises above the line when the sequence is hydrophobic, and drops when hydrophilic.

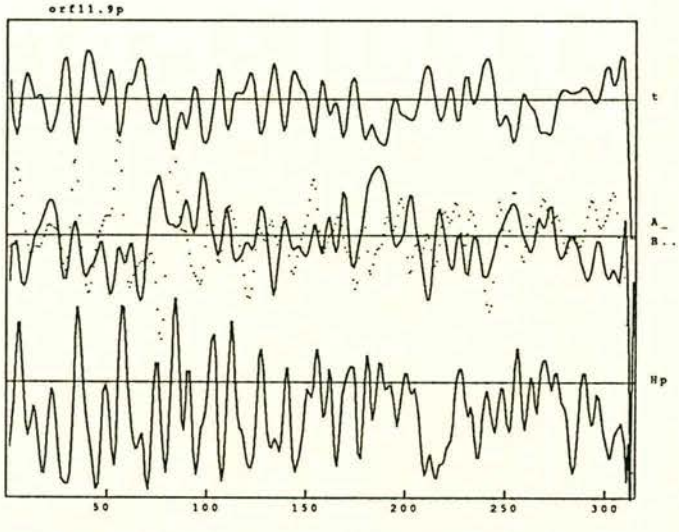

orf6.1p
t
A_
B..
Hp
50
100
150
200
250

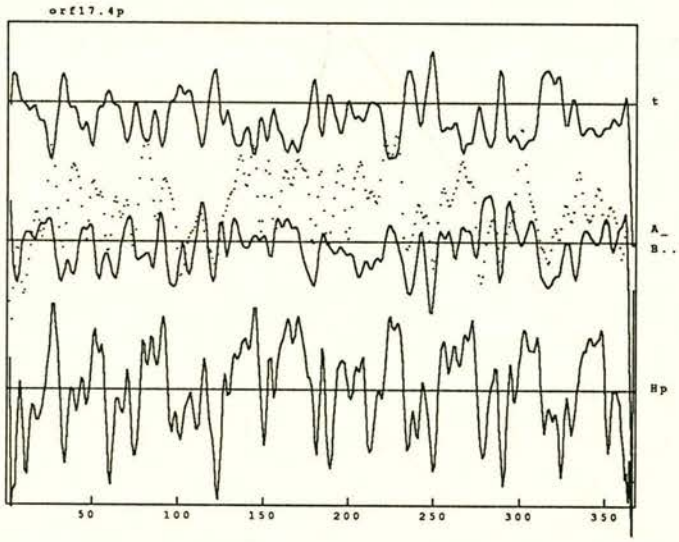

orf7.0p
t
A_
B..
Hp
25
50
75
100
125
150
175


orf7.9p
t
A-
B..
Hp
100
200
300
400
500


orf9.6p
t
A-
B..
Hp
50
100
150
200
250
300
350


orf10.9p
t
A_
B..
Hp
50
100
150
200
250
300
350


orf11.9p
t
A_
B..
Hp
50
100
150
200
250
300


orf17.4p
t
A_
B..
Hp
50
100
150
200
250
300
350

# TABLE 4.3

## Transmembrane segments of *orf7.9p*, *orf9.6p* and *orf17.4p*

Transmembrane segments were predicted using program ALOM based on the method of Klein *et al.* (1985).

**Table 4.3**

***orf7.9p*(M32): Residues, 1 - 512; Total of 512**

```
INTEGRAL  P:I odds = 6.7E-07  Transmembrane    93 - 109 ( 83 - 119)
INTEGRAL  P:I odds = 1.3E-04  Transmembrane   403 - 419 (398 - 426)
INTEGRAL  P:I odds = 3.3E-04  Transmembrane   380 - 396 (377 - 402)
INTEGRAL  P:I odds = 5.9E-04  Transmembrane   168 - 184 (164 - 199)
INTEGRAL  P:I odds = 2.6E-03  Transmembrane   348 - 364 (342 - 368)
INTEGRAL  P:I odds = 6.8E-03  Transmembrane   441 - 457 (439 - 465)
INTEGRAL  P:I odds = 6.8E-03  Transmembrane   473 - 489 (469 - 495)
INTEGRAL  P:I odds = 1.3E-02  Transmembrane     6 -  22 (  6 -  23)
INTEGRAL  P:I odds = 1.7E-02  Transmembrane   253 - 269 (250 - 271)
INTEGRAL  P:I odds = 2.4E-02  Transmembrane   272 - 288 (270 - 290)
INTEGRAL  P:I odds = 4.6E-02  Transmembrane   307 - 323 (307 - 324)
INTEGRAL  P:I odds = 5.7E-02  Transmembrane    44 -  60 ( 44 -  61)
PERIPHERAL  P:I odds = 4.9E+00
```

***orf9.6p*(M32): Residues, 1 - 396; Total of 396**

```
INTEGRAL  P:I odds = 3.2E-05  Transmembrane    85 - 101 ( 76 - 109)
INTEGRAL  P:I odds = 3.6E-05  Transmembrane   201 - 217 (196 - 223)
INTEGRAL  P:I odds = 4.6E-05  Transmembrane     4 -  20 (  1 -  28)
INTEGRAL  P:I odds = 5.7E-05  Transmembrane   158 - 174 (151 - 176)
INTEGRAL  P:I odds = 1.1E-04  Transmembrane   322 - 338 (318 - 342)
INTEGRAL  P:I odds = 3.6E-03  Transmembrane   370 - 386 (366 - 393)
INTEGRAL  P:I odds = 3.8E-03  Transmembrane   271 - 287 (265 - 291)
INTEGRAL  P:I odds = 4.7E-03  Transmembrane    45 -  61 ( 41 -  64)
INTEGRAL  P:I odds = 6.0E-02  Transmembrane   346 - 362 (345 - 363)
INTEGRAL  P:I odds = 6.3E-02  Transmembrane   180 - 196 (176 - 196)
PERIPHERAL  P:I odds = 1.0E+01
```

***orf17.4p*(M32): Residues, 1 - 368; Total of 368**

```
INTEGRAL  P:I odds = 5.0E-06  Transmembrane   159 - 175 (150 - 182)
INTEGRAL  P:I odds = 7.1E-05  Transmembrane   132 - 148 (126 - 152)
INTEGRAL  P:I odds = 4.8E-04  Transmembrane   259 - 275 (252 - 279)
INTEGRAL  P:I odds = 5.1E-04  Transmembrane    80 -  96 ( 76 -  98)
INTEGRAL  P:I odds = 1.1E-03  Transmembrane   334 - 350 (331 - 353)
INTEGRAL  P:I odds = 2.0E-03  Transmembrane   218 - 234 (214 - 235)
INTEGRAL  P:I odds = 8.9E-03  Transmembrane   294 - 310 (293 - 318)
INTEGRAL  P:I odds = 8.1E-01  Transmembrane   195 - 211 (195 - 211)
PERIPHERAL  P:I odds = 1.1E+00
```

(Green and Flanagan, 1976; Wallace *et al.*, 1986). The hydrophobic moment was calculated using angles of rotation of 100° or 180°, which are the appropriate values for α-helix and β-sheet structures respectively (Eisenberg *et al.*, 1984), and is listed in Table.4.4 together with the mean hydrophobicity. In both cases, twelve transmembrane segments were predicted, and most of them (9 and 8 for α-helix and β-sheet structures respectively) are predicted to be 'multimeric' transmembrane segments based on their hydrophobicity and hydrophobic moment values (Eisenberg *et al.*, 1984). The large proportion of transmembrane segments leaves little space for hydrophilic sequence of any significant length.

*orf9.6p* is also a hydrophobic protein with the hydrophobicity of 161.2. It is 391 amino acid residues in length, and has a molecular weight of 44,070 daltones. The algorithm of Eisenberg *et al.* (1984) predicts 10 and 11 transmembranes using angles of rotation of 100° or 180° respectively (Table 4.4), while the program ALOM predicts 10.

*orf10.8p* contains 361 amino acids and has a predicted molecular weight of 40,673. It is a hydrophilic protein with a hydrophobicity of 25.3. Secondary structure predictions indicate that it has both α-helix and β-sheet structures (Figure 4.7)

*orf11.9p* is 315 amino acids in length with a molecular weight of 35,901. The hydrophobicity of this protein is 1.1, indicating it is a hydrophilic protein. Second structure predictions indicate it consists of both α-helices and β-sheets (Figure 4.7)

*orf17.4p* is the third hydrophobic protein found in the sequenced group E1 *rfb* region. It consists of 368 amino acids with a molecular weight

# TABLE 4.4

## Predicted transmembrane segments and their hydrophobic moments

The prediction of transmembrane segments and calculation of their hydrophobic moments were carried out using the algorithm of Eisenberg *et al.* (1984). The hydrophobic moment was calculated using angles of rotation of 100° and 180°, which are the appropriate values for α-sheet and β-helix structures, respectively.

| Sequence (Position) | mean hydrophobicity | hydrophobic moment |
|---|---|---|
| ***orf7.9p*** | | |
| 100°: | | |
| F T S V L I L I T N F ( 15 to 25): | 0.662 | 0.358 |
| V L N L A E L G I S T (52 to 62): | 0.456 | 0.326 |
| G L L F F L L L N S I (102 to 112): | 0.780 | 0.221 |
| I L Q I F L M I C G V (168 to 178): | 0.826 | 0.321 |
| I T V I T S F S S Y M (261 to 271): | 0.481 | 0.322 |
| L I S A I A S P L G A (280 to 290): | 0.635 | 0.301 |
| I F C A Y M F Y I S S (320 to 330): | 0.623 | 0.164 |
| L S Q Q L V L L L S V (344 to 354): | 0.491 | 0.240 |
| I I N I I V S L V L V (386 to 396): | 0.902 | 0.334 |
| N I I V V M I L K P V (412 to 422): | 0.629 | 0.302 |
| F L L S I F F I T R A (450 to 460): | 0.574 | 0.310 |
| I Q V C M S S I T P L (473 to 483): | 0.426 | 0.341 |
| 180°: | | |
| L G S Q I G L F T S V (8 to 18): | 0.497 | 0.294 |
| L V V Q L M G V L N L (45 to 55): | 0.634 | 0.421 |
| I G L L F F L L L N S (101 to 111): | 0.780 | 0.233 |
| Y I L Q I F L M I C G (167 to 177): | 0.752 | 0.270 |
| I S V F L G I T V I T (255 to 265): | 0.794 | 0.282 |
| A S P L G A Y I G A K (285 to 295): | 0.360 | 0.276 |
| L G I F C A Y M F Y I (318 to 328): | 0.795 | 0.215 |
| L L L S V N C F V L I (350 to 360): | 0.754 | 0.327 |
| I N I I V S L V L V H (387 to 397): | 0.740 | 0.271 |
| S N I I V V M I L K P (411 to 421): | 0.515 | 0.231 |
| L S I F F I T R A K V (452 to 462): | 0.331 | 0.630 |
| L D F F I Q V C M S S (469 to 479): | 0.429 | 0.511 |

| Sequence (Position) | mean hydrophobicity | hydrophobic moment |
|---|---|---|
| ***orf9.6p*** | | |
| 100°: | | |
| I A I G I Y I I C G I (8 to 18): | 0.946 | 0.224 |
| G D V I M V V S T G V (39 to 49): | 0.561 | 0.317 |
| I I S A F L I F A L G (84 to 94): | 0.925 | 0.140 |
| I S I T N K V V I A V (148 to 158): | 0.499 | 0.292 |
| N L I L L I L C F V S (174 to 184): | 0.7818 | 0.318 |
| S I I C I I M F F I A (201 to 211): | 0.968 | 0.248 |
| V F T A I P F A V M Y (228 to 238): | 0.739 | 0.194 |
| D L F S P I L F I I P (266 to 276): | 0.709 | 0.377 |
| L G Y V G V F L H G L (315 to 325): | 0.712 | 0.230 |
| I I S L I F V A M R K (350 to 360): | 0.411 | 0.491 |
| F M M I F Y F I Y K L ( 375 to 385): | 0.699 | 0.438 |
| 180°: | | |
| L A W F I A I G I Y I (4 to 14): | 0.960 | 0.384 |
| G D V I M V V S T G V (39 to 49): | 0.561 | 0.228 |
| I S A F L I F A L G V (85 to 95): | 0.898 | 0.264 |
| I S I T N K V V I A V (148 to 158): | 0.499 | 0.505 |
| F V S F Y M L N Y K I (182 to 192): | 0.418 | 0.304 |
| I K V S I I C I I M F (198 to 208): | 0.765 | 0.453 |
| I P F A V M Y E T H F (232 to 242): | 0.481 | 0.437 |
| L D L F S P I L F I I (265 to 275): | 0.795 | 0.276 |
| G Y V G V F L H G L I (316 to 326): | 0.741 | 0.270 |
| F M M I F Y F I Y K L (375 to 385): | .0.700 | 0.306 |
| ***orf17.4p*** | | |
| 100°: | | |
| L I I G S I A L L I N (85 to 95): | 0.803 | 0.272 |
| A Y I F V S I L S V I (136 to 146): | 0.825 | 0.343 |
| L L S V L V A I S V L (161 to 171): | 0.829 | 0.161 |
| L N F I C Q V L T G M (192 to 202): | 0.500 | 0.456 |
| I Y N L L N F Y V F I (215 to 225): | 0.664 | 0.302 |
| I M S V I N A L I G I (255 to 265): | 0.767 | 0.399 |
| V Y V I L D I I A S I (302 to 312): | 0.776 | 0.448 |
| I T K W S V P V Y I A (329 to 339): | 0.455 | 0.184 |
| 180°: | | |
| F V L V G H I N N D I (27 to 37): | 0.435 | 0.421 |
| V T Y L I I G S I A L (82 to 92): | 0.770 | 0.255 |
| W F L F A Y I F V S I (132 to 142): | 0.907 | 0.244 |
| L S V L V A I S V L L (162 to 172): | 0.829 | 0.396 |
| L N F Y V F I L L T V (219 to 229): | 0.775 | 0.470 |
| I N A L I G I Y A V F (259 to 269): | 0.788 | 0.404 |
| V Y V I L D I I A S I (302 to 312): | 0.776 | 0.424 |
| T K W S V P V Y I A L (330 to 340): | 0.425 | 0.549 |

of 42,180. Its hydrophobicity is 150.5. Program ALOM predicts that it has 10 transmembrane segments, and the algorithm of Eisenberg *et al.* (1984) also predicts 10 using angles of rotation of 100° or 180° (Table 4.4).

## 4.6 SIMILARITY BETWEEN *RFB* GENES OF STRAIN M32 AND *RFB* GENES FROM OTHER GROUPS

The nucleotide and amino acid sequences of each ORF found in the sequenced group E1 *rfb* region were compared to these of the *rfb* genes from other groups studied in this laboratory using programs SEQH (Kanehisa, 1984) and BESTFIT (GCG package, Deveuraux *et al.*, 1984). DNA and amino acid alignments were carried out using the program COMPARE ( P. R. Reeves, personal communication). The percentages of similarity and identity given below were derived using BESTFIT and the parameters: gap weight, 3.0; gap length weight, 0.1. Comparison of structure was done based on the structure predicted by PEPPLOT and CHOU.

### 4.6.1 *ORF6.1* AND *ORF7.0*

*orf6.1* of strain M32 is almost identical to *orf6.1* of strain LT2 at both DNA and amino acid levels with 94.1% and 94.8% identity respectively. Figure 4.8 shows the alignment of these two genes at both DNA and amino acid levels. Among the 52 (5.9%) base substitutions, 37 (4.3% of total) are in the last 135 bases. Alignment of these two amino acid sequences indicates that of the 15 (5.2%) amino acid substitutions, 13 are within the last 34 amino acids of the carboxyl terminus. We conclude that these two genes have the same function.

# FIGURE 4.8

## Comparison of *orf6.1* of strains M32 and LT2

Alignment of the DNA and amino acid sequences of *orf6.1*(M32) with those of *orf6.1*(LT2). The alignment is based on the comparison of the amino acid sequences using program BESTFIT with a gap weight of 3.0 and a gap length weight of 0.1. The program COMPARE was used to generate this alignment. The sequences of *orf6.1*(M32) are on the top, and those of *orf6.1*(LT2) are only shown beneath where different from those of *orf6.1*(M32).

```
    MetLysThrArgLysGlyIleIleLeuAlaGlyGlySerGlyThrArgLeuTyrProVal
1   ATGAAAACGCGTAAGGGCATTATTTTAGCGGGGGGCTCCGGCACCCGTCTTTATCCGGTG   60

    ThrMetAlaValSerLysGlnLeuLeuProIleTyrAspLysProMetIleTyrTyrPro
61  ACCATGGCGGTAAGTAAGCAATTGCTACCAATTTATGATAAACCGATGATTTACTATCCC   120

    LeuSerThrLeuMetLeuAlaGlyIleArgAspIleLeuIleIleSerThrProGlnAsp
121 CTTTCCACACTTATGCTGGCAGGTATTCGGGATATCCTGATCATCAGTACGCCACAGGAC   180
          G               C

    ThrProArgPheGlnGlnLeuLeuGlyAspGlySerGlnTrpGlyLeuAsnLeuGlnTyr
181 ACGCCGCGTTTTCAACAACTGCTGGGAGACGGCAGCCAGTGGGGGCTGAATCTTCAATAT   240

    LysValGlnProSerProAspGlyLeuAlaGlnAlaPheIleIleGlyGluGluPheIle
241 AAAGTACAGCCAAGCCCGGATGGCTTAGCACAGGCGTTTATTATTGGTGAAGAGTTCATT   300

    GlyAsnAspAspCysAlaLeuValLeuGlyAspAsnIlePheTyrGlyHisAspLeuPro
301 GGTAATGATGATTGTGCATTAGTACTGGGTGACAATATCTTCTATGGTCATGATTTACCA   360
      C                   G
      His

    LysLeuMetGluAlaAlaValAsnLysGluSerGlyAlaThrValPheAlaTyrHisVal
361 AAGTTAATGGAAGCTGCCGTTAATAAAGAAAGTGGTGCTACCGTCTTTGCTTATCATGTA   420
                                                      C

    AsnAspProGluArgTyrGlyValValGluPheAspGlnSerGlyThrAlaValSerLeu
421 AACGATCCAGAGCGCTACGGTGTGGTTGAGTTTGACCAAAGTGGCACAGCCGTTAGTCTG   480
          G                                   AG
                                              Lys

    GluGluLysProLeuGlnProLysSerAsnTyrAlaValThrGlyLeuTyrPheTyrAsp
481 GAGGAAAAACCGTTACAACCGAAGAGTAATTACGCGGTAACGGGGCTGTATTTTTACGAT   540
      A       A                                                T

    AsnSerValValGluMetAlaLysAsnLeuLysProSerAlaArgGlyGluLeuGluIle
541 AACAGCGTGGTGGAGATGGCGAAAAATCTTAAGCCTTCCGCTCGCGGTGAGTTAGAAATC   600
      T

    ThrAspIleAsnArgIleTyrMetGluGlnGlyArgLeuSerValAlaMetMetGlyArg
601 ACGGATATTAACCGTATCTATATGGAGCAGGGAAGATTGTCTGTCGCTATGATGGGGCGC   660

    GlyTyrAlaTrpLeuAspThrGlyThrHisGlnSerLeuIleGluAlaSerAsnPheIle
661 GGTTATGCTTGGCTGGATACGGGAACGCATCAGAGTTTGATAGAGGCCAGTAATTTTATT   720
          C           A  G

    AlaThrIleGluGluArgGlnGlyLeuLysValSerCysProGluGluIleAlaTyrArg
721 GCAACCATCGAAGAACGCCAGGGTCTGAAAGTGTCTTGCCCGGAAGAGATCGCTTATAGA   780
                        G  A       C                 A T C T
                                                     Phe
```

```
    LysGlyPheIleAspAlaGluGlnIleLysAsnLeuAlaLysProLeuSerLysAsnAla
781 AAAGGGTTTATTGATGCAGAGCAGATTAAAAATCTGGCTAAACCGTTGTCGAAGAATGCT  840
      AAT     AA     C A   G   T G A     CGGG  A  A  A  A    A
      Asn     Asn    Gln   ValIleGlu     Gly                Asp

    TyrGlyGlnTyrLeuLeuAsnMetIleLysGlyTyr
841 TATGGGCAGTATCTCCTGAACATGATTAAAGGTTAT  876
       CA A   T G     G   G G       TA
        Lys         Lys   Val      Leu
```

*orf7.0p* of strain M32 is 7 amino acids shorter than *orf7.0p* of strain LT2, and can be aligned to *orf7.0p* of strain LT2 with 3 gaps of 1, 1, and 5 amino acid residues (Figure 4.9). Of the 176 amino acids which can be aligned, 120 (68.2%) are identical, and of the 56 (31.8%) which have been substituted, 19 (10.8% of total residues) show a conservative change. The structural predictions of these two proteins show the same pattern except for the last 10 amino acid residues at the C-terminal end (Figure 4.10). Again we conclude that these two genes have the same function.

### 4.6.2 *ORF7.9*

*orf7.9* has 12 transmembrane segments distributed throughout its length, suggesting that the whole protein is embedded in the membrane leaving no regions which are long enough to form a hydrophilic domain for the reaction sites for hydrophilic reactions. This ORF is the only *rfb* gene encoding a twelve transmembrane segment protein in strain M32, and *orf12.8* which is located in the same relative position is the only such gene in strain LT2 (Jiang *et al.*, 1991). The distribution of the twelve transmembrane segments along these two proteins is nearly identical (Figure 4.11). These data suggest that *orf7.9* of strain M32 and *orf12.8* of strain LT2 may have a similar function. *orf7.9p* is 82 residues longer than *orf12.8p*, and there is only 50.5% similarity and 20.8% identity between these two genes at amino acid level (Figure 4.12).

All the *rfb* gene clusters so far studied have a gene encoding a protein with 12 transmembrane segments. These genes are *orf11.45* in group C1, and *orf12.6* in group C2 (Lee *et al.*, 1992; Brown *et al.*, 1992). The alignments of *orf7.9p* with *orf11.45p* and *orf12.6p* are shown in

# FIGURE 4.9

## Comparison of *orf7.0* of strains M32 and LT2

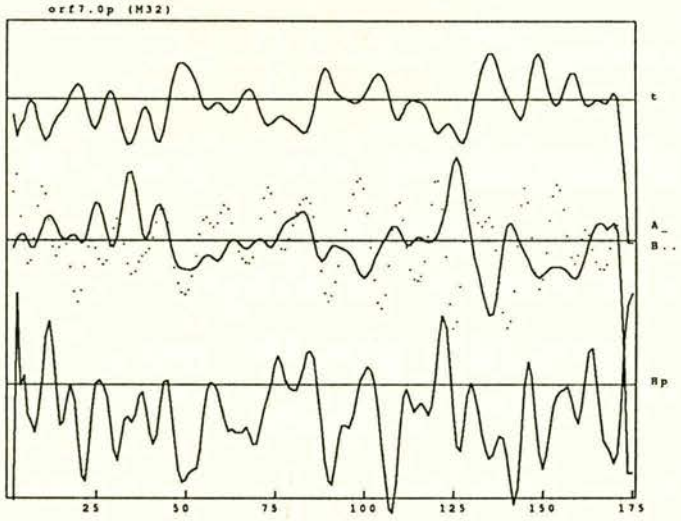
Alignment of the DNA and amino acid sequences of *orf7.0*(M32) with those of *orf7.0*(LT2). The alignment is based on the comparison of the amino acid sequences using program BESTFIT with a gap weight of 3.0 and a gap length weight of 0.1. The program COMPARE was used to generate this alignment. The sequences of *orf7.0*(M32) are on the top, and those of *orf7.0*(LT2) are only shown beneath where different from those of *orf7.0*(M32).

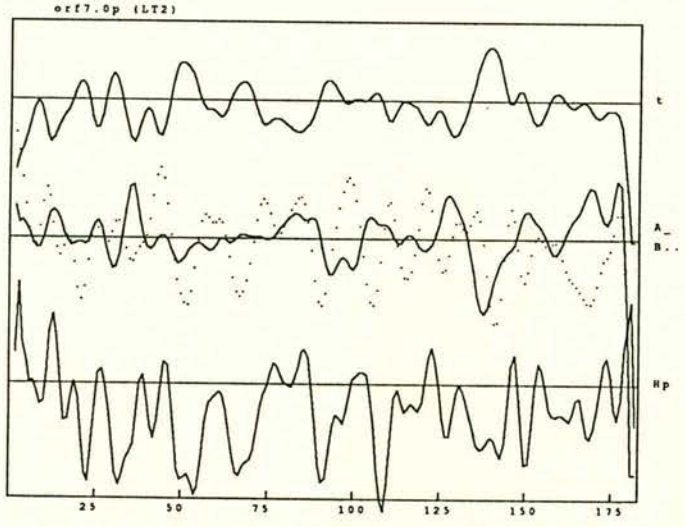
```
      MetAsnValIleLysThrGluIleProAspValLeuIlePheGluProLysValPhe
1     ATGAACGTTATTAAAACAGAAATTCCTGATGTATTAATTTTTGAACCTAAAGTGTTT   60
GTG    TT  G          C   A  A     C  G  C  A  G        T
Val   Ile             Ala                  Leu

      SerAspGluArgGlyPhePheMetGluSerPheAsnGlnLysValPheGluGluAlaVal
61    AGTGATGAACGTGGGTTCTTTATGGAAAGTTTTAATCAGAAAGTATTTGAGGAAGCGGTT  120
G C     GA G  A      T T       A   C   C GACC     A  GTT A
Gly                  Phe      Tyr      GlnThr         LeuIle

      GlyArgLysIleGluPheValGlnAspAsnHisSerLysSerThrLysGlyValLeuArg
121   GGTCGAAAGATTGAATTTGTTCAGGATAATCATTCAAAATCAACTAAAGGTGTGTTACGT  180
  A  T  AG  AC         A                 C AA  GAAC  AC CA A
         ValThr                           Lys   Asn

      GlyLeuHisTyrGlnValGluProTyrAlaGlnGlyLysLeuValArgCysIleAlaGly
181   GGTTTACATTATCAAGTTGAACCTTATGCTCAAGGGAAGCTTGTACGCTGTATAGCGGGA  240
  GC      T   GAGA G GAAA    A  G      T A  T  T   GCT TC  T
        Phe   ArgGlyGluAsn                          AlaVal

      GluValPheAspValAlaValAspIleArgAsnAspSerGluThrPheGlyLysTrpVal
241   GAAGTTTTTGATGTTGCTGTAGATATTCGCAACGATTCCGAAACGTTTGGTAAATGGGTT  300
  G             G  C     C  A  A  A  GCCT  T      C
                            LysGlu   Pro         Gln

      GlyValAsnIleSerSerGluAsnLysArgGlnLeuTrpIleProGluGlyPheAlaHis
301   GGTGTCAATATTTCTTCTGAAAACAAAAGGCAGTTGTGGATACCTGAAGGTTTTGCTCAT  360
     A   C G   G    G  T  GC A   C T     T  A
         Leu   Ala

      GlyPheLeuValLeuSerGluGluAlaGluPheValTyrLysThrSerAsnTyrTyrSer
361   GGGTTCTTAGTATTAAGTGAAGAAGCTGAATTTGTTTATAAGACATCAAACTATTACTCT  420
  T  TG TACTC T     GT T  A  G   C G  C  AG  A T  T        A
       ValThr         Tyr         Leu      AlaThr

      GlyGluHisGluArgGlyIleIleTrpAsnAspProAspIleAsnIleThrTrpGlyIle
421   GGCGAACATGAAAGAGGTATTATTTGGAATGATCCTGATATTAATATTACATGGGGAATA  480
CCTTC TCG   G TA C   C A         GAG CA  AGG    GA    CCTT T
ProSerSer   GlySer   Leu         GluAla   Gly   Glu   ProPhe

      AspSerProIleLeuSerLeuLysAspLysIleHisLysGlyLeuValLys
481   GATAGTCCAATTCTTTCATTAAAAGATAAGATTCATAAAGGTTTAGTAAAG  531
TC CAG TGCC GAGCTT C GC A AG TGC GCAGC CC    C GG T
SerGlnLeuProGluLeuSerAlaLysAspAlaAlaAlaPro   LeuAsp
```

# FIGURE 4.10

## Comparison of the predicted secondary structures of *orf7.0p* of strains M32 and LT2

The hydrohpobicity and secondary structure are predicted using program CHOU. The uppermost panel represents predicted turns using the algorithm of Chou and Fasman (1978): the curve rises above the central line when a turn promoter is encountered, and drops below the line at a turn inhibitor. The middle panel represents a prediction of α-helix and β-pleat structure using the algorithm of Chou and Fasman (1978): the dashed and solid lines are β-pleat and α-helix respectively. The bottom panel represents the hydrophobicity of the protein calculated using the algorithm of Kyte and Doolittle (1982): the curve rises above the line when the sequence is hydrophobic, and drops when hydrophilic.

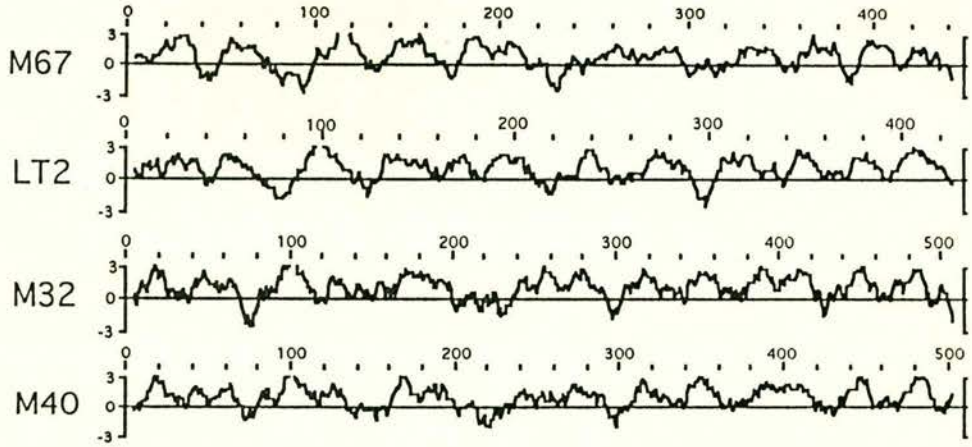

orf7.0p (M32)
t
A
B
Hp
25
50
75
100
125
150
175

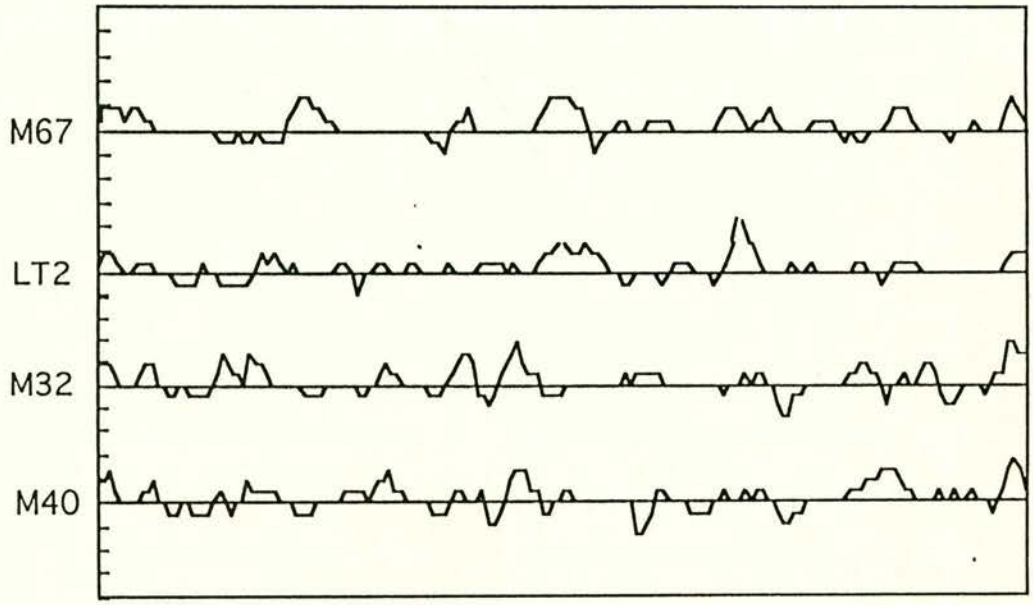

orf7.0p (LT2)
t
A
B
Hp
25
50
75
100
125
150
175

# FIGURE 4.11

## Structural comparison of *orf7.9p* with related proteins

Comparison of the predicted hydrophobicity profile (above) and the charge distribution (below) of the deduced amino acid sequences of *orf7.9* of strain M32, *orf12.8* of strain LT2, *orf12.6* of strain M40 (group C2) and *orf11.4* of strain M40 (group C1). The hydrophobicity was plotted using program PEPPLOT setting the window at 9 amino acids. Hydrophobicity and hydrophilic regions are indicated when the curve is in the upper and lower half respectively. The charge distribution was plotted using program NIP, and the positive and negative charge is indicated when the curve is above and below the central line, respectively. This figure was adopted from P. Brown's PhD thesis.


M67
LT2
M32
M40
0
100
200
300
400
500
3
0
-3


M67
LT2
M32
M40

# FIGURE 4.12

## Comparison of deduced amino acid sequence of *orf7.9* with those of related genes

The deduced amino acid sequence of *orf7.9* was aligned with those of *orf12.8* (group B), *orf11.45* (group C1) and *orf1020* (group C2) using the program BESTFIT. Identity is indicated by a vertical line, amino acids with a comparison value greater than or equal to 0.5 are indicated by a colon, and amino acids whose comparison value is greater or equal to 1.0 are indicated by a period.

```
      Gap Weight:  3.000      Average Match:  0.540
   Length Weight:  0.100   Average Mismatch: -0.396

         Quality:  184.9            Length:    453
           Ratio:  0.434              Gaps:     12
Percent Similarity: 50.495   Percent Identity: 20.792

orf12x8.pro x orf7.9_elxpro July 3, 1991 17:25  ..

                .         .         .         .         .
  5 LLKIPSHLIVAGSSWLSKIIIAGVQLASISYLISMLGEEKYAIFSLLTGL 54
    ::: .  | ... :::..::|  ..:   ||:.:.|.:  :.: |:..|
  1 MFRNKVAL.GSQIGLFTSVLILITNFFLRSYFVKVYGADLTGYYLLVVQL 49
                .         .         .         .         .
 55 LVWCSAVDFGIGTGLQNYISECRAKNKSYD.........AYIKSALHLS 94
    :.  . .::||:|:|  .: ..  :... :           |   || :
 50 MGVLNLAELGISTALTYILFKPLHRKENSELRQLYFIIKKIYHFIALGIL 99
                .         .         .         .         .
 95 FIAIIFFIALFYIFSGVISAKYLSSFHEVLQDKTRMLFFTSCLVFSSIGI 144
    .|:::||: |  |..: ||:. |    :|:  .|.: ::     :|. ::
100 VIGLLFFLLLNSIVNASISPENLYITWGVFVISTSLSYL.....YSAQSV 144
                .         .         .         .         .
145 GAIAYKILFAELVGWKANLLNALSYMIGMLGLLY.IYYRGISVDIKLSLI 193
      .| . ::  || : ..| ..|.|::.:: ::: : :: ::.   || :
145 ILTADQNVY..LVKLITGLTRSLAYILQIFLMICGVSFWIVCAIELLSNV 192
                .         .         .         .         .
194 VLYLPVGMISLCYIVYRYIKLYHVKTT....KSHYIAILRRSSGFFLFTL 239
    : .: ...:.|     :| .| .:..|    |.:.|.: .: .  : :|:
193 IQLILFNKLTL....KKYPQLVKLDITDTINKENIISVKSKIKREIKYTF 238
                .         .         .         .         .
240 L....SIVVLQTDYMVISQRLTPADIVQYTVTMKIFGLVFFIYTAILQAL 285
    :    ::||:.|||::||  |. . |. :.  |.::. : |:..||  :|
239 VHKIAGVVVFNTDYLIISVFLGITVITSFSSYMMLIQALAFLISAIASPL 288
                .         .         .         .         .
286 WPICAELRVKQQWKKLNKMIG........VNILLGSLYVVGCTIFIYLFK 327
     :..:.   .: ..|: ::|:        :.|: : :: ::::..|: |:.
289 GAYIGAKLYSDSKEKVIRIISLYNSLFFMLGIFCAYMFYISSSTFVSLWM 338
                .         .         .         .         .
328 EQIFSVIAKDINYQVSILSFMLIGIYFCIRVWCDTYAMLLQSMNYLKILW 377
    :. : |:..:: . :|: :|:||:     |:. |.  : |. |. :.
339 GKEI.VLSQQLVLLLSVNCFVLIA.....RISFDVAKVGLGYMSDIE... 379
                .         .         .         .         .
378 ILVPLQAIIGGIAQWYFSSTLGISGVLLGLIISFALTVFWGLPLTYLIKA 427
     |. |:|||. |. : :   ||:.|:::| |:|  :.|:   |: .. .|
380 .LPLLEAIINIIVSLVLVHYLGLNGIVIGTIVSNIIVVMILKPVFLYKNA 428

428 NKG 430
     |:
429 LKS 431
```

```
        Gap Weight:  3.000      Average Match:  0.540
      Length Weight:  0.100   Average Mismatch: -0.396

            Quality:  364.6             Length:    517
              Ratio:  0.722              Gaps:        6
Percent Similarity: 62.097   Percent Identity: 33.669

orf11.45xc1xpro x orf7.9_e1xpro July 3, 1991 17:36  ..

                .         .         .         .         .
      1 MRTIKAINNFKVDLFITFFLIALGFYLRTIFVSKMGADLTGVMLLFTQLT 50
        |   |.  . .::||...::: .|:||..||. ||||||  ||..||
      1 MFRNKVALGSQIGLFTSVLILITNFFLRSYFVKVYGADLTGYYLLVVQLM 50
                .         .         .         .         .
     51 AYLNLAELGIGVAAASLLYKPLSEGDYAKIKYLTLLLSTIYRYISFLVLL 100
        : ||||||||:.| . :|:|||   : ..:: | :::..||::|.: :|:
     51 GVLNLAELGISTALTYILFKPLHRKENSELRQLYFIIKKIYHFIALGILV 100
                .         .         .         .         .
    101 IGIVIGFGIYFFIDSVNAVSHVFIYWAFFVINTSLTYSYAKHSTLLTANQ 150
        ||::: : :  :::.  ...:::| |:.|||.|||.| |. :|.:|||:|
    101 IGLLFFLLLNSIVNASISPENLYITWGVFVISTSLSYLYSAQSVILTADQ 150
                .         .         .         .         .
    151 QYSVVRKIQGGGKILIIALQILLLVTTHNFLLYLLVETIGVIVQYFIFKN 200
        .  :|: | | .: | . |||:|::.. .|::   :| :: ::|.::|..
    151 NVYLVKLITGLTRSLAYILQIFLMICGVSFWIVCAIELLSNVIQLILFNK 200
                .         .         .         .         .
    201 IINNDIHFKV...VPQSISDDEKTTLKNELKIKIKNMFFHKIGGVLVLNT 247
        :. ....  |   :.:.|..:: ..:|..:| .||  |.|||:||:|:||
    201 LTLKKYPQLVKLDITDTINKENIISVKSKIKREIKYTFVHKIAGVVVFNT 250
                .         .         .         .         .
    248 DYLLVSKFLNLSYVTIYGSYMMVFQVVTVLMSSFVNAITASVGNFLINQN 297
        |||::| ||.:. :| ::|||||::|.:..|:|.:..::.| :|. |..:.
    251 DYLIISVFLGITVITSFSSYMMLIQALAFLISAIASPLGAYIGAKLYSDS 300
                .         .         .         .         .
    298 DDEVTSIAKQFNTVFIALATFISLNMYFLVNDFITSWIGEKFILGNGIVI 347
        .:.|..| . :|.:|: |:.|... :|:  ..|:. |:|..::|:..:|:
    301 KEKVIRIISLYNSLFFMLGIFCAYMFYISSSTFVSLWMGKEIVLSQQLVL 350
                .         .         .         .         .
    348 LMLVNVFISVIRIPCDIFKNATGFFGDVYYPLLEGVVNLFFSALLAFYIG 397
        |: ||.|: : ||. |: | : |:::|: .||||:::|::.| :|. |:|
    351 LLSVNCFVLIARISFDVAKVGLGYMSDIELPLLEAIINIIVSLVLVHYLG 400
                .         .         .         .         .
    398 LPGIIIGTIISNVLITLIAKPLYLYGKMFGRFNALKKYLSFVLKPLI... 444
        | ||:||||:||:::.:| ||::||       ||||.  .|::..||
    401 LNGIVIGTIVSNIIVVMILKPVFLYK......NALKSDNAFIIHELIRSW 444
                .         .         .         .         .
    445 ..FSFVIFAVFYFTREQIIFFKVSN.WFDFISKLTIVSLVSMIIVFAVFY 491
          :|:.::.:|::||..:|   ||| |:||: .:.: |:..::::: ::
    445 FFISIFLLSIFFITRAKVI...VSNEWLDFFIQVCMSSITPLLLLIMIY. 490
                .
    492 ADANFRSFVKRILRVVF 508
          . | | |:::: |:|
    491 ..SVFDSNVRKFMFVMF 505
```

```
          Gap Weight:  3.000       Average Match:  0.540
       Length Weight:  0.100    Average Mismatch: -0.396

             Quality:  176.4              Length:    486
               Ratio:  0.405                Gaps:     13
  Percent Similarity: 47.596   Percent Identity: 18.510

orf1020.pro x orf7.9_elxpro July 3, 1991 17:37  ..

                .         .         .         .         .
      10 VDKAIRYVIFGKIISVLTGLLLIMLISHHLSKDAQGYYYTFNSVVALQII 59
         |. : ...:|..:: ::|.::|  .: . .: |  |||. .  :::: :
       6 VALGSQIGLFTSVLILITNFFLRSYFVKVYGADLTGYYLLVVQLMGVLNL 55
                .         .         .         .         .
      60 FELGLSTVIIQFASHEMSALKYDYSERDIIGESKNKQRYLSLFRLAIKWY 109
          |||:||.:. :  ..:              ..|:.  .  |: :  | |
      56 AELGISTALTYILFKPL.............HRKENSELRQLYFIIKKIY 91
                .         .         .         .         .
     110 AVIALLIILIVGPIGYVFFTQKEGLGVPWQGAWL...LLTIVTAFNIFLV 156
          .|| | ||::| : ::::.   . ::. :. ::   ::.| |.:. :|
      92 HFIA.LGILVIGLLFFLLLNSIVNASISPENLYITWGVFVISTSLS.YLY 139
                .         .         .         .         .
     157 SVLSVAEGSGLITDVNKMRMYQSL...LAGILAVSLLISGFGLYATSAIA 203
         |. ||  ..:   :|  :::. :|   || ||.: |:|:|.::: .:||.
     140 SAQSVILTAD..QNVYLVKLITGLTRSLAYILQIFLMICGVSFWIVCAIE 187
                .         .         .         .         .
     204 ISGTIIFSIFSYKYFKKIFLQSLKHKNKYTEGGISWVNEIFPMQWRIALS 253
         : :.:|  |: |. | : | :| . . | .  . :.  .:.: | ..
     188 LLSNVIQLILFNKLTLKKYPQLVKLDITDTINKENIISVKSKIKREIKYT 237
                .         .         .         .         .
     254 WMSGYFIYFVMTPIAFKYFGAIYAGQLGMSLTLCNMVMATGLAWISTKYP 303
         ::   : .|:  ..: .::.:.:  :  |:. : |::..   :||.  .
     238 FVHK.IAGVVVFNTDYLIISVFLGITVITSFSSYMMLIQALAFLISAIAS 286
                .         .         .         .         .
     304 KWGVMVSNKQLAE.LSKSFKSAVMQSSFFVLTGL.........TGVYISL 343
         .:|. ::.| ..:  .| ::   : .|:|.: |:         .:.::||
     287 PLGAYIGAKLYSDSKEKVIRIISLYNSLFFMLGIFCAYMFYISSSTFVSL 336
                .         .         .         .         .
     344 WL.......................LKLSGSNIGERFLGLQDFFFLSLA 369
         |:                       :.:....:|  :::  :: :|. |
     337 WMGKEIVLSQQLVLLLSVNCFVLIARISFDVAKVGLGYMSDIELPLLE.A 385
                .         .         .         .         .
     370 IIGNHIVACFATYIRAHKTEKMTLASCIMALLTITTMLFV........AY 411
         ||.  :  .:. |:  : .   |:.| |:.::.:..:::         |:
     386 IINIIVSLVLVHYLGLNGIVIGTIVSNIIVVMILKPVFLYKNALKSDNAF 435
                .         .         .
     412 L..EYSRFYMLMYAALTWLYFVPQTYIIFKRFKSSY 445
         :  |. | ::::   | .::|:... :|... . .:
     436 IIHELIRSWFFISIFLLSIFFITRAKVIVSNEWLDF 471
```

Figure 4.12. *orf11.45p* shows 62.1% similarity and 33.7% identity with *orf7.9p*, and is the most related protein to *orf7.9p* among those 12 transmembrane segment proteins. *orf11.45p* is 507 amino acid residues in length, which is 5 residues shorter than *orf7.9p*. Although *orf12.8p* and *orf12.6p* are much shorter than *orf7.9p* and *orf11.45p* in length, the distribution of the predicted transmembrane segments is similar (Figure 4.11).

### 4.6.3 *ORF11.9*

*orf11.9p* shows 75.2% identity and 85% similarity to *orf16.5p* of strain LT2. *orf11.9p* is 1 amino acid residue longer than *orf16.5p*, and these two proteins can be aligned with 2 gaps in *orf16.5p* and *1 gap* in *orf11.9*, all of 1 amino acid residue (Figure 4.13). Of the 83 amino acid substitutions, 31 have undergone a conservative change, and also most of the substitutions are concentrated in the N-terminal half of the protein (Figure 4.13). At the DNA sequence level 694 of the 945 bases (73.4%) are conserved.

*orf16.5* of strain LT2 has been found to encode the rhamnose transferase which catalyzes the linkage between rhamnose and galactose in the O antigen (D. Liu, personal communication). The linkage between rhamnose and galactose of the LT2 O antigen is same as that of the M32 O antigen (Figure 1.3), and it is therefore expected that a gene similar to *orf16.5* will be found in strain M32. The high level of homology between *orf11.9* and *orf16.5* at both DNA and amino acid levels suggests that *orf11.9* is the rhamnose transferase gene in strain M32.

# FIGURE 4.13

## Comparison of *orf11.9* of strain M32 with *orf16.5* of strain LT2

Alignment of the DNA and amino acid sequences of *orf11.9* (M32) with those of *orf16.5* (LT2). The alignment is based on the comparison of the amino acid sequences using program BESTFIT with a gap weight of 3.0 and a gap length weight of 0.1. The program COMPARE was used to generate this alignment. The sequences of *orf11.9* (M32) is on the top, and those of *orf16.5* (LT2) are only shown beneath where different from those of *orf11.9* (M32).

```
    MetProSerPheThrLeuIleValProThrTyrAsnAlaGlyThrAlaArgTrpAsnGlu
1   ATGCCCTCCTTTACTTTAATTGTTCCTACATATAATGCTGGTACTGCTAGATGGAATGAG   60
    XXXATGAAAA A  A      A    C           A  GT GCT T GCCT    TT
       MetLysIle         Ile                  SerLeuTrpPro   Val

    TrpPheHisAlaPheGlnLysGlnThrAsnGlnProSerGlnLeuIleIleIleAspSer
61  TGGTTTCACGCTTTTCAAAAACAAACTAATCAGCCATCACAATTGATAATCATTGATTCA   120
    CT GA GCGAT AAG  GC  ACT TAT   C GATAA TTGA TG T  AGACTCAGGT
    LeuAspAlaIleLys   GlnThrIleTyrProAspLysLeuIleVal   AspSerGly

    SerSerThrAspAspThrArgAsnIleAlaSerThrTyrThrAsnAsnIleIleValIle
121 TCATCGACAGATGACACTAGAAATATTGCAAGCACTTATACCAACAATATTATAGTTATA   180
      TAAAGAT  AACGGT CCGTTAGCCT  GA CTGA A AT TATCA  AT TAA   T
       LysAspGluThrValProLeuAlaSerAspLeuLysAsnIleSer   PheAsn

    SerProSerGluPheAsnHisGlyGlyThrArgAsnLysAlaLeuAlaSerAlaGluGlu
181 TCACCAAGTGAATTTAATCACGGAGGGACGAGAAATAAAGCATTAGCTTCTGCCGAAGAA   240
    GACT T AA  T        T     A  C      TT    G T  AAAAA TCTG  C
    AspSerLysAsp                        Leu   Val   LysThrLeuAsp

    SerAspPheValValPheLeuThrGlnAspAlaIlePheGluAsnLysAsnAlaLeuGlu
241 AGTGATTTTGTTGTTTTTTTAACCCAGGATGCTATATTTGAAAACAAAAATGCCCTGGAA   300
    GC    G  A AA     C    G  A     A  TC C CGG TTCGG    AA TA
    Ala   ValIleIle                     LeuAlaAspSerAsp   IleLys

    GluIleLeuSerLeuPheAsnAspLysAsnValAlaAlaValCysGlyArgGlnLeuPro
301 GAAATATTATCATTATTTAATGACAAAAATGTCGCTGCAGTATGTGGCAGACAGCTTCCT   360
    A TT GG T AT AT   TCA  TCC TTGA A  A  G  T     T     A
    AsnLeuValTyrTyr   Ser   ProLeuIle

    HisHisAspAlaAsnProLeuAlaIleHisAlaArgAsnPheAsnTyrGlyThrGluThr
361 CATCATGATGCTAATCCTCTTGCTATACACGCAAGAAATTTTAATTATGGTACTGAAACT   420
       A A          C     AG G  T  C               A  T AA  T
       Lys               Val                      SerSerLysSer

    LeuIleLysSerLysAsnAspIleLysAsnLeuGlyIleLysThrValPheMetSerAsn
421 CTAATTAAAAGTAAAAATGATATAAAAAATCTCGGCATTAAAACCGTGTTTATGTCAAAT   480
    A TG          GGCA      G    AT G  T        T  A        C
    IleVal         Ala      GluLys

    SerPheAlaAlaTyrArgArgSerValPheGluAlaLeuGlyGlyPheProGluHisThr
481 TCTTTTGCAGCTTATCGGCGGTCTGTTTTTGAAGCACTTGGGGGGTTTCCTGAACATACT   540
           T  C     C  T  C          AGT AA T                  A
                                    Glu   Ser

    IleLeuAlaGluAspMetPheMetAlaAlaArgMetIleGlnAlaGlyTyrLysIleAla
541 ATCCTCGCTGAAGATATGTTTATGGCTGCGAGAATGATCCAGGCGGGGTATAAAATTGCT   600
      T  T  C  G              G  T AG     T        T     GG C  C
                                Lys                      Val

    TyrCysAlaGluAlaSerValArgHisSerHisAsnTyrThrProArgGlnGluPheGln
601 TATTGCGCAGAAGCCAGCGTAAGGCATTCGCATAATTATACCCCGAGACAAGAGTTCCAG   660
      C     T     GGTG      A  C  C                C  G       T  A
                  Val                               Glu

    ArgTyrPheAspThrGlyValPheHisAlaCysAsnProTrpIleGlnArgAspPheGly
661 CGGTATTTTGATACAGGGGTATTTCATGCCTGTAACCCGTGGATTCAGCGTGACTTTGGC   720
      A           T  T           T   TCT
                                     Ser

    GlyAlaGlyGlyGluGlyPheArgPheValLysSerGluIleGlnPheLeuLeuLysAsn
721 GGAGCCGGTGGTGAGGGTTTCCGCTTCGTAAAATCAGAGATTCAATTTCTGCTTAAAAAT   780
                                                  C
```

```
     AlaProPheTrpIleProArgAlaLeuLeuThrThrPheAlaLysPheLeuGlyTyrLys
781  GCACCGTTCTGGATTCCAAGAGCTTTATTAACAACCTTTGCTAAATTTCTGGGTTACAAA   840
                                                        CT

     LeuGlyLysHisTrpGlnSerLeuProLeuSerThrCysArgTyrPheSerMetTyrLys
841  TTAGGCAAGCATTGGCAATCTTTACCGTTGTCTACATGTCGCTATTTTAGCATGTACAAG   900

     SerTyrTrpAsnAsnIleGlnTyrSerSerSerLysGluIleLys
901  AGTTATTGGAATAATATCCAATATTCTTCGTCAAAAGAGATAAAA  945
```

## 4.7 SIMILARITY WITH OTHER PROTEINS

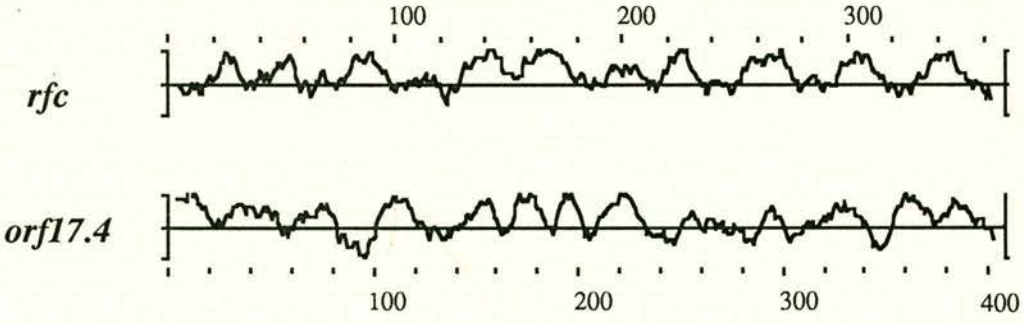
The deduced amino acid sequence of each ORF was used to search against GenPeptide, PIR PROTEINS and SWISS-PROT protein databases using the program FASTA (Pearson, 1990). *orf10.8p* and *orf9.6p* did not show significant similarity to any proteins in the databases. *orf6.1p*, *orf7.0p*, and *orf7.9p* only showed significant homology to proteins encoded by *rfb* genes from the other groups mentioned above.

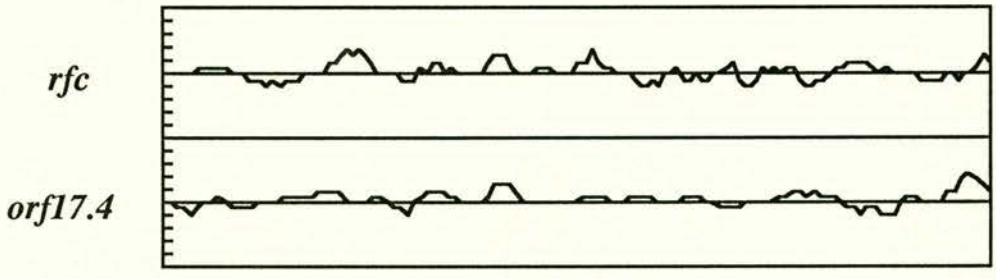
*orf17.4* shows significant similarity to the sequence of the *S. enterica* strain C5 (serovar typhimurium, group B) *rfc* gene (Collins and Hackett, 1991) which encodes the O antigen polymerase: having 21.2% identity and 54.6% similarity at the amino acid level. The predicted secondary structure and the hydrophobic plot showed a similar pattern for the two proteins encoded by *orf17.4* and *rfc* of strain C5 (Figure 4.14), both proteins being highly hydrophobic. To find out if the high level of similarity between these two proteins is because they both contains large number of certain hydrophobic amino acid residues, five thousand shuffle sequences of *orf17.4p* were generated and compared to the sequence of *rfcp* using BESTFIT. Of the five thousand comparisons, only eleven (0.22%) showed a level of similarity higher than 54% (the highest was 55.6% and the mean was 47.56%). Thus, the similarity between *orf17.4p* and *rfcp* is significant and we propose that *orf17.4p* catalyzes O antigen polymerization in strain M32.

# FIGURE 4.14

## Structural comparison of *orf17.4p* with *rfcp* of train C5

Comparison of the predicted hydrophobicity profile (above) and the charge distribution (below) of the deduced amino acid sequences of *orf17.4* of strain M32 and *rfc* of strain C5 (group B). The hydrophobicity was plotted using program PEPPLOT setting the window at 9 amino acids. Hydrophobicity and hydrophilic regions are indicated when the curve is in the upper and lower half respectively. The charge distribution was plotted using program NIP, and the positive and negative charge is indicated when the curve is above and below the central line, respectively.


100
200
300
rfc
orf17.4
100
200
300
400


rfc
orf17.4

# TABLE 4.5

## G+C content of sequenced M32 *rfb* genes

The G+C content of each gene and the values of P1, P2, and P3 as the corrected average G+C contents for codon position 1, 2, and 3 (Sueoka, 1988). The features of some LT2 *rfb* genes (Jiang *et al.*, 1991) which are underlined are presented in brackets for comparison.

**G+C contents of *rfb* genes and codon positions**

| Gene | G+C | P1 | P2 | P3 |
|---|---|---|---|---|
| *orf6.1* (M32) | 0.457 | 0.56 | 0.378 | 0.449 |
| [ *orf6.1* (LT2) | 0.457 | 0.567 | 0.371 | 0.447 ] |
| *orf7.0* (M32) | 0.339 | 0.511 | 0.30 | 0.20 |
| [ *orf7.0* (LT2) | 0.404 | 0.567 | 0.348 | 0.285 ] |
| *orf7.9* (M32) | 0.289 | 0.35 | 0.279 | 0.231 |
| *orf9.6* (M32) | 0.314 | 0.398 | 0.325 | 0.226 |
| *orf10.8* (M32) | 0.337 | 0.485 | 0.288 | 0.227 |
| *orf11.9* (M32) | 0.373 | 0.445 | 0.399 | 0.277 |
| [ *orf16.5* (LT2) | 0.375 | 0.460 | 0.374 | 0.299 ] |
| *orf17.4* (M32) | 0.294 | 0.359 | 0.277 | 0.254 |

## 4.8 G+C CONTENT AND CODON USAGE

### 4.8.1 G+C CONTENT

The sequenced region of M32 has a G+C content which ranges from 29% to 46% (Table 4.5). The M32 genes located in the regions which are nearly identical in strain LT2 are assumed to have the same G+C content as their corresponding LT2 genes, which range from 37% to 50% (Jiang *et al.*, 1991). The G+C content of the whole M32 *rfb* region was plotted (Figure 4.15), and shows that the G+C content is variable but less than 50% throughout the whole region.

### 4.8.2 CODON USAGE AND AMINO ACID COMPOSITION

Codon usage for the predicted proteins of sequenced M32 *rfb* genes is shown in Table 4.6. Due to the low G+C content of *rfb* region, there is a substantial bias towards the usage of codons with low G+C content. For example, of the 78 leucine residues in *orf7.9p*, 44% are encoded by TTA, 18% by TTG, 19% by CTT, 14% by CTA , and 5% by CTG, while none is encoded by CTC.

### 4.8.3 P1, P2 AND P3 VALUES

The values of P1, P2 and P3 of the sequenced M32 *rfb* genes are listed in Table 4.5 (P1, P2 and P3 are the corrected average G+C contents for codon positions 1, 2 and 3 (Sueoka, 1988)). The values observed are characteristic of low G+C content DNA with P3 being lower than P1 and P2. *orf7.0* (M32) and *orf11.9* (M32) have lower P3 values than the corresponding genes of strain LT2.

# TABLE 4.6

## Codon usage in the sequenced M32 *rfb* genes

Codon usages of sequenced *rfb* genes of strain M32 were calculated using the program NIP.

*orf6.1*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 78. | S TCT | 12. | Y TAT | 76. | C TGT | 50. |
| F | TTC | 22. | S TCC | 18. | Y TAC | 24. | C TGC | 50. |
| L | TTA | 23. | S TCA | 0. | * TAA | 0. | * TGA | 0. |
| L | TTG | 13. | S TCG | 6. | * TAG | 0. | W TGG | 100. |
| L | CTT | 17. | P CCT | 7. | H CAT | 100. | R CGT | 36. |
| L | CTC | 3. | P CCC | 7. | H CAC | 0. | R CGC | 36. |
| L | CTA | 3. | P CCA | 33. | Q CAA | 40. | R CGA | 0. |
| L | CTG | 40. | P CCG | 53. | Q CAG | 60. | R CGG | 9. |
| I | ATT | 59. | T ACT | 0. | N AAT | 69. | S AGT | 47. |
| I | ATC | 36. | T ACC | 31. | N AAC | 31. | S AGC | 18. |
| I | ATA | 5. | T ACA | 15. | K AAA | 65. | R AGA | 18. |
| M | ATG | 100. | T ACG | 54. | K AAG | 35. | R AGG | 0. |
| V | GTT | 20. | A GCT | 41. | D GAT | 73. | G GGT | 42. |
| V | GTC | 13. | A GCC | 14. | D GAC | 27. | G GGC | 23. |
| V | GTA | 33. | A GCA | 23. | E GAA | 44. | G GGA | 12. |
| V | GTG | 33. | A GCG | 23. | E GAG | 56. | G GGG | 23. |

*orf7.0*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 83. | S TCT | 25. | Y TAT | 80. | C TGT | 100. |
| F | TTC | 17. | S TCC | 8. | Y TAC | 20. | C TGC | 0. |
| L | TTA | 70. | S TCA | 33. | * TAA | 0. | * TGA | 0. |
| L | TTG | 10. | S TCG | 0. | * TAG | 0. | W TGG | 100. |
| L | CTT | 20. | P CCT | 83. | H CAT | 100. | R CGT | 29. |
| L | CTC | 0. | P CCC | 0. | H CAC | 0. | R CGC | 29. |
| L | CTA | 0. | P CCA | 17. | Q CAA | 40. | R CGA | 14. |
| L | CTG | 0. | P CCG | 0. | Q CAG | 60. | R CGG | 0. |
| I | ATT | 80. | T ACT | 20. | N AAT | 56. | S AGT | 33. |
| I | ATC | 0. | T ACC | 0. | N AAC | 44. | S AGC | 0. |
| I | ATA | 20. | T ACA | 60. | K AAA | 64. | R AGA | 14. |
| M | ATG | 100. | T ACG | 20. | K AAG | 36. | R AGG | 14. |
| V | GTT | 47. | A GCT | 67. | D GAT | 100. | G GGT | 57. |
| V | GTC | 6. | A GCC | 0. | D GAC | 0. | G GGC | 7. |
| V | GTA | 35. | A GCA | 0. | E GAA | 94. | G GGA | 14. |
| V | GTG | 12. | A GCG | 33. | E GAG | 6. | G GGG | 21. |

*orf7.9*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 85. | S TCT | 28. | Y TAT | 81. | C TGT | 20. |
| F | TTC | 15. | S TCC | 13. | Y TAC | 19. | C TGC | 80. |
| L | TTA | 44. | S TCA | 21. | * TAA | 0. | * TGA | 0. |
| L | TTG | 18. | S TCG | 2. | * TAG | 0. | W TGG | 100. |
| L | CTT | 19. | P CCT | 29. | H CAT | 100. | R CGT | 0. |
| L | CTC | 0. | P CCC | 14. | H CAC | 0. | R CGC | 0. |
| L | CTA | 14. | P CCA | 43. | Q CAA | 75. | R CGA | 0. |
| L | CTG | 5. | P CCG | 14. | Q CAG | 25. | R CGG | 8. |
| I | ATT | 33. | T ACT | 33. | N AAT | 83. | S AGT | 23. |
| I | ATC | 16. | T ACC | 5. | N AAC | 17. | S AGC | 13. |
| I | ATA | 51. | T ACA | 52. | K AAA | 87. | R AGA | 50. |
| M | ATG | 100. | T ACG | 10. | K AAG | 13. | R AGG | 42. |
| V | GTT | 45. | A GCT | 29. | D GAT | 91. | G GGT | 36. |
| V | GTC | 15. | A GCC | 21. | D GAC | 9. | G GGC | 14. |
| V | GTA | 34. | A GCA | 42. | E GAA | 71. | G GGA | 27. |
| V | GTG | 6. | A GCG | 8. | E GAG | 29. | G GGG | 23. |

*orf9.6*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 31. | S TCT | 7. | Y TAT | 23. | C TGT | 5. |
| F | TTC | 9. | S TCC | 2. | Y TAC | 1. | C TGC | 1. |
| L | TTA | 20. | S TCA | 9. | * TAA | 0. | * TGA | 0. |
| L | TTG | 6. | S TCG | 3. | * TAG | 0. | W TGG | 1. |
| L | CTT | 9. | P CCT | 5. | H CAT | 6. | R CGT | 5. |
| L | CTC | 5. | P CCC | 1. | H CAC | 0. | R CGC | 0. |
| L | CTA | 8. | P CCA | 3. | Q CAA | 4. | R CGA | 1. |
| L | CTG | 1. | P CCG | 0. | Q CAG | 2. | R CGG | 0. |
| I | ATT | 22. | T ACT | 6. | N AAT | 15. | S AGT | 2. |
| I | ATC | 7. | T ACC | 2. | N AAC | 2. | S AGC | 4. |
| I | ATA | 24. | T ACA | 4. | K AAA | 13. | R AGA | 4. |
| M | ATG | 10. | T ACG | 3. | K AAG | 3. | R AGG | 1. |
| V | GTT | 6. | A GCT | 9. | D GAT | 7. | G GGT | 12. |
| V | GTC | 5. | A GCC | 4. | D GAC | 1. | G GGC | 3. |
| V | GTA | 12. | A GCA | 10. | E GAA | 7. | G GGA | 9. |
| V | GTG | 5. | A GCG | 2. | E GAG | 4. | G GGG | 5. |

*orf10.8*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 89. | S TCT | 42. | Y TAT | 78. | C TGT | 80. |
| F | TTC | 11. | S TCC | 0. | Y TAC | 22. | C TGC | 20. |
| L | TTA | 41. | S TCA | 16. | * TAA | 0. | * TGA | 0. |
| L | TTG | 16. | S TCG | 5. | * TAG | 0. | W TGG | 100. |
| L | CTT | 35. | P CCT | 46. | H CAT | 81. | R CGT | 22. |
| L | CTC | 0. | P CCC | 8. | H CAC | 19. | R CGC | 11. |
| L | CTA | 5. | P CCA | 38. | Q CAA | 78. | R CGA | 0. |
| L | CTG | 3. | P CCG | 8. | Q CAG | 22. | R CGG | 11. |
| I | ATT | 48. | T ACT | 22. | N AAT | 80. | S AGT | 26. |
| I | ATC | 18. | T ACC | 11. | N AAC | 20. | S AGC | 11. |
| I | ATA | 33. | T ACA | 44. | K AAA | 73. | R AGA | 56. |
| M | ATG | 100. | T ACG | 22. | K AAG | 27. | R AGG | 0. |
| V | GTT | 51. | A GCT | 40. | D GAT | 85. | G GGT | 48. |
| V | GTC | 6. | A GCC | 16. | D GAC | 15. | G GGC | 14. |
| V | GTA | 34. | A GCA | 28. | E GAA | 50. | G GGA | 24. |
| V | GTG | 9. | A GCG | 16. | E GAG | 50. | G GGG | 14. |

*orf11.9*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 84. | S TCT | 24. | Y TAT | 85. | C TGT | 75. |
| F | TTC | 16. | S TCC | 4. | Y TAC | 15. | C TGC | 25. |
| L | TTA | 43. | S TCA | 32. | * TAA | 0. | * TGA | 0. |
| L | TTG | 10. | S TCG | 12. | * TAG | 0. | W TGG | 100. |
| L | CTT | 19. | P CCT | 33. | H CAT | 70. | R CGT | 7. |
| L | CTC | 10. | P CCC | 8. | H CAC | 30. | R CGC | 13. |
| L | CTA | 5. | P CCA | 25. | Q CAA | 54. | R CGA | 0. |
| L | CTG | 14. | P CCG | 33. | Q CAG | 46. | R CGG | 20. |
| I | ATT | 45. | T ACT | 42. | N AAT | 87. | S AGT | 16. |
| I | ATC | 18. | T ACC | 26. | N AAC | 13. | S AGC | 12. |
| I | ATA | 36. | T ACA | 26. | K AAA | 88. | R AGA | 53. |
| M | ATG | 100. | T ACG | 5. | K AAG | 12. | R AGG | 7. |
| V | GTT | 45. | A GCT | 48. | D GAT | 73. | G GGT | 35. |
| V | GTC | 9. | A GCC | 17. | D GAC | 27. | G GGC | 24. |
| V | GTA | 36. | A GCA | 28. | E GAA | 69. | G GGA | 12. |
| V | GTG | 9. | A GCG | 7. | E GAG | 31. | G GGG | 29. |

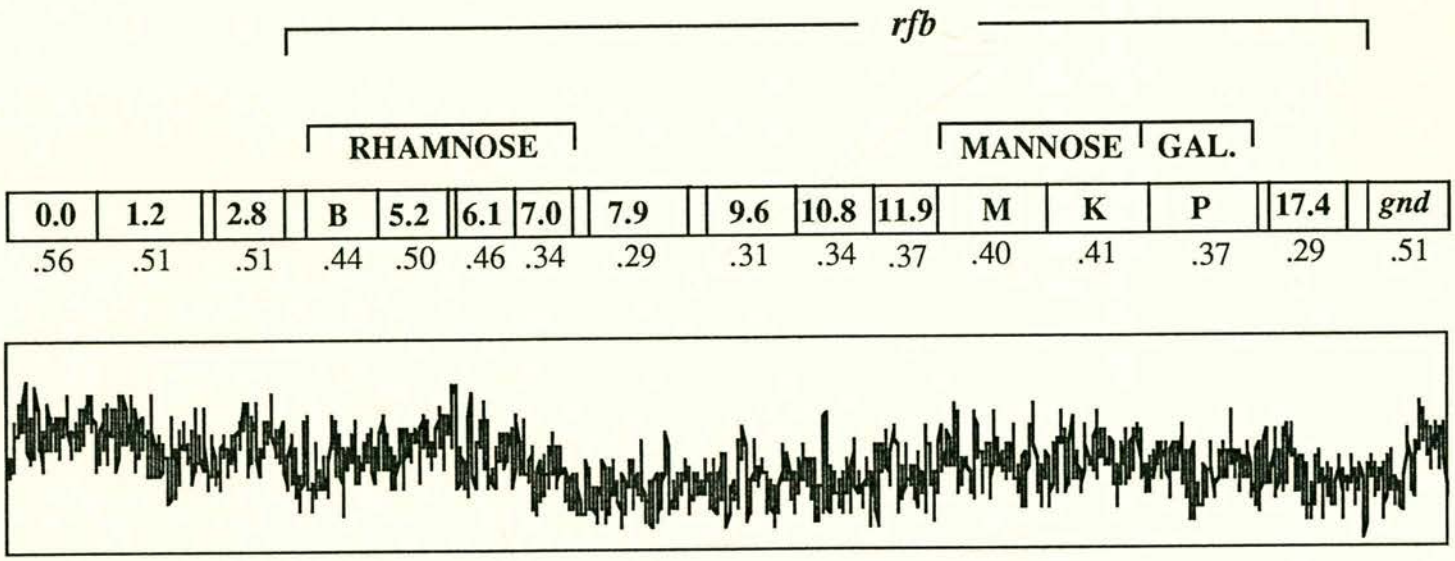
*orf17.4*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| F | TTT | 85. | S TCT | 18. | Y TAT | 93. | C TGT | 50. |
| F | TTC | 15. | S TCC | 11. | Y TAC | 7. | C TGC | 50. |
| L | TTA | 51. | S TCA | 18. | * TAA | 0. | * TGA | 0. |
| L | TTG | 11. | S TCG | 21. | * TAG | 0. | W TGG | 100. |
| L | CTT | 9. | P CCT | 57. | H CAT | 67. | R CGT | 12. |
| L | CTC | 0. | P CCC | 14. | H CAC | 33. | R CGC | 0. |
| L | CTA | 15. | P CCA | 14. | Q CAA | 29. | R CGA | 0. |
| L | CTG | 15. | P CCG | 14. | Q CAG | 71. | R CGG | 38. |
| I | ATT | 45. | T ACT | 28. | N AAT | 78. | S AGT | 14. |
| I | ATC | 10. | T ACC | 6. | N AAC | 22. | S AGC | 18. |
| I | ATA | 45. | T ACA | 56. | K AAA | 82. | R AGA | 38. |
| M | ATG | 100. | T ACG | 11. | K AAG | 18. | R AGG | 12. |
| V | GTT | 38. | A GCT | 53. | D GAT | 82. | G GGT | 55. |
| V | GTC | 9. | A GCC | 7. | D GAC | 18. | G GGC | 0. |
| V | GTA | 34. | A GCA | 20. | E GAA | 100. | G GGA | 25. |
| V | GTG | 19. | A GCG | 20. | E GAG | 0. | G GGG | 20. |

# FIGURE 4.15

## The G+C content plot of the *rfb* gene cluster of strain M32

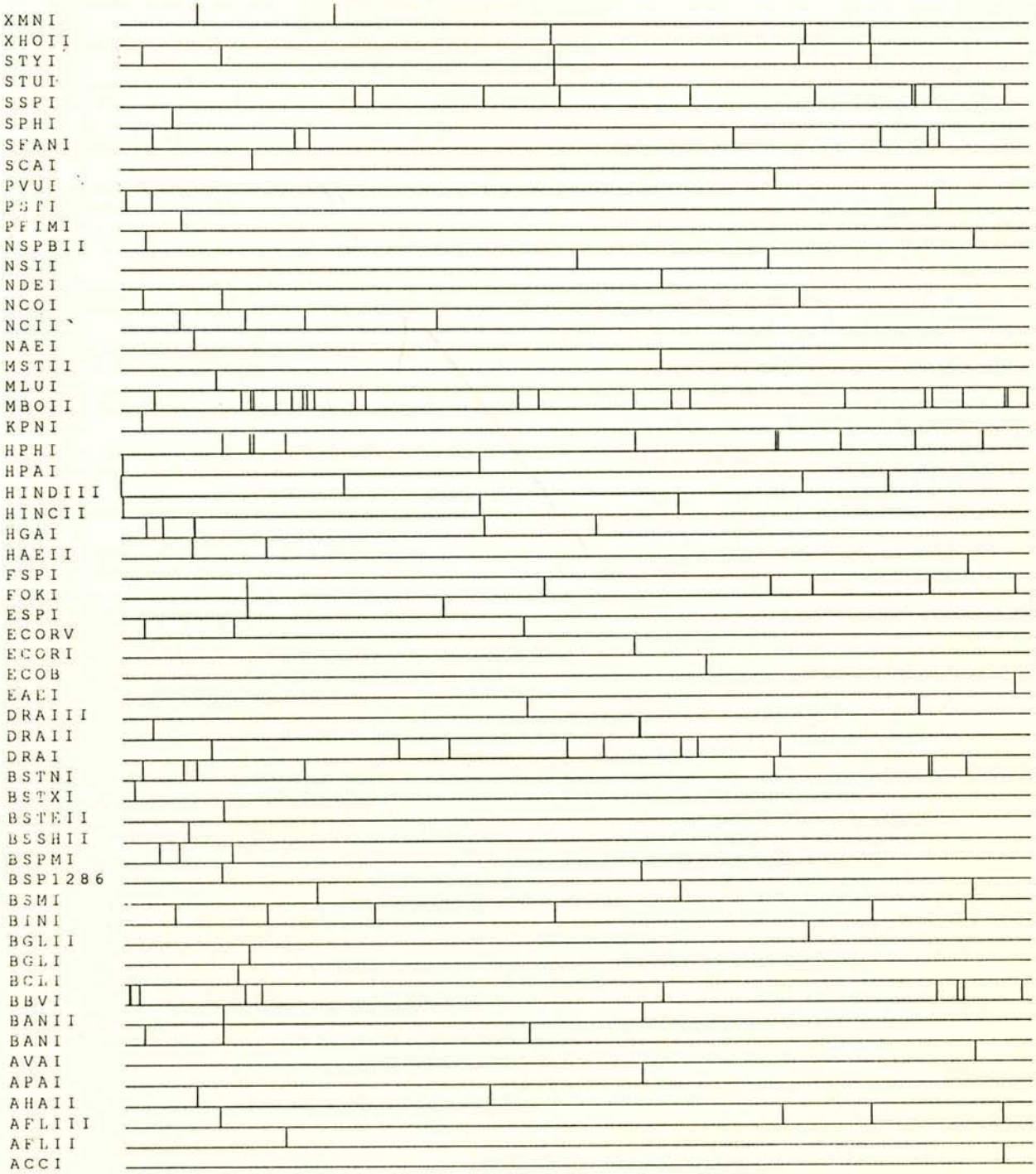
The G+C content of the *rfb* gene cluster of strain M32 was plotted from position 0.0 to the *gnd* gene using program NIP. Each point of the plot gives an average of over 199 bases. The G+C content of each gene is given underneath that gene.

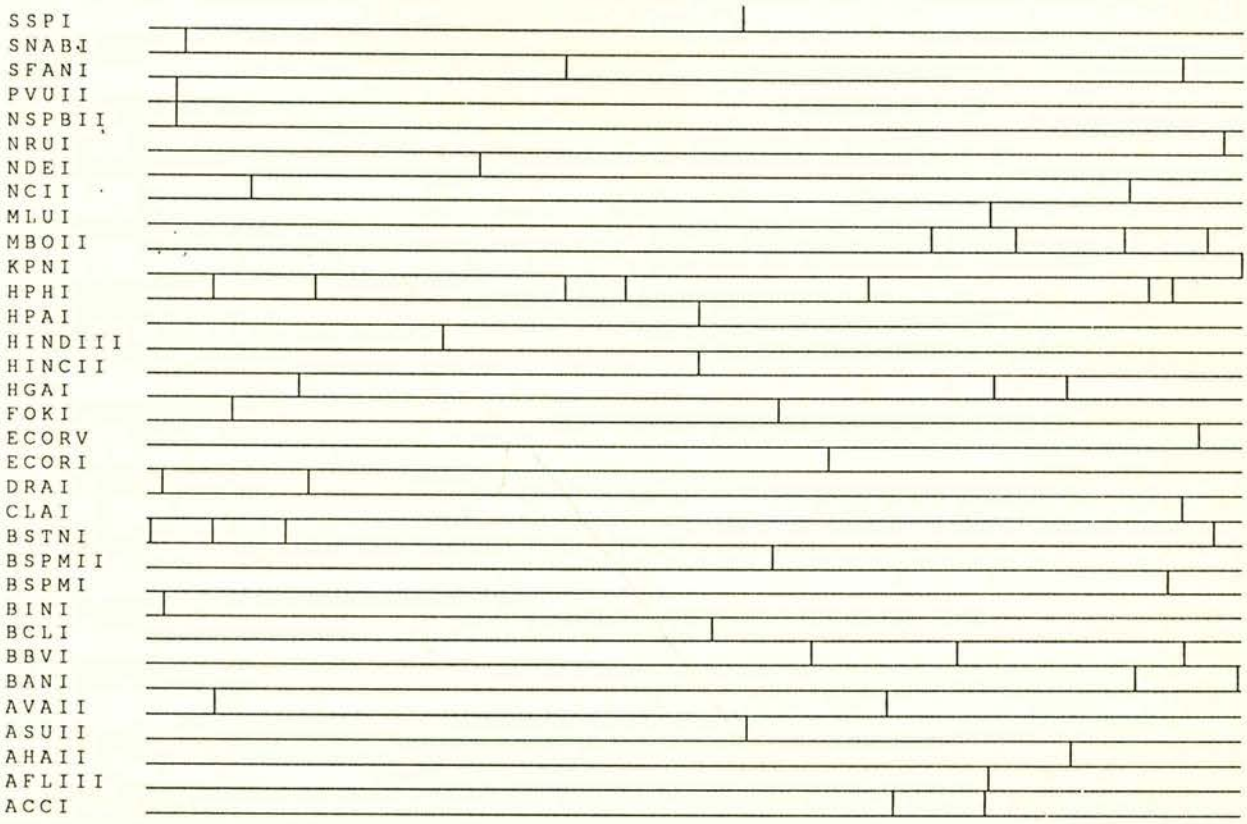

rfb
RHAMNOSE
MANNOSE
GAL.
0.0
1.2
2.8
B
5.2
6.1
7.0
7.9
9.6
10.8
11.9
M
K
P
17.4
gnd
.56
.51
.51
.44
.50
.46
.34
.29
.31
.34
.37
.40
.41
.37
.29
.51

# FIGURE 4.16

## Restriction maps of sequenced *rfb* regions of strain M32

Restriction maps of the sequenced *rfb* regions of strain M32. A: the map of region from position 5,337 to position 13,000. B: the map of the region from position 16,808 to position 19,094. Map was generated using the program NIP.


XMNI
XHOII
STYI
STUI
SSPI
SPHI
SFANI
SCAI
PVUI
PSTI
PFIMI
NSPBII
NSII
NDEI
NCOI
NCII
NAEI
MSTII
MLUI
MBOII
KPNI
HPHI
HPAI
HINDIII
HINCII
HGAI
HAEII
FSPI
FOKI
ESPI
ECORV
ECORI
ECOB
EAEI
DRAIII
DRAII
DRAI
BSTNI
BSTXI
BSTEII
BSSHII
BSPMI
BSP1286
BSMI
BINI
BGLII
BGLI
BCLI
BBVI
BANII
BANI
AVAI
APAI
AHAII
AFLIII
AFLII
ACCI


SSPI
SNABI
SFANI
PVUII
NSPBII
NRUI
NDEI
NCII
MLUI
MBOII
KPNI
HPHI
HPAI
HINDIII
HINCII
HGAI
FOKI
ECORV
ECORI
DRAI
CLAI
BSTNI
BSPMII
BSPMI
BINI
BCLI
BBVI
BANI
AVAII
ASUII
AHAII
AFLIII
ACCI

## 4.9 RESTRICTION MAP

Restriction endonuclease sites were located in the sequenced regions using the program NIP (Figure 4.16). All the restriction sites mapped earlier by restriction mapping, and used in subcloning, match with the sites predicted from the sequences.

## 4.10 COMPARISON OF THE *RFB* GENE CLUSTERS OF STRAINS M32 AND LT2

Comparison of the *rfb* gene cluster of strain M32 with that of strain LT2 by Southern hybridization and restriction mapping was described in the previous chapter. To make a detailed comparison, the *rfb* regions unique to strain M32 and the junction regions were sequenced. Here, the comparison based on the sequence analysis is presented combining with the results obtained by Southern hybridization and restriction mapping.

The region from position 0.0 to position 5.6 of strain M32 *rfb* gene cluster is nearly identical to that of strain LT2 with about 1 to 2% nucleotide difference based on restriction enzyme analysis. The region from position 5,337 to position 6,903 has 30 (1.9%) nucleotide difference between strains M32 and LT2 based on sequence comparison (Figure 4.3). The region from position 6,904 to position 7,541 shows considerable similarity to the LT2 sequence: of the 637 bases, 408 (64.1%) were identical to those in the region from position 6,904 to position 7,558 of the LT2 sequence.

The region from position 7,542 to position 11,877 contains *orf7.9* (M32), *orf9.6* (M32) and *orf10.8* (M32). No significant similarity

between this region and the LT2 *rfb* sequence was found at the nucleotide level using any of the available programs, and this region does not appear to be represented in strain LT2.

The region from position 11,878 to position 12,822 is similar to the sequence from position 16,441 to position 17,382 of strain LT2: there was 26.6% difference at the nucleotide level (Figure 4.3).

The region from position 12,823 to position 17,187 of strain M32 is nearly identical to the *rfb* region from position 17,383 to position 21,747 of strain LT2 sequence with about 5% nucleotide difference as shown by restriction enzyme analysis. The region from position 17,188 to position 18,697 of strain M32 is not presented in strain LT2, and the region from position 18,698 to position 19,094 of strain M32 is nearly identical to the region from position 21,889 to position 22,285 of strain LT2 with one base difference (Figure 4.4).

The comparison of the *rfb* gene clusters of strains M32 and LT2 is summarised in Figure 4.17.

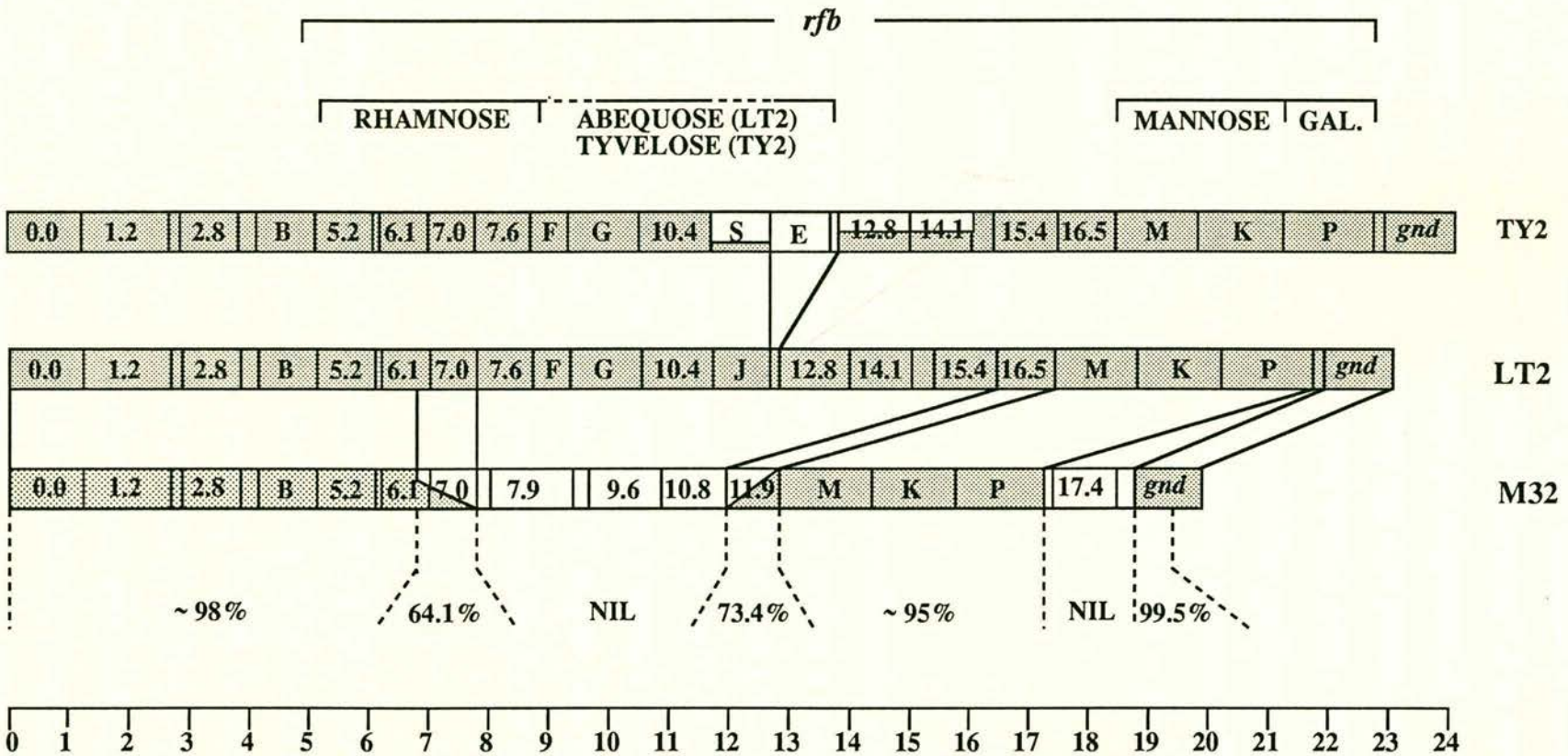
## 4.11 JUNCTIONS BETWEEN THE CENTRAL AND FLANKING REGIONS

The central part of the M32 *rfb* region shows no detectable sequence homology with the *rfb* region of strain LT2, but the regions flanking the central part are conserved in both strains. The junctions between the central and flanking regions are interesting. *orf7.0* and *orf11.9* are adjacent to the central region in strain M32, as are *orf7.0* and *orf16.5* in strain LT2 (Figure 4.17). The alignment of the *orf7.0* genes is quite clear up to within 24 and 39 bases of the ends of *orf7.0* (M32) and

# FIGURE 4.17

## Comparison of the *rfb* gene clusters of strains LT2, Ty2 (group D), and M32.

The biochemical pathways to which the genes of strains LT2 and Ty2 are related are indicated above the map, with dotted lines indicating uncertainty (Jiang *ett al.*, 1991). The percentages of similarity between segments of M32 and the corresponding segments of strain LT2 (as shown by connecting lines), based on restriction enzyme analysis and sequence comparison are given below the segments. Where there is no significant similarity, NIL is indicated. Map position in kilobase pairs is indicated in the lower part of the diagram.


rfb
RHAMNOSE
ABEQUOSE (LT2)
TYVELOSE (TY2)
MANNOSE
GAL.
0.0 1.2 2.8 B 5.2 6.1 7.0 7.6 F G 10.4 S E 12.8 14.1 15.4 16.5 M K P gnd
TY2
0.0 1.2 2.8 B 5.2 6.1 7.0 7.6 F G 10.4 J 12.8 14.1 15.4 16.5 M K P gnd
LT2
0.0 1.2 2.8 B 5.2 6.1 7.0 7.9 9.6 10.8 11.9 M K P 17.4 gnd
M32
~ 98%
64.1%
NIL
73.4%
~ 95%
NIL
99.5%
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

*orf7.0* (LT2) respectively (*orf7.0* (LT2) extends 15 bases beyond *orf7.0* (M32), Figure 4.9), whereas there is no sequence similarity between *orf7.9* (M32) and *orf7.6* (LT2) which follow *orf7.0* in strains M32 and LT2 respectively. However although the alignment of the two genes is clear and the junction between similar and non similar regions is sharp, the level of similarity decreases from within *orf6.1* of both strains up to the junction (Figure 4.18), and a similar gradient is also found at the other junction with the level of similarity between *orf11.9* (M32) and *orf16.5* (LT2) increasing from the junction up to the ends of these ORFs (Figure 4.18).

## 4.12 SUMMARY AND DISCUSSION

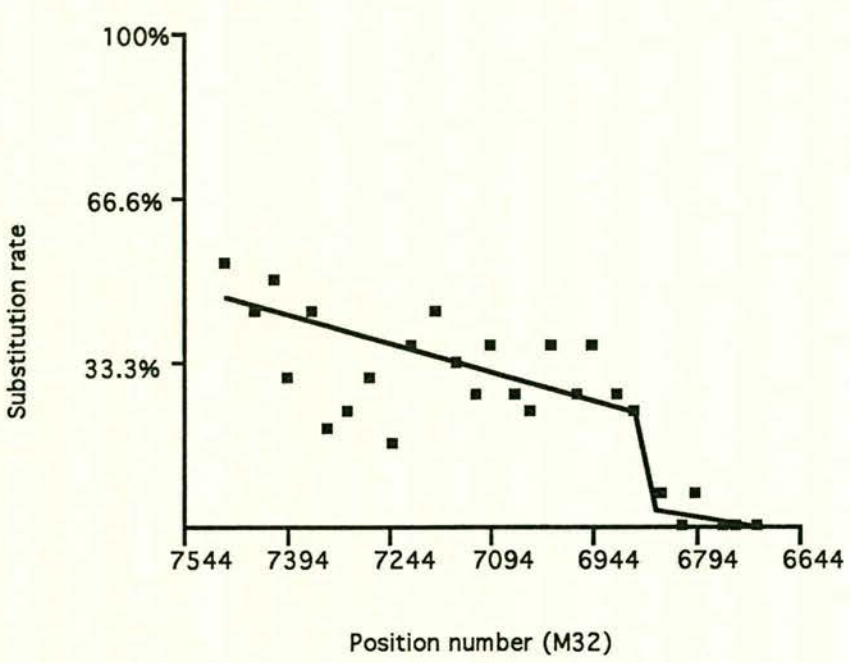
The *rfb* regions of strain M32 which show little or no homology with the *rfb* region of strain LT2 have been sequenced to enable a detailed comparison of the whole *rfb* regions of strains M32 and LT2. The sequenced regions of strain M32 are from position 5,337 to 13,000 and from position 16,808 to 19,094. The new sequences extended into the regions nearly identical in strain LT2 whose *rfb* region has been sequenced, and we thus have in effect the sequence of the entire M32 *rfb* region.

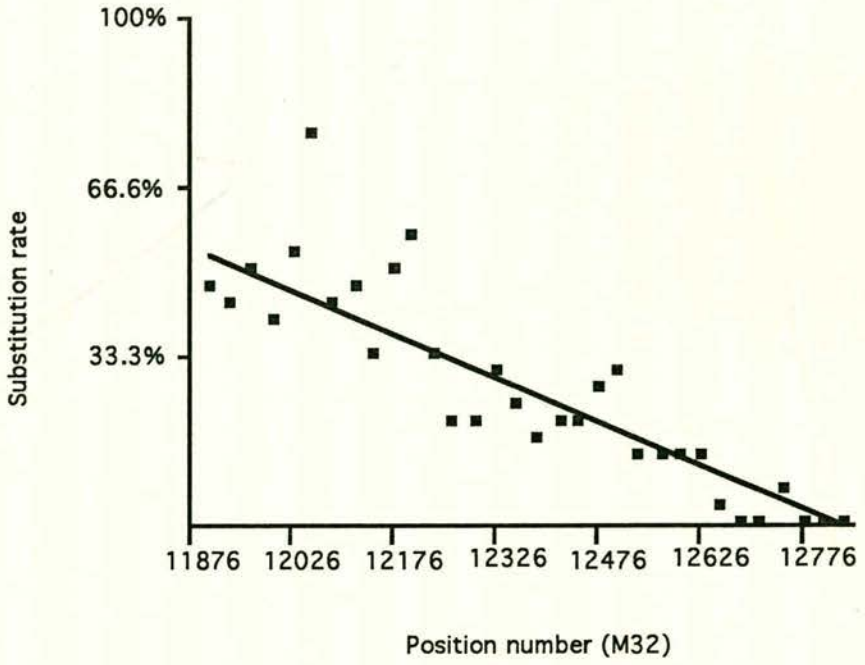
Analysis of the sequences revealed five complete ORFs (*orf6.1*, *orf7.0*, *orf7.9*, *orf9.6* and *orf10.8*) in the region from position 5,337 to 13,000 and one ORF (*orf17.4*) in the region from position 16,808 to 19,094. *orf6.1*, *orf7.0* and *orf11.9* are clearly homologous to *orf6.1*, *orf7.0* and *orf16.5* of strain LT2 respectively. However comparison of *orf7.9*, *orf9.6*, *orf10.8* and *orf17.4* with the ORFs of strain LT2 revealed no detectable homology at nucleotide sequence level. There is thus a sharp distinction between genes which can be aligned and genes with no

# FIGURE 4.18

## Levels of substitution at the junctions between homologous and nonhomologous sequences

The M32 *rfb* sequences from position 6,644 to position 7,544 and from position 11,876 to position 12,822 were compared with the LT2 *rfb* sequences from position 6,644 to position 7,561 and from position 16,441 to position 17,382 respectively. The data shown are the levels of substitution per block of 30 nucleotides.


100%
66.6%
33.3%
Substitution rate
7544
7394
7244
7094
6944
6794
6644
Position number (M32)


100%
66.6%
33.3%
Substitution rate
11876
12026
12176
12326
12476
12626
12776
Position number (M32)

detected similarity.

However, a gradient of similarity level was found in the homologous genes adjacent to junctions between central and flanking regions with the similarity (between the regions of strains M32 and LT2) decreasing markedly from a very high level as the junction and central region is approaching. There have been several reports of intragenic recombination during the evolution of *E. coli* (Dubose *et al.*, 1988; Bisercic *et al.*, 1991; Nelson *et al.*, 1991). Such events allow those regions of *rfb* gene clusters which are common to two forms to undergo random genetic drift as one, while the nonhomologous regions must drift independently, as each must be independently maintained and there is no possibility of genetic recombination within such regions. We suggest that at the time of the recombination event which generated the junctions, the M32 genes adjacent to the central region, *orf7.0* (M32) and *orf11.9* (M32), were very similar to their corresponding LT2 genes, *orf7.0* (LT2) and *orf16.5* (LT2) respectively, but proximity to the regions of non homology reduced homologous recombination, such that the neutral mutations which have accumulated in the two *rfb* gene clusters are different. The gradient reflects the increasing possibility of recombination with distance from the central region. The abrupt drop in the level of difference at about position 6,850 (Figure 4.18) could be due to a recent recombination between two forms of *rfb* clusters at this site.

The other region which shows no homology to that of strain LT2 is the *orf17.4* region, and this region is flanked at both ends by DNA which is very similar to LT2 sequence right up to the junctions. This may indicate that the junctions were generated by recombination event or

events quite recently (comparing to the age of the recombination events involved in the generation of the junctions between central and flanking regions).

The rhamnose biosynthetic pathway genes in strain LT2 comprise *rfbB*, *orf5.2*, *orf6.1* and *orf7.0* (Jiang *et al.*, 1991). Rhamnose is a constituent sugar of the M32 O antigen, and it is expected that 4 similar genes are found in the *rfb* region of strain M32. *orf5.2* and *rfbB* are present within the region from position 0 to 6.9 which is nearly identical in both strains LT2 and M32 as shown above. *orf6.1* of strain M32 is almost identical to its LT2 corresponding gene at both DNA and amino acid sequence levels, and *orf7.0p* of strain M32 shows significant similarity with *orf7.0p* of strain LT2. Thus we conclude that these four rhamnose biosynthetic pathway genes are conserved in both strains M32 and LT2.

*orf7.9p* has 12 predicted transmembrane segments. The fact that the distribution of these transmembrane segments is nearly identical to that of *orf12.8p* of strain LT2 suggests that these two proteins are structurally similar and may have a similar function although there is only 50.5% similarity between these two proteins. Based on their hydrophobicity and hydrophobic moment values, 8 or 9 of the 12 transmembrane segments of *orf7.9p* were predicted to be paired or bundled in the membrane. It thus appears that this protein may form a pore or a complex in the membrane by itself or with other proteins, and may be involved in the translocation of the O antigen.

*orf9.6* and *orf10.8* do not show any significant similarity with LT2 *rfb* genes at either nucleotide or amino acid level. *orf9.6* encodes a highly hydrophobic protein with ten potential transmembrane segments, but

none were found in *orf10.8p*. The linkage of mannose to rhamnose differs in the O antigens of strains M32 and LT2, and hence we expect the mannose transferase gene of strain M32 to differ from that in strain LT2, and either *orf9.6* or *orf10.8* could be the mannose transferase gene.

*orf17.4p* shows significant similarity to amino acid sequence of the *rfc* gene of strain C5, and it may catalyze O antigen polymerization in strain M32.

All the ORFs are transcribed in one direction, which is the same as that of strain LT2 *rfb* genes. The sequences were searched for promoter using the programs provided by Dr. O'Neill, and two potential promoters were found located in the intergenic gap between *orf7.9* and *orf9.6*. Thus, the M32 *rfb* gene cluster is highly likely to be a single operon translated from a promoter in the same position as that of strain LT2, at position 2.8 or 3.9, as this region is identical in the two strains, and this operon may have an additional internal promoter located in front of *orf9.6*.

Of the seven sequenced M32 *rfb* genes, *orf10.8* and *orf11.9* do not have good ribosome binding sites. *orf10.8* and *orf11.9* have start sites overlapping with the previous ORF or locating within three bases after the stop codon of previous ORF respectively, and they are presumably initiated by translational coupling. Another five ORFs all have ribosome binding sites, and they may be expressed by translation reinitiation.

The M32 *rfb* region is of low G+C content. The sequenced regions have a G+C content ranging from 0.29 to 0.37, and the ORFs have a

codon usage characteristic of low G+C species, with P3 being lower than P1 or P2 (Table 4.5). The G+C content of a gene reflects the long term effects of the balance between the mutation rates from G.C to A.T and A.T to G.C (Sueoka, 1988). We might expect that once a gene is in *S. enterica*, the mutational pressure which normally maintains a given G+C content would bring this value to about 0.5 which is usual for *S. enterica* genes. To account for the low G+C content and atypical (for *S. enterica*) codon usage of the group B *rfb* gene cluster, it was proposed that the *rfb* gene cluster had been transferred to the *S. enterica* strain LT2 from a low G+C content species (Wyk and Reeves, 1989; Jiang *et al.*, 1991), and the same argument applies to the *rfb* gene cluster of group E1.

There has in general been substantial synonymous substitution involving the third base since *Escherichia* and *Salmonella* diverged (Ochman and Wilson, 1987), and had the *rfb* gene cluster been in *Salmonella* during this period, then the G+C content of the third base would have been adjusted to approximate that of these two genera. The fact that P3 is typical of a low G+C content organism suggests that the *rfb* gene cluster has been in *S. enterica* for only a small part of the time since the divergence.

# CHAPTER 5: *RFBP(T)*

## 5.1 INTRODUCTION

As discussed in chapter one, LPS consists of three structural regions: the lipid A moiety, the oligosaccharide core and the O antigen. Different structural regions of LPS are synthesized by different biosynthetic pathways. Lipid A is synthesized first and core is then synthesized on the lipid A directed by *rfa* genes. O antigen is synthesized on a lipid carrier, undecaprenol, directed by *rfb* genes.

The synthesis of O antigen starts with synthesis of O unit on undecaprenol as discussed in chapter one. Single O units are polymerized into O polysaccharide on undecaprenyl pyrophosphate, and then transferred from undecaprenyl pyrophosphate to lipid A-core to form the complete LPS molecule. Synthesis of O unit on undecaprenyl pyrophosphate is known to be membrane associated and must be on the cytoplasmic face of the inner membrane because all the nucleotide sugars are synthesized by cytosolic enzymes as mentioned in chapter one (Osborn *et al.*, 1972a, 1972b). Also as discussed in chapter one, polymerization of the repeat O units as well as the ligation of the O antigen to lipid A-core occurs at the periplasmic face of the inner membrane (Osborn *et al.*, 1972b; Mulford and Osborn, 1983; McGrath and Osborn, 1991a). How the undecaprenyl pyrophosphate linked single O unit is translocated to the periplasmic face of the inner membrane where the polymerization and ligation occur is unknown.

As mentioned in chapter one, genes in the *rfa* locus are involved in the synthesis of the core. Because synthesis of O antigen occurs separately from that of core, O antigenic material can be synthesized in *rfa* mutants which do not have complete core (Beckmann *et al.*, 1964). In these circumstances O antigen can not be attached to incomplete core and it remains linked to undecaprenyl pyrophosphate, and can be detected in lysates as 'O hapten'. O hapten differs from LPS in solubility and other properties (Mäkelä and Stocker, 1984).

However, a mutant blocked in one of the genes located in the *rfa* gene cluster, *rfaL*, can synthesize core which appears to be complete on the basis of chemical composition, serological properties, and the ability to act as a receptor for bacteriophage Felix O (Mäkelä and Stocker, 1984). Bacteriophage Felix O is a *Salmonella*-specific phage that requires core terminating in N-acetylglucosamine (the complete core) as the adsorption site (Lindberg and Holme, 1969). *rfaL* mutants also produce O hapten, and apparently they were unable to transfer O antigen from undecaprenol to the core: this was verified by *in vitro* tests (Mäkelä and Stocker, 1984). Unable to synthesize the complete LPS, *rfaL* mutants are resistant to the smooth-specific phage P22 (P22 uses the *S. enterica* group B lipid A-core linked polymerised O antigen as a receptor). It was concluded that *rfaL* encodes the O antigen ligase which transfers the O antigen from undecaprenyl pyrophosphate to lipid A-core (Mäkelä and Stocker, 1984). *rfaL* from *S. enterica* and *E. coli* has recently been sequenced (Pradel *et al.*, 1992; Klena *et al.*, 1992).

Three mutants of *S. enterica* strain LT2 (serogroup B, serovar typhimurium) isolated by Wilkinson *et al.* were also characterized as

bacteriophage Felix O sensitive, bacteriophage P22 resistant, and producing O hapten (Wilkinson *et al.*, 1972). These mutants were further confirmed to make complete core by showing that their lipid A-core complex accepted O antigen *in vitro* (Wilkinson *et al.*, 1972). These mutations were located in the region close to *his* by co-transduction experiment (Wilkinson *et al.*, 1972). Thus like a *rfaL* mutation, these mutations near *his* also cause bacteria to accumulate O hapten and complete core. It was inferred that a gene, *rfbT*, located within the *rfb* region was also involved in transfer of the O antigen from undecaprenol to lipid A-core, and that both *rfbT* and *rfaL* genes were required for this reaction (Wilkinson *et al.*, 1972; Mäkelä and Stocker, 1984). Since then, no further work has been done on these *rfbT* mutations, and at the time this study was started it was not known if the mutations were in one of the known genes or there was indeed a separate *rfbT* gene.

This chapter describes the mapping and some functional studies of the *rfbT* mutants. No such mutants from group E *S. enterica* have been reported, and the mutants of group B *S. enterica* described above were used in the work presented in this chapter. As discussed in chapter one, the mechanism of LPS synthesis in group E1 strains is the same as that in group B strains, and thus the ligation and related functions should be common to strains from these two groups.

## 5.2 CHARACTERIZATION OF *rfbT* STRAINS

Strains P9285, P9286 and P9287 were isolated and characterized as *rfbT* mutants by Wilkinson *et al.* in 1972, and no other such strains have been reported since then. We obtained these strains from Professor Stocker, and bacteriophage sensitivity tests were carried out

using P22 and Felix O, the results being shown in Table 5.1. Strain P9287 showed resistance to both phages, and there must now be a secondary mutation in the *rfa* region in this strain. Strains P9285 and P9286 showed the expected pattern, and further work was carried out using these two strains.

To check if strains P9285 and P9286 produce any LPS, whole membrane fractions were prepared from those two strains, run on an SDS-PAGE gel and silver stained as described in chapter two (Figure 5.1). Membrane fraction of strains P9003 (wild-type) was also included in the same gel. Strain P9003 produces an LPS which gives a typical ladder pattern. Strain P9285 produces no LPS (except lipid A-core) detectable by silver staining, but strain P9286 gives a weak band which runs at the position which is expected for the LPS molecule with a single O unite on it. This band is weaker than the corresponding band produced by strain P9003. It is likely that the mutation is leaky in strain P9286, and this strain produces a small amount of O antigen.

The presence of LPS on the cell membrane was also tested by slide agglutination using *S. enterica* serovar typhimurium specific antisera O4 and O5. Cells of P9285 do not agglutinate with either of them, but cells of P9286 agglutinate with both very weakly. This also suggests that strain P9286 can produce a small amount of LPS (O antigen), indicating a leaky mutation in this strain.

## 5.3 *rfbT* MUTATION IS IN THE *rfbP* GENE

The whole *rfb* gene cluster of *S. enterica* strain LT2 has been cloned and sequenced, and most of the *rfb* genes have been identified in this laboratory (Jiang *et al.*, 1991). To locate the mutations, a series of *rfb*

**Table 5.1.** Sensitivity and O antiserum specificity patterns

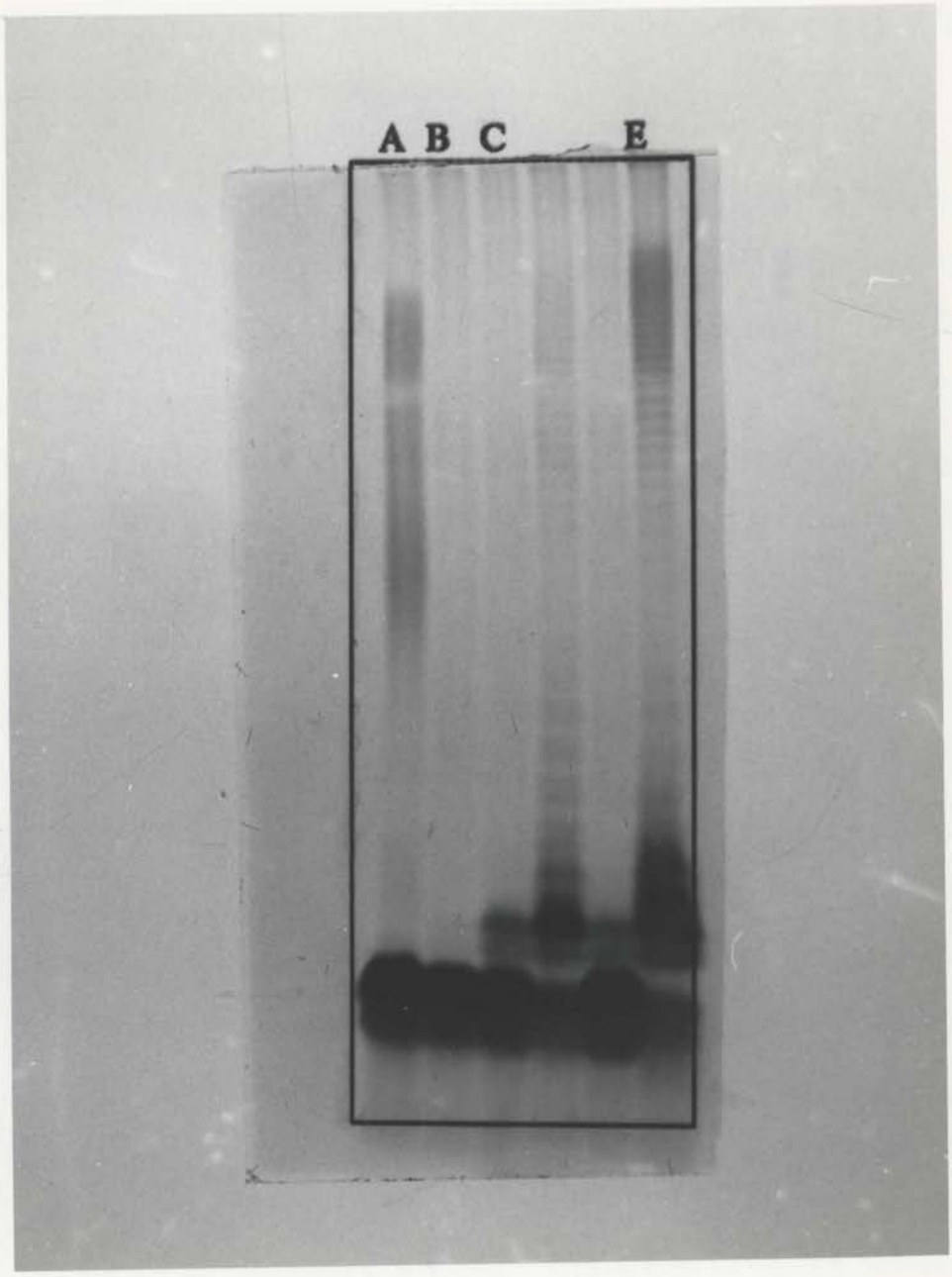
| Strains | Lysis by phages | | Reaction with O antiserum | |
|---|---|---|---|---|
| | P22 | Felix O | O4 | O5 |
| P9285 | - | + | - | - |
| P9286 | - | + | (+)* | (+)* |
| P9287 | - | - | - | - |
| P9324 | + | + | + | + |
| P9325 | + | + | + | + |
| P9003 | + | + | + | + |
| P9300 | - | + | - | - |
| P9073 | - | + | - | - |
| P9320[a] | - | + | - | - |
| P9321[a] | - | + | - | - |
| P9317[b] | - | + | - | - |

a, grown on the media supplied with galactose and glucose
b, grown on the media supplied with mannose
*, indicating a very weak agglutination

# FIGURE 5.1

## SDS-PAGE analysis of LPS produced by various strains

LPS profiles of various strains in silver-stained SDS-PAGE. Whole membrane fractions were boiled for 5 min then treated with proteinase K and analysed by SDS-PAGE with LPS-specific silver staining. The lane designations are as follows:A, P9073; B, P9285; C, P9286; E, P9003.


A B C E

clones were transformed into P9285 and P9286. Strains P9324 and P9325, which are P9285 and P9286 respectively carrying pPR1381, give a phage sensitivity pattern identical to that of wild type strain P9003 (Table 5.1), and strong positive slide agglutinations with *S. enterica* serovar typhimurium specific antisera O4 and O5.

To confirm that strains P9324 and P9325 can produce LPS, whole cell membrane was prepared from these two strains, and analysed by silver staining after SDS-polyacrylamide gel electrophoresis. P9324 and P9325 give a ladder pattern of O antigen distribution similar to that of wild type strain P9003, although the amount of high molecular weight LPS from strains P9324 and P9325 is less than that from strain P9003 (Figure 5.2).

pPR1381 carries the right end region of strain LT2 *rfb* gene cluster from positions 19.7 to 21.7, and the only complete gene located in this region is *rfbP* which encodes the galactose transferase (Jiang *et al.*, 1991). The '*rfbT*' mutations are thus complemented by *rfbP* and we conclude that there is no separate *rfbT* gene, but that the 'T' mutations are in *rfbP*: these mutants are from here on referred to as *rfbP(T⁻)* mutants. The mutation in strain P9285 is named as *rfbP4451(T⁻)*, and the leaky mutation in strain P9286 is named as *rfbP4452(T⁻)*.

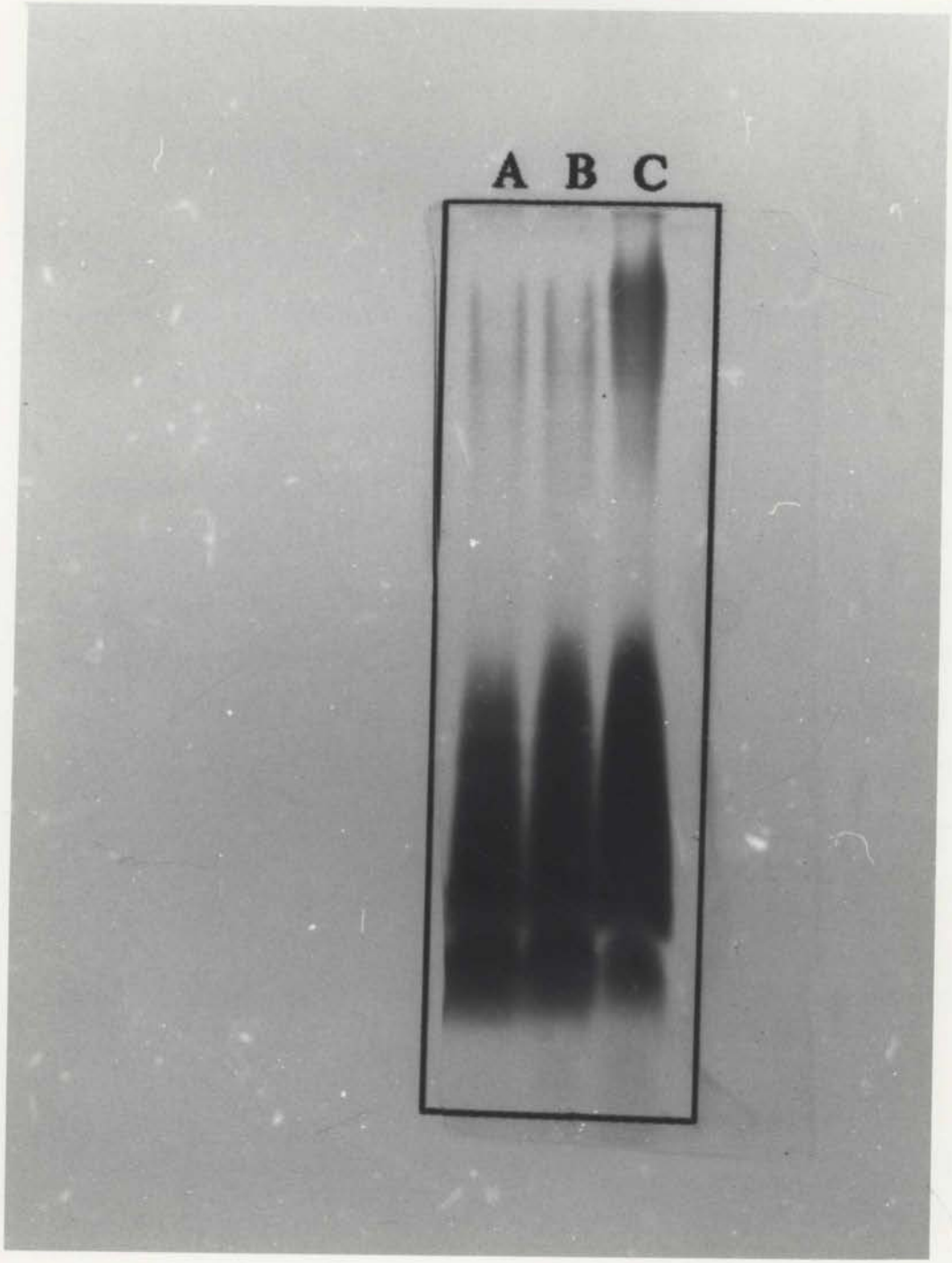
## 5.4 *rfbP(T⁻)* MUTANTS HAVE GALACTOSE TRANSFERASE ACTIVITY

The known function of the protein encoded by *rfbP* is to transfer galactose to undecaprenol with a diphosphate linkage (galactose transferase function). The 'T' function appears unrelated but it is important to see if there is any connection.

# FIGURE 5.2

## SDS-PAGE analysis of LPS produced by various strains

LPS profiles of various strains in silver-stained SDS-PAGE. Whole membrane fractions were boiled for 5 min then treated with proteinase K and analysed by SDS-PAGE with LPS-specific silver staining. The lane designations are as follows:A, P9324 (P9285 carrying pPR1381); B, P9325 (P9286 carrying pPR1381); C, P9003.


A B C

Galactose transferase activity of strains P9285 and P9286 was assayed using the whole membrane fraction of an overnight culture as described in chapter two. P9285 and P9286 have about 5-7% the activity of P9003 (Table 5.2). Thus, these two strains retained galactose transferase activity, as expected since it is needed to synthesise O hapten, but it was apparently reduced.

It is possible that the 'T' block, which leads to accumulation of O hapten, has a feedback effect on *rfbP* transcription, and that the reduced galactose transferase activity is due to reduced level of enzyme rather than reduced enzyme kinetics. We therefore introduced a *galE* mutation which would prevent accumulation of any O unit intermediate in these mutants in the absence of extrogenous galactose (see chapter one). It was found that in the wild type strain the *galE* mutation has little effect on the level of galactose transferase activity, but that in the *rfbP(T$^-$)* mutants the level of galactose transferase activity increased about two fold (Table 5.2). It appears that there is some feedback effect due to accumulation of O hapten, but that the intrinsic activity of the enzyme is reduced by the 'T' mutation.

## 5.5 CHARACTERIZATION OF THE O-SPECIFIC PRODUCTS OF *rfbP(T$^-$)* STRAINS

To further characterise the 'T' function of this protein, the quantity and the properties of the O hapten accumulated in *rfbP(T$^-$)* strains were determined. To do this, *pmi* and *galE* derivatives of *rfbP(T$^-$)* strains were isolated and used for the labelling of the O hapten, which was then isolated and analysed by chromatography. The use of *galE* and *pmi* mutants to label O antigen was pioneered by Kent and Osborn (1968a, 1968c), and has been widely used since then (Osborn *et al.*,

**Table 5.2.** Galactose transferase activity of *rfbP(T⁻)* strains

| Strains | Relavent genotype | In vitro transferase activity ($^{14}$C-UDP-gal. incorporated into CHCl3-methanol (2:1) soluble material /30 min/200 ng protein) |
|---|---|---|
| | | CPM |
| P9029 | *Δ(hisD-rfb)* | 18 |
| P9285 | *rfbP(T⁻)* | 209 |
| P9286 | *rfbP(T⁻)* | 140 |
| P9320 | *rfbP(T⁻), galE* | 548 |
| P9321 | *rfbP(T⁻), galE* | 577 |
| P9003 | wild type | 3378 |
| P9138 | *galE* | 3165 |

1972a, 1972b; Kopmann and Jann, 1975; Kanegasaki and Jann, 1979; Marino *et al.*, 1991; MacGrath and Orsborn, 1991a, 1991b). It is believed that under normal conditions of growth, O antigen or LPS can be specifically labelled by using extrogenous radioactive mannose or galactose.

### 5.5.1 INTRODUCTION OF A *galE* MUTATION INTO VARIOUS *S. ENTERICA* LT2 STRAINS

As mentioned in chapter one, the *galE* gene encodes the galactose epimerase which catalyzes the reversible interconversion of UDP-glucose and UDP-galactose. Galactose is a constituent of O antigen, LPS core and the M antigen capsule (Mäkelä and Stocker, 1984; Markovitz, 1977). M antigen capsule is made only under certain physiological conditions such as low temperature, high concentration of salts and excess of fermentable sugars (Markovitz, 1977). Thus, under the usual conditions of growth, bacteria do not produce M antigen capsule, and the O antigen and core of *galE* strains can be specifically labelled by adding radioactive galactose in the medium (Kent and Osborn, 1968c).

*S. enterica* LT2 strain P9138 carries a transposon tn10($Tet^r$) closely linked to *galE*$^-$, so the transfer of *galE*$^-$ of this strain could be detected easily. Phage ES18 was used to transduce the *galE*$^-$ region of P9138 to other LT2 strains: an ES18 lysate grown on strain P9138 was used to infect target strains, and recombination between *galE*$^-$ DNA in the lysate and *galE* DNA on the chromosome of target strain will introduce the *galE*$^-$ allele into the chromosome of the target strain, and it can be selected by selection for the closely linked tn10 tetracycline resistant gene.

*galE*$^-$ derivatives of strains P9285, P9286, P9073 and P9300 were isolated, and named P9320, P9321, P9318 and P9322 respectively. They all showed sensitivity to galactose, confirming that they are *galE* mutants. All four strains show the same phage pattern as that of their parent strains on galactose-glucose media.

### 5.5.2 INTRODUCTION OF A *pmi* MUTATION INTO VARIOUS *S. ENTERICA* LT2 STRAINS

As discussed in chapter one, the *pmi* gene encodes the phosphomannose isomerase which catalyses the isomerization of frucose-6-phosphate to mannose-6-phosphate which is then converted to GDP-mannose. The mannose of GDP-mannose then can be incorporated into the O antigen (Mäkelä and Stocker, 1984). Thus, *pmi* mutants can not synthesize O antigen without exogenous mannose. GDP-mannose can also be converted into GDP-L-fucose which is a constituent of M antigen capsule (Markovitz, 1977; Stevenson *et al.*, 1991). Under normal conditions of growth, bacteria do not produce M antigen capsule and O antigen of *pmi* mutants can be specifically labelled by adding radioactive mannose to the medium (Kent and Osborn, 1968a).

*S. enterica* LT2 strain P9306 has a phage Mu-d1 insertion in its *pmi* gene: Mu-d1 carries a ampicillin resistant gene. Because of the large size of Mu-d1, it is hard to move this *pmi* mutation around using phage ES18. Chromosomal DNA of strain P9306 was therefore transformed into target strains by electro-transformation, and transformants were selected on Mannose-EMB plates with ampicillin.

We attempted to introduce a *pmi* mutation into strains P9285, P9286, P9073 and P9300. Despite several attempts, only strain P9317, the *pmi*

derivate of strain P9286 was obtained. Strain P9317 shows the same phage pattern as that of its parent strain on mannose media.

### 5.5.3 LABELLING OF THE O-SPECIFIC PRODUCTS

10 ml cultures (about $3x10^8$ cells) of strains P9320, P9321, P9138, P9318 and P9322 were labelled with a mix of 1 nmol [$^3$H]galactose (23,265,840 cpm) and 4 nmol nonradioactive galactose for 1 min at 37°C as described in chapter two. Cells were washed with cold 50% acetone, and the radioactivity was counted. The labelled cells were then extracted with hot phenol as described in chapter two, and the aqueous phase which contains both the LPS and O antigen released from undecaprenyl pyrophosphate was collected. 2.5 volumes of cold ethanol was added to the aqueous phase, and pellet and supernatant from this ethanol treatment were separately collected. The counts of these two fractions are shown in Table 5.3. The pyrophosphate bond between O antigen and undecaprenol is broken by hot phenol: the labelled material in the ethanol supernatant is free O antigen released from undecaprenol, and the labelled material in the pellet is LPS including lipid A-core (MacGrath and Osborn, 1992a; Kent and Osborn, 1968b).

After a 1 min pulse, about 5,121,000 cpm (about 0.22 nmol [$^3$H]galactose) were incorporated into 10 ml cells of the wild-type strain P9138. 27%, 34%, 22% and 26% of that amount was incorporated into cells of strains P9320, P9321, P9322 and P9318 respectively (Table 5.3). The wild-type strain incorporated more galactose than all other strains showing its ability for rapid synthesis of the complete LPS. The leaky nature of the *rfbP(T$^-$)* mutation of strain P9321 may have contributed to this strain's higher level of galactose uptake than that of strain P9320.

# TABLE 5.3

## Incorporation of [$^3$H]galactose by *galE* strains under pulse and pulse-chase conditions

10 ml cultures (about $3x10^8$ cells) of stains P9320, P9321, P9138, P9318 and P9322 were labelled with a mix of 1 nmol [$^3$H]galactose (23,265,840 cpm) and 4 nmol nonradioactive galactose for 1 min at 37°C as described in chapter 2. The cells were harvested by centrifugation, washed twice with cold 50% acetone, and the total radioactivity incorporated was counted. The labelled cells were then extracted with hot phenol before the aqueous and phenol phases were separated by centrifugation. In each case there was about 8% to 20% incorporated radioactivity left behind in the phenol phase. 2.5 volumes of cold ethanol was added to the aqueous phase, and the pellet and supernatant from this ethanol treatment were separately collected and counted as described in chapter 2.

For pulse-chase experiment, cells were labelled as above before a 10000-fold excess of nonradioactive galactose was added to the culture and incubation continued for another 10 min. Cells were then washed and treated with hot phenol and 65% ethanol as above.

**Table 5.3.** Incorporation of [$^3$H]galactose by *galE* strains under pulse and pulse-chase conditions

| | | Pulse experiment | | | | Pulse-chase experiment | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 65% ethanol supernatant | | 65% ethanol pellet | | 65% ethanol supernatant | | 65% ethanol pellet | |
| Strains | Total counts incorporated | [$^3$H]Galactose (CPM) | % of Total | [$^3$H]Galactose (CPM) | % of Total | [$^3$H]Galactose (CPM) | % of Total | [$^3$H]Galactose (CPM) | % of Total |
| P9320 | 1,389,150 | 113,120 | 8.1% | 981,711 | 70.6% | 123,690 | 8.9% | 961,279 | 69.1% |
| P9321 | 1,768,557 | 148,120 | 8.4% | 1,489,444 | 84.2% | 44,100 | 2.5% | 1,605,100 | 90.7% |
| P9138 | 5,121,000 | 166,600 | 3.2% | 3,596,050 | 70.2% | 91,980 | 1.8% | 3,708,570 | 72.4% |
| P9322 | 1,148,781 | 714,980 | 62.2% | 240,833 | 20.9% | 707,140 | 61.5% | 231,686 | 20.2% |
| P9318 | 1,365,399 | 259,560 | 19.0% | 686,061 | 50.3% | 41,790 | 3.1% | 892,881 | 65.4% |

As mentioned above, the radioactivity in the pellet of the 65% ethanol treatment represents the complete LPS molecule or the lipid A-core. Much less radioactivity was found in the pellet of strains P9320, P9322 and P9318 than that found in wild-type strain P9138 (Table 5.3), as expected in the absence of complete LPS molecules in these strains. More radioactivity was found in the pellet of strain P9321 than those found in the pellets of strains P9320, P9322 and P9318, and it is presumably because the *rfbP(T⁻)* mutation in this strain is leaky and it can synthesize some complete LPS.

The radioactivity found in the 65% supernatant is derived from undecaprenol linked O antigen. 8.1%, 8.4%, 3.2%, 62.2% and 19.0% of the incorporated radioactivity was recovered from the ethanol supernatant of strains P9320, P9321, P9138, P9322 and P9318 respectively (Table 5.3). Wild-type strain P9138 does not accumulate undecaprenol linked O antigen, and 3.2% of the incorporated radioactivity was recovered from the ethanol supernatant in this strain as expected. Strain P9322 is the *galE* derivative of a *rfaL* strain, and 62.2% of the incorporated radioactivity was recovered from the ethanol supernatant. As it was believed that "*rfbT*" was involved in the same ligation reaction as *rfaL,* it was expected that a similar result would be found in strains P9320 and P9321. But only 8.1% and 8.4% of the incorporated radioactivity was recovered from the ethanol supernatant of strains P9320 and P9321 respectively. 19.0% of the incorporated radioactivity was recovered from the ethanol supernatant of deletion strain P9318, and this was also unexpected as this strain should not make any O antigen.

## 5.5.4 GEL FILTRATION OF THE ETHANOL SOLUBLE RADIOACTIVE MATERIALS

To further characterize the radioactive materials in the 65% ethanol supernatant, they were applied to a Sephadex G-15 column which had been calibrated using the synthetic disaccharide, trisaccharide and tetrasaccharide which were synthesized as described in chapter two.

The material from *rfaL* strain P9322 was excluded from the column (Figure 5.3), as expected for the long chain O antigen. Most of the material from the wild-type strain P9138 was also excluded from the column, but the radioactivity of the excluded material is about 10% of that of strain P9322.

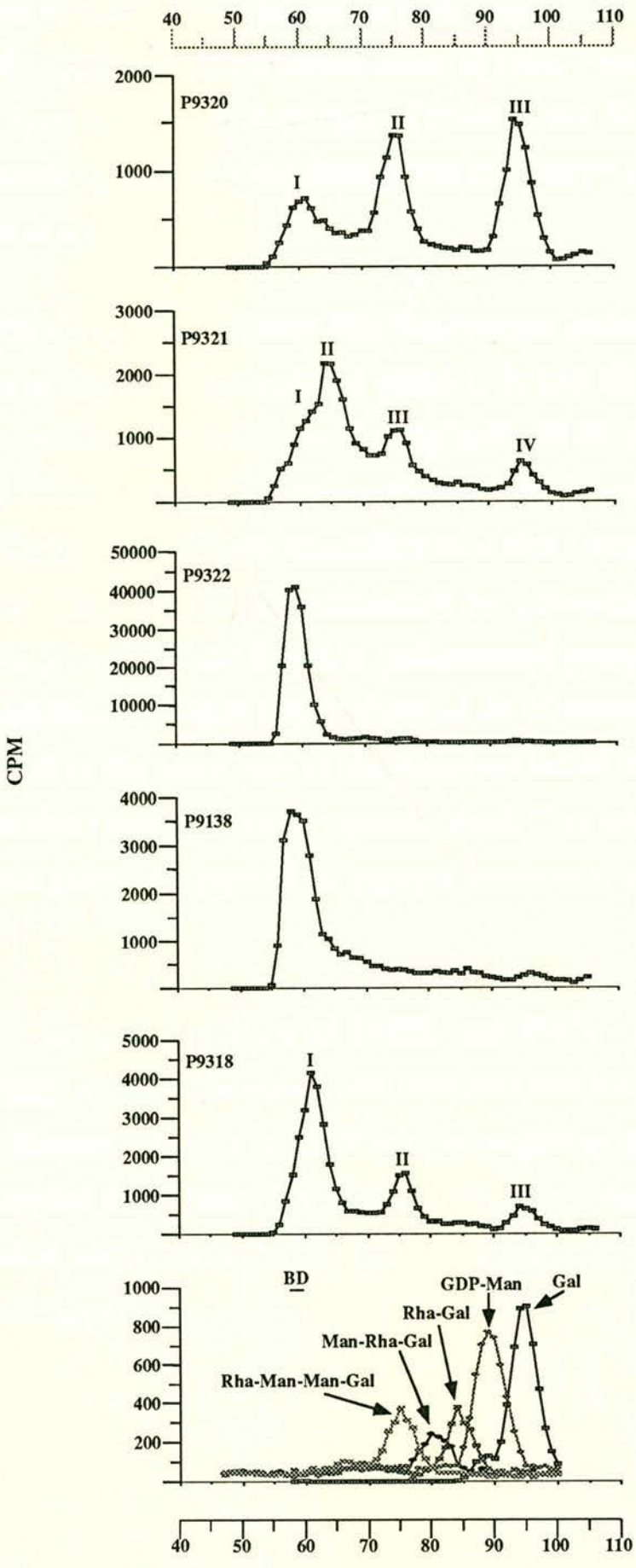
The material from strain P9320, the *galE* derivative of *rfbP(T$^-$)* strain P9285, gave three peaks (Figure 5.3). Peak I is slightly retarded with respect to blue detran, and could represent long chain O antigen or contaminating LPS or lipid A-core. The material in peak II runs at a position between a trisaccharide (Man-Rha-Gal) and a tetrasaccharide (Rha-Man-Man-Gal), and could be the tetrasaccharide single O unit (Abe-Man-Rha-Gal) of strain LT2. Peak III runs at the same position as galactose, and presumably is galactose. Peaks I, II and III contained approximately 18%, 38% and 44% of the total radioactivity in the 65% ethanol supernatant. The amount of radioactivity in peak I is about 0.2% of that found in the peak of strain P9322, and thus it can be concluded that this *rfbP(T$^-$)* strain makes none or extremely little long chain O antigen.

The material extracted from the other *rfbP(T$^-$)* strain, P9321, gave four peaks. Peak I runs at the same position as peak I of strain P9320.

# FIGURE 5.3

## Gel filtration of free O antigens accumulated by *galE* strains from the label experiment

Free O antigens accumulated by strains P9320, P9321, P9322, P9138 and P9318 from the label experiment were extracted with 45% phenol at 70°C as described in chapter two. The ethanol-soluble fraction was freeze-dried, and resuspended in 400 µl 10% CsCl, and 100 µl of which was then applied to a Sephadex G-15 column as described in chapter two. In vitro synthetic oligosaccharides were used as controls. The elution positions of galactose and GDP-mannose were determined by using [$^{3}$H]galactose and [$^{14}$C]GDP-mannose. The positions of disaccharide, trisaccharide and tetrasaccharide were determined by using in vitro synthetic [$^{14}$C-rhamnose]-galactose, [$^{14}$C-mannose]-rhamnose-galactose, and [$^{14}$C-rhamnose]-mannose-mannose-galactose, respectively. The excluded volume, determined with blue dextran (BD), is indicated.


40 50 60 70 80 90 100 110
2000
1000
P9320
I
II
III
3000
2000
1000
P9321
I
II
III
IV
50000
40000
30000
20000
10000
P9322
4000
3000
2000
1000
P9138
5000
4000
3000
2000
1000
P9318
I
II
III
1000
800
600
400
200
BD
GDP-Man
Gal
Rha-Gal
Man-Rha-Gal
Rha-Man-Man-Gal
40 50 60 70 80 90 100 110
CPM
Fraction No.

Peak II is slightly retarded with respect to peak I, and could represent the mixture of O polysaccharides with different O units. Peaks III and IV run at the same positions respectively as peaks II and III of strain P9320. This result aggress with the finding that the *rfbP(T-)* mutation in this strain is leaky.

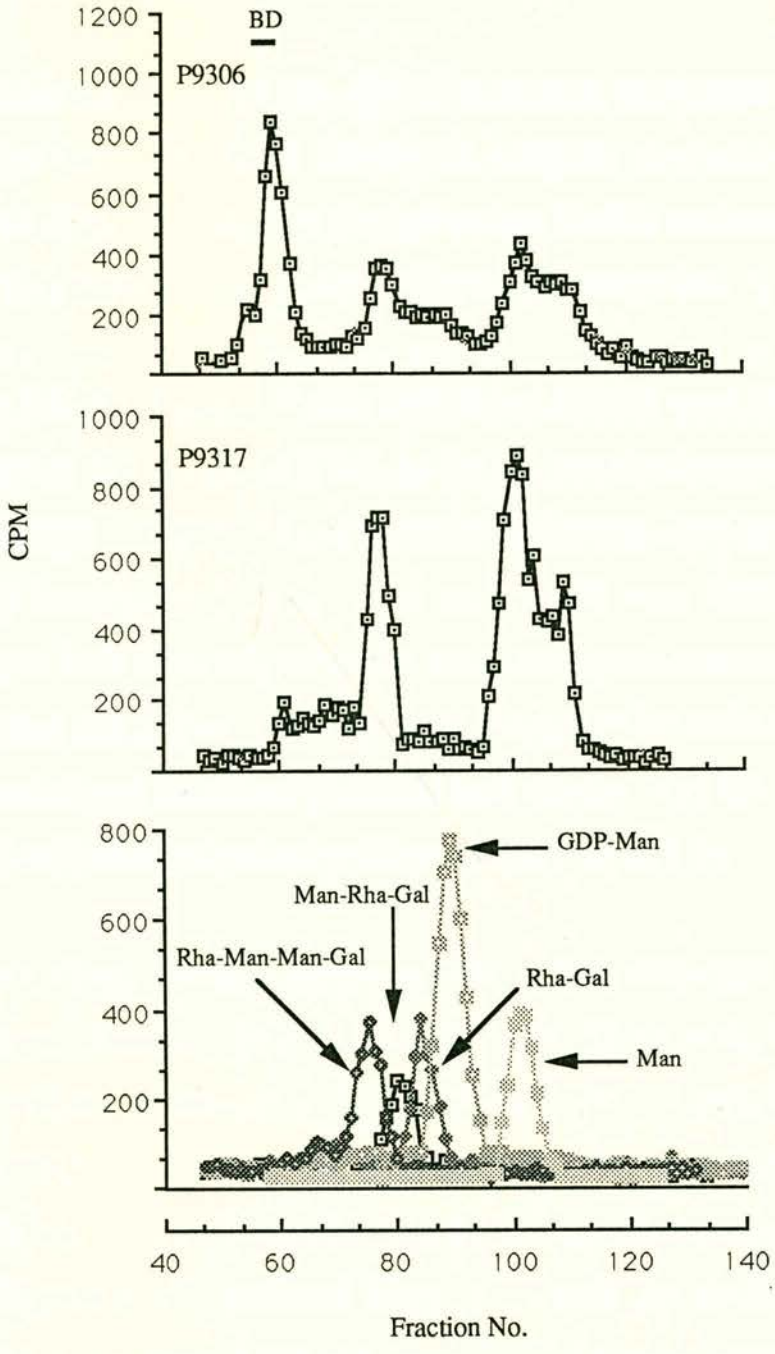
The ethanol soluble material of strain P9318 gave three peaks. Peaks I, II and III run at the same positions as peaks I, II and III of strain P9320. Strain P9318 does not produce O antigen due to the absence of *rfbP* and *rfbK* on its chromosome, and no ethanol soluble material was expected. As mentioned in chapter one, galactose is also a constituent of the M antigen capsule which is synthesized by polymerization of oligosaccharide repeating units linked to undecaprenyl pyrophosphate. Thus, these undecaprenol linked intermediates of the M antigen capsule will be extracted in the same way as undecaprenol linked O antigen, and the radioactivity seen here could represent them. If so strain P9318 must be at least partly derepressed for capsule production, although it is not mucoid.

Strains P9317, the *pmi* derivative of P9286, and P9306 (wild-type) were also labelled with [14C]mannose and extracted with hot phenol as described in chapter two. The ethanol soluble materials from these strains were applied to the same G-15 column and gave a similar result as that obtained with strains P9321 and P9138, respectively, using [3H]galactose (Figure 5.4).

# FIGURE 5.4

## Gel filtration of free O antigens accumulated by strains P9317 and P9306 from the label experiment

Free O antigens accumulated by P9317 and P9306 from the label were extracted with 45% phenol at 70°C as described in chapter two. The ethanol-soluble fraction was freeze-dried, and resuspended in 400 µl 10% CsCl, and 200 µl of which was then applied to a Sephadex G-15 column as described in chapter two. In vitro synthetic oligosaccharides were used as controls. The elution positions of mannose and GDP-mannose were determined by using [$^{14}$C]mannose and [$^{14}$C]GDP-mannose. The positions of disaccharide, trisaccharide and tetrasaccharide were determined by using in vitro synthetic [$^{14}$C-rhamnose]-galactose, [$^{14}$C-mannose]-rhamnose-galactose, and [$^{14}$C-rhamnose]-mannose-mannose-galactose, respectively. The excluded volume, determined with blue dextran (BD), is indicated.


BD
P9306
P9317
CPM
GDP-Man
Man-Rha-Gal
Rha-Man-Man-Gal
Rha-Gal
Man
Fraction No.

## 5.5.5 CHASE OF THE PREVIOUSLY ACCUMULATED ETHANOL SOLUBLE RADIOACTIVE MATERIALS

The above results suggest that the *rfbP(T⁻)* strains accumulate low molecular weight undecaprenol linked O antigen. A pulse-chase experiment was then carried out to determine if this accumulated material could be chased into high molecular weight material.

Cells of strains P9320, P9321, P9138, P9318 and P9322 were labelled with [3H]galactose for 1 min at 37°C as in the previous experiment before a 10,000-fold excess of nonradioactive galactose was added to the culture and incubation continued for another 10 min. Cells were washed with cold 50% acetone, extracted with hot phenol, treated with 65% ethanol, and the pellet and supernatant separately collected as before. Radioactivity of these two fractions was counted and the results were compared with those obtained without chase (Table 5.3).

After a 10 min chase, counts recovered from the wild-type strain P9138 as hapten decreased from 166,600 to 91,980 cpm (a loss of 74,620 cpm) while those in LPS increased from 3,596,050 to 3,708,570 (an accretion of 112,520, 37,900 cpm more than lost from the hapten). Thus at least 66.3% of the radioactivity incorporated into LPS during the chase was derived from previously accumulated O hapten, and the remainders from the free [$^3$H]galactose which had entered the pool. The results suggest that the previously accumulated O hapten can be chased into LPS.

In *rfbP(T⁻)* strain P9320, the counts of the radioactivity recovered from the pellet and the ethanol supernatant were similar to those

obtained after the 1 min label showing that the undecaprenol associated materials were not able to be chased into LPS.

In the leaky *rfbP(T⁻)* strain P9321, counts recovered from the ethanol soluble material decreased from 148,120 to 44,100 cpm (a loss of 104,020 cpm) while those in LPS increased from 1,489,444 to 1,605,100 (an accretion of 115,656, 11,636 cpm more than lost from the ethanol soluble material). Thus at least 89.9% of the radioactivity incorporated into LPS was derived from previously accumulated ethanol soluble material. This demonstrates effective chase of the previously accumulated hapten into LPS in the leaky *rfbP(T⁻)* strain.

In *rfaL* strain P9322, the counts of the radioactivity recovered from the pellet and the ethanol supernatant were similar to those obtained from the label experiment. The ligation step is blocked in this strain, thus the accumulated O hapten can not be chased into LPS.

In the deletion strain P9318, counts recovered from the ethanol soluble material decreased from 259,560 to 41,790 cpm (a loss of 217,770 cpm) while those in the pellet increased from 686,061 to 892,881 (an accretion of 206,820 cpm). Thus, at least 84% of the previously accumulated ethanol soluble material was chased into material which is insoluble in 65% ethanol.

## 5.6 SUMMARY AND DISCUSSION

In this chapter, the location and a functional analysis of the "*rfbT*" mutation is presented.

The two "*rfbT*" mutations were complemented by plasmid pPR1381 which carries *rfbP* gene encoding the galactose transferase, and it was concluded that the mutation must be in *rfbP* gene and these mutants were therefore referred to as *rfbP(T⁻)*.

Enzyme assay of the galactose transferase using membrane fractions showed that the two *rfbP(T⁻)* strains, P9285 and P9286, had galactose transferase activity of about 5-7% of that of wild-type strain.

*galE* and *pmi* derivatives of *rfbP(T⁻)* strains P9285 and P9286 were constructed and O antigen intermediates were labelled using radioactive sugars. The use of *galE* and *pmi* mutants to specifically label O antigen and LPS was pioneered by Kent and Osborn (1968a, 1968c), and has been widely used since then. Labelled cells were extracted with hot phenol, and LPS and O hapten were separated by differential ethanol precipitation and quantitated. To further characterize the O hapten, the ethanol soluble fraction was passed through a G-15 column. No undecaprenol linked long chain O antigen was found in one *rfbP(T⁻)* strain (P9320), and very small amount (compared to the amount found in a *rfaL* strain) was found in a leaky *rfbP(T⁻)* strain (P9321). This does not fit the model that "*rfbT*" and *rfaL* are both involved in ligation of O antigen to lipid A-core (Mäkelä and Stocker, 1984), in which case it would be expected both the *rfaL* and *rfbP(T⁻)* strains would accumulate undecaprenol linked long chain O antigen. Thus, it must be an earlier step in the O antigen biosynthetic pathway blocked rather than the ligation step.

Instead of accumulating undecaprenol linked long chain O antigen, *galE* derivative of strain P9285, P9320, accumulated undecaprenol linked low molecular weight materials (peak II) which run at a position

on the column close to that of the single O unit of group C2 strains (Rha-Man-Man-Gal) (Figure 5.3). The single O unit of *S. enterica* group B strains is also a tetrasaccharide (Abe-Man-Rha-Gal), but the sugar composition, linkages between sugars and thus the molecular shape is different to that of the group C2 single O unit. It would not be surprising if the mobility of these two tetrasacchrides is slightly different. Thus, we conclude that the low molecular weight material accumulated by strain P9285 must be the single O unit of group B. This accumulated material can not be chased through into 65% insoluble material as shown by the pulse-chase experiment, and it is likely that strain P9285 accumulates undecaprenol linked single O units, but the polymerization of single O units and ligation (to lipid A-core) are both blocked in this strain. The reason we say that both steps are blocked is because single O unit can be ligated to lipid A-core: thus the cells have to have both steps blocked to accumulate undecaprenol linked single O units. *rfc* is the only gene required for polymerization of O units (Mäkelä and Stocker, 1984; Collins and Hacket, 1991; Naide *et al.*, 1965), and the ladder pattern of LPS produced by strain P9324 (Strain P9285 containing pPR1381) shows that the *rfc* gene is normal in strain P9285. Thus, it must be that the undecaprenol linked single O units can not access the polymerization and ligation sites. As discussed in chapter one, the synthesis of O units on undecaprenol occurs at the cytoplasmic face of inner membrane, but the polymerization of O units and ligation both occur at the periplasmic face of the inner membrane. Thus, it is possible that in this strain the flipping of undecaprenol linked single O unit from cytoplasmic face to periplasmic face of the inner membrane is blocked, and that polymerization and ligation (to the lipid A-core) are therefor prevented.

Strain P9318 is the *galE* derivative of strain P9073 which has a deletion from *his* to *rfbK*, and it can not produce O antigen or LPS. After 1 min label, it produced significant amount of radioactive materials in both 65% ethanol supernatant and pellet, and the radioactivity in 65% ethanol can be chased into 65% ethanol insoluble materials. We suggest that these materials are the M antigen capsule or intermediates. The regulation of the *cps* genes is unclear, and it is possible that these genes are expressed in this strain under normal conditions. Some of the 65% ethanol soluble materials from this strain run at the same position as the peak II of strain P9320, and this raised the possibility that the low molecular weight materials accumulated in both *rfbP(T⁻)* strains are also intermediates of M antigen capsule. We believe it is highly unlikely that this is the case because the materials accumulated in strain P9320 can not be chased through. Further serological tests on these materials will be able to resolve this doubt. It does seem possible that the material in peak I of P9320 is also capsular hapten, as it has the same mobility as the material in peak I of P9318 (Figure 5.3). It should be possible to test this by mutating the *cps* gene cluster to block capsule synthesis, but unfortunately this could not be done in the time available. If peak I of P9320 was capsule, this would explain its presence in a strain which seems to have an absolute block in the 'T' function.

The other *rfbP(T⁻)* strain, P9286, showed weak agglutination with antisera O4 and O5, and produced a small amount of LPS, and it was concluded that the mutation in this strain is leaky. After 1 min label, this strain accumulated more low molecular weight materials than the wild-type strain, suggesting that the flipping procedure is blocked in this strain. It also accumulated materials (peak II) which are presumed

to be O polysaccharide with fewer O units than the full length O antigen, showing the block is not absolute.

Thus, we conclude that *rfbP* encodes a bifunctional protein and the 'T' function may be concerned with the flipping of the undecaprenol linked single O unit from the cytoplasmic face to the periplasmic face of the inner membrane.

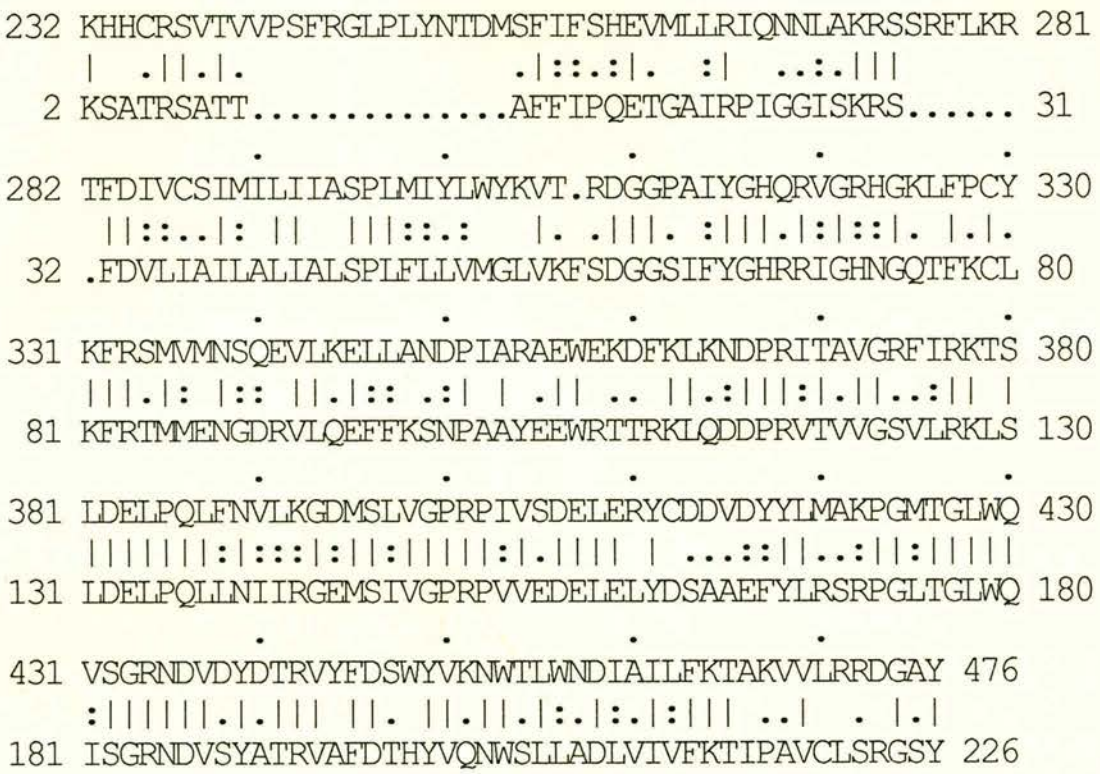
*rfbP* gene has been previously sequenced in this laboratory, and it is 1,428 bp in length (Jiang *et al.*, 1991). Analysis of the derived amino acid sequence reveals 4 transmembrane segments located at the N-terminal and another located in the middle of the protein with 140 residues in between. The C-terminal half of this protein including the last transmembrane segment is homologous to the whole amino acid sequence of *exoY* of *Rhizobium* with 48.7% identity and 66.1% similarity (Figure 5.5). The most likely function of the protein encoded by *exoY* is transfer of galactose to the lipid carrier which is the first reaction in the synthesis of the exopolysaccharide of *Rhizobium* (Reed *et al.*, 1991). A model for the protein encoded by *rfbP* in the inner membrane was predicted based on the "positive in" rule (Dalbey, 1990; Heijne, 1989), and is shown in Figure 5.6. The C-terminal end from residue 232 to 476 is likely to be the transferase domain as this part shows high level similarity to *exoY*. The N-terminal hydrophobic part of this protein may be embedded in the inner membrane as shown in Figure 5.6, and this part with the periplasmic domain (see Figure 5.6) may be concerned with 'T' function.

The discovery of a second function for *rfbP* raises the question as to whether it is really two genes, and that a sequencing error caused us to

# FIGURE 5.5

## Comparison of the deduced amino acid sequence of *rfbP* with that of *exoY*

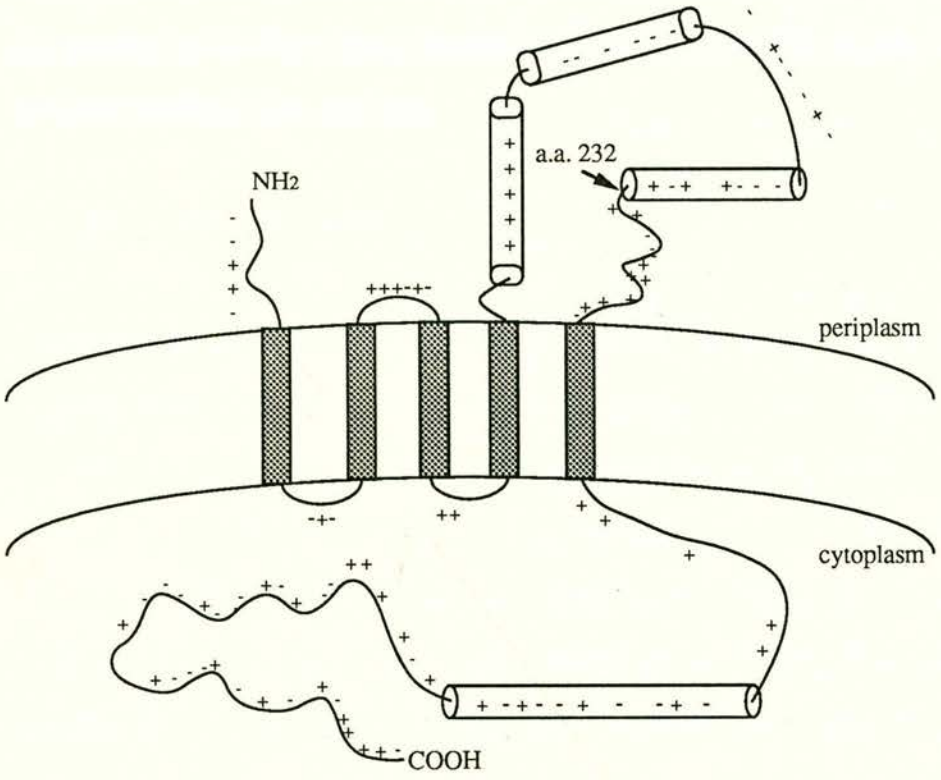
The deduced amino acid sequence of *rfbP* was aligned with that of *exoY* using program BESTFIT. The parameters for running BESTFIT were a gap weight of 3.0 and a gap length of 0.1. *rfbP* is on the top. Identity is indicated by a vertical line, amino acids with a comparison value greater than or equal to 0.5 are indicated by a colon, and those whose comparison value is greater than or equal to 0.10 are indicated by a period.


232 KHHCRSVTVVPSFRGLPLYNTDMSFIFSHEVMLLRIQNNLAKRSSRFLKR 281
2 KSATRSATT.....AFFIPQETGAIRPIGGISKRS..... 31
282 TFDIVCSIMILIIASPLMIYLWYKVT.RDGGPAIYGHQRVGRHGKLFPCY 330
32 .FDVLIAILALIALSPLFLLVMGLVKFSDGGSIFYGHRRIGHNGQTFKCL 80
331 KFRSMVMNSQEVLKELLANDPIARAEWEKDFKLKNDPRITAVGRFIRKTS 380
81 KFRTMMENGDRVLQEFFKSNPAAYEEWRTTRKLQDDPRVTVVGSVLRKLS 130
381 LDELPQLFNVLKGDMSLVGPRPIVSDELERYCDDVDYYLMAKPGMTGLWQ 430
131 LDELPQLLNIIRGEMSIVGPRPVVEDELELYDSAAEFYLRSRPGLTGLWQ 180
431 VSGRNDVDYDTRVYFDSWYVKNWTLWNDIAILFKTAKVVLRRDGAY 476
181 ISGRNDVSYATRVAFDTHYVQNWSLLADLVIVFKTIPAVCLSRGSY 226

# FIGURE 5.6

## Model for the arrange of the protein encoded by *rfbP* in inner membrane

Programs PepPlot of GCG and CHOU were used to predict the secondary structure. Program ALOM was used to predict transmembrane segments, and the algorithm of Eisenberg *et al.* (1984) was used to calculate the hydrophobic moment. This model is based on the above predictions. Each cylinder represents the region whose secondary structure is predicted as α-helix. The charges of the charged amino acid residues are shown.


NH2
a.a. 232
periplasm
cytoplasm
COOH

believe there to be one gene only. We have rechecked the sequence of *rfbP* and found no obvious site for the junction in terms of a possible stop codon and start codon: we have done this on the basis that an error (or mutation in the clone sequenced) is likely to be of the 'frameshift' or base substitution nature.

# CHAPTER 6: FINAL DISCUSSION AND CONCLUSION

## 6.1 INTRODUCTION

This thesis describes the molecular analysis of the *rfb* gene cluster of *S. enterica* serovar anatum strain M32 (group E1) as a representative of group E *S. enterica* strains. The *rfb* gene cluster of *S. enterica* groups B (strain LT2), D (strain Ty2), C1 (strain M40) and C2 (strain M67) have been studied in this laboratory (Jiang *et al.*, 1991; Verma *et al.*, 1989; Liu *et al.*, 1991; Lee *et al.*, 1992; Brown *et al.*, 1992). The roles of some of the M32 *rfb* genes are discussed. Comparison of the *rfb* gene cluster of strain M32 with those of other groups, and the genetic basis of O antigen variation in groups E1, B and D are also discussed.

This thesis also describes the finding of the second function (the 'T' function) of the protein encoded by *rfbP*, and the possible role of this protein in the flipping of the undecaprenol linked single O unit is discussed.

## 6.2 CLONING AND SEQUENCE OF THE *RFB* GENE CLUSTER OF STRAIN M32

The *rfb* gene cluster of *S. enterica* strain M32 (serovar anatum, group E1) has been cloned as four overlapping clones. Several attempts to clone the whole M32 *rfb* gene cluster in one piece were unsuccessful, and hence the region required for function has not been directly

determined. Comparison of the cloned *rfb* region of strain M32 with the *rfb* gene cluster of strain LT2 by Southern hybridization and restriction mapping showed that both ends of the LT2 *rfb* region are conserved in strain M32. This indicates that the cloned M32 DNA covers the entire *rfb* region of strain M32.

The regions unique to strain M32 and the regions showing only a low level of similarity to LT2 *rfb* genes were sequenced: the new sequences extended into the regions nearly identical in strain LT2 for which the entire *rfb* region has been sequenced, and thus in fact the sequence of the whole *rfb* gene cluster of strain M32 has been obtained.

The central region of the M32 *rfb* region from position 7,542 to position 11,877 does not show any sequence similarity to the LT2 *rfb* sequence. This central region contains *orf7.9*, *orf9.6* and *orf10.8*. Another small region of strain M32 from position 17,188 to position 18,697 is not present in strain LT2 either, and *orf17.4* is located in this region (see Figure 4.17).

The flanking regions are almost identical between strains M32 and LT2: the region from positions 0.0 to 6.9 of strain M32 shows 98% identity to that of strain LT2, and the region from position 12,823 to position 17,187 of strain M32 shows about 95% identity to the region from 17,383 to position 21,747 of strain LT2 (see Figure 4.17).

The DNA in the homologous regions adjacent to the junctions between central and flanking regions shows the level of similarity between strains M32 and LT2 decreasing markedly from a very high level as the junction and the central region is approached. The M32 DNA in the region from position 6.7 to position 7.5 can be easily aligned with the

LT2 DNA from position 6.7 to position 7.5 with the level of similarity decreasing from about 98% (around position 6.7) to about 50% (around position 7.5). The level of similarity between the M32 DNA from position 11.8 to position 12.8 and the LT2 DNA from position 16.4 to position 17.4 increases from about 50% (around position 11.8 of strain M32) to about 95% (around position 12.8 of strain M32) (see Figure 4.17).

Most of the *rfb* genes of strain LT2 have been identified in this laboratory (Wyk and Reeves, 1989; Jiang *et al.*, 1991). Comparing the O antigen structures of groups E1 and B (Chapter 1), we expect that the TDP-rhamnose and GDP-mannose biosynthetic pathway genes and the galactose transferase and rhamnose transferase genes to be present in both LT2 and M32; but the CDP-abequose biosynthetic pathway and transferase genes of strain LT2 to be absent in strain M32, and the mannose transferase genes of strains M32 and LT2 to be different. This expectation is borne out as discussed below.

## 6.3 GENES COMMON TO GROUPS E1 AND B

The genes which have been identified in group B as encoding the TDP-rhamnose and GDP-mannose biosynthetic pathways, the galactose transferase and the putative rhamnose transferase, are all conserved in strain M32 (Chapters 3 and 4).

Most of the conserved genes are essentially identical in the two strains. However the M32 genes (*orf7.0* and *orf11.9*) which are located in the junctions between flanking and central regions show substantial differences from their corresponding genes (*orf7.0* and *orf16.5*) of strain LT2. The deduced amino acid sequences of these two M32 genes show

high level of identity and similarity to those of their corresponding genes in strain LT2, and it is believed that their functions are the same as those of strain LT2.

## 6.4 STRAIN M32 LACKS ABEQUOSE BIOSYNTHETIC PATHWAY GENES

In strain LT2, three of the CDP-abequose pathway genes (*rfbF*, *rfbG* and *rfbJ*) have been identified, and they are located in the central region of LT2 *rfb* gene cluster (Jiang *et al.*, 1991). Comparison of the amino acid sequences of the protein products of these three genes with those of *orf7.9*, *orf9.6* and *orf10.8*, which are located in the central region of the M32 *rfb* region, did not reveal any homology. The other two CDP-abequose pathway genes in strain LT2 have not yet been identified but comparison of *rfb* clusters of groups B and C2 (Brown *et al.*, 1992) suggests that they are *orf7.6* and *orf10.4*, both of which are absent in strain M32. It appears that as expected the whole CDP-abequose biosynthetic pathway is absent in strain M32.

## 6.5 GENES UNIQUE TO GROUP E1

### 6.5.1 *orf7.9p* (A PUTATIVE MEMBRANE PROTEIN)

*orf7.9* is one of the three ORFs in the central region with no detectable similarity at nucleotide sequence level to genes of strain LT2. Structural analysis of *orf7.9p* predicts twelve transmembrane segments distributed throughout its length, suggesting that the whole protein is embedded in the membrane. No part seems likely to be hydrophilic, suggesting that this protein is not a transferase as we would expect a hydrophilic domain for the active site.

A gene encoding a protein with 12 predicted transmembrane segments has been found in all the *rfb* gene clusters of *S. enterica* that have been studied (Liu *et al.*, 1991; Jiang *et al.*, 1991; Lee *et al.*, 1992; Brown *et al.*, 1992). *orf7.9* of strain M32 fits this pattern and is in the same position on the *rfb* gene cluster as *orf12.8* and *orf12.6*, which are the corresponding genes in strains LT2 and M67 (group C2) respectively. These data suggest that these genes may have a similar function although they share no detectable similarity at DNA level between each other, and only little amino acid sequence similarity.

The presence of such a gene in all *rfb* gene clusters studied so far suggests that it has an important function, and the lack of good sequence similarity with the corresponding genes in other *rfb* regions suggests that this function is specific for the different O antigens produced.

### 6.5.2 *orf17.4* (O ANTIGEN POLYMERASE GENE ?)

The *rfc* gene encodes the O antigen polymerase. By crossing between *rfc*$^-$/*rfb*$^+$ and *rfc*$^+$/*rfb*$^+$ strains, and between *his*$^-$/*rfc*$^-$ and *his*$^+$/*rfc*$^+$ strains, the *rfc* gene was located in the *rfb* region in *S. enterica* strain SH1660 (serovar muenster, group E1) (Nyman et al., 1979). Thus, it was expected that a *rfc* gene would be found in the sequenced M32 *rfb* region.

*orf17.4* shows significant similarity to the sequence of *S. enterica* strain C5 (serovar typhimurium, group B) *rfc* gene (COLLINS and HACKETT 1991). We propose that *orf17.4p* catalyzes O antigen polymerization in strain M32.

### 6.5.3 *orf9.6* AND *orf10.8*

*orf9.6* and *orf10.8* do not show any significant similarity with LT2 *rfb* genes at either nucleotide or amino acid level.

Preliminary data shows that *orf10.8* is the mannose transferase gene in strain M32 (D. Liu and P.R. Reeves, unpublished data). The linkages of mannose to rhamnose differ in the O antigens of strains M32 and LT2, and hence we expect the mannose transferase gene of strain M32 to differ from that of strain LT2.

*orf9.6* encodes a highly hydrophobic protein with ten potential transmembrane segments. The function of this protein is unknown.

## 6.6 STRUCTURE OF THE *RFB* GENE CLUSTER OF STRAIN M32

The 19kb *rfb* gene cluster of strain M32 has been cloned as four overlapping clones, and hence the minimum region required for function has not been directly determined.

The strain LT2 *rfb* gene cluster extends from the end of *orf1.3* to the end of *rfbP* (position 21.7) with no DNA outside this region needed for function (Jiang *et al.*, 1991). The region from positions 0.0 to 6.9 is conserved in strain M32 and the *rfb* gene cluster of strain M32 probably starts from the end of *orf1.3* as in strain LT2. The 141 bp sequence between *rfbP* (the last *rfb* gene in LT2) and *gnd* in strain LT2 (Jiang *et al.*, 1991) is replaced by a 1,510 bp sequence in strain M32 (Chapter 4). The non identity begins 3 bases after the end of *rfbP* and ends 22 bases before the start of *gnd*. *orf17.4* (in M32) is located in this region and has

considerable homology with the *rfc* gene of group B (Chapter 4). We have not yet confirmed that *orf17.4* (in M32) encodes the O antigen polymerase, but if so then the '*rfb*' region (including *rfc*) of strain M32 will extend to position 18,465 (the end of *orf17.4* ).

Of five ORFs unique to the *rfb* region of strain M32, two (*orf10.8* and *orf11.9*) do not have Shine-Dalgarno sequences (Chapter 4 and Table 4.2). *orf10.8* starts with an ATG codon two bases upstream of the stop codon of *orf9.6*, and the putative start codon of *orf11.9* lies three bases after the stop codon of *orf10.8*: these two ORFs are presumably initiated by translational coupling. *orf7.9* and *orf9.6* have three base Shine-Dalgarno sequences and no possibility of translational readthrough, and are perhaps translated at low efficiency. *orf17.4* has a good Shine-Dalgarno sequence (Chapter 4 and Table 4.2).

In strain LT2, two potential promoters in the flanking conserved region with the mRNA startpoints at positions 2803 and 3927 were predicted using the five programs from Dr. O'Neill (Jiang *et al.*, 1991). Two with the mRNA startpoints at positions 9502 and 9563 were found in sequenced M32 *rfb* region using the same programs. We suggest that the M32 *rfb* gene cluster is a single operon translated from the same promoter as strain LT2, but may have an additional internal promoter with mRNA startpoint at position 9502 or 9563.

## 6.7 EVOLUTION OF THE *RFB* GENE CLUSTER

The M32 *rfb* region is of low G+C content. The sequenced region has a G+C content ranging from 0.29 to 0.37, and the ORFs have a codon usage characteristic of low G+C species, with P3 being lower than P1 or P2 (Chapter 4 and Table 4.5) (P1, P2 and P3 are the corrected average

G+C contents for bases 1, 2 and 3 of codons; Sueoka, 1988). The regions common to strains M32 and LT2 are also of low G+C content and have the codon usage of low G+C content strains (see Jiang *et al.*, 1991). The G+C content of a gene reflects the long term effects of the balance between the mutation rates from G.C to A.T and A.T to G.C (Sueoka, 1988). We might expect that once a gene is in *S. enterica*, the mutational pressure which normally maintains a given G+C content would bring this value to about 0.5 which is usual for *S. enterica* genes. To account for the low G+C content and atypical (for *S. enterica*) codon usage of the group B *rfb* gene cluster, it has been proposed that *rfb* gene cluster had been transferred to the *S. enterica* strain LT2 from a low G+C content species (Wyk and Reeves 1989; Jiang *et al.*, 1991), and the same argument applies to the *rfb* gene cluster of group E1.

There has in general been substantial synonymous substitution involving the third base since *Escherichia* and *Salmonella* diverged (Ochman and Wilson, 1987), and had the *rfb* gene cluster been in *Salmonella* during this period, then the G+C content of the third base would have been adjusted to approximate that of these two genera. The fact that P3 is typical of a low G+C content organism suggests that the *rfb* gene cluster has been in *S. enterica* for only a small part of the time since the divergence.

The regions common to strains LT2 and M32 show very little difference and must be derived from a common ancestral sequence, but the group specific central regions, in contrast, have little in common and at most one of the two forms can have been associated with the ancestral flanking regions. It is not possible with only the evidence of the two current forms to say if both central regions have recently become

associated with the flanking regions or if one has been associated for much longer and was replaced by the other in a relatively recent event.

The junctions between the central and flanking regions are interesting. *orf7.0* and *orf11.9* are adjacent to the central region in strain M32, as are *orf7.0* and *orf16.5* in strain LT2 (Figure 4.17). The level of similarity between these two gene clusters decreases from within *orf6.1* of both strains up to the junction, and a similar gradient is also found at the other junction with the level of similarity between *orf11.9* (M32) and *orf16.5* (LT2) increasing from the junction up to the ends of these ORFs (Chapter 4 and Figure 4.18). There have been several reports of intragenic recombination during the evolution of *E. coli* (Dubose *et al.*, 1988; Bisercic *et al.*, 1991, Neilson *et al.*, 1991). Such events allow those regions of *rfb* gene clusters which are common to two forms to undergo random genetic drift as one, while the non homologous regions must drift independently, as each must be independently maintained and there is no possibility of genetic recombination within these regions. We suggest that at the time of the recombination event which generated the junctions, *orf7.0* (M32) and *orf11.9* (M32) were very similar to *orf7.0* (LT2) and *orf16.5* (LT2) respectively, but proximity to the regions of non homology reduced homologous recombination, such that the neutral mutations accumulated in the two *rfb* gene clusters are different. The gradient reflects the increasing possibility of recombination with distance from the central region. The abrupt drop in the level of difference at about position 6,850 (Figure 4.18) could be due to a recent recombination between two forms of *rfb* clusters at this site (Event 3 in Figure 6.1).

The *rfb* gene clusters of strains M32 and LT2 have two regions of non homology. One of them, *orf7.9* to *orf10.8* of strain M32 and *orf7.6* to *orf15.4* of strain LT2, is flanked at both ends by homologous DNA with a gradient of similarity; but the other, *orf17.4* of strain M32, is flanked by DNA which is very similar to LT2 sequence right up to the junctions. This difference may reflect a difference in age of the recombination event involved; but if *orf17.4* does indeed encode the group E1 O antigen polymerase, it must have evolved with the *orf7.9* to *orf10.8* region which includes some of the transferase genes. The *rfc* gene of strain LT2 (also C5) maps at 32 min. (Sanderson and Roth, 1989), well away from the *rfb* region, and the same applies to strains in the related groups A and D. The separation of the *rfc* gene in these groups from the *rfb* region is perhaps secondary, as the *rfc* gene is also of low G+C content (Collins and Hackett, 1991), and probably evolved in association with the *rfb* gene cluster of these groups. The *rfc* gene of *S. enterica* strain M67 (group C2) is also located in the *rfb* gene cluster, and has been sequenced recently (Brown et al., 1992).

The level of difference in the common genes, away from those located in the junctions, is approximately 1% to 5% based on restriction enzyme data, and 1.2% based on sequence data where available: This is within the normal range for pairwise comparison between genes in the related *E. coli* (Bisercic *et al.*, 1991; Dubose *et al.*, 1988; Milkman and Crawford, 1983; Milkman and Bridges, 1990), and in *S. enterica* (Nelson *et al.*, 1991). This suggests that the genetic exchange, required for random genetic drift, occurs at normal frequency in the regions away from the junctions, although the sequences would only be present in strains carrying *rfb* regions of the closely related groups A, B, C2, D, and E.

It is not easy to put specific times to the recombination events involved, but if we are correct in our interpretation of the gradient of similarity, then there has been time for substantially more substitution (up to 50%) than has occurred in average genes (about 15% to 20%) since *Salmonella* and *Escherichia* diverged. We have then recombination events in the divergence of the two *rfb* gene clusters under discussion which predate divergence between *E. coli* and *S. enterica*, although both forms appear to have been transferred to *S. enterica* since that date. This supports the suggestion that the recombination events involved in the divergence of the *rfb* gene clusters of group B and D may have occurred in the donor low G+C species (Verma and Reeves, 1989; Jiang *et al.*, 1991). The differences between groups B and D, and between groups B and A lie entirely within the segment absent in group E1 (Chapter 4 and Figure 4.17). Thus although we have compared group E1 with group B, the recombination events could equally well have involved the *rfb* gene clusters of groups A and D. Comparison of the *rfb* gene cluster of groups B and C2 shows a similar pattern of similarity to that between groups B and E1: a central unique region of only limited homology is flanked by two similar regions (Brown *et al.*, 1991). We now have five *rfb* regions with substantial regions of homology, which are thought to have diverged in a low G+C species before lateral transfer to *S. enterica*. In the case of groups A, B, C2 and D, after one form was picked up by a *S. enterica* strain, the other form was able to enter by homologous recombination involved the homologous DNA at the two ends. However, the presence of *orf17.4*, unique to group E1, at the end of the gene cluster requires a more complex set of events unless the location of *orf17.4* is secondary.

Since many synonymous changes involve base 3 of the codon, the G+C content of base 3 can reflect the mutational pressure more directly. The corrected G+C contents for base 3 (P3) of the M32 genes located in regions with gradient of similarity (*orf7.0* (M32) and *orf11.9* (M32)) are slightly lower than those of their corresponding genes in strain LT2 (*orf7.0* (LT2) and *orf16.5* (LT2)) (Table 4 5). Looking at the synonymous substitutions, in 491 codons, 81 sites with A or T in strain M32 had G or C in strain LT2; and 61 sites with G or C in strain M32 had A or T in strain LT2. Synonymous substitution affects mostly coding base position 3, and it is expected that over time the value for P3 would approach 0.6, characteristic of *S. enterica* and *E. coli*: The data thus suggest that the *rfb* cluster of strain LT2 has been in *S. enterica* for longer than that of strain M32.

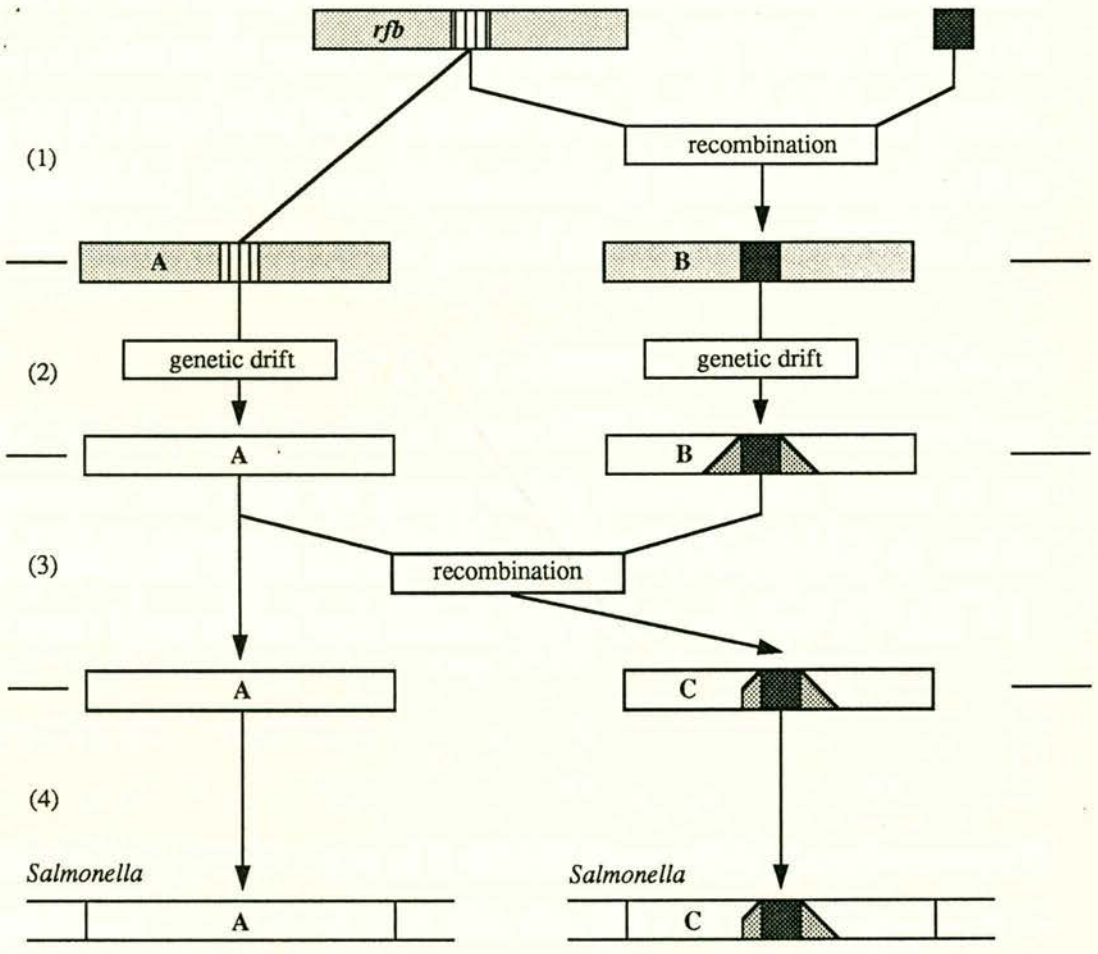
## 6.8 EVOLUTIONARY RELATIONSHIP OF *RFB* GENE CLUSTERS OF GROUPS E1 AND B

The evolutionary relationship of the *rfb* clusters of group B and group E1 *S. enterica* is shown in Figure 6.1. It appears that they arose from a common ancestral cluster in a low G+C content organism (represented by A in Figure 6.1), and that either one or both of these clusters had the central region replaced by recombination to give forms A and B. The homologous regions of the two forms could undergo recombination among different *rfb* gene clusters such that this part evolved by random genetic drift, with a level of neutral variation that is normal for bacteria. However, the two non homologous forms of the central regions have persisted, presumably reflecting selection for each form in different hosts, and hence each must drift independently. This independent drift must also apply to those parts of the homologous regions immediately

# FIGURE 6.1

## Possible origins of the *rfb* gene clusters of strains LT2 and M32

The two forms of the *rfb* gene cluster currently in strains LT2 and M32 are either of cluster A or cluster C, and they arose from a common ancestral. The whole evolutionary process is represented as four major events: (1), a recombination in which the central region of one cluster (shown crosslabeled) was replaced (the new central region is shown by black dots); (2), genetic drift in both clusters, with proximity to the new central region reducing recombination and giving rise to the nucleotide differences in the two clusters (level of difference is indicated by shading); (3), one recent recombination which led to form *rfb* cluster in form C; (4), transfer of clusters A and C into *S. enterica* strains.


rfb
recombination
(1)
A
B
genetic drift
genetic drift
(2)
A
B
recombination
(3)
A
C
(4)
Salmonella
A
Salmonella
C

adjacent to the non homologous central regions, to give the observed gradient of similarity. It should be noted that such a gradient was not observed at the junctions in the non homologous region from positions 17288 to 18697 of strain M32 or the junctions found when group D and group B *rfb* gene clusters were compared (Verma and Reeves, 1989; Verma *et al.*, 1988), perhaps because the recombination event in these regions were more recent. Both *rfb* gene clusters (forms A and C in Figure 6.1) must have been acquired by *S. enterica* by lateral transfer. We can not say which form represents the *rfb* gene cluster of group B or E1, but the one of group B was perhaps transferred into *S. enterica* first.

The putative O polymerase gene is adjacent to the *rfb* cluster in strain M32 but separated from it in strain LT2. The *rfc* genes must have coevolved with their *rfb* genes which determine the nature of the oligosaccharide which is polymerized, and the association of *rfc* and the *rfb* gene cluster seems more natural, but is not considered in Figure 6.1.

### 6.9 *RFBP(T)*

Both *rfaL* and *rfbT* mutants of *S. enterica* group B strains accumulate O hapten, and it was believed that the transfer of O antigen from undecaprenol to the core requires both *rfbT* and *rfaL* products. Three *rfbT* mutants of strain LT2 (group B) isolated by Wilkinson *et al.* (1972) are the only such strains available. Characterization of these three mutants by phage sensitivity tests and visualising the LPS produced by these strains on silver staining SDS-PAGE gel showed that strain P9285 has the expected phenotype but strain P9287 appears to have a secondary *rfa* mutation and the mutation in strain P9286 is leaky. Thus, strains P9285 and P9286 were used in this study. As discussed in chapter one, the mechanism of LPS synthesis in group E1 strains is the

same as that in group B strains, and the ligation and related function should be common to strains from these two groups.

### 6.9.1 *rfbT* MUTATION IS IN THE *rfbP* GENE

*rfbT* mutations were located in the region close to *his* by co-transduction experiments (Wilkinson *et al.*, 1972). To locate these mutations, a series of *rfb* clones which together cover the *rfb* region close to *his* was transformed into both mutants. It was found that pPR1381 can complement both the mutation strains. The only complete gene pPR1381 carries is *rfbP* which encodes the galactose transferase. Thus, these mutations must be in the *rfbP* gene and are referred to as *rfbP(T$^-$)* mutants.

The ligation function is not related to the galactose transferase function but it is important to see if there is any connection. Thus, the galactose transferase activity of the two *rfbP(T$^-$)* mutants and their *galE* derivatives was assayed. It was found that the two mutants, P9285 and P9286, have about 5-7% the activity of wild type strain P9003. It was also found that the activity increased about two fold when a *galE* mutation was introduced into the *rfbP(T$^-$)* mutant, while the *galE* mutation has little effect on the level of galactose transferase activity in the wild type strain. *galE* mutation prevents the accumulation of any O antigenic intermediate in the *rfbP(T$^-$)* mutants without extrogenous galactose. It appears that the 'T' block, which leads to accumulation of O hapten, has some feedback on *rfbP* transcription, but that the intrinsic activity of the enzyme is reduced by the 'T' mutation.

### 6.9.2 'T' FUNCTION IS NOT CONCERNED WITH THE LIGATION AS PREVIOUSLY BELIEVED

To study the 'T' function, the O hapten accumulated by *rfbP(T-)* strains was labelled and characterised. *galE* and *pmi* derivatives were used for the labelling, and the undecaprenyl pyrophosphate linked O antigen was released from the undecaprenyl pyrophosphate by hot phenol treatment and separated from LPS and core by 65% ethanol. The use of *galE* and *pmi* mutants to specifically label O antigen, and using hot phenol to release O antigen from undecaprenyl pyrophosphate and 65% ethanol to separate O antigen from LPS molecules were pioneered by Kent and Osborn (1968a, 1968c) and have been used since then as a standard method.

It was believed that the products of both "*rfbT*" and *rfaL* genes are required for the transfer (ligation) of O antigen from undecaprenyl pyrophosphate to lipid A-core, thus it was expected that both the *rfbP(T-)* and *rfaL* mutants would accumulate undecaprenol linked long chain O antigen. Characterization of the labelled O hapten by chromatography showed that the *rfaL* strain accumulated high molecular weight O antigen as expected, but almost no such high molecular weight O antigen was found in P9320 (*galE* derivative of *rfbP(T-)* strain P9285) and very little (compared to the amount found in *rfaL* strain) was found in P9321 (*galE* derivative of the leaky *rfbP(T-)* strain P9286). Thus, it must be an earlier step in the O antigen biosynthetic pathway blocked rather than the ligation step in *rfbP(T-)* strains, and the 'T' function is not concerned with the ligation step.

### 6.9.3 THE POSSIBLE 'T' FUNCTION

Instead of accumulating undecaprenol linked long chain O antigen, strain P9320 accumulates undecaprenol linked low molecular weight materials (peak II) which run at a position close to the position single O unit of group C2 strains runs from the column (Figure 5.3). The single O units of both groups B and C2 are tetrasaccharides but different to each other in the sugar composition and linkages between sugars, and it is expected that the mobility of them would be slightly different. Thus, it is likely that the materials (peak II) accumulated by strain P9285 are the undecaprenol linked single O units. We can not obtain radioactive CDP-abequose which is the last sugar in the tetrasaccharide single O unit of group B, and were unable to synthesise the complete single O unit and compare its mobility with that of the materials (peak II) accumulated by strain P9320. But, we have synthesized a trisaccharide with the same composition and linkages as the three sugar backbone of the group B single O unit, and have shown that it runs slower than the materials (peak II) accumulated in strain P9320 (Figure 5.3). A label-chase experiment was then carried out, and showed that the 65% ethanol soluble materials accumulated by strain P9320 can not be chased into LPS. Thus, we suggest that strain P9285 (parent strain of P9320) accumulates undecaprenol linked single O units.

As discussed in chapter one, *rfc* is the only gene responsible for the polymerization of undecaprenol linked single O units. The *rfc* gene is normal in strain P9285 as strain P9324 (P9285 containing pPR1381) produce LPS with long chain O antigen. Also, it is known that single O unit can be ligated to lipid A-core as *rfc* mutants produce LPS with a single O unit. Thus, the only step which can prevent the undecaprenol

linked single O units being polymerized and ligated to lipid A-core is to block the access of undecaprenol linked single O units to the polymerization and ligation sites. As discussed in chapter one the synthesis of O unis on the undecaprenyl pyrophosphate occurs at the cytoplasmic face of the inner membrane, but the polymerization and ligation occur at the periplasmic face of the inner membrane. Thus, it is likely that the flipping of the undecaprenol linked single O units from the cytoplasmic face to the periplasmic face is blocked in strain P9285, and 'T' function is concerned with this flipping procedure.

Strain P9321 (*galE* derivative of the leaky *rfbP(T$^-$)* mutant P9286) accumulated more low molecular weight materials than wild type strain did after 1 min label, showing the flipping step is also blocked in this strain. It also produced materials which are presumed to be O polysaccharide with less number of O units (compared to the full length O antigen of wild type strain), showing the block is not absolute.

Strain P9318 is the *galE* derivative of strain P9073 which has a deletion from *his* to *rfbK*, and it can not produce O antigen or LPS. It produced significant amount of labelled 65% ethanol soluble and insoluble materials after 1 min label, and it was also found that the radioactivity in 65% soluble materials can be chased into 65% insoluble materials. We suggest that the *cps* genes in this strain are at least partly derepressed for capsule production, and labelled materials found here are the capsular antigen or its intermediates. Some of the 65% ethanol soluble materials from this strain run at the same position as the peak II of strain P9320, and this raised the question that if the low molecular materials accumulated in both *rfbP(T$^-$)* strains are intermediates of capsular antigen. We believe that this is unlikely to be the case because

the materials accumulated in strain P9285 can not be chased through, and at least some of the materials accumulated by strain P9321 must be O antigen. It should be possible to mutate the *cps* genes to block the synthesis of the capsular antigen or to do the serological tests to confirm the nature of these materials, but unfortunately they could not be done in the time available.

Goldman and Hunt (1990) found that strain P9286 is a leaky 'T' mutant, and the amount of LPS produced by this strain is less than 10% of that produced by wild type strain. It agrees with our results. Using [$^{32}$p]orthophosphoric acid to label the LPS, they were able to detect the distribution of the small amount LPS molecules produced by this leaky mutant and showed the chain length of the O antigen produced by this strain is 20 to 22 O units compared to 28 to 29 of wild type strain (Goldman and Hunt, 1990). We failed to show the LPS molecules with long chain O antigen produced by strain P9286 (Figure 5.1) using silver staining, and the reason for this may be that silver staining is less sensitive than [$^{32}$P] labelling. Because more LPS molecules with single O unit are produced than LPS molecules with long chain O antigen, only the band represents LPS with single O unit can be visualised by silver staining in strain P9286 (Figure 5.1).

As discussed in chapter one, the chain length of the LPS molecule is controlled by the *cld* gene (Bastin *et al.*, 1992). It was proposed that the chain length determinant (CLD) (encoded by *cld* gene) and the O polymerase form a complex which has two states; "E" facilitating the extension of the O antigen, "T" facilitating the transfer of O antigen to core. The complex is postulated to enter the E state as O antigen polymerisation starts and to shift to the T state after a predetermined

time, CLD acting as a molecular clock. The *cld* gene is normal in strain P9286 as strain P9325 (P9286 containing pPR1381) produces LPS with the normal ladder pattern (Figure 5.2). We suggest that in strain P9286 the CLD would enter the T state after the same period of time as that in wild type strains, but the flipping of the undecaprenol linked single O units is partly blocked in this strain, and during that time less number of O units (compared to the wild type strain) are flipped across the inner membrane and become available for the polymerization. Thus, compared to the wild type strain, strain P9286 would have less O units being polymerized when the CLD starts to enter the T state to enable the ligation (to lipid A-core). Thus, full length O antigen on either lipid A-core or undecaprenol are not produced in strain P9286 (Figure 5.3).

The flipping function for 'T' put forward is clearly speculating as neither the single O unit nature of the material accumulated by both *rfbP(T⁻)* strains have been confirmed nor the direct evidence that these materials are accumulated on the inside of the inner membrane has been found. But it has the merit to explain the available experimental results.

### 6.9.4 DOES *rfbP* ENCODE A BIFUNCTIONAL PROTEIN ?

The C-terminal part of the derived amino acid sequence of *rfbP* shows high level of similarity to the entire amino acid sequence of *exoY* which was proposed to be the galactose transferase gene for the synthesis of exopolysaccharide in *Rhizobium*. We suggest that this part must be the galactose transferase domain in *rfbP*, and the N-terminal part could be concerned with 'T' function. The *rfbP* sequence has been rechecked, and no error was found. Thus, it is likely that *rfbP* gene encodes a protein with two domains responsible for two unrelated functions.

But, it is also possible that a mutation has occurred in the clone used for sequencing, and this mutation could be a deletion or a base substitution to join two open reading frames.

# APPENDIX

Material from this thesis has been published in the following journal:

Wang, L., L. K. Romana and P. R. Reeves. 1992. Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics. **130:**429-443

# BIBLIOGRAPHY

**Ames, G. F. and K. Nikaido**. 1976. Two-dimensional gel electrophoresis of membrane proteins. Biochemistry. **15:**616-623

**Anderson, M. S. and C. R. H. Raetz**. 1987. Biosynthesis of lipid A precursors in *Escherichia coli*: A cytoplasmic acyl transferase that converts UDP-N-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine. J. Biol. Chem. **262:**5159-5169

**Ardeshir, F., C. F. Higgins and G. F.-L. Ames**. 1981. Physical map of the histidine transport operon: correlation with the genetic map. J. Bacteriol. **147:**401-409

**Austin, E. A., J. F. Graces, L. A. Hite, C. T. Parker and C. A. Schnaitman**. 1990. Genetic analysis of lipopolysaccharide core biosynthesis by *Escherichia coli* K-12: insertion mutagenesis of the *rfa* locus. J. Bacteriol. **172:**5312-5325

**Bailey, M. J. A., V. Koronakis, T. Schmoll and C. Hughes**. 1992. *Escherichia coli* HlyT protein, a transcriptional activator of haemolysin synthesis and secretion, is encoded by the *rfaH* (*sfrB*) locus required for expression of sex factor and lipopolysaccharide genes. Mol. Microbiol. **6:**1003-1012

**Bastin, D. A., P. K. Brown, A. Haase, G. Stevenson and P. R. Reeves**. 1992. Repeat unit polysaccharides of bacteria: a model for polymerisation resembling those of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. **In Press**

**Bayer, M. E.** 1968. Areas of Adhesion between Wall and Membrane of *Escherichia coli*. J. Gen. Microbiol. **53:**395-404

**Bayer, M. E. and H. Thurow**. 1977. Polysaccharide Capsule of *Escherichia coli*: Microscope Study of Its Size, Structure, and Sites of Synthesis. J. Bacteriol. **130:**911-936

**Bayer, M. E.** 1979. The fusion sites between outer membrane and cytoplasmic membrane of bacteria; their role in membrane assembly and virus infection, p. 167-202. *In* M. Inouye (ed.), Bacterial outer membranes, John & Sons, New York.

**Beckmann, I., T. V. Subbaiah and B. A. D. Stocker**. 1964. Rough mutants of *Salmonella typhimurium*. Nature. **1:**1299-1301

**Birnboim, H. C. and J. Doly**. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7:**1513-1523

**Bisercic, M., J. Y. Feutrier and P. R. Reeves**. 1991. Nucleotide sequence of the *gnd* gene from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as contributing factor in the evolution of the polymorphic *gnd* locus. J. Bacteriol. **173:**3894-3900

**Boulnois, G. J. and K. Jann**. 1989. Bacterial polysaccharide capsule synthesis, export and evolution of structural diversity. Mol. Microbiol. **3:**1819-1823

**Boulnois, G. J. and I. S. Roberts**. 1990. Genetics of Capsular Polysaccharide Production in Bacteria. Cur. Top. Micro. and Immun. **150:**1-18

**Brahmbhatt, H., N. B. Quigley and P. Reeves**. 1986. Cloning part of the region encoding biosynthetic enzymes for surface antigen (O-antigen) of *Salmonella typhimurium*. Mol. Gen. Genet. **203:**172-176

**Brahmbhatt, H. N., P. Wyk, N. B. Quigley and P. R. Reeves**. 1988. Complete physical map of the *rfb* gene cluster encoding biosynthetic enzymes for the O antigen of *Salmonella typhimurium* LT2. J. Bacteriol. **170:**98-102

**Braun, V**. 1975. Covalent lipoprotein from the outer membrane of *Escherichia coli*. Biochim. Biophys. Acta. **415:**335-377

**Brenner, D. J.** 1984. Enterobacteriaceae, p. 408-420. *In* N. R. Krieg (ed.), Bergey's Manual of Systematic Bacteriology, Vol.1. Williams and Wilkins, Baltimore/London.

**Brown, P. K., L. K. Romana and P. R. Reeves**. 1991. Cloning of the *rfb* gene cluster of a group C2 *Salmonella*: comparison with the *rfb* regions of groups B and D. Mol. Microbiol. **5:**1873-1881

**Brown, P. K., L. K. Romana and P. R. Reeves**. 1992. Molecular Analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6:**1385-1394

**Bucholtz, C. A. and A. H. Reisner**. 1986. MBIS - an integrated system for the retrieval and analysis of sequence data from nucleic acids and proteins. Nucl. Acids Res. **14:**265-272

**Chou, P. Y. and G. D. Fasman**. 1974. Prediction of protein conformation. Biochemistry. **13:**222-245

**Churchward, G., D. Belin and Y. Nagamine**. 1984. A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. Gene. **31:**165-171

**Coleman, W. G**. 1983. The *rfa D* gene codes for ADP-L-glycero -D-manno- heptose-6-epimerase. J. Biol. Chem. **258:**1985-1990

**Coleman, J. and C. R. Raetz**. 1988. First committed step of lipid A biosynthesis in *Escherichia coli*: sequence of the *lpxA* gene. J. Bacteriol. **170:**1268-1274

**Collins, L. V. and J. Hackett**. 1991. Molecular Cloning, Characterisation, and Nucleotide Sequence of the *rfc* gene Encoding an O-antigen Polymerase of *Salmonella typhimurium*. J. Bacteriol. **173:**2521-2529

**Creeger, E. S. and L. I. Rothfield**. 1979. Cloning of genes for bacterial glycosyltransferase: selection of hybrid plasmid carrying genes for two glucosyltransferases. J. Biol. Chem. **254:**808-810

**Creeger, E. S., T. Schulte and L. I. Rothfield**. 1984. Regulation of membrane glycosyltransferases by the *sfrB* and *rfaH* genes of *Escherichia coli* and *Salmonella typhimurium*. J. Biol. Chem. **259:**3064-3069

**Crosa, J. H., D. J. Brenner, W. H. Ewing and S. Falkow**. 1973. Molecular Relationship Among the *Salmonella*. J. Bacteriol. **115:**307-315

**Crowell, D. N., W. S. Reznikoff and C. R. Raetz**. 1987. Nucleotide sequence of the *Escherichia coli* gene for lipid A disaccharide synthase. J. Bacteriol. **169:**5727-5734

**Dalbey, R. S**. 1990. Positively charged residues are important determinants of membrane protein topology. TIBS. **15:**253-256

**Darveau, R. P. and R. E. W. Hancock**. 1983. Procedure for isolation of bacterial lipopolysaccharide from both smooth and rough *Pseudomonas aeruginosa* and *Salmonella typhimurium* strains. J. Bacteriol. **155:**831-838

**Davis, D. B. and E. S. Mingoli**. 1950. Mutants of *Escherichia coli* requiring methionine or vitamine B12. J. Bacteriol. **60:**17-28

**Devereux, J., P. Haeberli and O. Smithies**. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res. **12:**387-395

**Dubose, R. F., D. E. Dykhuizen and D. L. Hartl**. 1988. Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA. **85:**7036-7040

**Eisenberg, D., E. Schwarz, M. Komaromy and R. Wall**. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179:**125-142

**Eliopoulos, E. E., A. J. Geddes, M. Brett, D. J. C. Pappin and J. B. C. Findlay**. 1982. A structural model for the chromophorebinding domain of ovine rhodopsin. Int. J. Biol. Macromol. **4:**263

**Felix, A. and B. R. Callow**. 1943. Typing of paratyphoid B bacilli by means of Vi bacteriophages. British Medical Journal. **II:**127-130

**Finke, A., D. Bronner, A. V. Nikolaev, B. Jann and K. Jann**. 1991. Biosynthesis of *Escherichia coli* K5 Polysaccharide, a Representative of Group II Capsular Polysaccharides: Polymerization In Vitro and Characteriization of the Product. J. Bacteriol. **173:**4088-4094

**Flemming, H. C. and K. Jann**. 1978a. Biosynthesis of the O9 antigen of *Escherichia coli*: Growth of the polysaccharide chain. Eur. J. Biochem. **83:**47-52

**Flemming, H. C. and K. Jann**. 1978b. Biosynthesis of the O8 antigen of *Escherichia coli*: Glucose at the reducing end of the polysaccharide and growth of the chain. FEMS Microbiol. Lett. **4:**203-205

**Gabriel, O. and L. V. Lenten**. 1978. The interconversion of monosaccharide. Int. Rev. Biochem. **16:**1-36

**Garger, S. J., O. M. Griffith and L. K. Grill.** 1983. Rapid purification of plasmid DNA by a single centrifugation in a two step caesium chloride-ethidium bromide gradient. Biochem. Biophys. Res. Commun. **177:**835-842

**Gold, L. and G. Stormo** 1987. Translational Initiation, p. 1302-1307. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*, 2. American Society for Microbiology, Washington, D.C.

**Goldman, R. C. and F. Hunt.** 1990. Mechanism of O-antigen distribution in Lipopolysaccharide. J. Bacteriol. **172:**5352-5359

**Green, N. M. and M. T. Flanagan.** 1976. The prediction of the conformation of membrane proteins from the sequence of amino acids. Biochem. J. **153:**729-732

**Heath, E. C., R. M. Mayer, R. D. Edstrom and C. A. Beaudreau.** 1966. Structure and biosynthesis of the cell-wall lipopolysaccharide of *Escherichia coli*. Ann. N. Y. Acad. Sci. **133:**315-333

**Heijne, G. V.** 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature. **341:**456-458

**Hellerqvist, G. G. and A. A. Lindberg.** 1971. Structure studies of the common-core polysaccharide of the cell wall lipopolysaccharide from *Salmonella typhimurium*. Carbohydr. Res. **16:**39-48

**Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene. **28:**351-359

**Hitchcock, P. J. and T. M. Brown.** 1983. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. J. Bacteriol. **154:**269-277

**Howley, P. M., M. A. Isral, M.-F. Law and M. A. Martin.** 1979. A Rapid Method for Detecting and Mapping Homology between Heterologous DNAs. J. Bio. Chem. **254:**4876-4883

**Hughes, K. T. and J. R. Roth.** 1984. Conditionally Transposition-Defective Derivative of Mu d1(Amo Lac). J. Bact. **159:**130-137

**Iseki, S. and S. Hamano.** 1958. The transduction and lysogenic conversion in *Salmonella* group E by means of Epsilon 34 phage. Gunma J. Med. Sci. **7:**1-6

**Ish-Horowicz, D. and J. F. Burke.** 1981. Rapid and efficient cosmid cloning. Nucleic Acids Res. **9:**2989-2998

**Jacobs, W. R., J. F. Barrett, J. E. Clark-Curtiss and R. Curtiss III.** 1986. In vivo repackaging of recombinant cosmid molecules for analyses of *Salmonella typhimurium, Streptococcus mutans,* and mycobacterial genomic libraries. Infect. Immun. **52:**101-109

**Jann, B., K. Reske and K. Jann.** 1975. Heterogeneity of lipopolysaccharides. Analysis of polysaccharide chain length by sodium dodecylsulfate-polyacrylamide gel electrophoresis. Eur. J. Biochem. **60:**239-246

**Jann, K. and B. Jann** 1984. Structure and biosynthesis of O-antigen, p. 138-186. *In* E. T. Rietschel (ed.), Handbook of Endotoxin, Vol. 1. Elsevier Scientific Publishers, Amsterdam.

**Jann, M.-L., B. Jann and K. Jann.** 1988. Structure and serological characteristic of the capsular K4 antigen of *Escherichia coli* O5:K4:H4, a fructose containing polysaccharide with a chondroitin bacbone. Eur J Biochem. **177:**117-124

**Jann, B. and K. Jann.** 1990. Structure and Biosynthesis of the capsular antigens of *Escherichia coli.* Curr. Top. Microbiol. **150:**19-42

**Jansson, P.-E., B. Lindberg, G. Bruse, A. A. Lindberg and R. Wollin.** 1977. Structural studies on the hexose region of the lipopolysaccharide from *Escherichia coli* C. Carbohydr. Res. **54:**261-267

**Jansson, P.-E., A. A. Lindberg, B. Lindberg and R. Wollin.** 1981. Structural studies on the hexose region of the core lipopolysaccharides from *Enterobacteriaceae.* Eur. J. Biochem. **115:**571-577

**Jiang, X. M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5:**695-713

**Johnson, E. M., B. Krauskopf and L. S. Baron.** 1966. Genetic analysis of the *ViA-his* chromosomal region in *Salmonella.* J. Bacteriol. **92:**1457-1463

**Kanegasaki, S. and K. Jann.** 1979. Demonstration by Membrane Reconstitution of a Butanol-Soluble Intermediate in the Biosynthesis of the O9 Antigen of *Escherichia coli.* Eur. J. Biochem. **95:**287-293

**Kanehisa, M.** 1984. Use of the statistical criteria for screening potential homologies in nucleic acid sequence. Nucleic Acids Res. **12:**203-213

**Kauffmann, F.** 1941. A typhoid variant and a new serological variation in the *Salmonella* group. J. Bacteriol. **41:**127-140

**Kauffmann, F. and R. Rohde.** 1961. Neue Befunde beim O-Formen-Wechsel der *Salmonella-species.* Acta. Pathol. Microbiol. Scand. **52:**211-216

**Kauffmann, F.** 1966. *Enterobacteriaceae.* Ejnar Munksgaard, Copenhagen

**Kenne, L. and B. Lindberg** 1983. Bacterial Polysaccharides, p. 287-363. *In* G. O. Aspinall (ed.), The Polysaccharides, 2. Harcourt Brace Jovanovich, New York.

**Kent, J. L. and M. J. Osborn.** 1968a. Haptenic O-antigen as a polymeric intermediate of in vivo synthesis of lipopolysaccharide by *Salmonella typhimurium.* Biochemistry. **7:**4419-4422

**Kent, J. L. and M. J. Osborn.** 1968b. Further Studies on Enzymatic Synthesis of O-Antigen in *Salmonella typhimurium.* Biochemistry. **7:**4409-4419

**Kent, J. L. and M. J. Osborn.** 1968c. Properties of the O-Specific Hapten Formed in Vivo by Mutant Strains of *Salmonella typhimurium.* Biochemistry. **7:**4396-4408

**Kido, N., M. Ohta, K.-I. Iida, T. Hasegawa, H. Ito, Y. Arakawa, T. Komatsu and N. Kato.** 1989. Partial deletion of the cloned *rfb* gene of *Escherichia coli* O9 results in synthesis of a new O-antigenic lipopolysaccharide. J. Bacteriol. **171:**3629-3633

**Kishi, K. and S. Iseki.** 1973a. Genetic analysis of the O antigens in *Salmonella.* I. Heredity of the blood group-active O antigens in *Salmonella* groups G, R and U. Jpn. J. Genet. **48:**89-97

**Kishi, K. and S. Iseki.** 1973b. Genetic analysis of the O-antigens in *Salmonella.* II. Heredity of the O antigens 4, 5 and 9 of *Salmonella.* Jpn. J. Genet. **48:**133-136

**Klein, P., M. Kanehisa and C. Delisi.** 1985. The detection and classification of membrane spanning proteins. Biochem. Biophys. Acta. **815:**468-476

**Klena, J. D., E. Pradel and C. A. Schnaitman**. 1992. Comparison of Lipopolysaccharide Biosynthesis Genes *rfaK*, *rfaL*, *rfaY*, and *rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium*. J. Bacteriol. **174:**4746-4752

**Kopmann, H. J. and K. Jann**. 1975. Biosynthesis of the O9 antigen of *Escherichia coli*. The polysaccharide component of *E. coli* O9:K29$^-$. Eur. J. Biochem. **60:**587-601

**Kroll, J. S. and E. R. Moxon**. 1990. Capsulation in distantly related strains of *Haemophilus influenzae* type b: Genetic drift and gene transfer at the capsulation locus. J. Bacteriol. **172:**1347-1379

**Kuo, T. T. and B. A. D. Stocker**. 1972. Mapping of *rfa* genes in *Salmonella typhimurium* by ES18 and P22 transduction and by conjugation. J. Bacteriol. **112:**48-57

**Kusecek, B., H. Wloch, A. Mercer, V. Vaisanen, G. Pluschke, T. Korhonen and M. Achtman**. 1984. Lipopolysaccharide capsule and fimbriae as virulence factors among O1, O7, O16, O18, O75, and K1, K5 or K100 *Escherichia coli*. Infect. Immun. **43:**368-379

**Kuzio, J. and A. M. Kropinski**. 1983. O-antigen conversion in *Pseudomonas aeruginosa* PAO1 by bacteriophage D3. J. Bacteriol. **155:**203-212

**Kyte, J. and R. F. Doolittle**. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105-132

**Le Minor, L. and M. Y. Popoff**. 1987. Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*. Int J Syst Bacteriol. **37:**465-468

**Lee, S. J., L. K. Romana and P. R. Reeves**. 1992. Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain. J. Gen. Microbiol. 1843-1855

**Leinonen, M.** 1985. Serological methods for the study of bacterial surface antigens., p. 179-206. *In* T. K. Korhonen, E. A. Dawes and P. H. Mäkelä (ed.), Enterobacterial surface antigens: methods for molecular characterisation., Elsevier Science Publishers, Amsterdam.

**Leive, L.** 1965. Release of lipopolysaccharide by EDTA treatment of *E. coli*. Biochem. Biophys. Res. Commun. **21:**290-296

**Leive, L.** 1974. The barrier function of the gram-negative envelope. Ann. N. Y. Acad. Sci. **235:**109-127

**Levinthal, M. and H. Nikaido.** 1969. Consequences of deletion mutants joining two operons of opposite polarity. J. Mol. Biol. **42:**511-520

**Lew, H. C., H. Nikaido and P. H. Mäkelä.** 1978. Biosynthesis of uridine diphosphate N-acetylmannosaminuronic acid in *rff* mutants of *Salmonella typhimurium*. J. Bacteriol. **136:**227-233

**Lindberg, A.** 1973. Bacteriophage receptors. Ann. Rev. Microbiol. **27:**205-241

**Lindberg, A. A., T. Holme, C. G. Hellerqvist and S. Svensson.** 1972. Studies of a LI-hapten isolated from cell-walls of the rough mutant *Salmonella typhimurium* 395 MR10. Acta. Pathol. Microbiol. Scand. Sect. **B80:**743-750

**Lindberg, A. A. and T. Holme.** 1969. Influence of O side chains on the attachment of the Felix O-I bacterio-phage to *Salmonella* bacteria. J. Bacteriol. **99:**513-519

**Liu, D., N. K. Verma, L. K. Romana and P. R. Reeves.** 1991. Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J. Bacteriol. **173:**4814-4819

**Luderitz, O., H. J. Risse, H. Schulte-Holthausen, J. L. Strominger, I. W. Sutherland and O. Westphal.** 1965. Biochemical studies of the smooth-rough mutation in *Salmonella minnesota*. J. Microbiol. **89:**343-354

**Lugtenberg, B., J. Meijers, R. Peters, P. v. d. Hoek and L. v. Alphen.** 1975. Electrophoretic resolution of the major outer membrane proteins of *Escherichia coli* K12 into four bands. FEBS Lett. **58:**254-258

**MacLachlan, P. R., S. K. Kadam and K. E. Sanderson.** 1991. Cloning, Characterization, and DNA sequence of the *rfaLK* region for lipopolysaccharide synthesis in *Salmonella typhimurium* LT2. J. Bacteriol. **173:**7151-7163

**Maloy, S., C. Ginsburgh, R. Simons and W. Num.** 1981. Transport of long and medium chain fatty acids by *Escherichia coli* K-12. J. Biol. Chem. **356:**3735-3724

**Maniatis, T., E. F. Fritsch and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratories, Cold Spring Harbour, N.Y.

**Markovitz, A.** 1977. Genetics and regulation of bacterial capsular polysaccharide synthesis and radiation sensitivity, p. 415-462. *In* I. Sutherland (ed.), Surface carbohydrate of the procaryotic cell, Academic press, London.

**Matsuhashi, S., M. Matsuhashi and J. L. Strominger**. 1966. Enzymatic synthesis of cytidine diphosphate 3, 6 dideoxyhexoses. I. Overall reactions. J. Biol. Chem. **241:**4267-4274

**Matsuhashi, S. and J. L. Strominger**. 1967. Enzymatic synthesis of cytidine diphosphate-3, 6-dideoxyhexoses. IV. Partial resolution of the enzymes required for the reduction of cytidine diphosphate-4-keto-6-deoxy-D-glucose and isolation of a new intermediate. J. Biol. Chem. **242:**3494-3500

**Mäkelä, P. H.** 1966. Genetic determination of the O antigens of *Salmonella* groups B (4,5,12) and C1 (6,7). J. Bacteriol. **91:**1115-1126

**Mäkelä, P. H. and O. Mäkelä**. 1966. *Salmonella* antigen 12$_2$: genetics of form variation. Ann. Med. Exp. Biol. Fenn. **44:**310-317

**Mäkelä, P. H., M. Jahkola and O. Lüderitz**. 1970. A new gene cluster *rfe* concerned with the biosynthesis of *Salmonella* lipopolysaccharide. J. Gen. Microbiol. **60:**91-106

**Mäkelä, P. H.** 1973. Glucosylation of lipopolysaccharide in *Salmonella*: mutants negative for O antigen factor 122. J. Bacteriol. **116:**847-856

**Mäkelä, P. H. and H. Mayer**. 1974. Enterobacterial common antigen. Bact. Rev. **40:**591-632

**Mäkelä, P. H., G. Schmidt, H. Mayer, H. Nikaido, H. I. Whang and E. Neter**. 1976. Enterobacterial common antigen in *rfb* deletion mutants of *Salmonella typhimurium*. J. Bacteriol. **127:**1141-1147

**Mäkelä, P. H. and B. A. D. Stocker** 1984. Genetics of lipopolysaccharide, p. 59-137. *In* E. T. Rietschel (ed.), Handbook of Endotoxin, Vol. I. Elsevier Science Publishers., Amsterdam.

**McGrath, B. C. and M. J. Osborn**. 1991a. Evidence for Energy-Dependent Transposition of Core Lipopolysaccharide Across the Inner Membrane of *Salmonella typhimurium*. J. Bacteriol. 3134-3137

**McGrath, B. C. and M. J. Osborn**. 1991b. Localization of the terminal steps of O antigen synthesis in *Salmonella typhimurium*. J. Bacteriol. **173:**649-654

**Meier-Dieter, U., R. Starman, K. Barr, H. Mayer and P. D. Rick.** 1990. Biosynthesis of Enterobacterial Common Antigen in *Escherichia coli*. J. Biol. Chem. **265:**13490-13497

**Meier-Dieter, U., K. Barr, R. Starman, L. Hatch and P. D. Rick.** 1992. Nucleotide Sequence of the *Escherichia coli rfe* Gene Involved in the Synthesis of Enterobacterial Common Antigen. J. Biol. Chem. **267:**746-753

**Melchoir, D. L. and A. Carruthers.** 1983. Studies of the relationship between bilayer water permeability and bilayer physical state. Biochemistry. **22:**5797-5807

**Milkman, R. and I. P. Crawford.** 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*. Science. **221:**378-380

**Milkman, R. and M. M. Bridges.** 1990. Molecular evolution of the *Escherichia coli* chromosome. III: Clonal Frames. Genetics. **126:**505-517

**Muhlradt, P. F., J. Menzel, J. R. Golecki and V. Speth.** 1973. Outer membrane of *Salmonella* sites of export of newly synthesized lipopolysaccharide on the bacterial surface. Eur. J. Biochem. **35:**471-481

**Mulford, C. A. and M. J. Osborn.** 1983. An intermediate step in translocation of lipopolysaccharide to the outer membrane of *Salmonella typhimurium*. Proc. Natl. Acad. Sci. USA. **80:**1159-1163

**Naide, Y., H. Nikaido, P. H. Mäkelä, R. G. Wilkinson and B. A. D. Stocker.** 1965. Semirough strains of *Salmonella.* Proc. Natl. Acad. Sci. USA. **53:**147-153

**Nelson, K., T. S. Whittam and R. K. Selander.** 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA. **88:**6667-6671

**Nelson, K., T. S. Whittam and R. K. Selander.** 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA. **88:**6667-6671

**Nikaido, H.** 1962. Studies on the biosynthesis of cell wall polysaccharide in mutant strains of *Salmonella*. Proc. Natl. Acad. Sci. USA. **48:**1542-1548

**Nikaido, H., M. Levinthal, K. Nikaido and K. Nakane**. 1967. Extended deletions in the histidine-rough-B region of *Salmonella* chromosome. Proc. Natl. Acad. Sci. USA. **57:**1825-1832

**Nikaido, K. and H. Nikaido**. 1971. Glucosylation of lipopolysaccharide in *Salmonella*: biosynthesis of O antigen factor $12_2$. II. Structure of the lipid intermediate. J. Biol. Chem. **246:**3912-3919

**Nikaido, H., K. Nikaido, T. Nakae and P. H. Mäkelä**. 1971. Glucosylation of lipopolysaccharide in *Salmonella*: biosynthesis of O antigen factor $12_2$.I. Overall reaction. J. Biol. Chem. **246:**3902-3911

**Nikaido, H.** 1973. Biosynthesis and assembly of lipopolysaccharide and the outer membrane layer of Gram-negative cell wall, p. 131-208. *In* L. Leive (ed.), Bacterial Membranes and Walls, 1. Dekker, Ney York.

**Nikaido, H.** 1976. Outer membrane of *Salmonella typhimurium*: transmembrane diffusion of some hydrophobic substances. Biochim. Biophys. Acta. **433:**118-132

**Nikaido, H., P. Bavoil and Y. Hirota**. 1977. Outer membranes of gram-negative bacteria. XV. Transmembrane diffusion rates in lipoprotein-deficient mutants of *Escherichia coli*. J. Bacteriol. **132:**1045-1047

**Nikaido, H. and T. Nakae**. 1979. The outer membrane of gram-negative bacteria. Adv. Microb. Physiol. **20:**163-250

**Nikaido, H. and M. Vaara** 1987. Outer membrane, p. 7-22. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular biology, 1. American Society for Microbiology, Washington D. C.

**Norrander, J., T. Kempe and J. Messing**. 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene. **26:**101-106

**Nyman, K., M. Plosila, L. Howden and P. H. Mäkelä**. 1979. Genetic determination of lipopolysaccharide: locus of O-specific unit polymerase in group E of *Salmonella*. Zbl. Bakt. Hyg., I. Abt. Orig. A. **243:**355-362

**O'Neill, M. C. and F. Chiafari.** 1989. *Escherichia coli* promoters-II. A spacing class-dependent promoter search protocol. J. Biol. Chem. **264:**5531-5534

**O'Neill, M. C.** 1989a. Consensus methods for finding and ranking DNA binding site-Application to *Escherichia coli* promoters. J. Mol. Biol. **207:**301-310

**O'Neill, M. C.** 1989b. *Escherichia coli* promoter-I. Consensus as it relates to spacing class, specificity, repeat structure, and three-dimensional organization. J. Biol. Chem. **264:**5522-5530

**O'Neill, M. C.** 1991. Training back-propagation neural networks to define and detect DNA-binding sites. Nucl. Acids Res. **19:**313-318

**Ochman, H. and A. C. Wilson** 1987. Evolutionary History of Enteric Bacteria, p. 1649-1654. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology, Vol II. American Society for Microbiology, Washington, DC.

**Orskov, I. and K. Nyman.** 1974. Genetic mapping of the antigenic determinants of two polysaccharide K antigens, K10 and K54, in *E. coli.* J. Bacteriol. **120:**43-51

**Orskov, I., F. Orskov, B. Jann and K. Jann.** 1977. Serology, chemistry and genetics of O and K antigens of *Escherichia coli.* Bacteriol. Rev. **41:**667-710

**Orskov, F., V. Sharma and I. Orskov.** 1984. Influence of growth temperature on the development of *Escherichia coli* polysaccharide K antigens. J. Gen. Microbiol. **130:**2681-2684

**Osborn, M. J. and I. M. Weiner.** 1968. Biosynthesis of a bacterial lipopolysaccharide. VI. Mechanism of incorporation of abequose into the O-antigen of *Salmonella typhimurium.* J. Biol. Chem. **243:**2631-2639

**Osborn, M. J. and R. Y. Tze-Yuen.** 1968. Biosynthesis of bacterial lipopolysaccharide. II. Enzymatic formation of the first intermediate in biosynthesis of the O-antigen of *Salmonella typhimurium.* J. Biol. Chem. **243:**5145-5152

**Osborn, M. J. and L. I. Rothfield** 1971. Biosynthesis of the core region of lipopolysaccharides, p. 331-350. *In* G. Weinbaum, S. Kadis and S. J. Ajil (ed.), Microbial Toxins, 4. Academic Press, New York.

**Osborn, M. J., J. E. Gander, E. Parisi and J. Carson**. 1972a. Mechanism of Assembly of the Outer Membrane of *Salmonella typhimurium*: Isolation and Characterization of Cytoplasmic and Outer Membrane. J. Biol. Chem. **247:**3962-3972

**Osborn, M. J., J. E. Gander and E. Parist**. 1972b. Mechanism of Assembly of the Outer Membrane of *Salmonella typhimurium*: Site of Lipopolysaccharide Synthesis. J. Biol. Chem. **247:**3973-3986

**Palva, T. and P. H. Mäkelä**. 1980. Lipopolysaccharide heterogeneity in *Salmonella typhimurium* analysed by sodium dodecylsulfate-polyacrylamide gel electrophoresis. Eur. J. Biochem. **107:**137-143

**Parker, C. T., E. Pradel and C. A. Schnaitman**. 1992. Identification and sequences of the lipopolysaccharide core biosynthetic genes *rfaQ, rfaP, and rfaG* of *Escherichia coli* K-12. J Bacteriol. **174:**930-934

**Pearson, W. R**. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Meth. Enzymol. **183:**63-98

**Pegues, J. C., L. S. Chen, A. W. Gordon, L. Ding and W. G. Coleman**. 1990. Cloning, expression, and characterization of the *Escherichia coli* K-12 *rfaD* gene. J. Bacteriol. **172:**4652-4650

**Peterson, A. A. and E. J. McGroarty**. 1985. High-molecular-weight components in lipopolysaccharide of *Salmonella typhimurium, Salmonella minnesota, and Escherichia coli*. J. Bacteriol. **162:**738-745

**Plosila, M. and P. H. Mäkelä**. 1972. Mapping of a gene *oafC* determining antigen 1 in *Salmonella* of group E. Scand. J. Clin. Lab. Invest. **29 (Suppl. 122):**55

**Pradel, E., C. T. Parker and Schnaitman**. 1992. Structures of the *rfaB*, *rfaI*, *rfaJ*, and *rfaS* Genes of *Escherichia coli* K-12 and their Roles in Assembly of the Lipopolysaccharide Core. J. Bacteriol. **174:**4736-4745

**Raetz, C. R. H**. 1978. Enzymology, genetics and regulation of membrane phospholipid synthesis in *Escherichia coli*. Microbiol. Rev. **42:**614-659

**Raetz, C. R. H.** 1982. Genetic control of phospholipid bilayer assembly, p. 435-477. *In* J. N. Hawthorne and G. B. Ansell (ed.), Phospholipids, Elsevier Biomedical Press, Amsterdam.

**Raetz, C. R. H.** 1986. Molecular Genetics of Membrane Phospholipid Synthesis. Ann. Rev. Genet. **20:**253-295

**Raetz, C. R. H.** 1990. Biochemistry of endotoxins. Annu. Rev. Biochem. **59:**129-170

**Reed, J. W., M. Capage and G. C. Walker.** 1991. *Rhizobium meliloti exoG* and *exoJ* Mutations Affect the ExoX-ExoY System for Modulation of Exopolysaccharide Production. J. Bacteriol. **173:**3776-3788

**Rehemtulla, A., S. K. Kadam and K. E. Sanderson.** 1986. Cloning and analysis of the *sfrB* (sex factor repression) gene of *Escherichia coli* K-12. J. Bacteriol. **166:**651-657

**Reisner, A. H. and C. A. Bucholtz.** 1986. The MTX package of computer programs for comparison of sequences of nucleotides and amino acid residues. Nucl Acids Res. **14:**233-238

**Rivera, M., L. E. Bryan, R. E. W. Hancock and E. J. McGroarty.** 1988. Heterogeneity of lipopolysaccharides from *Pseudomonas aeruginosa*: analysis of lipopolysaccharide chain length. J. Bacteriol. **170:**512-521

**Roantree, R. J., T.-T. Kuo and D. G. Macphee.** 1977. The Effect of Defined Lipopolysaccharide Core Defects upon Antibiotic Resistances of *Salmonella typhimurium*. J. Gen. Micro. **103:**223-234

**Robbins, P. W. and T. Uchida.** 1965. Chemical and Macromolecular Structure of O-Antigens from *Salmonella anatum* Strains Carrying Mutants of Bacteriophage $\varepsilon^{15}$. J. Biol. Chem. **240:**375-383

**Robbins, P. W., J. M. Keller, A. Wright and R. L. Bernstein.** 1965. Enzymatic and Kinetic Studies on the Mechanism of O-Antigen Conversion by Bacteriophage $\varepsilon^{15}$. J. Biol. Chem. **240:**384-390

**Robbins, P. W. and A. Wright** 1971. Biosynthesis of O-antigen, p. 351-368. *In* G. Weinbaum, S. Kadis and S. J. Ajl (ed.), Microbial Toxins, 4. Academic Press, New York.

**Roncero, C. and M. Casadaban.** 1992. Genetic analysis of the genes involved in synthesis of the lipopolysaccharide core in *Escherichia coli* K-12: three operons in the *rfa* locus. J Bacteriol. **174:**3250-3260

**Sanderson, K. E., J. V. Wyngaarden, O. Luderitz and B. A. D. Stocker**. 1974. Rough mutants of *Salmonella typhimurium* with defects in the heptose region of the lipopolysaccharide core. Can. J. Microbiol. **20:**1127-1134

**Sanderson, K. E. and J. R. Roth**. 1983. Linkage map of *Salmonella typhimurium*, edition VI. Microbiol. Rev. **47:**410-453

**Sanderson, K. E. and J. R. Roth**. 1988. Linkage map of *Salmonella typhimurium*. Microbiol. Rev. **52:**485-532

**Sanger, F., S. Nicklen and A. R. Coulson**. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. **74:**5463-5467

**Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith and B. A. Roe**. 1980. Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143:**161-178

**Schindler, M. and M. J. Osborn**. 1979. Interaction of divalent cations and polymyxin B with lipopolysaccharide. Biochemistry. **18:**4425-4430

**Schnaitman, C. A., C. T. Parker, J. D. Klena, E. L. Pradel, N. B. Pearson, K. E. Sanderson and P. R. MacLachlan**. 1991. Physical maps of the *rfa* loci of *Escherichia coli* K-12 and *Salmonella typhimurium*. J. Bacteriol. **173:**7410-7411

**Shepherd, J. C. W.** 1981. Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and if possible evolutionary justification. Proc. Natl. Acad. Sci. U.S.A. **78:**1596-1600

**Simon, R., J. Quandt and W. Klipp**. 1989. New derivatives of transposon Tn5 suitable for mobilization of replicons, generation of operon fusions and induction of genes in Gram-negative bacteria. Gene. **80:**161-169

**Smith, T. F. and M. S. Waterman**. 1981. Identification of the common molecular sequences. J Mol Biol. **147:**195-197

**Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids Res. **12:**505-519

**Staden, R.** 1986. The current status and portability of our sequence dandling software. Nucl. Acids Res. **14:**217-231

**Stevenson, G. S., S. J. Lee, L. K. Romana and P. R. Reeves.** 1991. The *cps* gene cluster of *Salmonella* strain LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the *rfb* gene cluster. Mol. Gen. Genet. **227:**173-180

**Sueoka, N.** 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA. **85:**2653-2657

**Sutherland, I. W., O. Luderitz and O. Westphal.** 1965. Studies on the structure of lipopolysaccharides of *Salmonella minnesota* and *Salmonella* R-strains. Biochem. J. **96:**439-448

**Takeshita, M. and P. H. Mäkelä.** 1971. Glucosylation of lipopolysaccharide in *Salmonella*: biosynthesis of O-antigen factor $12_2$. III. The presence of $12_2$ determinants in haptenic polysaccharides. J. Biol. Chem. **246:**3920-3927

**Trisler, P. and S. Gottesman.** 1984. *Lon* transcriptional regulation of genes necessary for capsular polysaccharide synthesis in *Escherichia coli* K12. J. Bacteriol. **160:**184-191

**Tsang, R. S. W., S. Schlecht, S. Aleksic, K. H. Chan and P. Y. Chau.** 1991. Lack of the *a*-1,2-linked N-acetyl-D-glucosamine epitope in the outer core structures of lipopolysaccharides from certain O serogroups and subspecies of *Salmonella enterica*. Res. Microbiol. **142:**521-533

**Uchida, T., T. Matsumoto and T. Sasaki.** 1974. Mutants of group D1 *Salmonella* carrying the somatic antigen of group A organisms: isolation and serological characterization. J. Bacteriol. **117:**8-12

**Uetake, H., T. Nakagawa and T. Akiba.** 1955. The relationship of bacteriophage to antigenic changes in group E *Salmonellas*. J. Bacteriol. **69:**571-579

**Uetake, H., E. Luria and J. W. Burrous.** 1958. Conversion of somatic antigens in *Salmonella* by phage infection leading to lysis or lysogeny. Virology. **5:**68-91

**Unger, F. M.** 1981. The chemistry and biological significance of 3-deoxy-D-manno-2-octulosonic acid (KDO). Adv Carbohydr Chem. **38:**323-388

**Verma, N. K., N. B. Quigley and P. R. Reeves.** 1988. O-antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. J. Bacteriol. **170:**103-107

**Verma, V. and P. R. Reeves**. 1989. Identification and sequence of *rfb S* and *rfb E* , which determine antigenic specificity of group A and group D *Salmonella*. J. Bacteriol. **171:**5694-5701

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

**Vieira, J. and J. Messing**. 1987. Production of Single-Stranded Plasmid DNA. Meth. Enzymol. **153:**3-11

**Wallace, B. A., M. Cascio and D. L. Mielke**. 1986. Evolution of methods for the prediction of membrane protein secondary structure. Pro. Nalt. Acad. Sci. U.S.A. **83:**9423-9427

**Wang, L., L. K. Romana and P. R. Reeves**. 1992. Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O antigen and the genetic basis of the major polymorphism. Genetics. **130:**429-443

**Weiner, I. M., T. Higuchi, L. Rothfield, M. Saltmarsh-Andrew, M. J. Osborn and B. L. Horecker**. 1965. Biosynthesis of bacterial lipopolysaccharide. V. Lipid-linked intermediates in the biosynthesis of the O-antigen groups of *Salmonella typhimurium*. Pro. Natl. Acad. Sci. USA. **54:**228-235

**Weisgerber, C., B. Jann and K. Jann**. 1984. Biosynthesis of the O9 antigen of *Escherichia coli*. Core structure of *rfe* mutant as indication of assembly mechanism. Eur. J. Biochem. **140:**553-556

**Wicken, A. J. and K. W. Knox**. 1980. Bacterial cell surface amphiphiles. Biochem Biophys. **604:**1-26

**Wilbur, W. J. and D. J. Lipman**. 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA. **80:**726-730

**Wilkinson, R. G., P. Gemski, J. Stocker and B. A. D. Stocker**. 1972. Non-smooth Mutants of *Salmonella typhimurium*: Differentiation by Phage Sensitive and Genetic Mapping. J. Gen. Microbiol. **70:**527-554

**Wright, A., M. Dankert, P. Fennessey and P. W. Roobins**. 1967. Characterization of a polyisoprenoid compound functional in O-antigen biosynthesis. Proc. Natl. Acad. Sci. U.S.A. **57:**1798-1803

**Wright, A. and S. Kanegasaki**. 1971. Molecular aspects of lipopolysaccharides. Physiol. Rev. **51:**748-784

**Wollin, R., B. A. D. Stocker and A. A. Lindberg**. 1987. Lysogenic conversion of *Salmonella typhimurium* bacteriophages A3 and A4 consists of O-acetylation of rhamnose of the repeating units of the O-antigenic polysaccharide chain. J. Bact. **169**:1003-1009

**Wyk, P. and P. R. Reeves**. 1989. Identification and sequence of the gene for abequose synthase which confers antigenic specificity on group B Salmonellae: homology with galactose epimerase. J. Bacteriol. **171:**5687-5693

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

**Yuasa, R., M. Levinthal and H. Nikaido**. 1969. Biosynthesis of bacterial wall lipopolysaccharide in mutants of *Salmonella*. J. Bacteriol. **100:**433-444

**Zinder, N. D**. 1953. Infective heredity in bacteria. Cold Spring Harbor Symposia on Quantitative Biology. **18:**261-269